

Didáctica de la Probabilidad y Estadística

El Caso de la Variable Aleatoria

Blanca R. Ruiz y José Armando Albert

ITESM, Campus Monterrey
México

bruiz@itesm.mx

Probabilidad, Estadística y Combinatoria – Nivel Superior

Resumen

Nuestro proyecto aborda una de las ideas fundamentales en los cursos de Probabilidad y Estadística en las instituciones de enseñanza universitaria: la variable aleatoria. Este concepto se apoya en muchos otros conceptos matemáticos y probabilísticos y sirve a su vez de soporte a gran parte de los temas en probabilidad y estadística. El proyecto didáctica de la variable aleatoria se fundamenta en la Teoría de Situaciones Didácticas y en este artículo profundizaremos en una exploración cognitiva como complemento del análisis preliminar que propone la metodología de la Ingeniería Didáctica. En esta primera etapa se buscará saber cuál es el estado de apropiación de algunas ideas fundamentales estocásticas relativas a la variable aleatoria en dos estudiantes que se acaban de integrar al nivel universitario.

1. Introducción

Existe un rezago en las instituciones educativas en comparación con la demanda que las sociedades exigen del uso de la estadística a sus ciudadanos y profesionistas. Se hace necesario abordar la enseñanza de la probabilidad y la estadística desde una perspectiva científica para trazar líneas que permitan abordar este tema en nuestras instituciones educativas. Dentro de estos esfuerzos, Heitele (1975) propone una lista de diez ideas fundamentales en la enseñanza de la estadística basándose en un enfoque epistemológico-pragmático, en donde una idea fundamental proporciona modelos explicativos en cada etapa del desarrollo del concepto a lo largo de la educación del individuo. La variable aleatoria está incluida dentro de esas diez ideas fundamentales a desarrollarse en los estudiantes desde niveles educativos básicos y que culminen en niveles superiores.

La pertinencia del desarrollo de un proyecto de investigación alrededor de la didáctica de la variable aleatoria también se sustenta en otras razones, tanto en dificultades en su didáctica y en su aprendizaje, como a razones propias del desarrollo del concepto en la probabilidad y la estadística como ciencias. El concepto de variable aleatoria propicia el paso de la estadística descriptiva y probabilidad básica hacia modelos probabilísticos, por lo tanto por un lado en ella convergen muchos conceptos estadísticos y probabilísticos provenientes de los niveles educativos básico y medio superior, y por otro constituye la unión principal entre el pensamiento probabilístico y el pensamiento estadístico. Así mismo, está presente en las ideas de distribución de probabilidad, al Teorema del Límite Central, muestreo, pruebas de hipótesis, correlación y asociación, etc. Por lo tanto, una incomprensión de ese concepto puede ocasionar una mala percepción de temas posteriores, y también es un indicador del aprendizaje de elementos estudiados anteriormente.

2. Metodología

Nuestro proyecto de investigación se circunscribe fundamentalmente en la Teoría de Situaciones Didácticas, iniciada por Brousseau (1986), y a su metodología, la Ingeniería Didáctica, reportada por Artigue M. (1995), aunque también se recurre a otras aportaciones a modo de herramientas metodológicas, tal como el recurso de las entrevistas. Nos restringimos a la didáctica de la variable aleatoria en el nivel superior, puesto que es en donde realizamos nuestra labor como profesores.

En este reporte se presenta el desarrollo del análisis preliminar con sus tres componentes: epistemológica, cognitiva y didáctica, así como una *exploración cognitiva*, que se realizó con la finalidad de complementar los elementos de la componente cognitiva que justifiquen un primer diseño de actividad para el salón de clase y sus hipótesis preliminares, que realizaremos en futuras investigaciones.

La exploración cognitiva consiste en una serie de entrevistas a un par de estudiantes a las que se les plantea un problema y lo resuelven a través de la discusión suscitada entre ellas hasta lograr un acuerdo en presencia de un investigador, cuyo único papel es aclarar o profundizar, a sí mismo o a la investigación, las ideas que las estudiantes expresan en voz alta. Las preguntas dirigidas a las estudiantes están condicionadas por la solución que van dando al problema como por una guía que el investigador preparó previamente.

3. Componente epistemológica

En un primer acercamiento de análisis preliminar, la componente epistemológica puede ser vista, a su vez, como disciplinar, histórica y social (Albert, 1998).

La idea de variable aleatoria ha sido causa de las múltiples aplicaciones actuales del cálculo de probabilidades, ya que pasó de ocuparse del estudio de la probabilidad de sucesos aislados al estudio de las distribuciones de probabilidad, y posteriormente al de los procesos estocásticos. La variable aleatoria y su distribución, así como el estudio de las familias de distribuciones y sus propiedades son una herramienta muy potente, porque permite trabajar con el aparato del análisis matemático (Batanero, 2001).

En palabras muy simples y de manera muy intuitiva, la variable aleatoria relaciona un suceso del espacio muestral con un número real, pero la relación no es tan sencilla, puesto que el suceso tiene que estar definido en un espacio de probabilidad. El que el suceso esté definido en un espacio de probabilidad es lo que a su vez permite posteriormente, establecer otra relación importante que es la de la función de probabilidad. Es decir dentro de la definición de variable aleatoria, otros dos conceptos que están íntimamente relacionados, el espacio de probabilidad y la función de distribución. Estos, a su vez, involucran otros conceptos más elementales, lo que hace de la variable aleatoria un concepto muy complejo desde la perspectiva epistémica. Así, la problemática de la variable aleatoria se verá modificada dependiendo si la noción de probabilidad que se trabaje en el espacio de probabilidad es frecuentista, clásica o subjetiva pero una vez definida la variable aleatoria y con ella la función de probabilidad no importará de

qué tipo de probabilidad se deriva, será tratada como probabilidad formal en la función de distribución en donde se relaciona con otro elemento matemático que es la variable aleatoria.

Otro tipo de consecuencia que se deriva a partir de esta complejidad epistémica es la que involucra a la variable aleatoria como elemento de modelación. La variable aleatoria es la que permite construir un modelo matemático a través de la probabilidad. Así cuando nosotros hablamos de una función de distribución, la variable aleatoria nos lleva a la realidad, el número 5 ó 6 nos conduce de inmediato al suceso al que nos estamos refiriendo, es decir, la variable aleatoria en ese contexto es la realidad. Pero en el proceso de transformar los sucesos del espacio muestral a números, la variable aleatoria es el modelo matemático, es la abstracción. Anteriormente también los sucesos, ligados a un espacio de probabilidad, son el modelo matemático. Así, la variable aleatoria a la vez que funge como “realidad” en un contexto, en otro contexto, muy cercano, funge como modelo matemático.

Por otro lado desde la perspectiva histórica, hay razones que justifican por qué los matemáticos pasados, que no conocían esta idea, tuvieron serios problemas con diversas paradojas matemáticas y, por ejemplo, Bernoulli necesitara 20 años para descubrir y probar su ley débil de los grandes números, temas que hoy se explica en unas pocas líneas. (Batanero, 2001).

Desde una perspectiva social, manejamos intuitivamente la idea de variable aleatoria cuando nos encontramos con juegos y experimentos en los que usamos dados, monedas, etc. Así mismo tenemos experiencias cotidianas con variables aleatorias continuas, como el tiempo de espera del autobús, o el necesario para llegar de nuestra casa al trabajo. Por el refuerzo de las múltiples, a veces inconscientes, experiencias con variables aleatorias en la vida cotidiana, la intuición de la magnitud aleatoria y de valor esperado, a veces aparece antes que la de probabilidad (Heitele, 1975).

4. Componente didáctica.

Los programas de estudio suelen ubicar el estudio de la variable aleatoria justo antes de ver Distribuciones de Probabilidad como un antecedente a éstas. Entrevistas con profesores manifiestan las peculiares dificultades que presentan los estudiantes para el tema de variable aleatoria tanto por su notación peculiar, como por fuerte tendencia a asociarla a la variable algebraica, y por tanto, diluirse lo estocástico. Los libros de texto usados en el ITESM en ciencias sociales e ingeniería, mencionan la variable aleatoria fuera de sus capítulos de probabilidad para abordarla como preámbulo al de distribuciones de probabilidad. Esta separación y orden pudiera estar generando algunas dificultades, pues autores como Heitele (1975) sostienen que la noción de variable aleatoria y su esperanza es, a veces, incluso anterior a la probabilidad. Además, a la par de la idea de variable aleatoria es necesario estudiar y retomar otras ideas que están vinculadas con ella, lo que hace que sea una idea didácticamente difícil de ser abordada.

En general, encontramos una falta de profundización sobre la idea de variable aleatoria en los libros de texto coincidiendo con algunos autores que han analizado libros de texto como Ortiz (2002) y Miller (1998), quienes en sus investigaciones concluyen que hay una ausencia de la

presencia de la noción tanto de variable aleatoria tanto en libros de texto en secundaria (Ortiz) y profesional (Miller). Así como con Tauber (2001) quien concluye que no hay una conexión entre el estudio del modelo probabilístico y los datos empíricos en los libros de texto, y con Oseguera (1994) quien reporta una falta de vinculación de la variable aleatoria con experiencias intuitivas de los estudiantes en los programas de estudio.

5. Componente cognitiva.

La literatura en este ámbito resultó muy escasa, así encontramos algunas citas principalmente en Batanero (2001) y Heitele (1975) quienes basándose en que el modelo de variable aleatoria hay tres conceptos básicos: su distribución, media y varianza, reúnen diversos estudios sobre estos tres conceptos. Así citan que algunos psicólogos sostienen que la habilidad para estimar la esperanza matemática de las variables aleatorias es puramente biológica puesto que mediante la experiencia es posible llegar a estimar el tiempo medio que tardaremos en preparar la comida o arreglarnos, lo que gastaremos en promedio al hacer la compra. Sin embargo, en un plano formal, de acuerdo con Piaget y otros, la esperanza matemática se interpreta como la media aritmética de los valores de una variable aleatoria, si el experimento se repitiese suficientemente en condiciones idénticas. También sostienen que los tres conceptos se deben enseñar de manera diferente puesto que mientras que la idea de media (esperanza matemática) es muy intuitiva, lo es menos la idea de distribución, especialmente cuando unos valores son más probables que otros. Ambos dan un lugar preponderante a la enseñanza de la distribución Normal y a las dificultades que los estudiantes tienen en su aprendizaje.

6. Exploración cognitiva

Fue pertinente realizar una exploración cognitiva que nos permitiera complementar la revisión bibliográfica sobre la componente cognitiva profundizando sobre el aprendizaje de la variable aleatoria en particular. Se pretendió que de esta exploración cognitiva se obtuvieran 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 un el atisbo de algunas trayectorias que siguen los estudiantes al querer resolver el problema propuesto.

La exploración consistió en que dos estudiantes (Brenda y Mónica) de nuevo ingreso a la universidad resolvieran un problema planteado a través de un cuestionario dividido en tres partes. El problema fue el siguiente:

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. 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 encomiendan a la trabajadora social de la empresa que decida cuántos boletos tiene que comprar (se proporcionan, en forma tabular, el número de hijos que tienen los trabajadores de la fábrica).

7. Algunos resultados.

PARTE I. Ellas comenzaron a trabajar la probabilidad como un cociente. No mostraron mucha dificultad para obtener la probabilidad de todos los eventos posibles. Sin embargo al cuestionarlas se notó una confusión entre lo que es la probabilidad y el número de trabajadores. Aunque están conscientes de que el número de trabajadores por sí mismo no es la probabilidad sino que tiene que ir “acompañada” con el total de trabajadores (es decir, el cociente entre ambos). No efectúan el cociente entre el número de trabajadores y total de trabajadores, más bien mencionan poco la operación que media a los dos números y lo manejan mejor como una relación de un cierto número de trabajadores de un total de otros tantos, como si la raya de división sólo sirviera como una nomenclatura para establecer la relación entre las dos cantidades, pero no la alcanzaran a ver como un solo número (el cociente). Creemos que su preferencia a no efectuar el cociente les dificulta interpretar la probabilidad como una variable que puede tomar un valor y por eso mismo les parece más natural que la variable sea el número de trabajadores. Una vez que tienen los decimales, ellas interpretan correctamente el significado de la probabilidad como la parte de una unidad, aunque se facilitan la interpretación con porcentajes.

De manera que ellas definen que la variable dependiente es el número de trabajadores y la variable independiente el número de hijos porque el número de hijos define el número de trabajadores y no al revés. Sin embargo la relación de dependencia que ellas manejan es “la probabilidad depende del número de hijos”. Con la intervención de los profesores, ellas se dan cuenta de que en realidad la probabilidad es la variable dependiente y no el número de trabajadores. Una vez más, a pesar de que en la primera actividad aparentemente llegaron a definir correctamente la probabilidad como un número (incluso en forma decimal), terminan por decir que el número de hijos es lo mismo que la probabilidad y cuesta un poco de trabajo desprenderse del número de trabajadores como variable y la probabilidad y que manejen la segunda en lugar de la primera.

PARTE II: En primera instancia Brenda y Mónica piensan que lo mejor es comprar 2 boletos y después comprar más boletos cuando ya se sepa cuántos hijos tendrá la familia ganadora. Después se inclinan por comprar 5 boletos, porque el número de hijos de los trabajadores vinculados a la mayor probabilidad es 3. La necesidad de la probabilidad acumulada surge de los mismos argumentos de las estudiantes por convencerse una a la otra de cuántos boletos comprar al observar a quiénes beneficiarían los boletos que compran, sin embargo, el profesor es el que surge que se elabore una tabla con las probabilidades acumuladas. La tabla es la que las hace discutir mejor sobre cuál sería la mejor opción, una de ellas está de acuerdo en que comprar 5 boletos sería lo mejor porque ello haría que la mayoría de los trabajadores saliera beneficiado, pero hay una cierta preocupación por no dejar desamparados a los trabajadores que tienen un mayor número de hijos. Si se compran 5 boletos, sólo el 58% de los trabajadores tienen posibilidades de no tener problemas con llevar a sus hijos, pero si se compran más (por ejemplo, 6 ó 5) es casi seguro que menos familias tendrán problemas con sus hijos. El profesor les aclara que han de ponerse de parte de la empresa, eso induce a las muchachas a decidirse por 5 boletos. Ellas usan la probabilidad acumulada antes de calcularla para convencerse mutuamente como un “*si tienes cubierto todo esto*”, sin embargo, no surge de ellas el calcularla, lo mismo pasa con la probabilidad del complemento, que de hecho no calculan

Al calcular las probabilidades exactas y acumuladas sólo necesitan volver a leer detenidamente el problema (¿se incluye a los que tienen exactamente 3 hijos?). Concluyen que no es posible asegurar que no habrá desperdicio de boletos, pero sí podemos saber qué es lo más probable:

PARTE III. En primera instancia, relacionan la notación funcional con la expresión algebraica del evento que están tratando de describir. “De hecho de la tabla se saca la función”, por lo tanto ven la tabla como una función (aunque no lo expresan propiamente), pero siguen pensando en encontrar la expresión algebraica. El dominio de la función lo obtienen como si fuera una función continua, aunque hay una confusión entre el contexto del problema (el número de hijos) y el contexto matemático (una gráfica en la que unieron los puntos y que, manifiestan, “*siempre han construido así*”). Se inclinan por mencionar que al interpretar no debemos usar decimales en la variable número de hijos, pero en la gráfica o al describir el dominio optan por la nomenclatura continua porque, dicen, “*así se acostumbra en matemáticas*”, como si las matemáticas fueran un ente ajeno a todo contexto y al interpretarlas en una situación ya dejaran de ser matemáticas. En el rango se deciden casi inmediatamente (casi sin pensarlo) por una variable continua.

Mencionan que la variable independiente de esta relación (probabilidad-número de hijos) es diferente a otras funciones que han trabajado en álgebra o en cálculo en varios aspectos: es limitada, porque no puede tomar todos los valores que quieran, sólo de 0 a 9; no puede tomar valores decimales, sólo enteros; aquí no se manejan ecuaciones, los valores de ‘y’ se dan por la ecuación y en esta relación es aleatoria porque no se puede sacar a través de un patrón; la diferencia la hace la variable dependiente (probabilidad) porque sus valores no se pueden obtener a través de una ecuación sino que son datos que se dan, son dadas por una situación no por una relación matemática, no hay exactitud (mencionan que de poder encontrar la ecuación, lo pueden hacer, pero ella no tendría sentido en la descripción del problema); además no es posible predecir qué sigue después de que ocurra un evento.

Conclusiones

El concepto de variable aleatoria tiene una gran complejidad epistémica, puesto que se apoya en conceptos previos complejos, está vinculada con un proceso de modelación matemática, está asociada con los conceptos de distribución de probabilidad y función de distribución, se vincula con otros conceptos no probabilísticos.

Se pudo observar que las estudiantes manejan eficientemente la noción de variable aleatoria en su contexto de significado, pero tuvieron dificultades para su asociación con el concepto de función, y por tanto, con su formalización. Se hicieron patentes las dificultades en el manejo de una variable discreta, surgieron dificultades de origen aritmético tal como asociar la probabilidad a dos números (el cociente de dos enteros) en lugar de reconocer que se trata de uno solo (decimal). La variable dependiente se mostró más débil en el contexto de la probabilidad que en un contexto algebraico debido a que la probabilidad resultó algo menos tangible y controlable como variable dependiente. La noción de azar se manifestó en su fase más elemental de lo incontrolable e impredecible. Así mismo se observa que encuentran difícil

pasar del contexto del problema al contexto matemático porque no ven a la variable aleatoria como modelo matemático sino sólo como parte de la realidad.

Referencias Bibliográficas

- Albert A. (1998). Introducción a la epistemología de la Escuela Mexicana de Matemática Educativa, Bogotá: *Actas de RELME XII*.
- Artigue, M., Douady, R., Moreno, L. y Gómez, P. (1995) *Ingeniería didáctica en educación matemática*. México: Iberoamericana.
- Batanero, C. (2001). *Didáctica de la Estadística*. Granada: Grupo de investigación en educación estadística, Universidad de Granada.
- Brousseau, G. (1986). Fondements et méthodes de la didactiques des mathématiques. *Recherches en Didactique des Mathématiques*, 7 (2), 33-115.
- Heitele, D. (1975). Un enfoque epistemológico sobre las ideas estocásticas fundamentales. *Educational Studies in Mathematics* 6, 187-205.
- Miller, T.K. (1998). The random variable concept in introductory statistics. In Pereira-Mendoza, L. et al (Eds) (1998). *Statistical education - expanding the network: Proceedings of the Fifth International Conference on Teaching Statistics*. 1998 (1221-1222) Singapore: International Statistical Institute.
- 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.
- Ortiz J. (2002). *La probabilidad en los libros de texto*. Tesis doctoral. 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 para obtener el grado de maestría. México: CINVESTAV-IPN.
- Tauber, L. (2001). *La construcción del significado de la distribución normal a partir de actividades de análisis de datos*. Tesis Doctoral. Universidad de Sevilla.