Aproximación frecuencialista de la Probabilidad

Apellidos, nombre	Boigues Planes, Francisco José[1] (fraboipl@mat.upv.es) Estruch Fuster, Vicente D.[1] (vdestruc@mat.upv.es)
Departamento	[1]Matemática Aplicada
Centro	Universitat Politècnica de València

1 Resumen de las ideas clave

Presentamos un enfoque de la **probabilidad** que resalte la relación entre la teoría y la experiencia: el enfoque frecuencialista. De esta manera se reforzará un aprendizaje más significativo de la estadística inferencial, más allá de la mera manipulación algorítmica. Además, el enfoque frecuencialista se tratará dentro de un entorno tecnológico.

2 Introducción

La probabilidad es un concepto nuclear en la estadística puesto que constituye la base de muchas otras nociones. Por ejemplo, una variable aleatoria no es más que un conjunto de valores a los cuales va asociada una probabilidad. Una función de distribución de una variable aleatoria asocia a cada valor de la variable, la probabilidad de alcanzar dicho valor, y los que son menores al mismo. Cuando se calculan los parámetros esperanza (valor esperado) o varianza (dispersión respecto al valor esperado) de una variable aleatoria, es necesario conocer la probabilidad asociada a la variable aleatoria. También cuando se estiman parámetros poblacionales, un intervalo de confianza se obtiene a partir de una probabilidad teórica conocida. El P-valor en los contrastes de hipótesis es una probabilidad. Es inmediato pues que encontraremos en todos los grandes temas de la estadística inferencial la noción de probabilidad. Consecuentemente, una comprensión sólida y rica en sus diversos significados de la noción de probabilidad ayudará a abordar y comprender mejor la estadística en general.

Los orígenes históricos de la noción de probabilidad se sitúan en los juegos de azar, y simplemente se refería a porcentajes asociados a los diferentes resultados que podían acontecer en un juego. Dichos porcentajes indicaban las posibilidades que había de que se concretase cierta experiencia del juego. Más tarde, la noción de probabilidad se formalizo matemáticamente como una aplicación que a cada suceso asocia un número real entre 0 y 1 cumpliéndose dos axiomas. Axioma 1: Al suceso seguro se le tiene que asociar el valor 1. Axioma 2: Si dos sucesos no tienen ningún elemento en común (son incompatibles), se tiene que cumplir que la probabilidad de la unión debe ser igual a la suma de probabilidades. A partir de estos axiomas se pueden ir deduciendo diversas propiedades de la probabilidad. Por ejemplo, la probabilidad del suceso contrario a determinado suceso *A* es igual a 1 menos la probabilidad del suceso *A*.

Pero, ¿Cómo asignar probabilidades a los sucesos? Generalmente se usa la regla de Laplace que establece que probabilidad de un suceso A es igual al número de situaciones en que se puede asegurar que sucede A, dividido por el número total de situaciones; siempre y cuando, todas estas últimas situaciones tengan las mismas posibilidades de suceder. Por ejemplo, si se lanza una moneda equilibrada, la probabilidad de obtener cara sería ½, es decir, 0'5 que en porcentaje sería el 50%. Pero hay experiencias donde aplicar la regla de Laplace resulta imposible. Por ejemplo, determinar la probabilidad de que una máquina produzca una pieza defectuosa o de que un paquete de harina tenga exactamente el peso mostrado en la etiqueta. En esos casos resulta útil, una aproximación frecuencialista, que obtiene valores de probabilidad repitiendo muchas veces una experiencia y calculando la frecuencia relativa del suceso para el cual se quiere estimar la probabilidad.

3 Objetivos

Al finalizar este artículo, un estudiante debe ser capaz de asociar probabilidades a sucesos desde una perspectiva frecuencialista.

4 Desarrollo

Se inicia la sesión con la siguiente pregunta ***¿Cuál es la probabilidad de obtener un uno al lanzar un dado?*** Que tiene una respuesta teórica. A continuación veremos, con ayuda de Excel y de Matlab, la visión frecuencialista. Acabaremos planteando un problema y su solución frecuencialista.

Si se aplica la Regla de Laplace (ecuación 1) tendremos que la probabilidad de sacar un uno al lanzar un dado será igual al número de casos favorables que es 1 caso, dividido por los casos totales o posibles que son 6, es decir, 1/6= 0'1$\hat{6}$≈16'67%.

$$p(A) = \frac{Casos\ \ favorables\ de\ A}{Casos\ totales\ \ de\ A}$$

Ecuación 1. Regla de Laplace

Comprobemos el resultado aplicando la aproximación frecuencialista. Para ello lanzamos un dado 60 veces. Contemos las veces que sale el 1 (*n*) y hallemos la probabilidad pedida *n*/60 que debe ser un porcentaje parecido a 16'67%. En la figura 1 podrás ver lo realizado por un estudiante de ciencias ambientales para obtener dicho valor.

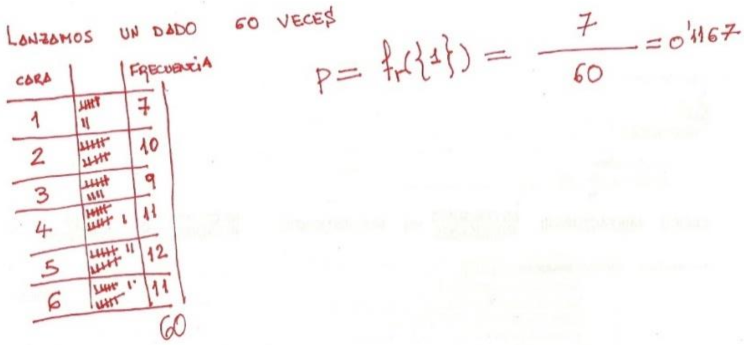


Figura 1. Lanzamiento de un dado por un estudiante de ambientales

Algunas observaciones a tener en cuenta. En primer lugar, tu resultado no tiene por qué ser igual al mostrado en la figura 1. Lo usual es que sea diferente por la componente aleatoria que tiene la experiencia del lanzamiento de un dado. Y en segundo lugar, tu resultado puede que sea muy distinto al obtenido por la regla de Laplace. Esto último podría estar motivado en que el dado que estás usando no sea perfecto (esté un tanto sesgado hacia algunos de los resultados), porque lo lances mal o porqué es necesario lanzar más veces el dado.

4.1 Trabajando con una hoja de Cálculo

Pasamos a continuación, con la ayuda de EXCEL, a lanzar 300 veces un dado. La hoja de cálculo EXCEL dispone de la función **=ALEATORIO.ENTRE(1;6)** que nos da un valor aleatorio entero entre el 1 y el 6, con la misma posibilidad para cada valor. Por lo tanto, dicha función proporciona una manera virtual de lanzar un dado.

En primer lugar, lanzamos 300 veces el dado

- *Nos situamos en una celda, A1 y escribimos* **=ALEATORIO.ENTRE(1;6)**
- *Copiar la fórmula anterior hasta la celda A300*

Mientras se completan los lanzamientos se observa que los números aleatorios que van apareciendo van cambiando. No te preocupes, al final de todo dispondrás de una tabla de frecuencias correspondiente a una muestra aleatoria del lanzamiento de un dado 300 veces. Comencemos a construir la tabla de frecuencias

- *Escribir* **Numero** *en la celda C1,* **1** *en la C2,* **2** *en la C3, … y* **6** *en la celda C7*
- *Escribir* **Frecuencia** *en la celda D1*
- *Escribir* ***=CONTAR.SI($A$1:$A$240; C2)*** *en la celda D2*
- *Copiar la formula anterior desde D3 hasta D7*
- *Sumar la columna de las frecuencias con la orden* **Σ Autosuma**

Por tanto, la probabilidad de obtener un uno sería 52/300= 0'1733 (ver figura 2) que ya se parece a lo calculado con la regla de Laplace.

	A	B	C	D	E	F
1	6		Número	Frecuencia		
2	2		1	52		
3	3		2	51		
4	6		3	56		
5	3		4	46		
6	1		5	52		
7	1		6	43		
8	4			300		
9	6					
10	1					
11	5	obabilidad de saccar un ur		0,17333333		
12	6					
13	3					

Figura 2. Lanzar 300 veces el dado virtual de la hoja de Cálculo

Veamos gráficamente lo que se denomina el teorema central del límite (o ley de los grandes números), que indica que cuantas más experiencias realicemos se observa como la frecuencia relativa tiende (se va acercando) al valor de la probabilidad teórica

Comencemos "lanzando", en una nueva hoja, esta vez 400 veces el dado virtual:

- *Nos situamos en una celda, A1 y escribimos* **=ALEATORIO.ENTRE(1;6)**
- *Copiar la fórmula anterior hasta la celda A400*

A continuación, calculemos las frecuencias absolutas que se van acumulando de las veces que aparece el 1.

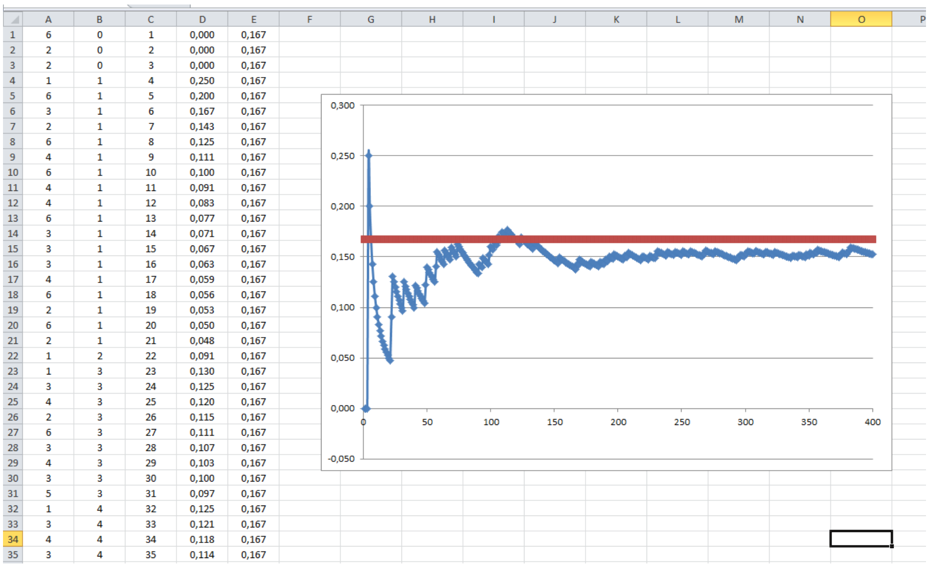
- *Escribir* ***=CONTAR.SI($A$1:A1; 1)*** *en la celda B1*
- *Copiar la formula anterior desde B2 hasta B400*

Ahora calculemos sus frecuencias relativas

- *Escribir en la celda C1,* **1** *y en la C2,* **2**
- *Seleccionar las celdas C1 y C2 y arrastrar hasta la celda C400.*
- *Escribir en la celda D1* **= B1/C1**
- *Copiar la formula anterior desde D2 hasta D400*

Finalmente representemos mediante un gráfico de dispersión las frecuencias relativas obtenidas y lo compararemos con el valor obtenido por la regla de Laplace (ver figura 3).

- *Escribir en la celda E1,* **=1/6**
- *Copiar la formula anterior desde E2 hasta E400*
- *Seleccionar las celdas desde A1, B1 y C1 hasta A400, B400 y C400*
- *Insertar Gráficos Dispersión*



	A	B	C	D	E
1	6	0	1	0,000	0,167
2	2	0	2	0,000	0,167
3	2	0	3	0,000	0,167
4	1	1	4	0,250	0,167
5	6	1	5	0,200	0,167
6	3	1	6	0,167	0,167
7	2	1	7	0,143	0,167
8	6	1	8	0,125	0,167
9	4	1	9	0,111	0,167
10	6	1	10	0,100	0,167
11	4	1	11	0,091	0,167
12	4	1	12	0,083	0,167
13	6	1	13	0,077	0,167
14	3	1	14	0,071	0,167
15	3	1	15	0,067	0,167
16	3	1	16	0,063	0,167
17	4	1	17	0,059	0,167
18	6	1	18	0,056	0,167
19	2	1	19	0,053	0,167
20	6	1	20	0,050	0,167
21	2	1	21	0,048	0,167
22	1	2	22	0,091	0,167
23	1	3	23	0,130	0,167
24	3	3	24	0,125	0,167
25	4	3	25	0,120	0,167
26	2	3	26	0,115	0,167
27	6	3	27	0,111	0,167
28	3	3	28	0,107	0,167
29	4	3	29	0,103	0,167
30	3	3	30	0,100	0,167
31	5	3	31	0,097	0,167
32	1	4	32	0,125	0,167
33	3	4	33	0,121	0,167
34	4	4	34	0,118	0,167
35	3	4	35	0,114	0,167

Figura 3. La Ley de los grandes números (aproximación gráfica)

4.2 Ahora utilizando MATLAB

Continuemos con el problema planteado pero lanzando 1200 el dado. Esta vez utilizaremos el programa MATLAB. Para generar los valores pseudo aleatorios uniformes discretos en el intervalo [1,6], que es la manera de simular un dado virtual, utilizaremos la función **unidrnd(6,*n*,1)** que genera un vector columna de *n* números aleatorios enteros entre 1 y 6. MATLAB dispone además de la función **tabulate(A),** que construye la tabla con las frecuencias absolutas y relativas a partir de los datos recogidos en **A**.

Abre el MATLAB y ves probando las posibilidades que te ofrecen funciones antes mencionadas. Recuerda que tus resultados no tienen por que coincidir con los que se muestran a continuación

```
>> A= unidrnd(6,4,1)

A =
     1
     1
     6
     6


>> tabulate(A)

  Value    Count   Percent
      1        2     50.00%
      2        0      0.00%
      3        0      0.00%
      4        0      0.00%
      5        0      0.00%
      6        2     50.00%
```

Lancemos 1200 veces el dado virtual de MATLAB y construyamos la tabla de frecuencias (ver figura 4).

```
>> A=unidrnd(6,1200,1); tabulate(A)

  Value    Count   Percent
      1      192     16.00%
      2      197     16.42%
      3      205     17.08%
      4      210     17.50%
      5      223     18.58%
      6      173     14.42%
```

Figura 4. Lanzar 1200 veces el dado virtual de MATLAB

Vemos como la probabilidad obtenida (16%) ya resulta más parecida a la que se obtiene siguiendo la regla de Laplace (16'67%). Prueba y lanza 10000 veces el dado y analiza lo que obtienes, es decir, si la probabilidad frecuencialista se parece a la teórica de Laplace.

4.3 Otro ejercicio

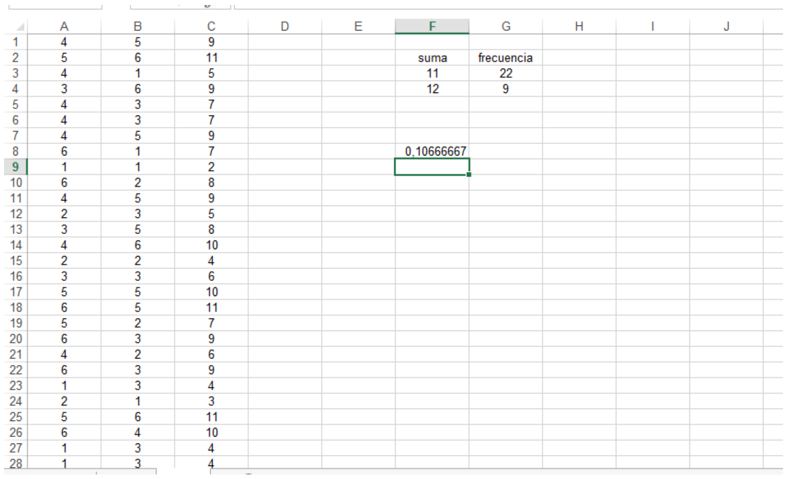
Sigamos con el siguiente ejercicio: *¿Cuál es la probabilidad de obtener más de 10 puntos, es decir, 11 o 12 puntos, al lanzar dos dados y sumar los resultados de cada uno de los dados?*

Esta vez, comenzaremos usando EXCEL

- *Nos situamos en la celda, A1 y escribimos* **=ALEATORIO.ENTRE(1;6)**
- *Copiar la fórmula anterior hasta la celda A300*
- *Nos situamos en una celda, B1 y escribimos* **=ALEATORIO.ENTRE(1;6)**
- *Copiar la fórmula anterior hasta la celda B300*

- *Nos situamos en una celda, C1 y escribimos* **=A1+B1**
- *Copiar la fórmula anterior hasta la celda C300*

De la misma manera que se ha procedido en la actividad anterior, construiremos la tabla de frecuencias para los datos de la columna C. Por tanto, se puede deducir que la probabilidad de obtener más de 10 puntos sería (20+12)/300=0'1067 (Figura 5).



	A	B	C	D	E	F	G	H	I	J
1	4	5	9							
2	5	6	11			suma	frecuencia			
3	4	1	5			11	22			
4	3	6	9			12	9			
5	4	3	7							
6	4	3	7							
7	4	5	9							
8	6	1	7			0,10666667				
9	1	1	2							
10	6	2	8							
11	4	5	9							
12	2	3	5							
13	3	5	8							
14	4	6	10							
15	2	2	4							
16	3	3	6							
17	5	5	10							
18	6	5	11							
19	5	2	7							
20	6	3	9							
21	4	2	6							
22	6	3	9							
23	1	3	4							
24	2	1	3							
25	5	6	11							
26	6	4	10							
27	1	3	4							
28	1	3	4							

Figura 5. Lanzamiento de dos dados virtuales con Hoja de Cálculo

Para afrontar el problema lanzando los dos dados., por ejemplo, 5000 veces, utilizando MATLAB, puedes escribir las siguientes sentencias

```
>> D1=unidrnd(6,5000,1);
>> D2=unidrnd(6,5000,1);
>> A= D1+D2; tabulate(A)
      Value    Count   Percent
          1        0     0.00%
          2      159     3.18%
          3      284     5.68%
          4      410     8.20%
          5      542    10.84%
          6      657    13.14%
          7      855    17.10%
          8      706    14.12%
          9      554    11.08%
         10      407     8.14%
         11      303     6.06%
         12      123     2.46%
```

Por lo tanto, la probabilidad de obtener más de 10 puntos sería, aproximadamente 0'0606+0'0246=**0'0852** Si recurres a la regla de Laplace, debes saber que los casos totales en el lanzamiento de dos dados son 6×6=36, es decir, {(1,1), (1,2), (1,3), …, (1,6), (2,1), (2,2),.., (6,5), (6,6)} de los cuales solo hay 3 que suman más de 10 puntos {(5,6), (6,5), (6,6)} por tanto, la probabilidad pedida sería 3/36=1/12=**0.083**.

5 Algunas conclusiones

Es obvio que la regla de Laplace es un buen instrumento para calcular probabilidades, pero también es cierto que la aproximación frecuencialista puede ayudar a superar determinados problemas que surgen a la hora de aplicar dicha regla. Contamos, además, con instrumentos tecnológicos que pueden simular la realidad y ayudarnos a utilizar con éxito la aproximación frecuencialista a la probabilidad.

Es importante destacar que el conocimiento matemático se construye repitiendo los procesos, interiorizándolos y finalmente aplicándolos a diferentes contextos. Por esta razón proponemos una serie de ejercicios para consolidar lo aprendido en este artículo docente.

1) Calcula la probabilidad de obtener una o más caras al lanzar dos monedas. **(Sol.- aproximadamente el 75%).**
2) Obtén la probabilidad de obtener más caras que cruces al lanzar 3 monedas? **(Sol.- aproximadamente el 50%).**
3) Problemas del Caballero De Meré:
   a. En el lanzamiento de 4 monedas ¿Que suceso tiene mayor probabilidad, obtener al menos un seis o su contrario? **(Sol.- sacar un seis)**
   b. ¿Tiene la misma probabilidad sacar un seis en el lanzamiento de dos dados que sacar dos seises al lanzar 12 veces el dado? **(Sol. No, es más probable- sacar un seis en el lanzamiento de dos dados).**

6 Bibliografía

BATANERO C. (2015). "Understanding randomness for research and teaching". *CERME 9: 9th Congress of European Research in Mathematics Education.* Praga.

BATANERO C. (2005). "Significados de la probabilidad en la educación secundaria" *Revista Latinoamericana de Matemática educativa*, 8(3), 247-264.

BOIGUES, F.J; ESTRUCH V.D. Y ROIG B. (2016). *Herramientas matemáticas para el estudio del Turismo*. Colección apuntes. Valencia: Universitat Politècnica de València.

BRASE C. H. Y BRASE C. P. (2016). *Understanding Basic Statistics*. 7ª edición, Metric version. Ed. Cengage Learning.

PECK, R. (2014). *Statistics. Learning from data*. Preliminary edition. Boston: Ed. Cengage Learning.

SÁNCHEZ SÁNCHEZ, E. Y VALDEZ MONROY, J.C. (2015). "El razonamiento probabilístico informal de estudiantes de bachillerato" en C. Fernandez y N. Planas (eds.), *Investigación en Educación Matemática* (pp.89-103). Alicante: SEIEM.