INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE INVESTIGACIÓN EN CIENCIA APLICADA Y TECNOLOGÍA AVANZADA

Un Acercamiento Cognitivo y Epistemológico a la Didáctica del Concepto de Variable Aleatoria

Tesis que para obtener el grado de
Maestra en Ciencias en Matemática Educativa

PRESENTA:

Blanca R. Ruiz Hernández

DIRECTOR DE TESIS:

Dr. José Armando Albert Huerta

CO-DIRECTOR DE TESIS:

Javier Lezama Andalón

Septiembre de 2006

SIP-14

INSTITUTO POLITECNICO NACIONAL
SECRETARIA DE INVESTIGACION Y POSGRADO

ACTA DE REVISION DE TESIS

En la Ciudad de México siendo las 11:00 horas del día 17 del mes de agosto del 2006 se reunieron los miembros de la Comisión Revisora de Tesis designada por el Colegio de Profesores de Estudios de Posgrado e Investigación de CICATA LEGARIA para examinar la tesis de grado titulada:

"Un acercamiento cognitivo y epistemológico a la didáctica del concepto de variable aleatoria"

Presentada por el alumno:

Ruiz	Hernández	Blanca Rosa
Apellido paterno	materno	nombre(s)

Con registro: A 0 1 0 6 8 7

aspirante al grado de:

Maestro en Ciencias en Matemática Educativa

Después de intercambiar opiniones los miembros de la Comisión manifestaron ***SU APROBACION DE LA TESIS***, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISION REVISORA

Director de tesis

Dr. José Armando Albert Huerta

Codirector

Dr. Francisco Javier Lezama Andalón

Dr. Apolo Castañeda Alonso

CICATA - IPN
Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada del Instituto Politécnico Nacional

Dr. Juan Ernesto Colunga Cavazos

Dra. Gisela Montiel Espinosa

EL PRESIDENTE DEL COLEGIO

Dr. José Antonio Irán Díaz Góngora

INSTITUTO POLITECNICO NACIONAL

SECRETARIA DE INVESTIGACION Y POSGRADO

CARTA DE CESION DE DERECHOS

En la ciudad de México, D.F. el día 16 del mes agosto del año 2006, el (la) que suscribe Blanca Rosa Ruiz Hernández alumno (a) del Programa de Maestría en Ciencias en Matemática Educativa con número de registro A010687 adscrito al Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección del Dr. J. Armando Albert Huerta y cede los derechos del trabajo titulado "Un acercamiento cognitivo y epistemológico a la didáctica del concepto de variable aleatoria" al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección bruiz@itesm.mx Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Blanca Rosa Ruiz Hernández

A
MI MADRE

porque su vivacidad alcanza para mantenernos vivos a todos

ÍNDICE

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

Un acercamiento cognitivo y epistemológico a la didáctica del concepto de variable aleatoria

Resumen

El concepto de variable aleatoria es uno de los conceptos base de la teoría de probabilidad y de la estadística inferencial. Desde una perspectiva matemática permite, junto con las funciones de distribución, el manejo de la probabilidad y los eventos aleatorios dentro del contexto de los números reales, desde una perspectiva de modelación permite acotar el contexto y trabajar específicamente con los elementos que son manipulables y de interés dentro de un problema que involucre algún fenómeno aleatorio.

En su trabajo acerca de las ideas estocásticas fundamentales, Heitele (1975) propuso este concepto como una idea fundamental dentro de la enseñanza escolar porque debe desarrollarse a través de un currículo en espiral desde las prefiguraciones en etapas tempranas del niño hasta estadios muy desarrollados en el profesionista. Sin embargo han sido pocas las investigaciones enfocadas explícitamente sobre la didáctica de este concepto.

Esta investigación pretende establecer algunas bases sobre las que se pueda apoyar el estudio de la didáctica de la variable aleatoria en el nivel universitario haciendo uso de la Ingeniería didáctica. Se enfoca al análisis preliminar de esa metodología de investigación, en particular a los análisis cognitivo y epistemológico y se sustenta en la Teoría de situaciones didácticas. Desde la vertiente cognitiva, se trabajó en una entrevista clínica a dos estudiantes recién ingresadas al nivel universitario y desde la perspectiva epistemológica se enfocó al análisis desde la disciplina que le da sustento al concepto (la probabilidad) y al análisis del contexto de su emergencia histórica.

El estudio permitió observar la complejidad epistémica del concepto de variable aleatoria. Las estudiantes vincularon la idea de variable aleatoria con conceptos probabilisticos complejos, la relacionaron con un proceso de modelación y la relacionaron con herramientas determinísticas. La complejidad epistémica de la variable aleatoria se ve reflejada en las dificultades y los aciertos de las estudiantes al resolver el problema, pero también en el contexto y los problemas históricos que lo hicieron surgir como concepto matemático. El análisis cognitivo proporcionó elementos y vertientes sobre las cuales se puede profundizar en el análisis epistemológico.

A cognitive and epistemological approximation to the didactics of the concept of random variable

Abstract

The random variable concept is one of the based concepts of the probability theory and the inferential statistic. From a mathematical point of view, random variable and distribution functions introduce probability and random events in real numbers context. From a model point of view, random variable delimits the context and allows relating random real phenomena to the mathematical context of probability distributions.

In his study about fundamental stochastical ideas, Heitele (1975) included the random variable as a fundamental one in school education because it must develop in the curriculum spiral from pre-figurations in early stages of children to developed stages in university students. However few researchers have explicitly focused on the didactics of this concept.

This research seeks to find elements to support the didactics of the random variable study in university, helping of the didactic engineering. The focus is on preliminary analysis of this research methodology, particularly, on cognitive and epistemological analysis, and It is based on theory of didactical situations. From the cognitive aspect, a clinical interview was carried out with two university students. From the epistemological aspect, a disciplinal and historical analysis was done.

The research allows observing the epistemological complexity of the random variable concept. Students linked the idea of a random variable to complex probabilistic concepts, related it to a mathematical modeling process and connected it to other stochastic and non stochastic mathematical tools. It was reflected in the difficulties and success that the students faced when solving the problem. But also in the historical context and problems that gave rise to the mathematical concept. The cognitive analysis provided elements and aspect which can be studied in depth in the epistemological analysis.

Porque una de las maravillas de enseñar,
es que, preguntas que nunca te hiciste,
te tienes que contestar.

Introducción

En los últimos años la didáctica de la probabilidad y la estadística ha tenido un fuerte realce e impulso. Muchos profesores que imparten esas materias se han preocupado por fundamentar su trabajo y por tener mejores resultados en la enseñanza de estos conocimientos. Al mismo tiempo, estadísticos, psicólogos e investigadores en didáctica de la matemática se han percatado de la ausencia de investigaciones que respalden esa labor. Este trabajo se une a ese impulso por fomentar la enseñanza de la probabilidad y estadística en el nivel universitario desde la perspectiva de la didáctica de la matemática.

En particular, en esta investigación se aborda una de las ideas fundamentales de los cursos de Probabilidad y Estadística en las instituciones de enseñanza universitaria: la variable aleatoria, que se apoya en muchos otros conceptos matemáticos y sirve a su vez de soporte a gran parte de los temas en probabilidad e inferencia estadística. En una primera aproximación, el concepto aparenta una simplicidad que se pierde apenas se introduce alguna breve formalidad o es necesario manipularlo en algún otro concepto. A la vez que es un concepto muy abstracto, también es el encargado de extraer del contexto del problema los elementos que permiten trabajar la situación en un contexto matemático. Como se pudo constatar en esta investigación, para los estudiantes universitarios no es tan sencillo asimilarlo en toda su dimensión al mismo tiempo que sí pueden trabajar con él y manipularlo desde una visión intuitiva.

Muchos estudiantes no pueden identificar la variable aleatoria involucrada en una situación problemática, lo que les ocasiona fuertes obstáculos al abordar su resolución. Otros también tienen una comprensión incompleta de algunos conceptos vinculados con su definición, o bien no establecen adecuadamente las relaciones entre los mismos. Todo ello repercutirá negativamente en el estudio de otros conceptos que el estudiante deberá abordar posteriormente, y que se basan en la variable aleatoria, como son las distribuciones de probabilidad (íntimamente ligadas a la noción misma de variable aleatoria), el teorema central del límite, el muestreo, la correlación y asociación y la inferencia estadística.

La motivación para profundizar sobre la didáctica de la variable aleatoria esta fundamentada en estas razones. Los objetivos se enmarcan en la teoría de situaciones didácticas y se apoyan en las metodologías de ingeniería didáctica y en la exploración cognitiva denominada entrevista clínica. Esta investigación se enfoca al análisis preliminar de la ingeniería didáctica y dentro de ella únicamente a las vertientes cognitiva y epistemológica con la intención de identificar elementos que proporcionen sustento para el diseño de una situación didáctica. Desde la componente cognitiva se plantea la realización de una exploración cognitiva y en la dimensión epistemológica se profundiza en la naturaleza del concepto de variable aleatoria a través del análisis del concepto en la disciplina que lo formaliza y lo justifica, la estadística, y del estudio de su emergencia histórica.

Para facilitar la ubicación de los capítulos, la tesis se dividió en tres partes. En la primera de ellas se describen los preliminares de la investigación. La segunda constituye propiamente el análisis cognitivo y la tercera se centra en el análisis epistemológico.

Los capítulos uno y dos constituyen los preliminares de la investigación y forman la primera parte de este reporte. En ellos se justifica el interés del trabajo, se analizan los antecedentes de investigación y se concluye con la pertinencia de continuar en esta línea. En el segundo capítulo, se presenta con detalle el marco teórico que fundamentará, no sólo el trabajo reportado en esta memoria, sino la línea de investigación alrededor de la didáctica de la variable aleatoria que sustenta proyectos que ya se tienen en pie.

La tercera parte de este trabajo está formada por los capítulos tres, cuatro y cinco. Esta parte está dedicada al análisis cognitivo visto a través de una entrevista clínica. En ella se describe el diseño de una actividad problemática orientada a explorar las concepciones de los estudiantes sobre la variable aleatoria y conceptos relacionados, así como sus estrategias en la modelación de la situación planteada y su resolución. Se reporta un análisis a priori detallado, formulando hipótesis iniciales y definiendo los objetos de observación, las variables de observación asociadas y los instrumentos para recoger información sobre las mismas. Finalmente, se presentan los resultados de la observación y entrevista, respecto a cada una de las variables del diseño y las conclusiones obtenidas.

La cuarta parte de este reporte la constituyen los capítulos seis, siete y ocho. En ellos se incursiona en la dimensión epistemológica del concepto, tanto desde la profundización sobre su aspecto teórico disciplinar como desde la historia. Esta parte se finaliza con las conclusiones de este análisis.

El reporte de esta investigación concluye con algunas ideas para investigaciones futuras, a manera de una conclusión general, y se completa con las referencias, la actividad diseñada, su solución y la transcripción de la entrevista clínica.

Este trabajo significa un punto de partida de una línea de investigación que continuará con la profundización del análisis preliminar de la ingeniería de situaciones didácticas y posteriormente con el diseño de una situación didáctica que permita la exploración de cómo se apropian los estudiantes de la noción de variable aleatoria en situación escolar a nivel universitario. Creemos que la presente investigación ha permitido poner a punto una metodología de trabajo para continuar con el estudio de la comprensión de la variable aleatoria y también refinar nuestras hipótesis iniciales sobre sus dificultades y concepciones.

Parte I

Preliminares

Capítulo 1

El problema de investigación

La educación en probabilidad y estadística es una demanda cada vez más urgente de nuestras sociedades modernas. Y en respuesta a ello, tanto profesores como investigadores han comenzado a abordar esta compleja problemática de manera científica. Este interés está justificado por la demanda social de la estadística puesto que es necesario tanto que los especialistas en el área produzcan y analicen estadísticas, como que profesionistas y ciudadanos sepan interpretarlas y tomar decisiones basándose en ellas. Pero también está vinculado con el rápido desarrollo de la probabilidad y estadística como ciencias, las dificultades intrínsecas de las asignaturas en sí mismas y su vinculación tan cercana, y en ocasiones tan específica, con otras profesiones. Así mismo dentro del sistema educativo estamos viviendo cambios, como la incorporación de nuevas estrategias de aprendizaje y usos de tecnología, que requieren una revisión cuidadosa de las modificaciones de los contenidos que con ellos se ocasiona.

Dentro de este contexto, en este capítulo situamos nuestro trabajo sobre el aprendizaje de la variable aleatoria como una idea fundamental dentro de la enseñanza de la probabilidad y la estadística sobre la que resulta pertinente detenernos a analizar. Revisaremos algunas investigaciones relativas al tema y precisamos el problema específico que se aborda en este trabajo.

1.1 Demanda social de educación en estadística

El acelerado desarrollo de la tecnología, de la ciencia y de los medios de comunicación en las sociedades actuales ha exigido la incorporación a la educación formal de una de las áreas matemáticas de más actualidad: probabilidad y estadística. Los gobiernos y empresas exigen de la estadística cifras más claras y fiables que permitan tomar decisiones más acertadas en el orden de lo económico, social y político, para ello es necesaria la cultura estadística básica del ciudadano que ha de colaborar en la recolección de datos (Batanero, 2002), pero que también será usuario de esos datos para establecer criterios al participar en las decisiones que toma su sociedad.

La UNESCO y otros organismos internacionales acordaron que es prioritaria la promoción de la formación estadística como clave para el crecimiento, particularmente, en los países en desarrollo. Desde 1885 se han fundado diversos organismos que promueven la estadística como el Instituto Internacional de Estadística (ISI) que posteriormente creó el Comité de Educación, en 1948. Este Comité impulsó los Centros internacionales de Educación estadística (ISEC) en Calcuta y Beirut y las International Conferences on Teaching of Statistics (ICOTS), iniciadas en 1982. En 1991 el ISI fundó la International Association for Statistical Education (IASE) para encargarse específicamente de impulsar las acciones educativas en estadística.

Por otro lado, el Comité Internacional para la Instrucción en Matemática inició los Intenational Congresses of Matemátics Education (ICME) a partir de 1973, donde, entre otros temas, se organizan grupos de trabajo en educación estadística. Tanto en PME (Psicología de la Educación Matemática), como en CERME (Congreso Europeo de Educación Matemática), CIBEM (Congreso Iberoamericano de Educación Matemática) y RELME (Congreso Latinoamericano de Educación Matemática) se están incluyendo grupos temáticos de educación estadística en los últimos años. Todas estas instancias dan cuenta de la importancia de la educación en probabilidad y estadística y de la preocupación de la sociedad por mejorarla.

1.2 Rezago de las instituciones educativas

Hasta ahora, el desarrollo social ha sido más rápido que la capacidad de las instituciones educativas para responder a la demanda de educación estadística. Particularmente en México, los estudiantes llegan al nivel universitario con serias deficiencias en el ámbito de probabilidad y estadística y de desarrollo del pensamiento estocástico, como lo manifiestan las academias de estadística del sistema ITESM (Instituto Tecnológico de Estudios Superiores de Monterrey), tomado en cuenta el desempeño escolar, los índices de reprobación y al rápido olvido de los contenidos estudiados en sus alumnos.

Sin embargo este problema no es propio de un nivel educativo. Batanero (2002) expresa que el mal uso de la estadística por investigadores de diversas áreas de la ciencia tiene su raíz en que la incorporación de la estadística desde la educación básica no es todavía un hecho. Esto se puede observar en los programas de educación matemática mexicanos del nivel básico, en donde, a pesar de que en los programas oficiales están especificados contenidos de probabilidad y estadística, suele ser la última unidad que, por falta de tiempo e incluso por dificultades del profesor en el tema, no son estudiados o lo son superficialmente (Alatorre, 1998 y Alquicira, 1998). En el nivel medio superior actualmente no hay una política educativa uniforme para todo el país respecto a la enseñanza de la Probabilidad y Estadística (Secretaría de Educación Pública, 2003), a pesar de que la misma Secretaria de Educción Pública (SEP) de México lo maneja como un tópico obligatorio desde 1982. De esta forma, en muchos planes de estudio de bachillerato no es incluida como asignatura, en otros es una asignatura optativa y en otros más es incluida sólo como tópico en alguna otra asignatura.

Con respecto al último punto, hay esfuerzos que pueden rendir frutos si son acompañados de medidas que hagan posible su incorporación y su constante revisión. La SEP de México está trabajando en la homogeneización de los planes de estudio del nivel medio superior, en donde se pretende incorporar la materia de Probabilidad y Estadística (Secretaría de Educación Pública, 2003) y en el Sistema ITESM se está trabajando, desde nuestro grupo de profesores de Probabilidad y Estadística de la misma institución, en el desarrollo de un programa de estudio sobre tópicos de probabilidad y estadística para sus bachilleratos. Por supuesto este tipo de medidas por sí mismas son insuficientes, como nos lo ha enseñado la experiencia en el nivel básico, pero la necesidad de revisión de los programas de estudio es real, sobre todo si no lo han sido desde hace algunos años.

En el nivel profesional, hay también una necesidad patente de revisión de los contenidos que se enseñan en la materia de probabilidad y estadística, puesto que hace mucho que la asignatura en genérico no satisface la demanda de las distintas ingenierías y licenciaturas, quienes exigen una orientación hacia sus áreas de trabajo. Actualmente en el Sistema ITESM la asignatura está diversificada en Estadística para Ciencias Sociales, Probabilidad y Estadística para Ingeniería, Estadística para Economistas, Estadística para Ciencias Biológicas y hay intentos encaminados a un programa de estudio sobre Estadística para Psicología.

Aunado a ello, tomemos en cuenta que ha habido un aumento notable en el uso de ideas estadísticas en diferentes disciplinas (Cox, 1997) y que la estadística es una ciencia útil en otras no sólo como técnica auxiliar de las mismas, sino también dentro de su investigación y su crecimiento. A su vez la estadística se alimenta de estas áreas y crece con ellas. Todo ello, junto con el impulso que está teniendo la computación, ha hecho que sea una ciencia que tiene un rápido y diverso desarrollo.

La adecuación de los contenidos de Probabilidad y Estadística a los distintos contextos y necesidades profesionales, la actualización permanente de los profesores que debe haber en esta ciencia en continuo cambio, así como las dificultades a las que tanto profesores como estudiantes universitarios se enfrentan al tratar de cubrir contenidos propios del nivel y subsanar deficiencias de niveles anteriores (como lo señala Batanero, 2002), dificulta que las instituciones educativas respondan adecuadamente a las demandas sociales de formación en Probabilidad y Estadística.

1.3 Problemas con la enseñanza y aprendizaje de la probabilidad y estadística

Existen también dificultades propias de la enseñanza y el aprendizaje de la probabilidad y estadística, que contradicen la falsa creencia de que el discurso escolar es *transparente*. Esto significa que no es fácil aprender una serie de contenidos ya organizados lógicamente y de menor a mayor grado de dificultad, como lo suelen presentar los textos.

El problema no sólo radica en la forma en cómo se les presenten los contenidos a los estudiantes. Muy diversas investigaciones han probado la existencia de dificultades intrínsecas a la naturaleza de los conocimientos referidos a la probabilidad y estadística, que se recogen, por ejemplo, en las actas de los seis Congresos ICOTS, y en las revistas Journal of Statistics Education y Statistics Education Research Journal, así como en revistas y congresos de didáctica de la matemática. Aunque algunos de estos problemas también pudieran ser producto de una enseñanza de la probabilidad y estadística no adecuada. Así, por ejemplo, en México la enseñanza de la estadística universitaria tiene una fuerte tendencia a prescindir de los fenómenos aleatorios y de «algebraizar» los resultados de la estadística, con la consecuencia, entre otras, que se fomenta muy poco el desarrollo del pensamiento estocástico.

Junto con estos problemas están muchos otros relativos a la elección de las mejores estrategias de aprendizaje, a los contenidos mismos y a las dificultades que se presentan ante las nuevas tecnologías. Es por eso que se hace imprescindible abordar la didáctica de la probabilidad y estadística desde una perspectiva científica.

1.4 El objeto de estudio: La variable aleatoria

La variable aleatoria es una de las ideas fundamentales del curso de probabilidad y estadística ya que constituye un pilar para el desarrollo de habilidades de modelación en los estudiantes, así como una base para el desarrollo de la teoría relativa a distribuciones de probabilidad y el teorema central del límite, entre otros temas fundamentales para la Estadística. No es un concepto sencillo, puesto que hemos tenido oportunidad de observar en nuestras clases que los estudiantes tienen especiales dificultades en el aprendizaje de este concepto y esto acarrea dificultades en el aprendizaje de otros. En una primera aproximación, el concepto aparenta una simplicidad que se pierde apenas se introduce alguna leve formalidad o se manipula como objeto dentro de otros conceptos matemáticos. En la variable aleatoria convergen muchos conceptos estadísticos y probabilísticos provenientes de los niveles educativos básico y medio superior, a la vez que está relacionada con conceptos abstractos propios de niveles superiores. Por tanto, el interés curricular de la variable aleatoria ya se ve manifiesto desde el nivel medio superior puesto que esta noción se puede considerar un indicador del aprendizaje de elementos revisados por el estudiante en cursos previos al universitario tales como experimento aleatorio, espacio muestral, evento, probabilidad y distribuciones de frecuencias, entre otros.

La variable aleatoria como idea central en la enseñanza estocástica fue propuesta por Heitele (1975) por primera vez y después ha sido retomada por diversos autores. Heitele se basa en la concepción de 'Idea fundamental' de Bruner (1960) como criterio para enunciar cuáles serían las ideas fundamentales dentro de la estocástica en un currículum escolar. Según Bruner, las ideas fundamentales son útiles desde la perspectiva curricular porque permiten definir cuáles son los tópicos a cubrir en cada grado escolar y con qué profundidad y vincularlos entre ellos y con otras ramas de las matemáticas, de modo que se genere una red de conocimientos que no sólo propicie el aprendizaje formal de los conceptos sino también que eduque la intuición y desarrolle conexiones significantes con la realidad. Para Heitele una idea fundamental puede ser enseñada en cualquier nivel escolar de manera exitosa porque puede ser tratada bajo diferentes niveles cognitivos y por lo tanto, crecer con el currículo y proporcionar un modelo explicativo en cada nivel de su desarrollo dentro del currículo. Su lógica está de acuerdo con un currículo en espiral en el que en cada etapa se desarrollen modelos explicativos sobre una idea fundamental que difieren en los niveles cognitivos, en su forma lingüística y en sus niveles de profundización, pero no en su estructura.

Heitele pone como ejemplo de este proceso curricular el desarrollo cognitivo de la idea de variable aleatoria. La primera aproximación es un modelo explicativo burdo en donde el niño observa repeticiones del fenómeno e interpreta que es lo que ocurrirá más frecuentemente, a través de actividades de juego en donde no se involucra una instrucción formal ni analítica. Una etapa intermedia enfatizaría en un modelo más cuantitativo en donde se enumeren los eventos posibles y el uso de la probabilidad teórica. Un modelo más elaborado daría cuenta de la identificación e interpretación de la variable aleatoria, el análisis de la situación completa y la observación de parámetros que definen el comportamiento del fenómeno. Heitele hace énfasis en que lo importante de este proceso es la constancia de la estructura del modelo explicativo a lo largo de todo el currículo escolar que permite la continuación hacia etapas más elaboradas.

Así, desde ese marco de referencia de Bruner, Heitele (1979) propone una lista de diez ideas fundamentales, dentro de las que considera, como octavo punto, a la variable aleatoria. Él manifiesta que esta noción, y todo el inventario conceptual relacionado con ella, juegan un papel básico en la matematización de la probabilidad, pero también en las aplicaciones de la estocástica a la vida diaria. Como idea fundamental también tiene un sostén psicológico porque considera que la intuición de magnitudes en las que participa el azar surge en una etapa más temprana que la de experimento aleatorio. Desde la perspectiva del modelo explicativo, Heitele considera que la variable aleatoria tiene un papel fundamental en tres aspectos: su distribución, su esperanza y las operaciones entre variables aleatorias.

Miller (1998) también considera el concepto de variable aleatoria como fundamental en la estadística. Él basa su afirmación en la importancia que este concepto tiene en posteriores nociones estocásticas, como las distribuciones de probabilidad, el modelo de regresión o la obtención de estimadores y considera, por lo tanto, que la confusión que se puede generar en este tema, acarreará repercusiones posteriores importantes en los estudiantes. Por su parte Batanero (2001) hace referencia al trabajo de Heitele y acepta su propuesta de la variable aleatoria como una idea estocástica fundamental. Ella afirma que la variable aleatoria es la responsable del paso del estudio de los sucesos aislados al estudio de las distribuciones de probabilidad, además de que, junto con las funciones de distribución, constituye una herramienta muy potente que permite hacer uso del análisis matemático en los fenómenos aleatorios. Batanero también está de acuerdo con Heitele en que la variable aleatoria es una idea que se maneja cotidianamente y esto es lo que hace posible que en ocasiones aparezca la idea intuitiva de magnitud aleatoria antes que la de probabilidad.

Así, de acuerdo con estos autores, en esta investigación nos introduciremos al estudio de la variable aleatoria con la finalidad de contribuir en la investigación sobre su aprendizaje y su enseñanza.

1.5 Estado actual de la Investigación

Hay muy poca literatura de investigaciones centradas específicamente en la didáctica de la variable aleatoria, aunque reconocemos que hay más abundantes investigaciones sobre cuestiones directamente vinculadas con nuestro tema de interés, como estudios sobre el aprendizaje y la enseñanza de las distribuciones de probabilidad (en particular de la

Normal), de la ley de los grandes números, del teorema central del límite, de las concepciones de probabilidad o bien de la enseñanza y el aprendizaje de temas no propiamente estadísticos, como la variable algebraica o la noción de función. La revisión de todos ellos es necesaria para introducirnos en el estudio de la enseñanza-aprendizaje de la variable aleatoria, sin embargo, como punto de inicio, en este apartado nos concentraremos en aquellos autores que, consideramos, se han interesado directamente en el estudio de la enseñanza o aprendizaje de la variable aleatoria o que han obtenido resultados en esta área aunque su investigación no se centre necesariamente en ella. Eso nos permitirá tener un panorama del estado actual de las investigaciones que aportan resultados sobre la didáctica de la variable aleatoria. Después la revisión bibliográfica se analiza y puntualiza de acuerdo a tres aspectos que nos parecen de interés: el didáctico, epistemológico y el cognitivo.

1.5.1 Revisión bibliográfica

Oseguera (1994) hace un análisis del material de apoyo para el alumno que sirve como base para el diseño de la enseñanza y la evaluación del tema de variable aleatoria en la asignatura de probabilidad y estadística en la Unidad Profesional Interdisciplinaria de Ingeniería. Ciencias Sociales y Administrativas (UPIICSA) del Instituto Politécnico Nacional en México. El análisis de este material de apoyo se basa en tres ejes principales. Aquí desglosamos los dos de mayor interés para nuestra investigación:

- *Los conceptos básicos de probabilidad relacionados con variable aleatoria*. Este criterio fue establecido por el mismo autor. Los conceptos básicos se refirieron principalmente al tipo de distribución que se ponía en juego en cada reactivo (normal, uniforme o exponencial) y al manejo de sus respectivos parámetros, así como a la aproximación de la distribución binomial a una normal. Dentro de estas categorías también se incluyeron el área en la que incide el reactivo y si se requiere alguna representación gráfica.

- *La propuesta epistemológica de Heitele*. Oseguera seleccionó las ideas fundamentales propuestas por Heitele que se relacionan con la variable aleatoria y detectó si estaban presentes en el material analizado.

En su trabajo, Oseguera observa que hay una tendencia predominante a interpretar los reactivos meramente como modelos matemáticos, olvidándose del contexto y de la forma en cómo los números fueron asignados a los valores de las variables aleatorias y cómo estos números pueden relacionarse con las mediciones (datos) y con el modelo matemático. Así mismo observa que hay una ausencia de vinculación entre el tópico de estocásticos y las experiencias intuitivas de los estudiantes dando prioridad a requerimientos puramente matemáticos.

Oseguera concluye asegurando que en el contexto que él analiza no hay aspectos del surgimiento de la idea de variable aleatoria en el proceso de enseñanza-aprendizaje, reduciéndolo al manejo metodológico de reglas y símbolos.

En su breve artículo, Miller (1998) hace una crítica al tratamiento de la variable aleatoria en algunos libros de texto. Él afirma que en los libros de introducción a la estadística casi no desarrollan esta idea y que cuando lo hacen, frecuentemente lo hacen de una forma errónea que acarrea problemas en temas posteriores. Critica, en particular, el tratamiento de la variable aleatoria como asociada al valor de los datos. Este tratamiento crea una confusión entre el resultado del fenómeno aleatorio propiamente y el valor o dato asignado a la variable aleatoria, como si ese dato no pudiera existir en sí mismo sin la presencia del fenómeno aleatorio. Miller toma dos ejemplos del libro de Moore (Moore, D.S. (1995). *The basic practice of statistics*. Freeman: New York) en los que se ilustra el concepto de variable aleatoria usando la distribución de las calificaciones de un grupo de estudiantes y la distribución del tamaño de un grupo de casas. En ambos casos el resultado del fenómeno aleatorio no es la calificación obtenida o la medida de la casa, sino el estudiante o la casa seleccionados aleatoriamente. El valor numérico relacionado con la casa o el estudiante están definidos para cada casa y para cada estudiante, no así el estudiante o la casa seleccionada a través de un fenómeno aleatorio. Los problemas son, afirma Miller, que (1) el valor numérico no determina claramente a la variable aleatoria (el tamaño de la casa no es un «resultado» numérico del fenómeno aleatorio) y (2) que en el discurso estadístico sí aparecen variables aleatorias dadas por un valor numérico calculado o determinado (como en el caso de la variable explicativa del modelo de regresión), lo que conlleva, tarde o temprano, a una contradicción en el discurso escolar.

Por su parte, Vallecillos (1996 y 1999) hace un análisis didáctico del contraste de hipótesis desde la doble perspectiva, teórica y experimental. Ella hace una investigación bibliográfica profunda sobre los aspectos matemáticos, históricos, filosóficos y didácticos del contraste de hipótesis en donde se pone de manifiesto la complejidad del significado de este concepto. Esta investigación sirve de referencia para plantear una investigación experimental sobre el aprendizaje del contraste estadístico de hipótesis en estudiantes universitarios. Su centro está en la detección de los errores y las concepciones de estos estudiantes. Entre sus principales conclusiones Vallecillos (1999) destaca la confusión entre estadístico y parámetro, ya que, indica «aunque los alumnos diferencian entre la media de la muestra y la de la población, no perciben la media muestral como una variable aleatoria» (p. 245). Así mismo indica que la comprensión del estadístico como variable aleatoria y su dependencia del parámetro puede ser una de las causas que influyan en la interpretación incorrecta del nivel de significancia en las pruebas de hipótesis.

En su tesis de doctorado, Tauber (Tauber, 2001 y Batanero, Tauber y Sánchez, 2001) estudió el aprendizaje de la distribución normal en una clase de estadística con el uso del ordenador. Su estudio no se centra en la variable aleatoria, pero es uno de los conceptos que considera como importantes en la descripción del significado[1] de la distribución normal.

En su propuesta de enseñanza, Tauber introduce la variable aleatoria continua como la generalización de la idea de la distribución de frecuencias de una variable estadística continua, es decir, presenta la función de densidad como el modelo matemático al que tendería el histograma de la variable estadística cuando se aumenta indefinidamente el tamaño de la muestra, haciendo progresivamente cada vez más pequeños los intervalos del histograma. Su análisis se centra en las diferencias y concordancias presentadas entre el significado de la distribución normal de los estudiantes que siguieron su propuesta y el significado institucional local pretendido. El trabajo se sustenta en el marco teórico del significado de los objetos matemáticos (Godino, 1996 y Godino y Batanero, 1998b) y su principal finalidad es evaluar los significados personales construidos por los alumnos al ser sometidos a una propuesta de enseñanza novedosa. La propuesta didáctica implementada también está basada en el análisis de los elementos de significado de la distribución normal que se presentan en una muestra de libros de texto universitarios (significado institucional de referencia).

[1] El significado de los conceptos matemáticos visto como un constructo teórico bajo la perspectiva de Godino y Batanero (1994 y 1998b).

Basándose en la confusión de los estudiantes entre el modelo y los datos reportada por autores como Vallecillos (1999), Tauber (2001), entre uno de sus objetivos de evaluación, se propone observar si los estudiantes discriminan correctamente los conceptos de parámetro y estadístico y la distribución teórica de la empírica. De modo que parte de su análisis desemboca en la comparación entre el manejo de la variable estadística y de la variable aleatoria.

Algunas de las conclusiones, importantes dentro del contexto que nos ocupa, que Tauber obtiene son las siguientes:

- En los libros de texto no siempre se relaciona el estudio de la estadística descriptiva con el de la variable aleatoria y con el de las distribuciones de probabilidad. Es decir, no se hace una conexión entre el estudio del modelo probabilístico y el análisis de datos empíricos, por lo que los modelos matemáticos pierden su razón de ser si no se relacionan con los datos que se quieren modelar.

- Los libros de texto enfatizan sobre todo en los elementos simbólicos, algebraicos y numéricos, particularmente las tablas de la distribución, y se dedica gran tiempo al cálculo de probabilidades a partir de las tablas.

- Pudieron observarse ciertas dificultades de los alumnos en la distinción de la distribución teórica y empírica, sobre todo cuando se ven en la necesidad de resolver problemas abiertos.

- Los alumnos llegan a un manejo razonable de la herramienta informática y a realizar correctamente cálculos aislados. Sin embargo, cuando se trata de poner en correspondencias diferentes elementos del significado para tomar una decisión -por ejemplo decidir si la variable estadística podría modelarse adecuadamente mediante la distribución normal- el número de respuestas correctas baja drásticamente.

Ortiz (2002), por su parte, hace un estudio sobre la caracterización del significado de los conceptos probabilísticos que se presentan, como resultado de la transposición didáctica, en los libros de texto del primer curso de bachillerato. Él se concentra en siete conceptos que considera fundamentales para determinar el significado del concepto de probabilidad que se presenta en los libros textos. Entre estos incluye el de variable aleatoria.

Siguiendo un procedimiento cualitativo, Ortiz analiza las definiciones, tipos de problemas y representaciones asociadas a los conceptos de aleatoriedad, frecuencia, probabilidad y variable aleatoria entre otros en una muestra de libros de texto. A partir de este análisis define los elementos de significado asociados a cada uno de ellos, siguiendo el marco teórico de Godino y Batanero (1994). Este procedimiento le permite obtener una categorización de los elementos de significado del concepto de variable aleatoria, así como la identificación de los elementos de significado de esos conceptos presentes en una muestra de libros de texto de primer año de bachillerato en España.

Ortiz reporta que en la mayoría de los textos que analizó no se trata el concepto de variable aleatoria de manera explícita, pero aquellos que lo hacen, lo ubican ya sea en los temas de probabilidad o en los de estadística y que sólo contemplan una mínima parte de sus elementos de significado. Él encontró que el concepto también es tratado de manera implícita, sobre todo cuando se describe la variable estadística en el tema de probabilidad, porque los valores de la variable estadística se obtienen a partir de los valores observados de una variable aleatoria en una muestra. Observa en uno de los textos cierta confusión entre los conceptos de variable aleatoria y variable estadística, ya que atribuyen tanto la distribución de probabilidad como la distribución muestral a la variable aleatoria. Recomienda que, dada la importancia que tiene, este concepto debiera ser tratado y relacionado con la probabilidad, aunque de una manera no excesivamente formalizada. Para lograrlo se deberá retomar con diferentes grados de profundización a lo largo de la educación secundaria.

En la caracterización del tipo de actividades (ejercicios, problemas o ejemplos) que se pueden resolver en el contexto de la variable aleatoria, define nueve tipos de actividades en los que se presenta el concepto de variable aleatoria, pero en los libros de texto analizados, cuatro de ellos no se presentan. Las actividades más frecuentemente presentes se refieren a la determinación de la distribución de probabilidad de una variable aleatoria y a la representación gráfica de dicha distribución. Así mismo, reporta que las actividades presentes sólo se manejan referidas a espacios muestrales finitos en contexto únicamente de juego, que la asignación de probabilidades predominante es mediante la regla de Laplace y que la forma de presentación más usada es la verbal, en ocasiones con tablas de datos. Ortiz recomienda diversificar las actividades en los libros de texto, sin olvidar que los conceptos deben introducirse acordes al nivel de enseñanza y edades de los alumnos. Él considera que la categorización de la variable aleatoria que realizó sería una buena guía para lograrlo.

Un trabajo reciente es el de Nardecchia y Hevia (2003), quienes realizaron una investigación bibliográfica histórica tendiente a encontrar los posibles obstáculos didácticos que el aprendizaje de la variable aleatoria pueda implicar. En uno de los resultados preliminares de su investigación, ellos concluyeron que el concepto de variable aleatoria está fuertemente vinculado con el de aleatoriedad, pero que la historia muestra que ha sido difícil aceptar esa relación, de modo que una de las principales dificultades con las que un estudiante se podría enfrentar en la construcción de la noción de variable aleatoria sería no visualizar la presencia del azar en el fenómeno aleatorio que se está tratando de modelar a través de la variable aleatoria. Ellos argumentan que, aunado a esto, los estudiantes tienen predominantemente desarrollado el pensamiento determinístico sobre el probabilístico y que esto puede influir aún más en la presencia de ese obstáculo principalmente en la enseñanza superior. Concluyen, también, que históricamente no ha sido simple la construcción de un modelo adecuado a partir de los datos observados, de modo que esta vinculación entre la realidad y la variable aleatoria (como modelo matemático) puede constituir otro obstáculo con el que se podría enfrentar un estudiante. Así mismo, ellos enfatizan en la importancia de realizar estudios que nos indiquen la transposición didáctica que el concepto matemático 'variable aleatoria' ha sufrido para poder ser incorporado a la enseñanza en las instituciones educativas.

A continuación, estos antecedentes de investigación se analizarán desde su aportación a tres aspectos de interés: hacia lo didáctico, lo cognitivo y lo epistemológico.

1.5.2 Aspecto didáctico

De la anterior revisión se puede concluir que la investigación sobre la variable aleatoria es aún muy somera, incluso cuando este tema forma parte fundamental del discurso escolar de la Probabilidad y Estadística universitaria. Los autores citados concuerdan en que el concepto es básico para introducirse en el estudio de la estadística universitaria y algunos de ellos recomiendan que sea retomado una y otra vez a lo largo de la educación previa a la universitaria con diferentes grados de profundidad (Heitele, 1975; Batanero, 2001; Miller, 1998 y Ortiz, 2002). Heitele (1975) propone que el éxito de esta idea en cada nivel escolar estará en ser tratada bajo diferentes niveles cognitivos que desarrollen modelos explicativos que difieran en su forma lingüística y en sus niveles de profundización. Heitele se pregunta qué será más conveniente si introducir el concepto de variable aleatoria vía el caso especial de equidistribución o de acuerdo al principio de Dienes de introducirlo hasta el final, vía las distribuciones generales. Sin embargo, no se encontraron más referencias a cuáles podrían ser esos estadios de la variable aleatoria que propiciarían el crecimiento del concepto de variable aleatoria en cada nivel escolar. Así pues, no se tiene muy claro que es lo que se espera que los estudiantes aprendan en cada nivel escolar, aunque se manifiesta una necesidad de que la noción sea introducida en niveles escolares primarios.

Algunos de los autores citados se enfocaron al estudio de este concepto en algunos libros de texto concluyendo que esos libros de texto casi no desarrollan la idea de variable aleatoria de manera implícita y cuando lo hacen sólo contemplan una mínima parte de sus elementos de significado o bien cometen errores en su interpretación (Oseguera, 1994; Miller, 1998; Tauber, 2001 y Ortiz, 2002). Ortiz (2002) reporta una confusión entre variable aleatoria y variable estadística en algunos libros de secundaria. Miller (1998) muestra la inconveniencia de asociar a la variable aleatoria con el valor de los datos directamente en los libros de introducción a la estadística universitarios. En este mismo sentido, Ortiz (2002) cita a Borovcnik y colaboradores (1991) con una definición práctica del concepto: «una variable aleatoria se determina como resultado de un experimento aleatorio» (p. 121) que alude al error que critica Miller de relacionar a la variable aleatoria con el resultado de un fenómeno aleatorio. Sin embargo, Ortiz considera que esa definición informal es la apropiada para el nivel de secundaria.

Así mismo, los autores citados concuerdan en que en el material de apoyo de la enseñanza hay un énfasis en el tratamiento de los elementos simbólicos, algebraicos y numéricos de la variable aleatoria y en que hay una ausencia de vinculación entre el tópico y las experiencias empíricas. Oseguera (1994) y Tauber (2001) encontraron que no siempre hay una conexión entre el estudio del modelo probabilístico y el análisis de los datos empíricos en el material de apoyo y textos universitarios que analizaron, olvidándose del contexto y de la forma en que se asignaron los valores a las variables aleatorias y cómo estos números pueden relacionarse con las mediciones (datos) y con el modelo matemático. Esto significa que hay un rompimiento entre el fenómeno aleatorio y la variable aleatoria o su distribución de probabilidad. Nardecchia y Hevia (2003) mencionan la dificultad de aceptar históricamente la relación entre la aleatoriedad y la variable aleatoria. Cosa por demás grave si a esta ausencia añadimos las dificultades que han sido señaladas en numerosas investigaciones sobre la percepción de la aleatoriedad en los estudiantes y las dificultades filosóficas alrededor del concepto de aleatoriedad. Algunas referencias a estas dificultades las encontramos en: Falk y Konold, 1997 y Batanero y Serrano, 1995 y 1999 y Batanero, 2001.

1.5.3 Aspecto epistemológico

Desde la perspectiva histórica se hace referencia a la importancia de la variable aleatoria en la comprensión de nociones que eran difíciles de entender antes de su surgimiento, tales como el Teorema Central del Límite (Heitele, 1975) y la dificultad en la aceptación histórica de su relación con la aleatoriedad. Nardecchia y Hevia (2003) mencionan la posibilidad de que los estudiantes se enfrenten a esta última cuando se está tratando de modelar a través de la variable aleatoria.

Desde el punto de vista conceptual, Heitele (1975) considera que la variable aleatoria juega un papel fundamental en tres aspectos que de alguna manera se pueden considerar campos de problemas en donde se pone en juego la noción de variable aleatoria:

- *Su distribución*, caracterizada por su esperanza y su desviación estándar. El análisis crítico de datos requiere de la caracterización apropiada de una función de distribución, a través de su valor esperado y de su desviación estándar.
- *Su esperanza*, que ha mostrado ser una idea muy intuitiva y que tiene un valor explicativo muy grande en un amplio campo de problemas.
- La *composición* de variables aleatorias para obtener otras nuevas.

Así mismo Ortiz (2002) hizo un desglose de los elementos de significado de la variable aleatoria, haciendo notar su complejidad epistémica, la cantidad de conceptos que es necesario poner en juego para su comprensión y los campos de problemas en los que participa. Ortiz detectó 11 *elementos intensivos*[2] y 9 *elementos extensivos*[3] de significado

[2] Desde el enfoque ontológico-semiótico de la cognición matemática (Godino y Batanero, 1994), los *elementos intensivos* del significado de un concepto matemático son sus conceptos, sus definiciones y sus propiedades.
[3] Desde el mismo marco teórico, los *elementos extensivos* son los problemas y los ejercicios vinculados al concepto.

de la variable aleatoria. Dentro de su análisis de los elementos extensivos, Ortiz también se ocupó del análisis de variables de tarea, dentro de ellas consideró dos líneas conceptuales que se retoman en distintos conceptos probabilísticos, el tipo de espacio muestral y la asignación de probabilidades, además de otras de acuerdo a la naturaleza de la tarea como contexto del problema y el tipo de presentación de la información. Sus elementos se ocupan de la variable aleatoria y de su distribución, así como de los elementos que la definen: su media y su varianza. La base de su análisis fue una definición formal del concepto, aunque su análisis en los libros de texto incluye nociones no dadas explícitamente, aquellas en las que el concepto se presenta y las relaciones que le son asociadas. En el apéndice 1 se pueden consultar los elementos de significado propuestos por Ortiz de manera extensa.

El trabajo de Ortiz resulta valioso por ser una primera propuesta de análisis epistémico del concepto de variable aleatoria además de que proporciona conclusiones importantes sobre la didáctica de la variable aleatoria. Sin embargo, se ocupa más de la caracterización de su función de distribución que del concepto de variable aleatoria en sí mismo, además de que no incluye la composición de variables aleatorias dentro de su problemática, posiblemente por el nivel escolar que trabajó en su investigación.

Vallecillos (1999) y Tauber (2001) hacen hincapié en la importancia de una de las concepciones de la media muestral como variable aleatoria. Miller (1998) por su parte, afirma que el tratamiento didáctico de las variables aleatorias definidas como el resultado de un experimento aleatorio generará confusión en el estudio de las variables aleatorias dadas por un valor numérico calculado. En ambas cuestiones están sumergidas dentro de la problemática de la composición de variables aleatorias.

Por otro lado, Tauber (2001) enfatiza en el papel de la variable aleatoria en los modelos matemáticos, puesto que estos pierden su razón de ser si no se relacionan con los datos que se quieren modelar. Heitele (1975) recalca la diferencia de la modelación en matemáticas y en probabilidad y estadística. Batanero (2002) considera que cualquier forma de representar la realidad puede considerarse modelo (un estadístico útil, un texto, un gráfico o una línea de regresión), siempre y cuando se diferencie el modelo de los datos y al mismo tiempo se relacione el modelo con los datos. Para Heitele esa diferenciación se facilita cuando las relaciones entre realidad y matemáticas se ven en estructura de estratos. Esto es, los datos numéricos en estadística descriptiva se pueden considerar como datos de un modelo, a la vez que desde un nivel más alto (en otro estrato), se les puede considerar parte de la realidad. Lo importante para él es definir cuál es el nivel del estrato en el que se está trabajando y qué es lo que en ese estrato puede fungir como «realidad» y qué como «modelo». Heitele no enfatiza en la variable aleatoria en particular, pero sí menciona la importancia de considerar esto en la modelación en probabilidad y estadística, puesto que en el contexto estocástico, el «modelo» puede modificar los datos que surgen de la «realidad».

1.5.4 Aspecto cognitivo

En este apartado encontramos tres aspectos reportados en los que los estudiantes se pueden enfrentar a problemas con el aprendizaje de la variable aleatoria:

- ❖ La intuición sobre magnitudes en las que participa el azar surge en etapas más tempranas que la de experimento aleatorio (Heitele, 1975 y Batanero, 2001).
- ❖ Los estudiantes tienen predominantemente desarrollado el pensamiento determinístico sobre el probabilístico y eso puede influir en la no visualización del azar en el fenómeno aleatorio cuando se está tratando de modelar haciendo uso de la variable aleatoria (Nardecchia y Hevia, 2003).
- ❖ Confusión entre estadístico y parámetro en los estudiantes debido a que no perciben la media muestral como una variable aleatoria y por consiguiente no distinguen la distribución teórica de la empírica (Vallecillos, 1999 y Tauber, 2001).

En estos tres puntos se destacan las dificultades del aprendizaje de la variable aleatoria cuando se trata de manera intuitiva y en su papel dentro de la modelación de los fenómenos aleatorios.

1.5.5 A manera de conclusión

La mayoría de los estudios consultados se enfocan al aspecto didáctico de la variable aleatoria, con mayor acento en el análisis de libros de texto. Hay pocas cuestiones, sin embargo, en los aspectos cognitivo y epistemológico. En este último hay pocas referencias al aspecto histórico de la variable aleatoria y una mayor mención de cuestiones referentes a la disciplina (probabilidad y estadística). Observamos también que algunos de los estudios sobre el aprendizaje o enseñanza de la distribución normal, de las distribuciones muestrales, de las distribuciones de probabilidad y de frecuencias, o de los contrastes de hipótesis, que no se centran específicamente sobre el concepto de la variable aleatoria, se enfrentan a las dificultades y errores que tienen los estudiantes, los textos o el material de apoyo sobre ese concepto.

Así, aunque hay algunos estudios preliminares sobre variable aleatoria y estudios sobre las distribuciones muestrales, la distribución normal, los contrastes de hipótesis y algunos otros conceptos básicos asociados con la variable aleatoria como la media, las distribuciones y la varianza, la enseñanza misma de la variable aleatoria no se ha estudiado en forma sistemática a nivel universitario. Sin embargo es un tópico que la investigación en didáctica de otros conceptos estadísticos universitarios forzosamente tiene que abordar.

1.6 El problema de investigación

Esta investigación aborda el problema didáctico de la variable aleatoria. La intención es conocer cómo se apropian los estudiantes de la noción de ese concepto en situación escolar a nivel universitario. Sin embargo, dado lo poco abordado que ha sido el tema en la investigación, afrontar este propósito resulta una tarea ardua y extensa. Así, esta etapa de la investigación se restringe a inmiscuirse en los preliminares de la didáctica de este concepto. La investigación se aborda únicamente desde las dimensiones cognitiva y epistemológica con la intención de identificar variables de observación relevantes, como obstáculos, trayectorias, conceptos matemáticos cercanos y prioridades entre conceptos, que aporten elementos para el estudio de la didáctica del concepto de variable aleatoria en el aula.

Desde la componente cognitiva se plantea la realización de una exploración cognitiva. Se trata de diseñar una situación experimental que parta de la resolución de un problema en un contexto familiar a los estudiantes y que permita encaminarlos hacia las ideas de experimento aleatorio, variable aleatoria asociada a dicho experimento, distribución de probabilidad y a su representación gráfica.

En la dimensión epistemológica se profundiza en la naturaleza del concepto de variable aleatoria a través del análisis del concepto en la disciplina que lo formaliza y lo justifica, la probabilidad, y del estudio de su emergencia histórica.

Capítulo 2

Fundamentos teóricos y metodológicos

Desde el aspecto teórico nos apoyamos en la *teoría de situaciones didácticas* planteada por Brousseau (1976/1981, 1986, 1990 y 1997) y ello nos conduce, desde el punto de vista metodológico, a la metodología conocida como Ingeniería Didáctica (Artigue, 1995a y 1995b). Nuestra investigación se concentra sólo en el análisis preliminar y se complementa con la entrevista clínica (Clement, 2000).

La teoría de situaciones didácticas toma como objeto de estudio al sistema didáctico y, más ampliamente, al sistema de enseñanza. El sistema didáctico se sustenta en el triángulo epistemológico: docente-saber matemático-alumno y sus relaciones. A partir de esta perspectiva, la ingeniería didáctica propone un análisis preliminar de los conceptos matemáticos desde tres dimensiones: la didáctica, la epistemológica y la cognitiva con el objetivo de estudiar el fenómeno didáctico en el aula. Los resultados aportados por las tres componentes dan lugar al diseño de una secuencia didáctica que permitirá el estudio de la didáctica del concepto matemático en el aula a través de variables de observación relevantes también fundamentadas en el análisis preliminar. Así, en esta teoría, el análisis preliminar, que constituye únicamente el preámbulo para la investigación en el aula, requiere de una investigación amplia que profundice en la naturaleza del concepto matemático desde cada uno de sus componentes y sus interrelaciones. Para ello, la teoría de situaciones didácticas se basa en la definición de una serie de constructos teóricos en los que se sustenta el análisis y la descripción del fenómeno didáctico.

En este capítulo, primero expondremos este marco teórico en el que se respalda nuestra investigación para abordar y plantear la problemática descrita en el capítulo anterior, así como su metodología. Estos elementos teóricos nos servirán de apoyo en la descripción y análisis del concepto, tanto desde la cognición matemática como desde su epistemología.

Así mismo describimos la herramienta metodológica de *entrevista clínica* que auxiliará la investigación en la vertiente cognitiva. Y finalmente, el capítulo se concluye con la clarificación de los objetivos de esta investigación desde la perspectiva de nuestro marco teórico.

2.1 Teoría de situaciones didácticas

Brousseau (1997) vislumbró por primera vez la necesidad de utilizar un modelo propio de la actividad matemática en la investigación en didáctica de la matemática a partir de la problematización y cuestionamiento del conocimiento matemático enseñado. A ese modelo le llamó *Teoría de situaciones didácticas*.

Brousseau está de acuerdo con la *Teoría de equilibración mayorante* de Piaget (1950/1991) en que el conocimiento se produce a través de la actividad humana, por la interacción entre el objeto y el sujeto. El reto está en construir *situaciones didácticas* con las que el estudiante se involucre con *situaciones a-didácticas* que permitan que un conocimiento matemático particular emerja de la actividad del estudiante. En los párrafos siguientes describiremos brevemente algunos de los aspectos más relevantes de esta teoría.

2.1.1 Teoría de la Equilibración mayorante de Piaget

Dentro de la teoría de situaciones didácticas, el aprendizaje del alumno se fundamenta en la *Teoría de la Equilibración mayorante* de Piaget (1950/1991). Una de las ideas centrales de esta última teoría es que tanto la naturaleza como la validez de los conocimientos dependen de su modo de formación. La postura de Piaget es que el conocimiento se adquiere mediante la interacción entre el sujeto y el objeto. De aquí que la acción sea para él un principio fundamental para el aprendizaje.

La naturaleza dialéctica de la teoría de Piaget es un proceso complejo de estructuraciones sucesivas a través de una jerarquía de niveles bien definidos. No se trata, afirma Piaget, de cortes arbitrarios en el seno de un proceso continuo o puramente aditivo, las estructuras adquiridas en un nivel dan lugar a una reconstrucción antes de que estas estructuras reconstruidas puedan ser integradas en las nuevas estructuras elaboradas sobre los niveles anteriores. Cada uno de los niveles constituye un estado de *equilibrio dinámico* a la manera de los estados de equilibrio estático (situaciones «estacionarias») en un sistema termodinámico. Piaget los llama *equilibración*. En la medida que el desarrollo del conocimiento es concebido como un juego, se entra en mecanismos de *desequilibración* del anterior nivel y de *reequilibración* en los nuevos niveles que se van alcanzando. Esto tiene repercusiones importantes en la didáctica porque no hay acumulación progresiva de saberes, sino una reorganización permanente de conocimientos: los nuevos saberes son integrados al saber anterior, aún a veces modificando o confrontando este último. Es necesario, por tanto que se establezca un *conflicto cognitivo* que provoque un estado de desequilibración-reequilibración para que surja un nuevo conocimiento.

La concepción básica más original de la teoría epistemológica de Piaget consiste en afirmar que la *acción* es constitutiva a todo conocimiento. El conocimiento es dependiente de la acción y la acción es productora de conocimiento. La actividad propia del sujeto no se ejerce forzosamente mediante la manipulación de objetos materiales, sino de una acción finalizada y problematizada que supone una dialéctica pensamiento-acción muy diferente a una simple manipulación guiada, orientada con frecuencia a una simple tarea de constatación.

En la acción elemental todavía no puede hablarse, en sentido estricto, ni de un sujeto ni de un objeto. Poner en el punto de partida la acción es, por un lado, sustituir las opciones clásicas (primacía del sujeto en el idealismo o del objeto en el empirismo) con un nuevo enfoque: la primacía es la del vínculo práctico, de la interacción efectiva, de la acción objetiva. Pero, por otro lado, es adoptar una perspectiva constructivista que dé cuenta de la constitución del sujeto en tanto sujeto cognoscente y del objeto en tanto objeto de conocimiento. Este papel de la acción rompe la vieja dicotomía entre pensamiento y acción. Tal como lo señala Piaget: «todas las teorías no-genéticas conciben al *pensamiento* como anterior a la *acción* y a ésta como una aplicación de aquél» (Piaget, 1950/1991, p 16). De este modo, la acción en la didáctica de las matemáticas juega un papel esencial, propone una forma de apropiación del conocimiento por parte del alumno.

2.1.2 Situaciones a-didácticas

De acuerdo con lo teoría descrita en el párrafo anterior, Brousseau (1997) sostiene que la concepción moderna de la enseñanza es pedir al maestro que provoque en el alumno las adaptaciones deseadas, con una elección acertada de los «problemas» que le propone. Estos problemas, elegidos para que el alumno pueda aceptarlos, deben hacerle actuar, hablar, reflexionar y evolucionar por sí mismo. Entre el momento en que el alumno acepta el problema como suyo y aquél en el que produce su respuesta, el maestro rehúsa a intervenir proponiendo los conocimientos que quiere ver aparecer. El alumno sabe bien que el problema ha sido elegido para hacerle adquirir un conocimiento nuevo, pero debe saber también que este conocimiento está enteramente justificado por la lógica interna de la situación y que *puede* construirlo sin atender a razones didácticas a través de la interacción con el problema. No sólo puede, sino que también debe construirlo, pues sólo habrá adquirido verdaderamente este conocimiento cuando él mismo sea capaz de ponerlo en acción, en situaciones que encontrará fuera de todo contexto de enseñanza, y en ausencia de cualquier indicación intencional. Tal situación es llamada *a-didáctica* por Brousseau.

Cada conocimiento puede caracterizarse por una o más situaciones a-didácticas que preservan su sentido y que llamaremos *situaciones fundamentales*. Pero el alumno no puede resolver de golpe cualquier situación a-didáctica, el maestro le propone, entre las situaciones a-didácticas, aquéllas que están a su alcance. Estas situaciones a-didácticas, ajustadas a fines didácticos, determinan el conocimiento enseñado en un momento dado y el sentido particular que este conocimiento va a tomar. Para Brousseau la definición del conocimiento matemático enseñado se establece a través de una *situación fundamental*, que, para un conocimiento matemático C en particular, definió como «un conjunto minimal de situaciones *a-didácticas* (específicas de C) que permiten engendrar, por manipulación de los valores que toman sus *variables didácticas*, un campo de problemas suficientemente extenso como para proporcionar una buena representación de C en relación a cómo ha sido construido C en la institución didáctica en cuestión» (Chevallard, Bosch y Gascón, 1997, p 216).

2.1.3 Situación didáctica

La situación o el problema elegido por el profesor es una parte esencial de la siguiente situación más amplia: el maestro busca *devolver* al alumno una situación a-didáctica que provoque en él una interacción lo más independiente y lo más fecunda posible. Para ello, comunica o se abstiene de comunicar, según el caso, informaciones, preguntas, métodos de aprendizaje, heurísticas, etc. En consecuencia, el enseñante está implicado en un juego con el sistema de interacciones del alumno y con los problemas que él le ha planteado. Esta situación más amplia es la *situación didáctica* (Brousseau, 1986).

Peltier (1993) añade que para construir buenas situaciones didácticas se consideren los siguientes aspectos:

- ❖ La actividad propuesta como punto de partida debe presentar un verdadero problema para los alumnos, pero, a la vez, ser comprendido por ellos. Es decir, los estudiantes deben poder pensar y planear la respuesta del problema.
- ❖ Debe permitir al alumno utilizar conocimientos anteriores (para que pueda introducirse en el problema).
- ❖ Y debe ofrecer una resistencia suficiente para llevar al alumno a hacer evolucionar sus conocimientos anteriores, a cuestionarlos, a elaborar nuevos conocimientos.
- ❖ Debe contener, en lo posible, su propia validación, es decir, el alumno debe poder por sí mismo – o confrontado con los otros alumnos – controlar su solución, decidir su validez de respuesta.
- ❖ El conocimiento previsto debe ser el más adaptado para resolver el problema

La situación didáctica implica una interacción dialéctica del estudiante con situaciones problemáticas, donde el sujeto anticipa y finaliza sus acciones y compromete sus conocimientos anteriores, los somete a revisión, los modifica, los complementa o los rechaza para formar concepciones nuevas. El objeto principal de la didáctica desde la perspectiva de la teoría de situaciones didácticas es estudiar las condiciones que deben cumplir las situaciones planteadas al alumno para favorecer la aparición, funcionamiento o rechazo de esas concepciones.

Para Brousseau (1986) el principio fundamental de la teoría de situaciones didácticas es el principio metodológico «definir todo 'concepto matemático' mediante una situación». Esto es, la «planeación documentada» de situaciones didácticas. Desde una perspectiva metodológica se habla de la búsqueda de fundamentos para diseñar secuencias didácticas que planteen problemas de los que surja de manera natural el concepto matemático al interactuar con ellos, del planteamiento de posibles conductas del estudiante al abordar esos problemas y del estudio de las interacciones del profesor con el estudiante al resolver el problema y de la verificación de tales hipótesis en el aula. Estas ideas se ponen de manifiesto en la *ingeniería didáctica*, que ha sido utilizada como la metodología de la teoría de situaciones didácticas.

2.2 Metodología de investigación

La metodología de esta investigación, acorde al marco teórico en el que se sustenta, es la ingeniería didáctica. Nos restringimos al análisis preliminar de esta metodología y dentro de él, a las componentes cognitiva y epistemológica, en este último aspecto nos enfocaremos al análisis del concepto desde la disciplina y desde la historia.

La vertiente cognitiva nos auxiliamos de la *entrevista clínica* (Clements, 2000; Von Glasersfeld, 1987 y Confrey, 1980a, 1980b y 1980c) y la vertiente epistemológica la abordaremos desde la revisión de libros históricos (originales, en la medida que tengamos acceso a ellos) y la consulta con expertos y libros sobre teoría de probabilidad.

Como lo señala Artigue (1995a), la profundización sobre el análisis preliminar del concepto matemático dentro de la ingeniería didáctica acepta el uso de herramientas metodológicas externas que ayuden en la búsqueda de elementos tendientes a proporcionar información al análisis a-priori y al diseño. Sin embargo, el objetivo explícito de las herramientas externas estará supeditado a proporcionar elementos que sean útiles a la ingeniería didáctica. La concordancia entre los sustentos teóricos de las herramientas metodológicas usadas y la teoría de situaciones didácticas es un aspecto que cuidamos y fundamentamos. Puntualizaremos cada una de las metodologías usadas en los siguientes apartados.

2.2.1 Ingeniería didáctica

Numerosos trabajos de investigación en didáctica –llamados *ingenierías didácticas*– se centran en la tarea de producir situaciones de aprendizaje destinadas a asegurar de manera controlada la emergencia de conceptos matemáticos en el contexto escolar. Tales situaciones abarcan varias fases descritas por Brousseau: fase de investigación o acción, fase de formulación, fase validación y fase de institucionalización. En esta sección nos enfocaremos a la ingeniería didáctica que ha sido propuesta como la metodología de investigación de la teoría de situaciones didácticas propuesta principalmente por Artigue (1995a).

La ingeniería didáctica, según describe Artigue (1995a y b), se aplica en *situación escolar,* su análisis es *cualitativo* y recurre a lo que llama *validación interna*. Sus formas de validación son básicamente cualitativas y están asociadas, se basan en el registro de los estudios de caso y es, en esencia, interna, basada en la confrontación entre el análisis *a priori* y *a posteriori*. La ingeniería didáctica, con respecto a su metodología experimental, está integrada por tres fases:

- *Análisis preliminar*. Es una investigación previa al planteamiento de una secuencia didáctica alrededor de un concepto matemático. Su objetivo es conocer más de cerca la naturaleza de este concepto desde la perspectiva didáctica, epistemológica y cognitiva con el propósito de identificar hipótesis sobre el proceso de construcción del concepto matemático por parte de los estudiantes en situación escolar, así como aportar elementos para el diseño de la secuencia didáctica. Proporciona información relevante para el diseño de la secuencia didáctica y el análisis a priori.

- *Diseño de la secuencia y análisis a priori*. El diseño de la secuencia supone una vinculación de los aspectos epistemológicos, cognitivos y didácticos, así como una planeación de la forma en la que el estudiante construirá el conocimiento a partir de ella. La secuencia didáctica pretende desarrollar el proceso de construcción del concepto matemático por parte de los estudiantes. Los objetivos del análisis a priori son (1) ser un elemento de validación de la secuencia didáctica, (2) poner a prueba las hipótesis que emergieron del análisis preliminar a través de una secuencia didáctica concreta y (3) aportar elementos de observación de la secuencia didáctica que de otro modo podrían pasar desapercibidos (obstáculos epistemológicos, rutas cognitivas, estadios del conocimientos, conceptos vinculados con la problemática específica).

- *Análisis a posteriori y validación interna*. A la puesta en práctica de las secuencias didácticas le sigue un análisis a posteriori que se basa en el conjunto de datos recogidos a lo largo de la experimentación, a saber, las observaciones realizadas de las secuencias. La validación interna es la confrontación de este análisis con el a priori.

Nuestra investigación se centrará en la profundización del análisis preliminar de esta metodología, de modo que únicamente nos detendremos en la descripción de esta fase.

2.2.1.1 Análisis preliminar

Para Chevallard (1985/1991) el objeto de estudio de la didáctica no puede ser visto como un fenómeno aislado, sino como un fenómeno sistémico. La didáctica se interesa en el juego que se realiza entre un *docente*, los *alumnos* y un *saber matemático*. Esta relación ternaria forma una relación didáctica, a la que Chevallard denomina *sistema didáctico* y que se sustenta en el triángulo epistemológico: *docente-saber matemático-alumno* y sus relaciones. Así el objeto de estudio de la didáctica es el sistema didáctico, más ampliamente se ocupa del *sistema de enseñanza*, que reúne los sistemas didácticos y se ocupa de todo lo que la sociedad organiza a través de instituciones para el funcionamiento de dichos sistemas didácticos.

El análisis preliminar dentro de situaciones didácticas está sustentado en el sistema didáctico. En él se analizan y se detectan aquellas restricciones o vínculos del conocimiento matemático a partir de distinguir tres dimensiones que surgen de dicho sistema:

- *Componente epistemológica*. Ésta se refiere a los aspectos relacionados con el conocimiento mismo. Pueden considerarse sus antecedentes históricos, estado actual o su referencia teórica.
- *Componente cognitiva*. Hace referencia a las concepciones de los estudiantes, a las dificultades y obstáculos que determinan la evolución de su aprendizaje.
- *Componente didáctica.* Se consideran los elementos didácticos tales como profesor, libros de texto, curriculum, método de enseñanza, etc.

2.2.1.2 Concepciones y errores

Para Piaget el desarrollo del conocimiento se realiza a través de mecanismos de equilibración y desequilibración con la consecuente reorganización de conocimientos en el sujeto. De manera tal que una persona logra un nuevo nivel de conocimiento cuando logra la re-equilibración después de una desequilibración del nivel de conocimiento anterior. En la teoría de situaciones didácticas, los nuevos estadios están vinculados con la asimilación y superación de obstáculos por los que el estudiante necesariamente debe pasar para provocar un estado de desequilibración-equilibración y por lo tanto surja un nuevo conocimiento.

Así, el estudio de las acciones y producciones de los estudiantes al enfrentarse a un problema también debe ser un punto a tomar en cuenta en el análisis preliminar. Es de esperar que de este análisis surjan hipótesis que sustenten el análisis a-priori y el diseño de la secuencia didáctica. De este modo, la didáctica de las matemáticas también se dedica a estudiar las *concepciones* del sujeto que aprende y de las condiciones en las cuales se construye el conocimiento.

Por concepción entenderemos (Peltier, 1993):

- La clase de problemas que dan sentido a un concepto para el estudiante;
- El conjunto de significantes que es capaz de asociar (imagen mental, expresión simbólica);
- Los instrumentos, teoremas, algoritmos, que es capaz de poner en marcha.

Se pretende buscar las condiciones bajo las cuales se construye el conocimiento con el fin de optimizarlas y reproducirlas en situación escolar. El error (entendido como aquel proceso operatorio que trastorna contenidos de conocimiento) constituye una fuente para encontrar esas condiciones. Es frecuente que el análisis de los errores producidos ayude al investigador a forjar hipótesis sobre las concepciones de los alumnos. No se concibe el error como una ausencia de conocimiento, sino como aquel conocimiento que se ejerce sobre un material ya dado. Conlleva siempre contenidos de conocimiento pero en su manipulación conceptual trastorna y ensambla los conocimientos de manera que no ajustan entre sí (Valverde, J. en Muñoz, J. y Valverde, J., 2000). En particular dentro de la ingeniería didáctica son importantes aquellos errores que son el testimonio de un conocimiento que tuvo su ámbito de validez, pero que la pierden en un nuevo contexto, a este tipo de errores, Brousseau (1976/1981) los llamó *obstáculos epistemológicos*.

2.2.2 Entrevista clínica

Esta técnica es considerada una herramienta metodológica que proporciona información sobre los procesos cognitivos de los estudiantes a través del cuestionamiento sobre sus concepciones por parte de un profesor (o investigador). Se enfoca al estudio de los procesos de razonamiento y de la formación de estructuras del conocimiento. En la década de los 80 el auge de investigaciones sobre resolución de problemas, trajo el surgimiento de estudios sobre los procesos que siguen los estudiantes y expertos al resolver ciertos problemas o sobre la comprensión de estudiantes de conceptos matemáticos, lo que llevó a la necesidad de emplear métodos de investigación cualitativos, como la entrevista clínica (Villareal, 2003). Sin embargo, esta técnica ha sido utilizada desde mucho tiempo antes en investigaciones psicológicas, como en el caso de Piaget (1951/1991), quien fue uno de sus pioneros. La entrevista clínica tiene una fuerte connotación constructivista porque contempla la recolección y el análisis de los procesos mentales de las ideas y significados de un estudiante. Tiene la habilidad de exponer las estructuras y procesos escondidos en el pensamiento de los individuos (Clement, 2000). De acuerdo con Piaget, (citado por Clement, 2000) la entrevista clínica supone que los procesos de razonamiento y estructuras de conocimiento que siguen los estudiantes no son los mismos que los seguidos por los especialistas. Los estudiantes tienen concepciones alternativas y usan procesos de razonamiento y aprendizaje no formales que pueden ser expuestos y detectados por la entrevista clínica. Esta técnica ha sido ampliamente usada y sustentada por investigadores como Schoenfeld (1985 y 2002), Nemirosky (1994) y Confrey (Confrey y Perry, 1980 y Confrey, 1980, 1995a, 1995b y 1995c), entre otros.

La entrevista clínica es una herramienta exploratoria que se propone el descubrimiento de lo que ocurre en la mente de los estudiantes, no de una forma estática, sino también con la intención de modificar la estructura de conocimiento del estudiante. Desde esta perspectiva, la entrevista clínica es una especie de «enseñanza experimento» cuya finalidad no sólo estará en proporcionar al investigador (o al profesor) una idea adecuada de «donde está» el estudiante, sino también una idea adecuada de «la dirección que debe tomar». Esto proveerá un modelo hipotético de la forma del razonamiento del estudiante, que complementado con otros estudios, puede proporcionar un modelo conceptual de la formación de las estructuras y operaciones que constituyen la competencia matemática[4] (Von Glasersfeld, 1987).

[4] Para Von Glaserfeld (1987) competencia está vinculado no sólo con un saber hacer sino también con el conocimiento que permite justificar por qué se hace y monitorear la actividad: «El hacer lo correcto no es suficiente, para ser competente uno debe saber lo que está haciendo y por qué es correcto» (p 8)

Para Confrey (1980 y 1995a) en las entrevistas clínicas se busca modelar la *expresión* de un estudiante a través de la *perspectiva* de un conocedor mejor informado, es decir, el modelo de las nociones del entrevistado será construido a partir de elementos conceptuales del entrevistador. Pero la forma de indagación provoca que también se modifique la *perspectiva* del investigador sobre el mismo conocimiento. Ambas, la expresión (del estudiante) y la perspectiva (del profesor-investigador), según Confrey «aportan importante contenido epistemológico a la interacción enseñanza-aprendizaje» (p 40). Puesto que no sólo proporciona al investigador un conocimiento sobre la forma en que piensa el estudiante sino también un nuevo conocimiento sobre el concepto matemático o problema sobre el que se está trabajando.

Hay una gran variedad de formas de llevar a cabo una entrevista clínica, desde entrevistas con preguntas abiertas hasta resolución de problemas en voz alta. Clement (2000) analiza la diversidad de entrevistas clínicas guiado por el protocolo de análisis que se planteé. De acuerdo con su clasificación, nuestra investigación se sitúa en un estudio *generativo/interpretativo* en donde se obtienen nuevas categorías de observación y nuevos elementos de modelación de acuerdo a las respuestas que se vayan obteniendo por parte de los estudiantes. Sus resultados son la formulación de variables relevantes de observación y modelos explicativos de procesos cognitivos basándose en un protocolo más o menos abierto y algunas veces en hipótesis generales y/o variables de observación a grandes rasgos que servirán para guiar y dirigir el estudio, sin que ello limite su interpretación. Son apropiados para tópicos poco estudiados o para los cuales se tiene poca información para establecer una teoría previa puesto que se concentran en la viabilidad y relevancia del modelo de acuerdo a las observaciones detectadas por el profesor/investigador y sólo prestan atención a aspectos no formales de fiabilidad. La generalización de los modelos producidos es hipotética y sólo será válida en escasos contextos y algunos individuos. En este tipo de entrevista, es necesario reproducir o confirmar las conclusiones así como ampliar la profundización sobre los aspectos a investigar.

En esta investigación se pretende conocer la forma en cómo un grupo pequeño de estudiantes, dos o tres, resuelven en "voz alta" una actividad propuesta sin ayuda del profesor, pero en presencia de él. En este caso el papel del profesor/investigador únicamente es intervenir para aclarar o profundizar, a sí mismo o a la investigación, las ideas que los estudiantes expresan. Los estudiantes tienen que llegar a resolver la actividad propuesta en consenso, expresando sus ideas en voz alta y argumentando sus posiciones.

2.3 Objetivo de investigación

De acuerdo a este marco teórico, nuestro objetivo de investigación se traduce en:

Proporcionar elementos de diseño y elementos predictivos que sustenten y proporcionen información para el análisis a priori y para el diseño de una secuencia didáctica alrededor del concepto de variable aleatoria en el nivel universitario. Tales elementos se concentrarán en las vertientes cognitiva y epistemológica y serán fuentes de información para la formulación de hipótesis y de variables de observación. Pueden ser prospectos de obstáculos epistemológicos, concepciones y errores, rutas cognitivas, estadios de conocimiento, conceptos vinculados con la problemática específica, prioridad entre conceptos, etc.

Se sustenta principalmente en el marco teórico de teoría de situaciones didácticas, por consiguiente la metodología principal es la ingeniería didáctica, en particular, se concentra en su análisis preliminar en sus componentes cognitivo y epistemológico y hace uso de elementos metodológicos externos.

Parte II

Análisis Cognitivo

Capítulo 3

Diseño e implementación de una entrevista clínica

Nos acercaremos a la componente cognitiva del análisis preliminar de la ingeniería didáctica a través de una exploración cognitiva, haciendo uso de la técnica de entrevista clínica descrita en el capítulo anterior (sección 2.2.3). La intención es obtener elementos que permitan discernir sobre la forma en que los estudiantes se aproximan al concepto, así como elementos que nos acerquen al análisis del concepto por medio de los argumentos, razonamientos, procedimientos y resultados que los estudiantes usen para resolver un problema en donde se ponga en juego el concepto de variable aleatoria.

La entrevista clínica es considerada una herramienta metodológica que proporciona información sobre los procesos cognitivos de los estudiantes. En esta primera parte de la investigación se pretende obtener algunas hipótesis sobre las posibles dificultades que surgen en estudiantes universitarios de los primeros semestres cuando abordan un problema estocástico relacionado con la idea de variable aleatoria, así como el atisbo de algunas trayectorias que siguen los estudiantes al querer resolver la actividad. Se espera que del análisis de esos resultados se obtengan algunas líneas que permitan explorar las vertientes epistemológicas y didácticas del análisis preliminar. Así, dentro del análisis preliminar, esta será la primera vertiente analizada que nos permitirá enfocar las vertientes epistemológica y didáctica, y posteriormente, en forma de espiral, regresar al análisis de la vertiente cognitiva. Por supuesto, sin olvidar el carácter sistémico del fenómeno que estamos analizando. En esta tesis únicamente encauzaremos la investigación hacia esta primera aproximación de la componente cognitiva (a través de la entrevista clínica) y a la componente epistemológica (a través del análisis histórico y de la disciplina).

En este capítulo describiremos la planeación de la entrevista clínica que finaliza con el protocolo de investigación. Primeramente describiremos dos antecedentes a la entrevista clínica, el contexto escolar de nuestra investigación y una primera aproximación al concepto de variable aleatoria. Posteriormente se definen los objetivos, las hipótesis de la clínica y las variables de diseño del protocolo. El protocolo se concreta en una actividad que, junto con las hipótesis, establecerán las bases de la entrevista. Finalmente se describe y analiza el problema base de la actividad desde la visión de la forma en que se manifiesta la variable aleatoria en él.

3.1 Preliminares a la entrevista

3.1.1 Contexto escolar

A pesar de que Heitele (1975) considera que la intuición sobre las magnitudes relacionadas con el azar aparece en los niños antes que la concepción de experimento aleatorio, consideramos que un planteamiento tan complicado como el concepto matemático de variable aleatoria no es sencillo de asimilar ni comprender. Actualmente los programas de estudio de nivel primaria (6 años) y secundaria (3 años) en México contemplan el estudio del azar desde las primeras etapas del desarrollo escolar de los niños, de modo que es factible que durante ese periodo los estudiantes desarrollen ideas sobre las variables vinculadas con el azar. Sin embargo, en el nivel bachillerato, que es donde podría esperarse que se trabajara este concepto de manera más formal, no está contemplado el tratamiento de la probabilidad y estadística de manera obligatoria, salvo en raras excepciones (como los bachilleratos del IPN y algunos tecnológicos regionales). En algunos otros planes de estudio es una materia optativa o es un tópico (en la mayoría de las preparatorias de México afiliadas a la SEP o a la UNAM) y en pocos de estos planes de estudio se contempla el estudio de la variable aleatoria. En el caso de nuestra institución, el ITESM, el estudio de esta materia se restringe a un modulo de estadística descriptiva en el nivel bachillerato. En la mayoría de las universidades el tema se introduce hasta el primer curso de probabilidad y estadística, como preámbulo a las distribuciones de probabilidad. De modo que esperamos que los estudiantes recién ingresados a la universidad no muestren condiciones de formalización del concepto. Las prefiguraciones sobre el concepto del estudiante universitario de nuevo ingreso estarán condicionadas por sus experiencias vividas en los niveles básicos y en su vida propia, así como por la influencia de sus cursos de matemática determinística durante toda su vivencia escolar.

3.1.2 Un primer encuentro con la noción de variable aleatoria

Una de las definiciones más simples de variable aleatoria alude a *la función que asigna un valor numérico a cada evento elemental del espacio muestral.* Sin embargo esa definición tan sencilla ya alude a la dificultad de definir una «variable» como una «función» (Meyer, 1970/1973, p. 55). Algunos autores como Borovcnik, Benz y Kapadia (1991) la relacionan con el «valor numérico que se asocia y determina por el resultado observado de un experimento [aleatorio]»[5] (p. 50). Añaden que la variable es aleatoria en el sentido de que está definida por el azar (en un experimento aleatorio) y que es necesario conocer el conjunto de todos los posibles resultados y sus probabilidades asociadas para modelar un fenómeno estocástico. Es decir, la aparente sencillez de esa definición también se desvanece cuando profundizamos un poco en ella y notamos que están involucrados otros conceptos que por sí mismos son complejos tales como experimento aleatorio, espacio muestral, evento o suceso, probabilidad y distribución de probabilidad.

Borovcnik, Benz y Kapadia la involucran además con el proceso de modelación del fenómeno aleatorio, lo cual trae consigo la definición de una magnitud de interés que centre el problema y defina el contexto matemático. La asignación de la variable aleatoria queda, por tanto, definida por el problema que se está interesado en resolver dentro un fenómeno aleatorio y en si es posible transformar el contexto cotidiano en un contexto matemático. Al mismo tiempo, es lo que hace que la función de probabilidad sea una función dentro de los números reales y por lo tanto pueda ser factible hacer uso de herramienta matemática compleja, como el análisis matemático, en los fenómenos aleatorios.

En la figura 1 se muestra, de manera simplificada, la noción de variable aleatoria que, previo a esta investigación, esperamos de los estudiantes del nivel universitario tengan al egresar de una carrera en ciencias sociales dentro del ITESM, campus Monterrey, y su relación con los conceptos con los que se vincula más estrechamente.

[5] El original en ingles: «..some numerical value which is associated with and determined by the observed outcome of the experiment».

La variable aleatoria asigna un valor numérico a cada suceso (evento) del espacio muestral, lo que hace posible que se defina una función de distribución, $F(x) = P(X \leq x)$, que, a su vez, caracterizará a la variable aleatoria vinculada a un experimento aleatorio en particular y concretará la definición de una relación funcional dentro de los números reales. A su vez, cada evento está definido en un espacio muestral de un experimento aleatorio. La variable aleatoria vincula al espacio muestral (en particular a cada evento simple) con un valor numérico definido en los reales, al mismo tiempo que la inversa de la variable aleatoria (recordemos que finalmente la variable aleatoria es una función) vincula el valor numérico con el experimento aleatorio (o con cada evento elemental). Es decir, la función de distribución se vincula con el experimento a través de operar con la variable aleatoria.

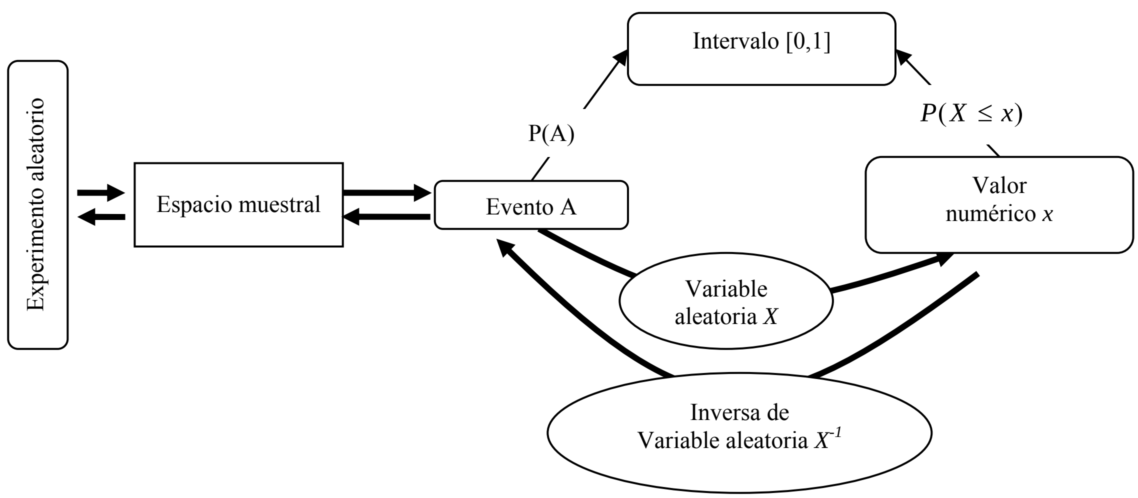


Figura 1. Esquema simplificado la noción de variable aleatoria esperado por los estudiantes que egresan del nivel universitario.

Desde una perspectiva más formal, el experimento aleatorio, el espacio muestral y los eventos asociados a él, están definidos en un espacio de probabilidad que nos relaciona con espacios medibles y con estructuras σ-álgebras. A la vez que la función de distribución vincula el aparato probabilístico con herramientas de cálculo y de análisis. El aparato matemático anterior y posterior a la definición de variable aleatoria es un tanto complicado por lo que no esperamos esa especialización en nuestros estudiantes, sin embargo en esta investigación es importante hacer notar que es esta definición la que nos permite pasar del trabajo matemático con conjuntos al trabajo matemático con números reales.

3.2 Objetivos de la entrevista clínica

Desde la perspectiva cognitiva, nos interesamos en conocer cuál es la estructura matemática que los alumnos recién egresados del bachillerato pueden construir sobre la variable aleatoria y cuáles son sus prefiguraciones. Así como observar la necesidad del surgimiento del concepto durante la resolución de un problema por parte de los estudiantes.

Definimos el objetivo de esta entrevista clínica como sigue:

En primer lugar deseamos identificar, en los estudiantes universitarios que inician un curso de estadística en ciencias sociales, los procesos de solución de la actividad propuesta cuando trabajan sin la ayuda directa del profesor. Queremos describir las dificultades por las que atraviesan, y con qué conceptos previos a la idea de variable aleatoria se relacionan. Nos interesa explorar las concepciones de los estudiantes sobre nuestro objeto de investigación y las ideas que usan y desarrollan al resolver una situación problema en donde se pone en juego el concepto de variable aleatoria.

Nos preguntamos también por la idoneidad de la situación planteada, en cuanto permita que a partir de ella los alumnos desarrollen la idea de variable aleatoria y sea una situación motivadora.

3.3 Diseño de la actividad

En nuestro estudio diseñamos algunas hipótesis que guiaron la entrevista. Éstas permitieron que el investigador y el profesor centraran las preguntas, así como que profundizaran en las nociones que las estudiantes utilizaban o mencionaban. También se plantearon algunas variables de observación en las que se basó el diseño del protocolo de investigación y la selección del problema a resolver. Estas variables también permitieron desglosar y concretar, a grandes rasgos, las hipótesis de investigación en el protocolo. En este punto describiremos las hipótesis de investigación de la entrevista clínica, sus variables de diseño y la descripción de la actividad que se usó como protocolo de investigación.

3.3.1 Hipótesis

La hipótesis que planeamos son generales en la medida que no tenemos suficiente información previa que nos permita definir hipótesis más específicas. El sustento teórico del método de investigación empleado (entrevista clínica) nos lo permite. Así, nos acercamos a las concepciones de los estudiantes con los siguientes prejuicios con respecto a su comportamiento ante la situación problema:

- *Falta de percepción de la aleatoriedad del proceso*.

  Nardecchia y Hevia (2003) nos alertan acerca de la falta de percepción de la aleatoriedad durante el desarrollo histórico del concepto de variable aleatoria. Esperamos que los estudiantes no se percaten de la vinculación entre la variable aleatoria y la aleatoriedad. Desde la perspectiva formal, esta vinculación se realizaría a través del espacio probabilístico del experimento aleatorio. Desde la perspectiva intuitiva pensamos en la relación que se establece entre una variable numérica (la variable aleatoria) y la aleatoriedad del proceso. En ese sentido, en los fenómenos aleatorios no se puede asegurar que se obtendrá el mismo *resultado*, no obstante se repita muchas veces el mismo experimento bajo las mismas condiciones, es decir que no son reversibles. Las dificultades de percepción de la aleatoriedad en estudiantes han sido señaladas en numerosas investigaciones, como, por ejemplo, en Falk y Konold (1997) y Batanero y Serrano (1999).

- *Tendencia a "algebraizar" y descontextualizar los procedimientos relacionados con la noción de variable aleatoria*.

  Esto es, que la componente aleatoria que se presenta en el contexto de un problema que se trata de matematizar tienda a diluirse para dársele prioridad a los procedimientos matemáticos. La variable aleatoria permite la definición de las funciones de distribución y la distribución de probabilidad y por lo tanto, también el análisis de las situaciones probabilísticas a través de la introducción del análisis matemático. Sin embargo el uso de esta herramienta puede repercutir en que se descontextualice el aparato matemático de las situaciones que hacen surgir a la variable aleatoria y que en cambio se le dé mayor importancia a los procedimientos matemáticos. Es decir, el proceso mencionado por Borovcnik, Benz y Kapadia (1991) y Tauber (2001) que enfatiza en el papel de la variable aleatoria en la modelación, se reduciría al trabajo con el modelo matemático, olvidando la parte de generación del modelo y de la interpretación de sus resultados. Oseguera (1994) y Tauber (2001) reportan esta ausencia en su análisis de material de apoyo y de textos universitarios.

- *Extrañeza de trabajar funciones en un contexto probabilístico*.

  El estudio de las funciones normalmente se asocia con el estudio del cálculo y por lo tanto con la matemática determinística. Es de esperarse que los estudiantes asocien las gráficas y las ecuaciones de las distribuciones de probabilidad con esa parte de la matemática en la que han profundizado en los últimos semestres de su escolaridad.

En particular esperamos que les sea extraño trabajar con una variable dependiente en una función tan poco tangible como lo es la distribución probabilidad, en el sentido de que no es una magnitud física.

- *Dificultades con la noción formal de variable aleatoria*.

  Heitele (1975) sugiere a la variable aleatoria como idea fundamental porque es posible que se proporcionen modelos explicativos de esa noción, que difieran tanto en sus niveles de profundización como en sus formas lingüísticas en cada etapa de desarrollo del individuo. Algunos otros autores (como Batanero, 2001; Miller, 1998 y Ortiz, 2002) se unen a su propuesta de enseñarla en diferentes niveles cognitivos, desde los primeros niveles escolares, sin embargo no nos queda claro cuáles son esos niveles cognitivos que dan lugar a diferentes prefiguraciones del concepto. Podemos esperar que en nuestra investigación los estudiantes tengan un manejo apropiado de la variable aleatoria de manera intuitiva y que en cambio tengan problemas al formalizarla, de modo que estaremos alerta acerca de cuáles son esas dificultades y si efectivamente esta noción (al nivel que la queremos enseñar)[6] representa una dificultad para los estudiantes.

- *Dificultades en los conceptos que intervienen en la noción formal de variable aleatoria*.

  De acuerdo a la noción que esperamos en el nivel que trabajaremos, esperamos dificultades en nociones tales como experimento aleatorio, probabilidad, la variable aleatoria y su distribución de probabilidad, como lo muestran diversas investigaciones en investigaciones realizadas en niveles escolares básicos y en bachillerato.

3.3.2 Variables de diseño

Cada una de las variables que analizaremos se centra en una concepción o un procedimiento del estudiante que deseamos explorar. Estas variables determinarían el diseño del protocolo. Este se planteó en forma de un cuestionario que girara alrededor de un problema, de manera que las preguntas, planeadas y definidas por las variables de diseño y las hipótesis estructuradas, retomaran y profundizaran los conceptos utilizados y los supuestos y las estrategias empleadas para resolver el problema por un grupo de estudiantes. Las variables se organizaron en cuatro grandes grupos, dependiendo de los conceptos matemáticos en que nos interesamos, que fueron los siguientes:

- Lo aleatorio
- La noción de probabilidad

[6] La noción de variable aleatoria que se pretende con los estudiantes de ciencias sociales se desglosa en el punto 3.1. Ver figura 1.

- La noción de variable aleatoria
- La solución del problema

3.3.2.1 Lo aleatorio

En este objeto incluiremos tanto la incertidumbre, en el sentido coloquial de la imposibilidad de predecir con exactitud por la intervención del azar, como la aleatoriedad propiamente, como un atributo matemático establecido por la sucesión de resultados de un experimento realizado repetida e independientemente, que proporciona, además, la factibilidad de poder ser analizado a través del cálculo de probabilidades (Batanero, 2001).

La actividad forzosamente tendrá que estar sumergida en un contexto de aleatoriedad. La aleatoriedad quedará definida por el mismo contexto del problema, de modo que esperamos no haya cuestionamientos a este respecto. Es de interés la concepción del alumno sobre la aleatoriedad en la medida que esperamos que esta concepción influya sobre la que tenga de variable aleatoria. Sin embargo no será para nosotros tan importante la formalización que sea capaz de hacer de la aleatoriedad, como su noción intuitiva, en el sentido de la intervención del azar, y la relación que establezca entre la incertidumbre y su decisión y con la variable aleatoria.

De esta forma, las variables de diseño dentro de esta categoría serán la definición de un experimento aleatorio, la intervención del azar en el problema trabajado y el cuestionamiento acerca de que manera está presente la incertidumbre en la función de probabilidad. En la tabla 1 presentamos el concepto matemático por el que nos interesamos en la primera columna y en la segunda las variables objetos de evaluación respecto a este objeto, que se refieren a concepciones y procedimientos de los estudiantes. La última columna indica la forma en que la variable se presentará en el protocolo de la entrevista.

Tabla 1. Variables relacionadas con el objeto aleatoriedad

Objeto de observación	Procedimientos y concepciones a evaluar	En la actividad
Experimento aleatorio	Concepción de experimento aleatorio La presencia de la aleatoriedad en el fenómeno aleatorio	Presentación de un fenómeno aleatorio Interpretación de la ocurrencia de eventos

Incertidumbre	La intervención de la incertidumbre dentro de una toma de decisión: • Noción de riesgo • La incertidumbre de la decisión • La medida de la incertidumbre	Resolución de un problema ligado a la recomendación de una toma de decisión: • ¿Cómo se fundamenta la decisión? • ¿Qué tanto se puede asegurar que la decisión será certera?
La incertidumbre en la función de probabilidad	La función de probabilidad como modelo matemático que resuelve el problema: • Definición e interpretación de las variables que intervienen en la función de probabilidad • ¿Qué variable se relaciona con la aleatoriedad dentro de la función matemática establecida?	Definición de la función de probabilidad representativa de la situación. • ¿Qué información proporciona la variable dependiente (probabilidad) de la variable independiente (variable aleatoria)?

3.3.2.2 La noción de probabilidad

Dentro de este rubro la probabilidad será importante en la medida que el alumno la relacione con la variable aleatoria, de manera que llegue a establecer y cuestionar los elementos que definen a la distribución de probabilidad como función. Unimos a esto una exploración de cómo interviene la noción de probabilidad que tiene el estudiante en la apropiación del concepto de variable aleatoria.

La concepción de probabilidad seleccionada en la actividad será la noción clásica porque es la que menos conflictos epistémicos presenta y porque esperamos que sea la noción con la que los estudiantes de nuevo ingreso a la universidad estén más familiarizados. Es de nuestro interés que la actividad explore la interpretación y el manejo de la probabilidad por los estudiantes así como la forma de cálculo de esta probabilidad en tres niveles principales: a un nivel intuitivo en eventos aislados (dentro de la situación); a un nivel intuitivo en la situación conjunta y su vinculación con la función de probabilidad; y a un nivel más formal en la función de probabilidad. De manera que las variables de diseño en este rubro quedarán definidas por esos tres momentos (Tabla 2).

Tabla 2. Variables relacionadas con el objeto probabilidad

Objeto de observación	**Procedimientos y concepciones a evaluar**	**En la actividad**
La probabilidad de un evento aislado	Uso de la regla de Laplace Interpretación de la probabilidad de un evento: • La probabilidad de un evento como medida de la incertidumbre asociada al evento	Cálculo de la probabilidad de eventos asilados dentro del fenómeno aleatorio dado Interpretación de la probabilidad de un evento
La distribución de probabilidad en el experimento aleatorio	Definición del espacio muestral y de los eventos posibles asociados a la característica de interés dentro del problema Uso de la distribución de probabilidad asociadas al experimento aleatorio en la situación completa: • Definición de espacio muestral y de los eventos posibles • Definición de los valores posibles de la variable aleatoria • Asignación de un valor de probabilidad a todos los valores de la variable aleatoria • Calculo de la función de distribución acumulada y su interpretación • Uso del contexto tabular Recomendación asociada a la toma de decisión: • Alcance de la decisión • Medida del riesgo	Definición del rango en la variable aleatoria Cálculo de la probabilidad puntual y acumulada asociada a los valores posibles de dicho rango Organización de la información en una tabla Interpretación de la función de probabilidad y de la función de distribución acumulada en el contexto del problema Resolución de un problema ligado a la recomendación de una toma de decisión: • ¿Cómo fundamenta la decisión? • ¿Qué tanto puede asegurar que la decisión será certera?
La distribución de probabilidad como función matemática	Concepción de distribución de probabilidad como función (variable dependiente) asociada a una variable aleatoria (variable independiente) Identificación de la distribución de probabilidad como una función y su interpretación en un contexto matemático: • Tránsito al contexto gráfico • Definición de la distribución de probabilidad como función matemática • Los elementos matemáticos de la función de probabilidad: su notación funcional, su dominio y su rango y su regla de correspondencia	Identificación de las variables que intervienen en la modelación mediante una variable aleatoria del fenómeno aleatorio: • ¿Existe una relación de dependencia entre ellas? • ¿Cuál de ellas es la variable independiente? ¿Cuál la dependiente? Las variables que intervienen en el problema dentro de un contexto matemático: • ¿Estamos trabajando con una función matemática? • La gráfica de la función • Definición de los elementos de la función: nomenclatura, dominio, y rango

3.3.2.3 La noción de variable aleatoria

Al considerar la distribución de probabilidad como función, decidimos analizar el concepto variable aleatoria porque en ella se circunscribe nuestra investigación. Es de esperar que los estudiantes no manejen el concepto formal de variable aleatoria, de manera que es nuestra intención centrar la actividad en la exploración de su noción intuitiva y una posible apropiación de la noción desde la perspectiva de lo que esperamos en este nivel. Es de nuestro interés la exploración de la concepción intuitiva de la variable aleatoria en la solución el problema, así como el nivel de dificultad que involucra su tratamiento formal tanto dentro de la función de probabilidad como a ella misma definida como la función que relaciona el espacio muestral con un valor numérico. Las variables de diseño contemplan estos tres puntos (Tabla 3).

Tabla 3. Variables relacionadas con el objeto variable aleatoria

Objeto de observación	Procedimientos y concepciones a evaluar	En la actividad
La variable aleatoria en la solución del problema	Manejo de la variable aleatoria con miras a la toma de decisión recomendada: • Asociación de los valores de probabilidad con los valores de la variable aleatoria (distribución de probabilidad) • Asociación de los valores de la probabilidad acumulada con los valores de la variable aleatoria (función de distribución acumulada) • Uso de las distribuciones de probabilidades puntual y acumulada • Uso e interpretación de las medidas de tendencia central de la variable aleatoria	Cálculo de la probabilidad puntual y acumulada asociada al problema. Interpretación de la probabilidad puntual y acumulada en el contexto del problema Resolución de un problema ligado a la recomendación de una toma de decisión: • ¿Cómo se fundamenta la decisión? • ¿Qué tanto puede asegurar que la decisión será certera?
La variable aleatoria y la distribución de probabilidad	La variable aleatoria en la función de probabilidad: • La variable aleatoria como variable independiente en la distribución de probabilidad • Notación • Valores que toma • Como variable independiente en la distribución de probabilidad • Vinculación entre la variable independiente y la aleatoriedad de la situación	Identificación de la variable aleatoria como variable independiente: • ¿Cuál es la variable dependiente en la función definida por una distribución de probabilidad: la probabilidad o la variable aleatoria? • ¿Qué información proporciona la variable independiente (variable aleatoria) sobre la variable dependiente (probabilidad) en la distribución de probabilidad?

		La variable aleatoria en la función (distribución) de probabilidad: • Describe del método a partir del cuál se obtiene la variable dependiente (probabilidad) a partir de la independiente (variable aleatoria) • Definición de los elementos de la variable aleatoria: nomenclatura, dominio y características del dominio • ¿Cuál es la diferencia entre esta variable independiente con las variables independientes que has manejado en tus cursos de matemáticas? • ¿Qué nombre le darías a la variable independiente en la función de distribución?
Concepción de la variable aleatoria como función sobre el espacio muestral	Manejo de la partición del espacio muestral en términos de la característica de interés Clasificación de los eventos que subdividen al espacio muestral en relación a la variable numérica. Familiaridad con la que se maneja un número en lugar de un contexto Identificación de la vinculación de la característica de interés con los eventos en los que se subdivide el espacio muestral y con los valores numéricos	Definición del espacio muestral y de los eventos posibles asociados a la característica de interés dentro del problema Interpretación de la partición del espacio muestral en términos de la característica de interés Interpretación de las probabilidades puntuales que definen al fenómeno aleatorio y las probabilidades puntuales y acumuladas que definen la variable aleatoria Justificación de la recomendación de la toma de decisión ¿Cuál es la diferencia entre la variable independiente en la variable aleatoria, considerada como función con las variables independientes que has manejado en tus cursos de matemáticas? ¿Qué nombre le darías a la variable independiente en la variable aleatoria?

3.3.2.4 La solución del problema

Alrededor del problema girará el protocolo de la entrevista, de manera que a la vez que constituya un reto accesible para los estudiantes, también debe permitir la vinculación e interacción de las otras tres categorías de diseño. Por lo tanto, el problema seleccionado será de una dificultad media para permitir la concentración en la exploración cognitiva, pero lo suficientemente rico como para que permita la exploración de las concepciones de los estudiantes. También deberá cumplir con las siguientes características:

- Manejar una situación en donde se involucre el azar de manera natural.
- La noción de probabilidad vinculada al problema será la noción clásica.
- Su solución estará vinculada a la recomendación de una toma de decisión.
- Su solución requerirá el análisis de la situación completa y por lo tanto del manejo de una función de probabilidad.
- El riesgo en la toma de decisión podrá ser cuantificable.

Se espera que la solución del problema pase por las etapas de solución representadas en la figura 2.

En el problema se pretende vincular los conceptos matemáticos que constituyen las variables de interés de esta investigación, pero también que se apliquen algunas estrategias sencillas de resolución de problemas. Así por ejemplo, esperamos que la verificación de la propuesta a partir del cuestionamiento del riesgo y de la visualización del alcance de la propuesta dada por los estudiantes, despierte cuestionamientos a su modelo planteado. Esto dará pié a redefinir o afianzarse en su propuesta basándose en argumentos matemáticos, proporcionados por las herramientas matemáticas que ponen en juego en esas etapas: la aleatoriedad, la función de probabilidad y la variable aleatoria.

3.3.3 El protocolo

Para realizar la exploración cognitiva se consideró apropiado tomar como base el problema que Castillo y Gómez (1998, p. 120) proponen para introducir el tema de variable aleatoria en su libro de texto Estadística Inferencial Básica, porque permitiría la introducción natural de las variables de diseño. Alrededor del problema se diseñó un cuestionario con preguntas abiertas, que, a su vez, guía la solución del problema y conduce a los entrevistadores en la exploración de los conocimientos de los estudiantes. La actividad así diseñada quedó conformada por tres secciones, cada una con un objetivo específico, unidas por su contribución al objetivo general de la actividad.

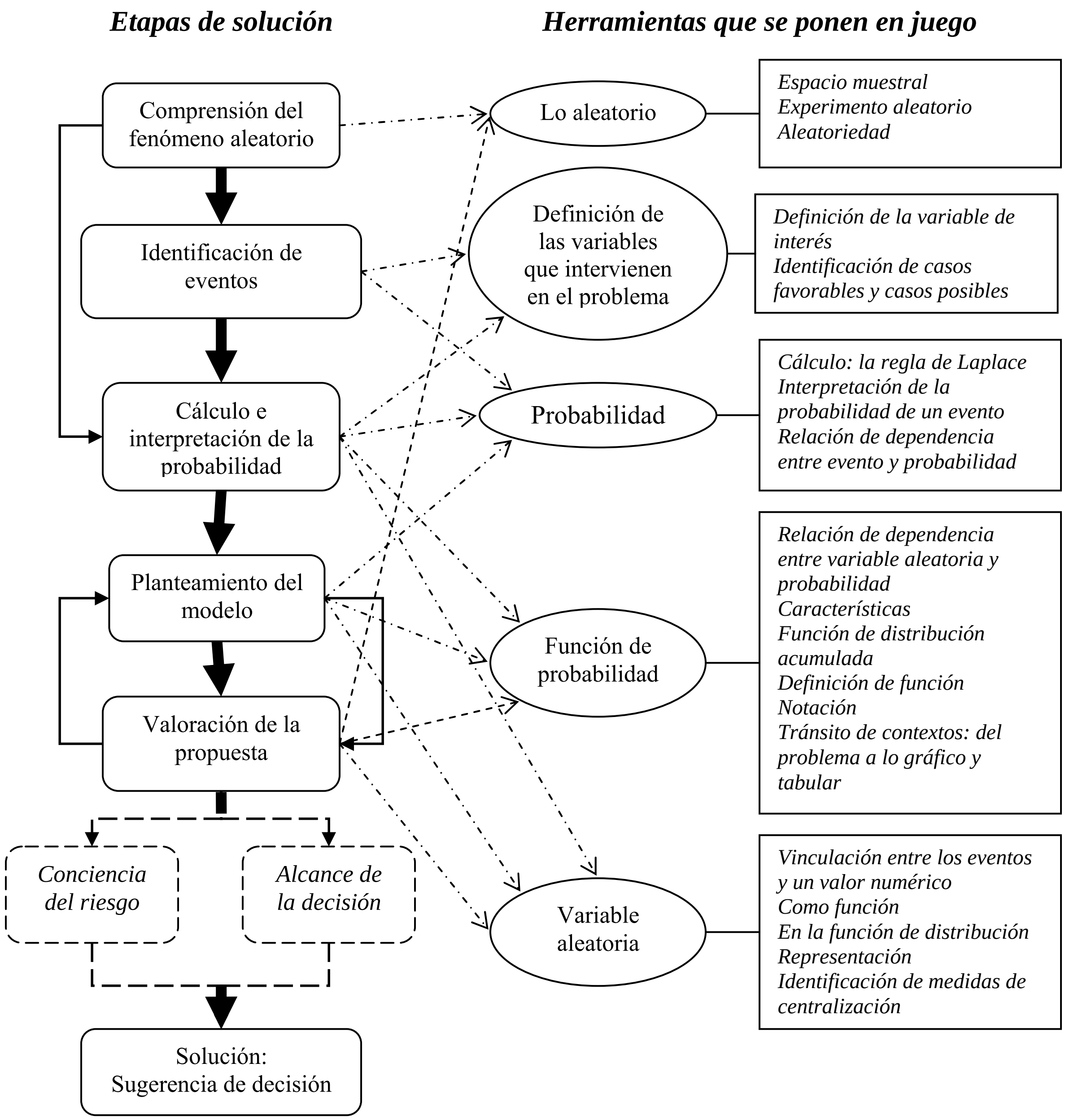


Figura 2. Etapas y herramientas esperadas en la solución del problema

La primera parte de la actividad es una introducción al problema y una exploración de la concepción de probabilidad puntual de los estudiantes. La segunda es propiamente “El problema”, constituye la parte central de la actividad y en torno a la que giran las otras dos secciones de la actividad. En la tercera parte se pretendió profundizar sobre los conceptos formales alrededor de la variable aleatoria y para ello se retomó el problema resuelto en la segunda sección. De manera que en cada una de las partes de la actividad los estudiantes se enfrentaran a diferentes cuestiones alrededor de la noción de variable aleatoria:

PARTE I: Introducción al problema. En esta parte se introduce el problema, se cuestiona a los estudiantes acerca de su noción de probabilidad teórica clásica (dada por la regla de Laplace); se cuestiona la relación entre probabilidad y los eventos y se pide la definición de las variables que intervienen.

PARTE II: El problema. Se les pide a los estudiantes que tomen una decisión, y que sustenten su posición. Para ello deberán exponer sus ideas acerca de riesgo y de probabilidad acumulada. También son cuestionadas sus concepciones acerca de espacio muestral y experimento aleatorio.

PARTE III: Cambio de registro de tabular a gráfico. También se explora su noción de variable aleatoria. Al final se pide una comparación entre situaciones estocásticas y deterministas.

La actividad completa se puede consultar en el Anexo 2. A continuación describimos el problema utilizado alrededor del cual se establece el protocolo y se genera la actividad, su solución y cómo vive la variable aleatoria en él.

3.3.3.1 El problema

El problema trabajado en la entrevista es el siguiente (Castillo y Gómez, 1998, p.120):

A raíz de los festejos del día del niño, el departamento de relaciones públicas de una fábrica desea efectuar una rifa que beneficie a los hijos de los trabajadores. Al obrero ganador se le premiará con boletos para que vaya al teatro con su familia completa. La intención es que la rifa beneficie a los hijos de los trabajadores.

Así que encomiendan a la trabajadora social de la empresa que decida cuántos boletos tiene que comprar para que se efectúe la rifa de manera exitosa, pero la empresa pierda lo menos posible. La distribución del número del número de hijos por trabajador se proporciona en la siguiente tabla 4.

Número de hijos	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Número de trabajadores que tienen ese número de niños	*16*	*22*	*33*	*45*	*31*	*20*	*12*	*9*	*7*	*5*

Tabla 4. Histograma de frecuencias de los trabajadores con respecto al número de hijos que tienen

El análisis que realice el alumno debe estar enmarcado en el contexto de esa fábrica y debe tomar en cuenta que es un festejo infantil y que la empresa no desea despilfarrar recursos. Así mismo existe la limitante de que los boletos se tienen que comprar antes de la rifa. Esta última condición no es explícita, sin embargo esperamos que los estudiantes la consideren porque es de esperar que al momento de la rifa se tengan los boletos para ser entregados al ganador.

La dificultad de este problema estriba en que en el momento de comprar los boletos, no se sabe cuántos hijos tiene la familia del obrero que saldrá ganador, puesto que no se sabe quién será el ganador. La solución oscila entre dos criterios de decisión:

- ❖ Por un lado la empresa se preocupa por que el trabajador ganador tenga posibilidades de llevar a toda su familia a la función de teatro. Debe tener boletos suficientes.
- ❖ Por otro lado, la empresa desea optimizar sus recursos. No quiere que sobren boletos.

Podemos suponer que todos los trabajadores de la fábrica entre los cuales se efectuará la rifa son casados, de modo que el número de miembros de la familia del trabajador estará determinado por el número de hijos que éste tenga. Así el número mínimo de miembros de una familia de esa fábrica será 2 (la pareja) y el número máximo de miembros será 11 (9 es el número máximo de hijos que tiene una pareja en esa fábrica). Los recursos gastados por la empresa estarán dados por el número de boletos que se compren, puesto que no nos proporcionan el precio de los boletos. No importa si los boletos de adultos y niños cuestan lo mismo o no porque lo que estaría en cuestión es cuántos boletos para los niños compraríamos (porque los adultos son siempre dos), pero para facilitar el razonamiento del problema supondremos que tanto los boletos de los niños como los de los adultos cuestan lo mismo.

La solución puede irse a uno o a otro extremo entre los dos criterios de decisión, pero el objetivo es tomar en cuenta los dos criterios y crear un ambiente de incertidumbre. Si se deja de tomar en cuenta uno de los dos criterios y sólo se toma en cuenta uno, las soluciones resultan triviales y determinísticas:

- ❖ **Se comprarían 11 boletos** si se quiere asegurar que el trabajador ganador lleve a toda su familia: dos para los papás y nueve para los hijos. Puesto que el número máximo de hijos que puede tener el trabajador ganador son nueve.

- ❖ **Se compran 2 boletos** para optimizar al máximo los recursos de la empresa. Estamos seguros que se ocuparían los dos boletos, puesto que todas las familias constan de al menos dos miembros: la pareja.

Pero la empresa quiere realizar esa rifa optimizando el dinero que invertirá en ello y al mismo tiempo quiere quedar bien con sus trabajadores (al menos con el ganador), que es diferente a que a la empresa no le importe cuanto se gasta o que la empresa no estuviera interesada en que el trabajador vaya con toda su familia. Plantear el problema con un solo criterio de decisión le quitaría el elemento de incertidumbre, lo mismo que si se supiera quién cuál será el nombre del trabador extraído de la urna. El estar entre los dos criterios de decisión introduce la incertidumbre y por lo tanto la necesidad de hacer uso de la herramienta matemática descrita en el punto 3.3.24 y en la figura 2. La aleatoriedad se introduce en la decisión cuando relacionamos los boletos a comprar con el número de hijos del trabajador extraído de la urna (que es el resultado del experimento aleatorio) y no con el número máximo o mínimo de hijos que tienen los trabajadores de la fábrica porque en éstos últimos resultados no hay intervención del azar, ya que en las condiciones de la fábrica, son fijos. De esta forma la decisión está condicionada por la incertidumbre si la rifa es aleatoria y no está «arreglada». Al ubicar la decisión en un contexto de incertidumbre es factible analizar el problema desde la teoría de la probabilidad.

Este tipo de problemas, que oscilan entre dos criterios de decisión son comunes cuando se pretende optimizar recursos. Proporcionamos a continuación dos ejemplos de situaciones problema que requieren una herramienta matemática más complicada que la que empleamos en esta investigación, pero que nos sirven para resaltar la importancia de este tipo de problemas:

- ❖ **En un restaurante.** Cuando se pretende decidir cuánta comida hacer. El elemento aleatorio está en no saber cuánta gente solicitará el servicio de comida. Podemos suponer que el restaurante tiene información acerca del comportamiento de sus clientes y puede esperar un número máximo de clientes y un número mínimo. El restaurante podría hacer una cantidad máxima de comida, pero casi con seguridad se le echaría a perder si no llega la cantidad máxima de clientes, lo cual repercutiría en su economía (tiene que optimizar). El otro extremo es comprar un número mínimo de comida, lo cual repercutiría en la reputación sobre el servicio que ofrece (tiene que brindar un servicio para atraer o conservar a sus clientes). La situación tiene elementos aleatorios y su solución oscila entre brindar un servicio porque eso atraería o le haría conservar a sus clientes y optimizar sus recursos.

- **Una compañía de autobuses**. Cuando se pretende decidir el número y horario de corridas que harán sus autobuses por día en un cierto recorrido (por ejemplo Monterrey-Distrito Federal). En tal caso la compañía de autobuses debe establecer sus horarios y corridas de acuerdo a la información que tiene sobre las necesidades de sus clientes. Por ejemplo, una corrida cada dos horas o tres corridas al día en distintos horarios. Una vez determinada una corrida se abre la venta de los boletos. Los boletos generalmente se venden con antelación, de modo que el autobús debe emprender su recorrido con el número de pasajeros que hayan comprado boletos hasta el momento de su salida. De nuevo hay la necesidad de tomar una decisión que oscilará entre dos criterios, el de brindar sus servicios (y por lo tanto no quedar mal con sus clientes, porque eso la desprestigiaría) y optimizar sus recursos planeando sus corridas de modo que los autobuses vayan lo más llenos que se pueda. Lo aleatorio está en no saber cuántos boletos se venderán y esa aleatoriedad influirá en la decisión del número de recorridos Monterrey-D. F. que ofrecerá la compañía.

3.3.3.2 Solución del problema

De acuerdo con el punto anterior, la solución del problema requiere del análisis de la distribución de probabilidad de que el trabajador premiado tenga un cierto número de hijos y de su función de distribución acumulada. Ambas se proporcionan en la tabla 5.

Tabla 5. Distribución de probabilidad y función de distribución acumulada del número del número de hijos del trabajador premiado

Número de hijos (x)	**Número de trabajadores**	**Distribución de Probabilidad**	**Función de distribución**
		Probabilidad de que el trabajador premiado tenga *x* número de hijos $p(x)=P(X=x)$	**Probabilidad de que al trabajador premiado le alcancen los boletos** $F(x)=P(X\le x)$
0	16	0.080	0.080
1	22	0.110	0.190
2	33	0.165	0.355
3	45	0.225	0.580
4	31	0.155	0.735
5	20	0.100	0.835
6	12	0.060	0.895
7	9	0.045	0.940
8	7	0.035	0.975
9	5	0.025	1.000

De acuerdo con la tabla, lo más probable es que el trabajador seleccionado tenga tres hijos o menos. Es decir, lo recomendable es que la trabajadora compre 5 boletos (2 para la pareja y 3 para los hijos), de esa manera la directiva de la empresa estará tendrá una probabilidad de 0.58 de que los boletos le alcancen al trabajador premiado y de que sobren los menos posibles.

La solución detallada del problema, junto con la solución de la actividad (protocolo) propuesta a los estudiantes desglosa en el Anexo 3.

En cambio hay otras nociones en las que la exploración cognitiva sí tiene objetivos explícitos de definición y cuestionamiento por parte de los estudiantes dentro del problema, como lo son el experimento aleatorio, el espacio muestral, la magnitud de interés, los posibles valores de la magnitud en el experimento, la partición que define en el espacio muestral y la variable aleatoria con sus tres componentes (dominio, regla de correspondencia y rango). Respecto a estos conceptos, aparte de las dificultades que los estudiantes puedan tener, nos interesamos por el modo en que los definan dentro del contexto del problema. De manera ideal, en el contexto del problema quedan definidas de la siguiente manera:

Experimento aleatorio:	Mezclar en una tómbola papeles con el nombre del trabajador y sacar al azar el nombre de uno de ellos, que será el ganador.
Espacio muestral:	Nombres de los trabajadores que laboran en esa fábrica. Hay 200 sucesos elementales equiprobables.
Característica de interés:	Hijos que tienen los trabajadores.
Variable aleatoria:	*Dominio*: sucesos elementales en el experimento, trabajadores factibles de que reciban el premio.
	Regla de correspondencia: Número de hijos que tiene cada trabajador de la fábrica.
	Rango (valores que toma): números 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.
Partición del espacio muestral:	Conjunto de trabajadores que tienen el mismo número de hijos.

La forma en que estas nociones se vinculan, y la forma en que vive la variable aleatoria en el problema se muestra en la figura 3. En esta figura se observa como se generan dos planos a partir del sorteo, en el de arriba se analiza el problema desde el contexto del sorteo en una fábrica. En la parte inferior se analiza la perspectiva matemática del problema. Ambos planos se retroalimentan uno al otro y definen la recomendación en el contexto del problema con argumentos matemáticos. A continuación describimos la forma en como se refleja este proceso en la figura 3.

El modelo matemático está condicionado por la magnitud de interés del problema. Para obtener esta pregunta es necesario hacer un razonamiento regresivo, es decir, enfocando el problema a la inversa, comenzando por lo que queremos alcanzar (Polya, 1965). Suponemos que ya compramos los boletos y que ya efectuamos la rifa. El resultado sería el nombre del trabajador ganador, ¿qué haríamos después? Nos preguntaríamos si le alcanzarán los boletos que ya compramos. De aquí, pensaríamos cuál es la característica de ese trabajador que hace que se responda la pregunta. Nos interesa fijarnos en los miembros de su familia, en particular, cuántos hijos tiene. Esa característica permitirá agrupar el espacio muestral en eventos compuestos de interés y asignar probabilidades a esos eventos. Al mismo tiempo, define la variable aleatoria que asignará un valor numérico a cada evento elemental del espacio muestral.

Observemos la cercanía entre la característica de interés y de la variable aleatoria y no obstante hacemos énfasis en la existencia de las dos. Esto es porque la característica de interés podría ser cualitativa, por ejemplo, si el trabajador premiado es hombre o mujer, o el medio de transporte que usa para llegar al trabajo. Es el atributo del resultado del experimento aleatorio que interesa de acuerdo al contexto del problema y va a ser el que va a permitir la partición del espacio muestral. La variable aleatoria tiene que tener naturaleza cualitativa forzosamente y tiene que transformar el resultado del experimento aleatorio en un número. Por supuesto, la partición del espacio muestral y la variable aleatoria tienen que estar vinculadas.

Una vez definida la variable aleatoria y sus valores, la función de probabilidad se obtiene al aplicar la función inversa de la variable aleatoria y asignar una probabilidad a cada valor de la variable aleatoria. Los pasos siguientes, en la solución del problema será hacer uso de las funciones de probabilidad y de distribución para poder proporcionar una solución de acuerdo al riesgo de que falten, sobren o queden exactos el número de boletos a la familia del trabajador ganador en el sorteo.

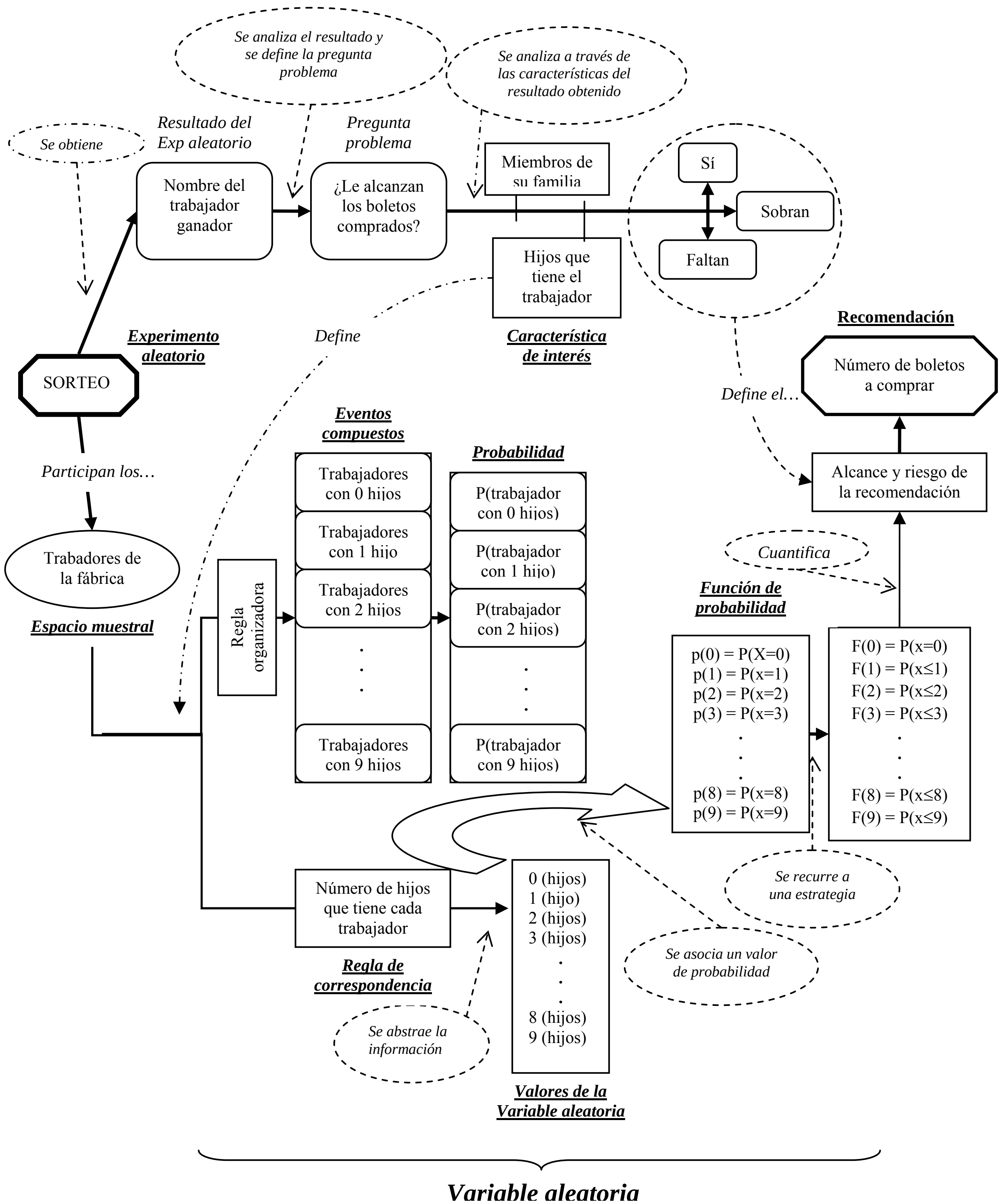


Figura 3. Representación de la relación entre los elementos teóricos vinculados con la variable aleatoria y la solución del problema

En el diagrama se observa como la variable aleatoria está condicionada por el contexto del problema, pero hace posible la utilización de herramienta matemática para cuantificar el riesgo de la solución que se proponga. Una vez comprendiendo y delimitando el problema, la regla de correspondencia de la variable aleatoria surge de manera natural. Las delimitaciones se refieren a eliminar otras posibilidades distintas a la de una familia común, como el trabajador sea divorciado o viudo, o que se considere como parte de la familia del trabajador a la abuela materna. Las suposiciones son necesarias para poder establecer un modelo matemático. El que la variable aleatoria (condicionada por el contexto del problema) vincule el espacio muestral con valores numéricos, hará posible que ese modelo matemático sea la función de probabilidad y/o la función de distribución, aunque la variable aleatoria en sí misma sea un modelo matemático en otro estrato de la estructura de la modelación que propone Heitele (1975).

3.4 Implementación

Se entrevistaron a dos estudiantes, Brenda y Mónica, que accedieron de manera voluntaria a hacerlo y que cursaban el primer semestre de universidad en el ITESM Campus Monterrey. Ninguna de las dos había llevado cursos de probabilidad o estadística en ese nivel. Una de ellas, Brenda, manifestó haber tomado un curso de probabilidad y estadística en el bachillerato y la otra, Mónica, no. En el curso que había tomado, Brenda llevó estadística descriptiva y reglas de probabilidad. No conocía el término de «variable aleatoria». En el momento de la entrevista, ambas cursaban matemáticas I para ciencias sociales (Cálculo diferencial) y eran consideradas buenas estudiantes por su profesor de matemáticas por el desempeño que tuvieron en su clase de ese semestre.

Se les entregó el problema por escrito, una sola copia para ambas, y se les pidió que trabajaran directamente sobre el pizarrón para que sus respuestas fueran abiertas entre ellas y a los entrevistadores (su profesor de matemáticas y la profesora investigadora). Así mismo que discutieran y proporcionaran una sola respuesta entre ambas.

La entrevista se dividió en las tres partes de que consta la actividad y tuvo una duración aproximada de tres horas de tiempo efectivo. A las estudiantes se les daba un tiempo en el que llegaban a una respuesta conjunta a cada inciso, una vez que ellas consideraban concluida su respuesta, los entrevistadores dialogaban con ellas para profundizar o esclarecer las ideas que exponían. Al concluir cada sección de la actividad, ellas borraban y pasaban sus resultados a una hoja de papel, de manera que su reporte estuvo constituido únicamente por sus conclusiones, incluso algunas respuestas las contestaron sólo verbalmente.

Se grabaron (tanto en video como en casete) ambas partes de la experiencia, tanto las discusiones entre ellas para llegar a su respuesta, como los diálogos establecidos con los profesores. La trascripción de las partes principales del diálogo sobre el desarrollo de la actividad se encuentra en el Anexo 4 de este trabajo.

Capítulo 4

Análisis de los resultados de la entrevista clínica

El análisis de los resultados de la entrevista clínica se llevó a cabo tomando como base las hipótesis y las variables de diseño de la actividad cognitiva, pero dirigido a explorar las nociones y dificultades que esperamos que los estudiantes tengan al enfrentarse con el problema propuesto relacionado con la variable aleatoria. Así, los objetos de análisis son esencialmente los mismos que los de diseño, pero las variables de análisis se modifican enfatizando el objetivo de la exploración cognitiva y se descartan cuestiones que no giran alrededor de las nociones relativas a la variable aleatoria o cuestiones del diseño que no tenían la intención explícita de ser observadas.

En lo que sigue presentamos el análisis realizado y los resultados obtenidos, organizados en torno a los objetos de estudio y con la descripción de cada variable de diseño. También presentamos los pasajes de la entrevista relacionados con cada una de las variables principales, analizando los conceptos matemáticos puestos en juego por el alumno en relación a dicha variable, las concepciones que se deducen y sus dificultades al resolver la tarea. Así, los pasajes que a continuación se presentan no están descritos cronológicamente y dentro de ellos se incluyen entre paréntesis observaciones de transcripción para facilitar la comprensión de los pasajes. La entrevista completa se presenta en el anexo 4.

Al final de este capítulo hacemos también, un análisis del protocolo la entrevista clínica tomando en cuenta los resultados obtenidos y el análisis de ellos. En ella podremos hacer notar qué tan efectivo resultó tanto el problema y la forma en que fue planteado como el cuestionario para cumplir el objetivo con el que fue planteado.

4.1 Lo aleatorio

Este apartado se centra en el estudio de las concepciones del estudiante sobre la idea de aleatoriedad y su reacción ante la presencia de la incertidumbre, tanto en el experimento como en su recomendación de cuántos boletos comprar. También interesa cómo vinculan la función de probabilidad con la incertidumbre. Esto es, se busca conocer el proceso que guía al estudiante a la utilización de términos cuantitativos en la solución de un problema en el que interviene el azar, a la vez que es deseable conocer la forma en que el estudiante redefine la situación de incertidumbre una vez que ésta está representada a través de términos matemáticos como la función de probabilidad (Figura 4). Preferimos tomar el aspecto informal de la aleatoriedad porque desde la perspectiva de la modelación de Heitele (1975) en este estrato, el experimento aleatorio es la «realidad», en cambio la función de probabilidad (o la función de distribución) es el «modelo matemático» que se está empleando para resolver el problema. Así mismo nos interesa la función de probabilidad porque en ella la probabilidad se vincula con la variable aleatoria.

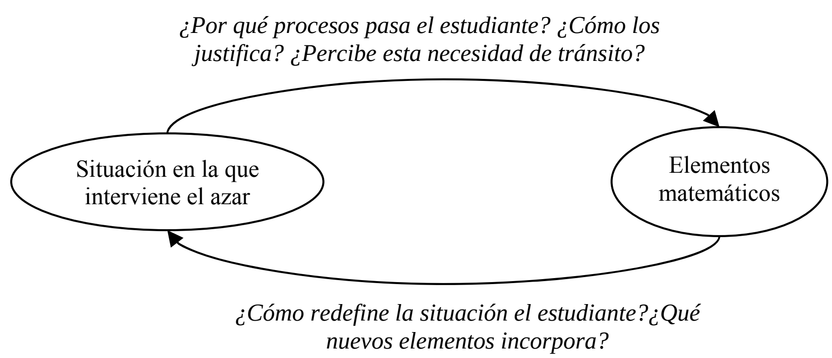


Figura 4. Descripción del objeto de análisis 'Lo aleatorio'

Así, las variables que analizaremos, que se desglosan con más detalle en la Tabla 6, quedarán definidas por:

- *Noción de aleatoriedad en la situación problema*. Atendiendo a la percepción de la aleatoriedad, interesa la forma en que la tiene en cuenta y cómo el estudiante justifica numéricamente su recomendación. También es de provecho observar las dificultades en este proceso.

- *La incertidumbre y la probabilidad*. Refiriéndonos a cómo el estudiante relaciona estos conceptos, identifica y redefine la incertidumbre con la función de probabilidad e interpreta una representación matemática de un problema en el que interviene la incertidumbre.

Tabla 6. Variables de observación asociadas al objeto 'Lo aleatorio'

Variables de observación	Desglose
Noción de aleatoriedad en la situación problema	Idea de aleatoriedad en el fenómeno aleatorio La noción de aleatoriedad en la recomendación La incertidumbre dentro de la recomendación Dificultades asociadas
La incertidumbre y la probabilidad	La incertidumbre y la noción de probabilidad La incertidumbre y la distribución de probabilidad involucrada en el proceso de solución del problema. Dificultades asociadas

4.1.1 Resultados en el objeto 'Lo aleatorio'

4.1.1.1 Aleatoriedad

En un primer enfrentamiento con el problema, hay una tendencia a dar una recomendación determinista (pasaje 8), principalmente por parte de Mónica que deja ver su tendencia a dar soluciones certeras, en las que no se involucre el azar. Hay dificultad en vincular el sorteo con la recomendación de cuantos boletos comprar y por lo tanto dejar al azar la definición de cuántos boletos comprar. Mónica lee la pregunta: «*¿Qué recomendación le darías a la trabajadora social de cuántos boletos comprar?*» y la resuelve diciendo que compraría dos (para la pareja) y que después, cuando se sepa cuantos hijos tiene el trabajador que ganó, compraría los demás. No aceptan la aleatoriedad en la recomendación ni hace uso de los datos, sólo quiere resolver el problema de la forma más exacta posible.

> Mónica: Pues yo para no perderle y para no errarle, compraría dos nada más porque sí hay familias que tienen 0 hijos, de hecho hay 16 trabajadores que tienen 0 hijos. Entonces nada mas compraría dos para que vayan él y la esposa, y ya después si no sale, por lo menos ya tengo asegurados dos. Por eso preguntaba, si ya tengo asegurado dos... porque dice que pierda lo menos posible. En caso de que compre 3, y si por ejemplo sale una familia de estas, ya perdí un boleto, ya estoy perdiendo un boleto ¿qué voy a hacer con ese boleto? Así pienso yo. Yo compraría dos.

La aceptación a dar una solución no segura es impuesta por las condiciones del problema: después ya no se podrán comprar boletos.

En el pasaje 11 Mónica (nuevamente) argumenta que hay una decisión que depende de la ética, que no tiene que ver nada con matemáticas y se quejan con el profesor de que el problema esté planteado así. Ella se desespera de tener que jugar un poco con lo incierto, de no poder dar una solución más precisa y, desde su perspectiva más justa.

> Mónica: Yo le pondría un 50% (*refiriéndose más bien al 58% de probabilidad de que el trabajador premiado tenga 3 hijos*) o si acaso un 70% (*refiriéndose al 73% de probabilidad de que el trabajador premiado tenga 4 hijos*), porque luego para que me salgan con que perdieron... aunque claro... maestro también está mal el problema porque no es cuestión de decirles oigan les vamos a dar boletos y si les toca pues que bueno porque es como decirles a los que tengan más (hijos) llevan las de perder porque no vamos a comprar para todos y luego que... Lo ideal sería que si la empresa está haciendo la campaña, pero eso ya no tiene que ver con matemáticas, sino que es más ético, si la empresa está ofreciendo pues que lo ofrezca bien.

En este mismo pasaje, ellas aún se resisten a aceptar que no puedan dar una solución con mayor seguridad, en cierto modo aceptan que no pueden controlar la situación con una manifiesta resignación.

> Mónica: Yo le pondría un 50% (*refiriéndose más bien al 58% de probabilidad de que el trabajador premiado tenga 3 hijos*), o si acaso un 70% (*refiriéndose al 73% de probabilidad de que el trabajador premiado tenga 4 hijos*), porque luego para que me salgan con que perdieron...
> Brenda: En todo caso si se está buscando que no vaya a gastar de más y si alcanzan bueno y si no, no, pues yo estaría por 4 hijos.
> Mónica: Ajá, yo también y si no alcanzan pues ya ni modo.

Posteriormente a lo largo del problema, ambas llegan a aceptar la incertidumbre de la decisión. Esto lo manifiestan en mayor medida en los pasajes 12 y 20:

> Profesor: Pero existirá alguna manera, ¿es ignorancia nuestra el no saber qué boleto va a salir?
> Brenda: No es algo que tú puedas controlar, es la suerte.

Así mismo en el pasaje 19 Brenda plantea una concepción de la aleatoriedad relacionándola con un patrón que no tiene una continuidad numérica:

> Brenda: ...son así como... aleatorios. No llevan un patrón así seguido.

4.1.1.2 La relación que las estudiantes establecen entre la aleatoriedad y la probabilidad

En el pasaje 9 ambas alumnas piensan que la mejor solución a la pregunta sobre cuántos boletos comprar sería 5 boletos, basándose en que esto es «*lo más probable*», por tanto hacen uso explícito de la idea de probabilidad. Están aplicando una concepción laplaciana, puesto que se basan en la consideración de casos favorables y posibles y lo vinculan con los datos de la tabla. En su recomendación usan el número de hijos que tuviera el mayor número de trabajadores (la moda), apareciendo por tanto la idea de promedio y una estimación de las diferentes probabilidades. Se introduce la condicionante de que no pueden comprar boletos después de la rifa, entonces Mónica está de acuerdo con Brenda en comprar 5 boletos. Seleccionan «*lo más probable*» porque es lo que más seguridad les da. De alguna manera interpretan la probabilidad como una forma de medir la incertidumbre

> Brenda: Yo compraría 5, porque de acuerdo a la tabla, hay más probabilidad de que sea una persona con tres hijos y luego sumándole el obrero y la esposa serían 5.
>
> ...
>
> Mónica: ...¡Ah! entonces compraría 5 boletos porque la mayor incidencia de familias es de 3 hijos, o sea, el mayor número de trabajadores que hay por hijo es 3, o sea, las familias con 3 hijos son las que mas abundan en la empresa, entonces como 3 hijos mas el papá y la mamá pues ya son 5.

En el pasaje 10 aparecen de nuevo estas ideas. Mónica relaciona la intervención del azar con una situación en la que no puedes saber qué ocurrirá (impredecibilidad del azar). Usa la idea de más probable (moda), relacionándola con lo que menos falla. Mónica ahora es la que trata de convencer a Brenda y argumenta a favor de comprar 5 boletos diciendo que es donde hay más probabilidades.

> Mónica: ...hay más probabilidades de que yo agarre un trabajador de estos 45 (*se refiere a los trabajadores que tendrían 3 hijos), o sea, en el círculo que dibujamos (señala la tabla*) éste (*señala el 45*) ocupa mas, hay mas probabilidades, es como que la que menos falla, es como que azar.

En el pasaje 12 Mónica de alguna manera continúa buscando el determinismo en la situación, pretendiendo afianzarse a algo seguro, y relaciona el porcentaje de trabajadores en la fábrica con un número menor o igual de hijos al de boletos comprados con la probabilidad de acertar en la decisión. Aparece aquí la idea implícita de probabilidad acumulada (por tanto función de distribución) y la idea de que la mediana induce una partición de la población en dos grupos con igual número de efectivos, como si con ello estuvieran siendo justas con todos los trabajadores (los que tienen más o menos hijos).

> Mónica: Aquí lo que le aseguramos es el 50%, el otro 50% definitivamente va a tener que comprar uno, dos, tres, cuatro boletos por su cuenta, pero lo que usted quiere también es congraciarse o llevarse bien con los trabajadores, entonces pues con más del 50% ya tiene probabilidad de que lo saque, es decir, tiene asegurada la mitad.

4.1.2 Análisis de resultados en el objeto 'Lo aleatorio'

Las interpretaciones y los pasajes antes anotados se recopilan en la Tabla 7, en donde se visualizan mejor los resultados obtenidos.

Tabla 7. Síntesis de resultados en el objeto 'Lo aleatorio'

Variable	Pasaje	Acción de las estudiantes	Interpretación
Noción de aleatoriedad en la situación problema	8	La primera solución al problema es comprar sólo dos boletos y comprar el resto cuando ya se conozcan los resultados de la rifa	Es una solución determinista, las estudiante quieren evitar dar una solución que no aporte una seguridad
	11	Argumentan que la decisión depende de la ética y no de las matemáticas	Hay una desesperación por tener que recomendar algo que es posible que no ocurra Continua la búsqueda del determinismo
	11, 12, 19 y 20	Manifiestan una solución con incertidumbre con un poco de resignación	Hay una aceptación resignada de la aleatoriedad en la decisión.
La incertidumbre y la probabilidad	9	Piensan que la mejor solución es la relacionada con lo más probable	Vinculan la mayor probabilidad (la moda) con proporcionar mayor seguridad de que eso ocurra.
	10	Mónica proporciona un argumento de porqué se deben comprar 5 boletos	Relacionan lo más probable (la moda) con lo que les proporciona menos incertidumbre
	12	Ellas explican al profesor el porqué de la solución del problema proporcionadas por ellas (comprar 5 boletos) utilizando la mediana de la distribución de probabilidad.	Relacionan el porcentaje de trabajadores con un número de hijos igual o menor a los boletos comprados con el porcentaje de acertar en la decisión y mediar la incertidumbre. Inducen la idea de la partición del espacio muestral en dos grupos con igual número de miembros (mediana) como una medida de equidad entre los trabajadores.

De aquí se puede observar cómo las estudiantes aceptan que no es posible dar una solución certera al problema. En un principio proporcionan una de las soluciones triviales, en la que no habría incertidumbre. Hay una aceptación de la aleatoriedad en la situación puesto que manifiestan no saber cuantos hijos tendrá el trabajador que será premiado pero no aceptan no poder dar una recomendación en la que todos los miembros del trabajador ganador puedan asistir a la función de teatro. Posteriormente, la reticencia a dar una respuesta de la que no están seguras es sustituida por la aceptación de la aleatoriedad en la solución del problema, al mismo tiempo que la relacionan con la probabilidad e interpretan a la probabilidad como una medida de lo incierto, lo que les permite obtener soluciones y transitar por diversas nociones, como la moda y la mediana de la distribución de probabilidad de la variable aleatoria.

Es posible que esta vinculación entre la probabilidad y la aleatoriedad es lo que, de alguna manera, les permita aceptar la aleatoriedad de la solución del problema y por lo tanto proporcionar una solución que les satisface. Puesto que el poder manipular la probabilidad al decidir cuántos boletos comprar es lo que les hace ver cómo pueden ser equitativas o favorecer a la parte que les parece éticamente más apropiada. La aceptación de lo incierto estaría, en tal caso, vinculada con una seguridad de que su solución es de alguna manera más equitativa y justa, ya que no pueden tener certeza, buscan equidad y justicia. Es importante puntualizar que las nociones de equidad y justicia son un primer paso en la construcción de una noción intuitiva de probabilidad (Cañizares et al., 2003).

4.2 La noción de probabilidad

En este apartado es deseable conocer la forma en que la concepción de probabilidad del estudiante influye en su apropiación del concepto de variable aleatoria. En el problema la probabilidad se maneja en dos contextos, la vinculada a los eventos y la vinculada a la variable aleatoria, de modo que las dos vertientes que nos permitirán tener elementos que hagan posible este objetivo son, la forma en cómo el estudiante se relaciona con el problema para definir la probabilidad como función sobre el espacio muestral. Y, segundo, una vez definida la función de probabilidad, profundizar sobre la idea que se genera en el estudiante de esa representación matemática y cómo la vincula en el contexto de incertidumbre.

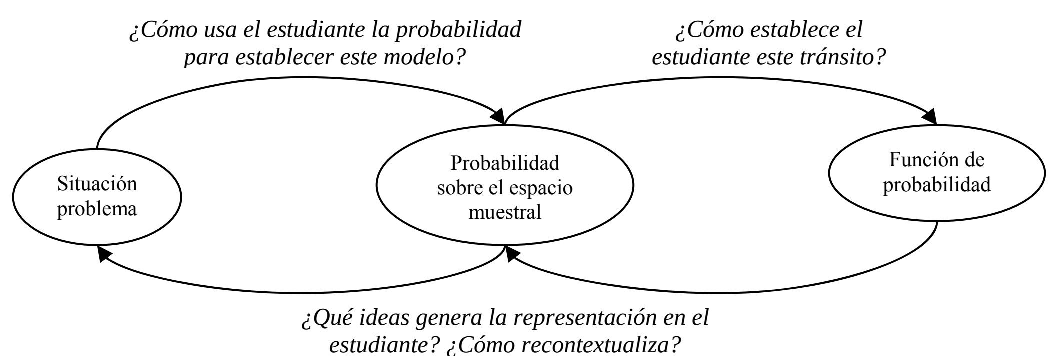


Figura 5. Descripción del objeto de análisis 'Noción de probabilidad'

Así, la noción de probabilidad será explorada a través de las siguientes variables, que se desglosan con más detalle en la Tabla 8:

- *La probabilidad sobre el espacio muestral.* Se busca esclarecer la forma en que el estudiante de manera intuitiva asigna la probabilidad y la función en el espacio muestral del contexto aleatorio.

- *La noción de distribución de probabilidad.* Nos interesa cómo el estudiante asigna la probabilidad a la variable aleatoria y establece la función de probabilidad. Pero también. Una vez establecida la representación matemática de la situación trataremos de analizar la interpretación que el estudiante le da a la distribución de probabilidad para contextualizar de nuevo el problema.

Tabla 8. Variables de observación asociadas a la noción de probabilidad

Variables de observación	Desglose
La probabilidad sobre el espacio muestral	Interpretación de la situación problema en términos del modelo probabilístico: • Definición del espacio muestral y de los eventos posibles • Interpretación de las probabilidades de los eventos posibles Presencia de los axiomas de probabilidad Dificultades asociadas
La función de probabilidad	Identificación de la función de probabilidad: • Los elementos matemáticos de la función probabilidad: su notación funcional, su dominio y su rango • Relación entre los contextos gráfico, numérico y algebraico Noción de función: • Qué es una función • Representación gráfica y algebraica Dificultades asociadas

4.2.1 Resultados en el objeto 'Noción de probabilidad'

4.2.1.1 Probabilidad sobre el espacio muestral

En el pasaje 13 a ambas estudiantes, en la medida que interpretan el problema, les queda claro el espacio muestral del experimento establecido por el sorteo, que sería el conjunto de los trabajadores de la fábrica (200):

Profesor: ¿es posible que pueda extenderse el número 200 a más valores? ¿por qué?
Mónica: No porque sólo son 200 trabajadores

Así mismo en el pasaje 21 se observa como están conscientes de que este espacio muestral está determinado por ciertas circunstancias: una fábrica, en un país, etc., que condicionan un ambiente y que el cambio de esas condiciones afectaría no sólo el espacio muestral sino también la distribución del número de hijos que tendrían los trabajadores.

Mónica: Ajá, porque (*los datos*) pueden estar determinados por diferentes factores por ejemplo, esa población puede haber tenido tres hijos porque se localizaba en México o en Estados Unidos en donde la mayoría de las personas tienen tres hijos, pero en poblaciones de otros lugares donde el número de hijos crece por persona, la gráfica ya no iba a ser igual…

En general se observa una buena interpretación y uso del concepto de probabilidad clásica. Hay evidencia de un buen manejo de dos axiomas de la probabilidad y tienen una interpretación correcta en el contexto del problema, pero tienen dificultades para despegarse de esa interpretación y vincular la probabilidad con un solo número. Así, en el pasaje 4, la probabilidad la expresan a través de la relación entre dos cantidades, «*31 de 200*». El número está escrito como una fracción y lo interpretan como esa relación, pero cuando hablan de ella separan las dos cantidades (pasaje 4):

Profesor: ¿Si tomamos un trabajador al azar cuál es la probabilidad de que tenga 4 hijos?
Ambas: 31 de 200.
Profesor: ¿Por qué no pusieron por ejemplo 4 entre doscientos u otro número porque 31?
Brenda: Porque bueno, en la tabla el número de hijos que son 4 tienen un número de trabajadores de 31 y el universo de trabajadores son 200. Entonces de esos 200 quiere decir que 31 tienen 4 hijos.

Esto repercute en que a pesar de que las alumnas interpretan correctamente el cociente (entre dos variables: número de trabajadores y número total de trabajadores) se les dificulta ver que ese cociente es un solo número y por lo tanto, un solo valor, que sería la probabilidad. Les resulta difícil dejar de situarse en el problema para encontrarse en un contexto matemático, no quieren dejar de «ver» el número de trabajadores. Señalan y hablan del número de trabajadores, pero lo llaman «probabilidad», a pesar de que cuando se les pregunta mencionan que no hay que olvidar el denominador, porque es la forma de relacionar al número de trabajadores con un cierto número de hijos con el total (pasaje 4):

Profesor: Es decir, la probabilidad de tener cuatro hijos...
Brenda: O sea, esta es la probabilidad de 31,... (le arrebata la palabra, señala la respuesta al inciso a), 31/200, escrita a un lado de la tabla).
Brenda: ...entonces estas son todas las probabilidades que hay (señala la columna de número de trabajadores)...
Mónica: A pues sí.

Brenda: ...y si las sumamos sí nos van a dar 200. Haz de cuenta que ésta (señala el 31 en la columna de número de hijos) es la probabilidad de tener 4 hijos. Está diciendo el profe que se refiere a estas probabilidades (señala la columna de número de trabajadores) y al sumarlas nos darían 200.

No se dan cuenta de la contradicción entre el que digan que la probabilidad tiene que ser menor que uno, con el hecho de decir: «*la probabilidad [de tener 4 hijos] es 31*». La confusión entre el número de trabajadores y la probabilidad está más presente en Brenda, pero está más segura de lo que dice que Mónica. Ésta es convencida por Brenda de usar el número de trabajadores como probabilidad. Pareciera que fuera más palpable para ellas la variable 'numero de trabajadores' que el de 'probabilidad'.

A pesar de eso reconocen que la probabilidad tiene que ser menor que uno y positiva (pasaje 4):

Profesor: Sí, por ejemplo el 31 sobre 200 es un número negativo, positivo...
Mónica: Positivo. Fracción.
Profesor: Fracción, mayor que uno, menor que uno,...
Ambas: Menor que uno.
Profesor: Menor que uno.¿Y las probabilidades son mayores que uno?
Mónica: Son menores que uno. Todas son menores que uno.

Y que la suma de todas las probabilidades de los eventos debe ser igual a 1 (pasaje 4):

Mónica: Yo digo que aún sumando todas estas (señala la columna de número de hijos) no nos da mayor que 1 porque para que fuera mayor que 1 la fracción tendría que ser 201 o más entre 200. O sea 201 sobre 200 y estos (refiriéndose a la columna de número de hijos) no van a dar más de 200.

De modo que el Profesor intenta utilizar el argumento de que una suma de probabilidades tendría que ser menor o a lo más igual a uno (porque es otra probabilidad) para hacerles ver que 200 no puede ser una probabilidad, esto las ayuda a darse cuenta que omiten el denominador (pasaje 4)

Profesor: Sí, ustedes me acaban de decir hace ratito que la probabilidad de tener 4 hijos es 31 sobre 200 y que esa era una fracción menor que uno ¿no?
Ambas: Ajá.
Profesor: Y luego me dijeron que todas las probabilidades son fracciones, fracciones menores que uno. Entonces si son 10 probabilidades a lo más que podríamos aspirar es a que su suma fuera menor que 10. Cuando yo les pregunto cuál es la suma de probabilidades, ustedes me dicen que 200 así que ¿Cómo le hacen para que llegue a 200?
Hacen exclamaciones de obviedad
Mónica: ¡Ay maestro!
Brenda: Lo que pasa es que esta fracción es con respecto a 200.
Mónica: Ajá. El común denominador es 200.

Sin embargo en el pasaje 6 nuevamente aparece la confusión entre las variables número de trabajadores y probabilidad. Toman a la probabilidad como si fuera el número de trabajadores. A pesar de que se dan cuenta de que la probabilidad debe ser un número relativo a un total. Ellas entienden la probabilidad como número de trabajadores porque de ahí sale la probabilidad, sin embargo tampoco son conscientes de su confusión.

Es posible que esa confusión empiece en el primer pasaje, cuando se les pide identificar una variable dependiente y otra independiente entre la probabilidad y el número de hijos. Al principio ellas establecen esta relación correctamente, pero cuando Mónica hace la afirmación: «*Yo digo que la probabilidad depende del número de hijos*», en Brenda se despierta una interrogante relacionada con los trabajadores. En esa fase, Brenda queda conforme hasta que logra establecer una frase en donde están incluidos: «*qué tan probable es que los trabajadores tengan ese número de hijos*». Como si en la relación establecida tuvieran que estar presentes los trabajadores porque son un elemento muy importante de la situación (es entre quienes se organizará la rifa) o como si les costara trabajo identificar la relación entre el número de hijos y los trabajadores.

Mónica: ...Yo digo que la probabilidad depende del número de hijos (*recurre a voltear a ver a los profesores para que le confirmen o le refuten esta última afirmación).*
Brenda: Pero más bien... el número...el...bueno... no, sí la probabilidad depende del número de hijos porque... *(dudando)*
Mónica: Es que la probabilidad...
Brenda: No, sí la probabilidad... *(muy segura)*
Mónica: Sí, porque según yo...
(se arrebatan la palabra una a la otra)
Brenda: depende del número de hijos porque...
Mónica: Ajá, porque qué tan probable es...
Brenda: porque qué tan probable es que los trabajadores tengan ese número de hijos (*en tono triunfalista*)

En el pasaje 5 se observa que interpretan correctamente la probabilidad expresada en forma de un número decimal, es decir llevan los doscientos trabajadores totales como si fuera una unidad.

Profesor: Nos podrían ayudar ustedes a encontrar un significado a ese decimal, ¿qué podría significar 0.155? ¿qué es eso?
Brenda: Yo digo que los doscientos trabajadores se están tomando como una unidad, entonces se supone que ese 0.155 es un pedacito de eso, a eso se refiere, entonces si sumamos cada fracción que da aquí, nos va a dar el entero que seria igual a uno. Bueno, yo entendí eso.

Otras concepciones de probabilidad se relacionan como grado de creencia, como una forma de expresar la confianza en lo que puede pasar. En el pasaje 10 se relaciona con lo que menos falla:

> Mónica:... Ocupa más, hay más probabilidades, es como que la que menos falla, es como que azar.

En el pasaje 11 con una medida del riesgo:

> Mónica: Yo le pondría un 50%, o si acaso un 70%, porque luego para que me salgan con que perdieron...

Y en el pasaje 12 se vincula con el grado de confianza:

> Brenda: Bueno, es que no se puede asegurar, pero de lo que se habla aquí es de probabilidad, lo que puede pasar.

4.2.1.2 La función de probabilidad

En el pasaje 1, ambas estudiantes cambian las variables sobre las que se les pide que trabajen en la función de probabilidad, para ellas sería más normal tomar a la variable «número de trabajadores» como la variable dependiente. En un principio ambas dudan de las variables, pero después Mónica responde acertadamente en la función de distribución. Para Brenda no es tan natural, como si el hecho de que la pregunta pida explícitamente definir a la probabilidad como variable dependiente o independiente choca un poco con sus concepciones.

> Brenda: Bueno... primero ¿qué depende de qué? pues el número de hijos... no...el número
> Ambas: de trabajadores...
> Brenda: ...depende del número de hijos, ¿no?
> Mónica: No, no es que dice: el número de hijos depende de la probabilidad, no del...
> Brenda: no.. pero la primera no es...
> Mónica: ¿qué depende de qué? ...pues...
> Ambas: el número de hijos depende de la probabilidad ¿o al revés?...
> Mónica: No, la probabilidad depende del número de hijos o sea que qué tan probable... o sea... depende... la probabi...Yo digo que la probabilidad depende del número de hijos (*recurre a voltear a ver a los profesores para que le confirmen o le refuten esta última afirmación*).
> Brenda: Pero más bien... el número...el...bueno... no, sí la probabilidad depende del número de hijos porque... (*dudando*)
> Mónica: Es que la probabilidad...

Al final de este pasaje definen la probabilidad como variable dependiente (dentro de la función de probabilidad), pero en realidad no están convencidas de la probabilidad como variable dependiente. El motivo por el que aceptan esa respuesta es porque el texto de la pregunta les condiciona las variables con las que tienen que trabajar, dejando a un lado lo que ellas consideran la verdadera variable dependiente que sería el número de trabajadores. Mónica es la más convencida, pero a lo único que recurre para convencer a Brenda es a leer una y otra vez la pregunta recalcando las variables que se les pide manejar.

En el pasaje 3 hay reincidencia, esta vez no se percatan que la relación en la que les piden trabajar es la misma que la de la pregunta anterior, así que se inclinan porque la variable dependiente sea el número de trabajadores.

Mónica: El número de hijos depende de la probabilidad o al revés y ya la pusimos. Ahora dice ¿Cuál de estas dos variables sería la variable dependiente y cuál la independiente?
Brenda: Pues yo digo que el número de trabajadores depende del número de hijos.

Esta confusión es más palpable en Brenda, pero eso desestabiliza a Mónica.

En el pasaje 6 se refieren a la probabilidad cuando mencionan el número de trabajadores. Expresan explícitamente su confusión en los términos, aunque también es posible que trabajen con el número de trabajadores porque a partir de él pueden pasar directamente a la probabilidad sin que eso implique un problema para ellas.

Brenda: No porque, si estamos diciendo que la probabilidad depende del número de hijos, quiere decir que la probabilidad es la dependiente, el número de trabajadores que es la probabilidad es la dependiente (*señala sus respuestas en el pizarrón*).
Investigadora: Pero ¿el número de trabajadores es la probabilidad?
Brenda: Si, son estas las probabilidades (*señala la columna de número de trabajadores en el pizarrón*).

En el pasaje 7 ellas se percatan de su confusión entre el número de trabajadores y la probabilidad de tener un número dado de hijos, pero ello no ocurre hasta que pueden ver que para la asignación de probabilidades a la variable aleatoria (número de hijos) se formulan dos relaciones de dependencia, que el número total de trabajadores es una variable (un parámetro, es decir podría cambiar en otro contexto) y que intervienen otras tres variables en ese proceso. Las cuatro variables son:

- El número de trabajadores: cardinalidad del evento compuesto en el espacio muestral del experimento (*¿cuántas personas en total?*),
- El número total de trabajadores: cardinalidad del espacio muestral (*¿de cuántas*?),
- Número de hijos: variable aleatoria vinculada a la característica de interés que agrupó el espacio muestral en eventos compuestos, (*¿Cuántos hijos*?) y
- La probabilidad de obtener un cierto número de hijos *(¿Cuántas personas de cuántas?*).

El ver el número total de trabajadores como otra variable más es lo que las ayuda a descubrir que lo omitían por considerarlo una constante que ahí estaba y que no era necesario mencionar a pesar de que, anteriormente (pasaje 4), manejaron el número total de trabajadores como una constante («*el común denominador es 200*»). En este pasaje Mónica dice «*tres variables*» pero en realidad está mencionando las cuatro.

Mónica: Es que aquí están las tres variables (*señala la pizarra*),... tres porque de 0 hijos hay 16 personas que tienen 0 hijos de las 200, o sea, es que hay tres factores: ¿cuántos hijos?, ¿cuántas personas? ¿y de cuántas?, pero también, ¿cuántas personas de cuántas?; como que estas dos (*se refiere al número de trabajadores y al total de trabajadores*) van relacionadas muy directamente porque son 16 de 200, 9 de 200, pero esas dos se relacionan con esta (*con el número de hijos*), porque esta es la variable de cuántos hijos son.

Por lo tanto, se hacen evidentes las dos relaciones que las estudiantes establecen para asignar al número de hijos, la variable aleatoria, una probabilidad:

Mónica: Según lo que yo entendí, es la probabilidad. Esta (*señala la columna de número de trabajadores en el pizarrón*).
Investigadora: Pero ¿esa es la probabilidad? ¿el 12 es probabilidad?
Brenda: No, de aquí se sacaría la probabilidad de 12 sobre 2.. (*señala el pizarrón*).
Mónica: Es que hay dos relaciones.
Investigadora: ¿Cuáles son las dos relaciones?
Mónica: Esta (*señala el número de trabajadores*) con esta (*señala el número de hijos*), y esta (*señala el número de trabajadores*) con el total (*se refiere al total de trabajadores*).

Las dos relaciones a las que ellas se refieren, son las siguientes (figura 6):

- La primera relación se establece cuando se vincula el número de hijos (la variable aleatoria) con el número de trabajadores (cardinalidad de los eventos compuestos). En realidad la relación es más compleja, se está vinculando a la variable aleatoria con la característica de interés (que es algo menos abstracto relacionado con los trabajadores) y posteriormente con el evento compuesto (trabajadores con un cierto número de hijos).

- La segunda es el cálculo de probabilidades. Se establece entre el subconjunto de trabajadores que tiene un cierto número de hijos (evento compuesto) y el valor de la probabilidad de tener un número dado de hijos (el valor de esta probabilidad depende del número de trabajadores, que serían los casos favorables entre el cociente), esto es $P(X=i)=N_i/N$, siendo N_i el número de trabajadores con i hijos y N el total de los trabajadores (200 en este caso). En este caso se define la probabilidad como función en el espacio muestral.

Lo cuál indica que las estudiantes para poder asignar un valor de probabilidad a la variable aleatoria tuvieron que recurrir a contextualizar en el problema el significado de la variable aleatoria, es decir a relacionarla con el evento compuesto al que estaba ligada para asignarle la probabilidad de ese evento. Para ellas no representaba ninguna dificultad el cálculo de probabilidades y es posible que eso es lo que les haya hecho obviarlo, en cambio la vinculación entre la variable aleatoria y el evento compuesto sí representaba un problema para ellas (recordemos que en el pasaje 1, Brenda tuvo que recalcar «*qué tan probable es que los trabajadores tengan ese número de hijos*» para poder vincular a la probabilidad con el número de hijos). Ellas representan el conjunto de los trabajadores que tienen el mismo número de hijos con la cardinalidad de ese conjunto, pero en realidad están pensando en los trabajadores (como conjuntos de personas), no en el número de trabajadores.

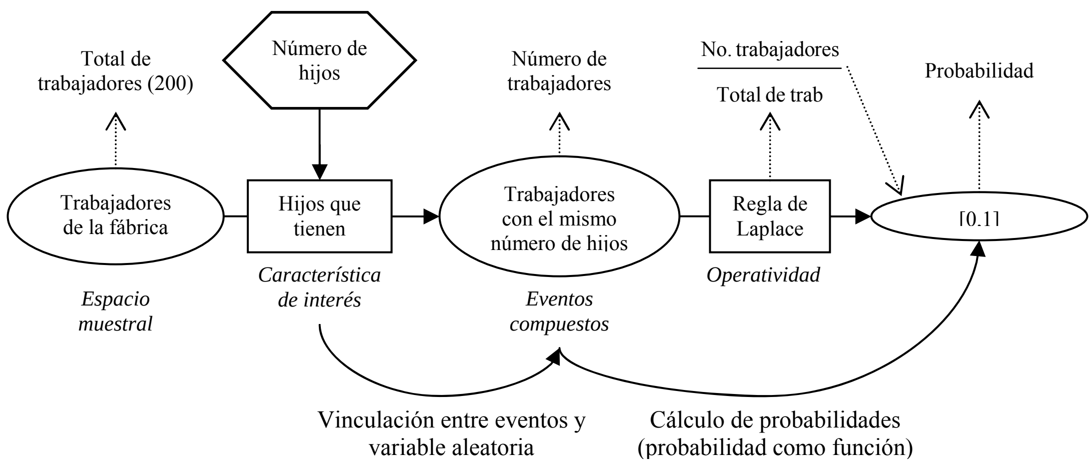


Figura 6. Proceso para vincular el número de hijos con una probabilidad

Observemos también que los eventos compuestos (los trabajadores con un mismo número de hijos) juegan un doble papel en esta descomposición en dos relaciones. Por un lado, en la vinculación del número de hijos con el subconjunto de trabajadores con un mismo número de hijos. Es el resultado de esa aplicación (la variable dependiente en esa relación). En cambio en el cálculo de probabilidades, es el argumento sobre el cuál se aplica la regla de Laplace para poder obtener la probabilidad. Es decir, es la variable independiente en esta otra aplicación. Observemos que ambas relaciones son de conjunto, pero que finalmente al vincular la variable aleatoria con la probabilidad se establece una relación entre números reales.

En el pasaje 14, al elaborar la gráfica, nuevamente se ven tentadas a usar el número de trabajadores como variable dependiente, pero ellas mismas se corrigen y usan la probabilidad como variable dependiente y lo resaltan «*¡Ah! Ya podemos saber que en vez de poner número de trabajadores, vamos a poner directamente la probabilidad*» como un logro, algo de lo presumen ya saber. En posteriores pasajes no vuelven a nombrar al número de trabajadores para referirse a la probabilidad.

En cuanto a la acepción de función, ellas manejan a lo largo de toda la entrevista diferentes concepciones. Una de las primeras de presenta en el pasaje 2, en donde Mónica momentáneamente cree que la función de probabilidad no es una función matemática porque ella cree que una función siempre crece. Cuando ella expresa la palabra *relación* se está refiriendo a una relación funcional.

> Mónica: Es que ve estás dos no pueden ser de que ¡ah! dependiente o independiente (*señala las columnas de la tabla*)... así, no. Yo no puedo entrar... de lo que el número de hijos y el número de trabajadores ... no puede ser, pon tú dependiente e independiente o independiente y dependiente (*señala las columnas de la tabla*) porque no, no hay relación... por ejemplo aquí (*señala la columna de número de trabajadores*) empieza un número 16, y luego aquí mayor, mayor, incrementa y luego decrementa y luego decrementa, así como que no hay rela...o sea no hay mucha relación, igual de cero hijos pudo haber un trabajador y de cuatro a cinco mil y así... como que no da resul...pero dice...

En el pasaje 3 tiene una mejor idea de lo que es una función matemática y se detiene a explicarle a Brenda que es una relación de dependencia entre dos variables:

> Brenda: Sí, pero si te fijas esta es la que.. esta es la que... (*señala la columna de número de hijos*)
> Mónica: Va determinando...
> Brenda: Ajá, va determinando al número de trabajadores. El número de trabajadores no determina al número de hijos.
> Mónica: Ah, ya entiendo.
> Brenda: Sino al revés, el número de hijos es lo que está determinando el número de trabajadores. Entonces está sería...

En ese mismo pasaje observamos como diferencian «determinar» de «dependencia». Creemos que el término «determinar» podría estar vinculado con una operación. Aunque esto no se observa muy bien en este pasaje, en posteriores recalcan la importancia de «*una función esté determinada por una ecuación*».

> Brenda: ... el número de hijos es lo que está determinando el número de trabajadores. Entonces está sería...
> Mónica: Pero eso no significa que depende de los hijos va a ver tantos trabajadores sino que esta *(señala la columna de número de hijos)* rige a esta *(señala la columna de número de trabadores).*
> Brenda: Entonces en este caso, el número de hijos es la variable independiente y el número de trabajadores es la variable dependiente.

En el pasaje 15, Mónica expresa que una función es una relación en las que una variable se corresponde con otra y manifiesta un buen manejo de la notación funcional. Ella expresa rápidamente el resultado pedido en notación funcional y eso le da pie a definir que su interpretación de función como una correspondencia entre dos números.

Profesor: ¿Sería posible eso (*se refiere a la gráfica de la función de probabilidad que está en el pizarrón*) escribirlo en notación de funciones? ¿qué es una función? Esa gráfica, suponiendo que fuera continua, ¿sería una función?
Mónica: Sí.
Profesor: ¿Por qué?
Mónica: Pues como todas (*se refiere a los valores involucrados*) se corresponden, por ejemplo f de 3 que corresponde a 0.225, ¿no? (*Escribe en el pizarrón (F(3) = 0.225)*)

En ese mismo pasaje se observa como relacionan la *F* de la notación funcional con la palabra función, de modo que piensan que esa letra se usa para expresar que la relación establecida sí es una función, de modo que cambian el nombre de la variable dependiente (anteriormente se llamaba ‘y’) por esa F.

Mónica: Pues… ‘x’ sería 3 y ‘y’ sería 0.225. La función de ‘x’ es ‘y’, es decir el resultado.
Brenda: Pero en todo caso… no lo sabes. Lo que yo digo es utilizar…
Mónica: ¿En vez de F, ‘y’, aquí (*señala la F en F(x)*)?
Brenda: No, es que de cualquier forma haría falta especificar que es una función de ‘y’. Lo que podríamos utilizar es, para tomar exactamente las mismas letras, ponerle allá F de ‘x’ (*se refiere al nombre del eje de las ordenadas*). Es decir, cambiar aquella, ponerle F(x) en lugar de ‘y’.

En el pasaje 17 Mónica expresa una definición incompleta, pero Brenda la completa. No hay discusión alrededor de ambas definiciones, aparentemente hay una aceptación inmediata de las dos por la definición de Brenda. Con esta definición se acepta que la distribución de probabilidades sí es una función.

Profesor: ¿La relación que establecieron es una función matemática? ¿o sea es función ¿Recuerdan las propiedades de la función? ¿qué es una función?
Mónica: Que hubiera dos variables.
Brenda: Que para el valor de ‘x’ haya un valor en ‘y’.
Profesor: ¿Y puede haber dos ‘x’ que vayan a dar a un mismo valor de ‘y’?
Ambas: Sí
Profesor: ¿Puede haber una ‘x’ que pueda dar dos valores de y? Es decir, ¿un número de hijos que de dos probabilidades diferentes?
Ambas: No.
Profesor: Entonces ¿es función o no?
Brenda: Sí, si es función porque para cada número de hijos hay una sola probabilidad.

También hay una comprensión de que la gráfica y la tabla son una forma de expresar una función. Lo mencionan en los pasajes 15 y 17. Relacionan la gráfica de una función con la continuidad del trazo (pasaje 14).

Brenda: Aquí (*señala el eje de las abscisas*) sería, uno, dos,… así hasta nueve.
Mónica pone la escala en el eje de las abscisas.
Brenda: Luego para arriba, ¿de qué la empezamos? ¿de 0.1, no?

> Brenda: Bueno, sería de cero hasta uno y luego ya sería irle poniendo las probabilidades. A ver, divídelo…
> Mónica: A ver, primero mitad y mitad. Punto cinco…
> *Brenda dicta los valores a Mónica y ella los va colocando en la gráfica.*
> Mónica: ¿Hacemos la gráfica?
> Brenda: Sí, hay que hacer la gráfica.
> *Brenda une los puntos.*

En el pasaje 16 vuelven a hacer énfasis en que no conciben una gráfica en la que no haya trazos, es decir, que haya sólo puntos.

> Brenda: Bueno es que sí es cierto, se supone que nada más serían los que te están dando. Las coordenadas que te está dando la gráfica únicamente son estos (*señala los puntos de la gráfica*) no son mitades ni nada de eso, sólo los enteros.
> Mónica: Entonces para qué es la gráfica si vamos a dejar los puntos así aislados.
> Brenda: Sí, pero es que realmente no puedes sacar la probabilidad de 0.5.
> Mónica: Sí, no puedes tener 8.5 hijos, es lo que yo digo.
> Brenda: Pero por ejemplo, o sea, o sea, la gráfica sí está abarcando todo, sí está abarcando todo este espacio.
> *Ambas se quedan calladas un rato.*
> Investigadora: ¿Entonces están de acuerdo en que ese es el dominio? (*refiriéndose al [0,9], que ellas habían escrito en el pizarrón*)
> Ambas: Sí.

En el mismo pasaje 16, ellas mencionan que su grafica sería igual si la variable fuera continua o fuera discreta, la diferencia sería que en una continua los decimales sí tienen sentido y en una discreta no, pero no les importa demasiado que la gráfica represente de una manera más fiel la situación problema.

> Investigadora: ¿Por qué lo encerraron en un corchete? ¿qué significa eso?
> Mónica: Que va de… incluye del 0 todos los números, todos los números, hasta el 9. Cerrado.
> Investigadora: ¿Todos los números?
> Mónica: Entre el cero y el nueve, sí.
> Investigadora: ¿Todos los números que hay entre cero y nueve?
> Mónica: Sí
> Investigadora: Cuatro punto cinco está cero y nueve.
> Mónica: Pues también es parte porque... No está en la gráfica, pero está en el dominio. Igual, esto (*señala el 4.5 en el eje de las abscisas y traza con el lápiz su proyección hacia la gráfica y después hacia el eje de las ordenadas*) no tiene un valor escrito pero sí existe.

Lo mismo ocurre con el dominio, se dan cuenta que hay una incoherencia cuando quieren interpretar la gráfica (porque unieron los puntos con líneas) en el contexto del problema puesto que no pueden calcular el valor de la probabilidad (de la variable dependiente) para valores fraccionarios. Sin embargo consideran necesario tanto la unión de los puntos como la inclusión de números decimales en el dominio con la aclaración de que no se tomarían en cuenta puesto que no pueden obtener las coordenadas de esos puntos. La continuidad en la gráfica y en el dominio la justifican a través de su experiencia en sus clases de matemáticas.

Se dan cuenta de que la variable independiente no puede tomar valores decimales, pero no eso no las decide a cambiar su gráfica.

Por otro lado, se dan cuenta de que la representación algebraica de esta función no es fácil de obtener, pero el argumento que dan para no obtenerla a través de una regresión apunta a que están pensando en una regresión lineal y a que su dominio está limitado (no va más allá de nueve hijos), además vinculan esta dificultad con el hecho de que los datos estén determinados por factores ajenos a las matemáticas. En el pasaje 17 dicen que podrían obtener la expresión algebraica de la tabla si aumentara el número de hijos conforme aumenta la probabilidad, porque usarían regresión, pero como eso no ocurre, entonces no pueden obtenerla.

> Mónica: Sin embargo estaba pensando ahorita ¿Cómo le podemos hacer? Lo primero que pensé fue regresión, hacer una regresión y encontrar una ecuación, pero después dije no, porque no tiene nada que ver el número de hijos con las personas, no es de que entre más hijos más personas haya sino de que pueda haber… si va disminuyendo, no va a llegar un momento en que se va a acabar, porque son hijos si fueran otros datos como dinero o alguna otra variable, puede que sí, pero como hijos no, pero como en, no sé, 15 ya se va a acabar, a partir de 3 va a ir disminuyendo.

En el pasaje 21 dicen que los datos que se manejan aquí no están determinados por una ecuación sino por las circunstancias, dicen, «*por el azar, yo creo*».

Ellas creen que la función de probabilidades es diferente a la función algebraica manejada por ellas en cursos de matemáticas porque el valor de ‘x’ no controla a la ‘y’, en el sentido de que no tiene un comportamiento que puedas predecir a través de una ecuación. Se extrañan que la función no sea en cierta medida conocida o controlable. Y creen que esto es lo que la hace diferente de otros contextos manejados por ellas en matemáticas. Es decir, vinculan las funciones «*normales*» (determinísticas) con una ecuación.

> Mónica: ... en otro tipo de gráficas normales a cada valor en ‘x’ corresponde un valor en y, es decir, tiene un comportamiento esperado, dependiendo de la gráfica, no nada más gráfica recta, que puede ser logarítmica o exponencial o polinómica, cualquier tipo de gráfica, pero para cualquier valor en ‘x’ se espera uno en ‘y’. Aquí no, para un valor en x, no sabes cuál va a ser el valor en ‘y’, puede ser más grande, más chico, es algo que no está controlado. La ‘x’ no controla a la ‘y’.

En el mismo pasaje 21, más adelante, lo vuelven a explicar.

> Mónica: Cualquier expresión, cualquier ecuación se esperaría que tuviera cierto comportamiento, aquí no, por ejemplo, si por ‘x’ o ‘y’ razón no hubiera familias con 3 hijos, ahí iría para abajo otra vez la gráfica y no por eso estaría mal la gráfica.

4.2.2 Análisis de los resultados obtenidos en el objeto 'Noción de probabilidad'

Recopilamos los resultados obtenidos sobre este objeto en la tabla 9. En ella están descritos los pasajes principales y su interpretación. Posteriormente hacemos una interpretación general de estos resultados.

Tabla 9. Síntesis de resultados de la noción de probabilidad

Variable	Pasaje	Acciones de las estudiantes	Interpretación
La probabilidad en la situación problema	13 y 21	Definen el espacio muestral como el conjunto de los 200 trabajadores de la fábrica Así mismo manejan el espacio muestral en la situación definida	Hay una comprensión del espacio muestral del problema
	4	Interpretan la probabilidad como un cociente de casos posibles entre casos totales. Lo interpretan como la relación de entre dos cantidades	Hay una buena interpretación de la probabilidad laplaciana en el contexto del problema Pero no la ven como un número sino como la relación entre dos
	4	Mencionan que todas las probabilidades son menores que uno y que su suma debe ser uno.	Hay un buen conociendo de dos axiomas de la probabilidad
	5	Interpretan la probabilidad expresada en decimales como si los 200 trabajadores fueran la unidad	Hay una correcta interpretación de la probabilidad en forma decimal
	4	Hablan y señalan el número de trabajadores pero la llaman 'probabilidad'. Aunque pareciera obvio que hay un común denominador aunque no lo mencionan	Ellas prefieren manejar el número de trabajadores en lugar de la probabilidad o bien hay una confusión entre la probabilidad y el número de trabajadores
	6	Vuelven a nombrar la probabilidad para referirse el número de trabajadores	Ellas no son conscientes de su confusión. Internamente ellas prefieren hablar de número de trabajadores porque está más vinculado con el problema
	10, 11 y 12	Relacionan la probabilidad con 'lo que menos falla', con una medida de riesgo y con un grado de confianza.	Utilizan la probabilidad para expresar su grado de certeza en las recomendaciones que van dando
La probabilidad como función	1, 3 y 6	Mencionan la variable dependiente como el número de trabajadores en lugar de la probabilidad en la distribución de probabilidad	Hay una confusión entre esas dos variables Prefieren analizar el problema tomando como base el número de

		Reinciden en esa confusión e incluso la hacen explícita	trabajadores porque ésta última está más cercana a un contexto
	7	Se dan cuenta de que hay 4 variables involucradas en este problema. Mismas que generan dos relaciones y en ambas interviene el número de trabajadores Se dan cuenta que no es lo mismo probabilidad que número de trabajadores. Dos cosas que contribuyeron a percatarse de su error es darse cuenta de que el número total de trabajadores es un parámetro y relacionar el número de hijos con los trabajadores que tienen ese número de hijos	Pueden percatarse de su confusión porque se dan cuenta de la existencia de que se establecen dos relaciones al asignar probabilidades a los valores de la variable aleatoria Ellas trabajan con el número de trabajadores como variable, pero están pensando en el conjunto de trabajadores con un mismo número de hijos Hay una dificultad en que los eventos compuestos funja como variable dependiente y como variable independiente en las dos relaciones vinculadas a la asignación de la probabilidad a los valores de la variable aleatoria
	14	Hay una reincidencia en la confusión entre probabilidad y número de trabajadores, pero ellas mismas corrigen	Hay una aceptación de ambos tipos de relaciones y por lo tanto deja de existir la confusión entre esas dos variables
	2, 3, 15 y 17	Expresan diferentes concepciones de lo que es una función matemática: es una relación que siempre es creciente (incorrecta) es una dependencia entre dos variables (correcta) es una correspondencia entre dos números (correcta) relación en la que para cada valor de ‘x’ le corresponde uno de ‘y’ (correcta)	Manejan diferentes concepciones de función. Todas ellas correctas, aunque en algún momento expresan alguna incorrecta
	2, 3, 15 y 17	Inicialmente dudan si la distribución de probabilidad es una función o no, pero finalmente terminan por asegurar que sí es una función	Ellas manejan apropiadamente la noción de función y por ello no tienen problemas en definir la distribución de probabilidad como una función.
	15	En la distribución de probabilidad, denotan a la probabilidad con ‘y’ y al número de hijos con ‘x’ Cambian la nomenclatura en el momento que saben que la relación es una función y utilizan la letra F	Utilizan las variables familiares para ellas para denotar las variables que intervienen en la distribución de probabilidad Usan la letra F porque ella expresa

		para denotar 'función'.	que hay una relación funcional
	15 y 17	Expresan que tanto la tabla como la gráfica son una función	Tienen una buena concepción de las representaciones gráficas y tabulares de una gráfica
	16	Hacen una gráfica continua, pero la interpretan como discreta Expresan que una gráfica con puros puntos aislados no tiene sentido Expresan el dominio como un intervalo continuo	Su noción de gráfica está relacionada con un trazo continuo. Están más familiarizadas con ese tipo de gráficas Están acostumbradas a expresar el dominio como un intervalo continuo
	16	Expresan que tanto la gráfica como el dominio deben ser continuos porque así son las matemáticas, aunque esto no les sirva para representar su situación problema	Desvinculan las matemáticas del contexto del problema. Hay una reticencia a que las matemáticas puedan representar una situación del contexto real
	17	Se dan cuenta que no es fácil obtener la ecuación de la distribución de probabilidad Creen que aunque la pudieran sacar esa ecuación no tendría sentido por el tipo de función con el que están trabajando que involucra datos obtenidos de una fábrica	Hay un choque entre el contexto algebraico y el de la situación problema. Piensan que no es posible obtener la ecuación de la relación porque los datos son extraídos de la realidad, están dados por circunstancias y no por la misma matemática. La matemática, para ellas, expresa sus relaciones en forma de ecuaciones

De la primera parte de la tabla se puede observar un buen manejo del cálculo e interpretación de la probabilidad clásica, tanto decimal como fraccionaria y de dos axiomas de la probabilidad. Así mismo tienen un correcto manejo de la probabilidad como el grado de certeza de la recomendación que darían. Sin embargo usan repetidamente el término 'número de hijos' para referirse a la probabilidad.

Es posible que la probabilidad no sea tan evidente como variable dependiente en la función de probabilidad y por eso prefieran trabajar con el número de hijos. Pero esa confusión también pudo ser ocasionada por la dificultad en asignar un valor de probabilidad a la variable aleatoria, es decir, en la definición de la función de probabilidad.

En la actividad propuesta, pedimos a las estudiantes, que, una vez realizado el cálculo de la probabilidad de los eventos compuestos, trabajaran con la función de probabilidad. Esto se observó en las preguntas *«¿Qué depende de qué: el número de hijos depende de la* *probabilidad o al revés? ¿Cuál de estas dos variables sería la variable dependiente y cuál la variable independiente?*» de la actividad (anexo 2). Pero eso constituyó un problema para ellas. No podían pasar directamente del valor de la variable aleatoria (un número de hijos) a la probabilidad porque en ello interviene una composición de funciones que ellas inconcientemente quieren obviar. Necesitaban regresar a preguntarse constantemente por el significado de la variable aleatoria en el espacio muestral. No interpretamos que las alumnas piensen que la probabilidad y el número de trabajadores son lo mismo, sino que saben que a partir del número de trabajadores ellas pueden calcular la probabilidad. La dificultad se traduciría en ellas no podían abstraer (descontextualizar, desde la perspectiva de Chevallard, 1985/1991) en dos sentidos:

- **En la probabilidad**: no podían extraerla de su cálculo a través de la regla de Laplace y por lo tanto olvidar que es la razón entre los favorables y los posibles (en sus palabras: «*¿cuántos de cuántos*?»).
- **En la variable aleatoria**: inconcientemente no podían desligar el valor de la variable aleatoria (número de hijos) con los eventos compuestos (trabajadores con ese número de hijos).

Así, desde la perspectiva de Chevallard, la asignación de probabilidades a los valores de la variable aleatoria y, por lo tanto, el establecimiento de la función de probabilidad (como concepto matemático) requiere dos procesos, la *contextualización* del valor de la variable aleatoria y *descontextualización* de la probabilidad de los eventos del espacio muestral (figura 7). En este proceso las estudiantes podían contextualizar bien los conceptos matemáticos que se les proporcionaron, pero tenían dificultades para descontextualizar, no recurrían a ello. Pudieron pasar de un número de hijos (valores de la variable aleatoria) al conjunto de trabajadores que tenían ese número de hijos (eventos compuestos), pero no era tan sencillo que representaran a los trabajadores con el mismo número de hijos a través del número de hijos que tenían. Así mismo, dada la probabilidad, podían interpretarla en el espacio muestral (*¿cuántos de cuántos?*), es decir, podían contextualizarla, pero no recurrían a usarla como un solo número (y por lo tanto desligarla del espacio muestral) o asignarla a la variable aleatoria, es decir, no podían descontextualizarla.

Desde esta perspectiva, la aplicación de la inversa de la variable aleatoria sobre los valores que toma, es una contextualización de esos valores en el espacio muestral.

Observemos, que es posible que su correcta interpretación de la probabilidad laplaciana las haga «ver» a la probabilidad como dos números y no como uno (*30 de 200*). Es decir, este conocimiento que podría llamarse exitoso en el contexto de probabilidad laplaciana, en este contexto de la función de probabilidad sería un obstáculo a superar.

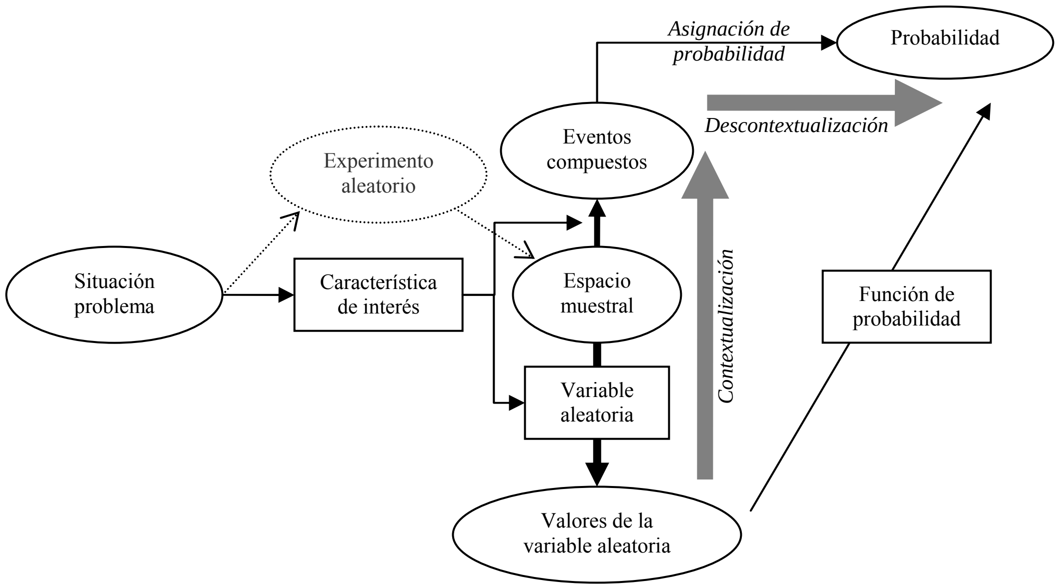


Figura 7. Proceso de contextualización y descontextualización de la asignación de probabilidades a la variable aleatoria y el establecimiento de la función de probabilidades.

Es posible que ellas no quieran descontextualizar el problema trabajando con una variable como la probabilidad, que es algo abstracto, y que en cambio prefieran trabajar con el número de trabajadores que tienen un cierto número de hijos (que son los eventos compuestos) porque de alguna manera es más «visible» para ellas.

También se presenta un choque entre la aceptación de que la matemática sirve para representar la realidad. Ellas piensan que las representaciones matemáticas no tienen porqué estar tan de acuerdo con la realidad sobre todo en este contexto discreto. Incluso no encuentran contradicción por considerar en la variable independiente números decimales (para el número de hijos) que no tienen sentido en el contexto del problema, pero que está de acuerdo con el contrato didáctico establecido en lo referido a las gráficas funcionales. Llegan a mencionar que la gráfica de una función sería la misma para variables independientes discretas o continuas. Igual les ocurre al interpretar el dominio de la función, donde unen con líneas los valores puntuales en la gráfica, aunque no puedan calcular la probabilidad para un número fraccionario de hijos. El dominio también lo ven limitado a los datos dados en el problema, sin considerar que el contexto matemático de la función de probabilidad da valor de probabilidad cero al número de hijos mayor de 9.

4.3 La noción de variable aleatoria

Aquí trataremos de explorar los tres momentos más importantes del papel que juega la variable aleatoria en la construcción del modelo de distribución de probabilidad: su concepción propiamente como función, la variable aleatoria como número vinculado a una probabilidad y como variable independiente dentro de la función de distribución. En el análisis contemplaremos estas mismas tres variables que manejamos en el diseño.

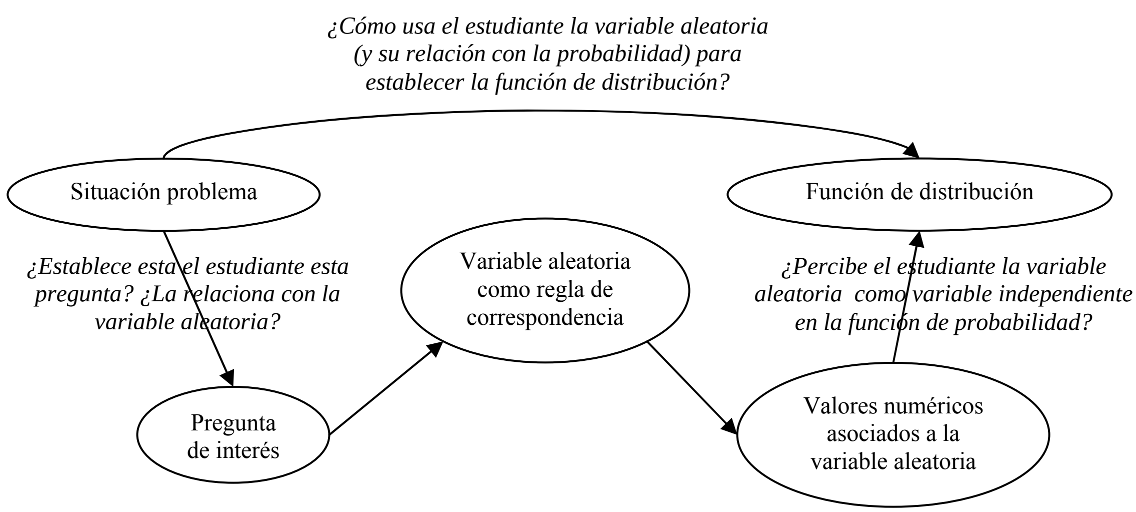


Figura 8. Descripción del objeto de análisis 'La variable aleatoria'

Las variables dentro de este apartado se describen más detalladamente en la tabla 5.3 y son:

- *Manejo de la variable aleatoria en la situación problema.* Lo relacionamos con el vínculo que se establece entre los valores que adopta la variable aleatoria con una probabilidad asociada en el contexto del problema y que por lo tanto, estará vinculado con la forma en que el estudiante maneje la probabilidad y la variable aleatoria para encontrar la solución del problema.

- *La variable aleatoria como función.* Nos interesa profundizar sobre la forma en que el estudiante maneja la variable aleatoria como función vista como una regla de correspondencia y como un número vinculado al espacio muestral en la situación problema.

- *La variable aleatoria y la función de probabilidad.* Buscamos observar la interpretación del estudiante de la variable independiente en de la función de probabilidad.

Tabla 10. Variables de observación asociadas a La noción de variable aleatoria

Variables de observación	Desglose
La variable aleatoria en la situación problema	La variable aleatoria en el proceso de solución del problema La variable aleatoria en la recomendación Dificultades asociadas
La variable aleatoria como función	Vinculación de la pregunta de interés con la variable aleatoria Relación de los eventos del espacio muestral con una variable numérica. Interpretación de la variable aleatoria vinculada con la situación problema Dificultades asociadas
La variable aleatoria y su distribución de probabilidad.	La relación de dependencia entre la variable aleatoria y la probabilidad Asociación de los valores de probabilidad puntual con los valores de la variable aleatoria Asociación de los valores de la probabilidad acumulada con los valores de la variable aleatoria Vinculación entre la variable aleatoria y la aleatoriedad Dificultades asociadas
Lenguaje	Dificultades, ambigüedades, equívocos y nomenclatura al referirse a: • La variable aleatoria • Las relaciones entre la variable aleatoria y la probabilidad • Las relaciones entre la variable aleatoria y el azar

4.3.1 Resultados en el objeto 'La variable aleatoria'

4.3.1.1 La variable aleatoria en la situación problema

En el pasaje 9, ya comentado, ambas alumnas piensan que la mejor solución a la pregunta de cuántos boletos comprar sería 5 boletos, basándose en que esto es «*lo más probable*», por tanto hacen uso explícito de la idea de que el número de hijos es una variable, al mismo tiempo que su interpretación de la probabilidad asociada indica que son concientes que el valor de esta variable está vinculado a la aleatoriedad.

En el pasaje 12, Brenda interpreta la probabilidad para explicar por qué recomiendan el valor que tiene la mayor probabilidad, de qué manera este valor soluciona el problema y como ésta se representa en la tabla de datos dada. Interpretar el contexto se le facilita más que manejar las frecuencias al señalar la acumulación de la mayor parte de los trabajadores.

> Brenda: Como además, se encuentran más concentrados los trabajadores en esa parte (*señala los renglones de la tabla correspondientes a trabajadores con 3 hijos o menos*) hay más posibilidades de que no vaya a perder tanto porque más del 50% (*de los trabajadores*) se juntan en ese espacio. No podemos asegurarlo pero es más factible, puede que suceda más, que salga más una de estas personas (*señala los renglones correspondientes a los trabajadores con 3 hijos o menos*), es más fácil que del resto porque son más, son mayoría.

En el pasaje 10 Brenda se comienza a fijar en otras características: la probabilidad acumulada del complemento. Brenda, ahora que Mónica está convencida de comprar 5 boletos, comienza a dudar porque ve cuántos son los trabajadores que a los que podrían no alcanzarles los boletos.

> Brenda: Bueno, yo diría eso, comprar los 5, pero también existe la probabilidad de que no vaya a completar después.

En el pasaje 10 Brenda relaciona las probabilidades acumuladas con la probabilidad que les interesa analizar en el problema. Ella introduce las probabilidades acumuladas (aunque usa también la frecuencia acumulada: «*el número de trabajadores hasta aquí*») para argumentar una postura diferente a los 5 boletos. No efectúa las operaciones, sólo hace aproximaciones.

> Brenda: Aunque bueno, también por decir un margen mas grande sería... bueno mas seguro sería comprar los 9, porque si te fijas, de aquí hasta acá (*señala en la tabla la columna de probabilidades las correspondientes de 0 a 6 hijos*) es donde se ocupa la mayor parte, de 0 a 6 es donde se esta ocupando la mayor parte inclusive (*se podría tomar*) un poco más, ya en la de 7 hijos que es donde llega a 9 (*es decir, añadiría los trabajadores que tienen 7 hijos*), o sea, es donde esta la mayor probabilidad, ésta es muy poquita (*señala la probabilidad de que se tengan 8 ó 9 hijos*). Estos dos tienen muy poquita probabilidad y serían 7 más los... que serían... Si te fijas realmente ya nada mas tendrías un margen de 13 o sea muy pocos.

Como ya notamos en el punto 4.2.2 la interpretación de la variable aleatoria representó un conflicto que requería que las estudiantes regresaran constantemente a los eventos compuestos. Posteriormente, cuando trabajan la función de probabilidad y la función de distribución, no tienen mayores conflictos en trabajar con los valores de la variable aleatoria. Sin embargo, recurren constantemente al manejo del conjunto de trabajadores (a la frecuencia) para interpretar sus resultados.

4.3.1.2 La variable aleatoria como función

La actividad no permitió que las estudiantes plantearan la variable aleatoria como regla de correspondencia puesto que los datos que se les proporcionaban estaban ya referidas a la variable aleatoria y a la partición del espacio muestral. Ellas tenían solamente que interpretarlos y manejarlos.

Como ya dijimos, la interpretación de la variable aleatoria fue exitosa en la medida que ellas pudieron visualizar bien lo que significaba en el contexto del problema pero eso hizo que se dificultara un poco la descontextualización de los conceptos matemáticos.

Al preguntarles por el rango y dominio en la función de probabilidad, en el pasaje 16 asocian las cotas superiores e inferiores del rango y el dominio con los valores más grandes y más pequeños que pueden tomar las variables al interpretar la grafica.

> Profesor: Bueno, la que sigue. Describe el dominio de la variable independiente y el rango de la variable dependiente.
> Mónica: Pues de 0 a 9 es el dominio de la independiente y el rango de la dependiente sería de 0.025 a 0.225 (*escribe en el pizarrón [0,9] y [0.025, 0.225]*).
> El dominio de la función de distribución correspondería a la imagen de la función variable aleatoria, de modo que no tendría sentido que la variable aleatoria tomara valores decimales en su imagen y sin embargo, como ya lo hemos mencionado, las estudiantes incluyen valores decimales en el dominio.

Se dan cuenta de que los valores de la variable aleatoria mantiene un orden ascendente, normal, como todos los números, en su tabla y en su gráfica, pero dicen que podrían estar ordenados de otra forma. Por ejemplo que se podría tomar como criterio de orden en la gráfica los valores correspondientes de probabilidad a cada valor de la variable independiente, de menor a mayor probabilidad o al revés o que incluso los valores de la variable aleatoria podrían haberse ordenado de manera descendente. Argumentan, sin embargo, que el contexto no tiene nada que ver con eso, los valores de la variable 'número de hijos' se presentarán aleatoriamente en caso de efectuar varias veces el experimento volviendo a introducir el papel en la tómbola (pasaje 20). Se observa, también un buen manejo de la noción de aleatoriedad y un salto directo de la variable aleatoria al experimento sin pasar por los conjuntos de trabajadores, esto es, vinculan el resultado del experimento aleatorio con el número de hijos directamente, lo que facilita que vinculen directamente la probabilidad con el número de hijos.

Brenda: … bueno en este caso sí va así seguido (*se refiere a los valores de la variable aleatoria*), pero a lo mejor hay casos en que no había nadie con 6 hijos entonces la variable ya no tendría que ser toda seguida.

Mónica: Pues obviamente sí se pueden ordenar siempre las variables (*se refiere a los valores que puede tomar la variable*), así de menor a mayor y de mayor a menor también (*señala los valores de la probabilidad*), pero eso no significa que de mayor a menor van a ir creciendo o van a ir disminuyendo igual (*si se repite muchas veces el experimento aleatorio*), o sea, el que aquí sea de menor a mayor (*se refiere a las valores de la variable aleatoria en la gráfica*), el que estén ordenadas, no significa que también estén creciendo uniformemente (*se refiere al efectuar varias veces el experimento*).

No creen en una necesidad del orden en la variable independiente de la función de distribución. Más bien ven a los números como etiquetas. Ordenan de menor a mayor porque así siempre lo han hecho en sus cursos de matemáticas.

En el pasaje 21 hay un momento en el que Mónica describe las relaciones que se establecen entre las variables que podemos obtener a partir de este problema para determinar la variable aleatoria. Menciona la regla de correspondencia y el proceso que se lleva a cabo para obtenerla, pero no es consciente de la naturaleza del concepto matemático que está manejando, sólo es una descripción de lo que pasa cuando se efectúa el proceso aleatorio. Ella utiliza un razonamiento regresivo (Polya, 1965).

Brenda: No, porque adentro de la tómbola están los trabajadores. Es que se supone que del número de hijos… por decir…¿cómo lo explicaré?

Mónica: O sea no está el número de hijos, de que uno, dos, sino que está un trabajador y tú, a ver cuántos hijos tienes. No pues que dos. Está implícito adentro del trabajador.

Brenda: Sí, ándale. O sea no hay papelitos que digan uno, dos, tres, …

Mónica: No, sino que sale, Juan Ramírez y tú, a ver, cuántos hijos tienes, no pues que tres, a corresponde a… (*señala el eje de las ordenadas en la gráfica*) y ya.

Brenda: O sea que dentro de los papelitos que están en la tómbola, están las posibilidades o sea estos (*señala la escala del eje de las abscisas*).

Se dan cuenta que el fenómeno aleatorio está relacionado con el número de trabajadores y que la situación problema es la que condiciona qué pregunta se le haría al trabajador, de modo que la partición del espacio muestral (la característica de interés) está condicionada por el objetivo del problema. Esto no logran establecerlo de manera más sistemática. Ellas se dan cuenta en este problema, pero no se puede garantizar que se efectúe de manera apropiada la descontextualización de la regla de correspondencia de la variable aleatoria.

4.3.1.3 La variable aleatoria y su distribución de probabilidad

Ellas dicen que el dominio de la variable aleatoria es 0 a 9 y el rango de la probabilidad sería de 0.025 a 0.225 (escriben en el pizarrón [0,9] y [0.025, 0.225]). A pesar de que anteriormente habían mencionado la probabilidad 0, a la hora de definir el dominio y el rango no lo toman en cuenta, tampoco mencionan la posibilidad de que haya 10 o más hijos. Para ellas la función de distribución comienza en 0 y acaba en 9 (pasaje 16). No lo hacen con fines de extender la probabilidad de la función de distribución, sino porque piensan que en el ámbito matemático todas las gráficas de las funciones son continúas. La continuidad está incluida en su concepción de gráfica. Esto se observa en el pasaje 16.

Brenda: Bueno es que sí es cierto, se supone que nada más serían los que te están dando. Las coordenadas que te está dando la gráfica únicamente son estos (*señala los puntos de la gráfica*) no son mitades ni nada de eso, sólo los enteros.
Mónica: Entonces para qué es la gráfica si vamos a dejar los puntos así aislados.
Brenda: Sí, pero es que realmente no puedes sacar la probabilidad de 0.5.
Mónica: Sí, no puedes tener 8.5 hijos, es lo que yo digo.
Brenda: Pero por ejemplo, o sea, o sea, la gráfica sí está abarcando todo, sí está abarcando todo este espacio.
Investigadora: ¿Entonces están de acuerdo en que ese es el dominio?
Ambas: Sí.

En el pasaje 18 las diferencias inmediatas que ambas estudiantes encuentran entre las variables manejadas en la función de probabilidad y otras variables independientes con las que ellas han trabajado radican principalmente en la rareza de los datos discretos y en que es muy restringido el intervalo de valores que puede tomar la variable independiente.

Profesor: ¿Qué diferencia encuentras entre la variable independiente de esta función de probabilidad y las variables independientes de las otras matemáticas que tú conoces, por ejemplo del álgebra y del cálculo?
Mónica: Qué es limitada.
Profesor: Supongamos que tiene un sentido «más hijos» indefinida, aun así ¿Hay más características? ¿Es la única característica que ustedes encuentran entre esta función y las otras variables independientes de álgebra y cálculo?
Mónica: Lo de los decimales.

También se observa que vinculan la noción de función con el manejo de un contexto algebraico, de manera que lo ven como otra diferencia entre las funciones que han trabajado en anteriores cursos de matemáticas y el contexto de esta actividad. En el pasaje 19 mencionan que una diferencia importante entre este modelo que están trabajando y los que han trabajado en sus cursos de matemáticas es que no pueden obtener la ecuación a través de la cual se relacione la variable dependiente con la independiente. Esto lo relacionan con la exactitud de la variable dependiente, es decir, a partir de la variable dependiente no pueden obtener (en el sentido de calcular) la dependiente de manera exacta (porque sólo puede ser obtenida a partir de la tabla o de la gráfica).

> Profesor: Qué más, piensen ¿Habrá alguna característica más que distinga esta variable de las otras que trabajan en matemáticas?
> Brenda: Normalmente en álgebra manejamos ecuaciones. Todas las gráficas se manejan por ecuaciones, todos los valores de 'y' se van a dar por la ecuación. Aquí los valores de y también se dan según una 'x', pero no tienen una ecuación, son así como… aleatorios. No llevan un patrón así seguido.

En los pasajes 21 y 22 también se observa esta tendencia a querer interpretar una función a través de una ecuación.

> Mónica: Las otras son constantes, continúas, se espera que sigan y que vayan, en cambio esta no sabes, no puedes predecir algo.

Expresan que aunque sí podrían sacar una ecuación, no tendría sentido en el contexto del problema (pasaje 19). Vinculan un «*patrón*» de comportamiento con obtener la ecuación y piensan que no lo tienen y por eso no pueden tener una ecuación. Relacionan esa carencia de patrón con un comportamiento aleatorio, pero vinculan ese comportamiento aleatorio a la variable dependiente (es decir a la probabilidad) porque ella es la que no pueden calcular. Están pensando en obtener la variable dependiente (probabilidad) a partir de la variable independiente (número de hijos).

> Brenda: Bueno más que la variable independiente diferente, como que es la variable dependiente la diferente, porque no podemos establecer cuál será la dependiente (*se refieren al valor que tomaría esa variable*). A diferencia del álgebra, con la independiente podías sacar la dependiente exactamente y en este caso no, son datos que te están dando y tú no puedes... es muy aleatorio, es más difícil que con álgebra. Con álgebra tú podías sacar 'y' con la ecuación y aquí no.

Ellas vinculan más la incertidumbre con la probabilidad más que con la variable aleatoria. Interpretan esa «aleatoriedad» con el no tener una ecuación para poder calcular la 'y' (probabilidad). Los valores de la variable aleatoria para ellas son números que se debería poder sustituir en una ecuación para obtener el valor de la probabilidad. En el pasaje 21 vuelven a retomar esto.

> Mónica: Yo también estoy de acuerdo con ella en que la que varia es la variable dependiente. Varía de forma inconstante o impredecible.
> Investigadora: Pero ¿por qué impredecible?
> Brenda: Sí, porque este valor no es algo que esté dado por una…
> Mónica: Función.

Brenda: Sí, por una función, por una ecuación, sino que porque son datos que así se dan o sea que, que pueden ser más o que pueden ser menos, que nos están determinados por algo sino por las circunstancias, por el azar yo creo.

En el pasaje 20, ellas mencionaron que los valores de la variable aleatoria no tienen que presentarse de manera continua en caso de realizar varias veces el experimento, ya que no pueden saber en realidad en que orden se presentará y que aún cuando puedan realizar el experimento varias veces, en un nuevo experimento no sabrán qué número de hijos tendrá el trabajador que gane.

Profesor: Entonces ¿habrá orden de aparición en el número que salga de la tómbola?
Brenda: O sea de que primero saque uno y salga uno, después vuelva a sacar y salga dos y así, pues no, no lo hay.

Sin embargo en el pasaje 21 no vinculan, de manera explícita, el número de hijos con la aleatoriedad, o con la intervención del azar. Ni tampoco esa incertidumbre les cuestiona que en el eje de las abscisas ellas colocaron los valores que puede tomar la variable aleatoria en forma ordenada. Cuando ellas mencionan la palabra «aleatorio» o «azaroso» se refieren al valor que tomará la probabilidad.

En otro momento de ese mismo pasaje aseguran que los valores de la variable aleatoria serían no predecibles sólo si la probabilidad de que tome ese valor fuera cero:

Mónica: ... a lo mejor hay casos en que no habría nadie con 6 hijos, entonces la variable (*se refiere a los valores de la variable aleatoria*) ya no tendría que ser toda seguida.

como si en tal situación este valor se tuviera que quitar del eje de las abscisas, puesto que no hay nadie que tenga ese número de hijos. Esta misma idea vuelve a aparecer en el pasaje 22 cuando intentan dar un nombre a las variables independientes de este tipo y Mónica sugiere llamarla «impredecible».

Brenda: No, la impredecible es y.
Mónica: No, pero también es x.
Investigadora: ¿Por qué esta también es impredecible?
Mónica: No, inexacta… ¡Porque! Por… porque a lo mejor podemos quitar el dos, a lo mejor no hay familias con dos hijos…. Incontrolable, no.

Es decir, vuelven a vincular lo impredecible del valor de la variable aleatoria en el sorteo con la probabilidad. En estos casos con la probabilidad cero que haría que no supieran cuáles son los valores que pueden poner en el eje de las abscisas.

La idea relacionar la aleatoriedad con una incapacidad de cálculo las induce a utilizar una noción errónea de aleatoriedad porque aunque ellas pudieran calcular todos los valores de probabilidad a través de una ecuación, eso no haría que supieran cuál es el trabajador que saldría seleccionado. Esta idea la argumentan varias veces a lo largo del pasaje 21.

Relacionan los datos «humanos» (es decir, los obtenidos a través de la experiencia) con la inexactitud de la función que están manejando. Piensan que no es posible modelar los datos que tienen a través de una ecuación porque son datos obtenidos a través de la experiencia y no de una ecuación. Están manifestando una creencia en que las ecuaciones no se obtienen a partir de «datos reales» y precisamente por eso creen no posible obtener una a partir de los datos que están manejando.

> Mónica: Aunque sí se observa (*señala la gráfica de la función de probabilidad*) que primero crece y luego ya disminuye, disminuye y como son hijos y como somos humanos son funciones más o menos, o sea va a ir disminuye y disminuye hasta que llegue a 12 y ya. En ese caso pero por ejemplo más datos humanos como… número de personas divorciadas o cuántos se han casado, ahí no sería viable sacar una ecuación porque no te iba a decir nada.
>
> Brenda: Bueno en realidad sí puedes sacar una ecuación, pero eso no te daría… no sería muy exacto. Esa es la diferencia, que no te pueden dar datos exactos.

Esta interpretación da pie a pensar que las estudiantes no están pensando únicamente en la situación planteada sino en una generalización para cualquier situación. Ellas piensan que una ecuación no debe servir para un caso sino para muchos (en este contexto para muchas fábricas), de manera que les causa conflicto pensar que no pueden determinar cuántos trabajadores tendrían un cierto número de hijos (y por lo tanto la probabilidad) en todas las fábricas con una ecuación porque eso sería algo muy particular de cada fábrica. Creen que una ecuación de número de hijos y probabilidad tendría que ser más general. En el pasaje 21 esto se observa con más claridad.

> Mónica: Yo también estoy de acuerdo con ella en que la que varia es la variable dependiente. Varía de forma inconstante o impredecible.
>
> Investigadora: Pero ¿por qué impredecible?
>
> Brenda: Sí, porque este valor no es algo que esté dado por una…
>
> Mónica: Función.
>
> Brenda: Sí, por una función, por una ecuación, sino que porque son datos que así se dan o sea que, que pueden ser más o que pueden ser menos, que nos están determinados por algo sino por las circunstancias, por el azar yo creo.
>
> Mónica: Ajá, porque pueden estar determinados por diferentes factores por ejemplo, esa población puede haber tenido tres hijos porque se localizaba en México o en Estados Unidos en donde la mayoría de las personas tienen tres hijos, pero en poblaciones de otros lugares donde el número de hijos crece por persona, la gráfica ya no iba a ser igual. En cambio si hablamos de dinero, bueno, no tanto de dinero sino de otro tipo de cosas como habíamos visto anteriormente, sí se espera algo, pero aquí el número de hijos… En este tipo de gráficas en donde están midiendo el azar, no podemos saber cuándo va a bajar y cuando va a subir, no tiene un comportamiento predestinado.

Además, se observa una vinculación entre una ecuación y la predicción, por eso para ellas la incapacidad de cálculo está vinculada con la incertidumbre (no saben qué pasará en otras circunstancias). En otras fábricas el valor del número de hijos (1, 2, 3,... 9) sería el mismo a menos que la probabilidad de tener un cierto número de hijos fuera cero, en cambio los valores exactos de la probabilidad cambiarían. Ya no serían 0.025, 0.035, 0.045,... 0.225 sino otros que dependerían de la educación del trabajador, de sus condiciones de salud, etc, es decir, de las circunstancias de la fábrica. Vinculan las ecuaciones con poder predeterminar resultados sin necesidad de adentrarse a la situación.

4.3.1.4 Dificultades con el lenguaje

En el pasaje 15 el nombre de las variables que proponen son las utilizadas comúnmente en los cursos de matemáticas, ‘x’ y ‘y’. Inicialmente, al mencionar notación funcional, Brenda relaciona la respuesta con una ecuación o con una expresión algebraica. Se pregunta si una función sólo está definida a través de una ecuación. Los profesores aprovechan para explorar sus concepciones de función, pero ellas no se muestran muy seguras de sus conocimientos al respecto. En un momento de ese diálogo, Brenda se da cuenta que no necesitan saber la ecuación para responder lo que se les está preguntando, así que se olvidan un poco de eso y se concentran en contestar la pregunta.

Profesor: Representa con una letra la variable dependiente y con otra la variable independiente.
Mónica: ‘x’ y ‘y’ (*las escribe en la gráfica que está en el pizarrón*). Ya.
....
Brenda: (*dirigiéndose a Mónica*) ¿A qué se refiere con notación funcional? ¿una ecuación?
Dudan..
Mónica: ¡Funciones!, F de x, f de…
Profesor: ¿Sería posible eso (*se refiere a la gráfica que está en el pizarrón*) escribirlo en notación de funciones? ¿qué es una función? Esa gráfica, suponiendo que fuera continua, ¿sería una función?
Mónica: Sí.
Profesor: ¿Por qué?
Mónica: Pues como todas (*se refiere a los valores involucrados*) se corresponden, por ejemplo f de 3 que corresponde a 0.225, ¿no? Escribe en el pizarrón (*F(3) = 0.225*)
Profesor: ¿Qué dice Brenda?
Brenda: Pero lo que pide es poner… ¿cómo es la ecuación?
Investigadora: ¿Una función nada más se expresa en forma de ecuación?
Brenda: No.
Investigadora: ¿De qué otra forma?
Mónica: Por ejemplo. No, una función no necesariamente tiene que ser una ecuación, por ejemplo está podría ser una función (*escribe F(x) = 3x*), siendo que no es una ecuación.
Profesor: Pero esa sería una expresión algebraica, ¿puede algo que no sea una expresión algebraica, una gráfica o una tabla, ser función? ¿tiene a fuerzas que ser una fórmula?
Mónica: La tabla o la gráfica es la representación de una función.

Brenda está especialmente preocupada por contestar las preguntas de acuerdo a las instrucciones planteadas, así, cuando se le pide escribir en notación funcional, no estará de acuerdo con la solución planteada por Mónica porque no está utilizando los nombres de las variables que originalmente les dieron. Después se dan cuenta que están utilizando diferentes nombres para la misma variable, *F(x)* y '*y*', optan por dejarle sólo *F(x)* porque consideran que con esa letra (*F*) harían notar que es una función. La notación funcional la utilizan bien, incluso mecánicamente. Recurren al lenguaje hablado para después usar el lenguaje escrito. Mónica deja ver que su concepción de función está vinculada con esta notación. Hasta aquí nos quedan ambiguos sus conocimientos acerca del concepto de función porque en realidad no lo necesitan para contestar las preguntas o para relacionarse con el problema.

En el pasaje 14 colocan de manera correcta las variables dependiente e independiente, probabilidad y número de hijos en la gráfica de la función de distribución, aunque espontáneamente hay un titubeo, una tendencia a coger de nuevo el número de hijos como la variable dependiente. Utilizan los valores decimales de la probabilidad, no los fraccionarios. No prevén los valores posibles de su variable dependiente para hacer su escala de una forma más visual, de manera que todo les queda amontonado en la parte de abajo. Unen los puntos por líneas.

> *Mónica hace unos ejes cartesianos en el pizarrón. En su sistema coordenado sólo incluyen el primer cuadrante.*
> Brenda: Aquí sería el número de hijos (*señala el eje de las abscisas*).
> *Mónica etiqueta los ejes: número de hijos y número de trabajadores.*
> Brenda: No, pero entonces ahí sería la probabilidad (*se refiere al nombre de la variable dependiente*).
> Brenda: Aquí (*señala el eje de las abscisas*) sería, uno, dos,… así hasta nueve.
> *Mónica pone la escala en el eje de las abscisas.*
> Brenda: Luego para arriba, ¿de qué la empezamos? ¿de 0.1, no?
> Brenda: Bueno, sería de cero hasta uno y luego ya sería irle poniendo las probabilidades. A ver, divídelo…

Así mismo se presentan otros problemas vinculados con el lenguaje. Por ejemplo ellas no ven la necesidad de expresar en la gráfica más fielmente el problema. Como si el lenguaje matemático no tuviera porqué expresar una situación en contexto. Muy probablemente se debe a que en el momento que hicieron el análisis ellas sabían perfectamente la vinculación entre el problema y la matemática. No vieron la necesidad de ser más fieles a la situación, de modo que la gráfica no resulta ser algo que les sirva para representar la situación problema (pasaje 16).

Un problema más es que ellas prefieren llamar probabilidad al número de trabajadores. Es posible que sólo lo usen para referirse de manera más fácil y ubicarse en el contexto del problema, aunque también puede deberse a que no ven esa necesidad de la matemática (es decir, de pasar a la probabilidad formal). Esto se observa en diversos pasajes, pero principalmente en el 4.

4.3.2 Análisis de los resultados obtenidos en el objeto 'La noción de variable aleatoria'

Recopilamos los resultados obtenidos sobre este objeto en la tabla 11 de manera resumida. Posteriormente se hace una síntesis general de los resultados obtenidos en este objeto.

Tabla 11. Síntesis de los resultados en la noción de variable aleatoria

Variable	Pasaje	Actitud por parte de las estudiantes	Interpretación
La variable aleatoria en la situación problema	9	No cuestionan el uso del número de hijos como la variable aleatoria	Usan la variable aleatoria de manera natural.
	9 y 12	Discuten el número de hijos más probable como solución Usan la probabilidad acumulada para introducir los sucesos «menos de tres hijos»	Usan la moda de la variable aleatoria como el suceso más probable Usan la función de distribución para resolver el problema
La variable aleatoria como función	16	Asocian las cotas superiores e inferiores del rango y el dominio de la función de probabilidad con los valores más grandes y más pequeños de la gráfica	Consideran que no es necesario indicar que la variable aleatoria no puede tomar valores decimales en el contexto matemático
	20	Consideran que el orden con el que acomodan los valores en el eje de las abscisas podría ser cualquiera	No se dan cuenta que la variable aleatoria de alguna manera ya es una abstracción matemática y que por lo tanto no podría ir en cualquier orden Ven los valores que puede tomar la variable aleatoria como etiquetas y no como números
	20	Dicen que no importa el orden de los valores de la variable aleatoria en el eje de las abscisas, al efectuar	Interpretan el número de hijos como el resultado del experimento aleatorio

		el experimento varias veces, no necesariamente tendrá que seguir ese orden	Asocian aleatoriedad a los valores de la variable aleatoria Vinculan la probabilidad con la variable aleatoria sin pasar por los eventos compuestos
	20	Hacen explícito que el número de hijos es lo que caracteriza a un grupo de trabajadores. Saben que lo que interesa del trabajador premiado son los hijos que tiene	Identifican la característica de interés como la condición que particiona el espacio muestral Identifican la regla de correspondencia de la variable aleatoria
La variable aleatoria y su distribución de probabilidad	16	Incluyen en el dominio y el rango a los valores decimales. No incluyen el cero	Para ellas la función de probabilidad está definida sólo en el trozo que grafican En su concepción, la gráfica de una función siempre es continua Para ellas no se presenta la probabilidad cero
	18	Una diferencia entre la distribución y otras funciones algebraicas es la rareza de datos discretos y acotados	Ellas están acostumbradas a trabajar funciones continuas cuyo dominio está en los reales
	19	Otra diferencia que observan en esta función es que no pueden obtener la ecuación Dicen que una ecuación en este caso no tendría sentido porque los datos son «reales» Relacionan los datos reales con la inexactitud	En sus cursos de matemáticas han trabajado poco con funciones que no estén representadas por una ecuación Una ecuación es una representación matemática que no se extrae de la realidad Para ellas lo exacto se obtiene de una ecuación
	19 y 21	Como no pueden obtener la ecuación, no pueden calcular la variable dependiente a partir de la independiente, lo que les hace pensar que la dependiente es aleatoria Deducen que la probabilidad es aleatoria	Sienten la necesidad de saber el valor de la probabilidad sin recurrir a consultar datos «reales» Vinculan la aleatoriedad con una incapacidad de cálculo (erróneamente)
	20 y 21	Mencionan que la probabilidad cero haría que la variable independiente fuera aleatoria, porque entonces no estaría en la gráfica Argumentan que la probabilidad cambiaría si la fábrica estuviera en otra ciudad o en otras condiciones y	Concluyen que en otras fábricas los valores del número de hijos sería el mismo (a menos que la probabilidad sea cero), en cambio los valores exactos de la probabilidad cambiaría Vinculan el uso de las ecuaciones

		por eso no se puede predeterminar el comportamiento de la función de probabilidad a través de una ecuación	con la capacidad de predicción o de predeterminar (¿qué pasaría en otras circunstancias?)
Lenguaje	1, 2 y 3	Escriben la probabilidad como una fracción. No efectúan la operación hasta que se les pide. Interpretan correctamente la probabilidad en el contexto del problema	La forma en que escriben la probabilidad les ayuda a su interpretación, como si el signo de división fuera una forma de escribir un razonamiento y no una indicación de la operación división
	14	Al elaborar la gráfica, usan valores decimales y unen los puntos con líneas, sin visualizar el contexto	No relacionan la gráfica (representación matemática) con el contexto del problema. La gráfica no la perciben como una forma de expresar la situación
	15	Nombran ‘x’ a la variable número de hijos y ‘y’ a la variable probabilidad en la distribución de probabilidad. Pero cuando se dan cuenta que es una función, cambian el nombre de la variable dependiente de ‘y’ a ‘F(x)’	Inicialmente usan el lenguaje común en algebra para denotar las variables Vinculan la variable F con indicar que una relación es una función, como si y(x), por ejemplo, no denotara una función
	1, 2, 6, 4, 7 y 14,	Llaman número de trabajadores a la probabilidad.	Es posible que ellas prefieran llamar número de trabajadores a la probabilidad porque es más fácil de contextualizar

La forma en cómo las estudiantes transitan de un contexto a otro al tratar de profundizar sobre el contexto matemático empleado en la solución del problema (a partir de las preguntas establecidas en el protocolo), nos hace proponer un modelo de construcción de la función de la probabilidad a partir de la situación problema a partir de la modelación desde la perspectiva de Heitele (1975). Este modelo tiene dos estratos de modelación en los que los valores de la variable aleatoria juegan dos papeles diferentes. Esto constituyó una dificultad cognitiva. Los dos estratos son:

- la vinculación de los eventos elementales con los valores de la variable aleatoria (la aplicación de la variable aleatoria como función)

- la asignación de un valor de probabilidad a cada valor de la variable aleatoria (la construcción de la función de probabilidad)[7]

En el primer estrato, los valores de la variable aleatoria son un «modelo matemático». Son el resultado de la aplicación de la variable aleatoria sobre el espacio muestral, que constituiría la «realidad». En el segundo estrato los valores de la variable aleatoria es lo que vincula al modelo matemático (la función de probabilidad) con la realidad, es decir, los valores de la variable aleatoria son vinculados directamente con el experimento aleatorio.

En el momento que las alumnas vinculan los valores de la variable aleatoria con la probabilidad, también trabajan con los valores de la variable aleatoria como si ellos fuera el resultado del experimento aleatorio. Eso les da una facilidad de interpretación y nos indica que han logrado dar un paso hacia la descontextualización de la relación entre la probabilidad y la variable aleatoria (figura 7). Pero cuando ellas aseguran que los valores de la variable aleatoria podrían estar en cualquier orden en la recta numérica no les están dando atributos de números reales, sino que sólo los están tomando como etiquetas. La razón que argumentan (les atribuyen la aleatoriedad del proceso), indica que los están interpretando como su «realidad».

Así, el doble papel que juegan los valores de la variable aleatoria (como resultado de la aplicación de la regla de correspondencia y como argumento de la función de distribución) las confunde porque en el primer momento son el «modelo matemático» de la aplicación de la variable aleatoria, pero en el segundo momento son su «realidad» en la función de distribución. Esto significaría que en el momento en que las estudiantes toman los valores de la variable aleatoria como el resultado del experimento aleatorio (un experimento aleatorio redefinido), nuevamente están pensando en eventos, no en números (reales) y por lo tanto la relación que establecen entre la probabilidad y los valores de la variable aleatoria sigue siendo una función entre la probabilidad y sus eventos y no una función vinculada con una variable numérica. Este paso implica la redefinición del experimento aleatorio en donde los resultados serían numéricos (el número de hijos).

[7] El proceso de construcción de la función de probabilidad seguido en este segundo estrato es bastante complejo. No es inmediato. Ya lo hemos descrito y discutido en la sección 4.2.2. Podría pensarse que el primer estrato también encierra una complejidad similar, sin embargo no tenemos elementos para afirmarlo.

Observemos que del proceso de modelación de la función de probabilidad, en este trabajo pudimos profundizar más en el segundo estrato porque, como ya mencionamos antes, la actividad no permitió profundizar en el primero, sin embargo, el doble papel de la variable aleatoria en los dos estratos de ese proceso permitió visualizar el primer estrato.

Consideramos que la idea de adjudicarle aleatoriedad a la probabilidad en la función de distribución, es parte de otro proceso, consecuencia de sus conocimientos y razonamientos matemáticos exitosos en contextos determinísticos y de la necesidad de hacer uso de la herramienta matemática determinística en un contexto aleatorio. Deducimos que tal idea surge de dos creencias primarias: (1) las ecuaciones no están vinculadas con datos extraídos de la realidad y (2) la generalización de una ecuación es que debe servir en varios contextos. Esquematizamos este proceso en la figura 9.

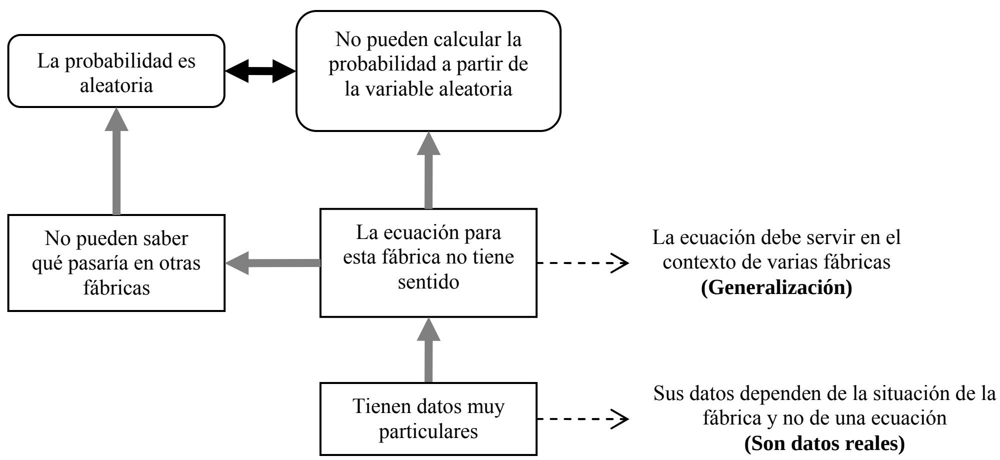


Figura 9. Proceso que siguieron las estudiantes para vincular la aleatoriedad con la incapacidad de cálculo de la probabilidad

En los párrafos siguientes discutiremos dos razonamientos alrededor de esta idea, que perjudican la descontextualización de la función de probabilidad.

Argumentan que la inutilidad de la ecuación de la función de distribución es porque no serviría para «predecir» o «predeterminar» el valor de la probabilidad de otras fábricas. Cosa por demás cierta. Sin embargo su argumento no apunta a una dificultad de encontrar una ecuación que satisfaga los datos que tienen sino a una inutilidad de aplicación de esa ecuación en otros contextos. Es posible que si la resolución del problema lo hubiera exigido, ellas hubieran determinado la ecuación sin preguntarse si serviría en otros contextos o no, pero como no fue así, ellas no sienten la necesidad de conocer la ecuación (suponiendo que se pudiera) simplemente como una forma de caracterizar mejor la situación planteada, cosa que serviría para generar herramientas teóricas a partir de la solución planteada. No vieron utilidad de descontextualizar la herramienta matemática que emplearon para resolver el problema, muy probablemente porque habían resuelto el problema sin necesidad de ello.

Por otro lado, interpretamos que cuando adjudican aleatoriedad a la probabilidad están pensando en un nivel más amplio. En realidad están cambiando el experimento aleatorio de nuestro problema por la selección de una fábrica al azar y nuestra característica de interés por observar la distribución de la proporción de trabajadores que tienen un mismo número de hijos. En tal caso, esa distribución de la proporción de trabajadores sería aleatoria.

Observemos que la proporción (*de trabajadores que tienen un mismo número de hijos*) y la probabilidad (*de que un trabajador seleccionado al azar tenga un determinado número de hijos*) tendrían el mismo valor numérico para un cierto número de hijos para cada fábrica del espacio muestral. De hecho las representaciones gráficas de sus funciones tendrían la misma forma en una misma fábrica. Sólo que la proporción no está vinculada con un experimento aleatorio y la probabilidad sí.

En su nuevo experimento aleatorio, la distribución de la proporción (*de trabajadores que tienen un mismo número de hijos*) de nuestra fábrica sería sólo la distribución de una muestra de una fábrica del total de fábricas. Es decir, sólo un dato de la función de probabilidad de las proporciones (*de trabajadores que tienen un mismo número de hijos*) y por lo tanto ésta última sería una distribución muestral. De esa forma, el tener exactamente la distribución de proporciones de nuestra fábrica en otra fábrica entre un montón de fábricas sería algo incierto.

Cuando adjudican aleatoriedad a la probabilidad en la función de probabilidad en realidad la están pensando como la distribución de proporciones de la fábrica y por lo tanto la función pierde su relación con el fenómeno aleatorio que nos interesa y se vuelve determinísta. Esta idea, que sería deseable en la construcción de la noción de distribución muestral, en nuestro tenor resulta un intento infructuoso por tratar de descontextualizar de la función de probabilidad, que desemboca en relacionar la aleatoriedad con una incapacidad de cálculo.

Desde la posición del estudio del aprendizaje de distribuciones muestrales, resulta interesante como la idea de que la proporción, aunque ellas le llamaban probabilidad, (*de trabajadores que tienen un mismo número de hijos*) es aleatoria en este estudio surgió a partir de ideas deterministas: que una ecuación debe servir en muchos contextos y de que la distribución de proporciones (para ellas función de probabilidad) estuviera definida por datos «reales».

Por último, creemos que las dificultades vinculadas al lenguaje están vinculadas a la dificultad que tienen para expresarse en lenguaje matemático, así de alguna manera, la gráfica la ven como un ente que no tiene porqué estar relacionado con el contexto del problema, es algo matemático que como que vive en sí mismo y no con un fin, digamos, práctico. Lo mismo ocurre con la probabilidad que ellas prefieren mantener como un cociente, porque eso les da facilidad de interpretación o que prefieren llamarlo «número de trabajadores» porque ellas se entienden más si vinculan con el contexto.

4.4 La solución del problema

En este apartado se analizará qué hicieron las estudiantes para resolver el problema. En particular, nos referimos a las estrategias y rutas de solución que siguieron, cuáles fueron las principales dificultades con las que se encontraron, a la simbología y al lenguaje utilizados al resolver el problema y a qué nociones matemáticas y relacionadas con la variable aleatoria recurrieron. Puesto que las estrategias exigen razonamientos que se discuten en pasajes muy largos de la entrevista, no incluimos pasajes en nuestro análisis, sino remitimos al anexo 4 para su lectura.

4.4.1 Resultados en el objeto 'La solución del problema'

En el pasaje 8 se observa que la solución inmediata es determinista. Desvinculan la decisión de comprar el número de boletos con el sorteo. Quieren que su recomendación a la empresa sea exacta. Se les dificulta un poco aceptar dar una recomendación en donde hay incertidumbre y que su respuesta pueda ser a lo más probable. En esta solución no usan los datos de la tabla, ni los datos que tienen, sólo expresan que sólo comprarían dos y esperarían el resultado de la rifa para comprar los que hagan falta. Esta postura es defendida principalmente por Mónica.

En el pasaje 9 Brenda dice que ella compraría 5 boletos y argumenta que es el valor más probable (puesto que 3 es la moda de la distribución de probabilidad y añade dos boletos para los papás de los niños). Mónica insiste en comprar sólo 2 para no hacer perder a la empresa, sin embargo cambia de opinión con un supuesto que antes no habían utilizado: los boletos tienen que comprarse por anticipado. El profesor es el que tiene que imponer ese supuesto, no son ellas las que excluyen la posibilidad de comprar boletos después de la rifa, pero es Mónica la que se da cuenta que necesita conocer ese dato extra del que no se habla implícitamente en el problema, necesita saber si se podrán comprar más boletos una vez que se efectúe la rifa.

> Mónica: Si, hay mucha probabilidad. Pero y si te salen por ejemplo estas (señala los trabajadores con menos de tres hijos). Ah profe, es que eso es lo que le preguntaba; lo que yo le estoy preguntando es muy importante porque, en caso de que yo compre dos me van a faltar, pero si tengo que comprar los exactos que lo menos pierda o puedo comprar ya después boletos.
> Profesor: Ya no, después ya no.

Mónica queda convencida de que se deben comprar 5 boletos porque es la solución más probable. Al parecer hay un acuerdo entre ambas. Pero en el pasaje 10 hay un cambio de opinión de Brenda, ahora ella duda porque comienza a hacer uso de la función de distribución de probabilidad (en la tabla que tiene escrita en el pizarrón). Lo primero que expresa es su preocupación por los trabajadores que «*no vaya a completar después*». Ella comienza a ver otras características a partir de la interpretación de la función de distribución, como el complemento de la probabilidad acumulada.

En este pasaje se observa como ambas estudiantes se posicionan en el problema, se apropian de él. Se entabla una discusión en que Brenda, para argumentar la compra de 5 boletos, usa de manera intuitiva tanto la probabilidad como la frecuencia acumulada, pero con mayor incidencia la frecuencia acumulada («*el número de trabajadores hasta aquí*»). Mónica en un principio usa la moda (como el valor más probable, «*el que menos falla*») y después el argumento de que se debe gastar lo menos posible, que no habían mencionado ninguna de las dos, pero que, probablemente, es lo que haya impulsado a Mónica anteriormente a su solución inicial inmediata de comprar sólo dos boletos. En contraste con el argumento de gastar menos, surge otra solución obvia posible que se descarta de inmediato: comprar 11 boletos para no introducir riesgo alguno («si *yo quiero que nadie se quede sin boletos, pues entonces de una vez compran los... 11...de una vez para ya no errarle*»). Mónica presta menos atención al argumento dado por el uso de la probabilidad (o de frecuencia) acumulada que a la preocupación de Brenda porque a la familia premiada tenga más miembros que los boletos que recibirán. Hay un pequeño debate entre los argumentos generados por los usos de probabilidad acumulada y de la moda y entre el argumento de dejar a los trabajadores sin boletos y de gastar lo menos posible. La discusión es desigual porque no usan la misma herramienta matemática, sino cada quien la suya (Mónica la moda, Brenda la probabilidad acumulada y la mediana). El profesor les pide que escriban el cálculo de la probabilidad acumulada (Brenda sólo había estado haciendo cálculos mentales aproximados) y se rompe la discusión entre ambas porque interpretan la intervención del profesor como una guía velada sobre lo que espera de ellas. No vuelven a utilizar a la moda como argumento. De manera natural ellas no deciden calcular la probabilidad acumulada aunque la usan, lo mismo pasa con la probabilidad del complemento, que de hecho, finalmente, no calculan. En los argumentos utilizados por Brenda en esta parte también hay una referencia a la mediana de la distribución.

En el pasaje 11, ahora Mónica también usa la probabilidad acumulada para argumentar cuántos boletos comprarían «*si tienes cubierto todo esto*». Dudan y no se convencen mutuamente. Ahora sí, una vez puestas en la misma herramienta matemática, se observa claramente la posición de cada una de las dos: Una, Mónica, argumenta que se debe ahorrar lo más posible (viendo la empresa) y la otra, Brenda, se enfoca a las posibilidades de que el número de boletos no alcance (tratando de ponerse del lado del trabajador). No pueden desligarse de su posición con respecto al contexto. Argumentan que hay una decisión que depende de la ética, que no tiene nada que ver con matemáticas y se quejan con el profesor de que el problema esté planteado así.

Aparentemente hay un acuerdo entre ambas, que implica también una aceptación de la aleatoriedad de la situación y aceptar un riesgo de ambas partes, de la empresa y de los trabajadores.

> Brenda: En todo caso si se está buscando que no vaya a gastar de más y si alcanzan bueno y si no, no, pues yo estaría por 4 hijos.
> Mónica: Ajá, yo también y si no alcanzan pues ya ni modo.

Ante la respuesta de comprar 6 boletos (4 para los hijos y dos para los padres), nuevamente el profesor interviene. Les dice que la recomendación tiene que estar en lo mínimo que se pueda gastar. De esa forma ambas llegan a un consenso: comprarían 5 boletos con una probabilidad de 0.58 de que los boletos les alcancen exactamente o sobren. Sin embargo esta respuesta está condicionada por la intervención del profesor que, a diferencia de la primera, ellas no la habían solicitado.

4.4.2 Análisis de los resultados obtenidos en el objeto 'La solución del problema'

Las estudiantes proporcionan diferentes soluciones utilizando diferentes herramientas y bajo distintos argumentos, también van incorporando supuestos a medida que descartan soluciones y se apropian del problema.

Tabla 12. Síntesis de las soluciones proporcionadas por las alumnas

Solución	Argumento	Herramientas	Motivo que las hace descartar la solución
2 boletos	La empresa debe perder lo menos posible	No usan los datos de la tabla de probabilidades Solución determinista	Cuestionan al profesor sobre si es posible comprar boletos después de la rifa. El indica que no
5 boletos	Es más probable que el trabajador premiado tenga 3 hijos	Usan la moda de la distribución de probabilidad	Se preocupan por los trabajadores que tienen muchos hijos
11 boletos	Es seguro que le alcanzarían los boletos al trabador premiado	Usan la probabilidad acumulada y la noción de evento seguro	La empresa correría el riesgo de perder demasiado. Es descartada de inmediato
6 boletos	Se busca favorecer a los trabajadores pero también que la empresa no gastarte de más	Hay una interpretación de la función de distribución a través del uso de la probabilidad acumulada y del complemento de la probabilidad acumulada En la discusión aparece el uso de la mediana de la función de probabilidad Hay una aceptación de la aleatoriedad como riesgo	El profesor interviene recordando que se debe gastar lo menos posible
5 boletos	Se busca favorecer a los trabajadores pero que la	Hay una interpretación de la función de distribución a	Está es la solución que ellas finalmente proporcionan.

	empresa se arriesgue lo menos posible.	través del uso de la probabilidad acumulada y del complemento de la probabilidad acumulada No queda muy claro si recurren a la moda o a la mediana o a ambas Hay una aceptación de la aleatoriedad como riesgo	

Observemos que su segunda solución de 5 hijos no es igual a su última solución (también de 5 hijos). En esta última ya tenían otro tipo de argumentos y hacían uso de otro tipo de herramientas estocásticas. Es posible que ellas necesitaran de ese tránsito entre varias soluciones para que hicieran uso de esas herramientas y fueran conscientes del riesgo de su decisión.

Así mismo hay una intervención del profesor que hace que ellas se sientan seguras de su solución y que no la cuestionen más. Muy probablemente si el profesor no hubiera intervenido, la solución con la que ellas se hubieran quedado hubiera sido la de comprar 6 boletos que fue su conclusión después de su pequeña discusión. De alguna manera fue la que les hizo sentirse bien con los trabajadores pero que no cumple con las condiciones del problema (porque en el problema se especifica una mínima inversión). Notemos que en realidad las dos soluciones finales no son muy distintas en cuanto a la argumentación matemática que utilizan, la diferencia está en la utilización de un mínimo riesgo a la empresa. El criterio de un mínimo riesgo para la empresa, no obstante, sigue estando por debajo del criterio de favorecer a los trabajadores, que es la respuesta que se esperaba.

La discusión de las estudiantes puso en claro los momentos en que se introducen los criterios de decisión y cómo se están poniendo en juego. En la figura 10 se desglosan este proceso.

Si el único criterio hubiera sido favorecer al trabajador, la decisión hubiera sido determinista, el sorteo no hubiera intervenido en la decisión y por lo tanto no habría incertidumbre en la compra de boletos. El problema aquí es que al mismo tiempo que se quiere favorecer al trabajador, también se quiere ahorrar lo más posible, en el sentido de no gastar en lo innecesario (no comprar boletos de más). La respuesta de comprar 6 boletos demuestra que ellas preferían que la empresa gastara más a dejar a una familia con boletos insuficientes. Así, ellas no usaban la función de distribución acumulada del número de hijos para medir la probabilidad de que la empresa gastara de más, sino para medir la probabilidad de que al trabajador ganador le alcancen los boletos. La función de distribución del número de hijos podía ser interpretada de esas dos maneras.

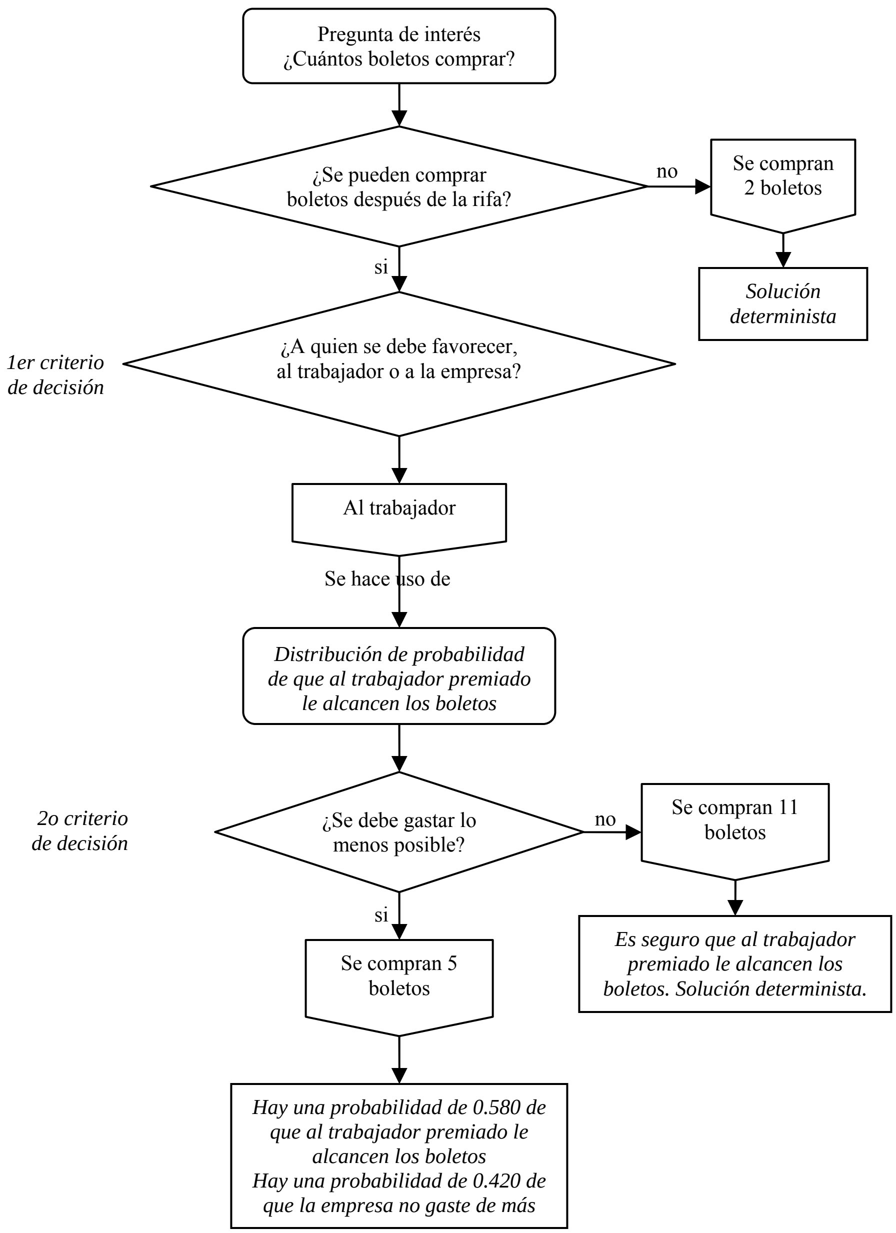


Figura 10. Ruta de solución poniendo explícito que se pretende favorecer a los trabajadores antes que a la empresa

Sin embargo, en la discusión de las alumnas también surgieron las interpretaciones de que la empresa no gaste de más y de que al trabajador ganador no le alcancen los boletos cuando se referían al complemento de la función de distribución (tabla 13). Esta función sería más importante si cambiamos el orden de los criterios de decisión y se quisiera favorecer a la empresa antes que al trabajador. En tal caso se comprarían 4 boletos. Este proceso se ilustra en la figura 11.

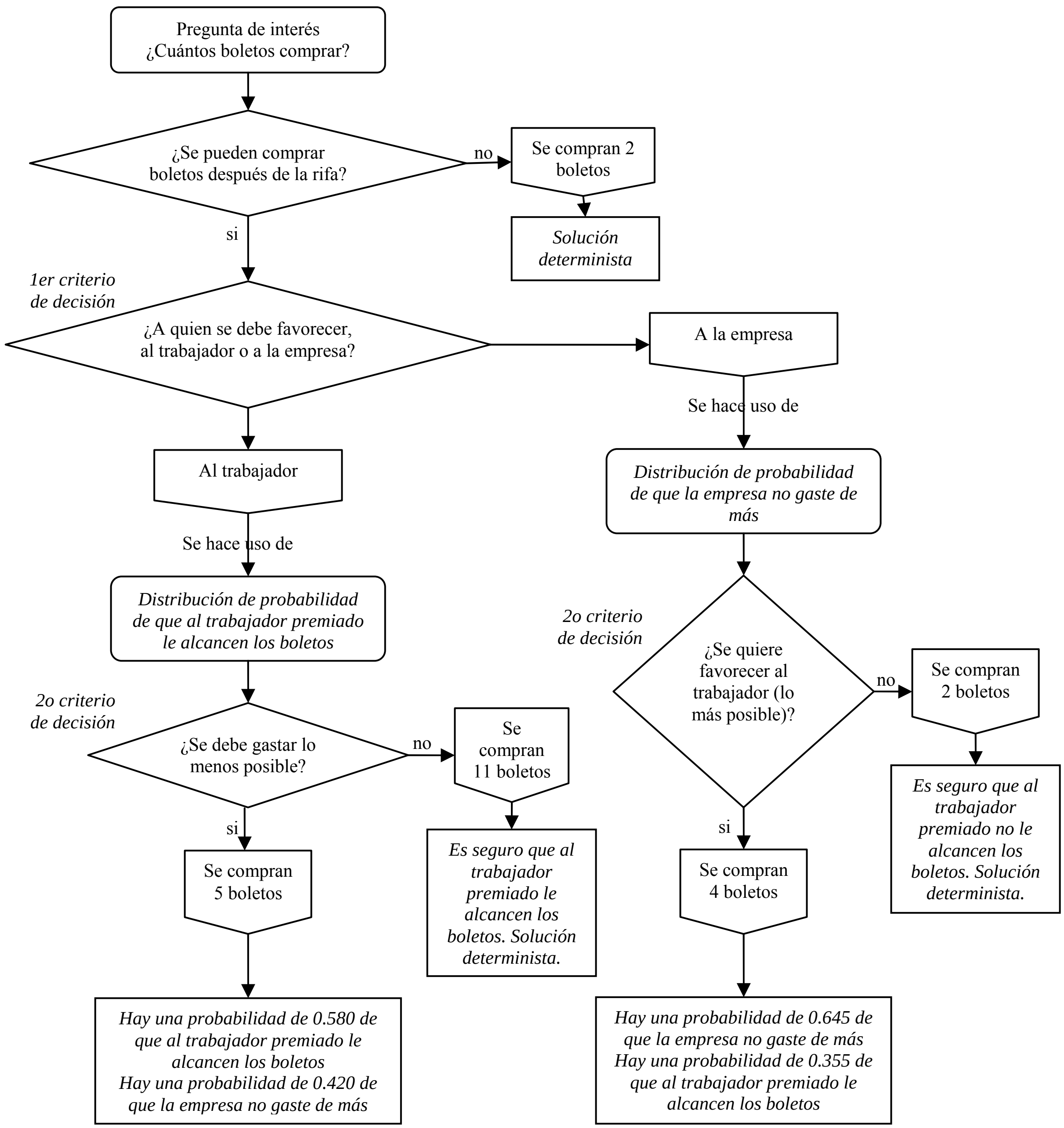


Figura 11. Ruta de solución tomando en cuenta el cambio de los criterios de solución

Es factible cambiar el orden de los criterios de decisión porque no podemos encontrar un número de boletos que proporcione exactamente el 50% (eso garantizaría ecuanimidad), de modo que se tiene que favorecer a una de las dos partes.

Desde la perspectiva ética, el problema se trabajó pensando en la recomendación a la empresa sobre cuál es la situación que favorecería más a los trabajadores y en la que ella gastaría menos. A partir de esa recomendación, los directivos decidirían cuántos comprarán.

Tabla 13. La función de probabilidad del número de hijos, función de distribución y el complemento de la función de distribución

Número de hijos (x)	**Distribución de Probabilidad**	**Función de distribución** $F(x)=P(X \le x)$	**Complemento** $1-F(x)$
	Probabilidad de que el trabajador premiado tenga *x* número de hijos $p(x)=P(X=x)$	**Probabilidad de que al trabajador premiado le alcancen los boletos (Probabilidad de que la empresa gaste de más)**	**Probabilidad de que la empresa no gaste de más. (Probabilidad de que el trabajador premiado no le alcancen los boletos)**
0	0.080	0.080	0.920
1	0.110	0.190	0.810
2	0.165	0.355	0.645
3	0.225	0.580	0.420
4	0.155	0.735	0.265
5	0.100	0.835	0.165
6	0.060	0.895	0.105
7	0.045	0.940	0.060
8	0.035	0.975	0.025
9	0.025	1.000	0.000

4.5 Observaciones al protocolo de la entrevista clínica

Observamos que el planteamiento del problema trajo algunos problemas en el análisis de la solución del problema. Por ejemplo, en el problema la moda era la misma que la mediana, de modo que algunas veces no podíamos saber si las argumentaciones de las estudiantes se basaban en el análisis de la moda o de la mediana. Aunque eso también favoreció la discusión de los dos criterios de decisión que se ponían en juego y los hizo evidentes como un factor que influye en la resolución del problema.

Así mismo nos dimos cuenta de la necesidad de hacer explícita la condición de que no se podían comprar boletos después de la rifa porque eso fue lo que hizo que la resolución del problema involucrara el sorteo y por lo tanto la aleatoriedad. En cambio el criterio no explícito de que había que favorecer a las familias de los trabajadores lo más posible no fue necesario porque ellas sintieron un deber ético de ponerse de su lado, eso hizo que en su solución última se olvidaran del criterio explícito de no despilfarrar recursos de la fábrica.

También vimos que al proporcionar los datos agrupados en la tabla de acuerdo a la variable aleatoria y al número de trabajadores, impidió la observación de la construcción de la variable aleatoria como función. Es posible que si el problema hubiera sido más abierto, se hubiera promovido la necesidad de los datos y la agrupación de los eventos compuestos y por lo tanto una mejor exploración de la generación de la característica de interés y de la regla de correspondencia.

Sin embargo también constatamos que la situación, a pesar de que debe ser mejorada, contiene algunos elementos interesantes como base para el diseño de una futura ingeniería didáctica que recoja esta y otras situaciones y sirva para ayudar al alumno a desarrollar la idea de variable aleatoria. Por un lado, la situación fue interesante para los estudiantes quienes trabajaron en ella hasta completarla. Pensamos que se consigue la devolución del problema a las estudiantes, quienes incluso llegan a pensar en un momento de que no se trata de un problema matemático.

Por otro lado la situación, junto con la forma en que es llevada a cabo la entrevista provoca el debate entre las estudiantes, quienes en todo momento están interesados en resolver las tareas y se sienten cómodas con la misma. El debate sirve para explicitar sus concepciones, dificultades y estrategias.

También la situación moviliza el aprendizaje de las alumnas, como hemos analizado al describir la solución y los pasos seguidos en la misma. Respecto a ello, encontramos ideas espontáneas correctas sobre aleatoriedad y probabilidad. Por ejemplo, no tuvieron dificultad en identificar correctamente el espacio muestral, y reconocen cómo depende de las circunstancias del experimento. Hay una buena interpretación y uso de la probabilidad, tanto desde un punto de vista laplaciano, como desde un punto de vista sujetivo, como grado de creencia, aunque no ocurre así con la probabilidad formal. También reconocen espontáneamente dos de los axiomas: que la probabilidad es positiva y menor que uno y que la suma de todas las probabilidades ha de ser igual a la unidad. Muestran correcta comprensión de los números decimales y de la probabilidad como decimal.

Las alumnas encuentran natural usar el número de hijos como valor de una variable aleatoria. Manejan bien el conjunto de valores que toma, aunque interpretan el rango y dominio sólo a los máximos y mínimos de estos conjuntos.

Usan correcta y espontáneamente las ideas de moda (como lo que menos falla), probabilidad acumulada (función de distribución), mediana (como división de la población en dos partes) y equitatividad (valor esperado). También se dan cuenta de cómo la probabilidad proporciona una solución al problema y de cómo el numerador de la misma se representa en la tabla de datos por medio de frecuencias. Posteriormente aparecen las ideas de probabilidad acumulada y de complemento de una probabilidad.

Capítulo 5

Conclusiones
del análisis cognitivo

Las conclusiones obtenidas a partir del análisis cognitivo se deducen principalmente del análisis de la entrevista clínica. Primeramente presentamos las conclusiones que se obtienen de una confrontación entre las hipótesis planteadas y sobre el diseño de la actividad que ha guiado la exploración cognitiva. Y posteriormente finalizamos proporcionando algunas líneas, en forma de sugerencias, para continuar nuestro trabajo con el análisis epistemológico.

5.1 Confrontación de la observación con las hipótesis

En las hipótesis de la entrevista clínica (punto 3.3.1) se manifestaron algunas de las ideas sobre lo que esperábamos encontrar en nuestro estudio empírico. A continuación discutimos las conclusiones que pueden deducirse del estudio de los resultados empíricos.

5.1.1 Falta de percepción de la aleatoriedad del proceso

La aleatoriedad del proceso aleatorio (el sorteo) no fue cuestionada. Sin embargo hubo una reticencia a aceptar dar una recomendación que no sabían si sería certera. En las primeras secuencias del trabajo las alumnas se negaban a incluir el sorteo en su análisis porque eso haría que la solución «heredara» la aleatoriedad del sorteo. Una vez aceptada la aleatoriedad en la solución, buscaron argumentos de equidad (entre los trabajadores y la directiva de la empresa) para proporcionar una solución con la que pudieran sentirse más confiadas.

Las estudiantes manejaron un buen concepto de aleatoriedad. Manifestaron no poder predecir cuál sería el número de hijos que tuviera el trabajador premiado y que si repitieran el sorteo muchas veces, no sabrían cuál es la secuencia que seguirían los resultados. Sin embargo en un intento por formular la función de probabilidad del problema, vincularon a la aleatoriedad con una incapacidad de cálculo (no tenían una ecuación que vinculara la probabilidad con la variable aleatoria y por ello no la podían «calcular»). De modo que asignaron aleatoriedad a la probabilidad (en realidad sería la proporción de trabajadores que tuvieran el mismo número de hijos).

5.1.2 Tendencia a «algebraizar» y descontextualizar los procedimientos relacionados con la noción de variable aleatoria

Las estudiantes manifestaron un buen manejo de las herramientas matemáticas necesarias para resolver el problema. Sin embargo, al contrario de lo esperado, para las alumnas fue más difícil pasar del contexto del problema al contexto matemático. Varias fueron las razones que nos llevan a concluir que el proceso de contextualización es más simple para ellas que el de descontextualización en el proceso de construcción del concepto de variable aleatoria:

- Hay una tendencia a centrarse preferentemente en el número de trabajadores, en lugar de en su probabilidad. Esto ocurrió cuando trataron de descontextualizar la función de probabilidad (al inicio de la actividad) y cuando argumentaban alguna propuesta de solución del problema.

- A pesar de que hay una interpretación apropiada de la probabilidad clásica, no dejan de ver a la probabilidad como un cociente de dos números en lugar de un solo número.

- Están constantemente recordando que la variable aleatoria 'Número de hijos' esta vinculada con el conjunto de trabajadores que tienen ese número de hijos.

- Grafican únicamente los datos que tienen. Así por ejemplo no incluyen al cero en el rango de la función de probabilidad a pesar de que habían mencionado que la probabilidad de que el trabajador premiado tuviera diez hijos sería cero.

- Manifiestan que no es forzoso que en el eje numérico (de la gráfica de la función de probabilidad) los valores de la variable 'número de hijos' tengan un orden ascendente, puesto que son aleatorios.

- Argumentan que no sería útil encontrar una ecuación de la función de probabilidad en el caso del problema planteado porque se trabaja con datos demasiado «reales».

- ❖ No encontraron una necesidad de descontextualizar la herramienta matemática empleada en la resolución del problema, muy probablemente porque el proceso de resolución no se los exigió.

A pesar de eso, también en algunos pasajes desvinculan las matemáticas del contexto del problema, mostrando una concepción de las mismas como algo que se justifica en sí mismo, sin tener una relación con una situación del contexto real. Algunos de los argumentos que expresan las alumnas y que usamos para sostener esta afirmación son:

- ❖ Piensan que una ecuación debe servir para hacer predicciones en distintas situaciones y no para una en particular.
- ❖ Creen que los datos de una función provienen de un contexto matemático y no de la «realidad».
- ❖ Piensan que una gráfica debe ser continua, por lo tanto, el dominio y el rango de una función también, no importa que el contexto del problema sea discreto.
- ❖ Piensan que una gráfica es infinita.

Estos puntos están vinculados con las evidencias de descontextualización, lo que refuerza la desvinculación entre los conceptos matemáticos y el contexto de un problema. Es decir, no solamente se mira a la matemática como algo sin contexto, sino que también se mira el contexto como algo que no puede ser representado a través de la matemática. Sin embargo en nuestro caso, fue más fácil para ellas contextualizar que descontextualizar, quizá porque estaban sumergidas en el contexto de un problema y hubo necesidad de descontextualizar para resolver el problema, es decir, de interpretar los datos que se les proporcionaron y que ya eran abstracciones.

5.1.3 Extrañeza de trabajar funciones en un contexto probabilístico

Las alumnas no encontraron extraño asociar el término función a la distribución de probabilidad. Esta facilidad de aceptación fue debida a que la pudieron escribir muy naturalmente en notación funcional y porque recurrieron a la definición y observaron que concordaba con ella. También fue inmediato el uso de la representación tabular para las frecuencias y probabilidades. Sin embargo sí encontraron extraño que hubiera funciones con las características encontradas para la función de probabilidad, por ejemplo:

- Encontraron raro trabajar con una función discontinua (una función puntual), donde la variable independiente es discreta.
- Hallan restringido el intervalo de valores de la variable independiente porque sólo toma valores entre 0 y 9.
- Encuentran excepcional no poder encontrar los valores de la variable dependiente a partir de la variable independiente a través de una ecuación.
- Se les hace extraño que la función esté definida a partir de datos «reales».
- Piensan que es novedoso que la ecuación de la función en caso de que pudiera ser obtenida, no les serviría para conocer la probabilidad de cualquier fábrica.

La idea de que si existiera la ecuación no les serviría para conocer datos de otras fábricas, junto con no poder encontrar la ecuación de la función las lleva a la idea errónea de que la aleatoriedad de la probabilidad está vinculada con una incapacidad de cálculo de la probabilidad.

5.1.4 Dificultades con la noción formal de variable aleatoria.

No pudimos observar la construcción de la regla de correspondencia de la variable aleatoria porque la forma en cómo estaba proporcionada la información en el problema no lo permitió, pero sí se tienen conclusiones en el manejo y la interpretación de las estudiantes de la variable aleatoria.

Las estudiantes se dieron cuenta de que la regla de correspondencia vinculaba el espacio muestral con los valores de la variable aleatoria. Pudieron percatarse, también, que esa regla de correspondencia provenía de una característica de interés (extraída del planteamiento del problema) que agrupaba a los trabajadores en eventos compuestos. Así mismo interpretaron correctamente que la probabilidad se obtenía a partir de los eventos compuestos y que esto era lo que permitiría asignarle un valor de probabilidad a cada valor de la variable aleatoria. Así, concluimos que se tuvo una buena interpretación y manejo de los valores de variable aleatoria. Dada la regla de correspondencia de la variable aleatoria y los valores que toma, contextualizan correctamente los elementos matemáticos en la situación problema.

Se presentaron dificultades, sin embargo, al asignarle un valor de probabilidad a cada valor de la variable aleatoria porque eso requería no sólo la formulación de la probabilidad como la relación entre dos conjuntos de personas (las favorables y las posibles) sino también el planteamiento de los valores de la variable aleatoria como un valor numérico y no como la característica de interés que agrupó a los trabajadores en los eventos compuestos. Esto provocó que las estudiantes prefirieran trabajar con la relación número de trabajadores-valores de la variable aleatoria, que con la relación probabilidad-valores de la variable aleatoria.

Pudimos observar que en el proceso de construcción de la función de probabilidad a partir de la situación problema, se generaron dos estratos de modelación desde la perspectiva de Heitele (1975):

- La aplicación de la variable aleatoria como función: Vinculación de los eventos elementales con los de la variable aleatoria.
- La construcción de la función de probabilidad: asignación de un valor de probabilidad a cada valor de la variable aleatoria.

De aquí surge la dificultad cognitiva de que los valores de la variable aleatoria es «modelo matemático» en el primer estrato y «realidad» en el segundo. Esto provoca que las estudiantes piensen que los valores de la variable aleatoria en el eje numérico pueden ir acomodados en cualquier orden porque los ven como el resultado de un experimento aleatorio y no como números con propiedades de campo.

Observamos también que en el intento por descontextualizar la función de probabilidad, las estudiantes pasan por el proceso de transformar el resultado del experimento aleatorio, ya no piensan que del sorteo se obtendrá el nombre de un trabajador sino un número de hijos. Esto hace que el experimento se transforme en un experimento con resultado numérico.

También en el intento por la descontextualización de la función de probabilidad, ellas desvinculan la función de probabilidad del evento aleatorio y la transforman en la distribución de frecuencias de la fábrica. Esto las conduce al planteamiento de una función de distribución muestral, en donde nuestra fábrica proporcionaría sólo un dato y los valores de sus frecuencias (probabilidades para ellas) serían aleatorios.

5.1.5 Dificultades respecto a los conceptos que intervienen en la definición formal de variable aleatoria

Consideramos que no se consiguió una definición formal de la variable aleatoria (al nivel que esperábamos) sin embargo sí hubo una buena aproximación a ella. En esa aproximación intervinieron diversos conceptos probabilísticos y estocásticos. Sobre algunos de ellos, como la aleatoriedad, ya hemos concluido.

Las alumnas usan la idea de probabilidad en forma natural, interpretándola como una medida de lo incierto, pero al aplicar una concepción clásica de probabilidad, la consideran una razón y no un único valor numérico. Esto contribuye a que no vinculen directamente a la probabilidad con el valor de la variable aleatoria, más que a través de los eventos compuestos. Es decir representa un obstáculo de descontextualización.

De lo anterior se deduce que el tipo de asignación de probabilidad puede influir en el proceso de planteamiento de la función de probabilidad y por lo tanto de la variable aleatoria.

Consideramos que la interpretación de los eventos compuestos juega un papel muy importante, en cuanto a que son los que permitieron la vinculación entre el experimento aleatorio y la probabilidad al construir la función de probabilidad.

Así mismo la definición de un nuevo espacio muestral numérico a partir del mismo experimento aleatorio, constituyó un paso hacia la conceptualización de la variable aleatoria.

También concluimos que la construcción de los conceptos de función de la probabilidad y de variable aleatoria, necesarios para la construcción de nuevos conceptos matemáticos, son procesos que están muy intrínsicamente vinculados y que se condicionan el uno al otro.

5.2 Idoneidad de la situación

La situación resultó exitosa en cuanto a que las estudiantes se posicionaron en ella y la llegaron a tratar como una situación a-didáctica. Lo que permitió explorar sus concepciones, dificultades y estrategias, así como poner en juego sus concepciones matemáticas en diversos aspectos, tanto en el contexto de la probabilidad y estadística como en el de la matemática determinística.

La actividad también permitió que las alumnas avanzaran hacia la conceptualización de la variable aleatoria y que se cuestionaran sobre sus concepciones que intervienen para construir el concepto de variable aleatoria.

Así mismo también nos pudimos percatar que el diseño de una situación problema más abierta podría proporcionarnos una mejor exploración de los procesos de construcción de la regla de correspondencia de la variable aleatoria.

5.3 Sugerencias para el análisis epistemológico

Es de interés analizar el concepto de variable aleatoria desde la teoría de probabilidad y hacer patentes las relaciones que en ella se establecen entre los conceptos que intervienen. En particular:

- La forma en que la teoría asigna probabilidad a la variable aleatoria.
- La diferencia que se establece entre la probabilidad cuando está definida en el espacio muestral y cuando está definida en la función de probabilidad.
- El papel de los eventos compuestos y de los eventos simples en la definición de variable aleatoria.

Así mismo dentro del análisis histórico sería interesante profundizar sobre dos tópicos en particular:

- la influencia de la herramienta matemática determinística en la definición del concepto de variable aleatoria.
- el papel de los fenómenos aleatorios con resultados numéricos en la construcción del concepto de variable aleatoria.

Así mismo consideramos de sumo interés estudiar a mayor profundidad la naturaleza multifacética de la probabilidad y cuáles son las dificultades que históricamente se han tenido en la formalización del concepto de probabilidad.

Parte III

Análisis Epistemológico

Capítulo 6

Análisis Epistemológico desde la Disciplina

El análisis de la variable aleatoria desde la disciplina pretende indagar sobre la naturaleza e interrelaciones del concepto que se pretende enseñar. Se trataría de delimitar el objeto de estudio, esto es, el *saber sabio*, desde la perspectiva de Chevallard (1985/1991) o el *significado de referencia* desde el marco teórico de Godino (1996: 2002). Nuestro trabajo se basará en la ingeniería didáctica, como en la concepción de herramienta-objeto de Douady (1986) y en la de campos conceptuales de Vergnaud (1990). Así en esta aproximación al concepto nos enfocaremos a una profundización desde la teoría de la probabilidad.

En cuanto al análisis del concepto tomamos libros que consideramos de referencia para nosotros porque nuestra intención en los cursos universitarios no está enfocada a la enseñanza de una teoría rigurosa de probabilidades.

En primera instancia recurrimos a libros de probabilidad, cuyo propósito principal está en la enseñanza de la teoría de probabilidades en niveles universitarios o de postgrado. Tres de ellos, Meyer (1970/1973), Feller (1968/1973) y Wackerly, Mendenhall y Scheaffer (2002), tratan de presentar, junto con el desarrollo de la teoría probabilística, aplicaciones, a la misma probabilidad, a la estadística o a diversas áreas de la ingeniería y a sí mismos se califican de no rigurosos. El resto de los libros: Krickeberg (1962/1973), Mood y Graybill (1952/1955), Hoel, Port, y Stone (1971), Petrov y Mordecki (2002) y Cuadras (1999) describen la teoría matemática desde una perspectiva más rigurosa y sin más aplicaciones que las propias de la misma teoría.

Los ocho libros que utilizamos para este análisis tienen en común que se presentan a sí mismos como adecuados para un primer curso de teoría de probabilidades, no presuponen conocimientos previos dentro de la probabilidad o conocimientos muy profundos de la teoría de la medida, más bien sólo piden como requisito profundización en el análisis matemático.

Recurrimos también a otros libros que añaden otros objetivos diferentes a la propia exposición del tema, que sin embargo proporcionan elementos importantes para concluir nuestro análisis. Nos referimos al Boudot (1979) que trata el tema con un corte filosófico y a Godino, Batanero y Cañizares (1996) que trata de mostrar los temas de azar y probabilidad desde el aspecto didáctico. De éste último tomamos elementos que completan solo la perspectiva epistemológica.

Son diversos los problemas relacionados con la variable aleatoria, desde tomar una buena decisión hasta describir una población. A partir de ellos se derivan otros problemas de tipo matemático, tanto en probabilidad como en estadística, como la estimación de parámetros, las pruebas de hipótesis y las distribuciones muestrales, o la definición de nuevas variables aleatorias a partir de una o más variables aleatorias, como la definición de suma, multiplicación, máximos y mínimos, o incluso la composición de variables aleatorias. La solución matemática a estos problemas pasa por definir cuál es la variable aleatoria vinculada a un experimento aleatorio. Consideramos así, a la variable aleatoria no sólo como un concepto matemático, sino como un instrumento de resolución de problemas definido por un proceso, consistente en encontrar la regla de correspondencia que cumpla con ciertos criterios y siempre vinculada a un contexto real. En ambos contextos la variable aleatoria está fuertemente vinculada con la definición de su función de distribución.

6.1 El concepto

El concepto que nos ocupa parte de una situación aleatoria (experimento), cuyo resultado es importante en la medida en que es la respuesta a un acontecimiento que puede ser único e irrepetible para una persona, o por ser el resultado esperado ante una predicción previa. Pero ante la imposibilidad de predecir con exactitud cada resultado en particular, el estudio de un experimento requiere del análisis de todos los eventos posibles que pueden ocurrir. Ese análisis nos brindará la posibilidad de establecer un criterio basado en las probabilidades de todos los resultados posibles.

Esto es, «para caracterizar el resultado de un experimento, no basta con decir que se ha producido un determinado suceso, sino que es preciso describir ese suceso dando cuenta de las diversas medidas que se hayan efectuado. La variable aleatoria, es decir, la magnitud que 'depende del azar' es entonces indispensable» (Boudot, 1979, p 332). La variable aleatoria es la herramienta matemática que permite pasar del estudio de sucesos aislados al estudio de las distribuciones de probabilidad y por lo tanto es también la responsable de la aplicación del análisis matemático y de otras herramientas matemáticas a la estadística (Batanero, 2000).

En palabras muy simples, la variable aleatoria es una *función que asocia un valor numérico a cada evento del espacio muestral asociado a un cierto experimento aleatorio* (Johnson, 1998/1999). Una vez situados en el contexto numérico, es factible trabajar con los valores que adopta la variable aleatoria en lugar de con eventos o sucesos y por lo tanto, también lo es aplicar diferentes herramientas de análisis matemático a las distribuciones de probabilidad, dependiendo de las características de la variable aleatoria. Esto significa que este concepto tan aparentemente sencillo transforma los sucesos a términos numéricos y permite modelar la relación del espacio muestral con la distribución de probabilidad en forma funcional. Pero su definición formal no resulta fácil de interpretar por lo que amerita que nos detengamos en ella.

Una definición formal de variable aleatoria se reproduce a continuación:

> Sea Ω, el espacio muestral asociado a un cierto experimento aleatorio, A un álgebra de sucesos definida en dicho espacio muestral y P una medida de probabilidad definida sobre A, es decir una aplicación:
>
> $$\begin{array}{rccc} P: & A & \rightarrow & [0,1] \\ & \mathrm{A} & \rightarrow & P(\mathrm{A}) \end{array}$$
>
> tal que se cumplen los tres axiomas:

a. $0 \leq$ P(A) ≤ 1, para todo suceso A del álgebra de sucesos *A*

b. P(Ω)=1

c. Si $(A \cap B) = \phi$, entonces p(A∪B)= P(A) + P(B)

En las condiciones anteriores se dice que (Ω, *A*, *P*) es un *espacio de probabilidad* o espacio probabilístico. Hacemos notar que el axioma c. se generaliza para cualquier número de sucesos incompatibles dos a dos.

Sea ahora (Ω, *A*, *P*) un espacio de probabilidad y R el cuerpo[8] de los números reales. Se dice que la aplicación:

$$\begin{array}{cccc} X: & \Omega & \to & R \\ & \omega & \to & X(\omega) \in R \end{array}$$

que a cada suceso elemental hace corresponder un número real, es una variable aleatoria[9] si para todo número real x, se verifica la relación:

$$A = \{\omega | X(\omega) \leq x\} \in A \quad \text{.................................}(1)$$

es decir, se verifica que A es un suceso. A se indica abreviadamente por $[X \leq x]$

(Cuadras, 1999).

La condición (1) indica, en palabras, que la imagen inversa en la aplicación «variable aleatoria» para todos los intervalos reales acotados superiormente es medible, puesto que, para cada conjunto del álgebra de sucesos está definida la probabilidad. Así, la definición matemática de la variable aleatoria exige que la imagen inversa de todo $X^{-1}(\omega)$ sea un elemento del álgebra *A*, porque una vez definida una medida de probabilidad P sobre (Ω, *A*), la variable aleatoria puede determinar una medida de probabilidad sobre (R, $\mathcal{B}$), en donde $\mathcal{B}$ es la sigma-álgebra construida por los conjuntos de Borel en R. De esta forma, la variable aleatoria induce una medida normada sobre los conjuntos que representan a los sucesos (Godino, Batanero y Cañizares, 1996).

[8] En algunos libros se le denomina campo.
[9] Algunos autores las denominan *magnitudes aleatorias*.

En realidad la variable aleatoria está definida para todo suceso del álgebra de sucesos, y no sólo para los puntos muestrales (elementos del conjunto Ω). Es decir se trata de una aplicación de A en R, lo que garantiza que la imagen inversa de cualquier elemento del conjunto imagen pertenezca a A y por tanto podamos posteriormente asignarle su probabilidad, que estaba definida previamente sobre A.

A veces las variables aleatorias están ya implícitas en los puntos muestrales, cuando el resultado del experimento aleatorio es numérico, por ejemplo, si el experimento consiste en observar el tiempo de espera a un autobús. Pero en otros casos, en un mismo experimento aleatorio podemos definir diferentes variables aleatorias. Por ejemplo, al lanzar tres monedas al aire podemos asignar a cada suceso la variable “número de águilas”, pero también el “número de soles”. Por ello no debe confundirse el experimento con la variable aleatoria ni el espacio muestral del experimento con el conjunto de valores de la variable.

El conjunto imagen de una variable aleatoria puede ser *discreto*, cuando toma un número finito o infinito numerable de elementos o *continuo*, si toma un número infinito no numerable de elementos. Las variables aleatorias definidas sobre espacios muestrales discretos se llaman discretas y las definidas sobre espacios muestrales continuos se llaman continuas.

6.2 Función de distribución de una variable aleatoria

La relación (1) hace factible establecer una relación funcional (*función de distribución*)[10] entre el conjunto de números reales y el intervalo [0,1]. Mediante esta función podemos asignar a cada elemento *x* de *R* la probabilidad del subconjunto del espacio muestral al que la variable aleatoria asigna un valor menor o igual al número *x* dado. Es decir:

$$F(x) = P(X \leq x) = P(\omega / X(\omega) \leq x)$$

Este proceso, haciendo uso de notación matemática, se ilustra en la figura 12.

[10] Los autores de los libros estadísticos consultados difieren en la terminología utilizada. La *función de distribución* de una variable aleatoria se denomina a veces *función de distribución acumulada*.

$$\begin{array}{lcccc} R: & \xrightarrow{\quad X^{-1}\quad} & A & \xrightarrow{\quad P\quad} & [0,\,1] \\ x: & \longrightarrow & \mathbf{A}=\{\omega \mid X(\omega)\le x\} & \longrightarrow & \mathrm{P(A)} \end{array}$$

Figura 12. Asignación de probabilidad a los valores de la variable aleatoria

Dicho de otro modo, mientras que un valor del conjunto imagen en una variable aleatoria era la *variable dependiente* en la función de conjunto (variable aleatoria), ahora juega el papel de *variable independiente* en la función numérica (función de distribución) y el valor numérico de la probabilidad asignada al valor imagen de la variable aleatoria, que es un número en el intervalo [0.1], seria la variable dependiente en la función de distribución.

Con la función de distribución podemos trabajar las operaciones usuales en las funciones de variable real, y por ello es posible hacer uso de la herramienta del análisis. Además, podemos extender la función a todos los números reales (y no sólo a los valores que toma la variable aleatoria), defininiendo $F(x)=0$ para x menor que el mínimo de la variable aleatoria y $F(x)=1$ para x mayor que el valor máximo. Intuitivamente esto indica que la probabilidad de que la variable aleatoria tome un valor menor que el mínimo es igual a cero y la probabilidad de que la variable aleatoria toma un valor inferior o igual al máximo es igual a uno.

6.3 Función de probabilidad de la variable aleatoria discreta

En el caso de la variable aleatoria discreta podemos calcular la probabilidad de que la variable tome un valor aislado. La función de probabilidad quedaría definida de una manera muy parecida a como se obtiene la función de distribución, pero haciendo uso sólo de la igualdad, puesto que en este caso se aplicaría la función inversa de la variable aleatoria sobre un solo valor numérico de la variable aleatoria. La inversa de la variable aleatoria en este caso arrojaría como resultado un solo evento en el espacio muestral.

$$p(x_i) = P(X = x_i) = P\left(\omega / X(\omega) = x_i\right)$$

Podemos, sin embargo, también obtenerla haciendo uso de la función de distribución. Bastaría para ello hallar la diferencia entre el valor de la función de distribución para dos valores consecutivos de la variable. Para cada valor x_i, podemos definir una nueva función, nuevamente de variable real, con imagen en el intervalo *[0, 1]* en la forma dada en la siguiente expresión.

$$p(x_i) = F(x_i) - F(x_{i-1})$$

El valor $p(x_i)$ no sería más que la probabilidad de que la variable tome el valor x_i. La definición formal de la función de probabilidad es:

> Sean $x_1, x_2, x_3, \ldots x_n$ los distintos valores que puede tomar la variable aleatoria discreta y $p(x_1), p(x_2), \ldots p(x_n)$ la probabilidad con que toma estos valores. El conjunto de pares de valores $(x_j, p(x_j))$ de una variable aleatoria discreta se denomina distribución de probabilidades de la variable aleatoria. La aplicación que a cada valor x en una variable aleatoria discreta le asigna el valor $p(x)$ es una función que algunos autores denominan *función de probabilidad* y otros, simplemente, *distribución de probabilidad* y debe cumplir con las siguientes propiedades:
>
> a. $0 \leq p(x_j) \leq 1$: $p(x)$ es una probabilidad, y por lo tanto debe tomar valores entre 0 y 1.
>
> b. $\sum_{j=1}^{n} p(x_j) = 1$: la suma de probabilidades repartidas entre todos los valores de la variable debe ser igual a 1.

De la misma forma que se calculan frecuencias acumuladas, podemos acumular probabilidades, obteniendo la *función de distribución de probabilidades*. La función de distribución para estas variables se obtiene en una forma alternativa mediante la expresión:

$$F(x) = \sum_{1}^{j} p(x_j)$$

6.4 Función de densidad de las variables aleatorias continuas

En el caso de variables aleatorias continuas, la probabilidad asociada a un evento elemental o simple es siempre igual a cero, por lo que no tiene interés considerar directamente la distribución de probabilidad (que en tal caso sería constante). Así, en lugar de trabajar con la probabilidad de valores particulares de la variable, resulta más apropiado trabajar directamente con la función de distribución (que continúa siendo una función de variable real para este caso) o bien calcular probabilidades asociadas a intervalos. Para esto último se usa una función que mide "concentración" de probabilidades alrededor de un punto, que se denomina *función de densidad de probabilidad* y se denota como *f(x).* Una función de densidad de probabilidad debe cumplir con las siguientes propiedades:

- $f(x) \geq 0$: la función es no negativa para cualquier valor de x, f(x) no es una probabilidad, y puede valer más de 1.
- $\int_{-\infty}^{\infty} f(x)dx = 1$: la acumulada para todos los valores de la variable suma 1, el área bajo la curva de la función vale 1.

La función de distribución para una variable aleatoria continua se calcula mediante la relación que liga con la función de densidad, que es la derivada de la función de distribución (para el caso de variables aleatorias continua), como se muestra en la expresión:

$$F(a) = P(X < a) = \int_{-\infty}^{a} f(x)dx$$

La probabilidad de que la variable esté dentro de un intervalo [a, b] en una variable aleatoria continua, también se calcula como diferencia de valores de la función de distribución en sus extremos, mediante la expresión:

$$P(a < X < b) = F(b) - F(a)$$

Como consecuencia, la probabilidad de que la variable tome un valor particular es nula, como se expresa en:

$$P(X = c) = P(c < X < c) = F(c) - F(c) = 0$$

Esto explica la idea de que para el caso de una variable aleatoria continua no tiene sentido trabajar con la probabilidad de un valor particular.

6.5 Promedios

A partir de estos los conceptos anteriores se desprenden otros relacionados con la variable aleatoria, tanto en probabilidad como en estadística, la mayoría de ellas más vinculadas con su función de distribución, aunque no desligados de su función de probabilidad. Entre ellos mencionaremos los momentos de la variable (medidas de posición central y dispersión), y los modelos de funciones que describen las familias de distribuciones de probabilidad (como la Binomial, Poisson, Uniforme, Normal, etc.) así como sus respectivos parámetros. Por ejemplo, para las variables aleatorias discretas podemos definir las medidas de posición central como sigue (Ortiz, 2002):

> Sea (x_i, p_i) donde $i \in I$, la distribución de probabilidad de una variable aleatoria discreta. Se define la *media o esperanza matemática* como $E[X]= \Sigma_{i\in I}\, x_i\, p_i$. Este concepto extiende la idea de media en una variable estadística.
>
> La *moda* es el valor más probable de la variable.
>
> La *mediana* es el valor de la variable para el cual la función de distribución toma el valor 1/2. Por tanto, la probabilidad de que una variable aleatoria tome un valor menor o igual a la mediana es exactamente 1/2.

6.6 La variable aleatoria como función

Recapitulando un poco, la variable aleatoria es una función numérica definida sobre un álgebra de sucesos, y en particular sobre cada elemento del conjunto Ω, tal que se puede hablar de la probabilidad de que esa función tenga un valor contenido en un intervalo determinado. La condición (1) que se explicita en su definición (apartado 6.1), para el caso de variables discretas, se puede traducir como $X^{-1}(x_i) \in A$ y significa que el conjunto de los elementos de Ω que por medio de X se les asigna un valor determinado x_i deberá pertenecer al álgebra de sucesos *A* y por lo tanto que su valor de probabilidad también estará (o podría ser) definido, en caso contrario X no sería variable aleatoria sobre *A* (Boudot, 1979). Nos auxiliaremos del esquema 13 para desglosar esta definición.

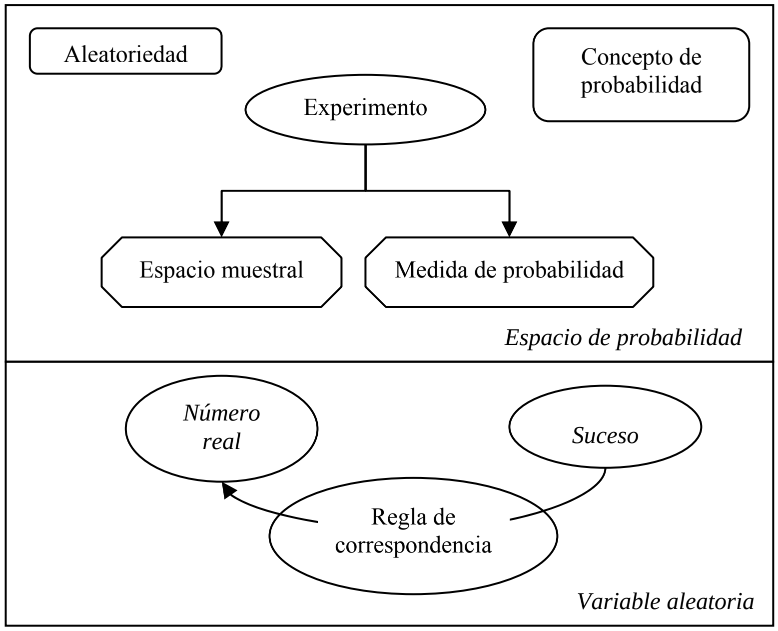


Figura 13. Interacciones asociadas a la definición de variable aleatoria

La variable aleatoria, como toda función matemática, tiene tres componentes: la *regla de correspondencia*, que es la "regla" que definirá la forma en que se vinculan el espacio muestral y al número real (perteneciente al conjunto imagen); el *valor numérico* (imagen), que es el número que toma la variable aleatoria y que será el mismo que ella, en su papel de regla de correspondencia, le asigna a cada evento del espacio muestral (dominio).

Hay que hacer notar que la regla de correspondencia que define la variable aleatoria es determinista. En cambio, su imagen, como valor numérico está cargada de aleatoriedad en la medida que es el número asociado al correspondiente suceso aleatorio en el espacio muestral (Meyer, 1970/1973). Por otro lado no admitimos cualquier posible función, sino que debe ser medible. Es decir, la definición de variable aleatoria no solamente vincula el espacio muestral con el conjunto de números reales y la regla de correspondencia sino que exige la estructura de espacio probabilístico. Al espacio muestral se añade, una asignación previa de probabilidades (y consecuente con los axiomas básicos), que ha sido asignada bajo una cierta noción de probabilidad.

El resultado final de todas estas interacciones son números reales (imagen de la variable aleatoria) relacionados con el espacio muestral, pero que además cumplen con todas las propiedades de campo del conjunto de los reales y que por lo tanto hacen factible establecer una relación funcional (la función de distribución) entre estos números. La variable aleatoria, cumple ahora el papel de *variable independiente* y la probabilidad asignada a cada uno de ellos, el de *variable dependiente* con la que es posible hacer uso de la herramienta del análisis.

6.7 Modelación y variable aleatoria

Krickeberg (1962/1973) sostiene que si estamos interesados en la variable aleatoria X, en realidad lo estamos en su función de distribución, puesto que ésta nos proporciona las probabilidades con las cuales el valor de X(ω) está comprendido dentro de los diferentes conjuntos de *A* en una observación aleatoria con resultado ω. Sin embargo otros autores (Meyer, 1970/1973, Feller, 1968/1973 y Hoel, Port y Stone 1971) sostienen que es en la variable aleatoria en la que radica la importancia de las distribuciones de probabilidad puesto que es ésta en la que se manifiesta el espacio muestral y el experimento y por lo tanto la «*realidad*». Lo cierto es que cuando se define la variable aleatoria queda automáticamente definida su función de distribución, así, el estudio de una implica el estudio de la otra y también que tanto en la definición de la variable aleatoria como en la de la función de distribución está implícito el concepto de modelación que vincula la realidad con una representación matemática.

Como regla de correspondencia y desde una perspectiva puramente matemática, la variable aleatoria podría ser arbitraria, siempre y cuando el espacio muestral definido satisfaga las condiciones establecidas en la definición (Mood y Graybill, 1952/1955), pero desde una perspectiva de modelación, la regla de correspondencia debe tener un sentido a partir de una pregunta que se pretenda resolver y por lo tanto, bajo ese contexto, no puede ser tan arbitraria: la variable aleatoria debe poder explicarse en el contexto del problema del que surge la pregunta.

Esto significa que desde una perspectiva exclusivamente teórico-matemática a un espacio muestral se le podrían asignar diferentes reglas de correspondencia (Godino, Batanero y Cañizares, 1996), pero éstas se ven limitadas cuando se condicionan a la pregunta de interés que nos mueve a ocuparnos del fenómeno aleatorio. Por tanto, una dificultad de la definición de la variable aleatoria radica en que esa pregunta, ubicada en el contexto del problema, debe poder ser planteada de tal manera que asigne un número a cada evento del espacio muestral pero no solamente como etiqueta, sino con las propiedades de los números reales, de otra forma no definiríamos una variable aleatoria.

Así por ejemplo, en un problema en el que nos interesa un fenómeno aleatorio en el que esté involucrado el sexo de las personas, la regla de correspondencia: «asignamos '0' a los hombres y '1' a las mujeres» no representaría propiamente una variable aleatoria, porque en tal caso los números 0 y 1 en el contexto del problema son considerados etiquetas, en cambio, la regla de correspondencia «número de mujeres» sí les asigna las propiedades de campo a los números, aun cuando desde el punto de vista práctico ambas reglas podrían proporcionar los mismos resultados. La regla de correspondencia también podría ser «número de hombres» y en tal caso modificaríamos la asignación primera porque la función de distribución sería diferente. La regla de correspondencia quedaría definida por el contexto del problema que nos hace ocuparnos de ese problema, aunque desde la perspectiva axiomática de la probabilidad, es indiferente cuál se escoja.

La importancia de la vinculación entre el contexto del problema y el matemático en la asignación de la regla de correspondencia también se puede ejemplificar con la tirada de dos dados. Supongamos que estamos interesados en los números que quedan en la cara superior de dos dados honestos cuando se tiran. Desde una perspectiva matemática daría lo mismo asignar la regla de correspondencia «la suma de los números de las caras superiores de los dados» o la regla «la resta de los números de las caras superiores de los dados» o incluso «la multiplicación de los números de las caras superiores de los dados» porque todos ellos asignarían números reales a la variable aleatoria, sin embargo la distribución de probabilidades resultante de cada una de ellas sería diferente. Aquí lo que determinaría cuál regla de correspondencia a seleccionar debe ser el problema que nos hace interesarnos en la tirada de esos dos dados porque el modelo debe poder interpretarse en el problema para que nos sea de utilidad.

Por lo tanto, desde la perspectiva de Heitele (1975) la modelación del proceso tendría dos estratos, una en la que se asigna un valor a cada evento del espacio muestral y otra en la que se le asigna un valor de probabilidad a cada valor de la variable aleatoria. Desde este punto de vista, la teoría de probabilidad apoyaría el modelo encontrado con las estudiantes y en el que los valores de la variable aleatoria juegan un doble papel, el de «realidad» en la función de probabilidad y el de «modelo matemático» en la asignación de un valor numérico a cada evento del espacio muestral.

6.8 Variable aleatoria y asignación de probabilidades

El problema de encontrar la regla de correspondencia en la variable aleatoria está vinculado con el problema de asignación de probabilidades, puesto que si bien es cierto que la teoría matemática nos permite formalizar y operar con las probabilidades, no nos soluciona el problema de cómo asignar las probabilidades. Desde un punto de vista matemático, la variable aleatoria nos permite definir un tipo de probabilidad a través de un sistema axiomático teórico, a la que Hawkins y Kapadia (1984) le denominan *probabilidad formal*, que proporciona el soporte del cálculo de probabilidades y que conecta la probabilidad con la corriente principal de la matemática moderna. Esta probabilidad formal también define la función de distribución en donde, de acuerdo con el análisis matemático, la probabilidad es la variable dependiente y la variable aleatoria es la variable independiente.

Sin embargo, esta definición formal no es tan simple porque, como señala Azcarate (1995), no mantiene relación con los fenómenos naturales, sino a través de otras concepciones de probabilidad. Godino, Batanero y Cañizares (1996, p 27) sugieren que "*la base matemática puede reflejar hipótesis hechas en las concepciones clásica, frecuencial o subjetiva*" y Fine (1973) hace un análisis de las perspectivas axiomáticas desde diferentes interpretaciones del significado de la probabilidad. Esto es, la definición de variable aleatoria, exige una interpretación previa de probabilidad y esto es precisamente lo que unirá a la probabilidad formal (o a la función de probabilidad) con los fenómenos aleatorios (con la *realidad*).

El problema de asignar la regla de correspondencia que vincule al espacio muestral con un número real se ataca de manera diferente dependiendo de la concepción de probabilidad de la persona que asigna la probabilidad inicial en el experimento aleatorio. La asignación de probabilidades por sí misma, involucra un problema epistemológico (Godino, Batanero y Cañizares, 1996; Batanero, Henry y Parzysz, 2005):

- En el caso de que la concepción de probabilidad en cuestión sea *clásica*, la probabilidad sería el cociente entre el número de casos favorables y posibles en la situación. Aparentemente esta concepción sería fácil de aplicar y es objetiva. Sin embargo el juicio sobre si los sucesos que intervienen en la situación son equiprobables es en cierto modo subjetivo y esta noción es difícilmente aplicable más allá del terreno de los juegos de azar. Además la definición de probabilidad es circular, porque para poder aplicarla hay que considerar la equiprobabilidad de los sucesos. Tampoco puede aplicarse al caso de variables aleatorias continuas, en el que la probabilidad de cada punto aislado es nula, nos parezcan o no todos los sucesos igualmente probables.

- Cuando la asignación de probabilidades sea *frecuencial* la probabilidad se estimaría a partir de la frecuencia de aparición de un valor de la variable en un número elevado de repeticiones del experimento (valores de la *variable estadística*). La variable aleatoria se concebiría entonces como un modelo teórico de la variable estadística, pero los valores de probabilidad que se obtienen son sólo estimaciones y nunca el valor exacto que es sólo teórico. Hay además la dificultad de decidir el número de experimentos necesarios para una buena aproximación y más aún de decidir si las condiciones en que se repiten los experimentos pueden considerarse como equivalentes.

- Finalmente, en la *concepción subjetiva*, la probabilidad es un grado de creencia que asigna un decisor a cada valor de la variable. No se requiere la repetición del experimento y puede aplicarse a situaciones únicas, pero inciertas. Sin embargo, el estatuto científico de una probabilidad asignada con criterios subjetivos ha sido objeto de debate.

Capítulo 7

Análisis epistemológico histórico

Un aspecto esencial para la comprensión didáctica del concepto de variable aleatoria es conocer su desarrollo histórico. La variable aleatoria, como muchos otros conceptos de la ciencia, ha surgido progresivamente a través de su historia y ha presentado etapas de mayor o menor desarrollo las cuales están marcadas por eventos que marcan algún progreso en su conceptualización como ente matemático.

En particular el concepto de variable aleatoria es difícil de seguir a lo largo del tiempo debido a la especificidad de su definición, a lo complejo de las herramientas matemáticas necesarias para formalizarlo, a la cercanía (en el tiempo) de su definición formal y a que la variable aleatoria está vinculada con problemas que la teoría de la probabilidad todavía no resuelve. En este análisis nos guiamos por las nociones intuitivas que percibamos en la historia de la probabilidad, por las propiedades distintivas de la formalización denotada en el capítulo anterior y por el uso de la nomenclatura actual.

Hacemos nuestra disertación de acuerdo a los estadísticos que pudimos dilucidar, intervinieron en la formulación del concepto. Algunos están agrupados exclusivamente por las fechas de aparición y no por que los caracterice alguna aportación en particular.

7.1 Primeros indicios

Hald (1984) menciona que uno de los primeros indicios del uso intuitivo de la variable aleatoria aparece en el planteamiento de un problema de distribución hipergeométrica

(Propositio IV) [11] en *De Ratiociniis in Ludo Aeae* de **Cristian Huygens** (1629-94) que aparece en el *Ars Conjectandi* de **Jacob Bernoulli** (1654-1705) en el capítulo primero:

> Proposición IV: Suponga que juego contra otra persona tres veces, y que ya he ganado dos juegos y él uno. Quiero saber qué proporción de la apuesta me correspondería, en caso de que decidiéramos no continuar el juego y dividiéramos la apuesta equitativamente entre nosotros[12].

En el problema ya está presente la variable como conteo de número de juegos ganados o perdidos.

En la proposición 14, Huygens plantea un problema donde se suman las caras superiores de dos dados:

> Proposición 14. Suponga que otro jugador y yo tomamos turnos para lanzar dos datos con una condición: yo gano si obtengo 7 puntos, y él gana si él obtiene 6 puntos, y yo le permito a él lanzar primero. Hallar la razón de mi oportunidad con respecto a la suya[13] (Hald, 1990, p. 70).

Donde puede observarse que Huygens trabaja con la suma de dos resultados numéricos del experimento aleatorio de lanzar dos dados como variable.

Sin embargo, es Jacob quien propiamente usa magnitudes aleatorias para resolver problemas involucrados con los ahora llamados *ensayos de Bernoulli* como el caso de la Distribución Binomial en su *Ars Conjectandi*:

> Para una serie de n ensayos con la misma probabilidad de éxito, p¸ dice [Bernoulli], la probabilidad de m éxitos y n-m fracasos en un orden dado es $p^m q^{n-m}$. Si el orden de éxitos y fracasos no importa, la probabilidad de m éxitos

[11] Lo comenta Moivre en su obra "Mensura Sortis" publicada en *Philosophical Transactions* (Numb.329, 1711). Citado en Hald (1984), p. 231.
[12] Proposition 4: "Suppose that I play against another person about three games, and that I have already won two games and he one. I want to know what my proportion of the stakes should be, in case we decide not to continue the play and divide the stakes equitably between us" (Hald, 1990, p. 70)
[13] Proposition 14. "Suppose that I and another player take turns in throwing with two dice on the condition, that I win if I throw 7 points, and he wins if he throws 6 points, and I let him have the first throw. To find the ratio of my chance to his" ((Hald, 1990, p. 70).

y *n-m* fracasos es $\binom{n}{m} p^m q^{n-m}$, puesto que hay $\binom{n}{m}$ diferentes ordenamientos de *m* éxitos y *n-m* fracasos. (Hald, 1990, p. 227)

Bernoulli usa el resultado anterior en su generalización de algunos problemas planteados por Huygens[14]. Pero también identifica una característica importante de sus magnitudes aleatorias: que tienen que ser *independientes*. De hecho, él parte del supuesto de independencia en sus ensayos. También trabaja con varios tipos de magnitudes aleatorias: Hay una magnitud aleatoria implícita que asocia 1 ó 0 a cada éxito o fracaso de un ensayo. Está también la magnitud aleatoria correspondiente a la suma de éxitos en *n* ensayos y otra más correspondiente a la proporción muestral obtenida del número de éxitos y el tamaño *n* de la muestra.

Hacking I. (1975/1995) sostiene al respecto que Bernoulli en su exposición del capítulo 4 de su *Ars conjectandi* usa de la variable aleatoria: "El resultado de *n* ensayos, s_n, es una variable aleatoria. Para cualquier estimador F, la estimación $F(S_n)$ es, entonces, también una variable aleatoria" (p. 192). Lo cierto es que J. Bernoulli consigue identificar de sus variables la condición de independencia, pero no consigue profundizar más en su naturaleza ni en la especificación de algunas de sus operaciones.

Abraham De Moivre (1667-1754)[15] publicó en 1730 su *Miscellanea Analytica* que contiene el primer tratamiento de la probabilidad integral y la esencia de la curva Normal. Pearson (1924) dice que De Moivre encuentra una ecuación general para la curva: $y = y_o e^{\frac{-x^2}{2\sigma^2}}$ (él usaba *l* en lugar de *x*). Halló el valor de $y_o = \frac{1}{\sigma\sqrt{2\pi}}$ y halla que su integral es una buena aproximación para el cálculo de probabilidades de la distribución binomial cuando *n* es grande. Con esto De Moivre consigue una relación importante entre magnitudes aleatorias *discretas* y *continuas*.

[14] Proposiciones 12 a 14 *De Ratiociniis in Ludo Aeae* de Huygens.
[15] Citado en Pearson, K. (1924), p. 403.

7.2 Las aportaciones de Laplace

Pierre Simon Laplace (1749–1827) aplicó con gran éxito el cálculo de probabilidades en problemas de Astronomía. Particularmente en problemas de las medidas astronómicas y el error que se comete en ellas. Se le atribuye, al igual que a De Moivre, el planteamiento de la Distribución Normal. Laplace llega a una mayor generalización que De Moivre sobre el planeamiento el Teorema Central del Límite para variables discretas y para algunos casos de continuas, pero no lo consigue para una función arbitraria. Aparte de darle un mayor uso y generalización a las magnitudes aleatorias con el planteamiento de distribuciones muestrales, identifica condiciones como independencia e *idénticamente distribuidas*. También opera con magnitudes aleatorias continuas. En el siguiente ejemplo, citado en su obra *Teoría Analítica de las probabilidades* (1812), se puede ver el nivel de formalización de la variable aleatoria:

> Sean *i* cantidades variables y positivas $t, t_1, t_2, \ldots, t_{i-1}$, de las cuáles su suma es *s*, y de las cuales la ley de posibilidad es conocida; se propone encontrar la suma de los productos de cada valor que pueda recibir una función dada $\Phi(t, t_1, t_2, \ldots)$ de esas variables, multiplicadas por la probabilidad correspondiente su valor (Laplace, 1812/1886, p. 266 §15)[16]

En el texto anterior podemos identificar el uso del término *variable* para referirse a la variable aleatoria *t* así como a su característica *numérica* y hace referencia a su dominio, que hoy entendemos como los números reales positivos. Por otra parte, podemos ver que Laplace identificaba claramente las *i* variables aleatorias $t, t_1, t_2, \ldots, t_{i-1}$ así como otra variable *s* correspondiente a su suma. Respecto a su *ley de posibilidad* la interpretamos como la posibilidad de asignación de probabilidades a cada valor de la variable *t*. En sus *Mémoire sur les probabilités* de 1778 utiliza mucho esta expresión, como, a modo de ejemplo, puede verse en la siguiente cita:

> Ahora, si conociéramos el límite y la ley de posibilidades de los valores de α, nada sería más fácil que resolver exactamente este problema; porque si

[16] Texto original : « Soient i quantités variables et positives $t, t_1, t_2, \ldots, t_{i-1}$, dont la somme soit s, et dont la loi de possibilité soit connue ; on propose de trouver la somme des produits de chaque valeur que peut recevoir une fonction donnée $\Phi(t, t_1, t_2, \ldots)$ de ces variables, multipliée par la probabilité correspondante à cette valeur ».

> nombramos *q* a este límite y si representamos por $\psi(\alpha)$ como la probabilidad de α, vemos primero que α tiene necesariamente que caer entre 0 y *q*, la función $\psi(\alpha)$ debe ser tal que tenemos: $\int d\psi(\alpha) = 1$. (Laplace, 1778/1893, p. 394)[17]

Finalmente, podemos observar operaciones sobre las variables y sus probabilidades. Todos son elementos esenciales de la composición de las variables aleatorias entendidas desde la perspectiva moderna. Sin embargo, en este estadío de la comprensión del concepto de variable aleatoria, es vista como una *magnitud aleatoria*, es decir, puede observarse en la obra de Laplace que cuando se refiere a variables, lo hace respecto a resultados numéricos de experimentos aleatorios. No se vislumbra, en esta etapa de desarrollo de la variable aleatoria, la comprensión de ésta como una función conjunto de valor real.

Cabe hacer notar que Laplace distingue entre la función de probabilidad ($\psi(\alpha)$) y la posibilidad de asignación de probabilidades (ley de posibilidades). De alguna manera en este último concepto cree necesario asegurarse de que es posible asignarle una probabilidad a las variables.

Por otra parte, la aleatoriedad no estaba aún bien consolidada. Como cuando Laplace concibe una probabilidad más bien como un mal menor, fruto de nuestra ignorancia:

> Todos los acontecimientos, incluso aquellos que por su insignificancia parecen no atenerse a las grandes leyes de la naturaleza, no son sino una secuencia tan necesaria como las revoluciones del Sol. Al ignorar los lazos que los unen al sistema total del universo, se los ha hecho depender de causas finales o del azar, según que ocurrieran o se sucedieran con regularidad o sin orden aparente, pero estas causas imaginarias han ido siendo descartadas a medida que se han ido ampliando las fronteras de nuestro conocimiento, y desaparecen por completo ante la sana filosofía que no ve en ellas más que la expresión de nuestra ignorancia de las verdaderas causas (Laplace, 1814/1995, p. 24).

---

[17] Texto original: "Cependant, si l'on connaissait la limite et la loi de possibilité des valeurs de α, rien ne serait plus facile que de résoudre exactement ce problème; car, si l'on nomme q cette limite et que l'on représente par $\psi(\alpha)$ la probabilité de α, on voit d'abord que, α devant nécessairement tomber entre 0 y *q*, la fonction $\psi(\alpha)$ doit être telle que l'on ait: $\int d\psi(\alpha) = 1$"(Laplace, 1778/1893, p. 394)

7.3 Las aportaciones de Poisson

Simon D. Poisson (1781-1840) en su *Recherches sur la probabilite des jugements en matiere criminelle et en matiere civile* (1937) pareciera, en un momento dado, que hace uso de la variable aleatoria, pero no es muy claro:

> Así, llamaremos $C_1, C_2, C_3, ..., C_n, ..., C_m$, las m causas posibles del envenenamiento E; siendo $p_1, p_2, p_3, ..., p_n, ..., p_m$, las probabilidades conocidas de sus correspondientes entradas, relativas a esta variedad de posibles causas; de manera que p_n expresa la probabilidad de que E tendría lugar si la causa C_n fuera única, o, lo que es la misma cosa, si ella fuera cierta, excluiría a todas las demás . (Poisson, 1837, p. 82)[18]

Poisson hace la analogía de las m causas posibles con el mismo número de urnas $A_1, A_2, A_3, ..., A_n, ..., A_m$. Cada una con una probabilidad de sacar bola blanca de $p_1, p_2, p_3, ..., p_n, ..., p_m$. Luego, las expresa con un común denominador μ. En sus palabras:

> Por eso, supongamos que reducimos las fracciones $p_1, p_2, p_3,$ etc., a un mismo denominador, tendríamos:
>
> $$p_1 = \frac{\alpha_1}{\mu},\ p_2 = \frac{\alpha_2}{\mu},\ p_3 = \frac{\alpha_3}{\mu}, ...\ p_n = \frac{\alpha_n}{\mu}, ... p_m = \frac{\alpha_m}{\mu};$$
>
> μ y los numeradores $\alpha_1, \alpha_2, \alpha_3,$ etc., son unos números enteros. No alteraríamos nada si tomamos un azar una bola blanca de la una A_n, reemplazando las bolas que contiene, para un número α_n de bolas blancas de un

[18] « Ainsi, appelons $C_1, C_2, C_3, ..., C_n, ..., C_m$, les m causes possibles de l'événement E ; soient $p_1, p_2, p_3, ..., p_n, ..., p_m$, les probabilités connues de son arrivée, relatives à ces diverses causes ; de manière que p_n exprime la probabilité de E qui aurait lieu si la cause C_n était unique, ou, ce qui est la même chose, si elle était certaine, ce qui exclurait toutes les autres ». (Poisson, 1837, p. 82)

> número μ de bolas, tanto blancas como negras; y lo mismo para las otras urnas. (Poisson, 1837, p. 82-83)[19]

Se puede observar una variable cualitativa: las causas posibles C_n (eventos) con su probabilidad conocida, pero también Poisson trabaja con una variable numérica α_n asociada, por una parte al evento C_n, y por una probabilidad p_n. α pudiera interpretarse como una variable aleatoria discreta, pero en Poisson no se lee la intención de plantear un nuevo concepto ni de desarrollarlo. El caso es singular porque, a diferencia de Laplace que estaba interesado en problemas de magnitudes aleatorias, Poisson parte de una variable cualitativa –las posibles causas de envenenamiento- a la cual se le asoció una variable cuantitativa α con probabilidades específicas conocidas. Sin embargo, el concepto de variable aleatoria no tuvo más desarrollo en Poisson.

7.4 Las aportaciones de Chebyshef y sus discipulos

Pafnuty L. Chebyshef (1821-1894). En 1867 publica su obra *Des valeurs moyennes* –sobre valores esperados- donde pueden verse teoremas como el siguiente:

> Teorema. Si se designa por *a,b,c,…* las esperanzas matemáticas de las cantidades *x, y, z, …,* y por $a_1, b_1, c_1, \ldots$ las esperanzas matemáticas de sus cuadrados $x^2, y^2, z^2, \ldots,$ la probabilidad de que la suma $x + y + z + \ldots$ esté dentro de los límites

$$a+b+c....+\alpha\sqrt{a_1+b_1+c_1+\ldots-a^2-b^2-c^2-....},$$
$$a+b+c....-\alpha\sqrt{a_1+b_1+c_1+\ldots-a^2-b^2-c^2-....},$$

---

19 «Pour cela, supposons que l'on réduise les fractions $p_1,\ p_2,\ p_3,$ etc., au même dénominateur, et que l'on ait ensuite : $p_1 = \frac{\alpha_1}{\mu},\ p_2 = \frac{\alpha_2}{\mu},\ p_3 = \frac{\alpha_3}{\mu}, \ldots p_n = \frac{\alpha_n}{\mu}, \ldots p_m = \frac{\alpha_m}{\mu};$

μ et les numérateurs $\alpha_1, \alpha_2, \alpha_3$, etc., étant des nombres entiers. On ne changera rien à la chance de tirer une boule blanche de l'urne A_n, en y remplaçant les boules qu'elle contient, para un nombre α_n de boules blanches et un nombre μ de boules, tant planches que noires; et de même pour toutes les autres urnes ». (Poisson, 1837, p. 82-83)

> será todavía más grande que $1-\frac{1}{\alpha^2}$, cualquiera que sea α. (Tchebychef, 1867/1899, p. 687)[20]

Kolmogorov comenta que Chebyshev fue el primero en estimar claramente y hacer uso de la noción de la variable aleatoria y su valor esperado[21]. Y, en efecto, puede observarse en el anterior teorema no sólo el uso claro de variables aleatorias, sino una de sus características más importantes que es su valor esperado. También opera sobre ellas al buscar las esperanzas matemáticas de sus cuadrados.

Andrei Markov (1856-1922) fue el más cercano discípulo de Chebyshev e hizo aportaciones importantes a la teoría de la probabilidad, entre otros, por su teorema límite para la suma de variables aleatorias independientes, así como de variables dependientes, en particular aquellas conectadas en cadena.

En su artículo «Sobre las raíces de la ecuación $e^{x^2}\frac{\partial^m e^{-x^2}}{\partial x^m}=0$» publicada en 1898, Markov prueba un teorema límite con la siguiente formulación:

> La probabilidad de que la suma $u_1+u_2+...+u_n$ de variables independientes $u_1, u_2, ..., u_n$ está dentro de los límites $\alpha[2(a_1+a_2+...+a_n)]^{1/2}$ y $\beta[2(a_1+a_2+...+a_n)]^{1/2}$. Donde $a_1, a_2, ..., a_n$ son las esperanzas matemáticas de las variables $u_1^2, u_2^2, ..., u_n^2$ y α y β son dos cantidades elegidas arbitrariamente, se aproxima cuando *n* tiende a infinito al límite a
>
> $$\pi^{-1/2}\int_\alpha^\beta \exp(-x^2)dx,$$

[20] El texto original citado dice : « Théorème. Si l'on désigne par *a,b,c,…* les espérances mathématiques des quantités *x, y, z, …*, et par $a_1, b_1, c_1, \ldots$ les espérances mathématiques de leurs carrés $x^2, y^2, z^2, \ldots$, la probabilité que la somme $x+y+z+\ldots$ est renfermée entre les limites

$$a+b+c....+\alpha\sqrt{a_1+b_1+c_1+...-a^2-b^2-c^2-....},$$
$$a+b+c....-\alpha\sqrt{a_1+b_1+c_1+...-a^2-b^2-c^2-....},$$

sera toujours plus grande que $1-\frac{1}{\alpha^2}$, quel que soit α. »

[21] Citado en Maistrov L. E. (1974), p. 207.

> a condición de que la secuencia infinita de variables independientes $u_1, u_2, ..., u_n$ satisfagan las siguientes condiciones:
>
> 1. Las esperanzas matemáticas de $u_1, u_2, ..., u_n$ son cero;
>
> 2. Las esperanzas matemáticas de $u_n^2, u_n^2, ..., u_n^2$ son finitas para finitos valores de *k* cuando *k* se incrementa indefinidamente;
>
> 3. La esperanza matemática de u_k^2 no llega a ser arbitrariamente pequeña cuando *k* se incremente indefinidamente. (Maistrov, 1974, p. 210)

Cómo puede observarse, Markov enuncia las variables aleatorias que utiliza y se vale de varias de sus propiedades como independencia y su valor esperado, operatividad de la suma y potencia, así como de convergencia de series de variables aleatorias.

Sin embargo, aún no las enuncia en sus teoremas como «variables aleatorias».

Aleksandr M. Lyapunov (1857-1918), también discípulo de Chebyshev, da continuidad a los trabajos de su maestro y de Markov en teoría de probabilidad. Él es el primero en usar la expresión «variable aleatoria» en su discurso y teoremas:

> Chebyshev ha mostrado en una de sus memorias que los resultados de sus investigaciones sobre los valores límite de integrales que conduzcan a la prueba del tan conocido Teorema de Laplace y de Poisson relativo a la probabilidad de la suma de grandes números de variables aleatorias independientes está contenida dentro de determinados límites. (Lyapunov, 1948, p. 220. Citado por Hazewinkel, 2002)

Una de sus principales contribuciones en este campo es el enunciado y demostración del Teorema Central de Límite. El enunciado de Lyapunov es el siguiente:

> Supóngase que las variables aleatorias $X_1, X_2, ...,$ tienen medias finitas EX_k, varianzas DX_k y momentos absolutos $E|X_k - EX_k|^{2+\delta}, \delta > 0$, y supóngase también que $B_n = \sum_{k=1}^{n} DX_k$ es la varianza de la suma de $X_1, ..., X_n$. Entonces

si para algún $\delta > 0$, $\lim_{n\to\infty} \frac{\sum_{k=1}^{n} E\left|X_k - EX_k\right|^{2+\delta}}{B_n^{1+\delta/2}} = 0$ [conocida como Condición de Lyapunov], la probabilidad de la desigualdad $x_1 < \frac{\sum_{k=1}^{n}(X_k - EX_k)}{\sqrt{B_n}} < x_2$ tiende al límite $\frac{1}{\sqrt{2\pi}}\int_{x_1}^{x_2} e^{-x^2/2}dx$ cuando $n \to \infty$, uniformemente con respecto a todos los valores de x_1 y x_2.(Lyapunov, 1954. Citado por Hazewinkel, 2002)

Teorema enunciado y probado por Lyapunov en 1901 fue el último paso de la investigación de Chebyshev y Markov.

Es notable el conocimiento que Lyapunov tenía de variables aleatorias: no sólo las distinguía claramente y las enunciaba sino que usaba sus caracterizaciones más importantes como son su valor esperando y sus varianzas, así como su operatividad, convergencia y composición.

7.5 Kolmogorov

Los nuevos resultados en probabilidad pronto tomaron gran importancia ante el auge de aplicaciones en ciencias como la Física con los trabajos de L. Boltzmann (1844-1906), entre otros. Esto generó la necesidad de una mayor precisión conceptual y lógica y, por otra parte, la emergencia de la teoría de conjuntos y de la medida hicieron nacer una nueva era para la probabilidad: la axiomatización. Pioneros sobresalientes fueron E. Borel (1871-1956) y S. Bernstein (1880-1968). Kolmogorov puntualiza al respecto:

> La primera aximatización sistemáticamente desarrollada de la teoría de probabilidad, basada en la noción de comparación cualitativa de eventos (aleatorios) de acuerdo a sus (más grandes o más pequeñas) probabilidades es debida a S. N. Bernstein. El valor numérico de la probabilidad aparece en esta

> concepción como una derivada más que como una noción primaria. (Maistrov, 1974, p. 272)[22]

Andrei Kolmogorov (1903-1987) consolidó una axiomatización de la teoría de la probabilidad con la incorporación de la teoría de la medida. En su obra *Fundations of the Theory of Probability* publicada por primera vez en 1933, dedica todo un capítulo, el tercero, a variables aleatorias. Sin embargo, desde el capítulo I da una definición inicial de variable aleatoria:

> Sea $\mathfrak{A}$ una descomposición del conjunto fundamental E: $E = A_1 + A_2 + \ldots + A_n$, y *x* una función real de los eventos elementales ξ, los cuales para cada conjunto A_q es igual a una correspondiente constante a_q. *x* es también llamada *variable aleatoria*, y la suma $E(x) = \sum_q a_q P(A_q)$ es llamada *esperanza matemática* de la variable *x*. La teoría de las variables aleatorias será desarrollada en los capítulos III y IV. No nos limitaremos allí simplemente a aquellas variables aleatorias que pueden asumir solamente un número finito de valores diferentes». (Kolmogorov, 1933/1956, p. 12)[23]

Su aporte aquí es que queda claramente establecida la naturaleza funcional de la variable aleatoria como una función conjunto de valor real.

En el capítulo III, Kolmogorov da la siguiente definición de variable aleatoria:

> Definición. Una función de valor real $x(\xi)$, definida sobre el conjunto básico E, es llamada una *variable aleatoria* si para cada elección de un número real *a* el

[22] Kolmogorov, A. N. (1947). The role of Russian science in the development of probability theory, *Ucen. Zap. Moskov. Gos. Inst.* 91. citado en Maistrov L. E. (1974), p. 252.

[23] Let $\mathfrak{A}$ be a decomposition of the fundamental set E: $E = A_1 + A_2 + \ldots + A_n$, and *x* a real function of the elementary event ξ, which for every set A_q is equal to a corresponding constant a_q. *x* is then called a *random variable,* and the sum $E(x) = \sum_q a_q P(A_q)$ is called the *mathematical expectation* of the variable *x*. The theory of random variables will be developed in Chaps. III and IV. We shall not limit ourselves there merely to those random variables which can assume only a finite number of different values. (Kolmogorov, 1933/1956, p. 12)

> conjunto $\{x < a\}$ de todos los ξ para los cuales la desigualdad $x < a$ es verdadera, pertenece al sistema de conjuntos $\Im$ ».[24]

Esta definición se sostiene en lo que Kolmogorov llamó Campo de probabilidad y sus axiomas:

> Sea E una colección elementos $\xi, \eta, \zeta, \ldots$, los cuales llamaremos *eventos elementales,* y $\Im$ un conjunto de subconjuntos de *E*; los elementos de el conjunto $\Im$ serán llamados *eventos aleatorios..*
>
> I. $\Im$ *es un campo de conjuntos.*
>
> II. $\Im$ *contiene el conjunto E.*
>
> III. Para cada conjunto A en $\Im$ es asignado un número real no-negativo P(A). Este número P(A) es llamado la probabilidad del evento A.
>
> IV. P(E) es igual a 1.
>
> V. Si A y B no tienen elementos en común, entonces $P(A+B) = P(A) + P(B)$.
>
> Un sistema de conjuntos, $\Im$, junto con una asignación definida de número P(A), que satisfacen los Axiomas I-V, es llamado un *campo de probabilidad*. (Kolmogorov, 1933/1956, p. 2)[25]

---

[24] El texto original citado dice: "DEFINITION. A real single-valued function $x(\xi)$, defined on the basic set E, is called a *random variable* if for each choice of a real number *a* the set $\{x < a\}$ of all ξ for which the inequality $x < a$ holds true, belongs to the system of sets $\Im$". (Kolmogorov, 1933/1956, p. 22)

[25] El texto original citado dice: «Let E be a collection of elements $\xi, \eta, \zeta, \ldots$, which we shall call *elementary events,* and $\Im$ a set of subsets of E; the elements of the set $\Im$ will be called *random events.*

I. $\Im$ *is a field of sets.*

II. $\Im$ *contains the set E.*

III. *To each set A in* $\Im$ *is assigned a non-negative real number P(A). This number P(A) is called the probability of the event A.*

IV. *P(E) equal 1.*

V. *If A an B have no element in common, then* $P(A+B) = P(A) + P(B)$.

A system of sets, $\Im$, together with a definite assignment of numbers P(A), satisfying Axioms I-V, is called a *field of probability*». (Kolmogorov, 1933/1956, p. 2)

De modo que ahora la variable aleatoria queda circunscrita en un *campo de probabilidad*. Kolmogorov desarrolla también en el capítulo III algunos tópicos sobre variables aleatorias como variables aleatorias equivalentes y tipos de convergencia.

7.6 Levy, Petrov y Parzen

Paul Lévy (1871-1956). El uso de la nueva terminología se extiende por Europa. En 1936, tres años después de Kolmogorov, Lévy publica el artículo *Sur quelques points de la théorie des probabilités dénombrables* donde utiliza el término «variable aleatoria» en enunciados y explicaciones. Al final del artículo hay una mención a Kolmogorov, por lo que se supondría que es influencia de éste último. Paul Lévy da también un importante impulso a la teoría de la variable aleatoria, particularmente sobre suma de variables aleatorias: *L'addition des variables aléatoires définies sur une circonférence* de 1939. Sobre la convergencia de series de variables aleatorias: *Sur la sommabilité des séries aléatoires divergentes*, en 1935. Y sobre la multiplicación de variables aleatorias: *Esquisse d'une théorie de la multiplication des variables aléatoires* en 1959.

V. Vladímirovich Petrov (1931-) [26], cuya disertación doctoral fue *Some Extremal Problems in the Theory of Summation of Independent Random Variables.* Destacan también sus trabajos *Sums of independent random variables* (1972), *Limit Theorems for Sums of independent random variables,* (1987) y *Limit Theorems of Probability Theory* (1995). Él define, en su libro *Teoría de probabilidades* (2002), la variable aleatoria como sigue:

> Consideremos un espacio de probabilidades (Ω, A, P). Llamamos *variable aleatoria* a una función real $X = X(\omega)$ definida en el espacio de sucesos elementales Ω y tal que satisface la condición
>
> $$\{\omega \in \Omega : X(\omega) \leq x\} \in A \qquad (3.1)$$
>
> para todo *x* real.

[26] Valentin Vladímirovich Petrov nació en Rusia en 1931. Fue profesor de la Universidad de San Petersburgo. Doctor en Ciencias (Instituto Steklov, 1962).

> En la terminología del análisis real, una función $X(\omega)$ que cumple la condición (3.1) para todo *x* se denomina *medible.* De esta forma, una variable aleatoria es una función real y medible de los sucesos elementales. Se puede verificar que la condición (3.1) para todo *x*, es equivalente a la condición
>
> $$\{\omega \in \Omega : X(\omega) \in B\} \in A \qquad (3.2)$$
>
> para cualquier conjunto boreliano[27] ***B*** de puntos de la recta real $\Re$. En el caso particular en que ***B*** es el intervalo $(-\infty, x]$, la condición (3.2) se convierte en la condición (3.1) .(Petrov, 2002, p. 55)

Emanuel Parzen (1929-)[28], destacado miembro de la Asociación Americana de Estadística, autor y coautor de más de 100 artículos de investigación, ha sobresalido, entre otros, por sus artículos *A Central Limit Theorem for Multilinear Stochastic Process* (Parzen,1957a) y *An Approach to Time Series Análisis* (Parzen, 1957b). Destaca también, de sus 6 libros, la obra *Modern Probability Theory And Its Applications*, publicada por primera vez en 1960, y que ha tenido un amplio reconocimiento, sobre todo en el contexto de ingeniería. (Newton, 2002) En esta última obra, él hace distinción entre *Fenómenos aleatorios con resultados numéricos* (Capítulo 4) y su teoría: *Media y varianza de una ley de probabilidades* (Capítulo 5), *Leyes de probabilidades Normal y de Poisson y otras relacionadas con ellas* (Capítulo 6). Y, por otra parte, desarrolla la teoría de la variable aleatoria en los capítulos 7 a 10 *Variables aleatorias, Esperanza de una variable aleatoria, sumas de variables aleatorias independientes* y *Sucesiones de variables aleatorias.*

Parzen formaliza su concepción de un *fenómeno aleatorio con resultados numéricos:*

> *Un fenómeno aleatorio con resultados numéricos* es un fenómeno aleatorio cuyo espacio de descripciones muestrales es el conjunto *R* (de los los números reales de $-\infty$ a ∞) en cuyos subconjuntos está definida una función $P[\cdot]$, que

[27] Nota del autor. La clase de los conjuntos boreleanos en la recta es la mínima σ-álgebra de conjuntos que contiene a todos los intervalos.

[28] Emanuel Parzen nació en la ciudad de New York en 1929. Se doctoró en la Universidad de California, Berkeley en 1951 en Matemáticas y Estadística. Ha sido autor y coautor de más de 100 artículos y 6 libros. Él ha trabajado sobre la Teoría de detección de señales y en Análisis de series de tiempo donde ha sido pionero. En 1994, él fue galardonado con Medalla conmemorativa S. S. Wilks por la Asociación Americana de Estadística.

> asigna a cada conjunto boreliano de números reales E (que también se llama evento) un número real no negativo, denotado por $P[E]$, de acuerdo con los axiomas siguientes:
>
> Axioma 1. $P[E] \geq 0$ para todo evento E.
> Axioma 2. $P[R] = 1$.
> Axioma 3. Para toda sucesión de eventos $E_1, E_2, ..., E_2, ...$, que sea mutuamente exclusiva, tenemos que $P\left[\bigcup_{n=1}^{\infty} E_n\right] = \sum_{n=1}^{\infty} P[E_n]$. (Parzen, 1960/1971, p. 173).

En esta lógica, la variable aleatoria es un cierto «valor observado», pero lo que cuenta es el uso de *x* (variable determinística) en el desarrollo de la teoría, como podemos ver en una definición de función de distribución que Parzen da:

> La función de distribución $F(\cdot)$ de un fenómeno aleatorio con resultados numéricos se define, para cualquier número real *x*, como la probabilidad de que un valor observado del fenómeno aleatorio sea menor o igual que el número *x*. En símbolos, para cualquier número real *x*.
>
> (3.1) $F(x) = P[\{\text{ números reales } x' : x' \leq x \}]$ ((Parzen, 1960/1971, p. 191)

Luego, Parzen extiende el modelo a *fenómeno aleatorio con resultados numéricos* en *n* dimensiones cuyo espacio muestral es el conjunto $\Re^n$ que consiste en todas las *n*-adas $(x_1, x_2, ..., x_n)$ cuyas componentes $x_1, x_2, ..., x_n$ son números reales de $-\infty$ a ∞.

En el capítulo 7, Parzen abre camino a la variable aleatoria diciendo:

> Sin embargo, este método [*fenómeno aleatorio con resultados numéricos*] no es muy satisfactorio, pues que requiere que se fije de antemano el número *n* de fenómenos aleatorios que vamos a considerar en un contexto dado. Además, no ofrece ningún medio conveniente de generar, por medio de varias operaciones algebraicas y analíticas, nuevos fenómenos aleatorios a partir de fenómenos aleatorios conocidos. Estas dificultades las evitamos mediante variables aleatorias." (Parzen, 1960/1971, p. 299)

Finalmente, Parzen define la variable aleatoria de la siguiente manera:

Definición de variable aleatoria. Decimos que un objeto X es una variable aleatoria i) si es una función de valores reales definida en un espacio de descripciones muestrales, sobre una familia de cuyos subconjuntos hayamos definido una función de probabilidades $P[\cdot]$, y ii) si para todo conjunto boreliano B de números reales, el conjunto $\{s : X(s) \text{ pertenece a } B\}$ pertenece al dominio de $P[\cdot]$" (Parzen, 1960/1971, p. 300)

Capítulo 8

Conclusiones del análisis epistemológico

Las conclusiones sobre el análisis epistemológico se agrupan en las dos vertientes analizadas, desde la disciplina y desde la historia.

8.1 Desde la disciplina

Aparentemente la definición de variable aleatoria es sencilla, como función que asigna a cada suceso en un experimento aleatorio un valor numérico. La sencillez es, sin embargo aparente, en cuanto el concepto se apoya en la comprensión de los de aleatoriedad, experimento aleatorio, eventos y operaciones con eventos y probabilidad.

La variable aleatoria está vinculada con un proceso de modelación matemática, en el sentido de que es la que sirve para relacionar un fenómeno aleatorio («realidad») con el contexto matemático. Esta modelación presenta dos estratos: (1) uno en la que la variable aleatoria actúa como «modelo matemático» (en de la definición de variable aleatoria como función) y dentro del cual el problema que nos ocupa es la «realidad» y (2) otro en la función de distribución donde la variable aleatoria actúa como la «realidad», vinculando a la probabilidad con el fenómeno aleatorio.

Asociados a la variable aleatoria e inseparables de ella encontramos los conceptos de distribución de probabilidad y función de distribución. El requisito de que la imagen inversa de un intervalo acotado en la variable aleatoria sea un suceso medible nos lleva a la función de distribución que determina unívocamente a la variable aleatoria y viceversa. El proceso es complejo, pero también permite que la probabilidad de la función sea una probabilidad ampliada al intervalo [0,1] y que los valores de la variable aleatoria puedan extenderse al infinito, lo que la hace infinitamente aditiva.

En el análisis de la variable aleatoria aparecen diversas aplicaciones del álgebra de sucesos:

- en el conjunto de números reales (la variable aleatoria),
- del conjunto de sucesos en el intervalo [0,1] (la probabilidad asignada en el experimento)
- del conjunto de números reales en el intervalo [0,1] (la función de distribución y la distribución de probabilidad).

Esta última puede también definirse como composición de las anteriores (invirtiendo la variable aleatoria). Es por ello que el alumno debe comprender los conceptos de función y aplicación, función inversa, elementos de una función (dominio, rango, correspondencia), diferencia entre variable dependiente e independiente en una función.

Añadiendo las diversas representaciones verbales gráficas y simbólicas, vemos que el concepto variable aleatoria tiene una gran complejidad semiótica y epistémica.

8.2 Desde la historia

Pueden identificarse dos paradigmas en torno a variable aleatoria en su desarrollo histórico: el *Paradigma de magnitudes aleatorias* (o como le llama Parzen, *Fenómeno aleatorio de resultados numéricos* –Laplace, Poisson, Parzen son importantes representantes-) y el *Paradigma de variables aleatorias* (Kolmogorov, Petrov, entre otros). Ambos vigentes, e incluso en convivencia, en las distintas prácticas ingenieriles y de ciencias.

Identificamos 9 momentos históricos del desarrollo de la variable aleatoria:

1. Se asocian valores numéricos discretos a resultados de experimentos aleatorios, como el es caso de varios problemas planteados por Huygens. A esos resultados numéricos les llamaremos magnitudes aleatorias.
2. Se identifican probabilidades asociadas a las magnitudes aleatorias discretas y sus distribuciones de probabilidad, como los trabajos de J. Bernoulli.

3. Comienzan las primeras operaciones entre magnitudes aleatorias como uso de la suma, el cociente y la convergencia de sucesiones de variables aleatorias, aunque no siempre explícitas. Aparece la condición de independencia entre magnitudes aleatorias (Ley de los grandes números en Ars Conjectandi de J. Bernoulli).

4. Se vinculan las magnitudes aleatorias con problemas reales como errores de medición o problemas de sobrevivencia (De Moivre, Laplace y Galton). Nace la magnitud aleatoria continua, así como su probabilidad identificada como la integral en un intervalo dado. Identificación de distribuciones de probabilidad para magnitudes aleatorias continúas. Particularmente la Normal y la Exponencial. Auge del uso del cálculo infinitesimal en Probabilidad. Primeros enunciados del Teorema central del límite. Era común considerar la Probabilidad como un mal menor ante nuestra incapacidad de conocer con certeza (De Moivre y Laplace entre otros).

5. Aparecen los primeros indicios explícitos de asociación de resultados no-numéricos de experimentos aleatorios con ciertas magnitudes con asignación de probabilidades (Poisson).

6. Se trabaja con muchas propiedades de las magnitudes aleatorias y rigor matemático (Chebyshef, Markov) e incluso se les llama "variables aleatorias" (Lyapunov).

7. Se mira las magnitudes aleatorias como un caso particular de las variables aleatorias. Éstas últimas son definidas como una función conjunto de valor real dentro de un espacio de probabilidad y que toma valores de un espacio de medible (Borel, Bernstein y Kolmogorov)

8. Se multiplican las aportaciones a la Teoría de las variables aleatorias (Levy, Feller, Meyer, Petrov,....)

9. Hay un nuevo equilibro entre paradigmas (Parzen). Los paradigmas de magnitudes aleatorias y el de variables aleatorias subsisten hasta la actualidad.

En estos nueve momentos identificamos cuatro etapas que marcan saltos cualitativos en la conceptualización de la variable aleatoria:

1. Desde la probabilidad en un contexto de teoría de juegos hasta problemas de aplicaciones como sobrevivencia, o teoría de errores (Huygens –De Moivre y Laplace)

2. Desde resultados numéricos de un fenómeno aleatorio hasta el reconocimiento semiótico de la variable aleatoria y operaciones básicas (Chebyshev a Liapunov)

3. Desde estadío anterior al reconocimiento funcional de la variable aleatoria y su formulación formal en base a teoría de la medida (Borel y Kolmogorov).

4. Desarrollo de la teoría formal de la variable aleatoria: sumas, productos, convergencia de sucesiones y series de variables aleatorias (Lévy, Petrov).

Así mismo, identificamos algunos obstáculos para el surgimiento y desarrollo del concepto de variable aleatoria:

- El predominio del *Paradigma magnitudes aleatorias* ha dificultado el surgimiento de un paradigma nuevo basado en variables aleatorias.

- Dificultad en la *naturaleza funcional de la variable aleatoria* y la composición de funciones vinculada con ella y la probabilidad. Tradicionalmente ha sido más tratada como *variable* (magnitud aleatoria) que como función.

- Dificultad para reconocer la *existencia de la variable aleatoria al hacer composición entre variables aleatorias.* (por ejemplo, Bernoulli en la Ley de los grandes números).

- Las dificultades planteadas por la Física cuántica y otras ciencias, como el caso de las funciones discontinuas, las paradojas de teoría de conjuntos, la matemática formal y los problemas planteados por Hilbert, pusieron en crisis la teoría de la probabilidad. Lo que dio paso a una nueva era con las aportaciones de la Escuela de San Petersburgo, particularmente de Kolmogorov.

- *Complejidad con sus operaciones.* Se ha tenido que desarrollar la teoría de conjuntos y de la medida para plantear formalmente a la variable aleatoria y abordar más profundamente sus operaciones como la suma entre variables aleatorias, producto y convergencia de sucesiones y series de variables aleatorias, teoremas límite. (Lévy, Petrov).

Así mismo, logramos identificar algunos obstáculos para el surgimiento y desarrollo del concepto de variable aleatoria:

1. El predominio del Paradigma magnitudes aleatorias ha dificultado el surgimiento de un paradigma nuevo basado en variables aleatorias.

2. Dificultad en la *naturaleza funcional de la variable aleatoria* y la composición de funciones vinculada con ella y la probabilidad. Tradicionalmente ha sido más tratada como *variable* (magnitud aleatoria) que como función

3. Dificultad para reconocer la *existencia de la variable aleatoria al hacer composición entre variables aleatorias.* (por ejemplo, Bernoulli en la Ley de los grandes números)

4. Las dificultades planteadas por la Física cuántica y otras ciencias, como el caso de las funciones discontinuas, las paradojas de teoría de conjuntos, la matemática formal y los problemas planteados por Hilbert, pusieron en crisis la teoría de la probabilidad. Lo que dio paso a una nueva era con las aportaciones de la Escuela de San Petersburgo, particularmente de Kolmogorov.

5. *Complejidad con sus operaciones.* Se ha tenido que desarrollar la teoría de conjuntos y de la medida para plantear formalmente a la variable aleatoria y abordar más profundamente sus operaciones como la suma entre variables aleatorias, producto y convergencia de sucesiones y series de variables aleatorias, teoremas límite. (Lévy, Petrov).

Capítulo 9

Algunas ideas para investigaciones futuras

9.1 <u>En el análisis cognitivo</u>

Sobre la posibilidad de que la variable aleatoria pueda ser aprendida escolarmente.

Este trabajo sugiere que es posible diseñar una enseñanza efectiva de algunas nociones intuitivas sobre variable aleatoria dirigida a alumnos universitarios que no tengan conocimientos previos en estadística. Esto se pone de manifiesto en la cantidad de conceptos movilizados por las alumnas a lo largo de la situación durante la entrevista clínica, así como en el progreso que se observa entre sus concepciones y estrategias iniciales y cómo la actividad les ayuda a superar algunas ideas incorrectas o incompletas. Las dificultades que se presentaron en estos alumnos es previsible que aparezcan en otros. Particularmente es de interés estudiar qué regularidades se presentan en estudiantes con la misma preparación, en situación escolar.

Consideraciones para nuevos diseños

Por otro lado, el estudio proporciona criterios para revisar la situación, perfilando mejor las preguntas, proporcionando a los alumnos ayuda específica en los puntos en que ha habido dificultad y diseñando otras situaciones referidas a otros campos de problemas relacionados con la variable aleatoria, por ejemplo el de ajuste de una función a una distribución de datos estadísticos. La situación ha estado muy centrada en una variable aleatoria discreta y algunos puntos de nuestro análisis conceptual y de la tarea deberán ser revisados para el caso de variables aleatorias continuas. Pareciera interesante diseñar una segunda situación en que la variable aleatoria a modelar fuera continua y analizar las diferencias en dificultades y estrategias respecto a las descritas en este estudio. Así por ejemplo, nos parece de sumo interés indagar sobre las diferencias encontradas con otras situaciones en donde la probabilidad estuviera definida en otro paradigma.

La modelación y la variable aleatoria

También es necesario profundizar en la idea de la variable aleatoria como vinculación entre la realidad del problema y su modelación a través de la función de probabilidad. Consideramos que un problema más abierto, planteado con la posibilidad de permitir que el estudiante proponga diversas variables aleatorias y él deba decidir cuál sería la mejor variable que ayudaría a resolver el problema propuesto, aportaría conocimiento sobre las posibilidades de construcción del concepto de variable aleatoria, pero también aportaría vertientes importantes sobre la forma en cómo se modela a través de la variable aleatoria. Así mismo se vincularía más el modelo matemático con el contexto del problema, ayudando a superar el obstáculo de la reducción de la aleatoriedad a matemática determinística.

La componente semiótica: un área de oportunidad

Asimismo, en este trabajo no se ha aplicado toda la potencia del análisis semiótico, que nos permitiría identificar la naturaleza específica de los conflictos semióticos que causan las dificultades percibidas por los alumnos. Algunos de estos conflictos semióticos quedan implícitamente descritos (por ejemplo confusión entre variable dependiente o independiente en las diversas funciones involucradas o no interpretación de la probabilidad como un número sino como una razón), sin embargo es necesario un estudio más detallado de estos conflictos y de la forma en que podrían ser resueltos.

9.2 Vinculación entre el análisis cognitivo y el epistemológico

Algunos de los resultados obtenidos en el análisis epistemológico apoyan y sustentan algunos de los resultados obtenidos en el análisis cognitivo, lo que sirve como guía para dirigir investigaciones cognitivas tendientes a concretar obstáculos epistemológicos. Pero también para profundizar históricamente en las etapas del desarrollo del concepto a partir de los cuestionamientos provenientes del análisis cognitivo.

Así por ejemplo, es deseable indagar qué tanto coinciden las etapas históricas que se detecten en un análisis epistemológico con las etapas formuladas por los estudiantes para construir el concepto. En nuestro análisis encontramos una coincidencia en la formulación de experimentos aleatorios con resultados numéricos.

El análisis epistemológico apoyó las conclusiones obtenidas en el análisis cognitivo sobre cómo los conocimientos de matemática determinística, particularmente el álgebra y cálculo, se vuelven un obstáculo para el surgimiento y uso de ideas matemáticas probabilísticas. El trabajo de Laplace es representativo en esto. Él, que hizo muchas aportaciones en el cálculo, retomó a la probabilidad como una alternativa en su búsqueda determinística del conocimiento del mundo. También puede observarse que el desarrollo matemático de la teoría de probabilidades fue establecido fundamentalmente con el recurso de la matemática determinística de su época. El paradigma de la magnitudes aleatorias es, de hecho, una consecuencia de la influencia del cálculo, pues en este paradigma se busca lo más posible expresar los resultados y procedimientos en términos de *x*, no de la variable, aleatoria, X que toma el valor *x*.

Beta epistemológica

Otra coincidencia importante está en que en el desarrollo de la teoría de la variable aleatoria se presentaron dificultades como la expresión de la función de probabilidad con la igualdad: $p(x_i) = P(X = x_i)$, para variables aleatorias discretas, que es una expresión de convivencia entre los dos paradigmas, pero que, finalmente, expresan visiones distintas de la probabilidad. Las alumnas en el análisis cognitivo no distinguieron ambas probabilidades y por lo tanto no veían la necesidad de desprenderse de la interpretación clásica de probabilidad. Las relaciones y diferencias entre la probabilidad vinculada a la función de variable aleatoria en $\boldsymbol{R}$ y la probabilidad vinculada con espacios muestrales finitos es algo sobre lo que se tiene que profundizar cognitiva y epistemológicamente.

Desde la perspectiva semiótica, el término mismo de *variable aleatoria*, tiene sus dificultades, como dice Feller:

> El término variable aleatoria es un poco confuso; función aleatoria sería más apropiado (pues la variable independiente es un punto en un espacio muestral, es decir, constituye el resultado de un experimento) (Feller, 1968/1973, p 221)

9.3 Algunas otras consideraciones a la investigación

1. Observar en los libros de texto universitarios al definir y tratar la variable aleatoria la presencia y posible convivencia de los paradigmas de magnitud aleatoria y variable aleatoria.

2. Valorar la conveniencia didáctica de introducir la variable aleatoria a partir del desarrollo primero del paradigma de magnitud aleatoria.

3. Se esperan dificultades en el proceso de aprendizaje del concepto de variable aleatoria:

   - provenientes de la naturaleza funcional de la variable aleatoria.
   - por la simbología utilizada.
   - por la interferencia de la matemática determinística en la emergencia del concepto de variable aleatoria.
   - por sus dificultades asociadas a la comprensión de funciones conjunto de valor real.
   - por la complejidad del formalismo actual en torno a la variable aleatoria.

Glosario

Aleatoriedad: La aleatoriedad se puede interpretar desde dos perspectivas, la informal y la forma. La informal está relacionada con la incapacidad de predecir con exactitud. Se le relaciona con el azar en el sentido de se le atribuye ser «la causa» de la aleatoriedad y es fuente de innumerables controversias. Desde la perspectiva formal se analiza en su proceso de generación (matemáticamente, el experimento aleatorio) y en el patrón de las secuencias que se generan al repetir muchas veces el proceso. Se le concibe como un modelo matemático de una sucesión de resultados de un mismo experimento realizado repetida e independientemente. Bajo esta visión, es posible aplicar el cálculo de probabilidades a los fenómenos denominados aleatorios (Batanero y Serrano, 1995 y Batanero, 2001).

Distribución de probabilidad: Es la colección de pares ordenados $(x_i, p(x_i))$ con $i = 1, 2,....$ definidos de la siguiente manera: «Dada una variable aleatoria discreta X, que toma los valores $x_1, x_2, x_3,...$, a lo más infinitos numerables, a cada resultado posible x_i se le asocia un número $p(x_i) = P(X = x_i)$ llamado probabilidad de x_i, que cumple con que $p(x_i) \geq 0$ para todo i y que $\sum_{i=1}^{\infty} p_i = 1$». Meyer (1970/1973) establece que en la mayoría de los casos, los números $p(x_i)$ se determinarán por «la función de probabilidades asociada con sucesos en el espacio muestral sobre el que se define X. Esto es $p(x_i) = P[\omega / X(\omega) = x_i]$» (p.61).

Concepciones: De acuerdo con Peltier (1993), las concepciones de los estudiantes serán la clase de problemas que dan sentido a un concepto para el estudiante; el conjunto de significantes que es capaz de asociar (imagen mental, expresión simbólica) y los instrumentos, teoremas, algoritmos, que es capaz de poner en marcha.

Epistemología: Desde la teoría de los griegos, la *episteme* constituye en la teoría del conocimiento, el auténtico conocimiento. Desde la tradición francesa, epistemología se refiere a la filosofía de las ciencias y se limita a un solo conocimiento: el científico. Temas propios de la epistemología son, entonces, la organización de las ciencias, su unidad, su división, sus principios, sus métodos, etc. En particular, a la epistemología que nos referimos en esta memoria es a la de las matemáticas (en particular a la de la probabilidad y más específicamente a la del concepto de variable aleatoria). Su campo también lo constituye el estudio de las ideas que de de algún modo son previas, y eran independientes de la noción de estudio, pero que al traspasar ciertos límites inciden en el surgimiento, la organización (o reorganización) y desarrollo del campo de estudio. Entran en el interior de las nociones formándolas y constituyéndolas. Los estudios epistemológicos no sólo dan cuenta de la organización, coherencia y validez pasada y actual de las ciencias, sino también de la forma en cómo los conceptos propios de una epistemología trascienden y se aplican en otras disciplinas, aunque este último punto no lo abordamos en esta memoria.

Error: La idea de error es una idea específica de la epistemología. Desde la concepción de esta investigación, entenderemos por erróneo aquel proceso operatorio que trastorna contenidos de conocimiento. Lo diferenciamos de la ignorancia, que es un no-conocimiento, en que el error se ejerce sobre un material ya dado, es decir, conlleva conocimiento, pero en su manipulación conceptual lo trastorna y lo ensambla incorrectamente, de modo que no ajustan entre sí. El error aparece, por tanto, como algo inherente al sistema de conocimientos (Valverde, J. en Muñoz, J. y Valverde, J., 2000).

Espacio muestral: Es el conjunto de todos los resultados posibles de un experimento aleatorio ε. El espacio muestral puede ser finito, infinito numerable o infinito innumerable. Generalmente el espacio muestral se representa con la letra Ω.

Evento (o suceso): Un evento (respecto a un espacio Ω en particular, asociado a un experimento ε definido) es un conjunto de resultados posibles de un experimento aleatorio. En terminología de conjuntos, diremos que es un subconjunto del espacio muestral. Los sucesos pueden ser elementales o compuestos cuando se pueden descomponer en sucesos más simples. Se les representa con letras mayúsculas del alfabeto: A, B, C,... En ocasiones cuando se habla de sucesos elementales se emplea la letra minúscula ω.

Experimento aleatorio: Nos referiremos a un experimento aleatorio o no determinístico como aquel que es posible repetir indefinidamente sin cambiar esencialmente sus condiciones, pero del cual no podemos predecir cuál será su resultado en particular. Generalmente es posible describir el conjunto de los resultados posibles del experimento. Una característica importante que menciona Meyer (1970/1973) es que si un experimento aleatorio se repite, los resultados aparentan un comportamiento caprichoso, sin embargo, a medida que se incrementa el número de veces que se repite, aparece un patrón regular. Esta regularidad es la que hace posible la construcción de un modelo matemático que permite analizar el experimento. Generalmente el espacio muestral se representa con la letra ε.

Función de distribución (o función de distribución acumulada): Es la función definida por como $F(x) = P(X \leq x)$ para una variable aleatoria *X*, continúa o discreta. Caracteriza totalmente a la variable aleatoria y debe cumplir: (1) ser una función definida en $(-\infty, +\infty)$ y monótona no decreciente; (2) su límite cuando *x* tiende a infinito negativo es cero y su límite cuando *x* tiende a infinito positivo es 1; (3) es continua por la derecha.

Función de probabilidad: En algunos libros distinguen la función de probabilidad de la distribución de probabilidad. A la función de probabilidad la definen como la función, *p*, de una variable aleatoria discreta, *X*, en el mismo contexto que la distribución de probabilidad: «Dada una variable aleatoria discreta *X*, que toma los valores x_1, x_2, x_3,..., a lo más infinitos numerables, a cada resultado posible x_i se le asocia un número $p(x_i) = P(X = x_i)$ llamado probabilidad de x_i, que cumple con que

$p(x_i) \geq 0$ para todo *i* y que $\sum_{i=1}^{\infty} p_i = 1$». La diferencia es muy sutil y algunas veces la función de probabilidad la reducen a la expresión algebraica de la distribución de probabilidad. En esta memoria las tomaremos como sinónimos.

Probabilidad: Desde la perspectiva formal, la probabilidad se define a través de los tres axiomas de la probabilidad: (1) la probabilidad del suceso A está entre 0 y 1; (2) la probabilidad del evento seguro es uno; (3) La probabilidad de la unión de dos eventos excluyentes, tienen como intersección el conjunto vacío, es igual la suma de las probabilidades de los dos eventos. Sin embargo, el vocablo es mucho más complejo porque estos tres axiomas no incluyen realmente la asignación de probabilidades. A este respecto, muchos autores (entre ellos Fine, 1973; Boudot, 1979, Cabriá, 1992; Székely, 1982/1986 y Borovcnik, Benz y Kapadia, 1991) amplían el concepto y mencionan la existencia de diversos tipos de probabilidad. Los más consensuados son la (1) *probabilidad clásica*: se calcula de acuerdo a la regla de Laplace y que se define como la proporción del número de casos favorables entre el número de casos posibles; (2) *la probabilidad lógica*: es un grado de creencia racional, se calcula como una relación lógica entre una aseveración y otra que presenta una evidencia; (3) *la probabilidad frecuencial o empírica*: se calcula a partir de las frecuencias relativas en un experimento aleatorio que se repite muchas veces; y (4) *la probabilidad subjetiva*: que es obtenida como una expresión del grado de creencia de una persona. Cada uno de estos tipos de probabilidad, incluyendo la formal, tiene sus propias dificultades, epistémicas, didácticas y cognitivas, así como tiene su particular dificultad la interacción y convivencia entre diferentes formas de interpretar la probabilidad. Estas dificultades han sido estudiadas y reportadas desde la didáctica de la probabilidad por un número muy amplio de autores se han dedicado a investigarlas. Un buen compendio de algunos de estos resultados está en Shaughnessy, 1992, en Batanero, 2001 y 2005 y en Batanero, Henry y Parzysz 2005. A la probabilidad de un evento A se le representa como P(A).

Probabilidad como función: Con este término denominamos a la función que se establece entre los eventos y la probabilidad. Por consiguiente es una función en donde se relacionan conjuntos (los eventos) con un valor acotado entre 0 y 1 (la probabilidad). Esta terminología no está estipulada en la teoría de probabilidad, sin embargo hemos tenido la necesidad de introducirla para diferenciar la asignación de probabilidades a los eventos (*probabilidad como función*), de la asociación de probabilidad a la variable aleatoria (*función de probabilidad*).

Variable aleatoria: Es la función (*X*) que asigna a cada uno de los elementos $\omega \in \Omega$, un número real $X(\omega)$ en un experimento aleatorio ε asociado a un espacio muestral Ω. La variable aleatoria puede ser *discreta* (si toma una cantidad, finita o infinita, numerable de valores) o *continua* (si puede tomar una cantidad innumerable de valores). A los valores (números reales) que toma la variable aleatoria generalmente se les representa por letras minúsculas x_i.

σ-Álgebra: Dado un conjunto *Z,* una familia *A* de subconjuntos de *Z* forman una σ-álgebra si satisface las condiciones siguientes: (1) Los conjuntos vacío y *Z* pertenecen a *A*; (2) *A* es cerrada para le paso a complemento, es decir que el complemento de cualquier conjunto que pertenezca a *A*, también pertenecerá a *A*; (3) *A* es cerrada con respecto a uniones numerables, es decir, la unión de cualquier sucesión de conjuntos que pertenezcan a *A*, también pertenecerá a *A*. Se puede deducir que, dadas estas características, *A* también es cerrada para las intersecciones numerables y diferencias de conjuntos. Este concepto asegura que a cualquier evento generado en un espacio muestral dentro de una estructura σ-álgebra sea factible asignarle una probabilidad puesto que la estructura en una σ-álgebra de eventos (o sucesos) es cerrada para la unión e intersección numerable.

σ-Álgebra de Borel: Es la familia más pequeña de conjuntos con las propiedades de una σ-álgebra que existe en la recta real. La clase de subconjuntos de Borel está engendrada por uniones e intersecciones numerables de intervalos semiabiertos. A todo evento definido como conjunto boreliano y generado en un fenómeno aleatorio con resultados numéricos es posible asignarle probabilidades (Parsen, 1991).

Referencias Bibliográficas

Alatorre, S. (1998). Acerca del tratamiento didáctico de la probabilidad [Versión electrónica]. *Correo del Maestro 26*, 23-35. Obtenido el 20 de septiembre de 2003 en http://www.correodelmaestro.mx.

Alquicira, Z. M. I. (1998) *Probabilidad: Docencia y Praxis. Hacia una fundamentación epistemológica*. Tesis de maestría no publicada. CINVESTAV, México.

Artigue, M. (1995a). Ingeniería Didáctica. En P. Gómez (Ed.), *Ingeniería didáctica en educación matemática* (pp. 39-59). México: Iberoamericana.

Artigue, M. (1995b). Une recherche d'ingénierie didactique sur l'enseignement des mathématiques, Trabajo presentado en la Reunión *Recherches en didactique des mathématiques et de l'informatique.* Grenoble: IMAG-LSD y en el *XXVIII ème Rencontre de la CIEAEM,* Louvain la Neuve, 1989.

Azcarate, P. (1995). *El conocimiento profesional de los profesores sobre las nociones de aleatoriedad y probabilidad. Su estudio en el caso de la educación primaria*. Tesis doctoral no publicada. Universidad de Cádiz, España.

Batanero, C. (2001). *Didáctica de la estadística*. Granada: Grupo de investigación en educación estadística, Universidad de Granada.

Batanero, C. (2002). *Los retos de la cultura estadística*. Conferencia Inaugural de Jornadas Interamericanas de Enseñanza de la Estadística. Buenos Aires, Argentina. Obtenido el 20 de junio de 2004 en http://www.ugr.es/~batanero/ARTICULOS/CULTURA.pdf

Batanero, C. (2005). Significados de la probabilidad en la educación secundaria. *Relime*, 8 (3), 247-264.

Batanero, C. y Serrano, L. (1995). La aleatoriedad, sus significados e implicaciones educativas. *Uno*, 5, 15-28.

Batanero, C. y Serrano, L. (1999). The meaning of randomness for secondary students. *Journal for Research in Mathematics Education*, 30, 558-567.

Batanero, C., Henry, M. y Parzysz, B. (2005). The nature of chance and probability. En G. Jones (Ed.). *Exploring probability in school: challenges for teaching and learning* (pp. 16-42). Dordrecht: Kluwer.

Batanero, C., Tauber, L. y Sánchez V. (2001). Significado y comprensión de la distribución normal en un curso introductoria de análisis de datos. *Quadrante,* 10 (1), 59-92.

Boudot, M. (1979). *Lógica inductiva y probabilidad*. Madrid: Paraninfo.

Borovcnik, M., Bentz, H. J. and Kapadia, R. (1991). A probabilistic perspective. En R. Kapadia and M. Borovcnik (Eds), *Chance Encounters: Probability Education* (pp. 27-71). Netherlands: Kluwer Academic Publishers.

Brousseau, G. (1981). *Los obstáculos epistemológicos y los problemas en matemáticas* (Trad. Departamento de Matemática Educativa). México: Cinvestav. (Original en francés, 1976).

Brousseau, G. (1986). Fondements et méthodes de la didactique des mathématiques. *Recherches en Didactique des Mathématiques*¸ 7 (2), 33-115.

Brousseau, G. (1990). Le contrat didactique: le milieu. *Recherches en didactique des mathématiques,* 9 (3), 309-336.

Brousseau, G. (1997). *Theory of didactical situations in mathematics. Didactique des mathématiques, 1970-1990*. Dordrecht: Kluwer Academic Publishers.

Bruner, J. S. (1960). *The process of Education*. Massachusetts: Cambridge.

Bunge, M. (1973). *La investigación científica*. Barcelona: Ariel.

Cabriá, S. (1992). *Filosofía de la probabilidad* Valencia, España: Tirant lo Blanc.

Cañizares, M. J., Batanero, C., Serrano L., and Ortiz J. J. (2003) Children's understanding of fair games in CERME 3.

Castillo, J. y Gómez, J. (1998). *Estadística inferencial básica*. México: Grupo Editorial Iberoamérica.

Chaput, B., Dantal, B., Girard, J. C., Grange, J. P., Henry, M., Janvier, M., Parzicsz, B., Pichard, J. F., Raumondaud, H., Thienard, J. C. y Vendrely, M. (1997). *Enseigner les probabilités au Licée. Ouvertures statistiques, enjeux epistemologiques, questions didactiques et idées d'activités*. Reims: Commission Inter-Irem.

Chevallard, Y. (1991). *La transposición didáctica* (Trad. C. Gilman). Argentina: Aique. (Original en francés, 1985).

Chevallard, Y. (1992). Concepts fondamentaux de la didactique: perspectives apportées par une apache anthropologique. *Recherches en Didactique des Mathématiques*, 12 (1), 73-112.

Chevallard, Y., Bosch, M. y Gascón, J. (1997). *Estudiar matemáticas. El eslabón perdido entre la enseñanza y el aprendizaje*. Barcelona: Horsori.

Clement, J. (2000). Analysis of Clinical Interviews: Foundations and Model Viability. En A. E. Kelly and R. A. Lesh (Eds), *Handbook of Research Design in Mathematics and Science Education* (pp. 547-589). London: Lawrence Erlbaum Associates, Publishers.

Confrey, J. (1980). Clinical interview: Its potential to reveal insights in mathematical education. En R. Karplus (ed.) *Proceedings of the 4th International Conference for the Psychology of Mathematics Education* (400-408). Berkeley, C.A.

Confrey, J. (1995a). A theory of intellectual development (Part 1). *For the Learning of Mathematics*, 14 (3), 2-8.

Confrey, J. (1995b). A theory of intellectual development (Part 2). *For the Learning of Mathematics*, 15 (1), 38-48.

Confrey, J. (1995c). A theory of intellectual development (Part 3). *For the Learning of Mathematics*, 15 (2), 36-35.

Confrey, J., Perry, L. (1980). Students' mathematical abilities: A focus for the improvement of teaching general mathematics. *School science and mathematics,* 80 (7), 549-556.

Cox, D. (1997). The current position of statistics: A personal view. *International Statistical Review,* 65 (3), 261-290.

Cuadras, C. (1999). *Problemas de Probabilidades y Estadística.* (Vol. 1). Barcelona: EUB.

Douady, R. (1986). Jeux de cadres et dialectique outil-objet. *Recherches en Didactique de Mathématique*, 7(2), 5-31.

Falk, R. y Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review,* 104, 310-318.

Feller, W. (1973). *Introducción a la teoría de probabilidades y sus aplicaciones* (Vol. 1) (Trad. S. Morales). México: Limusa. (Original en inglés, 1968).

Fine, T. L. (1973). *Theories of probability. An examination of foundation*. New York: Academic Press.

Gascón, J. (1998). Evolution de la didactique des mathématiques comme discipline scientifique. *Recherches en Didactique des Mathématiques*, 18/1(52), 7-34.

Gimeno, J. (1984). *Teoría de la enseñanza y desarrollo del currículo*. Madrid: Paraninfo.

Godino, J. D. (1996). Mathematical concepts, their meaning and understanding. En L. Puig y A. Gutierrez (Eds.), *Proceedings of XX Conference of the International Group for the Psychology of Mathematics Education,* (Vol. 2, pp. 417-425). Valencia, España.

Godino, J. D. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathématiques*, *22*(2/3), 237-238.

Godino, J. D. (2003). *Teoría de funciones semióticas: Un enfoque ontológico-semiótico de la cognición y la instrucción matemática*. Granada: Facultad de Ciencias de la Educación. Obtenido el 20 de enero de 2004 de: http://www.ugr.es/~jgodino/indice_eos.htm.

Godino, J. D. y Batanero, C. (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactique des Mathématiques,* 14(3), 325-355.

Godino, J. D. y Batanero, C. (1998a). Building and experimenting a model for a meaningful instrucion on data analysis. En L. Pereira et al. (Eds.), *Proceedings of the V International Conference on Teaching Statistics*, (Vol. 12, pp. 905-912). Voorburg: Internacional Statistical Institute.

Godino, J. D. y Batanero, C. (1998b). Clarifying the meaning of mathematical objects as a priority area of research in mathematics education. En A. Sierpinska y J. Kilpatrick

(Eds.), *Mathematics Education as a research domain: A search for identity* (pp. 177-195). Dordrecht: Kluwer.

Godino, J. D., Batanero, C. y Cañizares, M. J. (1996). *Azar y probabilidad*. Madrid: Síntesis.

Hacking I. (1995). *El surgimiento de la probabilidad. Un estudio filosófico de las ideas tempranas acerca de la probabilidad, la inducción y la inferencia estadística.* (Trad. J. Álvarez). Barcelona: Gedisa. (Original en ingles, 1975).

Hald A. (1984). A. De Moivre: 'De Mensura Sortis' or 'On the Measurement of Chance', *International Statistical Review*, 52 (3), 229-262.

Hald, A. (1990). *A History of Probability and Statistics and Their Applications before 1750*. New York: John Wiley & Sons.

Hawkins, A. y Kapadia (1984). Children's conceptions of probability: A psychological and pedagogical review. *Educational Studies in Mathematics* 15, 349-377.

Hawkins, A., Jolliffe, F. y Glickman, L. (1992). *Teaching statistical concepts.* Essex: Longman.

Hazewinkel, M. (Ed.). (2002), *Encyclopaedia of Mathematics* [en línea]. Springer. Obtenido el 18 de junio de 2006 de: http://eom.springer.de/L/l061200.htm.

Heitele, D. (1975). An epistemological view of fundamental stochastic ideas. *Educational Studies in Mathematics* 6, 187-205.

Hoel, P., Port, S., y Stone, Ch. (1971). *Introduction to probability theory*. Boston: Houghton Mifflin Company.

Johnson, R. y Kubv, P. (1999). *Estadística elemental: Lo esencial* (2ª Ed). (Trad. H. Villagómez). Ciudad de México: Thompson editores. (Original en inglés, 1998).

Kolmogorov A. (1956). *Foundations of the Theory of Probability* (2ª Ed). (Trad. N. Morrison). New York: Chelsea Publishing Company. (Original en alemán, 1933).

Krickeberg, K. (1973). *Teoría de la Probabilidad* (Trad. B. G. Verlagsgesellschaft). Barcelona: Teide. (Original en Alemán, 1962).

Laplace P. S. (1893). Mémoire sur les probabilités. En L'Académie des sciences (Ed.), *Oeuvres complètes de Laplace* (Vol. 9). Paris: Gauthier-Villars. (Original publicado en 1778).

Laplace P. S. (1886). Théorie Analytique des Probabilités. En L'Académie des sciences (Ed.), *Oeuvres complètes de Laplace* (Vol. 7). Paris: Gauthier-Villars. (Original publicado en 1812).

Laplace, P. S. (1995), *Ensayo filosófico sobre las posibilidades* (Trad. P. Castillo). Barcelona: Ediciones Altaya. (Original en francés, 1814).

Lévy, P. (1935). Sur la sommabilité des séries aléatoires divergentes. *Bulletin de la S. M. F., 63,* 1-35. Obtenido el 20 de marzo de 2006 de: http://www.numdam.org/item?id=BSMF_1935_63__1_0.

Lévy, P. (1936). Sur quelques points de la théorie des probabilités dénombrables. *Annales de l'I. H. P.*, 6 (2), 153-184. Obtenido el 19 de marzo de 2006 de: http://www.numdam.org/numdam-bin/item?id=AIHP_1936__6_2_153_0

Lévy, P. (1939). L'addition des variables aléatoires définies sur une circonférence. *Bulletin de la S. M. F.* , *67,* 1-41. Obtenido el 21 de marzo de 2006 de: URL: http://www.numdam.org/item?id=BSMF_1939__67__1_0.

Lévy, P. (1959). Esquisse d'une théorie de la multiplication des variables aléatoires. *Annales scientifiques de l' É. N. S., 3 série,* 76 (1), 59-82. Obtenido el 22 de mayo de 2006 de: http://www.numdam.org/item?id=ASENS_1959_3_76_1_59_0

Maistrov L. E. (1974). *Probability Theory. A historical sketch* (Trad. S. Kotz). New York : Academic Press. (Original en ruso, 1967).

Meyer, P. (1973). *Probabilidad y aplicaciones estadísticas* (Trad. C. Prado). México: Fondo Educativo Iberoamericano. (Original en inglés, 1970).

Miller, T.K. (1998). The random variable concept in introductory statistics. In L. Pereira-Mendoza (Ed.). *Proceedings of the Fifth International Conference on Teaching Statistics*. (pp. 1221-1222). Singapore: IASE.

Mood, A. y Graybill, F. (1955). *Introducción a la teoría de la estadística* (2ª Ed) (Trad. F. Azorín). Madrid: Aguilar. (Original en inglés, 1952).

Muñoz, J. y Valverde, J. (2000). Compendio de Epistemología. Madrid: Trotta.

Nardecchia, G. y Hevia, H. (2003). Dificultades en la enseñanza del concepto de variable aleatoria. Trabajo presentado en el *V Simposio de Educación Matemática*. Chivilcoy, Argentina.

Nemirovsky, R. (1993). *Symbolizing motion, flow, and contours: the experience of continuous change*. Tesis doctoral (Ed. D.). USA: Harvard University.

Newton J. H. (2002). A conversation with Emanuel Parzen. *Statistical Science*, 17 (3), 357-378.

Ortiz, J. J. (2002) *La probabilidad en los libros de texto.* Tesis doctoral (Ed. D.). España: Grupo de Educación Estadística de la Universidad de Granada.

Oseguera, F. (1994). *El concepto de variable aleatoria en el contexto del currículo. Análisis y Alternativas*. Tesis de Maestría no publicada. México: Cinvestav, México.

Parzen, E. (1957a). A Central Limit Theorem for Multilinear Stochastic Process. *The Annals of Mathematical Statistics*, 28 (1), 252-256.

Parzen, E. (1957b). An Approach to Time Series Analysis. *The Annals of Mathematical Statistics*, 32 (4), 951-989.

Parzen, E. (1971). *Teoría Moderna de Probabilidades y sus aplicaciones* (Trad. E. Berumen). México: Limusa. (Original en inglés, 1960).

Pearson, K. (1924). Historical Note on the Origin of the Normal Curve of Errors. *Biometrika*, 16 (3/4), 402-404.

Peltier, M. L. (1993). Una visión general de la didáctica de las matemáticas en Francia. *Educación matemática*, 5(2), 4-10.

Petrov, V. y Mordecki, E. (2002). *Teoría de probabilidades*. Moscú: Editorial URSS.

Piaget, J. (1991). *Introducción a la epistemología genética* (Vol. 1. El pensamiento matemático) (Trad. M.T. Cevasco y V. Fishman). México: Paidós. (Original en francés, 1950).

Piaget, J. y Inhelder, B. (1975) *The origin of the idea of chance in children* (Trad. Por L. Leake, P. Burrell y H. Fishbein). New York: Norton & Company Inc. (Original en francés, 1951).

Poisson, S. D. (1837). Recherches sur la probabilité des jugements en matière criminelle et en matière civile. Paris: Bachelier. Obtenido el 23 de junio de 2006 de: http://gallica.bnf.fr/ark:/12148/bpt6k110193z.pagination.

Polya, G. (1965). *Cómo plantear y resolver problemas* (Trad. J. Zagazagoitia). México: Trillas. (Original en ingles, 1965).

Rubin, A., Bruce, B. y Tenney, Y. (1991). Learning about sampling: Trouble at the core of statistics. En D. Vere-Jones (Ed.) *Proceedings of the Third International Conference on Teaching Statistics* (pp. 314–319). Voorburg, Holanda: IASE.

Shaughnessy, J. M. (1992). Research in probability and statistic: Reflections and directions. En Grouws D. A. (Ed.) *Handbook of Research on Mathematics Teaching and Learning* (pp. 465-494). New York: Macmillan Publishing Company.

Schoenfeld, A. (1985). *Mathematical Problem Solving*. New York: Academic Press.

Schoenfeld, A. (2002). Research Methods in (Mathematics) Education. En M. Bartolini, et al. (Eds.) *Handbook of International Research in Mathematics Education* (435-488) Mahwah: Lawrence Erlbaum Associates, Publishers.

Secretaría de Educación Pública (2003). *Plan y Programa de Estudios de Educación Secundaria* (en elaboración) [En línea]. Obtenido el 20 septiembre de 2003 en la dirección: http://www.sep.gob.mx/.

Steinbring, H. (1997). Epistemological investigation of classroom interaction in elementary teaching. *Educational Studies in Mathematics*, 32, 49-92.

Székely, G. J. (1986). *Paradoxes in probability theory and mathematical statistics* (Trad. M. Alpár y É. Unger). Dordrecht-Boston: Reidel-Kluwer. (Original en húngaro, 1982).

Tauber, L. (2001). *La construcción del significado de la distribución normal a partir de actividades de análisis de datos*. Tesis Doctoral no publicada. Universidad de Sevilla, España.

Tchebychef P. L. (1899). Des Valeurs Moyennes (Trad. M. N. De Khanikof). En A. Markoff y N. Sonin (Eds.) *Oeuvres de P. L. Tchebychef* (Tomo I). St. Petesburgo: Académie Imperiale des Sciences (pp. 687-694). (Original en ruso, 1867). Obtenido el 12 de junio de 2006 de la University of Michigan Historical Math Collection. http://name.umdl.umich.edu/AAS7805.0001.001.

Vallecillos, A. (1996). *Inferencia estadística y enseñanza: Un análisis didáctico del contraste de hipótesis estadísticas*. Madrid: Comares.

Vallecillos, A. (1999). Some empirical evidence on learning difficulties about testing hypotheses. *Bulletin of the International Statistical Institute: Proceedings of the Fifty-Second Session of the International Statistical Institute* (Tome LVIII, Book 2, pp. 201-204). Helsinki: International Statistical Institute.

Villareal, M. (2003, Febrero). La investigación en educación matemática. *Boletín de la SOAREM,* 16. Obtenido el 9 de mayo de 2006 de: http://soarem.org.ar/publicaciones/htm.

Vergnaud, G. (1990). La théorie des champs conceptuels. *Recherches en Didactique de Mathématiques*, 10 (2-3), 133-170.

Von Glasersfeld, E. (1987). Learning as constructive activity. En C. Janvier (Ed.) *Problems of representation in the teaching and learning of mathematics* (Cap. 1). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Wackerly, D., Mendenhall, W. y Scheaffer L. (2002). *Estadística matemática con aplicaciones* (6ª Ed) (Trad. J. Yescas). México: Thompson. (Original en inglés, 2002).

Wild, C. y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry (con discussion). *International Statistical Review*, 67 (3), 223-265.

Anexos

ANEXO 1

Elementos de Significado de la variable aleatoria. De acuerdo a Ortiz (2002)*

Elementos intensivos del significado (p.121):

Definiciones y propiedades de la variable aleatoria.

VA1: La variable aleatoria toma sus valores dependiendo de los resultados de un experimento aleatorio.
VA2: Es una función del espacio muestral en *R*.
VA3: Queda caracterizada mediante la distribución de probabilidad: Conjunto de valores que toma junto con su probabilidad.
VA4: Se requiere que, para cada intervalo *I* de *R* el conjunto original sea un suceso del espacio muestral.
VA5: Una variable aleatoria define una medida de probabilidad sobre el conjunto de números reales.
VA6: Para cada variable aleatoria podemos definir una función de distribución de la forma siguiente:

$$R :\to [0,1]$$
$$x :\to F(x) = P(\xi \leq x)$$

VA7: La función de distribución de una variable aleatoria es una función real de variable real, monótona no decreciente.
VA8: La función de distribución de una variable aleatoria determina en forma biunívoca la distribución de probabilidad.
VA9: Sea $(x_i, p_i) i \in I$ la distribución de probabilidad de una variable aleatoria discreta. Se define la media o esperanza matemática como $E[\xi] = \sum_{i \in I} x_i p_i$. Este concepto extiende la idea de media en una variable estadística.
VA10: La moda es el valor más probable de la variable.
VA11: La mediana es el valor de la variable para el cual la función de distribución toma el valor 1/2. Por tanto, la probabilidad de que una variable aleatoria tome un valor menor o igual a la mediana es exactamente 1/2.

* Ortiz se basa en el marco teórico de Godino y Batanero (1994 y 1998a) para describir sus categorías de análisis.

Elementos Extensivos del significado (pp. 199-202)

Actividades: problemas y ejercicios característicos de la variable aleatoria.

SVA1: Determinar la distribución de probabilidad de una variable aleatoria.
SVA2. Determinar la probabilidad de que una variable aleatoria tome un valor determinado.
SVA3. Dada una lista de variables aleatorias, decidir si son cualitativas o cuantitativas y cuáles son discretas y cuáles continuas.
SVA4: Representar gráficamente la distribución de probabilidad de una variable aleatoria.
SVA5: A partir de la representación gráfica, hallar la distribución de probabilidad.
SVA6: Obtener la función de distribución dada la distribución de probabilidad.
SVA7: Obtener alguno de los valores de posición central, media, mediana o moda o de dispersión.
SVA8: Dada una distribución de probabilidad comprobar si verifica los axiomas de probabilidad.
SVA9: Calcular el valor de la variable aleatoria al que corresponde una cierta probabilidad.

Variables adicionales (pp 50-51)

En la descripción de los ejercicios, además se consideran las siguientes variables:

V1: Tipo de actividad que se pide al alumno: Dentro de esta variable hemos considerado cuatro categorías: 1) Ejemplo introductorio; 2) ejemplo después de la definición; 3) ejercicio introductorio; 4) ejercicio después de la definición. y 5) ejercicio después de haber hecho un ejemplo similar.
V2: Concepto al que el ejercicio se refiere explícitamente o implícitamente, entre los citados anteriormente.
V3: Tipología del ejercicio o ejemplo dentro de cada concepto. Esta es la principal variable dentro de nuestro estudio, pues describe la actividad especifica que debe realizar el alumno para resolver el ejercicio o que se ejemplifica en el ejemplo. Esta tipología de situaciones se describirá con detalle, para cada uno de los conceptos, en las secciones 3.4 a 3.9 y permitirá describir los elementos extensionales de los conceptos probabilísticos elementales.
V4: Tipos de espacio muestral. Hemos considerado de interés analizar el tipo de espacio muestral del experimento aleatorio que interviene en las situaciones propuestas al alumno, diferenciando entre espacio muestral infinito, finito, con dos elementos equiprobables; finito, con más de dos elementos equiprobables, finito, con sucesos no equiprobables e impreciso.
V5: Posible asignación de probabilidades a los sucesos. Un punto importante es la asignación de probabilidades a los sucesos dentro de los experimentos que intervienen en la situación, porque ello está directamente relacionado con los distintos significados del término probabilidad descritos en la sección 2.5.
V6: Contexto del enunciado del ejercicio: Se refiere al campo de aplicación mostrado.
V7: Presentación de la información: Verbal, numérica o gráfica.

ANEXO 2

La Actividad

PARTE I

A raíz de los festejos del día del niño, el departamento de relaciones públicas de una fábrica desea conocer el número de hijos que tiene cada uno de los 200 obreros que ahí laboran. La intención es efectuar una rifa que beneficie a los hijos de los trabajadores. Los resultados de la encuesta realizada para tal efecto son:

Número de hijos	Número de trabajadores
0	16
1	22
2	33
3	45
4	31
5	20
6	12
7	9
8	7
9	5

a) Si tomamos un trabajador al azar, ¿cuál es la probabilidad de que tenga 4 hijos? ¿Y 6 hijos? ¿Y 7? ¿Cuál es la probabilidad de que no tenga ningún hijo?
b) ¿Qué depende de qué: el número de hijos depende de la probabilidad o al revés? ¿Cuál de estas dos variables sería la variable dependiente y cuál la variable independiente?

PARTE II

En la fiesta se premiará a la familia de un obrero con boletos para el teatro, pero los boletos de teatro se tienen que reservar con días de anticipación, así que le encomiendan a la trabajadora social de la empresa que decida cuántos boletos tiene que comprar.

En la siguiente columna calcula la probabilidad de que el trabajador premiado tenga un hijo, dos hijos, tres hijos, … y nueve hijos.

Número de hijos	Número de trabajadores	Probabilidad
0	16	
1	22	

2	33	
3	45	
4	31	
5	20	
6	12	
7	9	
8	7	
9	5	

Indica las operaciones que efectuaste para llenar la tabla y contesta:

c) ¿Qué recomendación darías a la trabajadora social de cuántos boletos comprar para que la empresa pierda lo menos posible? Explica detalladamente porqué le darías esa recomendación.
d) De acuerdo a tu recomendación ¿con qué probabilidad el trabajador premiado ocuparía exactamente el número de boletos que compró la trabajadora social?
e) De acuerdo a tu recomendación ¿qué tanto puedes asegurar que se ocuparan exactamente los boletos que se compraron?
f) De acuerdo a tu recomendación, ¿con qué probabilidad al trabajador premiado le alcanzarán los boletos y hasta le sobrarán?
g) De acuerdo a tu recomendación, ¿con qué probabilidad el trabajador premiado no le alcanzarán los boletos para poder llevar a todos sus hijos al teatro?

PARTE III

Analicemos matemáticamente la situación…

Utiliza la información que obtuviste en la Parte I acerca de la variable dependiente e independiente y la tabla que hiciste en la segunda parte para hacer una gráfica que represente la probabilidad de que un trabajador con un cierto número de hijos salga premiado.
h) Representa con una letra la variable dependiente y con otra la variable independiente y utiliza estas letras para representar en notación funcional el resultado obtenido en el inciso b) de la parte II.
i) Describe, en el contexto del problema, con tus propias palabras la información que proporciona la variable dependiente de la variable independiente.
j) Describe el dominio de la variable independiente y el rango de la variable dependiente.
k) Describe el método a partir del cuál puedes obtener el valor de la variable dependiente, a partir del valor de la variable independiente.
l) La relación que estableciste ¿es una función matemática? ¿Por qué?
m) ¿Qué diferencias encuentras entre la variable independiente de esta función de probabilidad y las variables independientes de las otras matemáticas que tú conoces, por ejemplo, de álgebra o cálculo?
n) ¿Qué nombre le darías a este tipo de variable?

ANEXO 3

Solución de referencia de la Actividad

PARTE I

a) *Si tomamos un trabajador al azar, ¿cuál es la probabilidad de que tenga 4 hijos? ¿Y 6 hijos? ¿Y 7? ¿Cuál es la probabilidad de que no tenga ningún hijo?*

Hay tres posibles respuestas correctas, todas ellas siguiendo el razonamiento de la probabilidad clásica (o la regla de Laplace):

- De 4 hijos: $\frac{31}{200}$; de 6 hijos: $\frac{12}{200}$; de 7 hijos: $\frac{9}{200}$; de ningún hijo: $\frac{16}{200}$.

- De 4 hijos 0.155; de 6 hijos: 0.06; de 7 hijos: 0.045; de ningún hijo: 0.08.

- De 4 hijos 15.5%; de 6 hijos: 6%; de 7 hijos: 4.5%; de ningún hijo: 8%.

b) *¿Qué depende de qué? ¿El número de hijos depende de la probabilidad o al revés? ¿Cuál de estas dos variables sería la variable dependiente y cuál la variable independiente?*

Aquí hay una ambigüedad en cuál es la función a la que nos referimos. Podría ser a la función de probabilidad (dada por la variable aleatoria y la probabilidad) o a la formada por la probabilidad y los conjuntos de trabajadores que tienen un cierto número de hijos. En el primer caso la probabilidad depende del número de hijos, por lo tanto el número de hijos es la variable independiente y la probabilidad es la variable dependiente. En el segundo, la probabilidad depende de los conjuntos formados por los trabajadores con un cierto número de hijos. Esperamos la primera respuesta por las variables que indicamos.

PARTE II

Cálculo de la probabilidad de que el trabajador premiado tenga el número de hijos dado.

Número de hijos	*Número de Trabajadores*	*Probabilidad*	
0	16	$\frac{16}{200}$	0.08
1	22	$\frac{22}{200}$	0.11
2	33	$\frac{33}{200}$	0.165

3	45	$\frac{45}{200}$	0.225
4	31	$\frac{31}{200}$	0.155
5	20	$\frac{10}{200}$	0.1
6	12	$\frac{12}{200}$	0.06
7	9	$\frac{9}{200}$	0.045
8	7	$\frac{7}{200}$	0.035
9	5	$\frac{5}{200}$	0.025

Esta misma tabla se puede llenar con las probabilidades en forma de porcentaje.

c) *¿Qué recomendación darías a la trabajadora social de cuántos boletos comprar para que la empresa pierda lo menos posible? Explica detalladamente porqué le darías esa recomendación.*

Lo más probable es que el trabajador seleccionado tenga tres 3 hijos o menos, así que es necesario que se compren tres boletos para los niños más los dos boletos del trabajador y su esposa es decir, 5 boletos en total, de ese modo, la mayoría de los trabajadores (el 58%) no tendría problemas para llevar a sus hijos al teatro si salen premiados, al mismo tiempo que la empresa no perdería demasiado si sobran boletos.

Hay otras soluciones posibles, como elegir un valor de la variable mayor, pero esperamos que el alumno elija esta, ya que en este problema la mediana es muy próxima a la moda, aunque la media da un valor algo mayor (3,4) se acerca más al valor 3 que al 4. Por tanto todos los promedios coinciden aproximadamente en este contexto.

d) *De acuerdo a tu recomendación ¿con qué probabilidad el trabajador premiado ocuparía exactamente el número de boletos que compró la trabajadora social?*

Con una probabilidad de 0.225, es decir, en realidad es poco probable que se ocupen exactamente los 5 boletos.

e) *¿Qué tanto puedes asegurar que se ocuparan exactamente los boletos que se compraron?*

Es poco probable, es decir es más probable que no se ocupen exactamente puesto que sólo 45 de los trabajadores (de los 200) tienen oportunidad de ocupar exactamente el número de boletos recomendado. Es más alta la posibilidad de que sobren boletos (0.355) o bien de que falten (0.42).

f) De acuerdo a tu recomendación, ¿con qué probabilidad al trabajador premiado le alcanzarán los boletos y hasta le sobrarán?

Con una probabilidad de 0.58 (0.08+0.11+0.165+0.225=0.58), es decir, es más probable que alcancen y hasta sobren los boletos a que hagan falta boletos. Esto significa que de 100 trabajadores, 58 tendrían posibilidades de no tener problemas con sus hijos en caso de que salgan premiados.

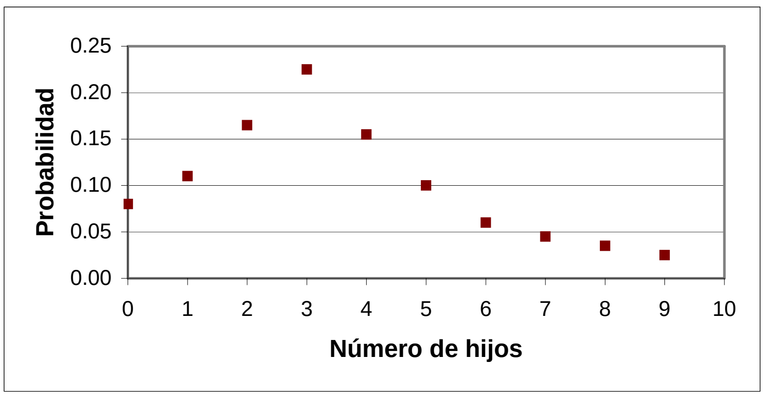
g) De acuerdo a tu recomendación, ¿con qué probabilidad al trabajador premiado no le alcanzarán los boletos para poder llevar a todos sus hijos al teatro?

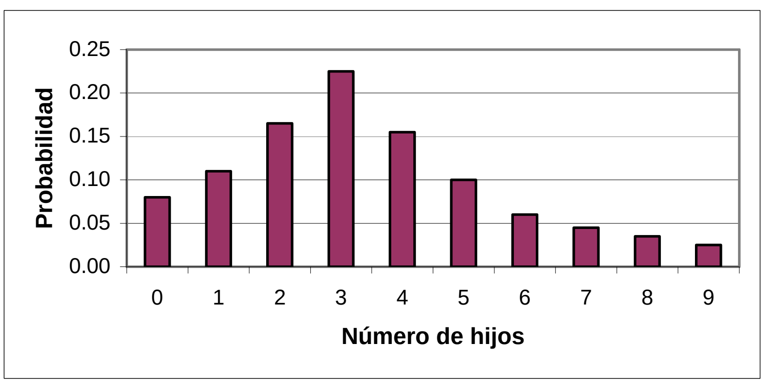
Con una probabilidad de 0.42. Es más probable que sí alcancen a que no alcancen. El 0.42 puede provenir de dos operaciones: la suma de 0.155, 0.1, 0.06, 0.045, 0.035 y 0.025 ó bien de la resta de 1 menos 0.58.

PARTE III

Gráfica que representa la probabilidad de que un trabajador con un cierto número de hijos salga premiado.

Hay dos posibles gráficas correctas:

❖ 

❖ 

h) Representa con una letra la variable dependiente y con otra la variable independiente y utiliza estas letras para representar en notación funcional el resultado obtenido en el inciso b) de la parte II.

El resultado del inciso b) de la parte dos se refiere a con qué probabilidad el trabajador premiado ocuparía exactamente el número de boletos que compre la trabajadora social, por

lo tanto el número de hijos será de 3, $x = 3$, $y = 0.225$, en notación funcional: $f(3) = 0.225$.
Esperamos que los alumnos sugieran una notación aproximada como la siguiente:

> x: variable independiente (número de hijos) (también podría ser n o cualquier otra).
>
> y: variable dependiente (probabilidad) (o P o cualquier otra letra)

En notación funcional: y = $f(x)$ (utilizando las letras que escogieron para denotar a las variables).

i) Describe, en el contexto del problema, con tus propias palabras la información que proporciona la variable dependiente de la variable independiente (en la función que acabas de representar).

La variable dependiente indica la probabilidad de que el trabajador que salga premiado tenga 'x' número de hijos. También se puede expresar como las posibilidades que hay de que un trabajador que tenga un número 'x' de hijos sea el premiado.
Nótese que lo que se espera es que regresen al contexto del problema.

j) Describe el dominio de la variable independiente y el rango de la variable dependiente.

En esta pregunta se suscita un conflicto con la teoría. Desde la perspectiva estricta de la teoría de probabilidades, la función de probabilidad toma valores de [0,1] en su rango y valores enteros comprendidos entre $[0, \infty)$ en su dominio. Esto es porque la probabilidad definida en la función de probabilidad es una extensión de la probabilidad definida en el espacio muestral (en realidad no son iguales). En este caso no esperamos que los estudiantes respondan de ese modo, puesto que no han tomado un curso de teoría de probabilidad sino que esperamos que interpreten en el contexto de una variable discreta, tomando en cuenta solamente los valores de la probabilidad que le asignaron a los valores de la variable aleatoria que estuvieron trabajando.

Dominio D $=\{ x \mid x \in \{0,1,2,3,4,5,6,7,8,9\}\}$
Rango R $= \{y \mid y \in \{0,0.025,0.035,0.045,0.06,0.08,0.1,0.11,0.155,0.165,0.225\}\}$
O descripciones semejantes.

k) Describe el método a partir del cuál puedes obtener el valor de la variable dependiente, a partir del valor de la variable independiente.

Las respuestas pueden ser diversas, dos que consideramos correctas serían:

- ❖ El valor de la probabilidad (variable dependiente) se obtiene dividiendo el número de trabajadores correspondiente al número de hijos de interés (variable independiente) entre 200. Para conocer el número de trabajadores correspondiente se recurre a la tabla dada.
- ❖ También se puede obtener directamente a partir de la tabla o de la gráfica que se hizo.

l) La relación que estableciste ¿es una función matemática? ¿por qué?

Sí es una función matemática porque a cada valor de la variable independiente (número de hijos) le corresponde un solo valor de la variable dependiente (probabilidad).

m) ¿Qué diferencias encuentras entre la variable independiente de esta función de probabilidad y las variables independientes de las otras matemáticas que tú conoces, por ejemplo, de álgebra o cálculo?

En la función (distribución de probabilidad) el valor de la variable independiente depende de un proceso aleatorio.
Los valores de la variable independiente (número de hijos) no necesariamente se presentarían uno tras otro si repetimos una y otra vez la rifa en las mismas condiciones. Más aún, algunos valores es posible que no se presenten después de varios intentos. En este tipo de experimentos no se puede predecir el valor que se obtendrá ni siquiera de manera aproximada, sólo podemos saber qué es lo más probable que ocurra, mientras que en los problemas tradicionales de álgebra y cálculo hay un orden de aparición de los valores de las variables.

n) ¿Qué nombre le darías a este tipo de variable?

Variable aleatoria o variable estocástica o variable azarosa u otro nombre similar en el que se haga énfasis en que el azar está vinculado con ella.

ANEXO 4

Transcripción de la entrevista

En la transcripción del diálogo de las personas que intervienen en la entrevista, se añaden en cursivas notas explicativas necesarias para la comprensión de la acción durante su lectura, así como tiempos de grabación y notas de los archivos de video y audio. Durante la entrevista participaron: Brenda (*alumna*), Mónica (*alumna*), Profesor (*de matemáticas de Brenda y Mónica de primer semestre*) y Blanca (*investigadora*).

PRIMERA PARTE

Brenda y Mónica transcribieron los datos (en forma de tabla) al pizarrón. La tabla contenía tres columnas correspondientes al número de hijo y al número de trabajadores y a la probabilidad. Contestan el primer inciso directamente en la tabla. La probabilidad la escriben como el cociente de dos números enteros en todos los casos. También escribieron en el pizarrón sus respuestas a cada inciso de la primera parte.

En el pizarrón:

Número de hijos	*Número de trabajadores*	*Probabilidad*
0	16	$\frac{16}{200}$
1	22	$\frac{22}{200}$
2	33	$\frac{33}{200}$
3	45	$\frac{45}{200}$
4	31	$\frac{31}{200}$
5	20	$\frac{20}{200}$
6	12	$\frac{12}{200}$
7	9	$\frac{9}{200}$
8	7	$\frac{7}{200}$
9	5	$\frac{5}{200}$

Primer pasaje: ***Definición de la probabilidad como variable dependiente o independiente.***

ByM01 (video) -- 0:22.8

Mónica tiene la actividad por escrito y se lo lee a Brenda. Brenda, en tanto, está terminando de escribir sus respuestas en el pizarrón.

Mónica: ¿Qué depende de qué: el número de hijos depende de la probabilidad o al revés? Y luego ¿cuál de estas dos variables será la variable independiente e independiente?
Brenda: Bueno... primero ¿qué depende de qué? pues el número de hijos... no...el número
Ambas: de trabajadores... *(muy seguras de su respuesta)*
Brenda: ...depende del número de hijos, ¿no?
Mónica: No, no es que dice: el número de hijos depende de la probabilidad, no del...
Brenda: no.. pero la primera no es...
Mónica: ¿qué depende de qué? ...pues...
Ambas: el número de hijos depende de la probabilidad ¿o al revés?...
Mónica: No, la probabilidad depende del número de hijos o sea que qué tan probable... o sea... depende... la probabi...Yo digo que la probabilidad depende del número de hijos. (*recurre a voltear a ver a los profesores para que le confirmen o le refuten esta última afirmación)*
Brenda: Pero más bien... el número...el...bueno... no, sí la probabilidad depende del número de hijos porque... *(dudando)*
Mónica: Es que la probabilidad...
Brenda: No, sí la probabilidad... *(muy segura)*
Mónica: Sí, porque según yo...
(se arrebatan la palabra una a la otra)
Brenda: depende del número de hijos porque...
Mónica: Ajá, porque qué tan probable es...
Brenda: porque qué tan probable es que los trabajadores tengan ese número de hijos.
Mónica: Ajá.
Brenda: Entonces sí *(escribe la conclusión en el pizarrón).*

En el pizarrón:

Figura 14. b) La probabilidad depende del número de hijos

-- 01:31.1

Segundo pasaje: ***Dan una primera aproximación a la noción de función.***

ByM01(video)--- 1:38.9

Nuevamente Mónica tiene la actividad por escrito y se la lee a Brenda.

Mónica: ¿Cuál de estas dos variables sería la variable dependiente y cual la independiente? Ésta. ¿Hablan de aquí *(se refiere al inciso anterior)* o de... ?
Brenda está terminando de escribir la conclusión en el pizarrón, así que no le hace mucho caso a Mónica.
Mónica: Es que ve estás dos no pueden ser de que ¡ah! dependiente o independiente *(señala las columnas de la tabla)...* así, no. Yo no puedo entrar... de lo que el número de hijos y el número de trabajadores ... no puede ser, pon tú dependiente e independiente o

independiente y dependiente *(señala las columnas de la tabla)* porque no, no hay relación... por ejemplo aquí *(señala la columna de número de trabajadores)* empieza un número 16, y luego aquí mayor, mayor, incrementa y luego decrementa y luego decrementa, así como que no hay rela...o sea no hay mucha relación, igual de cero hijos pudo haber un trabajador y de cuatro a cinco mil y así... como que no da resul...pero dice...
-- 2:22.9

Tercer pasaje: *Definen cuál es la variable independiente y cuál es la variable dependiente en el problema.*

ByM01(video)-- 2:32.0
Nuevamente Mónica tiene la actividad por escrito y se la lee a Brenda.
Mónica: El número de hijos depende de la probabilidad o al revés y ya la pusimos. Ahora dice ¿Cuál de estas dos variables sería la variable dependiente y cuál la independiente?
Brenda: Pues yo digo que el número de trabajadores depende del número de hijos.
Mónica: ¿Por qué? Porque es más probable que sea del tipo 3 ó 4
Brenda: Sí, porque, bueno se supone que... o sea aquí por decir, x... o sea'x'sería la... la variable independiente, o sea los número de hijos son independientes, porque según o sea... como... Sí has de cuenta... ya ves que sería como poner 'x' y 'y'...
Mónica: Ajá. A ver ponlo.
(Brenda escribe una'x'al final de la columna del número de hijos y una 'y' al final de la columna del número de trabajadores)

En el pizarrón:

Número de hijos	*Número de trabajadores*	*Probabilidad*
0	16	$\frac{16}{200}$
1	22	$\frac{22}{200}$
2	33	$\frac{33}{200}$
3	45	$\frac{45}{200}$
4	31	$\frac{31}{200}$
5	20	$\frac{20}{200}$
6	12	$\frac{12}{200}$
7	9	$\frac{9}{200}$
8	7	$\frac{7}{200}$

9	5	$\frac{5}{200}$
x	y	

Brenda: Entonces está sería la 'x' *(señala la columna de número de hijos)* porque realmente... o sea si te fijas...
Mónica: Bueno es que aquí sí crece y luego vuelve para abajo otra vez *(señala la columna de número de trabajadores).*
Brenda: Sí, pero si te fijas esta es la que.. esta es la que... *(señala la columna de número de hijos)*
Mónica: Va determinando...
Brenda: Ajá, va determinando al número de trabajadores. El número de trabajadores no determina al número de hijos.
Mónica: Ah, ya entiendo.
Brenda: Sino al revés, el número de hijos es lo que está determinando el número de trabajadores. Entonces está sería...
Mónica: Pero eso no significa que depende de los hijos va a ver tantos trabajadores sino que esta *(señala la columna de número de hijos)* rige a esta *(señala la columna de número de trabadores).*
Brenda: Entonces en este caso, el número de hijos es la variable independiente y el número de trabajadores es la variable dependiente.
(Brenda escribe la conclusión en el pizarrón)

En el pizarrón:

Número de hijos = V. independiente Número de trabajadores = V. dependiente

Figura 15.

-- 4:23.8

Cuarto pasaje: *Cuestionamiento sobre los motivos para definir las variables independiente y dependiente.*

ByM01(video)-- 4: 42.9
Profesor: ¿Si tomamos un trabajador al azar cuál es la probabilidad de que tenga 4 hijos?
Ambas: 31 de 200.
Profesor: ¿Por qué no pusieron por ejemplo 4 entre doscientos u otro número porque 31?
Brenda: Porque bueno, en la tabla el número de hijos que son 4 tienen un número de trabajadores de 31 y el universo de trabajadores son 200. Entonces de esos 200 quiere decir que 31 tienen 4 hijos.
Mónica: Es como que al sumar todos estos *(señala la columna de número de trabajadores)* y te dan 200, pero pues ya te lo da de antemano el problema, te dice que son 200 obreros en total. Este es el número de trabajadores obreros *(señala la columna)* que si hacemos la suma te da 200. Entonces por ejemplo de 5, a pues a ver 5 *(busca en la tabla)*, aquí están los 5. Hay 20 trabajadores de 200 que tienen 5 hijos.
Profesor: La cuestión sería ¿Qué pasaría si sumáramos todas las probabilidades? La probabilidad de que tenga un hijo, dos hijos, así..., ¿qué pasaría?

Brenda: ¿Sumar el número de hijos?
Profesor: Las probabilidades. Miren por ejemplo, cuál es la probabilidad de que... haya 4 hijos.
Brenda: 31 sobre 200.
Profesor: Bueno entonces 31 sobre 200 es una probabilidad. ¿Qué pasaría si sumo cada caso?
Brenda: Deben de dar 200, o sea debe de dar todo el universo.
--- 6:27.5
ByM01(video)-- 6:52.1
Profesor: Bueno, entonces la pregunta es si la probabilidad puede valer 200.
Brenda: Sí, bueno, ¿qué la probabilidad dé algo o sea pueda valer 200?
Profesor: Sí, por ejemplo el 31 sobre 200 es un número negativo, positivo...
Mónica: Positivo. Fracción.
Profesor: Fracción, mayor que uno, menor que uno,...
Ambas: Menor que uno.
Profesor: Menor que uno.¿Y las probabilidades son mayores que uno?
Mónica: Son menores que uno. Todas son menores que uno.
Profesor: Entonces si sumamos todas esas probabilidades, ¿qué creen que pueda pasar?
Brenda: Sería igual a 200. O sea te quedaría 200 y sería igual a... o sería la unidad uno. Quedaría 200...
--- 7:27.5
ByM01 (video)-- 8:04.4
Profesor: Es decir, la probabilidad de tener cuatro hijos...
Brenda: O sea, esta es la probabilidad de 31,... *(le arrebata la palabra, señala la respuesta al inciso a), 31/200, escrita a un lado de la tabla).*
Profesor: 31 de 200 *(ninguna de las dos lo escucha, enfrascadas en su discusión).*
Brenda: ...entonces estas son todas las probabilidades que hay *(señala la columna de número de trabajadores)...*
Mónica: A pues sí.
Brenda: ...y si las sumamos sí nos van a dar 200. Haz de cuenta que ésta *(señala el 31 en la columna de número de hijos)* es la probabilidad de tener 4 hijos. Está diciendo el profe que se refiere a estas probabilidades *(señala la columna de número de trabajadores)* y al sumarlas nos darían 200.
Mónica: Pues sí
Profesor: ¿16 sería una probabilidad?
Ambas: ¿16? *(extrañadas)*
Brenda: ¿así, sola?
Profesor: Sí, así sola.
Brenda: No, así solo, no.
Profesor: ¿Qué le faltaría?
Mónica: 16 de cuántos.
Profesor: ¿De cuántos sería?
Ambas: De 200.
Profesor: Entonces la cuestión es, ¿si sumando todas? ¿qué pasa si sumo probabilidades? ¿qué se esperaría que diera?
Brenda: El universo en general, los 200.
(Mónica asiente con la cabeza)

Profesor: Denme una respuesta entre las dos. Lleguen a un acuerdo.
Mónica: Pero es que.. probabilidad sí yo estoy de acuerdo que sí son estos *(señala la columna de número de trabajadores)* porque es todo.. es un porcentaje... es que tanto, que tan probable es que salgan... el número de hijos. Pues sí, aquí *(señala la columna de número de trabajadores)* de hecho serían 200. Pero lo que estábamos discutiendo de que... ya nos había preguntando de 200... y luego que tan probable de que si sumo ¿éstos *(señala la columna de número de hijos)* nos preguntó ahorita?
Profesor: Las probabilidades.
Ambas: No, las probabilidades son 200.
Mónica: Las probabilidades son 200. Ya.
Brenda: Al sumar las probabilidades te debe dar todo el conjunto que estás.. o sea, todo el universo.
(Las dos asienten, están de acuerdo y convencidas)
Profesor: Nada más la pregunta es si una probabilidad puede valer 200. Es que me dicen que la probabilidad suma 200.
Mónica: Pues ya no sería probabilidad sino un total *(riéndose).*
Brenda: Bueno, lo que podría ser es que el número... es decir, podría ser la probabilidad de que los trabajadores tuvieran de 0 a 9 hijos. Esa sería la única forma en que hubiera una probabilidad de 200.
Mónica: Ajá. Total ya, sin dividir. En conjunto.
Profesor: Sin embargo, en la primera respuesta de cuatro hijos dice 31 sobre 200. Y ustedes me dijeron que todas las probabilidades son fracciones menores que uno. Entonces en total, hay 10 fracciones porque son 10 casos y si todas son menores de uno, ¿cómo le hacen para que llegue a 200?
Desconcertadas Miran las tablas en el pizarrón.
Mónica: ¿Cómo?... o sea ¿cómo maestro?
Profesor: Sí, ustedes me acaban de decir hace ratito que la probabilidad de tener 4 hijos es 31 sobre 200 y que esa era una fracción menor que uno ¿no?
Ambas: Ajá.
Profesor: Y luego me dijeron que todas las probabilidades son fracciones, fracciones menores que uno. Entonces si son 10 probabilidades a lo más que podríamos aspirar es a que su suma fuera menor que 10. Cuando yo les pregunto cuál es la suma de probabilidades, ustedes me dicen que 200 así que ¿Cómo le hacen para que llegue a 200?
Hacen exclamaciones de obviedad
Mónica: ¡Ay maestro!
Brenda: Lo que pasa es que esta fracción es con respecto a 200.
Mónica: Ajá. El común denominador es 200.
-- 11:28.3
Profesor: ¿Tú que piensas Mónica?
Mónica: Yo digo que aún sumando todas estas *(señala la columna de número de hijos)* no nos da mayor que 1 porque para que fuera mayor que 1 la fracción tendría que ser 201 o más entre 200. O sea 201 sobre 200 y estos *(refiriéndose a la columna de número de hijos)* no van a dar más de 200.
Profesor: O sea ¿estás de acuerdo con ella?
(Ninguna de las dos contesta. Se miran confundidas)
Brenda: No, pero la probabilidad se refiere a ésta *(señala la columna de número de trabajadores).*

Brenda voltea a ver al profesor, busca su apoyo. Mónica está confundida. Brenda tiene mayor seguridad en lo que dice.
Profesor: Decidan ente ustedes.
Mónica: No, ¡ya me dejó... ¿cómo la probabilidad va a ser ésta?! *(extrañada, señala la columna de número de trabajadores).*
ByM01 (video) -- 13:04.8
Profesor: ¿Podría tener sentido decir que la probabilidad tiene un tope, un límite, un número al cuál no puede rebasar? ¿o puede tomar cualquier valor positivo? Ustedes me acaban de decir que no tiene sentido que sea negativo, pero por ejemplo, ¿puede ser cero?
Mónica: Sí.
Brenda: Pues sí.
Mónica: Por ejemplo, de 10 hijos es cero, la probabilidad.
Profesor: Entonces, ¿la probabilidad puede tomar cualquier valor? ¿Puede haber probabilidad 2, probabilidad 3... probabilidad 200, puede haber probabilidad 1000 por ejemplo? ¿Tiene sentido?
Mónica: Dependiendo del número total, ¿no? No se puede exceder del 200, en este caso. ¿Sí estás de acuerdo?
Brenda: Sí.
Mónica: Dependiendo del valor total. También ni modo que haya 201 de 200. Pues no. O sea mayor que 200, no.
-- 14:15.7

Quinto pasaje: *Interpretación de la probabilidad expresada con números decimales*.

ByM02 (video) -- 00:10.9
El Profesor solicita a Brenda y Mónica que efectúen la división en cada caso, añaden otra columna a la tabla en donde escriben el número decimal resultado de cada división.

En el pizarrón:

Número de hijos	*Número de trabajadores*	*Probabilidad*	*Probabilidad*
0	16	$\frac{16}{200}$	0.08
1	22	$\frac{22}{200}$	0.11
2	33	$\frac{33}{200}$	0.165
3	45	$\frac{45}{200}$	0.225
4	31	$\frac{31}{200}$	0.155
5	20	$\frac{20}{200}$	0.1
6	12	$\frac{12}{200}$	0.06

7	9	$\frac{9}{200}$	0.045
8	7	$\frac{7}{200}$	0.035
9	5	$\frac{5}{200}$	0.025

Profesor: Nos podrían ayudar ustedes a encontrar un significado a ese decimal, ¿qué podría significar 0.155? ¿qué es eso?
Brenda: Yo digo que los doscientos trabajadores se están tomando como una unidad, entonces se supone que ese 0.155 es un pedacito de eso, a eso se refiere, entonces si sumamos cada fracción que da aquí, nos va a dar el entero que seria igual a uno. Bueno, yo entendí eso.
Mónica: Si yo también, o se puede hacer como le hacemos, o sea, se puede dividir todo en doscientas unidades y ya sumas cada cosita. O sea los 31 representan... es que es como un porcentaje, es un porcentaje.
Brenda: No, bueno yo diría que no, en este caso no sería porcentaje, o sea, has de cuenta que los doscientos lo estas tomando como un entero, o sea como una sola unidad y esto sería una fracción, porque si lo quisiéramos sacar como porcentaje, necesitaríamos sacar lo proporcional, o sea 200 es a 100 y 31 es a...
Mónica: Ah si, por eso, no es exactamente el porcentaje pero si es la parte, no es porcentaje porque es de 100, pero es una fracción. Es la palabra porcentaje.
Profesor: Entonces que fue ese 0.155.
Mónica: Una parte del entero.
Brenda: Si, o sea, que el universo, o sea los doscientos trabajadores se están tomando como uno solo, entonces esta probabilidad se refiere a una parte, es una fracción, una parte de ese entero que se esta tomando, entonces al sumar todo debe de dar un uno que sería el entero que se esta considerando.
Profesor: Muchísimas gracias.
-- 2:07.5

Sexto pasaje: ***Surgimiento, nuevamente, de la discusión en la definición de la variable dependiente e independiente.***

ByM02 (video)-- 3:32.7
Profesor: Respecto a esa última pregunta, el inciso b) dice: ¿qué depende de qué? ¿el número de hijos depende de la probabilidad o al revés?
Brenda: Que la probabilidad depende del número de hijos.
Blanca: Si pero... ahí fíjate, ustedes están manejando "la probabilidad depende del número de hijos" y cuando ponen variable independiente ponen "número de hijos", y variable dependiente "número de trabajadores"
Mónica: Ah si es cierto, esta al revés, nos estamos contradiciendo aquí.
Brenda: No porque, si estamos diciendo que la probabilidad depende del número de hijos, quiere decir que la probabilidad es la dependiente, el número de trabajadores que es la probabilidad es la dependiente *(señala sus respuestas en el pizarrón).*
Blanca: Pero ¿el número de trabajadores es la probabilidad?

Brenda: Si, son estas las probabilidades *(señala la columna de número de trabajadores en el pizarrón).*
Mónica: ¿no?
Blanca: No sé.
Profesor: Bueno, yo mas bien lo que quería decir es que la pregunta esta parece que se esta refiriendo nada mas a la probabilidad y al número de hijos, porque dice ¿qué depende de qué? ¿el número de hijos depende de la probabilidad o al revés?, y eso ya lo contestaron.
Mónica: Ajá, que es la probabilidad es la que depende, entonces sería la dependiente, los trabajadores, dependen del número de hijos.
Profesor: Y luego donde dice ¿cuál de estas dos variables sería la variable dependiente y cuál la independiente?
Mónica: Aquí. *(señala su respuesta a esta pregunta en el pizarrón)*
Profesor: Entonces el número de hijos sería la independiente, y entiendo que el número de trabajadores es, como de ahí se sacan las probabilidades lo hacen equivalente, es como si estuviéramos hablando de las probabilidades, ¿así lo están entendiendo?
Ambas: Sí.
-- 5:29.6

Séptimo pasaje: *Discusión entre la diferencia entre probabilidad y número de trabajadores. Surgimiento de las dos relaciones importantes: probabilidad-número de trabajadores y número de hijos-número de trabajadores*.

ByM02 (video) --- 6:41.4
Profesor: Podemos estar de acuerdo en que entendemos por probabilidad, ¿simplemente un número suelto?
Brenda: No, es un número que, o sea, de la unidad que se está tomando, no puede ser así...
Profesor: O sea, "es relativo a". Entonces cuando ustedes hablan allá abajo *(se refiere a la respuesta al inciso b) que está escrita en el pizarrón)* del número de trabajadores como variable dependiente, también sería "relativo a" .
Mónica: Relativo al total de trabajadores.
Profesor: ¿O no? ¿cómo ven?
Mónica: ¿éste maestro? *(señala la respuesta a la variable dependiente "numero de trabajadores" escrita en el pizarrón).*
Profesor: Sí, sí, como variable dependiente.
Mónica: O sea, es que el número de trabajadores es variable dependiente del número de hijos porque aquí ponemos la 'x' es el número de hijos y dependiendo de, si hay 4 hijos va a haber... el número de trabajadores.
Profesor: Si, si hay una asociación de dependencia, ahora nada mas el detalle es sobre probabilidad, es que dice ¿el número de hijos depende de la probabilidad o al revés? Pero de la probabilidad.
Brenda: O sea, de esto *(señala los valores de probabilidad obtenidos en el inciso a) escritos en el pizarrón).*
Blanca: Es que en la función de dependencia que ustedes establecen donde dice: "la probabilidad depende del número de hijos", ¿cuáles son las variables que manejan ustedes?
Mónica: Según lo que yo entendí, es la probabilidad. Esta *(señala la columna de número de trabajadores en el pizarrón).*
Blanca: Pero ¿esa es la probabilidad? ¿el 12 es probabilidad?

Brenda: No, de aquí se sacaría la probabilidad de 12 sobre 2.. *(señala el pizarrón).*
Mónica: Es que hay dos relaciones.
Blanca: ¿Cuáles son las dos relaciones?
Mónica: Esta *(señala el número de trabajadores)* y esta *(señala el número de hijos)*, y esta *(señala el número de trabajadores)* y el total *(se refiere al total de trabajadores)*
Blanca: Esa, y esa, y esa del total.
Brenda: Bueno, lo que pasa es que por ejemplo, se supone que el número de trabajadores depende del número de hijos, pero por ejemplo, de 31 esto sería la probabilidad, entonces de acuerdo a esta tabla *(señala la tabla)* se esta sacando que los hijos, o sea, la probabilidad de que haya esos hijos es el número de trabajadores. A ver ¿cómo te lo puedo explicar?
Mónica: Es que aquí están las tres variables *(señala el pizarrón),* tres porque de 0 hijos hay 16 personas que tienen 0 hijos de las 200, o sea, es que hay tres factores: ¿cuántos hijos?, ¿cuántas personas? ¿y de cuántas?, o sea, ¿cuántas personas de cuántas?; como que estas dos *(se refiere al número de trabajadores y al total de trabajadores)* van relacionadas muy directamente porque son 16 de 200, 9 de 200, pero esas dos se relacionan con esta *(con el número de hijos)*, porque esta es la variable de cuántos hijos son.
Blanca: Pero esas dos, 16 sobre 200, ¿son dos números o es un número?
Mónica: Es un número.
Blanca: Es un número. Entonces ¿cuáles son las dos variables que ustedes tienen?
Brenda: Lo que pasa es que la probabilidad, este número *(el número de trabajadores)* que te da sobre el universo, está dependiendo del número de hijos que tú le estas poniendo, porque por decir, allá arriba *(se refiere al inciso a) escrito en la parte de arriba del pizarrón)* te esta pidiendo: "quiero la probabilidad de 4 hijos", entonces tú estas sacando la probabilidad de acuerdo a esta tabla *(señala la columna del número de trabajadores)*, que te da 31, o sea, 31 de todo el universo y esa probabilidad esta dependiendo del número de hijos. Por eso la probabilidad esta dependiendo de los hijos. Esa probabilidad es de todo el universo, pero lo estas tomando de acuerdo al número de hijos que tú estas buscando.
Mónica: Esta probabilidad *(señala la palabra probabilidad en la frase "La probabilidad depende del número de hijos" en el pizarrón)* es este número *(señala el 16/200, referida a la respuesta a la probabilidad de que el trabajador seleccionado tenga cero hijos)* que es el mismo que este *(señala el 31/200 respuesta referida a la probabilidad de que tenga cuatro hijos)*, o sea, cerrado en un solo número, y esa es la probabilidad, este *(nuevamente 16/200)* que es igual que este *(nuevamente el 31/200).*
Blanca: Entonces ¿cuáles son las variables que ustedes manejan ahí?
Mónica: Yo diría que es el número de trabajadores entre total de trabajadores.
Blanca: ¿Es decir?
Mónica: Probabilidad.
Blanca: ¿cuál sería entonces la variable dependiente?
Mónica: La probabilidad.
Brenda: Sí, sería la variable dependiente.
-- 11:46.7

SEGUNDA PARTE

Ellas transcriben sus respuestas a la hoja de su reporte y borran sus respuestas a la primera parte del pizarrón, pero no la tabla de los datos.

Octavo pasaje: ***Primera respuesta a la recomendación que darían a la trabajadora social: Hay uso escaso de los datos de la tabla.***

ByM03 (video)-- 4:35.4
Mónica lee el problema
Mónica: ¿qué recomendación le darías a la trabajadora social de cuántos boletos comprar para que la empresa pierda lo menos posible? ¡Ah, pero dice que se tiene que comprar por anticipado! ¿no? Pero si faltan se pueden compara ahí.
Profesor: El caso es, que quieren comprar de tal manera que no falten...
Mónica: Pero tampoco que no pierdan.
Profesor: Ajá, porque ¿qué pasa si compran nueve?. Nueve mas dos son... ¿qué pasa si compraran 11 boletos?
Mónica: Puede perder, pero ¿y si sale?
Profesor: Pero resulta que 9... a ver ustedes ¿cómo ven?. Pero ¿será posible que salgan los 9?
Mónica: Pues yo para no perderle y para no errarle, compraría dos nada mas porque sí hay familias que tienen 0 hijos, de hecho hay 16 trabajadores que tienen 0 hijos. Entonces nada mas compraría dos para que vayan él y la esposa, y ya después si no sale, por lo menos ya tengo asegurados dos. Por eso preguntaba, si ya tengo asegurado dos... porque dice que pierda lo menos posible. En caso de que compre 3, y si por ejemplo sale una familia de estas, ya perdí un boleto, ya estoy perdiendo un boleto ¿qué voy a hacer con ese boleto?. Así pienso yo, yo compraría dos.
--- 6:03.4

Noveno pasaje: ***Establecimiento del primer supuesto implícito en su decisión de cuántos boletos comprar: los boletos deben comprarse con anticipación.***

ByM03 (video)-- 6:03.8
Brenda: Yo compraría 5, porque de acuerdo a la tabla, hay más probabilidad de que sea una persona con tres hijos y luego sumándole el obrero y la esposa serían 5.
Mónica: Si, hay mucha probabilidad. Pero y si te salen por ejemplo estas *(señala los trabajadores con menos de tres hijos)*. Ah profe, es que eso es lo que le preguntaba; lo que yo le estoy preguntando es muy importante porque, en caso de que yo compre dos me van a faltar, pero si tengo que comprar los exactos que lo menos pierda o puedo comprar ya después boletos.
Profesor: Ya no, después ya no.
Mónica: Ah pues entonces si tiene razón ella, se tienen que comprar 5.
Profesor: ¿Por qué?
Mónica: Porque yo había dicho que 2 para no fallarle y no perder nada, porque dije: compro dos mínimo y ya, si me sale una familia de 0 hijos, o sea, que nada mas iría la pareja pues ya se los doy y ya no perdí yo nada, pero si me salen mas pues vuelvo a comprar, no hay problema. Pero no, porque tengo que comprar una vez. ¡Ah! entonces compraría 5 boletos porque la mayor incidencia de familias es de 3 hijos, o sea, el mayor número de trabajadores que hay por hijo es 3, o sea, las familias con 3 hijos son las que mas abundan en la empresa, entonces como 3 hijos mas el papá y la mamá pues ya son 5.
--- 7:33.9

Décimo pasaje: *Discuten la decisión sobre cuantos boletos comprar usando más fuertemente los datos que la tabla de probabilidad les proporciona. Surge la necesidad de calcular la frecuencia acumulada.*

ByM03 (video)-- 7:47.4
Profesor: ¿Tú que dices Brenda de eso?
Brenda: Bueno, yo diría eso, comprar los 5, pero también existe la probabilidad de que no vaya a completar después.
Mónica: Pero va a ser un sorteo, yo siempre me lo imagino como una canasta con papelitos, entonces, hay mas probabilidades de que yo agarre un trabajador de estos 45 *(se refiere a los trabajadores que tendrían 3 hijos)*, o sea, en el círculo que dibujamos *(se refiere a la tabla)* éste *(señala el 45)* ocupa mas, hay mas probabilidades, es como que la que menos falla, es como que azar.
Brenda: Aunque bueno, también por decir un margen mas grande sería... bueno mas seguro sería comprar los 9, porque si te fijas, de aquí hasta acá *(señala en la tabla la columna de probabilidades las correspondientes de 0 a 6 hijos)* es donde se ocupa la mayor parte, de 0 a 6 es donde se esta ocupando la mayor parte inclusive (*se podría tomar)* un poco más, ya en la de 7 hijos que es donde llega a 9 *(es decir, añadiría los trabajadores que tienen 7 hijos)*, o sea, es donde esta la mayor probabilidad, ésta es muy poquita *(señala la probabilidad de que se tengan 8 ó 9 hijos).* Estos dos tienen muy poquita probabilidad y serían 7 más los... que serían... Si te fijas realmente ya nada mas tendrías un margen de 13 o sea muy pocos.
Mónica: Pero dice ¿cómo le hace la empresa para gastar lo menos posible? Requieres gastar lo menos posible.
Brenda: Por eso, por decir, sí, este es el que tiene más *(se refiere a los trabajadores que tienen 3 hijos)*, pero si compras nada más 5, resulta que, tienes todavía 31 posibilidades de que salgan con 4, 20 posibilidades de que salgan con 5 y 12 todavía con 6 y así, entonces si te fijas hasta acá *(señala el renglón con 6 hijos)* esto abarca la mayor posibilidad, si tienes cubierto todo esto, ya no vas a tener problemas de que la familia (*premiada)* se vaya a quedar sin boletos. Si te fijas realmente todo esto esta cubriendo la mayor probabilidad.
Mónica: Pero es que entonces, yo quiero que nadie se quede sin boletos, pues entonces de una vez compran los... 11...de una vez para ya no errarle.
Brenda: No pero la cuestión es...
Mónica: Ahorrar más es la cuestión, ¿qué no?
Brenda: Sí ahorrar de modo que no se vaya a..., pero (*con mi propuesta)* también va a quedar abierta la posibilidad de que no vayan a alcanzar los boletos. Si en todo caso fuera, ay pues ya compra todos, pero no ¿verdad?, ya no habría problema; pero en todo caso, yo diría que si sería hasta aquí *(trabajadores con 6 hijos)* o hasta aquí *(trabajadores con 7 hijos)*, cualquiera de los dos, porque entonces abarcaría la mayor parte de la probabilidad.
Profesor: Una sugerencia, me gusta la idea que están ustedes desarrollando pero sería bueno como registrarla tantito, ustedes dicen: "porque hasta ahí abarcaría la mayor parte de la probabilidad", ese fue su último argumento, entonces ¿por qué no ven que tanta probabilidad abarca según el número de boletos?, o sea por ejemplo, si no compran ningún boleto ¿qué probabilidad abarcaron?, si compran un boleto ¿qué probabilidad abarcaron? Y así...
-- 11:23.7

Undécimo pasaje: ***Uso de la probabilidad acumulada para argumentar cuántos boletos comprar. Ambas se preocupan en diferentes intereses: en los trabajadores o en la empresa.***

ByM04 (video)--- 0:4.18
En el pizarrón, a la tabla anterior le añadieron la columna de la probabilidad acumulada:

Número de hijos	*Número de trabajadores*	*Probabilidad*	*Probabilidad*	*Probabilidad Acumulada*
0	16	$\frac{16}{200}$	0.08	0.08
1	22	$\frac{22}{200}$	0.11	0.19
2	33	$\frac{33}{200}$	0.165	0355
3	45	$\frac{45}{200}$	0.225	0.58
4	31	$\frac{31}{200}$	0.155	0.735
5	20	$\frac{20}{200}$	0.1	0.835
6	12	$\frac{12}{200}$	0.06	0.895
7	9	$\frac{9}{200}$	0.045	0.94
8	7	$\frac{7}{200}$	0.035	0.975
9	5	$\frac{5}{200}$	0.025	1

Mónica: Es que yo también estoy de acuerdo contigo porque sí es cierto por ejemplo hasta aquí (*señala el renglón correspondiente a 7 hijos*) ya es casi seguro que les toquen, un 94% (*es la probabilidad acumulada hasta 7 hijos*) ya casi seguro que le toque a todos, más casi no estás ahorrando porque imagínate que te toque ya no éste (*señala el renglón de 4 hijos*) sino éste ya estás perdiendo (*señala el renglón correspondiente a 6 hijos*). Un escalón menos ya estás perdiendo, dos más, imagínate que te toqué éste (*refiriéndose a 4 hijos*) o éste (*refiriéndose a 2 hijos*), ya perdiste mucho más. Imagínate que, a lo mejor es una empresa chica y no sé, a lo mejor los boletos están en mil pesos, así que no es lo mismo gastar...
Mónica: Yo le pondría un 50% (*refiriéndose más bien al 58% de probabilidad de que el trabajador premiado tenga 3 hijos*), o si acaso un 70% (*refiriéndose al 73% de probabilidad de que el trabajador premiado tenga 4 hijos*), porque luego para que me salgan con que perdieron... aunque claro, maestro también está mal el problema porque no es cuestión de decirles oigan les vamos a dar boletos y si les toca pues que bueno porque es

como decirles a los que tengan más (*hijos*) llevan las de perder porque no vamos a comprar para todos y luego que... Lo ideal sería que si la empresa está haciendo la campaña, pero eso ya no tiene que ver con matemáticas, sino que es más ético, si la empresa está ofreciendo pues que lo ofrezca bien.
Brenda: En todo caso si se está buscando que no vaya a gastar de más y si alcanzan bueno y si no, no, pues yo estaría por 4 hijos.
Mónica: Ajá, yo también y si no alcanzan pues ya ni modo.
Profesor: Y si se pusieran codas, si se pusieran así estrictas, la lana de veras es que... vamos a ofrecer este beneficio pero... trabajadora social de veras trata de hacernos gastar lo menos posible, ¿cuál sería el valor que escogerían?
Brenda: En la codez, en la codez sería tres hijos.
Mónica: Yo también porque es donde hay más trabajadores.
Profesor: ¿Y en términos de probabilidades?
Brenda: Hasta aquí alcanzaría más de la mitad, estaríamos hablando de más de la mitad. Es decir tiene bastante probabilidad.
--- 2:54.6

Décimo segundo pasaje: *Sobre su percepción y aceptación del azar.*

ByM04 (video)--- 9:57.1
Profesor: Supongamos que yo soy el patrón y ustedes son las trabajadoras sociales y me dicen 'te convienen cinco boletos' y me van a convencer: 'cinco boletos es mucho, ¿por qué tanto?', ustedes ¿qué argumento me darían?
Brenda: Bueno con cinco boletos tendría asegurado que la mayor parte de… o sea más la mitad de la probabilidad de que la familia que resultara seleccionada complete con esos boletos. Abarcaría a la mayor parte de sus trabajadores.
Profesor: Oigan pero eso es posible, a lo mejor eso ni sucede porque es azar…
Mónica: Pues es lo que yo decía…
Profesor: …acuérdense que será una ruleta o una tómbola en la que echaremos los boletos y es azar y pues, quién sabe si...
Mónica: Pero por la misma razón que es azar igual pueden salir estos (*señala en el pizarrón los renglones referidos a los trabajadores que tendrían de tres a nueve hijos*).
Profesor: Claro, como es azar, pueden salir nueve.
Mónica: Bueno, lo que diríamos aquí lo que le aseguramos es el 50%, el otro 50% definitivamente va a tener que comprar uno, dos, tres, cuatro boletos por su cuenta, pero lo que usted quiere también es congraciarse o llevarse bien con los trabajadores, entonces pues con más del 50% ya tiene probabilidad de que lo saque, es decir, tiene asegurada la mitad.
Profesor: Pero yo sigo así como renuente, porque es azaroso, como que es la tómbola, como están tan seguras de esos números, porqué le creen a esos números si es todo azar, ¿es posible que en medio de tanto azar ustedes, que nadie sabe que va a salir, ustedes puedan asegurarme qué va a salir?
Mónica: No, no le podemos asegurar nada. Yo digo que en los azares no se puede asegurar nada.
Brenda: Bueno, es que no se puede asegurar, pero de lo que se habla aquí es de probabilidad, lo que puede pasar. Como además, se encuentran más concentrados los trabajadores en esa parte (*señala los renglones de la tabla correspondientes a trabajadores con 3 hijos o menos*) hay más posibilidades de que no vaya a perder tanto porque más del

50% (*de los trabajadores*) se juntan en ese espacio. No podemos asegurarlo pero es más factible, puede que suceda más, que salga más una de estas personas (*señala los renglones correspondientes a los trabajadores con 3 hijos o menos*), es más fácil, que del resto porque son más, son mayoría.
Profesor: O sea que ustedes me están diciendo que uno puede predecir qué boleto va a salir.
Mónica: Cuál es más probable pero no cuál va a salir. No hay manera.
Profesor: Pero existirá alguna manera, ¿es ignorancia nuestra el no saber qué boleto va a salir?
Brenda: No es algo que tú puedas controlar, es la suerte.
-- 13:50.6

Décimo tercer pasaje: *Sobre el espacio muestral dentro del contexto del problema.*

ByM04 (video)--- 15:13.2
Profesor: ¿Cuántos casos posibles pueden haber de hijos? Puede ser que en una de esas salga una familia de 11 hijos.
Mónica: No, 9.
Profesor: ¿y mínimo?
Brenda: ¿Mínimo? Cero hijos.
Profesor: ¿es posible que pueda extenderse el número 200 a más valores? ¿por qué?
Mónica: No porque sólo son 200 trabajadores.
Profesor: ¿o sea que la probabilidad sólo tiene que ver con esos trabajadores o con otras cosas ahí?
Mónica: No, nada más con esos trabajadores.
-- 16:00.5

TERCERA PARTE

Décimo cuarto pasaje: *En la elaboración de la gráfica.*

Mónica tiene el gis, ella hace la gráfica mientras Brenda le va dando indicaciones, de manera que lo que Mónica va haciendo está consensuado por las dos.
Mónica hace unos ejes cartesianos en el pizarrón. El sistema coordenado sólo incluye el primer cuadrante.
ByM04 (video)--- 16:48.0
Brenda: Aquí sería el número de hijos (*señala el eje de las abscisas*).
Mónica etiqueta los ejes: número de hijos y número de trabajadores.
Brenda: No, pero entonces ahí sería la probabilidad (*se refiere al nombre de la variable dependiente*).
Mónica: ¡Ah! Aquí ya podemos saber que en vez de poner número de trabajadores, ya vamos a poner directamente la probabilidad. Vamos a trabajar con la probabilidad directamente.
Brenda: Aquí (*señala el eje de las abscisas*) sería, uno, dos,… así hasta nueve.
Mónica pone la escala en el eje de las abscisas.
Brenda: Luego para arriba, ¿de qué la empezamos? ¿de 0.1, no?

<u>Brenda</u>: Bueno, sería de cero hasta uno y luego ya sería irle poniendo las probabilidades. A ver, divídelo…
<u>Mónica</u>: A ver, primero mitad y mitad. Punto cinco…
Brenda dicta los valores a Mónica y ella los va colocando en la gráfica.
<u>Mónica</u>: ¿Hacemos la gráfica?
<u>Brenda</u>: Sí, hay que hacer la gráfica.
Brenda une los puntos.

En el pizarrón:

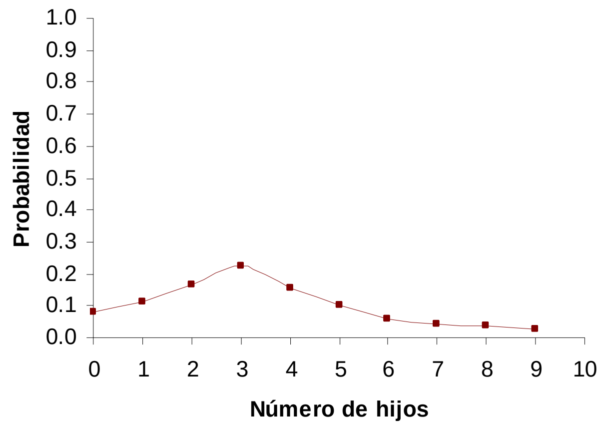


Figura 16.

ByM05 (video) -- 05:36.8

Décimo quinto pasaje: *Sobre la noción de función.*

ByM05 (video) -- 05:36.8
<u>Profesor</u>: Representa con una letra la variable dependiente y con otra la variable independiente.
<u>Mónica</u>: ‘x’ y ‘y’ (*las escribe en los ejes de la gráfica que está en el pizarrón*). Ya.

En el pizarrón:

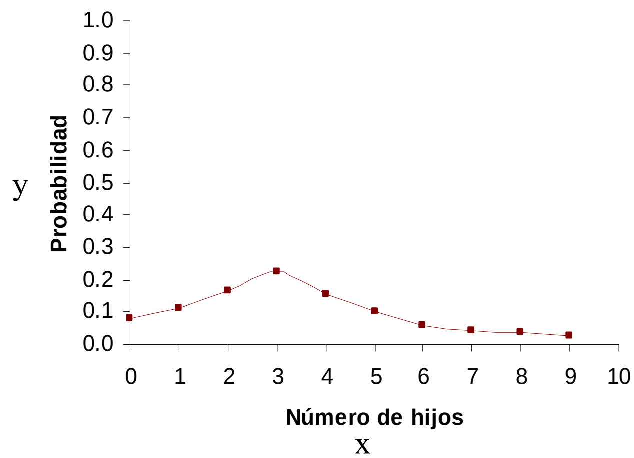


Figura 17.

Profesor: Utiliza estas letras para representar en notación funcional el resultado obtenido en el inciso b) de la parte 2. Es decir con que probabilidad el trabajador premiado ocuparía exactamente el número de boletos que compró la trabajadora social.
Dudan, leen una y otra vez la pregunta.
Mónica: Sería esté no. (*Señala el punto máximo en la gráfica* y *escribe en el pizarrón y = 0.225*), ¿No está bien?

En el pizarrón:

y = 0.225

Figura 18.

Brenda: (*dirigiéndose a Mónica*) ¿A qué se refiere con notación funcional? ¿una ecuación?
Dudan..
Mónica: ¡Funciones!, F de x, f de…
Profesor: ¿Sería posible eso (*se refiere a la gráfica que está en el pizarrón*) escribirlo en notación de funciones? ¿qué es una función? Esa gráfica, suponiendo que fuera continua, ¿sería una función?
Mónica: Sí.
Profesor: ¿Por qué?
Mónica: Pues como todas (*se refiere a los valores involucrados*) se corresponden, por ejemplo f de 3 que corresponde a 0.225, ¿no? (*Escribe en el pizarrón F(3) = 0.225*)

En el pizarrón:

y = 0.225 F(3) = 0.225

Figura 19.

Profesor: ¿Qué dice Brenda?
Brenda: Pero lo que pide es poner… ¿cómo es la ecuación?
Blanca: ¿Una función nada más se expresa en forma de ecuación?
Brenda: No.
Blanca: ¿De qué otra forma?
Mónica: Por ejemplo. No, una función no necesariamente tiene que ser una ecuación, por ejemplo está podría ser una función (*escribe F(x) = 3x*), siendo que no es una ecuación.

En el pizarrón:

y = 0.225 F(3) = 0.225 F(x) = 3x

Figura 20.

Profesor: Pero esa sería una expresión algebraica, ¿puede algo que no sea una expresión algebraica, una gráfica o una tabla, ser función? ¿tiene a fuerzas que ser una fórmula?
Mónica: La tabla o la gráfica es la representación de una función.
Profesor: ¿Qué dice Brenda?
Brenda: No, porque pueden variar los datos o que estén dados.
Mónica: De hecho la fórmula se saca de la tabla y la gráfica también.

<u>Profesor</u>: ¿Qué opinas de eso que escribió Mónica de F(3) = 0.225? ¿Eso sería una notación funcional?
<u>Brenda</u>: Bueno pues es que… que aquí dice que nombremos las variables y utilice estas las letras para escribir en notación funcional el resultado obtenido, o sea que tendríamos que utilizar 'x' y 'y', puesto que esas fueron las letras que escogimos.
<u>Profesor</u>: ¿Quién sería 'x'?
<u>Brenda</u>: En este caso, pues sería tres.
<u>Profesor</u>: Pregúntale a Mónica, ¿quién sería 'x' y quien sería 'y'?
<u>Mónica</u>: Pues… 'x' sería 3 y 'y' sería 0.225. La función de 'x' es 'y', es decir el resultado.
<u>Brenda</u>: Pero en todo caso… no lo sabes. Lo que yo digo es utilizar…
<u>Mónica</u>: ¿En vez de F, 'y', aquí (*señala la F en F(x)=0.225*)?
<u>Brenda</u>: No, es que de cualquier forma haría falta especificar que es una función de 'y'. Lo que podríamos utilizar es, para tomar exactamente las mismas letras, ponerle allá F de 'x' (*se refiere al nombre del eje de las ordenadas*). Es decir, cambiar aquella, ponerle F(x) en lugar de 'y'.
<u>Mónica</u>: Pero necesitamos ponerle F porque es lo que dice ella (*se refiere a Blanca*) que hace falta ponerle la función.
<u>Brenda</u>: Por eso te digo, mejor cambiar aquella (*señala la y en el eje de las ordenadas de la gráfica*).
<u>Mónica</u>: (*Borra la y en el eje de las ordenadas escribe F(x)*) Así nos queda una sola variable.

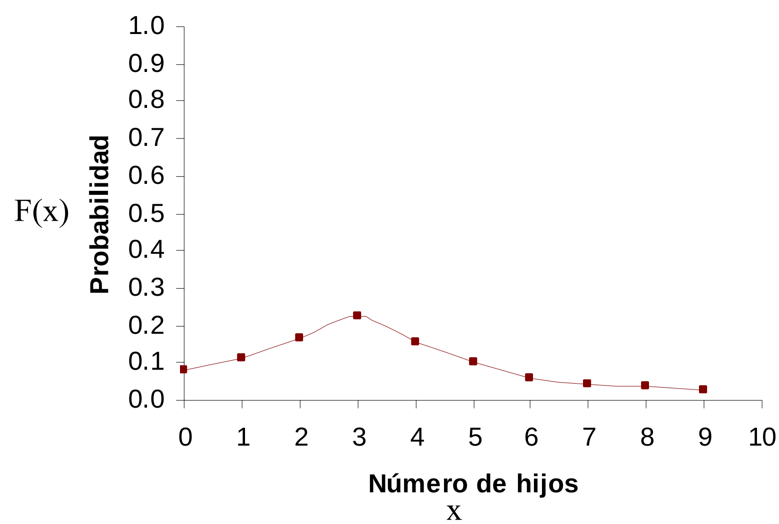


Figura 21.

ByM06 (video) -- 02.50.4

Décimo sexto pasaje: *Definición del dominio de la función.*

ByM06 (video) --- 03.03.7
<u>Profesor</u>: Describe en el contexto del problema, con tus propias palabras, la información que proporciona la variable dependiente de la variable independiente.
<u>Mónica</u>: ¿La variable dependiente (*señala en el pizarrón el eje de las ordenadas*) de la… (*Señala en el pizarrón el eje de las abscisas*)?

Brenda: Bueno pues se supone que esta (*señala en el pizarrón el eje de las ordenadas*) te está diciendo que probabilidad hay de que… cuál es la probabilidad de la variable independiente.
Mónica: Lo que dice es que todos estos números (*señala la escala del eje de las abscisas*), ninguno pasa de esto (*señala el punto máximo de la gráfica*). Es decir, ninguno va más para arriba.
Brenda: Pues básicamente la variable dependiente te está diciendo cuál es la probabilidad.
Profesor: ¿Llegan a un consenso?
Ambas: Sí
Profesor: Bueno, la que sigue. Describe el dominio de la variable independiente y el rango de la variable dependiente.
Mónica: Pues de 0 a 9 es el dominio de la independiente y el rango de la dependiente sería de 0.025 a 0.225 (*escribe en el pizarrón [0,9] y [0.025, 0.225]*).

En el pizarrón:

[0,9] [0.025,0.225]

Figura 22.

Blanca: ¿Por qué lo encerraron en un corchete? ¿qué significa eso?
Mónica: Que va de… incluye del 0 todos los números, todos los números, hasta el 9. Cerrado.
Blanca: ¿Todos los números?
Mónica: Entre el cero y el nueve, sí.
Blanca: ¿Todos los números que hay entre cero y nueve?
Mónica: Sí
Blanca: Cuatro punto cinco está cero y nueve.
Mónica: Pues también es parte porque... No está en la gráfica, pero está en el dominio. Igual, esto (*señala el 4.5 en el eje de las abscisas y traza con el lápiz su proyección hacia la gráfica y después hacia el eje de las ordenadas*) no tiene un valor escrito pero sí existe.
Blanca: ¿Si podemos calcularlo?
Mónica: Sí.
Brenda: Yo diría que no, se supone que estamos…
Mónica: Pero es que, es que… viéndolo así para no poner cuatro punto cinco hijos… pues es de que... sí son valores enteros, pero es que en las gráficas también te sale, por ejemplo población 1970.3, ni modo que haya 0.3 de humano, pero sí se puede calcular, es como un valor de…
Brenda: Bueno es que sí es cierto, se supone que nada más serían los que te están dando. Las coordenadas que te está dando la gráfica únicamente son estos (*señala los puntos de la gráfica*) no son mitades ni nada de eso, sólo los enteros.
Mónica: Entonces para qué es la gráfica si vamos a dejar los puntos así aislados.
Brenda: Sí, pero es que realmente no puedes sacar la probabilidad de 0.5.
Mónica: Sí, no puedes tener 8.5 hijos, es lo que yo digo.
Brenda: Pero por ejemplo, o sea, o sea, la gráfica sí está abarcando todo, sí está abarcando todo este espacio.
Blanca: ¿Entonces están de acuerdo en que ese es el dominio?
Ambas: Sí.

<u>Profesor</u>: ¿Y el rango? ¿Cómo lo encontraron?
<u>Mónica</u>: Igual, vimos cual es el valor más chiquito y cuál era el más grande y no puede salirse de ahí, ni para abajo ni para arriba.
<u>Profesor</u>: ¿Qué significa el 4, 5, 6…?
<u>Ambas:</u> El número de hijos.
<u>Profesor</u>: Entonces no tienen sentido los decimales.
<u>Ambas</u>: No.
<u>Blanca</u>: Pero a mí me acaban de decir que sí tenían sentido.
<u>Brenda</u>: No, lo que pasa es que no podemos saber cuál sería el punto, la coordenada aquí (*señala un punto que está entre 0 y 1 en la gráfica*) en este punto sería decimal, pero en este caso uno no usa los decimales, no sé la probabilidad de la mitad de un hijo.
<u>Mónica</u>: Si fuera dinero sí se podría saber.
<u>Brenda</u>: Sí podemos usar los decimales, porque entran dentro de la misma gráfica, pero en ciertas cosas no usas decimales.
<u>Profesor</u>: En cambio en el eje vertical ¿si tienen sentido los decimales?
<u>Brenda</u>: Es la probabilidad.
<u>Profesor</u>: Por ejemplo, puede haber 0.255, o sea valores intermedios.
<u>Ambas:</u> Sí.
<u>Profesor</u>: Ahí sí se puede escribir como intervalo cerrado. El que está dando lata es el de arriba (*se refiere al orden en que escribieron ambos intervalos en el pizarrón, el de arriba es el dominio, el de abajo es el rango*), el de abajo como que no hay discusión, el de arriba con la salvedad que dicen ustedes. "bueno, así se suele manejar". Ese sería el argumento, ¿si?
<u>Ambas:</u> Sí.
-- 10.57.8

Décimo séptimo pasaje: *Definición de función en la relación establecida.*

ByM06 (video)--- 10:57.0
<u>Profesor</u>: Describe el método para encontrar el valor de la variable dependiente a partir de la variable independiente.
<u>Mónica</u>: ¿La ecuación?
<u>Profesor</u>: La F, es decir la variable dependiente, a partir del valor de la variable independiente.
<u>Mónica</u>: Pues fijarse en la gráfica o sea checar.
<u>Brenda</u>: ¿Con un método más allá del gráfico?
<u>Profesor</u>: No dice un método más allá del gráfico. ¿Qué opinas de la respuesta de Mónica?
<u>Blanca</u>: Sí, por ejemplo si 'x' es igual a dos, ¿cuánto vale la variable independiente?
<u>Mónica</u>: Pues se vería aquí, (*se dirige a la gráfica y señala el dos en el eje de las abscisas, la proyección a la gráfica y posteriormente al eje de las ordenadas*) o sea dos y checar el resultado aquí, que es el mismo que de la tabla.
<u>Blanca</u>: ¿Lo sacarías de ahí o de la tabla?
<u>Mónica</u>: Es lo mismo. Bueno, si estuviera esto (*señala el eje de las ordenadas*) perfectamente numerado, no así tan… sería lo mismo.
<u>Brenda</u>: Es decir si se pudiera tener mayor precisión, pues sí, aquí pues está un poco más complicado.
<u>Mónica</u>: Ajá, si tuviera cuadrícula, pues si, la tabla y la gráfica es lo mismo.

Silencio
Mónica: Sin embargo estaba pensando ahorita ¿Cómo le podemos hacer? Lo primero que pensé fue regresión, hacer una regresión y encontrar una ecuación, pero después dije no, porque no tiene nada que ver el número de hijos con las personas, no es de que entre más hijos más personas haya sino de que pueda haber… si va disminuyendo, no va a llegar un momento en que se va a acabar, porque son hijos si fueran otros datos como dinero o alguna otra variable, puede que sí, pero como hijos no, pero como en, no sé, 15 ya se va a acabar, a partir de 3 va a ir disminuyendo.
Profesor: ¿La relación que establecieron es una función matemática? ¿o sea es función? ¿Recuerdan las propiedades de la función? ¿qué es una función?
Mónica: Que hubiera dos variables.
Brenda: Que para el valor de 'x' haya un valor en 'y'.
Profesor: ¿Y puede haber dos 'x' que vayan a dar a un mismo valor de 'y'?
Ambas: Sí
Profesor: ¿Puede haber una 'x' que pueda dar dos valores de y? Es decir, ¿un número de hijos que de dos probabilidades diferentes?
Ambas: No.
Profesor: Entonces ¿es función o no?
Brenda: Sí, si es función porque para cada número de hijos hay una sola probabilidad.
-- 15:24.3

Décimo octavo pasaje: *La continuidad como principal diferencia entre la variable independiente usada en el contexto manejado en esta actividad y las variables usadas en sus cursos normales de matemáticas.*

ByM06 (video)--- 15:25.0
Profesor: ¿Qué diferencia encuentras entre la variable independiente de esta función de probabilidad y las variables independientes de las otras matemáticas que tú conoces, por ejemplo del álgebra y del cálculo?
Mónica: Qué es limitada.
Profesor: Supongamos que tiene un sentido "más hijos" indefinida, aun así ¿Hay más características? ¿Es la única característica que ustedes encuentran entre esta función y las otras variables independientes de álgebra y cálculo?
Mónica: Lo de los decimales.
Profesor: Que se suele manejar variable continúa.
Mónica: Enteros.
Profesor: Aquí enteros. ¿Te pareció que esa es una variable discreta? A esa se le llama variable discreta, cuando nada más puede tomar valores enteros.
Mónica: Ah, pues es así, discreta.
-- 16:50.0

Décimo noveno pasaje: *El problema de obtener una ecuación a partir de los datos establecidos como otra diferencia entre las funciones que han trabajado en anteriores cursos de matemáticas y el contexto de esta actividad.*

-- 16:50.0

Profesor: Qué más, piensen ¿Habrá alguna característica más que distinga esta variable de las otras que trabajan en matemáticas?
Brenda: Normalmente en álgebra manejamos ecuaciones. Todas las gráficas se manejan por ecuaciones, todos los valores de'y'se van a dar por la ecuación. Aquí los valores de y también se dan según una 'x', pero no tienen una ecuación, son así como… aleatorios. No llevan un patrón así seguido.
Mónica: Es lo que decía, no podemos sacar una ecuación porque no nos va a dar nada preciso o sea de número de hijos no se puede sacar una ecuación.
Brenda: Bueno más que la variable independiente diferente, como que es la variable dependiente la diferente, porque no podemos establecer cuál será la dependiente. A diferencia del álgebra, con la independiente podías sacar la dependiente exactamente y en este caso no, son datos que te están dando y tú no puedes... es muy aleatorio, es más difícil que con álgebra. Con álgebra tú podías sacar 'y' con la ecuación y aquí no.
ByM06 (video)-- 10:57.0
En audio-- Segundo casete
Mónica: Aunque sí se observa que primero crece y luego ya disminuye, disminuye y como son hijos y como somos humanos son funciones más o menos, o sea va a ir disminuye y disminuye hasta que llegue a 12 y ya. En ese caso pero por ejemplo más datos humanos como… número de personas divorciadas o cuántos se han casado, ahí no sería viable sacar una ecuación porque no te iba a decir nada.
Brenda: Bueno en realidad sí puedes sacar una ecuación, pero eso no te daría… no sería muy exacto. Esa es la diferencia, que no te pueden dar datos exactos.
--- Segundo casete

Vigésimo pasaje: *Sobre el valor esperado del número de hijos al realizar la rifa de los boletos*.

En audio-- Segundo casete
Profesor: Sí tomo la tómbola y le doy vueltas, ¿cuál sería el número de hijos que creen que tanga el trabajador seleccionado?
Brenda: Tres hijos.
Profesor: ¿Por qué?
Mónica: Es el de mayor incidencia.
Profesor: Supongamos que sale un trabajador que tiene dos hijos, si salió dos, la siguiente salida de la tómbola ¿cuántos hijos creen ustedes que saldría?
Mónica: Tres.
Profesor: Supongamos que sale tres otra vez y volvemos a sacar otro papelito, ¿cuántos hijos tendría el trabajador ahora?
Brenda: Pues podría salir dos o cuatro.
Profesor: Vamos a suponer que sale 4, y volvemos a seleccionar un papelito, ¿cuánto creen ustedes que salga?
Mónica: Pues es que nomás está entre esos números. Puede salir cualquier número, pero las mayores probabilidades están entre esos números. Pero ya después de varias veces se van a acabar los 3.
Profesor: No, siempre estamos regresando el papelito que sale.
Mónica: Ah, pues tres siempre, es el que se espera.

Profesor: Pero si saco uno al azar y sale que el trabador tiene un hijo. Lo regreso y vuelvo a revolver, ¿cuál es el valor esperado?
Mónica: Pues tres maestro, ¿es que por qué no?
Profesor: Y bueno, si sacamos uno más, ¿podemos saber cuál es el que le sigue? ¿Alguna forma de conocer cuál va a salir antes de sacarlo?
Ambas: No.
Profesor: ¿Es posible que vuelva a salir el uno?
Brenda: Es poco probable, es más probable que salga entre el 2 y el 4 a que vuelva a salir el uno.
Mónica: Más sí puede volver a salir.
Profesor: Entonces ¿habrá orden de aparición en el número que salga de la tómbola?
Brenda: O sea de que primero saque uno y salga uno, después vuelva a sacar y salga dos y así, pues no, no lo hay.
--- Segundo casete

Vigésimo primero pasaje: *La variable dependiente como la principal diferencia entre las funciones que han trabajado en anteriores cursos de matemáticas y el contexto de esta actividad.*

En audio-- Segundo casete
Blanca: Comparen con los modelos que hayan trabajado en álgebra o cálculo. Por ejemplo, que ustedes van en un vehículo con una velocidad constante y ustedes quisieran tratar de describir esa situación, ¿cuáles serían las dos variables que manejarían?
Brenda: Sería el tiempo y la distancia.
Blanca: ¿Cuál sería la variable dependiente ahí?
Brenda: La dependiente es la distancia y la independiente el tiempo.
Blanca: Y díganme, si nosotros estamos ahorita en el tiempo uno, después ¿en cuál estaría?
Ambas: En el dos.
Blanca: ¿Y después?
Ambas: En el tres.
Blanca: ¿Por qué no pasa lo mismo aquí?
Mónica: Porque es azar.
Brenda. Ajá.
Profesor: La pregunta inicial era qué diferencia había entre esta variable y las variables como las que menciona Blanca.
Mónica: Las otras son constantes, continúas, se espera que sigan y que vayan, en cambio esta no sabes, no puedes predecir algo.
Brenda: Sí, a lo mejor hay personas, como dices, bueno, voy al tiempo uno, dos y tres, voy una hora, una hora, pero a lo mejor… bueno en este caso sí va así seguido, pero a lo mejor hay casos en que no había nadie con 6 hijos entonces la variable ya no tendría que ser toda seguida.
Mónica: Pues obviamente sí se pueden ordenar siempre las variables (*se refiere a los valores que puede tomar la variable*), así de menor a mayor y de mayor a menor también, pero eso no significa que de mayor a menor van a ir creciendo o van a ir disminuyendo igual, o sea, el que sea de menor a mayor, el que estén ordenadas, no significa que también estén creciendo uniformemente.

Blanca: Y entonces ¿está variable independiente es igual a las otras variables que ustedes han manejado?
Mónica: No.
--- En audio
ByM07 (video) -- 00:00.0
Profesor: Perdón, pero, ¿podrían sintetizar alguna idea de lo que están discutiendo hasta ahora?
Brenda: Yo decía que si lo comparamos con el ejemplo del carro, ahí la 'x' siempre va a ser continua, tiene que haber uno y luego dos, o sea siempre una hora va a seguir a otra hora y así, y aquí por ejemplo, podría haberse dado el caso de que a lo mejor no había hijos, no había una persona con cierto número de hijos. Como no está que tiene que ser así, como es muy aleatorio todo esto, no tiene porque ser así, creciendo igual.
Mónica: Lo que yo dije es que la variable independiente está ordenada y se pueden ordenar tanto la dependiente, como la independiente, pero en otro tipo de gráficas normales a cada valor en 'x' corresponde un valor en y, es decir, tiene un comportamiento esperado, dependiendo de la gráfica, no nada más gráfica recta, que puede ser logarítmica o exponencial o polinómica, cualquier tipo de gráfica, pero para cualquier valor en 'x' se espera uno en 'y'. Aquí no, para un valor en x, no sabes cuál va a ser el valor en 'y', puede ser más grande, más chico, es algo que no está controlado. La 'x' no controla a la 'y'.
ByM07 (video) -- 02:34.2
ByM08 (video) -- 00:00.0
Blanca: Bueno, estábamos en si era diferente o no era diferente y si es diferente ¿en qué?
Brenda: Bueno yo digo que básicamente la diferencia, lo mismo que decía ella, pero más en todo caso, más que nada la diferencia radica en la variable dependiente porque, en las funciones matemáticas o algebraicas que hemos visto, con la ecuación podemos saber el valor de la variable dependiente, pero con esta varía mucho. Básicamente la que varía más es la dependiente.
Mónica: Yo también estoy de acuerdo con ella en que la que varia es la variable dependiente. Varía de forma inconstante o impredecible.
Blanca: Pero ¿por qué impredecible?
Brenda: Sí, porque este valor no es algo que esté dado por una…
Mónica: Función.
Brenda: Sí, por una función, por una ecuación, sino que porque son datos que así se dan o sea que, que pueden ser más o que pueden ser menos, que nos están determinados por algo sino por las circunstancias, por el azar yo creo.
Mónica: Ajá, porque pueden estar determinados por diferentes factores por ejemplo, esa población puede haber tenido tres hijos porque se localizaba en México o en Estados Unidos en donde la mayoría de las personas tienen tres hijos, pero en poblaciones de otros lugares donde el número de hijos crece por persona, la gráfica ya no iba a ser igual. En cambio si hablamos de dinero, bueno, no tanto de dinero sino de otro tipo de cosas como habíamos visto anteriormente, sí se espera algo, pero aquí el número de hijos… En este tipo de gráficas en donde están midiendo el azar, no podemos saber cuándo va a bajar y cuando va a subir, no tiene un comportamiento predestinado.
Blanca: ¿La variable dependiente? ¿la probabilidad?
Ambas: Sí.
--- 02:07.3
ByM09 (video) -- 00:00.0

Blanca: O sea lo que ustedes piensan es que lo que hace diferente a esta función del resto es la variable dependiente y no la independiente.
Brenda: Sí, yo digo que esa sería la gran diferencia. Porque igual por ejemplo, aquí (*señala un punto sobre el eje de las abscisas*) puedes no tener un dato, pero esto (*señala la altura de ese punto sobre el eje de las abscisas*) es lo que siempre va a estar variando. En las otras tienes que seguir cierto orden y a lo mejor aquí (*señala el eje de las abscisas*) te puedes brincar un dato, pero esto no lo puedes tener determinar.
Mónica: Cualquier expresión, cualquier ecuación se esperaría que tuviera cierto comportamiento, por ejemplo, si por 'x' o 'y' razón no hubiera familias con 3 hijos, ahí iría para abajo otra vez la gráfica y no por eso estaría mal la gráfica.
Brenda: Es como lo que estaba preguntando ahorita el profe, que si podemos saber exactamente qué número va a salir, pues no, no se puede saber. Y toda la situación es de que no puedes saber cuándo va a salir, a diferencia de la matemática que nos decía, sí podemos saber cuánto se va a recorrer en cierto tiempo y en esta no, no puedes decir: "va a salir este número".
Mónica: Ajá.
Brenda: Como que más bien es una aproximación, no es más… es probabilidad.
--- 02:31.0

Vigésimo segundo pasaje: *Poniéndole nombre a la variable independiente, lo que permite resumir las características importantes de esa variable.*

ByM09 (video)--- 02:31.4
Profesor: Como ven esta variable del eje de las 'x' es muy distinta a la que están acostumbradas a trabajar, entonces, ¿qué nombre le darían? Si es distinta, pues vale la pena ponerle un nombre a la variable.
Brenda: ¿o sea a la x?
Profesor: Porque la y no hay problema ¿no? ¿cómo se llama la y?
Mónica: Probabilidad.
Profesor: Entonces ya nada más nos queda la x, ¿no?
Mónica: Estadística. Probabilidad y Estadística.
Risas
Mónica: Pues sí, es que… datos.
Blanca: Datos ¿Variable datos?
--- 03:03.3
ByM010 (video) --- 00:00.0
Mónica: Inconstante.
Blanca: ¿Por qué inconstante?
Mónica: No, impredecible.
Brenda: No, la impredecible es y.
Mónica: No, pero también es x.
Blanca: ¿Por qué esta también es impredecible?
Mónica: No, inexacta… ¡Porque! Por… porque a lo mejor podemos quitar el dos, a lo mejor no hay familias con dos hijos…. Incontrolable, no.
Profesor: ¿De dónde provendría este valor de las x?
Brenda: ¿De la x? En este caso es el número de hijos
Mónica: Pues del número de trabajadores.

Profesor: Pero ya sabemos el número de trabajadores, ya sabemos el número de hijos. Pero se acuerdan que se trataba de hacer un sorteo y el sorteo tenía una tómbola y de la tómbola sale un trabajador seleccionado con cierto número de hijos. Se hacía un experimento. Bueno si eso es así, ¿dónde estaría esa x, en los trabajadores, en la tómbola, en los boletos? ¿dónde quedó la x?
Mónica: En los boletos. No, en lo que hay adentro de la tómbola, en las bolitas.
Brenda: No, porque adentro de la tómbola están los trabajadores. Es que se supone que del número de hijos… por decir…¿cómo lo explicaré?
Mónica: O sea no está el número de hijos, de que uno, dos, sino que está un trabajador y tú, a ver cuántos hijos tienes. No pues que dos. Está implícito adentro del trabajador.
Brenda: Sí, ándale. O sea no hay papelitos que digan uno, dos, tres, …
Mónica: No, sino que sale, Juan Ramírez y tú, a ver, cuántos hijos tienes, no pues que tres, a corresponde a… (*señala el eje de las ordenadas en la gráfica*) y ya.
Brenda: O sea que dentro de los papelitos que están en la tómbola, están las posibilidades o sea estos (*señala la escala del eje de las abscisas*).
Mónica: …están los trabajadores y el número de hijos es una característica de ellos.
Profesor: Entonces esa ‘x’ es obtenida de la tómbola, ¿es cierto eso?
Mónica: Pero no directamente ¿verdad? Indirectamente, sí. Hay que juntarlo con un trabajador.
Profesor: Y entonces ahí sale, y si en el primer boleto salió un trabajador, vamos a suponer, con dos hijos, ¿se puede saber en el siguiente boleto el número de hijos que va a salir?
Mónica: No. Ni siquiera podemos saber qué trabajador va a salir.
Profesor: Entonces ¿hay orden de aparición en esos números?
Mónica: No.
Profesor: Entonces ¿por qué los pusieron así seguidos, uno, dos, tres, cuatro,…?
Brenda: Pues porque es un orden y así va la numeración, pero orden pues así como la mayoría...
Mónica: Pudimos haberlos ordenado, primero tres, luego cuatro y así, y luego iba a salir una gráfica así (*dibuja en el aire una gráfica decreciente*) porque van por orden en que salieron. Cómo se esperaría que salieran porque se espera que salgan muchas veces el tres, pero no sale, igual sale más veces otro, pero se espera que salga más veces el tres.
Profesor: ¿Es posible, de veras, que a la mera hora, supongamos que se sacan varios boletos, resulte que salen varios nueves?
Mónica: Sí, es poco probable, pero sí es posible. Es como lo de los boletos de las casas, a lo mejor no tiene nada que ver, pero por ejemplo de que se va a rifar una botella y una sola persona compra muchos boletos y uno dice, pues se espera que gane, pero puede no ganar.
Brenda: Sí.
Blanca: Entonces, ¿cómo llamarían a esa variable?
Mónica: Suerte.
-- 05:43.4
ByM11 (video)-- 00:00.0
Brenda: La implícita.
Mónica: Sí.
Brenda: Por que está dentro de la tómbola. Porque está dentro de la tómbola, lo que estábamos explicando ahorita, no está dentro de los papelitos. O ¿qué otro nombre le ponemos?

<u>Mónica</u>: Implícita determinada. Implícita por lo que ya quedamos y determinada porque está determinada por el trabajador. Está implícita por el trabajador y determinada por el trabajador. ¿Te parece buen nombre?
<u>Brenda</u>: Yo creo que sí.
--- 01:42.6