

Experimento de Bernouilli

En este Tema 1 comenzaremos el uso del Simulador de distribuciones de probabilidad. Aunque parezca prematuro, puede ser el lugar oportuno para ir aclarando algunos conceptos referentes a la relación entre probabilidad y frecuencia. Especialmente está indicado por las ideas preconcebidas que podamos tener respecto a este tema. Todos tenemos la experiencia de encuestas de opinión que no aciertan, o pronósticos de lluvia que no se cumplen. Solemos exigir a la Teoría de la Probabilidad y a la Estadística resultados que no nos pueden dar. Con esta primera simulación intentaremos comprobar las propiedades ya sabidas de los experimentos aleatorios:

- Repetido un experimento aleatorio en las mismas condiciones, no tiene que dar los mismos resultados.
- Los resultados son imprevisibles.
- En general (no siempre), cada experimento suele ser independiente de los anteriores y no se ve influido por ellos.
- A pesar de lo afirmado, en largas series de repeticiones de un experimento aleatorio se observan unas ciertas regularidades.

Usaremos la hoja de cálculo Simulador, programada en LibreOffice Calc y Excel. En el momento de redactar este texto aún requiere algunos retoques, pero para la simulación que se propondrá funcionará perfectamente.

<http://www.hojamat.es/estadistica/tema1/tema1.htm>

Está alojada en

Versión Excel: <http://www.hojamat.es/estadistica/tema1/open/simulador.xslm>

Versión LibreOffice: <http://www.hojamat.es/estadistica/tema1/open/simulador.ods>

y la puedes descargar para tu uso.

Para entender algo mejor la relación entre probabilidad y frecuencia simularemos una distribución de Bernouilli, que es la más sencilla de todas. Consiste en imaginar un suceso, por ejemplo tirar un dado y que resulte un 6, y construir una variable cuyo valor sea 1 si ocurre ese suceso y 0 si no ocurre. Así, el 1 tendrá probabilidad $1/6$ de salir y el 0, suceso contrario, $5/6$, es decir, mucho más probable. Suponemos que en una serie de intentos, el resultado de cada uno no se ve influido por los anteriores. Por ahora nos quedaremos ahí y pondremos a funcionar nuestro simulador.

Primera simulación

Imagina que disponemos de una bola roja y dos blancas en una bolsa, y que extraemos varias veces una bola, la identificamos y la devolvemos a su sitio. Esperaremos que por cada vez que salga una roja aparecerán dos blancas, es decir, intuitivamente asignamos a la roja (que representaremos por el 1) una probabilidad de $1/3$ y a las blancas una de $2/3$. Si repetimos el

experimento muchas veces tendremos la expectativa de que la frecuencia de las blancas será el doble que la de la roja. Observa estas tiradas que hemos conseguido con el simulador:

0	0	0	0	0
1	1	1	0	1
1	1	1	1	0
1	0	0	1	0
0	0	0	0	0
1	0	0	0	0
1	0	0	0	0
0	1	0	1	1
0	1	1	0	1
0	0	0	0	0
1	1	0	0	0
0	1	0	0	0

En ella han aparecido 39 ceros y 21 unos. Como contiene 60 tiradas, habríamos esperado 20 unos (rojas) y 40 ceros (blancas). Existe, pues una diferencia de una unidad, o expresado en porcentaje del 1,67%.

La primera idea que debes tener respecto a la relación entre probabilidad (tu expectativa previa) y la frecuencia observada después de un experimento es que presentan valores bastante aproximados, pero no exactos, salvo casualidades.

Si no se tiene en cuenta esto, dejaremos de creer en los sondeos de opinión o los pronósticos meteorológicos, porque pensaremos que se equivocan. Se debe pensar siempre que estos experimentos miden nuestras expectativas, y no la realidad. ¿Se puede intentar disminuir la diferencia entre frecuencia y probabilidad? Pues sí y no. Seguimos hablando de expectativas: si aumentas el número de experimentos esperarás que el error en porcentaje disminuya, pero tampoco esto te da seguridad. Siempre medimos la esperanza y no la certeza. Inténtalo tú:

Abre el simulador y trabaja en su primera hoja *Simulación*. Sigue estos pasos:

Como tipo de simulación elige **Bernouilli** con el desplegable:

Tipo de simulación	
Bernouilli	▼

Como parámetro A, situado en el bloque *Otros parámetros*, y que contendrá la probabilidad, escribe esto: $=1/3$, y te dará un valor de 0,3333

Otros parámetros	
A	0,333
B	

En el ejemplo hemos simulado 60 repeticiones del experimento. Puedes aumentar a 30 filas por 9 columnas (o bien otro valores) para que nos resulten 270 intentos

Número de filas (de 1 a 1000)		30
Número de columnas (si procede) (de 1 a 12)		9

Con esto tienes definido un nuevo experimento con más repeticiones. Según el carácter de los fenómenos aleatorios, esperaremos un error menor, **pero eso nunca es seguro**.

Pulsa el botón *Simulador*



Te aparecerán muchos unos y ceros. A nosotros nos resultó esto:

0	0	0	1	0	0	0	0	1
0	0	0	0	0	0	0	0	0
0	0	1	1	0	0	1	0	0
0	0	0	1	0	0	0	1	1
0	0	1	0	0	1	0	0	1
0	1	1	0	0	1	0	0	1
0	1	0	1	0	0	0	0	0
1	1	0	0	0	0	0	0	1
1	0	0	0	0	0	0	1	0
0	0	1	0	0	1	0	0	0
1	0	0	1	1	1	0	1	0
1	0	0	0	0	0	0	0	1
1	1	0	1	1	0	0	0	0
0	0	0	0	0	1	0	0	0
0	0	0	0	0	1	0	0	0
0	1	0	0	0	1	0	1	0
0	1	0	0	0	0	1	0	0
0	1	1	0	0	1	0	0	0
0	1	1	1	0	0	1	0	0
1	0	1	0	0	0	1	0	0
1	0	1	0	1	0	0	0	0
0	0	0	0	1	0	1	1	1
0	0	0	0	1	0	1	0	1
1	1	0	0	0	0	1	0	0
1	1	1	1	0	0	1	1	0
0	1	1	0	0	0	0	0	0
0	0	1	0	0	1	0	1	1
1	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	1	0

No vas a contar todos esos ceros y unos. Pasa a la siguiente hoja, Estadísticos, y lee las frecuencias:

-0.5 - 0.5	184
0.5 - 1.5	86

En nuestro caso, en lugar de 180 y 90, que sería lo esperado, hemos obtenido 184 y 86, con un error de 4 unidades, 4,44% de error. **Luego el error ha aumentado. No te puedes fiar de los fenómenos aleatorios.**

Es muy probable que tú hayas obtenido mejor resultado. Lo hemos repetido y los errores obtenidos han sido 10, 1, 2, 5, 0, 8, 7, 6,... Como ves, muy variables y con poca fiabilidad. Es posible que tú también hayas obtenido resultados bastante dispares.

Si deseas introducirte en la Estadística debes aceptar este hecho. Sólo son seguros los resultados posteriores a un experimento, y no nuestras expectativas previas.

Seguimos insistiendo. Ese error hay que reducirlo. Para ello aumentaremos aún más el número de experimentos. Busca la casilla de Repeticiones de la simulación y marca 20 repeticiones, lo que equivale a 5400 experimentos.

Pulsa de nuevo en el botón Simulador y verás oscilar los datos 20 veces. Pasa a la hoja Estadísticos. Lee las frecuencias. A nosotros nos han resultado 1738 bolas rojas y 3662 blancas, en lugar de las esperadas 1800 y 3600, con un error de 62 bolas, que representa un 1,1% de error. Es una mejora, pero no espectacular. Como regla empírica se suele tomar un 3% de error esperado en 1000 experimentos (más adelante se fundamentará esto).

Hemos aprendido que cuando se realiza un experimento aleatorio el error que hay que medir es el relativo, el porcentaje, que calculas dividiendo el error absoluto (aquí 62 bolas) entre el total de experimentos (540) y pasarlo, si se desea, a porcentaje.

Hemos repetido el experimento de los 5400 intentos y nos han aparecido estos errores:

0,02%, 0,07%, 1,02%, 0,46%

En conjunto es más fiable que el primer experimento.

Si deseas simular experimentos en los que confíes que se dé una proximidad entre probabilidad y frecuencia, deberás aumentar el número de experimentos (con el consiguiente gasto. Por ello son populares los sondeos con 1000, 5000 o 10000 encuestados, pero no más)

En este ejemplo de Bernouilli conocemos la probabilidad, pero en un sondeo o una previsión meteorológica no lo sabemos y tendremos que manejarnos con cotas de error (ya lo irás viendo)

Por curiosidad, hemos aumentado el número de repeticiones a 200 (tarda un poco), es decir, 54000 experimentos, consiguiendo un error de 0,2%.

Si deseas afinar más tus resultados deberás aumentar el número de repeticiones, con el consiguiente gasto en tiempo y dinero.

NOTA: En estas simulaciones estamos juzgando la exactitud de la función ALEATORIO de las hojas de cálculo. Si esta no está bien programada sufriremos errores sistemáticos, pero como nuestro objetivo es aclarar conceptos, no parece muy grave.

Segunda simulación

Cuando se interviene en procesos aleatorios en la vida real llaman mucho la atención las rachas: tener cuatro varones seguidos, sacar un seis en el parchís tres veces o que veas en la calle varios coches aparcados de la misma marca. Simularemos experimentos de Bernouilli para descubrirlas. Como todo esto es aleatorio, nunca sabremos si aparecerán o no.

Tener cuatro varones seguidos

Ya sabes cómo se programa el simulador: como parámetro A escribiremos $=1/2$, que es la probabilidad aproximada de nacer varón. Luego concretaremos una sola columna para leer mejor y, por ejemplo, 200 filas. ¿Has obtenido una racha de cuatro unos (también valdría de ceros, pues tienen la misma posibilidad)?

En nuestro intento han aparecido una racha de cuatro unos y otra de cinco muy cercana a la anterior:

1
0
1
1
1
1
0
0
0
1
1
1
1
1
0
n

Esto significa que no es un suceso tan raro. Aparecieron en 200 intentos más rachas de cuatro elementos (ceros o unos) e incluso una racha de ocho unos.

Tercera simulación

Recordarás la Ley de Murphy en sus distintas variantes: “Si algo puede salir mal, probablemente saldrá mal”. En efecto, a veces falla todo lo que era susceptible de fallo. Por eso, en instalaciones en las que la seguridad es prioritaria se suelen duplicar o triplicar los equipos. Imagina que en un quirófano existen tres alimentadores de energía, cada uno con una fiabilidad del 90%. ¿Podrán fallar todos a la vez?

Simularemos Bernouilli con parámetro A igual a 0,1 (probabilidad de fallo) y planificaremos tres columnas, que representarán a los tres equipos, y 1000 filas. ¿Aparecerá un fallo triple? Lo sabremos si aparece una fila con tres unos. Lo hemos intentado y logrado al primer intento:

0	0	0
1	1	1
0	1	0
n	1	n

Los que entendéis la probabilidad habréis comprendido que se espera un fallo cada mil intentos, porque $0,1 * 0,1 * 0,1 = 0,001$

Podéis plantearos otras simulaciones de sucesos cuya probabilidad podáis estimar y comprobar después si la simulación se acerca a las expectativas previas.