Barcelona, a 2 de julio de 2014

Examen de Introducción a la Probabilidad

- **Problema 1**

El 70 % de los pacientes que acuden a una clínica especializada en enfermedades óseas son mujeres. Una de las pruebas diagnósticas que se realizan es una radiografía que puede dar tres tipos de resultado: *negativo* para el 20 % de las mujeres (40 % para los hombres), *indicios* apreciables para el 30 % de las mujeres (40 % para los hombres), siendo el tercer tipo de resultado cuando el diagnóstico es claramente *positivo*.

La proporción real de casos en los que cierta patología se ha desarrollado depende de las condiciones anteriores, como se muestra a continuación:

Género	negativo	indicios	positivo
Hombres	0.1	0.3	0.8
Mujeres	0.5	0.7	0.9

1. represente el árbol de eventos y probabilidades de la experiencia anteriormente descrita (2 pts.)

   H: hombre. M: mujer.

   I: radiografía presenta indicios. +: resultado positivo. −: resultado negativo.

   E: enfermo. S: sano.

2. De 1000 pacientes visitados en la clínica, diga el número que correspondería a: (2 pts.)

   - pacientes con la patología.
   - mujeres con la patología.
   - radiografías que presentan indicios (a cada paciente se le hace una sola radiografía).
   - hombres con la patología cuyas radiografías no mostraba indicios de la enfermedad.

   $P(E) = P(E \cap - \cap H) + P(E \cap I \cap H) + P(E \cap + \cap H) + P(E \cap - \cap M) + P(E \cap I \cap M) + P(E \cap + \cap M) =$

   $0.1\ 0.4\ 0.3 + 0.3\ 0.4\ 0.3 + 0.8\ 0.2\ 0.3 + 0.5\ 0.2\ 0.7 + 0.7\ 0.3\ 0.7 + 0.9\ 0.5\ 0.7 = 0.628$

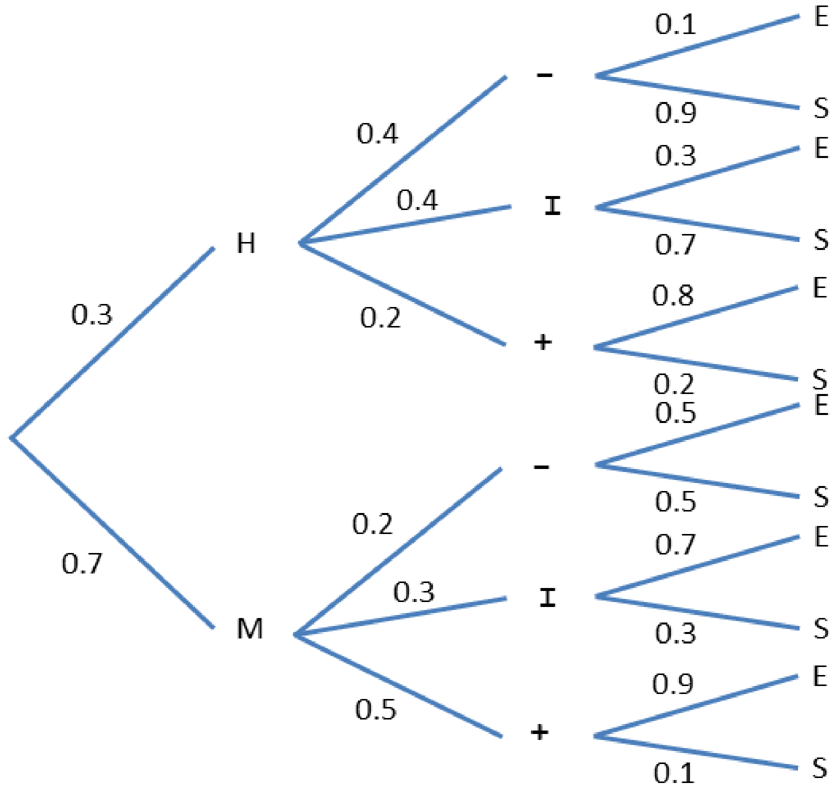
   De cada 1000, 628 son pacientes enfermos.

   $P(M \cap E) = P(E \cap - \cap M) + P(E \cap I \cap M) + P(E \cap + \cap M) = 0.5\ 0.2\ 0.7 + 0.7\ 0.3\ 0.7 + 0.9\ 0.5\ 0.7 = 0.532$

   De cada 1000, 532 son mujeres enfermas.

   $P(I) = P(H \cap I) + P(M \cap I) = 0.4\ 0.3 + 0.3\ 0.7 = 0.33$

   De cada 1000 radiografías, 330 presentan indicios.



P(H ∩ E ∩¬ I) = P(E ∩ − ∩ H) + P(E ∩ + ∩ H) = 0.1 0.4 0.3 + 0.8 0.2 0.3 = 0.06

De cada 1000 pacientes, 60 son hombres enfermos sin indicios en la prueba.

3. Un radiólogo hace una consulta: quiere saber si el género de los pacientes está asociado con la patología (si son independientes, o no). Redacte una respuesta detallada. (2 pts.)

   Podemos comprobarlo con un caso concreto: ¿P(M ∩ E) igual a P(M) P(E)? Vemos que no es así. P(M) P(E) = 0.7 0.628 = 0.4396. Si fueran independientes, de cada 1000 pacientes habría unas 440 mujeres enfermas, en vez de 532 que hemos encontrado. También podemos encontrar que, siendo mujer, la probabilidad de tener la patología es mayor que siendo hombre: P(E | M) > P(E | H).

4. Si de un paciente se sabe que no presenta la patología, ¿qué probabilidad tenemos entonces de que sea una mujer? (2 pts.)

   Escribimos: P(M | S), que podemos desarrollar por medio de Bayes.

$$
\begin{aligned}
P(M|S) &= \frac{P(S|M)P(M)}{P(S)} = \frac{(P(S\cap -|M) + P(S\cap I|M) + P(S\cap +|M))P(M)}{1-P(E)} \\
&= \frac{(0.2\ 0.5 + 0.3\ 0.3 + 0.1\ 0.5)0.7}{1-0.628} = 0.4516
\end{aligned}
$$

Esto también nos dice que género y patología no son independientes, ya que nos sale que si un paciente está sano disminuye la probabilidad de que sea una mujer (inicialmente, un 70 %).

5. Cuando se observan en la radiografía de un paciente indicios de la enfermedad, ¿es más probable que presente la patología un hombre o una mujer? (2 pts.)

   Se puede replantear la cuestión de la siguiente forma: ¿cuál es la probabilidad de que presente la patología (E) un hombre que presentaba indicios en el diagnóstico? ¿Y es mayor que si fuese mujer? Es importante darse cuenta de que lo que es incierto es la patología y no el género.

   Por tanto, hallaremos $P(E \mid I \cap H)$ y $P(E \mid I \cap M)$. Pero no hay que calcular nada, porque estas probabilidades se encuentran directamente en el árbol (o en la tabla): $P(E \mid I \cap H) = 0.3$
   $P(E \mid I \cap M) = 0.7$.

   Así que la respuesta es: con una mujer es más probable que haya enfermedad.

- **Problema 2**

  Un determinado tipo de patología ósea provoca que se produzca una rotura en el cuello del fémur con una tasa anual de 1 caso por cada 25 mujeres de más de 70 años que padecen dicha patología. En la clínica hay un equipo que sigue constantemente a grupos de pacientes sin relación entre sí.

  1. ¿Qué modelo de probabilidad puede aplicarse a las siguientes variables aleatorias? (2 pts.)

     - el número de roturas observado en un año para un grupo de 200 mujeres de más de 70 años.
     - el número de roturas observado en seis meses para un grupo de 80 mujeres de más de 70 años.
     - el tiempo en años que transcurre entre rotura y rotura, dentro de un grupo de 100 mujeres de más de 70 años.

     - Poisson, con tasa $\lambda_1 = 200\ 1/25 = 8$.
     - Poisson, con tasa $\lambda_2 = 80\ 1/25\ 1/2 = 1.6$ (1/2, por medio año).
     - En un grupo de 100 mujeres la tasa será de $\lambda_3 = 100\ 1/25 = 4$; pero el tiempo entre roturas será una variable Exponencial, con el mismo parámetro λ_3.

  2. Escogida una persona de esas características y que posee este tipo de patología, ¿qué probabilidad posee de que sufra una rotura en el plazo de un año? Y si fueran 20 personas, ¿qué probabilidad habría de que al menos una de ellas sufra una rotura? Halle el número esperado y la desviación tipo del número de personas (sobre 20) que sufren una rotura en un año. (2 pts.)

     En primera aproximación, podríamos coger $1/25 = 0.04$ como probabilidad para una sola persona. Otra opción sería considerar que se trata de un proceso de Poisson observado solo un solo sujeto en vez de muchos, y esta variable X_1 sería —como antes— Poisson$(1/25)$. De donde hallaríamos $\mathrm{P}(X_1 > 0) = 1 - \mathrm{P}(X_1 = 0) = 1 - e^{-1/25} = 0.0392$.

Para 20 personas, modelizaremos con una distribución Binomial el número de ellas que sufre rotura: $Y_{20} \sim B(20, 0.04)$. Entonces calcularemos $\mathrm{P}(Y_{20} > 0) = 1 - \mathrm{P}(Y_{20} = 0) = 1 - 0.96^{20} = 0.558$.

$\mathrm{E}(Y_{20}) = np = 20 \quad 0.04 = 0.8$
$\sigma = \sqrt{np(1-p)} = 0.8764$

3. En ocasiones, una paciente se fractura los dos fémures en el periodo de seguimiento, aunque lo normal es que solo le afecte a uno y, sobre todo, generalmente la fractura no llega a producirse. Respetando estos principios, invéntese una distribución de probabilidad para el número de fracturas por paciente, sabiendo que tiene una esperanza igual a 0.15. (2 pts.)

   Sea F la variable aleatoria que representa el número de fracturas de fémur en una paciente. hemos de tener en cuenta que para su función de probabilidad $P_F()$:

   - $P_F(0)$ ha de ser un valor alto,
   - $P_F(1)$ ha de ser un valor más alto que $P_F(2)$,
   - $\sum_{k=0}^{2} P_F(k) = 1$
   - $\sum_{k=0}^{2} k\, P_F(k)$ aproximadamente 0.15

   Tanteando un poco, los valores $\{0.87, 0.11, 0.02\}$ satisfacen completamente estos requerimientos (pero no es la única solución). El enunciado incluyendo estos valores quedaría mucho más preciso: “En ocasiones *(2 %)*, una paciente se fractura los dos fémures en el periodo de seguimiento, aunque lo normal *(11 %)* es que solo le afecte a uno y, sobre todo, generalmente *(87 %)* la fractura no llega a producirse.”

4. El centro participa en un ensayo clínico para probar la eficacia de un nuevo principio activo contra la descalcificación de los huesos. Supongamos que con este tratamiento la densidad de calcio medido en el fémur de las mujeres sigue una distribución Normal de media 1.15 mg/cm^2 y desviación tipo 0.06 mg/cm^2. Se pide determinar: (2 pts.)

   - la probabilidad de observar una densidad inferior a 0.97 en una participante seleccionada al azar.
   - el valor de densidad que representa el tercer cuartil de la variable (el 75 % de las mujeres presenta una densidad menor)

   Sea D la variable aleatoria *densidad de calcio [mg/cm^2]*, $D \sim N(1,15, 0.06)$. La probabilidad pedida es $\mathrm{P}(D < 0.97) = \mathrm{P}(Z < \frac{0.97-1.15}{0.06}) = \mathrm{P}(Z < -3)$.

   Si la tabla no ofrece probabilidades para el lado negativo, es equivalente a: $\mathrm{P}(Z < -3) = 1 - \mathrm{P}(Z > 3) = 1 - 0.9987 = 0.0013$.

   Por otro lado, buscamos el valor d_{75} tal que $\mathrm{P}(D < d_{75}) = 0.75$. Para una normal estándar, z_{75} vale aproximadamente 0.6745 (en las tablas solo veremos que para 0.67 la función de distribución vale 0.7486, y para 0.68

0.7517, así que tomar un punto medio como 0.675 es aceptable). La relación entre d_{75} y z_{75} es: $d_{75} = \mu + \sigma z_{75} \approx 1.15 + 0.06\ 0.675 = 1.1905$ mg/cm^2.

5. 9 mujeres han sido asignadas al tratamiento nuevo (N), y otras 7 están siendo tratadas con un tratamiento más antiguo (A) que presenta la misma desviación tipo para la densidad de calcio pero una media inferior (de 1.11 mg/cm^2). Ambos grupos son independientes, y se evalúa la media de la densidad de las 9 pacientes del tratamiento N con la media de las 7 pacientes del A. Calcúlese la probabilidad de que N presente una media superior a la de A. (2 pts.)

   Variables a considerar:

   - Densidad media para el grupo N ($n_N = 9$), $\overline{D}_N \sim N(1.15, 0.06/\sqrt{9})$;
   - Densidad media para el grupo A ($n_A = 7$), $\overline{D}_A \sim N(1.11, 0.06/\sqrt{7})$;
   - Diferencia entre medias, $E = \overline{D}_N - \overline{D}_A \sim N(1.15 - 1.11, \sigma_E)$, donde:

     $$\sigma_e = \sqrt{\frac{0.06^2}{9} + \frac{0.06^2}{7}} = 0.030237$$

     (recordad que la variancia de la diferencia es suma de variancias en variables independientes)

   Por tanto, la pregunta es cuánto vale $\mathrm{P}(\overline{D}_N > \overline{D}_A) = \mathrm{P}(E > 0) = \mathrm{P}(Z > -0.04/0.030237) = \mathrm{P}(Z < 1.3229) = 0.907$.

- **Problema 3** Suponga que estamos estudiando el tiempo útil de funcionamiento de dos modelos de tubo fluorescente. Para ello encendemos diversos fluorescentes y observamos, para cada uno de ellos, el tiempo de funcionamiento ininterrumpido, X, en meses, hasta que dejan de alumbrar. Supongamos que, si nos restringimos al modelo A, la variable X sigue una distribución absolutamente continua con función de densidad igual a: $f(x) = e^{-x}\ \mathbf{1}_{\mathbb{R}^+}(x)$ mientras que si nos restringimos al modelo B, la variable X sigue una distribución absolutamente continua con función de densidad igual a: $f(x) = 2e^{-2x}\ \mathbf{1}_{\mathbb{R}^+}(x)$. Bajo estos supuestos,

  1. Elegimos un tubo fluorescente al azar, de tipo A o B de forma equiprobable. Exprese la función de distribución de la variable X, duración del tubo expresada en meses de funcionamiento ininterrumpido. (2 pts.)

     Para calcular la función de distribución de X, $F(x)$, observemos que $F(x) = P(X \leq x)$. Por tanto, por el Teorema de las Probabilidades Totales

     $$F(x) = P(X \leq x) = P(A)P(X \leq x|A) + P(B)P(X \leq x|B)$$

     pero $P(A) = P(B) = 1/2$ y

     $$P(X \leq x|A) = \begin{cases} 0 & \text{si } x \leq 0 \\ \displaystyle\int_0^x e^{-t}\, dt = 1 - e^{-x} & \text{si } x > 0 \end{cases}$$

$$P(X \le x|B) = \begin{cases} 0 & \text{si } x \le 0 \\ \displaystyle\int_0^x 2e^{-2t}\, dt = 1 - e^{-2x} & \text{si } x > 0 \end{cases}$$

por tanto

$$F(x) = \begin{cases} 0 & \text{si } x \le 0 \\ \frac{1}{2}(1-e^{-x}) + \frac{1}{2}(1-e^{-2x}) = 1 - \frac{1}{2}(e^{-x} + e^{-2x}) & \text{si } x > 0 \end{cases}$$

2. ¿Cuál es la probabilidad de que un tubo A elegido al azar tenga una duración superior a 2 meses? (2 pts.)
   Tendremos

$$P(X > 2|A) = 1 - P(X \le 2|A) = 1 - (1 - e^{-2}) = e^{-2} = 0,135335$$

3. ¿Cuál es la probabilidad de que un tubo B elegido al azar tenga una duración inferior a 1 mes? (2 pts.)
   Tendremos

$$\begin{aligned} P(X < 1|B) &= P(X \le 1|B) = 1 - e^{-2\times 1} = 1 - e^{-2} = 1 - 0,135335 \\ &= 0,864665 \end{aligned}$$

4. Si elegimos un tubo A y un tubo B al azar y con independencia, ¿cuál es la esperanza de la variable Z igual a la duración del tubo A menos la duración de tubo B? ¿y su varianza? (2 pts.)
   Llamemos Y a la duración del tubo A y Z a la duración del tubo B. Resulta

$$\begin{aligned} E(Y) &= \int_0^\infty y e^{-y} dy = \Gamma(2) = 1! = 1 \\ E(Y^2) &= \int_0^\infty y^2 e^{-y} dy = \Gamma(3) = 2! = 2 \\ \text{var}(Y) &= E(Y^2) - E(Y)^2 = 2 - 1^2 = 1 \end{aligned}$$

$$\begin{aligned} E(Z) &= \int_0^\infty z e^{-2z} dz = \frac{1}{2}\int_0^\infty t e^{-t} dt = \frac{1}{2}\Gamma(2) = \frac{1}{2}1! = \frac{1}{2} \\ E(Z^2) &= \int_0^\infty z^2 e^{-2z} dz = \frac{1}{2^2}\int_0^\infty t^2 e^{-t} dt = \frac{1}{4}\Gamma(3) = \frac{1}{4}2! = \frac{1}{2} \\ \text{var}(Z) &= E(Z^2) - E(Z)^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4} \end{aligned}$$

Por tanto si llamamos $U = Y - Z$ resulta

$$\begin{aligned} E(U) &= E(Y) - E(Z) = 1 - \frac{1}{2} = \frac{1}{2} \\ \operatorname{var}(U) &= \operatorname{var}(Y) + \operatorname{var}(Z) - 2\operatorname{cov}(Y,Z) = 1 + \frac{1}{4} = \frac{5}{4} = 1,25 \end{aligned}$$

ya que $\operatorname{cov}(Y,Z) = 0$ al ser Y y Z variables independientes.

5. Sea U la variable aleatoria número de tubos de tipo A que hemos de escoger, al azar y con independencia, hasta encontrar uno que dure más de 2 meses. Escriba explícitamente la probabilidad del suceso $U = k$ donde $k \in \mathbb{N}$ (2 pts.)
   Sea p la probabilidad de que un tubo tipo A dure más de 2 meses:

   $$p = P(X > 2|A) = 1 - P(X \leq 2|A) = 1 - (1 - e^{-2}) = e^{-2} = 0,135335$$

   La variable U es una variable aleatoria discreta, cuyo soporte son los naturales positivos $U - 1$ seguirá una distribución de Pascal o geométrica:

   $$P(U = k) = P(U - 1 = k - 1) = (1 - p)^{k-1} p \qquad k = 1, 2, 3, \ldots$$

   y $P(U = k) = 0$ en caso contrario.

- **Problema 4** Sea X una variable aleatoria absolutamente continua con función de densidad de probabilidad igual a

  $$f(x) = \beta\,(1 - x)^2 x^2 \quad \text{para} \quad x \in [0, 1] \quad \text{y} \quad f(x) = 0 \quad \text{para} \quad x \notin [0, 1]$$

  donde $\beta > 0$ es una constante de normalización.

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

     $$1 = \int_0^1 \beta\,(1 - x)^2 x^2\, dx = \beta B(3, 3) = \beta \frac{\Gamma(3)\Gamma(3)}{\Gamma(6)} = \frac{2!2!}{5!}\beta = \frac{4\beta}{120} = \frac{\beta}{30}$$

     y por tanto $\beta = 30$.

  2. ¿Cuál es la probabilidad aproximada de que X tome valores superiores a 0,5 (2 pts.)
     Dicha probabilidad será

     $$\begin{aligned} P(X > 0,5) &= \int_{0,5}^1 30(1 - x)^2 x^2\, dx = 30 \int_{0,5}^1 (x^4 + x^2 - 2x^3)\, dx \\ &= 30 \left[\frac{1}{5}x^5 + \frac{1}{3}x^3 - \frac{1}{2}x^4 \right]_{0,5}^1 \\ &= 30 \left(\frac{1}{5} + \frac{1}{3} - \frac{1}{2} - \frac{1}{5}\frac{1}{2^5} - \frac{1}{3}\frac{1}{2^3} + \frac{1}{2}\frac{1}{2^4} \right) = \frac{1}{2} \end{aligned}$$

3. Halle la función de distribución de X. (2 pts.) Al ser X una variable aleatoria absolutamente continua, la función de distribución vendrá dada por:

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \displaystyle\int_0^x 30t^2(1-t)^2\,dt = 6x^5 - 15x^4 + 10x^3 & \text{si } x \in (0,1) \\ 1 & \text{si } x > 1 \end{cases}$$

4. Halle E(X) y var(X). (2 pts.)
Resulta

$$\begin{aligned} E(X) &= \int_0^1 xf(x)\,dx = 30\int_0^1 x^3(1-x)^2\,dx \\ &= 30B(4,3) = 30\frac{\Gamma(4)\Gamma(3)}{\Gamma(7)} = 30\frac{3!2!}{6!} = \frac{1}{2} = 0,5 \end{aligned}$$

$$\begin{aligned} E(X^2) &= \int_0^1 x^2 f(x)\,dx = 30\int_0^1 x^4(1-x)^2\,dx \\ &= 30B(5,3) = 30\frac{\Gamma(5)\Gamma(3)}{\Gamma(8)} = 30\frac{4!2!}{7!} = \frac{2}{7} = 0,285714 \end{aligned}$$

$$\mathrm{var}(X) = E(X^2) - E(X)^2 = \frac{2}{7} - \left(\frac{1}{2}\right)^2 = \frac{1}{28} = 0,035714$$

5. $X_1,\ldots,X_n$ son variables aleatorias independientes e idénticamente distribuidas con momentos de primer y segundo orden finitos e idénticos $E(X_i) = 0.5$, $\mathrm{var}(X_i) = 0.25$ para $i = 1,\ldots,n$ ¿Cuál es la probabilidad aproximada de que $S_n = X_1 + \ldots + X_n$ tome valores entre 55 y 65, para $n = 100$? (2 pts.)
Usaremos el Teorema del Límite Central. Teniendo en cuenta que

$$\mathrm{E}(S_n) = \mathrm{E}(X_1 + \ldots + X_n) = \mathrm{E}(X_1) + \ldots + \mathrm{E}(X_n) = 0,5n$$

$$\mathrm{var}(S_n) = \mathrm{var}(X_1 + \ldots + X_n) = \mathrm{var}(X_1) + \ldots + \mathrm{var}(X_n) = 0,25n$$

$$\sigma(S_n) = \sigma(X_1 + \ldots + X_n) = \sqrt{0,25n}$$

Por tanto, como

$$\frac{S_n - 0,5n}{\sqrt{0,25n}} \xrightarrow{\mathcal{L}} Z$$

donde Z sigue una normal estandarizada, entonces tendremos

$$p = P(55 \leq S_{100} \leq 65) = P\left(\frac{55-50}{5} \leq \frac{S_{100}-50}{5} \leq \frac{65-50}{5}\right)$$

y, de forma aproximada,

$$
\begin{aligned}
p &= P(55 \leq S_{100} \leq 65) \approx P(1 \leq Z \leq 3) \\
&= \Phi(3) - \Phi(1) \\
&= 0,998650 - 0,841345 = 0,157305
\end{aligned}
$$

donde

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

es la función de distribución normal estandarizada.