Barcelona, a 30 de enero de 2014

Examen de Introducción a la Probabilidad

Atención: es preciso definir previamente los sucesos relevantes. Todas las variables aleatorias utilizadas se han de especificar con claridad. Los cálculos tendrán que estar debidamente justificados.

- **Problema 1**

  Los ociosos de la oficina pasan largo tiempo observando a quienes acuden a la máquina del café y sus costumbres. Conocen que los hombres eligen dos veces más el café con leche que el capuchino, tres veces más el cortado que el capuchino, y el más elegido es el café solo (dos veces más que el café con leche). En cambio, para las mujeres las proporciones son las mismas, pero en orden inverso: de menos a más, café solo, cortado, con leche y capuchino.

  Vamos a suponer que el fenómeno de observar quién y qué tipo de café elige un/a trabajador/a de este lugar tiene naturaleza aleatoria, y sus probabilidades se rigen por las anteriores condiciones.

  1. Haga una tabla “Género” x “Tipo café” y complétela con las probabilidades condicionadas por el género. (2 pts.)

	solo	cortado	con leche	capuchino
P(* | Hombres)	0.4	0.3	0.2	0.1
P(* | Mujeres)	0.1	0.2	0.3	0.4

  2. Tomemos el consumo de cierto mes, en el que se despacharon 800 bebidas de la máquina, de las que 248 fueron capuchinos. Con esta información, diga cuál es la proporción de hombres y mujeres en la oficina. (2 pts.)

     P(capuchino) = 248/800 = P(capuchino | H) P(H) + P(capuchino | M) P(M) =
     P(capuchino | H) x + P(capuchino | M) $(1-x)$
     $248/800 = 0.1x + 0.4(1-x) \Rightarrow 0.3x = 0.4 - 248/800 \Rightarrow x = 0.3$

     Hombres: 30%; mujeres: 70%.

  3. Imaginemos que en este mes el reparto de géneros corresponde a dos mujeres por cada hombre. Con esta situación, halle la probabilidad de que un ocioso observe que el próximo servicio corresponde a un hombre que se pide un café solo o cortado. (2 pts.)

     Ahora supondremos que P(H) = 1/3 y P(M) = 2/3.

     P(H ∩ (solo ∪ cortado)) = P(solo ∪ cortado | H) P(H) = 0.7 1/3 = 0.2333...

  4. Nos hemos encontrado un vaso recién utilizado que huele a café solo o a cortado. ¿Cuál es la probabilidad de que su usuario haya sido un hombre? Diga si ha sido relevante tener alguna pista del tipo de bebida consumido, y por qué. (2 pts.)

     La probabilidad pedida ahora es condicional, ya que conocemos alguna cosa del tipo de café.

     $$\mathrm{P(H \mid solo \cup cortado)} = \frac{\mathrm{P(H \cap (solo \cup cortado))}}{\mathrm{P(solo \cup cortado)}}$$

     El denominador: P(solo ∪ cortado) = P(H ∩ (solo ∪ cortado)) + P(M ∩ (solo ∪ cortado)) =
     0.7 1/3 + (0.1+0.2)2/3 = 0.4333...

     Luego: P(H | solo ∪ cortado) = 0.2333.../0.4333... = 7/13 ≈ 0.5385. Observamos que la probabilidad se ha incrementado, desde P(H)=1/3. Esto significa que si sabemos que ha tomado café solo o cortado, las probabilidades de ser hombre aumentan, por tanto, no son sucesos independientes (ya lo sabíamos, desde el momento que las probabilidades condicionadas por el género eran diferentes).

  5. Y si los trabajadores de la oficina escogen su bebida sin influirse unos a los otros, ¿cuál es la probabilidad de que dos de ellos que coincidan en la máquina elijan la misma bebida? (2 pts.)

     Vamos a definir los eventos de la cuestión:

     - A_x, con $x \in$ {solo, cortado, con leche, capuchino} significa que el trabajador número 1 ha escogido la opción x.
     - B_x, ídem para el trabajador número 2.

     Así que la pregunta se formula como:

     $$\sum_{\forall x} P(A_x \cap B_x)$$

Puesto que nos dicen que las opciones son independientes:

$$\sum_{\forall x} P(A_x)P(B_x) = \sum_{\forall x} (P(x))^2$$

Solo nos falta hallar las probabilidades para cada opción (sin contemplar el género): P(x) = P(x | H)P(H) + P(x | M)P(M). La tabla resume todos los cálculos:

	solo	cortado	con leche	capuchino
P($x \cap$ Hombres)	0.133...	0.1	0.066...	0.033...
P($x \cap$ Mujeres)	0.067...	0.133...	0.2	0.266...
P(x)	6/30=0.2	7/30=0.233...	8/30=0.266...	9/30=0.3
P(x)^2	0.04	0.0544...	0.071...	0.09

El resultado final es la suma de la fila inferior: 0.255...

- **Problema 2**

  Siguiendo con el supuesto anterior, el dueño de la máquina de café obtiene un beneficio por unidad vendida que se muestra en la siguiente tabla:

	solo	cortado	con leche	capuchino
(ct. de euro)	22	18	14	16

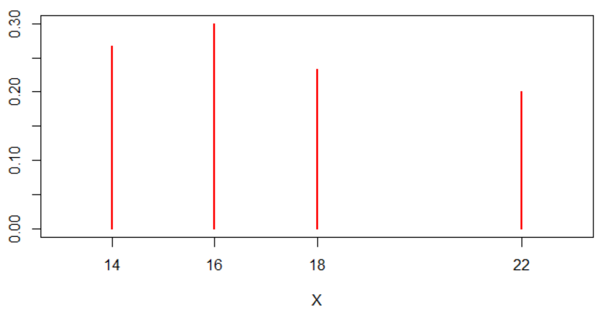
1. Represente gráfica y tabularmente la función de distribución de la variable "Beneficio por unidad vendida". (2 pts.)

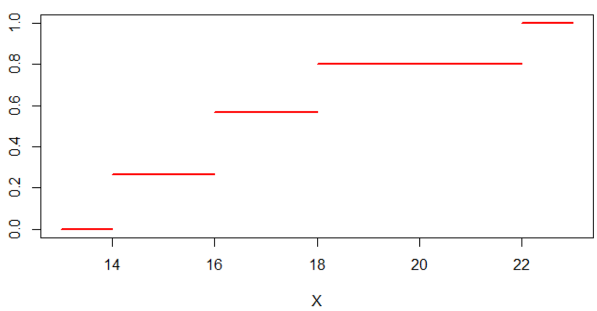
   Gracias a las probabilidades que habíamos hallado previamente podemos construir la correspondencia entre *resultados* y *valores* de la variable aleatoria B = "beneficio". Ordenando por valor:

resultado (ω_i)	b_i	P($B = b_i$) = $f_B(b_i)$
con leche	14	0.266...
capuchino	16	0.3
cortado	18	0.233...
solo	22	0.2

Función de distribución de B:

b	P($B \leq b$) = $F_B(b)$
$b < 14$	0
$14 \leq b < 16$	0.266...
$16 \leq b < 18$	0.566...
$18 \leq b < 22$	0.8
$22 \leq b$	1





2. Calcule el beneficio esperado y la desviación típica por unidad vendida. (2 pts.)
$E(B) = \sum_i b_i f(b_i) = 1/30(14 \times 8 + 16 \times 9 + 18 \times 7 + 22 \times 6) = 17.13...$ céntimos de euro.
$E(B^2) = \sum_i b_i^2 f(b_i) = 1/30(14^2 \times 8 + 16^2 \times 9 + 18^2 \times 7 + 22^2 \times 6) = 301.466...$
$V(B) = E(B^2) - E(B)^2 = 301.466... - 293.551 = 7.915...$; $\sigma_B = 2.813$ céntimos de euro.

3. ¿Cuál sería la distribución de probabilidad del beneficio total de un mes, suponiendo que tomamos una base de 800 bebidas vendidas? ¿Puede responder de forma exacta o aproximada? (2 pts.)

M: beneficio en un mes (800 consumiciones) = $\sum_{i=1}^{800} B_i$ siendo B_i el beneficio dado por la i-ésima consumición del mes. Dado que la distribución de probabilidad de B es discreta, con 4 valores irregularmente distribuidos, la distribución de M sería muy difícil de establecer con exactitud, pero al ser la suma de un gran número de variables (vamos a asumir que independientes), mediante el Teorema Central del Límite podemos decir que:
M $\approx N(\mu = 800 \times 17.13..., \sigma^2 = 800 \times 7.915...)$, es decir, y pasando a euros:
M $\approx N(\mu = 137.07, \sigma = 0.796)$.

4. Siendo realistas, no podemos fijar por anticipado el numero exacto de bebidas que se venderán, pero es sensato afirmar que las consumiciones se producen durante la jornada laboral (8 horas) con una media de 5 por hora e independientemente entre sí. ¿Con qué modelo de probabilidad representaría el número de consumiciones en un día? ¿Y en un mes (asuma 21 días laborables por mes)? Explícite la esperanza y la desviación típica en ambos casos. (2 pts.)
Un modelo apropiado parece ser el modelo de Poisson, dado que nos dicen que la consumición se hace esporádicamente en el tiempo pero con una tasa regular de 5 por hora, que se convierten en 40 por jornada laboral, o en 840 por mes de 21 días laborables. Evidentemente, estos son valores esperados, y existe cierta variabilidad en el resultado observado. Es decir:

periodo	modelo	esperanza	desviación
1 día	$P(40)$	40	$\sqrt{40} = 6.32$
1 mes	$P(840)$	840	$\sqrt{840} = 28.98$

5. Con el modelo de la pregunta anterior, determine la probabilidad de que en una semana laborable de cinco días se vendan entre 200 y 205 bebidas. (2 pts.)
S: "número de consumiciones bebidas en una semana o 5 días", S $\sim P(200)$.
P$(200 \le S \le 205) = \sum_{k=200}^{205} e^{-200} \frac{200^k}{k!}$
Sin embargo, el cálculo de esta expresión no es un requisito para el apartado. Es difícil poder evaluar numéricamente esto con una calculadora de mano. Por esta razón normalmente se acude a la aproximación de la Poisson por una normal, de media 200 y desviación tipo = $\sqrt{200}$.
Si a pesar de todo se quisiera resolver la probabilidad anterior a mano, nótese que hay cierta redundancia que puede servir para hacer factor común:

$$P(200 \le S \le 205) = e^{-200} \frac{200^{200}}{200!} \left(1 + \frac{200}{201}\left(1 + \frac{200}{202}\left(1 + \frac{200}{203}\left(1 + \frac{200}{204}\left(1 + \frac{200}{205}\right)\right)\right)\right)\right)$$

La expresión del paréntesis se puede hallar, y vale 5.830767. El número 200! puede evaluarse con la aproximación de Stirling:

$$e^{-200} \frac{200^{200}}{200!} \approx e^{-200} \frac{200^{200}}{\sqrt{2\pi 200} 200^{200} e^{-200}} = \frac{1}{\sqrt{2\pi 200}} = 0.02821$$

Así que la solución es aproximadamente 0.02821×5.830767 = 0.1645. La solución exacta (obtenida con R) es 0.1644144; en cambio, usando la aproximación normal sin corrección de continuidad obtenemos 0.1382.

- **Problema 3**

Suponga que estamos estudiando la talla, X, en centímetros, de los habitantes de un determinado municipio y nos indican que podemos suponer que si nos restringimos a los hombres X sigue una distribución aproximadamente normal $N(\mu = 175, \sigma = 10)$ y si consideramos las mujeres, X sigue también una distribución aproximadamente normal pero con diferente valor medio: $N(\mu = 165, \sigma = 10)$. Suponga, además, que hay igual proporción de hombres que de mujeres. Bajo estos supuestos,

1. Exprese la función de distribución de X (la talla en centímetros de una persona al azar del municipio), en función de la función de distribución normal estandarizada

$$\Phi(x) = \int_{\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$$

(2 pts.)

Sea M el suceso "ser hombre", y H el suceso "ser mujer", entonces

$$F_X(x) = P(X \le x) = P(M)P(X \le x|M) + P(H)P(X \le x|H) = \frac{1}{2}P(X \le x|M) + \frac{1}{2}P(X \le x|H)$$

Condicionado a M, X sigue una distribución normal $N(\mu = 175, \sigma = 10)$ y, por tanto $(X - 175)/10$ seguirá una normal estandarizada. Si condicionamos a H, $N(\mu = 165, \sigma = 10)$ y, por tanto $(X - 165)/10$ también sigue una normal estandarizada. Por consiguiente

$$\begin{aligned} F_X(x) &= P(X \le x) = \tfrac{1}{2}P((X-175)/10 \le (x-175)/10|M) + \tfrac{1}{2}P((X-165)/10 \le (x-165)/10|H) \\ &= \tfrac{1}{2}\Phi((x-175)/10) + \tfrac{1}{2}\Phi((x-165)/10) \end{aligned}$$

2. ¿Cuál es la probabilidad de que la talla de un hombre elegido al azar sea superior a 170cm? (2 pts.)

Con la misma notación del apartado anterior, nos piden:

$$P(X \ge 170|M) = P((X-175)/10 \ge (170-175)/10|M) = 1 - \Phi(-0,5) = \Phi(0,5) = 0,691462461$$

ya que, condicionado a M, X sigue una distribución normal $N(\mu = 175, \sigma = 10)$ y, por tanto $(X - 175)/10$ sigue una normal estandarizada.

3. ¿Cuál es la probabilidad de que la talla de una mujer elegida al azar sea inferior a 170cm? (2 pts.)

Con la misma notación del apartado anterior, nos piden:

$$P(X \le 170|H) = P((X-165)/10 \le (170-165)/10|M) = \Phi(0,5) = 0,691462461$$

ya que, condicionado a H, X sigue una distribución normal $N(\mu = 165, \sigma = 10)$ y, por tanto $(X - 165)/10$ sigue una normal estandarizada.

4. Si elegimos un hombre y una mujer del municipio al azar y con independencia, cuál es la probabilidad de que la talla del hombre sea superior a la talla de la mujer? (2 pts.)

Si llamamos ahora X a la talla del hombre e Y a la talla de la mujer, resultara que X sigue una distribución normal $N(\mu = 175, \sigma = 10)$ e Y una distribución normal $N(\mu = 165, \sigma = 10)$ Por tanto, $W = X - Y$ seguirá también una distribución normal (al ser suma de normales independiente), de esperanza $E(W) = E(X) - E(Y) = 175 - 165 = 10$ y varianza igual a $\text{var}(W) = \text{var}(X) + \text{var}(Y) = 100 + 100 = 200$, debido a la independencia de X e Y: Por tanto la desviación típica de Z es $\sigma_W = \sqrt{200} = 14,14213562$. Por consiguiente $(W - 10)/14,14213562$ seguirá una normal estandarizada y

$$\begin{aligned} P(X > Y) = P(W > 0) &= P((W-10)/14,14213562 > -10/14,14213562) \\ &= 1 - \Phi(-0,707106781) = \Phi(0,707106781) = 0,760249939 \end{aligned}$$

5. Sea Z la variable aleatoria número de hombres que hemos de escoger del municipio, al azar y con independencia, hasta encontrar uno con talla superior a 180cm. Escriba explícitamente la probabilidad del suceso $Z = k$ donde $k \in \mathbb{N}$ (2 pts.)

Calculemos en primer lugar la probabilidad de que un hombre tenga una talla superior a 180cm

$$\begin{aligned} p &= P(X > 180|M) = P((X-175)/10 > (180-175)/10|M) = P((X-175)/10 > 0.5|M) \\ &= 1 - \Phi(0.5) = 1 - 0,691462461 = 0,308537539 \end{aligned}$$

Por tanto Z sigue una distribución geométrica de parámetro $p = 0,308537539$

$$P(N = k) = (1-p)^k p = 0,691462461^k 0,308537539 \qquad k \in \mathbb{N}$$

- **Problema 4**

Sea X una variable aleatoria absolutamente continua con función de densidad de probabilidad igual a

$$f(x) = \beta x^\alpha \quad \text{para} \quad x \in [0,2] \quad \text{y} \quad f(x) = 0 \quad \text{para} \quad x \notin [0,2]$$

donde $\alpha > 0$.

1. Halle el valor de β para que $f(x)$ sea realmente una densidad de probabilidad. (2 pts.)
Al ser una densidad de probabilidad tendremos

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_0^2 \beta x^\alpha\,dx = \beta\left[\frac{x^{\alpha+1}}{\alpha+1}\right]_0^2 = \beta\frac{2^{\alpha+1}}{\alpha+1}$$

Por tanto

$$\beta = \frac{\alpha+1}{2^{\alpha+1}}$$

2. Halle la función de distribución de X. (2 pts.)
Tengan en cuenta que, en el caso absolutamente continuo, la función de distribución es

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

Por tanto

$$F(x) = \begin{cases} 0 & x \le 0 \\ \left(\frac{x}{2}\right)^{\alpha+1} & x \in [0,2] \\ 1 & x > 2 \end{cases}$$

3. Halle $\mathrm{E}(X)$ y $\mathrm{var}(X)$. (2 pts.) Para ello

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^2 x\frac{\alpha+1}{2^{\alpha+1}}x^\alpha\,dx = \frac{\alpha+1}{2^{\alpha+1}}\int_0^2 x^{\alpha+1}\,dx = \frac{\alpha+1}{2^{\alpha+1}}\left[\frac{x^{\alpha+2}}{\alpha+2}\right]_0^2 = 2\frac{\alpha+1}{\alpha+2}$$

Además

$$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^2 x^2\frac{\alpha+1}{2^{\alpha+1}}x^\alpha\,dx = \frac{\alpha+1}{2^{\alpha+1}}\int_0^2 x^{\alpha+2}\,dx = \frac{\alpha+1}{2^{\alpha+1}}\left[\frac{x^{\alpha+3}}{\alpha+3}\right]_0^2 = 4\frac{\alpha+1}{\alpha+3}$$

y, por tanto,

$$\mathrm{var}(X) = 4\frac{\alpha+1}{\alpha+3} - 4\left(\frac{\alpha+1}{\alpha+2}\right)^2 = 4\frac{(\alpha+1)((\alpha+2)^2-(\alpha+3)(\alpha+1))}{(\alpha+3)(\alpha+2)^2} = \frac{4(\alpha+1)}{(\alpha+3)(\alpha+2)^2}$$

4. Si $\alpha = 1$ y $X_1,\ldots,X_n$ son variables aleatorias independientes e idénticamente distribuidas según X ¿Cuál es la probabilidad aproximada de que $S_n = X_1+\ldots+X_n$ tome valores entre 135 y 145, para $n = 100$? (2 pts.)
Usaremos el teorema del límite central. Teniendo en cuenta que

$$\mathrm{E}(S_n) = \mathrm{E}(X_1+\ldots+X_n) = \mathrm{E}(X_1)+\ldots+\mathrm{E}(X_n) = 2n\frac{\alpha+1}{\alpha+2} = \frac{400}{3} = 133,3333$$

$$\mathrm{var}(S_n) = \mathrm{var}(X_1+\ldots+X_n) = \mathrm{var}(X_1)+\ldots+\mathrm{var}(X_n) = \frac{4n(\alpha+1)}{(\alpha+3)(\alpha+2)^2} = \frac{200}{9} = 22,2222$$

$$\sigma(S_n) = \sigma(X_1+\ldots+X_n) = \sqrt{22,2222} = 4,7140$$

Por tanto, como

$$\frac{S_n - nE(X)}{\sqrt{n\mathrm{var}(X)}} \xrightarrow{\mathcal{L}} Z$$

donde Z sigue una normal estandarizada, entonces tendremos, de forma aproximada,

$$p = P(135 \le S_n \le 145) = P\left(\frac{135 - 133,333}{4,7140} \le \frac{S_n - 133,333}{4,7140} \le \frac{145 - 133,333}{4,7140}\right)$$

$$\begin{aligned} p &= P(135 \le S_n \le 145) \approx P\,(0,353553391 \le Z \le 2,474873734) \\ &= \Phi(2,474873734) - \Phi(0,353553391) \\ &= 0,993335836 - 0,638163195 = 0,355172641 \end{aligned}$$

5. Halle el momento de cuarto orden de una normal estandarizada, justificando los cálculos realizados para ello. (2 pts.)
   Si Z sigue una distribución normal estandarizada, tenemos que calcular

$$\mathrm{E}(Z^4) = \int_{-\infty}^{\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx = 2\int_0^{\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

al ser la integral de una función *par*. Efectuando el cambio de variable

$$\frac{1}{2}x^2 = t \qquad x = \sqrt{2t} \qquad dx = \frac{1}{\sqrt{2t}}dt$$

resulta

$$\mathrm{E}(Z^4) = \frac{\sqrt{2}}{\sqrt{\pi}}\int_0^{\infty} (2t)^2 e^{-t} \frac{1}{\sqrt{2t}}dt = \frac{1}{\sqrt{\pi}}4\int_0^{\infty} t^{3/2} e^{-t} dt = \frac{1}{\sqrt{\pi}}4\Gamma(5/2) = \frac{1}{\sqrt{\pi}}4\frac{3}{2}\frac{1}{2}\Gamma(1/2) = 3$$

ya que, como es bien sabido, $\Gamma(1/2) = \sqrt{\pi}$.