EXPLORACIÓN BIBLIOGRÁFICA SOBRE LA ENSEÑANZA A NIVEL UNIVERSITARIO DE LOS CONCEPTOS DE VARIABLE ALEATORIA Y FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD

Calandra, María Valeria[1]; Costa, Viviana Angélica[2]
[1]UIDET GAMEFI, Departamento de Ciencias Básicas, Facultad de Ingeniería, UNLP.
[2]UIDET IMApEC, Departamento de Ciencias Básicas, Facultad de Ingeniería, UNLP.
mava@mate.unlp.edu.ar, vacosta@ing.unlp.edu.ar

Resumen

Este trabajo se enmarca en la etapa inicial de una tesis de maestría que está centrada en la problemática de la enseñanza y aprendizaje de los conceptos de variable aleatoria y función de distribución probabilidad, en especial en carreras de Ingeniería. Se presenta una exploración bibliográfica que da cuenta de las dificultades en la enseñanza y el aprendizaje de esos conceptos, y las distintas estrategias didácticas y metodológicas utilizadas para revertirlas. El objetivo es revisar la importancia de esta problemática e identificar las distintas dificultades reportadas de modo de poder establecer las bases para proponer una actividad didáctica superadora en una fase posterior.

Palabras clave: Enseñanza, Probabilidades, Variable aleatoria, Función de distribución de probabilidad.

Abstract

This study forms part the initial stage of a master's thesis is focused on the problems of teaching and learning the concepts of random variable and probability distribution function, especially in engineering schools. A literature examination that accounts for the difficulties in teaching and learning of these concepts, and different didactic and methodological strategies used to reverse them presents. The aim is to review the importance of this problem and identify the various difficulties reported so that we can establish the basis for proposing a that overcomes teaching activities at a later stage.

Keywords: teaching, probabilities, random variable, probability distribution function.

1. Introducción

En las carreras de Ingeniería se presenta en general la asignatura Probabilidades en los primeros años. En particular en la Facultad de Ingeniería de la Universidad Nacional de La Plata se corresponde con el segundo año para todas las carreras. Los conceptos matemáticos relativos a Probabilidades son una herramienta matemática importantísima para un estudiante de ingeniería. Les permite avanzar en el estudio de la estadística inferencial, y en su futura profesión les sirve para abordar el estudio de poblaciones a partir de muestras, obtener conclusiones acerca de las mismas, y a partir de ellas poder diseñar procedimientos, tomar decisiones, controlar productos y procesos, auditar organizaciones y muchas actividades propias de la profesión. Pero el aprendizaje de las temáticas involucradas puede presentar cierto grado de dificultad. Resulta difícil iniciar a los estudiantes en el pensamiento probabilístico el cual incluye técnicas de conteo y manejo de operaciones con conjuntos entre otros. Además de la dificultad extra generada por el hecho de que aunque la introducción a la temática de probabilidades está incluida en los diseños curriculares de la enseñanza preuniversitaria, en general la misma no es tratada con la profundidad necesaria o simplemente se la omite. Por otra parte, modelar una situación de la vida cotidiana en términos de un experimento aleatorio, de un espacio muestral y de una o más variables aleatorias, no resulta natural para los alumnos de los primeros años de las carreras universitarias. Esta problemática es la que motiva la investigación presentada en la que revisamos su importancia, e identificamos y analizamos las distintas dificultades reportadas, y las propuestas para su enseñanza realizada por diversos investigadores. Estos se han relevado de una selección realizada consultando libros, artículos en revistas científicas, de divulgación, en actas de congresos, simposios y jornadas, de las últimas décadas, que abordan la problemática de la enseñanza y aprendizaje de los conceptos de variable aleatoria y función de distribución de probabilidad, a nivel universitario. El objetivo es establecer las bases para proponer una actividad didáctica superadora en una etapa posterior.

2. Los conceptos de variable aleatoria y de función de distribución de probabilidad

Según Batanero (2001) *"La variable aleatoria ha sido responsable de las múltiples aplicaciones actuales del cálculo de probabilidades, puesto que el cálculo de probabilidades pasó del estudio de sucesos aislados al estudio de las distribuciones de probabilidad (y posteriormente al de los procesos estocásticos)"*. La variable aleatoria por un lado constituye el dominio de las funciones de probabilidad y por otro se trata de una función que relaciona los sucesos o eventos de un espacio muestral con los números reales, sirve para cuantificar las probabilidades de los eventos de un experimento aleatorio y es el sustento de la teoría de la probabilidad. Las variables aleatorias y las funciones que las caracterizan dan sistematicidad a los conceptos de espacio muestral, suceso y probabilidad, representan el modelo matemático que servirá para expresar el comportamiento de los fenómenos aleatorios.

2.1 Problemáticas en la enseñanza y aprendizaje de los conceptos.

Se identificaron en las publicaciones seleccionadas algunas problemáticas comunes y han sido agrupadas y clasificadas por su naturaleza. En lo que sigue hacemos una descripción.

a) Dificultades para reconocer la variable aleatoria: se recopilaron tres publicaciones que mencionan como problema la correcta identificación de la variable aleatoria en el contexto de una situación problemática, esto implica extraer del problema la información necesaria para modelar la situación matemáticamente. La identificación de la variable aleatoria es importante para el tratamiento de las distribuciones de probabilidad lo cual permite el cálculo de probabilidades y de otras medidas características como su esperanza, varianza, etc. Nardecchia y Hevia (2003), señalan 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 en términos de la variable aleatoria (como modelo matemático) puede constituir un obstáculo con el que se podría enfrentar un estudiante. La modelación requiere identificar la variable, es decir, definir la regla de correspondencia que asigne valores numéricos a los resultados de un experimento aleatorio. Ruiz Hernández (2006) como así también Calandra y Costa (2015) afirman que algunos estudiantes no pueden identificar la variable aleatoria involucrada en una situación problemática, lo que obstaculiza su resolución.

b) Tendencia a realizar cálculos mecánicos: muchas veces ocurre que una vez modelada la situación los alumnos lo tratan desde el punto de vista matemático dándole importancia a los procedimientos pero olvidándose del contexto del problema a tratar lo que les impediría interpretar adecuadamente los resultados (Ruiz Hernandez y otros, 2006). También en varias oportunidades los estudiantes son capaces de seleccionar la fórmula o modelo probabilístico pertinente a la situación, por ejemplo el modelo binomial, hipergeométrico, geométrico, etc. pero lo usan inadecuadamente para obtener la probabilidad solicitada, o no siempre comprenden por qué la fórmula es así o de dónde proviene (Andrade y otros, 2013). Entonces, trabajar las situaciones se convierte en un ejercicio de memorización y de reemplazo de fórmulas.

c) Desconocimiento de la génesis de la función de densidad asociada a una variable aleatoria continua: un autor menciona que también a los alumnos les resulta extraño trabajar con funciones de densidad de probabilidad ya que las consideran como una función poco tangible dado que no es una magnitud física. Tauber (2001), centra su estudio en el aprendizaje de alumnos universitarios, de la distribución Normal, y advierte sobre la existencia de ciertas dificultades de éstos para distinguir la distribución teórica y empírica, sobre todo cuando se ven en la necesidad de resolver problemas abiertos.

d) Dificultad en la naturaleza funcional de la variable aleatoria: el espacio muestral o el dominio de la variable aleatoria asociada a un experimento aleatorio, no siempre resulta clara (Fernandez, 2011). La asociación funcional entre los elementos del espacio muestral (no necesariamente numérico) y los valores que asume la variable aleatoria es importante para el caso de variables aleatorias discretas. Solo este autor menciona esta problemática.

e) Dificultades de tipo semióticas: han sido reportada por cinco autores. Una variable aleatoria se define como una función que a cada elemento del espacio muestral le asigna un número real y se denota usando las letras mayúsculas: X, Y. Los estudiantes no están preparados para usar esa simbología para denotar funciones *"consideran que con la letra (F) harían notar que es una función"* (Ruiz Hernandez, 2006). En la confusión de los estudiantes también puede incidir el nombre de 'variable aleatoria' pues tal como señala Meyer (1992) *"X es una función y todavía la llamamos variable aleatoria"*. Y como lo señala Goldberg (1974) "Llamar variable aleatoria a una función de valores numéricos definida respecto a un espacio muestral es inadecuado". Cabe aclarar que también Meyer (1992) advierte que la terminología variable aleatoria es algo confusa, pero es universalmente aceptada. Feller (1973) menciona que el término variable aleatoria es un poco confuso y opina que función aleatoria sería más adecuado. En Ruiz Hernández (2006) se reportan dificultades desde el punto de vista cognitivo con la convivencia de distintas expresiones de la función de probabilidad para variables aleatorias discretas con la igualdad: $P(x_i)=P(X=x_i)$ que implican la convivencia de dos paradigmas: la probabilidad como una función del rango de la variable aleatoria en el intervalo [0,1] y la probabilidad como una función del espacio muestral en el intervalo [0,1].

f) Surgimiento de concepciones erróneas: seis autores fueron identificados en esta descripción. Es común que los alumnos tomen como definición de variable aleatoria continua como aquella cuyo rango es un conjunto no numerable (generalmente un intervalo o una combinación de intervalos) lo cual es falso pues se podría tratar de una variable aleatoria mixta. Otra concepción errónea de los alumnos es pensar a una variable aleatoria continua como una variable con una función de distribución acumulada continua. La condición de continuidad de la función de distribución es necesaria pero no suficiente, en general esto no es tratado en los cursos de grado (Kachapova, 2012). En cuanto al aprendizaje de la distribución normal en alumnos universitarios Huck y otros (1986) han identificado dos concepciones erróneas sobre las puntuaciones normales tipificadas: algunos alumnos consideran que todas las puntuaciones tipificadas han de tomar un valor comprendido entre -3 y +3, lo que podría estar relacionado con que el 99,7 % de las observaciones se encuentra entre la media ± 3

desviaciones típicas o a la interpretación inadecuada de las tablas de distribución. Hawkins y otros (1992) describen errores que cometen alumnos universitarios en la aproximación de una distribución binomial mediante la distribución normal ya que aplican la corrección por continuidad de una forma mecánica, sin comprender su significado. Asumen también que una función de densidad no puede tomar valores mayores que 1. Algunos estudiantes confunden la estandarización de una variable aleatoria con la transformación en Normal estándar (Calandra y Costa, 2015; Tauber, 2001).
Es común que consideren que si el rango de una variable aleatoria no incluye valores enteros entonces es continua (Stoyanov, 1997). Por ejemplo, si la longitud de una varilla puede tomar los siguientes valores:

Longitud de la varilla (redondeada a la décima más cercana)	**5,1**	**5,2**	**5,3**	**5,4**	**5,5**	**5,6**
Número de varillas	**0**	**3**	**35**	**40**	**18**	**18**

y consideramos a la variable aleatoria X= "longitud de una varilla seleccionada al azar", es en este caso en el que los estudiantes podrían dudar en afirmar si esa variable aleatoria es continua o discreta.
g) Problemas con el uso de la función de distribución: han sido detectados en dos publicaciones. Desde la perspectiva disciplinar para calcular $P(a < X \leq b)$ siendo $a < b$ se utiliza su función de distribución, que se denota $F(x)$ y se define como $F(x)=P(X\leq x)$, por lo tanto $P(a < X \leq b)= F(b)-F(a)$. En particular si se trata de una variable continua, $F(x)$ es una función continua, por lo tanto $P(a \leq X \leq b) = P(a < X \leq b) = F(b)-F(a)$. Algunos autores han detectado problemas con el uso de $F(x)$ para variables continuas como: Kachapova (2012), Calandra y Costa (2015) y Stoyanov (1997). Por ejemplo: si denotamos con X a una variable aleatoria continua definida como "tiempo de vida de una lamparita en horas", la probabilidad de que una lamparita dure entre 100 y 200 horas inclusive se denota como $P(100 \leq X \leq 200)$ y se calcula como $F(200)-F(100)$. Algunos alumnos la calculan erróneamente como $F(200)-F(99)$ ya que consideran que de otro modo no incluyen las lamparitas que duran 100 horas.

2.2 Metodologías de enseñanza y aprendizaje sugeridas para estos conceptos.

Los siguientes autores proponen diversas metodologías de enseñanza que fueron agrupadas en los siguientes puntos comunes:
a) Enfatizar el carácter funcional de la variable aleatoria: Fernández y otros (2011) sugieren proponer situaciones para la conceptualización de la variable aleatoria desde una perspectiva que enfatice su carácter funcional entre un espacio muestral y un conjunto numérico.
b) Valorar la conveniencia didáctica los conceptos de estadística descriptiva y probabilidades: Gutiérrez y otros (2013) consideran que la definición de variable aleatoria continua, es muy poco intuitiva e introduce la función de densidad de probabilidad de manera muy artificial. Sugiere desarrollar la idea de función empírica de densidad, al momento de tratar la representación gráfica de variables de tipo continuo, a través de una definición apropiada de histograma haciendo la conexión entre estadística descriptiva y probabilidades relacionando la variable estadística con la variable aleatoria.
c) Los contraejemplos en la enseñanza: Kachapova (2012) propone usar contraejemplos en la enseñanza del concepto de variable aleatoria para erradicar los conflictos que los alumnos poseen con la propia definición. También alerta sobre la necesidad de introducir contraejemplo para la comprensión de las propiedades de la

distribución acumulada de una variable aleatoria continua. El rol de los contraejemplos en probabilidades en general puede ser encontrado en el libro de Stoyanov (1997).

d) La contribución de la historia en la enseñanza: según Polola y otros (2006) la historia de la génesis de los distintos conceptos usados en probabilidades y estadística juega un papel importante ya que la descripción y el análisis de los pasos constructivos que se han dado en su evolución pueden reeditarse en favor del proceso de enseñanza. Para reeditar y vivenciar los procesos históricos, como recurso didáctico en este área, las estrategias fundamentales de trabajo podrían ser la experimentación y la simulación. La experimentación permite que los alumnos logren visualizar e imaginar, en definitiva vivenciar el problema en estudio tal como lo realizaron quienes participaron en su desarrollo. En este trabajo los autores hacen hincapié en que el estudio de la historia de la matemática influye en el diseño de experiencias didácticas y formulan una actividad para arribar a la distribución Normal partiendo de la distribución Binomial, utilizando la experimentación en el aula para que los alumnos "revivan" el proceso real seguido por Bernoulli y por De Moivre quien en 1773 descubrió la función de densidad de probabilidad de la distribución normal como una forma límite de la función Binomial. Por su parte Maltz (2015) desarrolla una forma de motivar la aparición de la Distribución Normal en los primeros cursos universitarios de probabilidades y comenta: *"Muchos autores la definen sin motivarla y esperan que sus buenas propiedades convenzan a los alumnos de su importancia"*. Para ello presenta un esquema que era usual hace varias décadas (basado en el tipo de ideas estadísticas de Gauss), *"con un pequeño maquillaje"* (Maltz, 2015). Por otro lado Ruiz Hernández (2006) realiza un análisis epistemológico histórico del concepto variable aleatoria y encuentra algunos obstáculos en el surgimiento que apoyan y sustentan algunos resultados cognitivos obtenidos sobre una experiencia didáctica realizada a un grupo de alumnos.

e) Las simulaciones como recurso para la enseñanza: el potencial de la simulación en la enseñanza de la probabilidad data desde la década de los años 70 y 80 del siglo pasado. Biehler (1991, 1997) ha sido referente para muchos trabajos de investigación actuales. Investigadores han desarrollado diversas ideas interesantes y han realizado un análisis didáctico en torno a la perspectiva de modelación y simulación en probabilidad, por ejemplo: Bouleau (1986), Dantal (2001), Batanero (2003) y Tauber (2001).

5. Conclusiones

La exploración realizada sobre la problemática de la enseñanza y aprendizaje de estos conceptos nos muestra que no es sencilla, las dificultades con las que se enfrentan los estudiantes son muchas y de distinta naturaleza. Además, nos abre el camino para la exploración de nuevos obstáculos sobre el aprendizaje de estos temas que no hayan sido estudiados. Por ejemplo, en relación al conocimiento de la génesis histórica de un saber cómo punto de partida para un análisis didáctico, hemos identificado que existen aportes en la enseñanza sobre la emergencia de la distribución Normal, pero no sobre la génesis de distribuciones de probabilidad de otras variables aleatorias continuas importantes como la gama y otras. Finalmente en virtud de las problemáticas encontradas en relación a la enseñanza, esta revisión nos brinda información útil para elaborar en una etapa posterior un marco conceptual que sea adecuado para el desarrollo de una estrategia didáctica a ser implementada en cursos de Probabilidades en carreras de ingeniería; que haga posible plantearnos preguntas de investigación pertinentes, objetivos relevantes e hipótesis compatibles con lo que se sabe e ignora.

6. Referencias

Andrade, L., Fernandez, F., Sarmiento, B. (2013). Capítulo 4: La búsqueda del espacio muestral 'original': una necesidad para la enseñanza. A. Salcedo (Ed.), *Educación estadística en América Latina: tendencias y perspectivas* (pp. 81-98). Caracas, Venezuela: Universidad Central de Venezuela.

Batanero, C. (2001). Didáctica de la estadística. Grupo de investigación en educación estadística. Departamento de Didáctica de la Matemática, Universidad de Granada, España. http://dv.fosjc.unesp.br/ivan/downloads/Aulas%20em%20PDF*Didactica_Estadistica.pdf

Batanero, C. (2003). La simulación como instrumento de modelización en probabilidad. *Revista Educación y Pedagogía,* Medellín: Universidad de Antioquia, Facultad de Educación. 15(35), 37 -54.

Biehler, R. (1991). Chapter 6: Computers in probability education. R. Kapadia (Ed.), Chance encounters: probability in education. A review of research and pedagogical perspectives. (pp. 109-211). Amsterdam: Reidel Amsterdam.

Biehler, R. (1997). Software for learning and for doing statistics. *International Statistical Review*, 65(2), 167-189.

Bouleau, N. (1986). *Probabilités de l'ingénieur: variables aléatoires et simulation.* París, Francia: Hermann, Éd. des Sciences et des Arts.

Calandra, M. V., Costa, V. A. (2015). La problemática de la enseñanza y aprendizaje del concepto de variable aleatoria continua y de función de densidad de probabilidad. *IV Jornadas de Enseñanza e Investigación Educativa en el Campo de las Ciencias Exactas y Naturales.* Ensenada, Argentina: Departamento de Ciencias Exactas y Naturales Facultad de Humanidades y Ciencias de la Educación (UNLP).

Dantal, B. (2001). Deuxième partie: 1- Les enjeux de la modélisation en probabilité. M. Henry (Ed.) *Collection «Didactiques» «Mathématiques» Autour de la modélisation en Probabilités* (pp. 137-140) Besançon, Francia: Presses universitaires de Franche-Comté (PuFC)

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

Fernández, F., Andrade, L., Montañez, J. (2011). Hacia una posible aproximación comprensiva de la variable aleatoria. *XIII CIAEM-IACME*, Recife, Brasil: Universidade Federal de Pernambuco.

Goldberg, S. (1974). *Cálculo de probabilidades*. Bilbao: Ediciones Urmo S.A.

Gutiérrez, R.B., y Grima Cintas, P. (2013). El histograma como un instrumento para la comprensión de las funciones de densidad de probabilidad. *Probabilidad Condicionada. Revista de didáctica de la Estadística,* (2), 229-235.

Hawkins, A., Joliffe, F. y Glickman, L. (1992) *Teaching statistical concepts.* London: Longman Publishers.

Huck, S., Cross, T.L. y Clark, S.B. (1986). Overcoming misconceptions about z-scores. *Teaching Statistics,* 8(2), 38-40.

Kachapova, F. (2012). A general approach to teaching random variables. *Mathematics Teaching-Research Journal (MTRJ) Online.* 5(2), 1-16.

Maltz, A. L (2015). Un enfoque en la enseñanza de la distribución Normal. *Revista de Educación Matemática,* 30(1), 34-37.

Meyer, P.L. (1992) *Probabilidad y aplicaciones estadísticas*. Delaware, EE.UU.: Addison-Wesley Iberoamericana.

Nardecchia, G. y Hevia, H. (2003). Dificultades en la enseñanza del concepto de variable aleatoria. *V Simposio de Educación Matemática*. Chivilcoy, Argentina: Centro Regional Chivilcoy de la Universidad Nacional de Luján.

Polola, L., Pagano, L., Brunetti, S., Ecalle, M., y Borgna, E. (2006). Génesis y evolución histórica de los conceptos de probabilidad y estadística como herramienta metodológica. http://economicas.unlam.edu.ar/descargas/5_b107.pdf

Ruiz Hernández, B.R. (2006). Un acercamiento cognitivo y epistemológico a la didáctica del concepto variable aleatoria. Tesis de Maestría. Instituto Politécnico Nacional, México. http://www.matedu.cicata.ipn.mx/tesis/maestria/ruiz_2006.pdf

Ruiz Hernández, B.R., Albert Huerta, J.A. y Batanero, C. (2006). An exploratory study of students difficulties with random variable. Invited paper in ICOTS-7.Proceedings of Seventh International Conference on Teaching Statistics, (S/paginación). Salvador (Bahía), Brasil: International Association on Statistical Education.

Stoyanov, J.M. (1997). *Counterexamples in Probability (2 ed.)*. England: Wiley.

Tauber, L. (2001). La construcción del significado de la Distribución Normal a partir de actividades de análisis de datos. Tesis Doctoral. España: Universidad de Sevilla. http://www.ugr.es/~batanero/pages/ARTICULOS/Tesisliliana.pdf