Distribución uniforme

En otro tema posterior de este curso podrás estudiar la distribución uniforme. Por ahora te basta con la idea de que representa experimentos aleatorios en los que todos los elementos presentan una misma probabilidad de ocurrir, como tiradas de dados o las loterías. Se suele distinguir entre distribución uniforme discreta, cuando sólo existe un número finito de posibilidades, o continua, cuando pueden aparecer infinitos sucesos, o al menos, tantos, que sea preferible tratarlos como infinitos. Con eso podemos comenzar a realizar algunas simulaciones.

La lotería primitiva

En esta lotería española pueden salir los números del 1 al 49, y los apostadores eligen entre varios tipos de apuestas. En el sorteo aparecerán seis números y se gana más o menos dinero según las coincidencias entre nuestra apuesta y el resultado del sorteo. Suponemos la apuesta más sencilla, compuesta de seis números.

Sorteo 100, 15/12/2016

Combinación ganadora

16 17 21 26 31 38

Nuestro objetivo es el de comprobar la dificultad de acertar varios resultados, que suelen tenerse en cuenta entre tres o seis aciertos, cambiando la cuantía del premio según el número de aciertos.

Abrimos el simulador desde las direcciones

http://www.hojamat.es/estadistica/tema1/open/simulador.ods (versión LibreOffice Calc)

http://www.hojamat.es/estadistica/tema1/open/simulador.xlsm (versión Excel)

Le concretamos los siguientes datos:

Simulador — A. Roldán – Versión 2.2 – Año 2016

Repeticiones de la simulación	1
Número de filas (de 1 a 1000)	6
Número de columnas (si procede) (de 1 a 12)	1

Tipo de simulación	Muestreo sin
Decimal – Entero	Entero

Extremos (si procede)	
Mínimo	1
Máximo	49

Parámetros	
Media	200
Sigma	8

- Máximo y mínimo, de 1 a 49, que son los números del sorteo
- Distribución Muestreo sin y con enteros
- Seis filas y una columna, para simular bien el sorteo
- Criterios Máximo-Mínimo

Concretado de esta forma, se simulará el resultado de un sorteo sin repeticiones, en el que aparecerán seis números distintos del 1 al 49. Para hacer ver la dificultad de acertar, hemos adjuntado nuestra apuesta a la izquierda de la simulación:

Apuesta	Resultado
3	10
9	18
12	21
22	6
29	7
40	23

Ahora se trata de usar reiteradamente el botón de simulación y comprobar cuántos números hemos acertado. Nosotros hemos estado un buen rato jugando a simular y siempre hemos obtenido menos de tres aciertos. Valga esta experiencia para enseñarnos a jugar con prudencia.

El seis triple

En muchos juegos de mesa, sacar un 6 tres veces seguidas es un acontecimiento importante. Según la teoría, la probabilidad de que esto ocurra es de 1/(6*6*6)=1/216. Podemos esperar que de cada 216 tiradas triples que efectuemos, una de ellas presente 6-6-6. Podíamos simularlo para ver qué ocurre.

Para ello concretamos el Simulador:

- Distribución uniforme entera
- Mínimo 1 y máximo 6
- 200 filas y 3 columnas (las filas son opcionales. Podrían ser 216 o el doble)
- Criterio Máximo-mínimo

Simulador	A. Roldán – Versión 2.2 – Año 2016		
	Repeticiones de la simulación		1
	Número de filas (de 1 a 1000)		200
	Número de columnas (si procede) (de 1 a 12)		3
Tipo de simulación		Extremos (si procede)	
Uniforme		Mínimo	1
		Máximo	6
Decimal – Entero		Parámetros	
Entero		Media	200
		Sigma	8
Criterios		Otros parámetros	
Máximo-Mínimo			

Así, a simple vista percibiremos bien si resulta un 6 triple.

Nosotros hemos efectuado dos simulaciones sin obtener el seis triple, que sólo ha aparecido en el tercer intento:

[illegible]	1	6		
4	3	4		
6	6	6		1
2	5	3		
3	3	5		
5	5	1		

¿Por qué, entonces, creemos que es más fácil conseguirlo? Todo el que ha jugado recuerda haber enlazado tres seis seguidos. Ocurre que recordamos mejor las jugadas favorables, y no las desfavorables, como cuando en el juego de la Oca nos toca la cárcel o la muerte, o las veces que necesitamos un cinco para salir en el parchís y no nos sale en muchas jugadas.

Uso de la distribución uniforme para probabilidades dadas por una tabla

Con el Simulador y el uso de la distribución uniforme continua (con decimales) se pueden simular probabilidades dadas por una tabla. Por ejemplo, supongamos que disponemos de tres bolas rojas, dos verdes y dos amarillas para efectuar experimentos aleatorios. Lo normal sería introducirlas en una bolsa e ir sacando una a una con reposición. Esto, en las aulas, podría ser divertido y algo caótico. Una alternativa es simularlo con ordenador.

Partimos de la tabla de probabilidades

Roja	3/7
Verde	2/7
Amarilla	2/7

En primer lugar, la cambiamos a probabilidades acumuladas, es decir, que cada una sea la actual más todas las anteriores. Así:

Roja	3/7
Verde	5/7
Amarilla	7/7

Esto es para organizar unas desigualdades más adelante.

El proceso es como sigue:

En el simulador organizamos una distribución uniforme con decimales, mínimo 0 y máximo 1 (los 7/7). Como criterio le damos que use el Máximo y mínimo, y fijamos, por ejemplo, 100 filas y una columna.En el parámetro A se ha escrito 3/7 y en el B 5/7 (esto no es obligatorio).

Copiamos la imagen de la hoja:

Simulador	A. Roldán – Versión 2.2 – Año 2016		
	Repeticiones de la simulación		1
	Número de filas		100
	(de 1 a 1000)		
	Número de columnas (si procede)		1
	(de 1 a 12)		
Tipo de simulación		**Extremos (si procede)**	
Uniforme			
		Mínimo	0
		Máximo	1
		Parámetros	
Decimal – Entero			
Decimal		**Media**	10
		Sigma	2
Criterios			
Máximo-Mínimo		**Otros parámetros**	
		A	0,429
		B	0,71428571

Si iniciamos la simulación obtendremos una columna con 100 números entre 0 y 1 expresados con decimales:

0,01
0,18
0,34
0,64
0,9
0,83
0,35
0,52
0,95
0,04
0,74
0,68
0,7
0,49
0,65
0,14

Estos números (y esto no lo consigue la herramienta) debemos convertirlos en Rojo, Verde y Amarillo. Para eso añadimos una fórmula a la derecha de cada número, de forma que si este no llega a 3/7, es que se trata de la Roja, y en caso contrario, si tampoco llega a 5/7 (probabilidad acumulada), se trata de la Verde, y si no, es Amarilla. La primera vez que se organiza esto resulta complicado, pero hay que insistir y buscar otros ejemplos. Esta fórmula, en nuestro caso, es:

=SI(G5<E$21;"Roja";SI(G5<E$22;"Verde";"Amarilla"))

(hemos copiado la de G5)

En ella se ve claramente el mecanismo: Si el número contenido en G5 es menor que el parámetro A (3/7), se trata de Roja. Si no es Verde o Amarilla según quede menor o mayor que el otro parámetro. El resultado lo tienes en la imagen:

0,01		Roja				
0,18		Roja		38	0,38	0,43
0,34		Roja		33	0,33	0,29
0,64		Verde		29	0,29	0,29
0,9		Amarilla				
0,83		Amarilla		100		
0,35		Roja				
0,52		Verde				
0,95		Amarilla				
0,04		Roja				
0,74		Amarilla				
0,68		Verde				

Los números que no llegan a 3/7 producen una roja, los siguientes hasta 5/7 la verde y el resto amarilla. Para ver si la simulación es razonable hemos contado los colores, resultando 38 rojas, 23 verdes y 29 amarillas. A la derecha ves la comparación con la teoría: las amarillas han resultado muy aproximadas y las otras con errores de centésimas, luego aceptamos el procedimiento.

Media y desviación típica

En teoría (no en una simulación concreta), si la distribución uniforme es continua entre los extremos a y b, su media es (a+b)/2 y su varianza (b-a)^2/12. Con el simulador podríamos aproximarnos a esos valores. Construimos un ejemplo:

Estimar la media y desviación típica de una distribución uniforme generada entre los extremos 10 y 20

Concretaremos distribución uniforme con decimales, extremos 10 y 20, y para obtener una buena aproximación, usaremos 100 repeticiones del experimento, con una sola columna y 20 filas (todo esto ha sido opcional. Puedes cambiarlo)

Simulador		A. Roldán – Versión 2.2 – Año 2016	
	Repeticiones de la simulación		1
	Número de filas		100
	(de 1 a 1000)		
	Número de columnas (si procede)		20
	(de 1 a 12)		
Tipo de simulación		**Extremos (si procede)**	
Uniforme			
		Mínimo	10
		Máximo	20
		Parámetros	
Decimal – Entero			
Decimal		**Media**	10
		Sigma	2
Criterios			
Máximo-Mínimo		**Otros parámetros**	

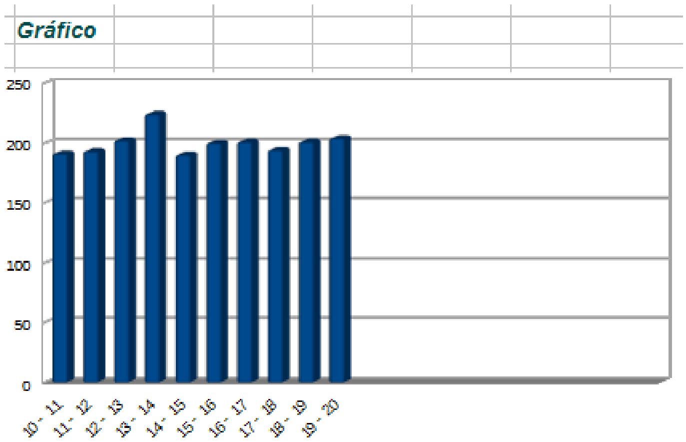
En este caso también concretaremos la opción de usar máximo y mínimo y fijaremos los intervalos en 10:

De esta forma generaremos 2000 datos uniformes entre 10 y 20. Su media debería ser (10+20)/2=15 y su varianza (20-10)^2/12=8,33. La desviación típica esperada será la raíz cuadrada de esta cantidad, 2,88.

Iniciamos la simulación y obtenemos:

Estadísticos de momentos			Tabla de frecuencias	
Media	15,0210			
Desviación típica	2,8839		10 - 11	191
Asimetría	0,0147		11 - 12	193
Curtosis	-1,1938		12 - 13	202
			13 - 14	224
			14 - 15	190
			15 - 16	200
			16 - 17	201
			17 - 18	194
			18 - 19	201
			19 - 20	204

Ha resultado bastante aproximado: media 15,02 y desviación típica 2,8839. Si repites el experimento no es fácil que consigas mejores aproximaciones. Ha sido una casualidad. En el gráfico se percibe bastante bien la uniformidad:



Es buena práctica, en cursos elementales de Estadística, construir estas simulaciones, porque ayudan a fijar conceptos.