Aleatoriedad, Modelización, Simulación

Carmen Batanero
Departamento de Didáctica de la Matemática, Universidad de Granada
Presentado en la X Jornadas sobre el Aprendizaje y la Enseñanza de las Matemáticas, Zaragoza, 2001

La aletoriedad es un modelo matemático que permite describir un gran número de fenómenos en forma más adecuada que otros modelos deterministas. A partir de dos ideas muy simples: repetibilidad de la situación en las mismas condiciones e independencia de resultados en dos repeticiones, surgen una serie de modelos de complejidad progresiva, que permiten resolver problemas de inferencia y de predicción en presencia de incertidumbre. Estas dos ideas son, en si mismas, una simplificación de la realidad y pueden ser más o menos aceptables en cada problema particular.
La probabilidad es un campo donde los modelos simples se componen entre si de una forma muy potente partiendo de unas pocas ideas estocásticas fundamentales. Aún así, la matemática de la probabilidad es muy compleja, más allá de unos pocos desarrollos elementales y sus resultados son bastante contraintuitivos. La simulación o sustitución de un experimento aleatorio por otro equivalente, como modelo pseudo-concreto de la situación modelizada, permite prescindir del aparato matemático para analizar situaciones estocásticas. Como recurso didáctico, puede ayudar a comprender la diferencia entre modelo y realidad y a mejorar las intuiciones sobre la aleatoriedad. Como contrapartida, la simulación no nos proporciona justificaciones ni demostraciones, que debemos buscar de nuevo en el modelo matemático.

Estadística y Modelización

Gran parte de la actividad matemática (y particularmente la estadística) puede ser descrita como proceso de modelización. En términos de Henry (1997) "*un modelo es una interpretación abstracta, simplificada e idealizada de un objeto del mundo real, de un sistema de relaciones o de un proceso evolutivo que surge de una descripción de la realidad* (pg. 78). La construcción de modelos, su comparación con la realidad, su perfeccionamiento progresivo intervienen en cada fase de la resolución de problemas estadísticos, no sólo en el análisis de datos en situaciones prácticas, sino también en el trabajo de desarrollo teórico. Un ejemplo notable de modelización estadística a partir de un problema práctico son las distribuciones de probabilidad, que permiten describir en forma sintética el comportamiento de las distribuciones empíricas de datos estadísticos y hacer predicciones sobre su comportamiento.

Por ejemplo, el modelo de la curva normal de media cero describe con notable precisión la distribución empírica de los errores al realizar medidas repetidas de una misma cantidad (o al tratar de fabricar un cierto producto con una medida fija). Una vez aceptado este modelo como una buena descripción de nuestros datos, basta obtener una estimación de la dispersión en los errores de medida (una estimación de la desviación típica σ en la población de los posibles errores) para tener el modelo completamente determinado. Las aplicaciones prácticas que podemos deducir seguidamente son muy variadas: previsión del número de defectos (piezas que se saldrán de los límites considerados como "normales"); decisión sobre límites de garantías; coste de la política de garantía; criterios para considerar que el proceso de producción está "controlado", etc.

Otro modelo muy potente, deducido esta vez a partir de un problema teórico en análisis multivariante, es representar una unidad estadística (por ejemplo, un alumno al que hemos dado un cuestionario) por un punto de un espacio vectorial, cuyas coordenadas son los valores de las diferentes variables incluidas en el estudio (las respuestas dadas a los diferentes ítems del cuestionario). Considerados los sujetos como "puntos" y las variables como "ejes" en dicho espacio vectorial, esta "metáfora" nos permite definir distancias para estudiar la proximidad entre dos puntos (análisis de aglomerados) o analizar la dimensión del espacio vectorial, tratando de reducirla (análisis factorial).

Estos dos ejemplos muestras que la modelización forma parte inseparable de la actividad estadística. Como indica McLean (2001), en este sentido, el razonamiento estadístico no es muy diferente del razonamiento científico o incluso del razonamiento "cotidiano". Constantemente usamos modelos, aunque no seamos conscientes de ello, y así nos formamos modelos de las personas de acuerdo a su raza, sexo, condición social o aspecto. Sin embargo, es poco habitual que la enseñanza de la estadística en secundaria o incluso en la universidad haga un énfasis en la modelización. Esto puede explicar el hecho de que los alumnos tengan dificultad en reconocer el modo particular de razonamiento estadístico, en el cual la modelización juega un papel muy importante.

Modelización y Enseñanza de la Probabilidad

Enseñar la probabilidad como un actividad de modelización y no como un conjunto de teoremas matemáticos que se deducen de una serie de axiomas no es una tarea sencilla. Para facilitar este trabajo a los profesores, la Comisión Inter- IREM para la enseñanza de la estadística y probabilidad organizó hace unos años un proceso de reflexión sobre el trabajo de modelización en la enseñanza de las probabilidades, produciendo algunas sugerencias que trataremos de reflejar a lo largo de este trabajo. Dentro de ellas, Dantal (1997) señala los siguientes pasos para la enseñanza de la probabilidad en secundaria con el enfoque recomendado en los nuevos currículos franceses:

1. Observación de la realidad
2. Descripción simplificada de la realidad
3. Construcción de un modelo
4. Trabajo matemático con el modelo
5. Interpretación de resultados en la realidad

Indica también que los profesores solemos tener prisa por llegar a los pasos 3 y 4 (los que podrían parecer "verdaderas matemáticas"), puesto que son los más sencillos de enseñar a nuestros alumnos. Sin embargo todas las etapas son igualmente importantes en el aprendizaje si queremos que realmente los alumnos lleguen a comprender la utilidad y la razón de ser de las matemáticas. Estas etapas son cuidadosamente analizadas por Coutinho (2001) en una experiencia de enseñanza de la actividad de modelización de situaciones aleatorias elementales.

En el paso 1 una primera dificultad es diferenciar, dentro de los fenómenos en los que el azar está presente, entre situación aleatoria y contingente. Una situación aleatoria es, por lo menos potencialmente, reproductible. Tiene diferentes resultados posibles, sin que sepamos con seguridad cuál será el que ocurrirá en una experiencia particular (por ejemplo, no sé con seguridad el tiempo que debo esperar en la parada a que aparezca el autobús). Pero puedo reproducir la situación en las mismas condiciones para analizar los diferentes resultados que se producen (puedo hacer un recuento del tiempo de espera al autobús). En una situación contingente (por ejemplo, el hecho de que yo me haya casado con una cierta persona) también una presencia del azar, pero no es reproductible en las mismas condiciones.

Una vez aceptada la aleatoriedad de la situación, en el paso 2 debemos realizar una descripción simplificada de la misma que nos permita pasar de la realidad observada (paso 1) a la construcción del modelo (paso 3). Para ello tomamos unos aspectos de ésta y prescindimos de otros. En el ejemplo del autobús deberemos decidir qué parada elegimos, si esperamos un autobús dado (el número 1) o si contamos el tiempo hasta que aparezca en la parada cualquier autobús. También si diferenciamos la hora del día o el día de la semana o no.

Todos estos puntos serán importantes para tratar de encontrar un modelo que nos describa el tiempo de espera y se requiere del juicio personal y del análisis cuidadoso de la situación para decidir cuáles de los aspectos de la realidad son los que realmente nos parecen importante para resolver el problema que tengamos entre manos y de cuáles podemos prescindir. Surge así la idea de "experimento aleatorio" y como en todo experimento científico Countinho (2001) sugiere que es importante establecer las condiciones de realización u observación, es decir, describir con detalle el protocolo experimental.

Al comenzar la construcción del modelo (punto 3) de nuevo se precisan una serie de decisiones: ¿aceptamos que el tiempo de llegada del próximo autobús es independiente del que acaba de llegar? ¿trataremos los tiempos como variable continua? ¿cuáles son otras variables, además del tiempo de espera que me podrían interesar en el trabajo con el modelo?

Una vez que hemos construido un modelo matemático para la situación (por ejemplo aceptamos que los tiempos de espera siguen una distribución uniforme entre un autobús y otro) y obtenidas las conclusiones a partir del modelo queda todavía la parte más importante: comparar estas conclusiones con el comportamiento real de la situación analizada y decidir que el modelo matemático nos proporciona una buena descripción de la realidad.

En esta breve descripción de los pasos esenciales en la actividad de modelización reconocemos el funcionamiento de la actividad de investigación y también de la forma en que aprendemos en la vida diaria. Nuestro conocimiento científico (y también el obtenido en la vida diaria) se compone de datos o hechos (la parte observable de nuestro universo) y teorías o modelos (que tratan de explicar y relacionar estos hechos entre si).

El propósito de construir un modelo es obtener una mejor comprensión de una parte de nuestro universo y, así, poder predecirla y si es posible controlarla. Un modelo no es "real", ni tampoco "verdadero";

en el mejor de los casos es consistente y concordante con las observaciones. Esto se olvida con facilidad y se suele confundir "modelo" y "realidad".

Es importante resaltar que el modelo mejor no es aquél que se aproxima mejor a los datos, sino que muchas veces el mejor modelo es uno que, siendo simple, nos proporcione una buena aproximación. Por ello mismo, incluso fenómenos que tiene naturaleza determinista y no siempre pueden medirse con precisión, se describen a veces mejor con un modelo probabilístico. En otros casos (Cabriá, 1997) el fenómeno tiene una naturaleza de "azar absoluto", como ocurre con la mecánica cuántica. por lo que las leyes que lo gobiernan tienen naturaleza aleatoria y el cálculo de probabilidades se convierte en un instrumento esencial para analizarlos.

La Aleatoriedad como Modelo Matemático

Una característica particular de la estadística y probabilidad es la existencia de problemas filosóficos ligados a la definición e interpretación de algunos de sus conceptos básicos. Aunque la teoría matemática correspondiente no presenta ninguna contradicción y está bien asentada a partir de la axiomática de Kolmogorov, sin embargo, al intentar aplicar en situaciones prácticas las ideas estocásticas fundamentales, nos encontramos con frecuencia dificultades y paradojas. Esto ocurre precisamente, porque olvidamos la diferencia entre el modelo y la realidad y porque no podemos encontrar en la matemática un criterio que aplicado mecánicamente nos lleve a la decisión de aceptar o rechazar un modelo para describir una cierta parcela de la realidad.

Como argumentamos en Batanero y Serrano (1995) la misma idea de aleatoriedad no tiene una definición simple y es difícil de aplicar en las situaciones prácticas. A lo largo de la historia ha recibido diferentes interpretaciones; ninguna satisfactoria y la definición estadística no siempre resulta convincente. Zabell (1992) diferencia dos aspectos distintos que, a veces, pueden no coincidir, en la idea de aleatoriedad

- el proceso de generación, que es lo que, matemáticamente se conoce como experimento aleatorio;
- el patrón de la secuencia aleatoria producida como consecuencia del experimento.

Una paradoja es que, por un lado, en la acepción de Kolmogorov (Fine, 1971) una secuencia sería aleatoria si no se puede codificar en forma más abreviada, por lo que la ausencia de patrones periódicos sería su característica esencial. Por otro, la visión estadística establece que un fenómeno es aleatorio si se ajusta al cálculo de probabilidades. Pero, aceptar incluso el modelo aleatorio más simple posible: experimento realizado repetida e independientemente, con sólo dos posibles resultados (sucesión de ensayos de Bernoulli) tiene como consecuencia que la secuencia aleatoria producida *debe* ajustarse a una serie de modelos deducidos del mismo. Consideremos el siguiente ítem propuesto a los niños por Green (1983):

Se pidió a algunos niños lanzar una moneda 150 veces. Algunos lo hicieron correctamente. Otros hicieron trampas. Anotaron con la letra C la aparición de una cara y con X una cruz. Estos son los resultados de Clara y Luisa:

Clara: c+c++cc++cc+c+c++c++c+ccc+++ccc++c++c+c+c++cc+ccc+
c+c+cc+++cc++c+c++cc+c++cc+c++cc+cc+c+++c++cc++c++
c+c+cc+c++cc+c+c++ccc+cc++c+c++cc+++c+++c+c++ccc++
Luisa: +cc+++c++++c+cc+++cc+cc+++cc+ccc+++c++++++c+c+c+c+
+++cccccc+ccc+c+cc+ccccc+ccc++ccc+c+ccccccccc++c+
ccccccc+++++cccc++c+c+cc+cc+cc++++++c+cc++ccc++ccc

¿Hicieron trampas Clara o Luisa? ¿Por qué?

Resolver este problema implica asumir un modelo de comportamiento de un proceso aleatorio y comparar los resultados de Daniel y Diana con los que esperaríamos si el modelo fuese cierto. Una idea intuitiva que aplican los alumnos a los que se proporciona este ítem es contar la frecuencia de caras y cruces. Puesto que las secuencias tienen una longitud de 150 lanzamientos, teóricamente esperaríamos una frecuencia aproximada de 75 caras y 75 cruces. En la Tabla 1 mostramos las frecuencias de caras y cruces obtenidas para Clara y Luisa. En principio nos parece que Luisa se desvía más de lo esperado en una secuencia aleatoria, pero quizás no tanto para poder acusarla de hacer trampas.

En lugar de contar en las secuencias los resultados del experimento simple consistente en lanzar una moneda, podríamos considerar que las niñas lanzan las monedas de 2 en 2 (por tanto hay 75 repeticiones del experimento). Hay cuatro resultados equiprobables al lanzar una moneda, por lo que, redondeando un poco, podemos esperar que aparezcan aproximadamente 18 veces cada uno de los posibles resultados.

Tabla 1. Frecuencias de caras y cruces en las secuencias

	C	+
Clara	72	78
Luisa	67	83
Teórica	75	75

En la Tabla 2 mostramos los resultados obtenidos para Clara y Luisa y ahora parece que es Clara la que más se aparta de lo esperado. Esto se confirma en la Tabla 3, cuando hacemos una prueba de todos los resultados en el experimento consistente en lanzar tres monedas. Ahora nuestras sospechas indican que es Clara la que hacía trampas (una prueba formal de contraste de Chi- cuadrado daría un valor significativo de la prueba de pares y triples para el caso de Clara, mientras que el resultado no sería significativo en ninguna de las tres pruebas para Luisa.

Tabla 2. Frecuencias de los diferentes resultados al lanzar la monedas 2 a 2

	CC	C+	+C	++
Clara	12	30	18	15
Luisa	25	21	12	17
Teórica	18	18	18	18

Tabla 2. Frecuencias de los diferentes resultados al lanzar las monedas 3 a 3

	CCC	CC+	C+C	C++	+CC	+C+	++C	+++
Clara	0	2	14	9	6	7	10	1
Luisa	8	11	6	4	6	4	5	6
Teórica	6	6	6	6	6	6	6	6

El ejemplo que hemos analizado nos muestra, por un lado, la dificultad de aplicar la idea de aleatoriedad en una situación práctica, por lo engañoso de nuestras intuiciones, incluso a pesar de haber estudiado estadística. Suelo proponer este problema a mis alumnos del curso de didáctica de la estadística (en la licenciatura de ciencias y técnicas estadística) y desde luego es frecuente que, al dar una primera solución al problema, muchos piensen que es Luisa la que hace trampas, ya que la frecuencia de caras y cruces difiere más en ella que en Clara, respecto a lo esperado.

Por otro, el ejemplo nos muestra que la aleatoriedad no admite una definición simple y ni siquiere puede describirse por un único modelo. En la situación propuesta necesitamos varios para decidir sobre la aleatoriedad de las dos secuencias. En realidad, la secuencia de Bernoulli contiene muchos modelos de probabilidad asociados, como la distribución binomial o geométrica. Aunque ya tenemos la sospecha de que Clara hace trampas, serían necesarias todavía nuevas pruebas para dejar a Luisa libre de sospechas.

El ejemplo también muestra con claridad la diferencia entre modelo y realidad. De hecho no esperamos que la sucesión de Clara o Luisa sean "exactamente iguales al modelo" y admitimos como "razonables" unas desviaciones respecto a lo esperado en el modelo, siempre que sean pequeñas.

Una cuarta consecuencia es la dualidad estadística- probabilidad. Sólo con la recogida y análisis de datos (análisis de los patrones en las secuencias de Clara y Luisa) podemos poner a prueba si el modelo probabilístico es adecuado para describir la situación (para decidir si las niñas hacen trampa o no).

Por último, observamos que la matemática no nos da ninguna regla inamovible para decidir si aceptamos o no el modelo como bueno. Incluso cuando llegásemos a realizar formalmente los contrastes de Chi cuadrado, lo más que obtendríamos es un valor de probabilidad p (la probabilidad de obtener el resultado dado) para cada una de las tres pruebas. Pero, el decidir cuan de pequeña debe ser esta probabilidad para asegurar que Clara hace trampas, es el riesgo que tengo que asumir, fruto de una decisión personal.

Y puesto que esta probabilidad nunca será igual a cero, siempre puedo equivocarme, incluso aunque el riesgo sea pequeño. Esta es precisamente una característica importante del razonamiento estadístico. La estadística nos proporciona unos modelos para controlar la incertidumbre y conocer los riesgos que asumimos de antemano, pero nunca puede anular la incertidumbre, que es una característica del mundo en el que nos desenvolvemos.

La Simulación: Entre Modelo Matemático y Realidad

En la sección anterior hemos resuelto el problema de decidir cuál de los dos niñas hace trampas con recursos matemáticos sencillos. En concreto hemos empleado las ideas de experimento y suceso aleatorio, equiprobabilidad de resultados, la regla de construcción del espacio muestral en un experimento compuesto de dos y tres experimentos simples, frecuencias esperadas y observadas en la repetición de un experimento.

Un estudio más completo de la actividad implicaría una matemática más compleja. Por ejemplo, el estudio del número de caras en n lanzamientos de la moneda llevaría a la distribución binomial $B(n, p)$. El efecto que sobre ella tendría sobre la secuencia el introducir un sesgo en la moneda, requeriría el estudio de la variación de los coeficientes binomiales en función de p. Los casos de p muy pequeña (distribución de Poisson) o n muy grande (teoremas de límite) quedaría seguramente fuera del alcance de un alumno de secundaria. No obstante la exploración de estas posibilidades y de sus aplicaciones a casos reales son los que dan mayor interés al estudio de la probabilidad.

Una pregunta que sin duda se plantea el profesor es si sería disponible realizar un estudio intuitivo de estos temas con ayuda del material concreto, calculadoras u ordenadores. Afortunadamente, contamos con la simulación., que para Heitele (1970) es en estadística algo parecido a lo que constituye el isomorfismo en otras ramas de las matemáticas.

En la simulación ponemos en correspondencia dos experimentos aleatorios diferentes, de modo que a cada suceso elemental del primer experimento le corresponda un suceso elemental del segundo y sólo uno, y los sucesos puestos en correspondencia en ambos experimentos sean equiprobables. Como indica Girard (1997) al trabajar mediante simulación estamos ya modelizando, porque debemos no sólo simplificar la realidad, sino fijar los aspectos de la misma que queremos simular y especificar unas hipótesis matemáticas sobre el fenómeno estudiado.

Por ejemplo, podemos "simular" el experimento aleatorio consistente en observar el sexo de un recién nacido mediante el experimento aleatorio consistente en lanza una moneda al aire. Ahora bien son muchos los aspectos que podríamos estudiar sobre un recién nacido, como el grupo sanguíneo, su peso o su raza, que no podrían simularse con el lanzamiento de la moneda. También hacemos una hipótesis (matemática) sobre equiprobabilidad para los dos sexos, independientemente de la raza, sexo y antecedentes familiares. Sólo una vez que hemos hecho estos supuestos, podremos comenzar el trabajo con la simulación.

Lo importante de ésta es que podemos operar y observar resultados del segundo experimento y utilizarlos para obtener información del primero. Por ejemplo, si queremos saber cual es la probabilidad que entre 100 recién nacidos hay más de un 60% de varones, podemos lanzar, por ejemplo 1000 veces 100 monedas al aire, estudiar en cada uno de los 1000 experimentos si hubo o no más de un 60% de nacimientos y obtener una estimación para la probabilidad pedida. La ventaja de la simulación es obvia, incluso en este ejemplo tan sencillo, pues permite condensar el experimento en un tiempo y espacio concreto.Vemos además que la simulación es, en si misma un modelo de la realidad simulada, puesto que simplifica la propia realidad y supone un trabajo de abstracción sobre la misma. Es además un modelo material (o bien algorítmico si usamos un simulador de una calculadora u ordenador), que nos permite reproducir físicamente el experimento y observarlo y por tanto, permite un trabajo intuitivo sobre el modelo sin necesidad del aparato matemático.

Entre el *dominio de la realidad* en que se encuentra la situación que queremos analizar y en la que interviene el azar y el *dominio teórico* donde, con ayuda de la matemática construimos un modelo teórico de probabilidad que debe, por un lado, simplificar la realidad y abstraer sólo sus aspectos esenciales y, por otro, ser útil para interpretar los caracteres retenidos en la modelización, Coutinho (2001) sitúa el *domio pseudo-concreto* en el que podríamos trabajar con los alumnos por medio de la simulación.

Mientras que en el dominio de la realidad se efectúa una acción o experiencia concreta y en el dominio teórico es característico la representación formal o simbólica, en el dominio pseudo concreto se opera mentalmente. En este dominio alumno ya ha salido de la realidad y trabaja con una situación abstracta idealizada. Por ejemplo, se imagina que está trabajando con dados perfectos, prescinde de las condiciones del lanzamiento. Al mismo tiempo conserva la denominación de las caras del dado real para nombrar los resultados del dado idealizado. El papel didáctico del modelo pseudo-concreto es inducir implícitamente el modelo teórico a los alumnos, incluso aunque su formulación matemática formalizada no sea posible (Henry, 1997).

Para presentar un modelo se pueden utilizar diversos tipos de lenguajes o representaciones. Incluso podemos usar palabras de la vida común, a las que atribuimos nuevos significados más precisos (como el caso que hemos descrito), usando la analogía.

Un caso muy interesante son los modelos de urna. En el caso de observar el sexo de un recién nacido podríamos simularlo sustituyéndolo por el experimento que consiste en elegir al azar con reemplazamiento una bola de una urna en la que introducimos dos bolas de diferente color para representar los dos sexos. Si queremos simular otro experimento aleatorio con dos sucesos en forma que sus probabilidades sean p y q ($p+q=1$), basta usar una urna en que se mantengan las proporciones p y q para los dos colores de bolas. Simular un experimento con r sucesos diferentes solo requiere usar bolas de r colores distintos, respetando las probabilidades correspondientes.

Cualquier problema probabilístico implica una serie de experimentos aleatorios compuestos de una determinada manera. Cada uno de estos experimentos puede ser "simulado" con un modelo de urnas convenientemente escogido (de una forma algo más compleja y usando una transformación inversa de la función de distribución, incluso los modelos continuos de probabilidad podrían simularse indirectamente, mediante este procedimiento).

Es decir, es posible asignar el experimento consistente en extraer al azar una bola de una urna con una cierta composición de bolas de colores. El experimento compuesto de varios experimentos simples se obtiene componiendo las "urnas" correspondientes a los experimentos simples (obteniendo una hiperurna) y la repetición del experimento global, junto con el análisis de los datos producidos permite una solución aproximada del problema.

En este sentido la urna con bolas de colores (fichas, tarjetas) es un "material universal", válido para estudiar cualquier problema o concepto probabilístico. Por ello la simulación proporciona un método "universal" para obtener una estimación de la solución de los problemas probabilísticos, que no tiene paralelo en otras ramas de la matemática. Además de la simulación con modelos de urnas y otros materiales manipulativos, las tablas de números aleatorios son también un instrumento de simulación universal, como hemos mostrado con algunos ejemplos presentados en nuestro libro *Azar y probabilidad* (Godino, Batanero y Cañizares, 1997).

Software para la Simulación en Internet

Aunque es importante que los alumnos realicen algunas actividades de simulación con apoyo de material manipulativo, como moneda, dados o ruletas, con tablas de números aleatorios y calculadoras, es realmente el ordenador el que proporciona una mayor potencia de simulación, construcción de modelos y exploración de los mismos. Las hojas electrónicas proporcionan generadores de números aleatorios, así como de valores de diferentes distribuciones de probabilidad, que pueden, una vez generados, ser analizados con ayuda de los recursos de cálculo y representación gráfica de la hoja.

Las posibilidades son todavía mayores con los programa estadísticos que suelen incluir módulos de estudio de las diferentes distribuciones de probabilidad con representación gráfica y cálculo de valores críticos y áreas bajo la función de densidad. Unido esto a la posibilidad de extracción de muestras de valores de estas distribuciones de tamaño dado, almacenamiento de las mismas en nuevos ficheros de datos, que pueden ser analizados, proporciona una herramienta muy interesante para la introducción de ideas de probabilidad y particularmente de inferencia.

Finalmente existen en Internet programas didácticos específicos para explorar conceptos estocásticos, desde los más elementales a los más avanzados, como, por ejemplo los procesos estocásticos. En lo que sigue mostramos algunos ejemplos de cómo podrían usarse algunos de los simuladores disponibles en Internet para complementar el estudio de la actividad que hemos analizado en la segunda sección. La versatilidad de estos programas hace que sean los mismos alumnos los que puedan formular y explorar sus propios problemas e introducirse de una forma experimental en la modelización estocástica de un gran número de situaciones reales.

Para comenzar, es posible que la solución dada en la sección 2 al problema sobre Clara y Luisa no haya convencido a algunos alumnos, quienes piensen que las rachas de Luisa eran demasiado largas y en una moneda real debiera haber mas alternancias entre caras y cruces. Podemos dar a los alumnos la oportunidad de experimentar con el simulador de lanzamientos de monedas de *Virtual Laboratories in Probability and Statistics* **(**http://www.math.uah.edu/stat/) que hemos reproducido en la Figura 1.

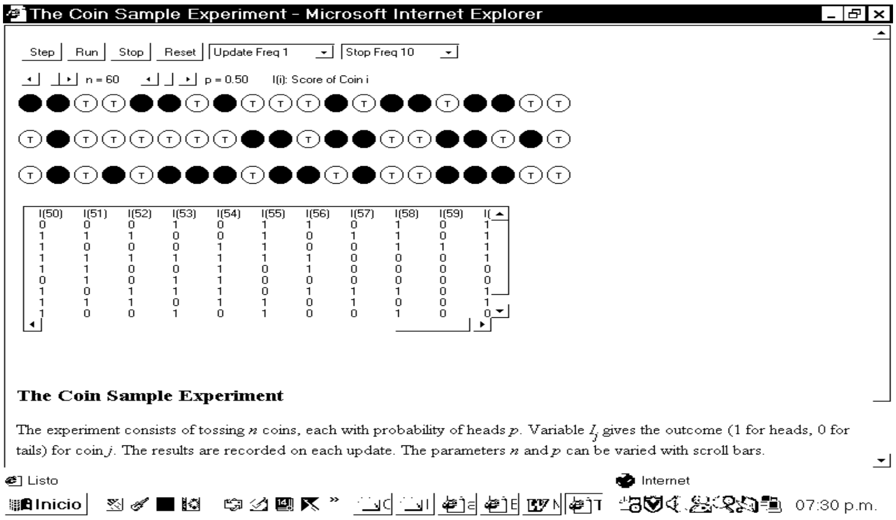
El programa permite elegir la longitud de la serie de monedas que queremos simular, y el número de simulaciones. Por ejemplo, podríamos simular 10 o 100 o 1000 secuencias aleatorias de longitud dada, que quedan almacenadas en la ventana y se pueden analizar posteriormente para estudiar la longitud de las rachas. Podríamos también observar el efecto de cambiar la probabilidad de cara, es decir, introducir un sesgo en la moneda.

En las Tablas 2 y 3 hemos analizado los lanzamientos de monedas de 2 en 2 o de 3 en 3. Podríamos, en general interesarnos por el lanzamiento de n monedas y estudiar la distribución del número de caras en n lanzamientos (modelo binomial).

El programa permite elegir la longitud de la serie de monedas que queremos simular, y el número de simulaciones. Por ejemplo, podríamos simular 10 o 100 o 1000 secuencias aleatorias de longitud dada, que quedan almacenadas en la ventana y se pueden analizar posteriormente para estudiar la longitud de las rachas. Podríamos también observar el efecto de cambiar la probabilidad de cara, es decir, introducir un sesgo en la moneda.

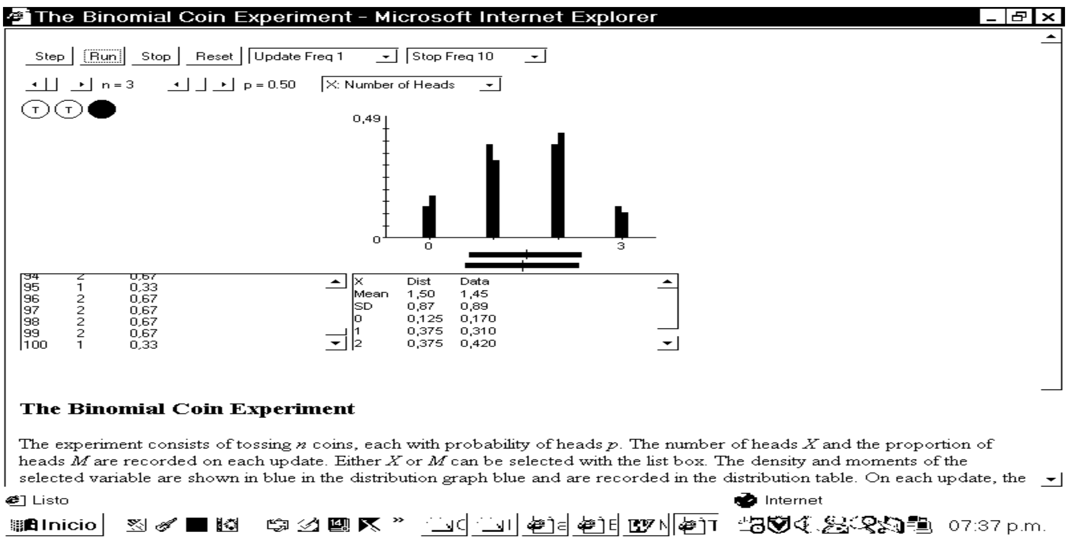
En las Tablas 2 y 3 hemos analizado los lanzamientos de monedas de 2 en 2 o de 3 en 3. Podríamos, en general interesarnos por el lanzamiento de n monedas y estudiar la distribución del número de caras en n lanzamientos (modelo binomial).

Figura 1. Simulación de experimentos con monedas



En la Figura 2 hemos reproducido la salida de otros de los programas del mismo laboratorio (Bernouilli trials: http://www.math.uah.edu/stat/bernoulli/index.html). En este caso podemos experimentar tanto con los valores de *p* como de *n* para ver el efecto sobre la forma de la distribución, los valores más probables, o la dispersión.

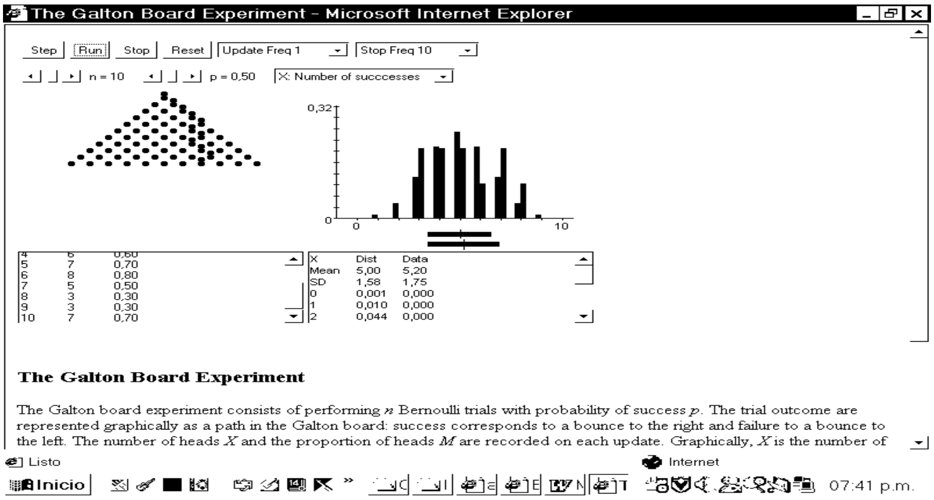
Figura 2. Simulación de experimentos binomiales



Un experimento relacionado con el del lanzamiento de monedas es el aparato de Galton (Figura 3). Al dejar caer una bola por dicho dispositivo, en cada uno de los pasos la bola puede ir a derecha o izquierda con igual probabilidad. Podemos, entonces sustituir una serie de 10 lanzamientos de una moneda por una trayectora de una bola en un aparato de Galton de 10 filas.

Nuestra intuición nos indica que en los orificios centrales llegarán más bolas y que la distribución obtenida será aproximadamente simétrica. Sin embargo, nuestra intuición nos juega malas pasadas en algunos otros puntos. Esperamos, por ejemplo, que la trayectoria que sigue la bola al caer corte con frecuencia el eje imaginario que pasa por el centro del dispositivo. Podemos utilizar el simulador disponible en (http://www.math.uah.edu/stat/bernoulli/index.html) que reproducimos en la Figura 3, para comprobar que esta intuición es errónea. Lo mas probable es que la trayectoria esté a uno de los dos lados del eje casi todo el tiempo. Asimismo podemos experimentar el efecto de añadir más o menos filas de canales, cambiar el valor de p (probabilidad de que la bola vaya a la derecha cada paso) y preguntarnos por qué la distribución que obtenemos se parece tanto a la obtenida en el punto anterior (distribución binomial). Y ¿qué relación tienen estos dos modelos con los coeficientes binomiales? ¿Cómo intervienen las combinaciones de m elementos tomados n a n en el cálculo de estas probabilidades y por qué?

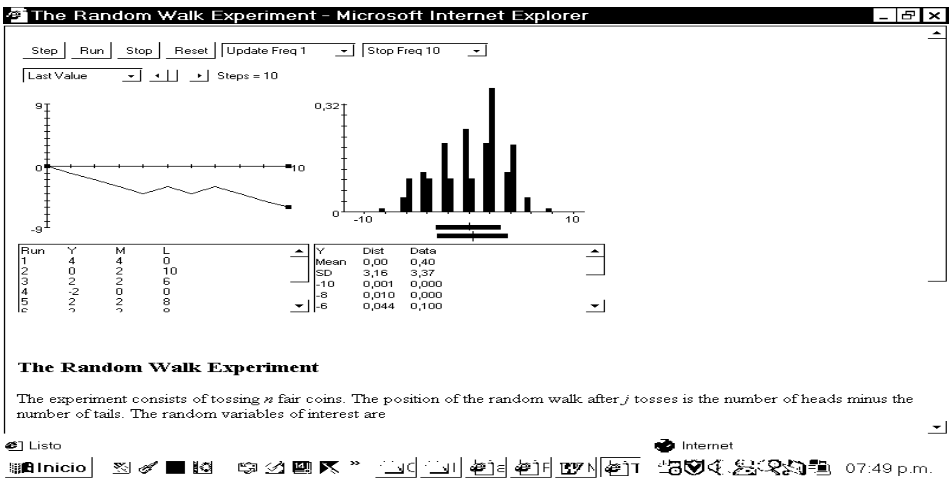
Figura 3. Simulador del aparato de Galton



La trayectoria seguida por una bola en el aparato de Galton nos remite a la caminata aleatoria sobre un plano (Figura 4). En vez de dejar caer una bola por un canal imaginamos que avanzamos de izquierda a derecha a lo largo de un eje horizontal. En cada paso, lanzamos una moneda para decidir si subimos o bajamos a lo largo de un eje vertical. Después de n pasos, ¿ a qué distancia nos encontramos del eje horizontal? ¿Cuántos cortes habremos dado al eje horizontal? ¿Cuánto pasos habrá que recorrer entre dos cortes sucesivos? Estas y otras preguntas similares con probabilidades iguales o diferentes de ir hacia arriba o hacia abajo en cada paso pueden explorarse en el simulador que hemos tomado del mismo laboratorio anterior.

Del problema de la caminata aleatoria podemos pasar al de la ruina del jugador. Si los valores positivos y negativos en el eje vertical representan las pérdidas y ganancias en un juego equitativo, la observación de la caminata nos indica que, en contra de nuestras expectativas es más probable que la mayor parte del tiempo de juego un jugador se mantenga ganando (y el otro perdiendo). Ahora bien, si la cantidad de dinero que uno de los jugadores quiere apostar es limitada (ponemos una cota a los valores negativos) ¿cuál es la probabilidad de ruina del jugador?

Figura 4. Simulador de la caminata aleatoria



En algunas situaciones estamos interesados por distribuciones aleatorias espaciales, por ejemplo, la distribución de conchas en la arena de la playa, de erratas en las páginas de un texto, de nidos o árboles en el campo. Si queremos explorar las distribuciones aleatorias de puntos en el espacio (modelo de Poisson) podemos utilizar el simulador que reproducimos en la Figura 5 que hemos obtenido del simulador disponible en (http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/bookapplets/chapter5/LondonBombs/LondonBombs.html)y analizar de nuevo el efecto de aumentar el número de puntos y la relación con la distribución binomial. Tanto en este caso como en los dos anteriores, para valores de *n* grandes aparecerá la característica forma de campana de la distribución normal y llevará a la pregunta de si es posible aproximar distribuciones discretas con una función continua y en qué sentido es posible usar esta aproximación. Otros programas disponibles en algunas de las direcciones que incluimos al finalizar este artículo proporcionan simuladores de la distribución normal, así como de la aproximación de las distribuciones binomial y Poisson mediante la misma.

Algunas precauciones

En este trabajo hemos tratado de mostrar el interés de la simulación en la enseñanza de la estadística y de la probabilidad, que es destacada entre otros en los trabajos de Biehler (1997). Como describe este último autor hay fenómenos que, por sus características espacio - temporales o de otra índole son difíciles de observar. Por medio de la simulación se construye, con el ordenador, un modelo simplificado de dicho fenómeno, eliminando las variables irrelevantes a su estructura, condensado en el tiempo y manipulable por el alumno. El uso de programas de simulación permite poner en manos del alumno un nuevo instrumento que hace posible la exploración y el descubrimiento de conceptos y principios que de otro modo serían mucho más abstractos, contribuyendo a paliar el problema de la falta de experiencia estocástica y a la mejora de la intuición probabilística que, en general, no se desarrollan espontáneamente (Fischbein, 1975).

Muchos problemas complejos se resuelven hoy día mediante simulación y el mostrar a los alumnos ejemplos sencillos de esta técnica puede servir para mostrar su aplicabilidad a campos y problemas reales.

En la enseñanza de la estocástica en secundaria la simulación cobra papel importante, ya que ayuda al alumno a conocer las diferencias entre la probabilidad experimental y la teórica.

No hay, sin embargo, que ser demasiado optimista respecto a la sencillez de la actividad de simulación. En el trabajo que hemos Citado de Countinho (2001) se observaron diferentes dificultades de los alumnos al proponérseles situaciones en que debian construir un modelo de urnas para simular ciertas experiencias concretas o juegos probabilísticos:

- Dificultades de manejo del software si el alumno no está familiarizado, por lo que se recomienda usar programas fácilmente manipulables que no añadan complejidad innecesaria a la actividad de simulación;
- Resistencia a usar la simulación y la aproximación experimental para resolver un problema de probabilidad en los casos en que es posible resolver el problema mediante cálculo directo;
- Dificultad en aceptar datos de simulaciones que no han llevado a cabo personalmente para obtener estimaciones de la probabilidad;
- Dificultad en diferenciar la estimación de la probabilidad que proporciona la simulación del verdadero valor teórico de la probabilidad (que solo es accesible por cálculo en los casos que sea posible).

Nosotros añadiríamos que la simulación y el enfoque frecuencial a la probabilidad, aunque proporciona una solución al problema no nos da la razón por la cual la solución es válida y por tanto carece de valor explicativo que solo puede obtenerse en el enfoque clásico y el cálculo formal de probabilidades.

No podemos contentarnos con que el alumno logre el paso del dominio de experiencia real al dominio pseudo-concreto, aunque este paso cumple una función didáctica importante y prepara al alumno para la comprensión del dominio formal en el que puede finalmente llevar a cabo una actividad matemática de formalización.

Referencias

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

Batanero, C., Serrano, L. y Green, D. R. (1998). Randomness, its meanings and implications for teaching

probability. *International Journal of Mathematics Education in Science and Technology,* 29 (1), 113-123.
Biehler, R. (1997). Software for learning and for doing statistics. *International Statistical Review,* 65(2), 167-190.
Cabriá, S. (1997). *Filosofía de la estadística.* Valencia.
Countinho, C. (2001). *Introduction aus situations aléatoires dés le Cxollége: de la modélisation à la simulation d'experiences de Bernoulli dans l'environment informatique Cabri-géomètre-II.* Tesis Doctoral. Universidad de Grénoble.
Dantal, B. (1997). Les enjeux de la modélisation en probabilité. En *Enseigner les probabilités au lycée* (pp. 57-59). Reims: Commission Inter-IREM.
Fischbein (1975). *The intuitive sources of probabilistic thinking in children.* Dordrecht: Reidel.
Girard, J. C. (1997). Modélisation, simulation et expérience aléatoire. En *Enseigner les probabilités au lycée* (pp. 73-76). Reims: Commission Inter-IREM.
Green, D. R (1983). A Survey of probabilistic concepts in 3.000 pupils aged 11 -16 years. En D. R. Grey et al. (Eds.), *Proceedings of the ICOTS* 1 University of Sheffield, 2, 766-783.
Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics*, 6, 187-205.

Henry, M. (1997). Notion de modéle et modélization en l'enseignement. En *Enseigner les probabilités au lycée* (pp. 77-84). Reims: Commission Inter-IREM.
McLean, A. (2001). Statistics in the catwalk. The importance of models in training researchers in statistics. In C. Batanero (Ed), *Training Researchers in the Use of Statistics.* Granada, Spain: International Association for Statistics Education and International Statistical Institute.

Zabell, S. L (1 992) Randomness and statistical applications. En F. Gordon and S. Gordon (Eds,), *Statistics for the XX1 Century.* The Mathematical Association of America.

Recursos en Internet para la Simulación

Gasp: The **G**lobally **A**ccessible**S**tatistical **P**rocedures, http://www.stat.sc.edu/rsrch/gasp/
Statiscope, http://www.df.lth.se/~mikaelb/statiscope/statiscope-enu.shtml.
Statistics Applet made for the WISE Project, http://www.grad.cgs.edu/wise/appletsf.shtml.
Statistics UCLA, http://www.stat.ucla.edu/textbook/.
Java Applets for Visualization of Statistical Concepts: http://www.kuleuven.ac.be/ucs/java/index.htm
Elementary Statistical Java Applets and Tools: http://www.stat.uiuc.edu/~stat100/cuwu/.
Rice Virtual Lab in Statistics , http://www.ruf.rice.edu/~lane/rvls.html
Virtual Laboratories in Probability and Statistics, Alabama http://www.math.uah.edu/stat/.
SticiGui: : Statistical Tools for Internet and Classroom Instruction with a Graphical User Interface, http://www.stat.berkeley.edu/~stark/SticiGui/index.htm.
STEPS - STatistical Education through Problem Solving, http://www.stats.gla.ac.uk/steps/home.html
WebStat, http://www.stat.sc.edu/webstat/.
Vista - The Visual Statsistical System, http://forrest.psych.unc.edu/research/index.html.
Zielman B. (University of), *Statistical Page*, http://huizen.dds.nl/~berrie/.
CAST Computer Assisted Statistics Teaching
http://www-ist.massey.ac.nz/CAST/CAST/CASTprog/HstartFiles/startCAST.html