

## **Conexiones entre las aproximaciones clásicas y frecuencial de la probabilidad en un ambiente de modelación computacional**

Santiago Inzunza Cazares, Universidad Autónoma de Sinaloa (México)

Recibido el 11 de Noviembre de 2016; aceptado el 25 de Febrero de 2017

---

### **Conexiones entre las aproximaciones clásicas y frecuencial de la probabilidad en un ambiente de modelación computacional**

#### **Resumen**

*Reportamos resultados de una investigación con un grupo de estudiantes de ciencias sociales que tomaban un curso introductorio de probabilidad, cuyo propósito era explorar el razonamiento ligado a las conexiones entre la probabilidad teórica y la estimación frecuencial de la probabilidad de eventos simples y compuestos en un ambiente de modelación computacional. Los resultados muestran que el proceso de modelación fue una actividad sencilla para la mayoría de los estudiantes cuando las probabilidades del problema eran conocidas (del enfoque clásico al enfoque frecuencial). Sin embargo, en la dirección opuesta (del enfoque frecuencial al enfoque clásico), la variabilidad de los resultados fue uno de los principales obstáculos para identificar el modelo teórico subyacente, particularmente en aquellos estudiantes que utilizaron menos de 1,000 simulaciones. Los estudiantes que utilizaron 5,000 o más simulaciones, desarrollaron una comprensión intuitiva de la ley de los grandes números y ajustaron las frecuencias al modelo teórico correcto que ellos generaron.*

**Palabras clave.** Probabilidad, modelación, ambiente computacional, simulación.

### **Conexões entre probabilidade teórica e a probabilidade frequencial num ambiente de modelação computacional**

#### **Resumo**

*Reportamos resultados de uma pesquisa com um grupo de estudantes de ciências sociais que tomavam um curso introdutório de probabilidade, cujo fim era explorar o raciocínio ligado às conexões entre a probabilidade teórica e probabilidade frequencial de eventos simples e compostos num ambiente de modelação computacional. Os resultados mostram que o processo de modelação não foi uma atividade complexa para a maioria dos estudantes quando as probabilidades do problema eram conhecidas (do enfoque teórico ao enfoque frequencial), mas na direção oposta (do enfoque frequencial ao enfoque teórico), a variabilidade dos resultados foi um dos principais obstáculos para identificar o modelo teórico subjacente, particularmente naqueles estudantes que utilizaram menos de 1,000 simulações. Os estudantes que utilizaram 5,000 ou mais simulações desenvolveram uma compreensão intuitiva da lei dos grandes números e ajustaram as frequências ao modelo teórico certo que eles geraram.*

**Palavras chave.** Probabilidade, modelação, ambiente computacional, simulação.

Para citar: Inzunza Casares, S. (2017). Conexiones entre las aproximaciones clásicas y frecuencial de la probabilidad en un ambiente de modelación computacional. *Avances de Investigación en Educación Matemática*, 11, 69 – 86.

## Connections between classical probability and the frequentist estimate of probability in a computer modelling environment

### Abstract

We report results of a study with a group of social science students taking an introductory course in probability. The purpose was to explore the reasoning linked to the connections between the theoretical probability and experimental probability in an environment modelling. The results show that the modelling process was not a complex activity for most students when the probabilities of the problem are known (from theoretical approach toward the frequency approach), but in the opposite direction (from frequency approach toward the theoretical approach), the variability of the results was one of the main obstacles to identify the theoretical model underlying, particularly in those students who used 1,000 or fewer runs. The students who used 5,000 or more runs developed an intuitive understanding of the law of large numbers and adjusted the frequencies to the correct theoretical model that generated.

**Key words.** Probability, modelling, computer environment, simulation

## Connexions entre la probabilité théorique et la probabilité fréquentielle dans un environnement de modélisation informatique

### Résumé

Nous présentons les résultats d'une étude avec un groupe d'étudiants en sciences sociales qui prenant un cours d'introduction en probabilité, dont le but était d'explorer les liens liés entre la probabilité théorique et la probabilité fréquentielle d'événements simples et composés dans un environnement de modélisation informatique. Les résultats montrent que le processus de modélisation n'a pas été une activité complexe pour la plupart des étudiants quand les probabilités du problème ont été connus (l'approche théorique à l'approche fréquentielle), mais dans le sens opposé (l'approche fréquentielle à l'approche théorique), la variabilité des résultats a été un des principaux obstacles pour identifier le modèle théorique sous-jacent, en particulier dans les étudiants qui ont utilisé moins de 1,000 simulations. Les étudiants qui ont utilisé 5,000 ou plusieurs simulations, ont développé une compréhension intuitive de la loi des grands nombres et ont ajustées les fréquences au modèle théorique correct qu'ils ont généré.

**Mots clés.** Probabilité, modélisation, environnement informatique, simulation

## 1. Introducción

Entre los principales obstáculos identificados para el desarrollo del razonamiento probabilístico de los estudiantes, se encuentra el excesivo énfasis que algunos profesores hacen en el enfoque clásico de la probabilidad, con frecuencia centrado en el uso de procedimientos rutinarios y técnicas combinatorias para el cálculo de probabilidades (Godino, Batanero y Cañizares, 1996); restando importancia a los enfoques frecuencial y subjetivo de la probabilidad, que de manera conjunta pueden contribuir al desarrollo de una mejor intuición y razonamiento probabilístico. Diversos investigadores (Batanero, Henry & Parzysz, 2005; Eichler & Vogel, 2014) proponen que la enseñanza de la probabilidad debe tener en cuenta los enfoques clásico, frecuencial y subjetivo de la probabilidad de manera interrelacionada; en particular, se deben explorar y hacer explícitas las conexiones entre los enfoque clásico y frecuencial (Chaput, Girard y Henry, 2011; Fischbein y Gazit, 1984). En estas ideas, el enfoque de modelación y simulación en la enseñanza de la probabilidad surge como el elemento integrador de estas propuestas.

Con el desarrollo de la tecnología y herramientas de software educativo cada vez más dinámicas, interactivas, con representaciones múltiples ligadas entre datos y

resultados, y con capacidad de simulación flexible de modelos, el enfoque de modelación y simulación ha llamado la atención de investigadores y profesores como alternativa de enseñanza de la probabilidad en todos los niveles educativos (Konold & Kazak, 2008; Pratt, Davies & Connor, 2011). Es así como estudios didácticos recientes apoyan una enseñanza basada en el proceso de modelación y simulación (Biehler & Prömel, 2010; Pratt, 2005). En esta misma dirección, Batanero, Henry y Parzysz, (2005, p. 32) señalan que “interpretar situaciones aleatorias en términos de modelos probabilísticos pueden ayudar a superar la controversia entre enfoque clásico, subjetivo y frecuencial”. De esta manera, “es posible presentar a la probabilidad como una herramienta de modelación que refleja la mentalidad de un experto cuando utiliza probabilidad para modelar comportamientos aleatorios en contextos del mundo real” (Prodromou, 2014, p. 417). Sin embargo, Borovcnick (2012) señala que aunque la respuesta usual a muchas dificultades experimentadas en el aprendizaje de la probabilidad es simplificar los conceptos a través del método de simulación, el cual explota la relación cerca entre probabilidades y frecuencias relativas; previene que el uso extensivo que se está haciendo de la simulación actualmente puede inducir un sesgo hacia la interpretación frecuencial de la probabilidad y mostrar sólo una cara de la probabilidad.

Los modelos probabilísticos que se asumen en los enfoques clásico y frecuencial y los resultados que se generan a través de ellos, pueden ser analizados a través de un proceso bidireccional, comparando las frecuencias observadas en los experimentos o simulaciones con las probabilidades obtenidas con el enfoque clásico y viceversa (Pfannkuch & Ziedine, 2014; Prodromou, 2012). En dicho proceso se ponen en juego importantes conceptos y propiedades (por ejemplo: aleatoriedad, variabilidad, tamaño de muestra, ley de los grandes números) que pueden ayudar a los estudiantes a desarrollar su razonamiento probabilístico y comprender el papel de los modelos en la predicción de la incertidumbre y la aleatoriedad. En este mismo sentido, García, Medina y Sánchez (2014) recomiendan evitar la tendencia a refugiarse sólo en el cálculo de probabilidades evitando los difíciles conceptos de aleatoriedad y variabilidad, sugieren complementar la enseñanza, planteando a los estudiantes actividades donde hagan observaciones sobre variables aleatorias y reflexionen sobre lo que ocurre con las frecuencias cuando un experimento se repite muchas veces.

La revisión de la literatura muestra pocos estudios sobre la perspectiva de modelación y simulación para la enseñanza de la probabilidad en el nivel universitario, y la conexión conceptual entre enfoque clásico y enfoque frecuencial, por lo que nuestra investigación se enfoca a explorar las relaciones y conexiones conceptuales que estudiantes universitarios de ciencias sociales establecen cuando abordan la probabilidad desde las perspectivas clásica y la frecuencial en un ambiente computacional como el proporcionado por el software TinkerPlots (Konold y Miller, 2011). En particular exploramos la forma como los estudiantes razonan y establecen relaciones entre la probabilidad clásica y la estimación frecuencial de la misma en situaciones de modelación que involucran eventos simples y compuestos. En específico nos planteamos las siguientes preguntas de investigación:

- ¿Qué conexiones establecen los estudiantes entre modelos, probabilidades y frecuencias relativas, cuando resuelven problemas donde utilizan el enfoque clásico y el enfoque frecuencial de la probabilidad en forma complementaria?
- ¿Pueden estudiantes universitarios establecer una conexión conceptual

correcta entre las aproximaciones clásica y frecuencial de la probabilidad en un ambiente de simulación computacional?

## **2. Antecedentes**

La revisión de la literatura muestra algunos trabajos que exploran el uso de modelos y las conexiones entre los enfoques clásico y frecuencial de la probabilidad en ambientes computacionales; la mayor parte de ellos se han realizado en niveles preuniversitarios. Por ejemplo, Ireland y Watson (2009), en una investigación con estudiantes de grado 5 y 6, en la que utilizaron el software TinkerPlots, señalan que muchos estudiantes pudieron conectar adecuadamente los resultados experimentales con el modelo teórico correspondiente y demostraron una comprensión intuitiva de la ley de los grandes números. Sin embargo, la falta de comprensión de las ideas subyacentes al concepto de equidad en los eventos, fue una barrera para que muchos estudiantes conectaran la probabilidad teórica con su estimación frecuencial.

En otro contexto, y con estudiantes para profesores como sujetos de estudio, Prodromou (2012) se propuso explorar el razonamiento y la manera como establecían la conexión bidireccional entre la probabilidad teórica (distribución de probabilidad) y su estimación frecuencial (distribución de frecuencias). Los resultados señalan que al hacer la conexión desde la grafica de la distribución de frecuencias hacia la grafica de la distribución de probabilidad, los profesores en formación identificaron la tendencia como un patrón global que podía emerger desde la distribución de frecuencias. Los profesores atendieron la variabilidad omnipresente en las gráficas de frecuencias y apreciaron la repetición de ensayos que se tradujeron en la estabilización de las frecuencias. Al hacer la conexión desde la distribución de probabilidad hacia la distribución de frecuencias, los profesores en formación percibieron la distribución de probabilidad como el resultado esperado y la distribución de frecuencias como el resultado observado. Esto sugiere que en la conexión realizada, la probabilidad teórica genera en algún sentido las frecuencias experimentales. Cuando los estudiantes compararon las frecuencias relativas experimentales con probabilidades teóricas, la mayoría logró distinguirlas, pero no siempre pudieron coordinar estos dos enfoques diferentes de la probabilidad debido a sus diferentes roles.

Con estudiantes de primaria de grado 6 (12-13 años), Nilsson (2014) realizó un experimento en el cual se propuso explorar aspectos críticos de la enseñanza de la probabilidad que se involucran en la producción y experimentación con datos. Entre los resultados de su investigación destaca varios desafíos para la enseñanza de la probabilidad a través de experimentaciones en el nivel de primaria. Por ejemplo, los estudiantes enfatizaron en frecuencias absolutas y relaciones parte-parte, la cual les hizo difícil comprender la condición de reemplazo de los elementos de la urna y uso de valores numéricos para estimar probabilidades. Concluye que para comparar y ver la relación entre probabilidad teórica y su estimación empírica, los estudiantes necesitan desarrollar una comprensión de la relación parte-todo, y deben iniciarse en la experimentación con una buena comprensión de proporciones y porcentajes. El estudio muestra que una enseñanza de la probabilidad basada en experimentación no es una empresa fácil –sobre todo si se realiza sin uso de tecnología- y que la experimentación en sí misma no necesariamente estimula a los estudiantes a reflexionar sobre el propósito de los datos, ver la información estadística como evidencia útil para conseguir una imagen de una población o hacer predicciones de probabilidad para situaciones aleatorias.

Por su parte, Konold et al. (2011) realizan una crítica a la práctica de muchos profesores de introducir a los estudiantes a la probabilidad teórica y su estimación frecuencial, con la expectativa que los estudiantes observen sin mayor dificultad, que las frecuencias convergen a la probabilidad conforme crece el número de observaciones o repeticiones de un experimento. Sin embargo, al igual que Nilsson (2014), advierten que la conexión entre ambos enfoques de la probabilidad representa desafíos conceptuales a muchos estudiantes. Señalan además algunos errores en materiales curriculares sobre el tema, en los cuales definen la “probabilidad experimental” (concepto que no existe en matemáticas) como la razón del número de ensayos favorables con el número total de ensayos, omitiendo la pieza crítica de información, que esta razón (frecuencia relativa), es una estimación de una probabilidad no conocida del resultado de interés.

### **3. Marco teórico**

En el enfoque clásico de la probabilidad, “la probabilidad es obtenida por la fracción de resultados favorables a un evento en un espacio muestral; bajo el supuesto implícito de igual probabilidad de todos los resultados individuales del espacio muestral” (Borovnick, Bentz & Kapadia 1991, p. 41). Por su parte, el enfoque frecuencial de la probabilidad está basado en la Ley de los Grandes Números, y define la probabilidad de un evento como la probabilidad estimada a partir de las frecuencias relativas de los resultados favorables en un experimento aleatorio que es repetido bajo las mismas condiciones un número suficientemente grande de veces.

En el contexto de la probabilidad, Henry (1997, p. 78) señala que “un modelo es un ente abstracto, simplificado e idealizado de un objeto real, un sistema de relaciones o un proceso evolutivo dentro de una descripción de la realidad”. El autor considera la simulación como un proceso de modelización que consta de diferentes etapas: a) descripción y simplificación de la realidad (partir de observaciones empíricas y decidir qué variables o hipótesis retener), b) construcción de un modelo, que puede ser la simulación; en esta etapa comienza la matematización y formalización (traducir las hipótesis de trabajo al modelo, en nuestro caso de probabilidad) c) trabajo con el modelo para obtener los resultados matemáticos o de simulación; d) validación de los resultados e interpretación en el contexto (si es necesario realizar ajustes del modelo de probabilidad).

Promodou (2012) basada en las componentes de una investigación estadística propone un marco para describir la relación bidireccional entre la estimación empírica de la probabilidad en la aproximación frecuencial y la probabilidad clásica (ver figura 1).

### **4. Metodología**

Para el desarrollo de la presente investigación, hemos elegido una metodología de tipo cualitativo (Miles y Huberman, 1994), en tanto se adapta bien para responder las preguntas que nos hemos planteado. Nos permite interpretar, describir y comprender significados que los estudiantes construyen cuando realizan conexiones conceptuales entre las aproximaciones clásica y frecuencial de la probabilidad en un ambiente de simulación computacional. El estudio tiene un alcance exploratorio y se desarrolló en el contexto de una aula de clase en la que el investigador funge también como docente.

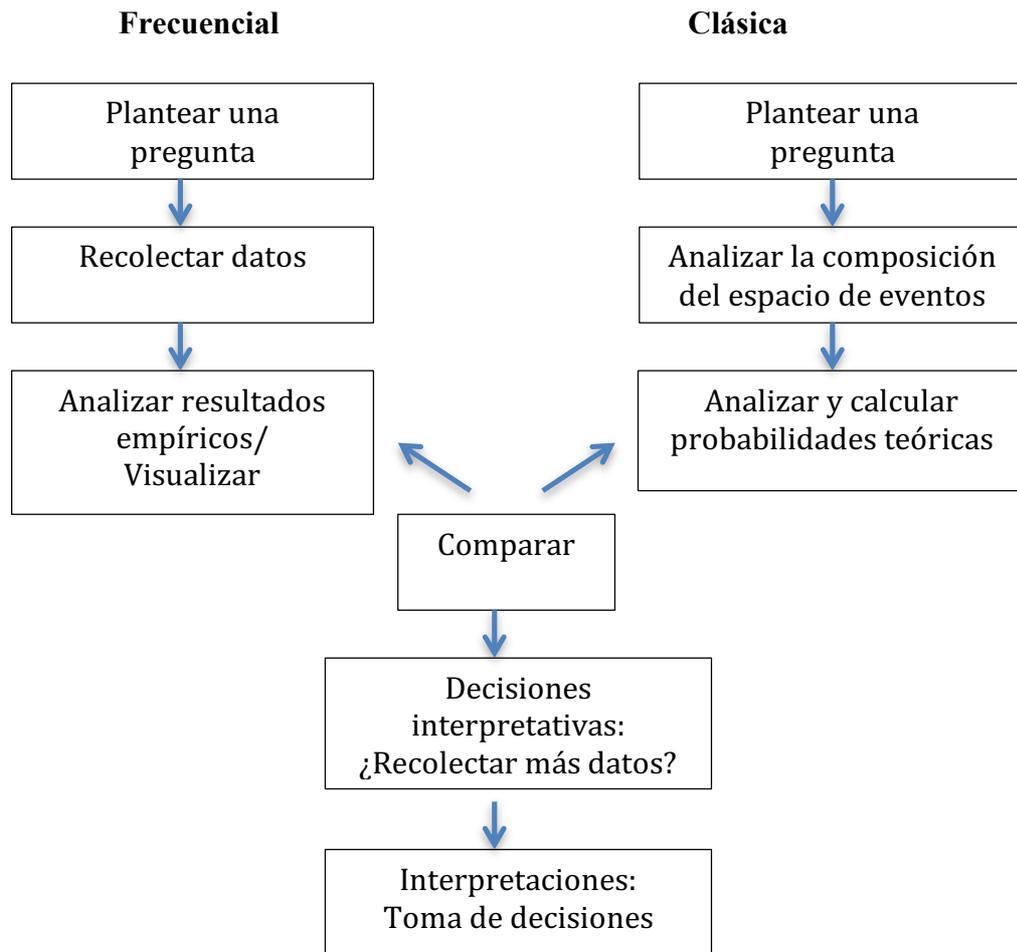


Figura 1: Un modelo para conectar las perspectivas frecuencial y clásica de probabilidad

El estudio se llevó a cabo con un grupo de 15 estudiantes (18-19 años) que cursaban la carrera de Estudios Internacionales en la Universidad Autónoma de Sinaloa, mientras tomaban el curso de probabilidad en el semestre II del ciclo 2015-2016. El curso fue diseñado para abordar el enfoque clásico y el enfoque frecuencial de la probabilidad en forma complementaria con apoyo del software TinkerPlots e iniciar a los estudiantes en el enfoque de modelos y simulación. Los antecedentes de los estudiantes en temas de probabilidad eran muy básicos en tanto apenas se iniciaba el curso, y únicamente tenían ideas superficiales sobre el enfoque clásico de la probabilidad adquiridos en el bachillerato.

Los resultados que presentamos en este artículo, fueron obtenidos de una actividad al inicio del curso. La primera parte de la actividad trataba de tres líneas de revisión de seguridad (A, B y C) en un aeropuerto a las que los pasajeros eran asignados en forma aleatoria antes de su ingreso a la sala de espera; la segunda parte consideraba además del cruce de los pasajeros por las líneas de seguridad, el cruce por un semáforo de aduanas para revisión de mercancías (ver Anexo).

La actividad fue desarrollada en dos sesiones de 1.5 horas cada una. La primera parte de la actividad tenía el propósito de explorar las conexiones desde el enfoque frecuencial hacia el enfoque clásico e investigar si los estudiantes podían aislar la variabilidad de los resultados e identificar el patrón de tendencia de las frecuencias,

relacionándolo en forma correcta con la cantidad de observaciones (cruces por las líneas de revisión), y si era posible además, que hicieran una buena estimación de las probabilidades del modelo implementado en el simulador que generaba los resultados (frecuencias), mismo que se encontraba oculto a la vista de los estudiantes y que debían identificar tal cual o uno equivalente. En este contexto, consideramos a un simulador construido por los estudiantes (por ejemplo, una ruleta con diferentes áreas) como un “modelo” que permite generar los datos proporcionados.

La segunda parte de la actividad tenía el propósito de explorar las conexiones que los estudiantes establecen desde el enfoque clásico hacia el enfoque frecuencial, para lo cual, a partir de probabilidades conocidas de asignación de pasajeros a las líneas de seguridad y cruce del semáforo, se solicitaba a los estudiantes la construcción de tres modelos de simulador equivalentes utilizando los diferentes dispositivos del software. Se pedía que realizaran simulaciones con los modelos construidos por ellos, contrastaran los resultados y validaran su equivalencia, tomando en consideración la similitud o diferencia entre las frecuencias observadas en cada uno de ellos; esto con la idea que comprendieran que modelos con estructuras diferentes pueden producir resultados similares. La idea era que relacionaran de forma correcta la relación entre el número de observaciones o corridas del modelo con la estabilidad de las frecuencias relativas, base para hacer buenas estimaciones de la probabilidad. La idea de equivalencia de modelos de simulador tenía la intención de que relacionaran patrones de variabilidad y tendencia equivalentes para modelos equivalentes. Adicionalmente, en esta segunda parte, se solicitaba que calcularan la probabilidad teórica de que un pasajero cruzara por la línea A y que encendiera la luz verde del semáforo, para lo cual debían aplicar la regla del producto de probabilidades. La comparación de la probabilidad teórica con las frecuencias generadas por los modelos era otro elemento de validación para la equivalencia de los modelos, además de la comprensión de las conexiones entre probabilidades teóricas y frecuencias relativas a medida que aumentaba el número de simulaciones.

Los instrumentos de recolección que se utilizaron para la recopilación y análisis de la información fueron una hoja de trabajo con la actividad y preguntas planteadas que requerían algunas justificación (ver Apéndice), y los archivos con la actividad desarrollada con el software. Al final de las actividades algunos estudiantes fueron entrevistados con la idea de tener más elementos sobre su razonamiento probabilístico. Para la selección de los entrevistados se tomó en cuenta las respuestas que dieron a los diferentes preguntas de la actividad y el trabajo realizado con el software, tales como el tipo de modelo de simulador construido y la cantidad de simulaciones realizadas, las cuales a juicio del investigador incluían aspectos conceptuales que debían ser investigados con mayor profundidad. Respecto al uso del software al comienzo de las actividades, cabe señalar que los estudiantes habían tenido una par de sesiones de familiarización con sus principales dispositivos, pues la investigación tuvo lugar apenas iniciando el curso de probabilidad.

## **5. Resultados y discusión**

### **5.1. Conexiones desde el enfoque frecuencial (datos) hacia el enfoque teórico (modelo)**

Para explorar estas conexiones, se presentaron a los estudiantes tres distribuciones de frecuencias que fueron obtenidas cuando pasaron 100, 500 y 1,000 pasajeros a

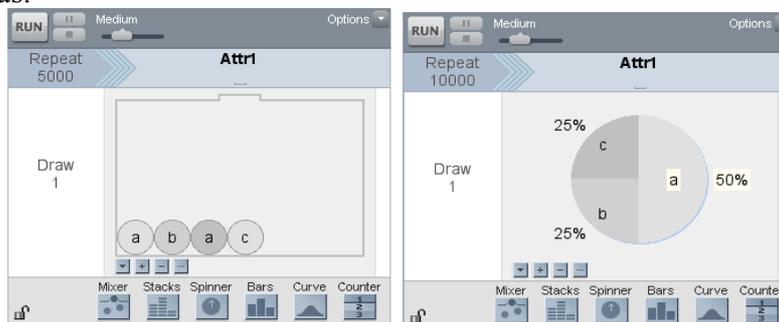
través de las líneas de revisión A, B y C (ver Apéndice). Se solicitaba a los estudiantes que con base en la información de las gráficas, estimaran la probabilidad con la que el sistema asigna a los pasajeros a cada una de las líneas de revisión. Una vez hecha la estimación, los estudiantes abrieron el archivo TinkerPlots que contenía un simulador (considerado en este trabajo como modelo) que generaba los resultados de las gráficas, el cual estaba oculto y los estudiantes no lo podían ver; enseguida hicieron varias corridas del simulador, para finalmente proponer un nuevo simulador (modelo) que generara resultados similares a los que generaba el oculto, es decir, un modelo de simulador equivalente.

En cuanto a la estimación de las probabilidades del modelo, todos los estudiantes identificaron que la línea A es mucho más probable que las líneas B y C, y que esta últimas tenían la misma probabilidad; pero sólo cuatro estudiantes lograron identificar la convergencia a los valores correctos (50%-25%-25%), los cuales tienen en común haber utilizado más de 5,000 corridas del modelo. Cabe señalar que para 5000 corridas se pueden presentar valores ligeramente diferentes a los anteriores, sin embargo, a los estudiantes les pareció que los valores correctos podrían ser 50%-25%-25%. Entre estos estudiantes se encuentran Rafael y Anaid, quienes mostraron una adecuada comprensión de la variabilidad como una propiedad intrínseca de los eventos aleatorios que guarda relación con la cantidad de observaciones o corridas de un modelo. A continuación mostramos sus respuestas:

Rafael: La línea A tiene una probabilidad cercana a 50%, las líneas B y C; parecen tener la misma probabilidad.

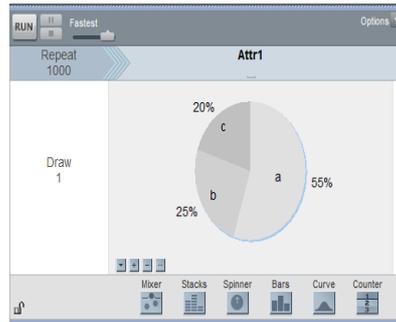
Anaid: De las tres gráficas se observa que, conforme pasan más pasajeros, las frecuencias son más precisas, y aparece el 50%-25%-25%, como una tendencia del 51%-25%-24% que se observa en la última gráfica. Tiene un rango de variabilidad porque es aleatoria.

En cuanto a la identificación del modelo de simulador que generaba los resultados (frecuencias), sólo cuatro fueron correctos (ver la Figuras 2a y 2b). Se identificaron dos tipos de modelos incorrectos: los que contemplaban probabilidades iguales a las tres líneas (equiprobabilidad) y los que contemplaban probabilidades cercanas a los resultados obtenidos en las corridas del modelo (ver Figura 2c). Los dispositivos utilizados para construir los modelos de simulador consistieron principalmente de ruletas y urnas.



a) Modelo correcto

b) Modelo correcto



c) Modelo Incorrecto

Figura 2: Algunos ejemplos de modelos (simulador) construidos por los estudiantes

Los estudiantes que generaron un modelo de simulador similar al 2c, no lograron superar la barrera que representa la variabilidad para identificar la convergencia de las frecuencias a las probabilidades, aún cuando el software les permitía hacer una gran cantidad de corridas para estabilizar las frecuencias e identificar su patrón de comportamiento de manera más precisa. Tal fue el caso de Katya quien hizo tres corridas de 1,000 casos y propuso un modelo aproximado al correcto. En una entrevista con el investigador respondió lo siguiente:

Investigador: ¿Cómo obtuviste el modelo 54% -25% -21%?

Katya: Hice tres corridas de 1,000 y observé que las frecuencias de la línea C eran las más bajas, así que puse un 4% debajo de 25%. Mi modelo considera una pequeña variación en cada resultado debido a que hay 50% -25% -25%.

De lo anterior puede verse que Katya es consciente que la tendencia de los resultados del simulador era 50%-25%-25%; sin embargo, la variabilidad que observa en los resultados le hacen proponer el modelo 54%-25%-21%, -como agregando un factor de variabilidad al modelo correcto-.

Por su parte, Naxely realizó 10,000 corridas del modelo de simulador construido por ella y respondió las siguientes preguntas:

Investigador: ¿Por qué utilizaste 10,000 corridas para explorar el modelo?

Naxely: Porque entre más repeticiones del modelo, los resultados tienen menos variabilidad.

Investigador: ¿En qué te basaste para proponer el modelo 50% -25% -25%?

Naxely: En la última gráfica que tiene frecuencias 50% -25% -25%.

Investigador: ¿Qué pasa si realizas 10,000 corridas del modelo en forma repetida?

Naxely: Se mantendrán las frecuencias cercanas. Por ejemplo las frecuencias de la línea A serán cercanas al 50% y las otras cercanas al 25% y 25%.

En el caso de Anaid utilizó 5,000 corridas del modelo:

Investigador: Para 5,000 corridas obtuviste los resultados 50%-25%-25%. ¿En qué te basaste para construir el modelo 50%-25%-25%?

Anaid: Cuando hice 1,000 corridas las frecuencias fueron 53% -23% -24%, entonces incrementé a 5,000 corridas y obtuve 50% -25% -25% casi siempre, así que considero que el modelo es 50% -25% -25%.

La respuesta Anaid muestra que comprende el efecto del número de corridas o cantidad de observaciones en la variabilidad de los resultados, por eso incrementó de

1,000 a 5,000 corridas del modelo del modelo de simulación. Reconoce la variabilidad que se presenta aún para 5,000 corridas, pero identifica claramente la tendencia hacia la cual los resultados se estabilizan.

De los resultados anteriores puede observarse que los estudiantes lograron identificar, en las exploraciones del simulador que generaba los resultados de las tres líneas de seguridad, que la línea A tenía mayor probabilidad y que las líneas B y C podrían tener probabilidades iguales, lo cual constituye una primera aproximación al modelo teórico del sistema de asignación de pasajeros para su revisión de seguridad. Los estudiantes que incrementaron el número de corridas del simulador, identificaron que la variabilidad de los resultados disminuía, y así lograron estimar las probabilidades verdaderas del modelo; es decir, relacionaron de manera correcta el tamaño de muestra número de corridas con la variabilidad y la precisión de la estimación de la probabilidades, lo cual representa una noción correcta de la Ley de los Grandes Números. Ello fue la base para que los estudiantes identificaran correctamente el modelo oculto o propusieran un modelo equivalente. Sin embargo, la variabilidad constituyó una barrera que muchos estudiantes no lograron pasar para estimar las probabilidades de forma precisa y plantear el modelo subyacente al sistema.

Los resultados anteriores comparen algunas similitudes con los obtenidos por Ireland y Watson (2009) quienes en una investigación sobre la construcción de conexiones entre el enfoque clásico y frecuencial de la probabilidad con estudiantes de grado 5 y 6, y en la cual se utilizó el mismo software TinkerPlots, encontraron que un tercio de los estudiantes tuvieron una buena apreciación de la probabilidad y algunos intuiciones sobre la variación involucrada en un escenario experimental. Un porcentaje cercano al 70% apreció correctamente el comportamiento de los resultados experimentales conforme se incrementa el tamaño de muestra.

## **5.2 Conexiones desde el enfoque teórico (modelo) hacia el enfoque frecuencial (datos)**

Para analizar estas conexiones, utilizamos una extensión de la actividad; además de la líneas de revisión de seguridad, los pasajeros debían pasar por un semáforo de aduana. Las probabilidades de las líneas de revisión (A, B, C) ahora eran conocidas (50%, 25% y 25% respectivamente), y consideramos que la probabilidad de que un pasajero al pasar por el semáforo encendiera la luz verde era del 80%. Los estudiantes debían construir tres modelos diferentes de simulador para representar y simular la situación previa. Además se solicitaba que verificaran si los tres modelos generaban los mismos resultados para 1,000 y 10,000 corridas, además que explicaran si los modelos eran equivalentes. Los estudiantes hicieron varias corridas del modelo para estimar la probabilidad de que un pasajero pasara a través de la línea A y se encendiera la luz verde del semáforo; esta estimación fue comparada con la probabilidad teórica calculada mediante la regla del producto de probabilidades.



Figura 3: Ejemplos de modelos correctos de simulador desarrollados por los estudiantes en la segunda parte de la actividad

A diferencia de la sesión previa, en la cual los estudiantes requerían identificar el modelo aleatorio a partir de las frecuencias que generaba, debían partir ahora de probabilidades conocidas. Por este motivo fue más sencillo que los estudiantes construyeran el modelo de simulador, pues 12 estudiantes lo construyeron correctamente utilizando de manera diferenciada los dispositivos del software. No obstante, que ahora el modelo implicaba la concatenación de dos eventos (el cruce por las líneas de revisión de seguridad y el semáforo de la aduana) (ver Figura 3).

Otro aspecto relevante en esta parte de la actividad fue la exploración de modelos equivalentes, pues la diversidad de dispositivos del software permite construir diferentes simuladores que generan resultados similares. Se buscaba con ello que los estudiantes observaran la variabilidad en la dirección del modelo hacia los datos y que haciendo abstracción de ella, identificaran la equivalencia de los modelos. Las respuestas de algunos estudiantes se muestran a continuación:

Katya: Aunque varían los resultados de los tres modelos, presentan un porcentaje mínimo de variación para 1,000 repeticiones. Para 10,000 corridas los resultados son casi idénticos, así que los modelos son equivalentes porque los resultados varían muy poco.

María José: Únicamente el modelo 1 y 2 parecen ser equivalentes, porque los resultados del tercero son muy diferentes.

Enseguida mostramos algunas respuestas de los estudiantes en la entrevista con el investigador (ver Figura 4):

Investigador: ¿Crees que los tres modelos producen los mismos resultados?

Fernanda: Sí, porque escribí los mismos datos (porcentajes) en cada sección de la ruleta y en cada barra, pero en las urnas no me permitió escribir los porcentajes [el software no dispone de esta opción] y entonces dudé si el modelo era correcto, pero cuando corrí los modelos me dieron los mismos resultados y

entonces me di cuenta que los tres modelos eran equivalentes.

Investigador: ¿Los modelos generaban exactamente los mismos resultados?

Fernanda: Había diferencia más o menos del 1%, muy poca variabilidad.

Investigador: Cuándo corriste el modelo 10,000 veces, ¿tuviste alguna duda si eran equivalentes los modelos?

Fernanda: Por el contrario, pude confirmar que eran equivalentes

De las respuestas de arriba podemos ver que María José identifica que uno de sus modelos de simulador no es equivalente a los otros dos, porque la distribución de frecuencias es diferente, mientras que Katya identifica que la variabilidad decrece con el tamaño de muestra (número de corridas) y que los tres modelos son equivalentes. Por su parte, Fernanda asegura que los modelos son equivalentes porque fue cuidadosa al escribir los datos, pero duda del modelo de urna hasta que realiza la simulación y obtiene resultados similares. Cuando Fernanda corrió el modelo 10,000 veces ya no tuvo dudas sobre la equivalencia de los modelos. La justificación ofrecida por estos estudiantes es una importante componente de razonamiento estadístico

En la última parte de la actividad los estudiantes estimaron la probabilidad que un pasajero pasara por la línea A y se encendiera la luz verde del semáforo, la frecuencia obtenida fue comparada con la probabilidad teórica calculada usando la regla del producto de probabilidades. Los resultados de Fernanda y Naxely, quienes usaron 1,000 y 10,000 corridas respectivamente, son presentados en la Figura 5.



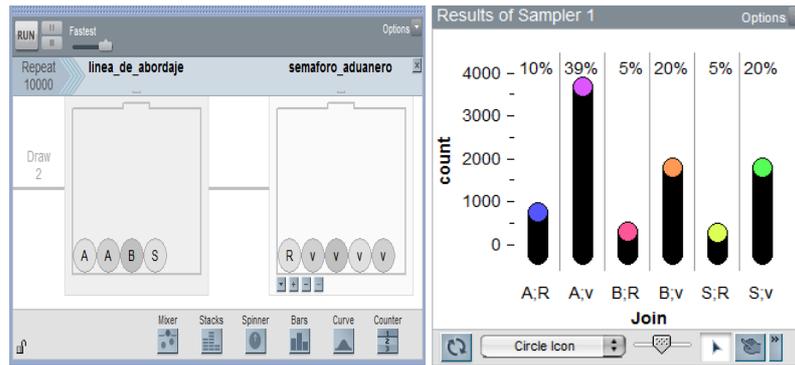
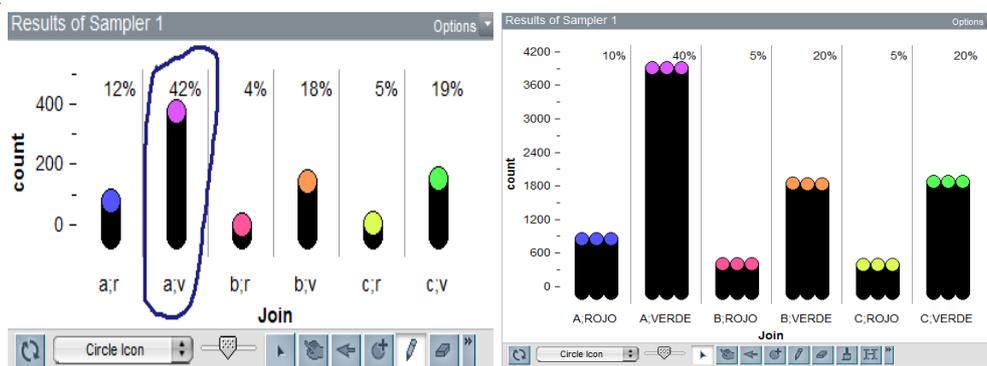


Figura 4: Tres modelos de simulador y sus respectivos resultados para 10,000 simulaciones desarrollados por Fernanda



Fernanda (1,000 corridas)

Naxely (10,000 corridas)

Figura 5: Resultados generados por los modelos de simulador construidos por Fernanda y Naxely.

Investigador: ¿Consideras que los tres modelos son equivalentes?

Naxely: Si, porque en cada modelo fui cuidadosa de introducir los datos correctos del problema.

Investigador: De qué otra forma verificaste que los modelos eran correctos?

Naxely: Haciendo varias corridas y observando que los resultados no cambiaban, o cambiaban muy poco, cuando tuve poca diferencia en los porcentajes de las tres gráficas.

Investigador: ¿No tuviste duda sobre los modelos cuando aparecían diferencias?

Naxely: Si, quizá cuando utilicé pocas repeticiones 100 o 1,000, observé que había diferencia, pero no mucha en las tres gráficas.

Investigador: ¿Cuándo usaste las formulas de probabilidad y comparaste qué observaste?

Naxely: Noté que los resultados eran iguales para el caso de 10,000 repeticiones

Investigador: ¿Tuviste confianza sobre que los modelos eran correctos?

Naxely: Si, fue cuando estuve segura que los modelos eran correctos, porque había verificado los datos de entrada al modelo cuidadosamente y además las corridas producían resultados similares en los tres modelos.

Las respuestas de Naxely muestran es consciente de la variabilidad intrínseca de los modelos que simulan situaciones aleatorias, de igual manera hace notar que esta decrece en la medida que se incrementa el número de simulaciones, lo que representa

una comprensión intuitiva de la ley de los grandes números. También da cuenta que modelos diferentes pueden generar resultados similares, y establece su equivalencia en función de la convergencia hacia valores estables generados por 10,000 repeticiones en el mismo sistema de representación de la simulación, aún más utiliza la probabilidad teórica como elemento de validación para asegurar que los modelos son correctos.

## **6. Conclusiones**

En la conexión desde el enfoque frecuencial hacia el enfoque clásico, la variabilidad de las gráficas fue uno de los principales obstáculos para identificar el modelo de simulador que generaba las frecuencias, particularmente en el caso de los estudiantes que utilizaron 1,000 o menos simulaciones. En este sentido, ajustar las observaciones empíricas en un modelo de trabajo –una de las etapas en el proceso de modelación que señala Henry (1997)- fue una tarea difícil para estos estudiantes. Sin embargo, quienes utilizaron más de 5,000 simulaciones del modelo, notaron que la variabilidad de las frecuencias decrecía conforme el número de simulaciones se incrementaba, a su vez, que las frecuencias convergían al valor de las probabilidades. Estos estudiantes tuvieron éxito en identificar el modelo de simulador correcto que generaba las frecuencias. Resultados similares fueron reportados en una investigación de Ireland y Watson (2009), en la cual muchos estudiantes pudieron conectar adecuadamente los resultados experimentales con el modelo teórico correspondiente, además que demostraron una comprensión intuitiva de la Ley de los Grandes Números; sin embargo, varios estudiantes no lograron comprender la relación entre frecuencias relativas y probabilidades y el efecto del tamaño de muestra en la variabilidad de la distribución de frecuencias, como sucedió con nuestros sujetos de estudio.

En la conexión desde el enfoque clásico hacia el enfoque frecuencial, los estudiantes modelaron y simularon un evento compuesto de forma exitosa en la mayoría de los casos, utilizando varios dispositivos aleatorios que proporciona el ambiente computacional utilizado. En este sentido, les resultó más fácil partir de probabilidades conocidas de los eventos para construir el modelo de un simulador, que identificarlas a partir de las frecuencias observadas en la primera parte de la actividad. La construcción de los tres modelos del mismo problema, sirvió como mecanismo de validación para muchos estudiantes, quienes notaron la similitud en el patrón de comportamiento de las distribuciones de frecuencias, especialmente cuando el número de simulaciones se incrementaba, desarrollando con ello una idea correcta del efecto del tamaño de muestra en la variabilidad de los resultados y la equivalencia de modelos para representar un fenómeno aleatorio. Similarmente, la comparación entre probabilidades calculadas con la fórmula de la regla del producto, sirvió como un mecanismo de validación de los modelos y conexión cercana entre frecuencias y probabilidades. Las características interactivas y flexibilidad del ambiente computacional para incrementar el número de simulaciones, repetirlas de forma reiterada y observar los resultados de manera inmediata, fue un elemento que ayudó a que muchos estudiantes observaran la tendencia de convergencia de las frecuencias hacia valores estables (probabilidades), estableciendo así una relación correcta entre tamaño de muestra, variabilidad, estimación de probabilidades y modelos equivalentes.

Los resultados obtenidos por Prodromou (2012) con profesores en formación, donde se utilizó la misma herramienta computacional guardan algunas similitudes pero también diferencias con los resultados de este estudio. En la dirección del enfoque

frecuencial hacia enfoque clásico, -contrario a lo que sucedió en nuestro caso-, la mayoría de los sujetos de estudio identificaron la correctamente la variabilidad en las frecuencias y su relación con el tamaño de muestra, por lo que no tuvieron dificultad para identificar la tendencia a la que convergían las frecuencias relativas. Al hacer la conexión desde el enfoque clásico al enfoque frecuencial, cuando compararon las frecuencias relativas con probabilidades teóricas, la mayoría logró distinguirlas. En nuestro caso, la mayoría de los sujetos de estudio también construyó sin dificultad el modelo de simulación e hizo conexiones correctas entre frecuencias relativas y probabilidades.

No obstante el carácter restringido de este estudio y su carácter exploratorio, los resultados sugieren que es posible que estudiantes con pocos antecedentes matemáticos y de probabilidad, -como es el caso de los estudiantes de ciencias sociales-, pueden realizar conexiones correctas entre el enfoque clásico y frecuencial de la probabilidad, así como apreciar el papel de los modelos en el pronóstico de probabilidades de un evento aleatorio. Se observa sin embargo, como señalan Konold et al (2011) y Nilsson (2014) que la conexión entre ambos enfoques de la probabilidad representa desafíos conceptuales para muchos estudiantes, por lo que se requieren estudios mucho más extensos sobre el origen de estas dificultades. El potencial que proporcionan los ambientes computacionales para la conexión entre estos ambos enfoques de la probabilidad, -estrategia de enseñanza ampliamente sugerida en la literatura de didáctica de la probabilidad- constituye una amplia línea de investigación que debe ser estudiada con profundidad.

## Referencias

- Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. En G. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning*, (pp. 15-37). Springer Science+Business Media.
- Biehler, R., & Prömmel, A. (2010). Developing students' computer-supported simulation and modelling competencies by means of carefully designed working environments. *Proceedings of ICOTS 8*, Ljubljana, Slovenia.
- Borovcnik, M. (2012). Multiple perspectives on the concept of conditional probability. *Avances de Investigación en Didáctica de la Matemática*, 2, 5-27.
- Borovcnik, M., Bentz, H. J., & Kapadia, R. (1991). Empirical research in understanding probability. En R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: probability in education* (pp. 73-105). Dordrecht: Kluwer.
- Chaput, B., Girard, J. C., & Henry, M. (2011). Frequentist approach: Modelling and simulation in statistics and probability teaching. En C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education: A joint ICMI/IASE study*, (pp. 85-95). Heidelberg: Springer
- Eichler, A., & Vogel, M. (2014). Three approaches for modelling situations with randomness. En E. J. Chernoff, & B. Sriraman (Eds.), *Probabilistic thinking: presenting plural perspectives* (pp. 75-100). Dordrecht: Springer Science+Business Media
- Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions?. *Educational Studies in Mathematics*, 15, 1-24.
- García, J. I., Medina, M., & Sánchez, E. (2014). Niveles de razonamiento de estudiantes de secundaria y Bachillerato en una situación-problema de probabilidad. *Avances de Investigación en Educación Matemática*, 6, 5-23.

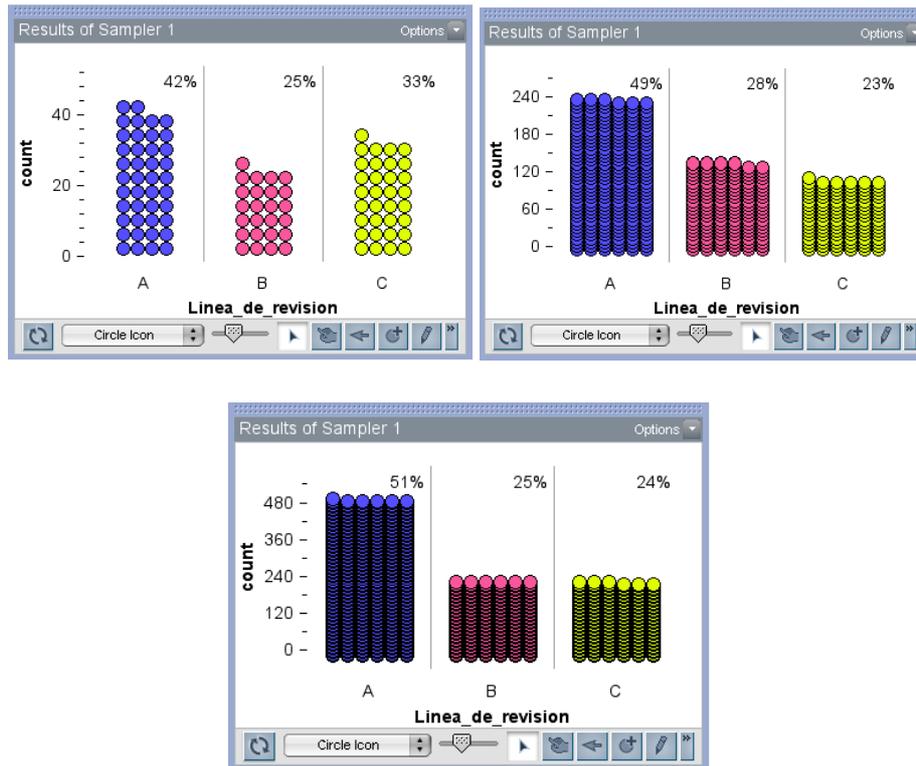
- Godino, J., Batanero, C., & Cañizares M. J., (1996). *Azar y probabilidad*. Madrid, España. Editorial Síntesis.
- Henry, M. (1997). Notion de modèle et modélisation en l'enseignement. En *Enseigner les probabilités au lycée* (pp. 77-84). Reims: Commission Inter-IREM
- Ireland, S. y Watson, J. (2009). Building a connection between experimental and theoretical aspects of probability. *International Electronic Journal of Mathematics Education*, 4(3), 339– 370.
- Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education*, 2(1). Recuperado de <http://escholarship.org/uc/item/38p7c94v>.
- Konold, C., & Miller, C. (2011). *TinkerPlots Version 2* [Computer Software]. Emeryville, CA: Key Curriculum Press.
- Konold, C., Madden, S., Pollatsek, A., Pfannkuch, M., Wild, C., Ziedins, I., Finzer, W., Horton, N., & Kazak, S. (2011). Conceptual challenges in coordinating theoretical and data-centered estimates of probability. *Mathematical Thinking and Learning*, 13, 68–86.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*. London: Sage.
- Nilsson, P. (2014). Experimentation in probability teaching and learning. En E.J. Chernoff, & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 509-532). Dordrecht: Springer Science+Business Media.
- Pfannkuch, M., & Ziedine, I. (2014). A modelling perspective on probability. En E.J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 101-116). Dordrecht: Springer Science+Business Media.
- Pratt, D. (2005). How do teachers foster students' understanding of probability? En G. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning* (pp. 171–190). New York: Springer.
- Pratt, D., Davies, N., & Connor, D. (2011). The role of technology in teaching and learning statistics. En C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics-Challenges for teaching and teacher education: A joint ICMI/IASE study*, (pp. 97-107). Heidelberg: Springer
- Prodromou, T. (2012). Connecting experimental probability and theoretical probability. *ZDM: The International Journal on Mathematics Education*, 44(7), 855-868.
- Prodromou, T. (2014). Developing a modelling approach to probability using computer-based simulations. En E.J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives*, (pp. 417-440). Dordrecht: Springer Science+Business Media.

## Apéndice

### Actividad

En un aeropuerto, un sistema electrónico asigna en forma aleatoria a los pasajeros que pasan por el control de seguridad a tres líneas para revisión (A, B, y C). Las líneas tienen diferente capacidad, y por consiguiente diferentes probabilidades de asignación.

- a) Las siguientes gráficas fueron obtenidas cuando habían pasado 100, 500 y 1000 pasajeros respectivamente. Con base en esta información, ¿puedes estimar la probabilidad con la cual el sistema electrónico está asignando a los pasajeros a cada una de las líneas de revisión?



- b) En el fichero *aeropuerto*, aparece oculto (no lo puedes ver) un modelo de simulador equivalente al que tiene el sistema electrónico para asignar pasajeros a las líneas de revisión. Realiza diversas series de simulaciones con el simulador y explora las frecuencias que se obtienen para cada línea de revisión; a continuación con base en los resultados, estima la probabilidad de asignación en cada línea. Finalmente describe un modelo de probabilidad que podría generar las frecuencias que se han observado.
- c) Las probabilidades de las líneas de revisión (A, B, C) son ahora conocidas, 0.50, 0.25 y 0.25 respectivamente. Considera que los pasajeros deben pasar además por un semáforo de aduana que es independiente de las líneas de revisión, el cual tiene una probabilidad de encender la luz verde de 0.80 y de encender la luz roja de 0.20.
- Construye tres modelos de simulador diferentes para representar y simular la situación previa.
  - Verifica si los tres modelos generan los mismos resultados para 1,000 repeticiones. Después incrementa a 10,000 repeticiones de los tres modelos y explica si los modelos son equivalentes.
  - Realiza el número de repeticiones que consideres necesario para obtener resultados precisos y estima la probabilidad de que un pasajero pase a través de la línea A y la luz del semáforo sea verde.

## Referencias a los autores

Santiago Inzunza Casares, Universidad Autónoma de Sinaloa (México)

## **Connections between classical and frequentist approaches to probability in a computer modelling environment**

Santiago Inzunza Cazares, Universidad Autónoma de Sinaloa (México)

In this paper we report results of a research with a group of 15 social science students taking an introductory course in probability. The course was designed to analyse the classical approach and the frequentist approach to probability with the support of a computer environment (TinkerPlots), and to initiate students in the modelling and simulation approach. The results presented were obtained from an activity at the beginning of the course (see Appendix). The first part of the activity consisted of three lines of review security (A, B and C) in an airport where passengers are assigned randomly before admission to the boarding room; the second part also considers the crossing of the passengers by a traffic light for inspection of tax payment.

The purpose of the first part was to explore the connections from the frequentist approach toward the classical approach and investigate whether students could isolate the variability of the results and identify the limit trend of frequencies presented in three graphs, in correctly relating it with the number observations or simulations. In addition, it was we were interested to know if they were able to provide accurate estimations of the probabilities for the model of simulator that generated the results of the graphs, which was hidden from the students and should be identified by them (or an equivalent model).

The purpose of the second part was exploring the connections from the classical approach to the frequentist approach, for which, the probabilities of assignment of passengers on security lines and traffic light are known, and the students should build three equivalent models of simulator. Students should perform simulations with the models built, and test the results and validate their equivalence, taking into account the similarity or difference between the frequencies observed in each model. The aim was that they relate correctly the relationship between the number of observations and model runs with the stability of the relative frequencies. Additionally, in this second part, the students were asked to estimate the theoretical probability that a passenger crossed the line A and to turn on the green light, where they should apply the product rule of probabilities.

The results show that the modelling process was not a complex activity for most students when the problem probabilities are known (from the classical approach towards the frequentist approach), but in the opposite direction (from the frequentist approach towards the classical approach), the variability of the results was a main obstacles to identify the underlying simulator model; particularly in those students who used 1,000 or fewer simulations. The students, who used 5,000 or more simulations, developed an intuitive understanding of the Law of Large Numbers and adjusted the frequencies to the correct theoretical model.

Despite the exploratory nature of the study, which took place just at the beginning the probability course, the results show that students can make adequate connections between the classical and the frequentist probability approaches when interacting with a computer environment. This perspective of teaching probability is a very broad line of research that should be continued.