Exámenes de Estadística

Julio Muñoz

Septiembre 2006

Universidad de Castilla-La Mancha
Facultad CC Químicas
Primer Curso de Ingeniería Química
Estadística. Examen de Evaluación. 09-05-2003

1. Un sistema eléctrico consta de dos bloques en paralelo de forma que si falla uno de los bloques entonces falla el sistema. Se supone que la probabilidad de que funcione correctamente el primer bloque durante un intervalo de tiempo t es p_1, y la del segundo bloque es p_2. Si se ha comprobado que durante un tiempo t, el sistema ha fallado, ¿cuál es la probabilidad de que este fallo sea debido al primer bloque y no al segundo?. Se supone independencia del funcionamiento entre los dos bloques.

   (1.5 *puntos*)

2. Sea ξ v.a. de tipo continuo cuya función de densidad de probabilidad es
$$f_\xi(x) = \begin{cases} Ce^{-2x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}.$$

   (a) ¿Qué valor hemos de asignar a C para que f_ξ sea realmente una densidad de probabilidad?

   (0.5 puntos)

   (b) Halla la función de distribución así como la esperanza. (1 punto)

   (c) Calcular $P\left\{\xi > 2|_{\xi<4}\right\}$ (0.5 puntos)

3. Una fábrica produce fusibles eléctricos, resultando defectuosos el 3%. Los fusibles se empaquetan en cajas de 24 unidades. Se pide:

   (a) Calcular la probabilidad de que una caja elegida al azar contenga al menos un fusible defectuoso.

   (0.5 puntos).

   (b) Si selecionamos 5 cajas al azar, ¿cuál es la probabilidad de que exactamente en 2 de ellas no haya ningún fusible defectuoso?

   (0.5 puntos).

   (c) Vamos selecionando cajas al azar y comprobando si sus fusibles son defectuosos o no. ¿Cuál es el número medio de cajas que tendremos que inspeccionar hasta encontar algún fusible defectuoso?

   (0.5 puntos).

4. Un laboratorio ha desarrollado una prueba para el diagnóstico de la hepatitis de tipo C. Esta prueba tiene un 95 % de exactitud tanto en los que tienen hepatitis C como entre los que no la tienen. Si el 0.5 % de la población realmente tiene hepatitis C, calcular la probabilidad de que un determinado individuo que tenga tal enfermedad, si la prueba dice que la tiene.

   (2.5 puntos).

5. Una partícula se mueve en un fluido de manera rectilínea, digamos que lo hace a lo largo del eje coordenado OX del plano $\mathbb{R}^2$. Sobre este movimiento se sabe lo siguiente: al cabo de un segundo la partícula se mueve a la derecha 0.001 mm con una probabilidad de $\frac{1}{60}$ y -0.001 mm a la izquierda con una probabilidad $1-\frac{1}{60}$. Se supone además que cada movimiento en un segundo es independiente del movimiento realizado en segundos anteriores. Se pide:

   (a) Calcular la probabilidad de que la partícula realice 15 movimientos a la derecha al cabo de 360 segundos. (0.75 puntos)

   (b) Calcular la probabilidad de que la partícula, al cabo de 360 segundos, esté en el punto $(-0.3, 0)$ si se supone que en el instante inicial (justamente antes de empezar a contabilizar los 360 segundos) se encontraba en el origen de coordenadas $(0, 0)$. (0.75 puntos)

   (c) Encontrar la posición esperada de la partícula al cabo de los 360 segundos. (1 punto)

Universidad de Castilla-La Mancha
Facultad Ciencias Químicas
Estadística. Primer Curso de Ingeniería Química
Examen de convocatoria ordinaria, 27-06-2003

INSTRUCCIONES

- Escribe claramente tus apellidos y nombre, en todas y cada una de las hojas de examen.
- No se admitirá más de una hoja por problema. No escribas partes de problemas diferentes en una misma hoja.
- En función de si uno tiene el examen parcial de evaluación suspenso, la duración, los ejercicios y su puntuación, quedan distribuidos del siguiente modo:

	Ejercicios	Puntuación	Duración
Con parcial aprobado	2,3,4,6	2'5 ptos cada uno	2h.30m
Sin parcial aprobado	1,2,3,5,6,7	2,1'5,2,1,1'5,2 ptos. resp.	2h.30m

EJERCICIOS

1. En tres grandes poblaciones A, B y C la proporción de individuos infectados por un determinado virus es del 30, 60 y 10% respectivamente. Se toma al azar una de las 3 poblaciones (las tres son igual de probables), y de ella elegimos a 10 individuos al azar, resultando que 2 de ellos están infectados. ¿A qué población es más probable que pertenezcan?

2. El número de partículas en suspensión que había en Madrid el día 22 de Febrero estaba distribuido de manera uniforme entre 20000 y 30000 particulas por decímetro cúbico (dm^3). Sabiendo que en cada inspiración introducimos 0'5 dm^3 de aire en nuestros pulmones y que realizamos 10 inspiraciones por minuto, calcular la probabilidad de que el número de partículas inspirado en una hora sea superior a 7510000.

3. Un investigador está estudiando un parámetro θ que representa la duración media (en días) de vida de una rata desde que se le suministra cierto tipo de veneno. Se sabe que esta duración viene dada por una variable aleatoria de distribución $N(\theta, 0.2)$. Les suministra el veneno a 100 ratas resultando que la suma de los tiempos que sobreviven dichas ratas es de 30 días. En base a esto construir un intervalo de confianza del 95 % para el parámetro θ.

4. Una población sigue un distribución de probabilidad cuya función de densidad de masa viene dada por

$$P_\xi^*(x) = p(1-p)^{x-1} \; x = 1, 2, ...$$

siendo p cierto parámetro del intervalo $(0, 1]$. Se pide

   (a) Calcular un estimador de p mediante el método de los momentos (ayuda: $E[\xi] = 1/p$).

   (b) Determinar un estimador de p mediante el método de máxima verosimilitud.

5. Lanzamos un dardo sobre el plano al objeto de acertar en una diana de radio $\sqrt[2]{1.39}$ y centro en el punto de coordenadas $(3, 2)$. Se supone que los lanzamientos tienen carácter aleatorio y que cada impacto (ξ_1, ξ_2) es un vector aleatorio, de suerte que las componentes ξ_1 y ξ_2 son independientes y además $\xi_1 \sim N(3, 1)$ y $\xi_2 \sim N(2, 1)$. Calcular la probabilidad de que el dardo impacte dentro de la diana.

6. Sean, ξ_1 la v.a. que representa a los gastos por impagados en millones de euros de una determinada empresa, y ξ_2 la v.a. de los ingresos netos. Se sabe que la función de densidad del vector (ξ_1, ξ_2) es

$$f_{(\xi_1,\xi_2)}(x, y) = \begin{cases} 5(1+x)e^{20(2-y)} & \text{si } x \in (0,2) \text{ e } y > 2 \\ 0 & \text{en el resto} \end{cases}$$

   (a) Calcular la probabilidad de que los gastos por impagados sean menores a un millón de euros.

   (b) ¿Son independientes los gastos por impagados y los ingresos netos?

   (c) ¿Qué cantidad de ingresos netos se esperan obtener si los impagados ascendieron hasta 1 millón de euros?

7. Una empresa desea determinar la proporción de clientes dispuestos a adquirir uno de sus productos. Estima que dicha proporción es 0.45 ó 0.5. Decidir en base al Principio de Máxima Verosimilitud, una estimación de dicha proporción si después de realizar una muestra aleatoria simple de tamaño 15 entre sus clientes potenciales, 6 de ellos afirmaron estar dispuestos a la adquirir y los 9 restantes no estaban dispuestos a optar por dicho producto.

Resolución del examen convocatoria de 27-06-2003

1. Del enunuciado del problema se desprende lo siguiente: si Inf denota estar infectado y A, B y C haber elegido a dichas poblaciones, entonces

$$P(Inf|\,A) = 3/10,\ P(Inf|\,B) = 6/10,\ P(Inf|\,C) = 1/10,$$

y

$$P(A) = P(B) = P(C) = \frac{1}{3}.$$

Se sabe que S el suceso consistente en la realización de 10 obsevaciones con resultado de 2 infecciones ha tenido lugar. La pregunta es saber cuál de las siguientes cantidades es mayor:

$$P(A|\,S),\ P(B|\,S),\ P(C|\,S).$$

Calculemos: se sabe que

$$P(A|\,S) = \frac{P(A\cap S)}{P(S)} = \frac{P(S|\,A)P(A)}{P(S|\,A)P(A) + P(S|\,B)P(B) + P(S|\,C)P(C)},$$

ahora bien, $P(S|\,A)$ es la probabilidad de que hayan tenido lugar 2 infecciones de 10 individuos sabiendo que la probabilidad de infección en A es $3/10$. Esta probabilidad no es otra cosa que el valor de la masa de un binomial de parámetros $p = 3/10$ y $n = 10$; análogamente para las otras probabilidades condicionadas. Por tanto

$$P(S|\,A) = \binom{10}{2}\left(\frac{3}{10}\right)^2\left(\frac{7}{10}\right)^8$$

y así

$$P(A|\,S) = \frac{\binom{10}{2}\left(\frac{3}{10}\right)^2\left(\frac{7}{10}\right)^8}{\binom{10}{2}\left(\frac{3}{10}\right)^2\left(\frac{7}{10}\right)^8 + \binom{10}{2}\left(\frac{6}{10}\right)^2\left(\frac{4}{10}\right)^8 + \binom{10}{2}\left(\frac{1}{10}\right)^2\left(\frac{9}{10}\right)^8}.$$

Puesto que el denominador que aparece en el cálculo de $P(A|\,S)$, $P(B|\,S)$ y $P(C|\,S)$ es el mismo, basta con comparar

$$\left(\frac{3}{10}\right)^2\left(\frac{7}{10}\right)^8,\left(\frac{6}{10}\right)^2\left(\frac{4}{10}\right)^8,\left(\frac{1}{10}\right)^2\left(\frac{9}{10}\right)^8.$$

El valor más grande es el correspondiente a $P(A|\,S)$ [1]. La población con más probabilidad de haber sido elegida es A.

[1] Para cerciorarse de ello basta con esbozar la gráfica de la función $\psi(p) = p^2(1-p)^8$ y notar que el máximo se da para $p = 0.2$. Los valores en 0.6 y 0.1 están más lejos de este máximo que 0.3.

2. Supongamos que X es la v.a. que proporciona el número de partículas en suspensión (expresada en miles de unidades por dm^3). Entonces $X \sim U(20, 30)$ y por ende

$$E[X] = 25 \text{ y } V[X] = \frac{25}{3}.$$

Realizamos 600 inspiraciones y en cada una introducimos 0'5 dm^3 de aire, esto es, nos llevamos a los pulmones 300 dm^3. Sean $X_1, X_2, ..., X_{300}$ las v.a. que dan el número de partículas por cada una de las 300 veces que almacenamos 1 dm^3. Nos están pidiendo calcular $p = P(\sum_{i=1}^{300} X_i > 7510)$:

$$p = P(\frac{\frac{\sum_{i=1}^{300} X_i}{300} - 25}{\sqrt{\frac{1}{300}\frac{25}{3}}} > \frac{\frac{7510}{300} - 25}{\sqrt{\frac{1}{300}\frac{25}{3}}})$$

pero en virtud del TCL podemos aproximar p por $P(\phi > 0.2)$ con $\phi \sim N(0, 1)$, con lo cual acudimos a las tablas de la normal para aproximar la probabilidad pedida como

$$p \approx P(\phi > 0.2) = 0.5793$$

3. Sea la variable aleatoria X que da el tiempo de vida después de ingerir el veneno: $X \sim N(\theta, 0.2)$. Sea la muestra $X_1, X_2, ..., X_{100}$. Si pretendemos la cosntrucción de un intervalo tomaremos como cantidad pivotal a

$$Q = \frac{\frac{\sum_{i=1}^{100} X_i}{100} - \theta}{\frac{0.2}{\sqrt{100}}}$$

Sabemos que Q está distribuido como una $N(0, 1)$. Desarrollamos y llegamos a que el intervalo es

$$I = \left[\frac{\sum_{i=1}^{100} X_i}{100} - Z_{0.025}\left(\frac{0.2}{\sqrt{100}}\right), \frac{\sum_{i=1}^{100} X_i}{100} + Z_{0.025}\left(\frac{0.2}{\sqrt{100}}\right)\right]$$

Usandolas tablas y los datos del investigador llegamos a que $Z_{0.025} = 1.96$ y a que el intervalo obteniendo:

$$I = [0.260\,8, 0.339\,2].$$

4.

(a) Mediante el método de los momentos: $\overline{\xi} = E[\xi] = 1/p$, lo que da lugar a $p = \frac{1}{\overline{\xi}}$; por tanto el estimador pedido es

$$p^* = \frac{1}{\overline{\xi}}$$

(b) El EMV es el que maximiza la función de verosimilitud

$$\begin{aligned} l &= \log(p(1-p)^{x_1-1}p(1-p)^{x_2-1}...p(1-p)^{x_n-1}) \\ &= n\log p - n\log(1-p) + \left(\sum_{i=1}^{n} x_i\right)\log(1-p) \end{aligned}$$

Derivamos con respecto a p e igualamos a cero:

$$l' = \frac{n}{p} + \frac{n - \left(\sum_{i=1}^{n} x_i\right)}{1-p} = 0$$

La solución es $p = \frac{n}{\left(\sum_{i=1}^{n} x_i\right)} = \frac{1}{\overline{\xi}}$. Puesto que $l'' = \frac{-n}{p^2} + \frac{\left(n - \left(\sum_{i=1}^{n} x_i\right)\right)}{(1-p)^2} < 0$ (¿por qué?) el estimador máximo verosimil es $p^* = \frac{1}{\overline{\xi}}$.

5. Si usamos el Teorema de Pitágoras nos daremos cuenta de que nos están pidiendo que evaluemos la probabilidad

$$p = P\left\{(\xi_1 - 3)^2 + (\xi_2 - 2)^2 \leq 1.39\right\},$$

y puesto que $(\xi_1 - 3)^2$ y $(\xi_2 - 2)^2$ son cuadrados de normales $(0,1)$, entonces $\Psi = (\xi_1 - 3)^2 + (\xi_2 - 2)^2$ se distribuye como una $\mathcal{X}_2^2$. Buscamos y obtenemos

$$p = P(\Psi \leq 1.39) = 0.5$$

6.

(a)

$$\begin{aligned} P(\xi_1 &\leq 1, \xi_2 \text{ cualquiera}) = P(\xi_1 \leq 1, -\infty < \xi_2 < \infty) \\ &= \int_{-\infty}^{1}\left(\int_{-\infty}^{+\infty} f_{(\xi_1,\xi_2)}(x,y)\,dy\right)dx \\ &= \int_{0}^{1} 5(1+x)\left(\int_{2}^{+\infty} e^{40-20y}dy\right)dx \\ &= \frac{3}{8} \end{aligned}$$

(b) Las densidades marginales son

$$\int_0^2 5(1+x)\left(\int_2^{+\infty} e^{40-20y}dy\right)dx$$

$$\begin{aligned}
f_{\xi_1}(x) &= \int_{\mathbb{R}} f_{(\xi_1,\xi_2)}(x,y)\,dy \\
&= \begin{cases} \int_2^{+\infty} 5(1+x)e^{40-20y}dy & \text{si } x \in (0,2) \\ 0 & \text{si } x \notin (0,2) \end{cases} \\
&= \begin{cases} \frac{1}{4}(1+x) & \text{si } x \in (0,2) \\ 0 & \text{si } x \notin (0,2) \end{cases}
\end{aligned}$$

y

$$\begin{aligned}
f_{\xi_2}(y) &= \int_{\mathbb{R}} f_{(\xi_1,\xi_2)}(x,y)\,dx \\
&= \begin{cases} \int_0^2 5(1+x)e^{40-20y}dx & \text{si } y \in (2,\infty) \\ 0 & \text{si } y \notin (2,\infty) \end{cases} \\
&= \begin{cases} 20e^{40-20y} & \text{si } y \in (2,\infty) \\ 0 & \text{si } y \notin (2,\infty) \end{cases}
\end{aligned}$$

Puesto que $f_{(\xi_1,\xi_2)}(x,y) = f_{\xi_1}(x)\, f_{\xi_2}(y)$ las variables son independientes.

(c) Nos piden $E[\xi_2 | \xi_1 = 1]$ pero dado que hay independencia

$$\begin{aligned}
E[\xi_2 | \xi_1 = 1] &= E[\xi_2] \\
&= \int_2^{+\infty} 20ye^{40-20y}dy \\
&= \frac{41}{20}
\end{aligned}$$

7. θ es 0.45 ó 0.5. Habremos de evaluar $P_\xi^*(6)$ cuando $\xi \sim B(15,\theta)$:

	$P_\xi^*(6)$
$\theta = 0.45$	$\binom{15}{6}(0.45)^6(0.65)^9$
$\theta = 0.5$	$\binom{15}{6}(0.5)^6(0.5)^9$

El valor mayor se corresponde con $\theta = 0.45$, luego esta sería la estimación máximo verosimil.

Universidad de Castilla-La Mancha
Facultad Ciencias Químicas
Estadística. Primer Curso de Ingeniería Química (Incluido P.A.)
Examen de Septiembre, 11-09-2003

INSTRUCCIONES

- Escribe claramente tus apellidos y nombre, en todas y cada una de las hojas de examen.
- No se admitirá más de una hoja por problema. No escribas partes de problemas diferentes en una misma hoja. La puntuación de cada uno de los problemas es de 2.5 puntos.
- Razona todas las respuestas y da a conocer los resultados que estás utilizando en la resolución de los ejercicios que a continuación se proponen.

EJERCICIOS

1. En tres grandes poblaciones A, B y C la proporción de individuos infectados por un determinado virus es del 30, 60 y 10% respectivamente. Se toma al azar una de las 3 poblaciones (las tres son igual de probables), y de ella elegimos a 10 individuos al azar, resultando que 2 de ellos están infectados. ¿A qué población es más probable que pertenezcan?

2. El número de partículas en suspensión que había en Madrid el día 22 de Febrero estaba distribuido de manera uniforme entre 20000 y 30000 particulas por decímetro cúbico (dm^3). Sabiendo que en cada inspiración introducimos 0'5 dm^3 de aire en nuestros pulmones y que realizamos 10 inspiraciones por minuto, calcular la probabilidad de que el número de partículas inspirado en una hora sea superior a 7510000.

3. Un investigador está estudiando un parámetro θ que representa la duración media (en días) de vida de una rata desde que se le suministra cierto tipo de veneno. Se sabe que esta duración viene dada por una variable aleatoria de distribución $N(\theta, 0.2)$. Les suministra el veneno a 100 ratas resultando que la suma de los tiempos que sobreviven dichas ratas es de 30 días. En base a esto construir un intervalo de confianza del 95 % para el parámetro θ.

4. Una población sigue un distribución de probabilidad cuya función de densidad de masa viene dada por

$$P_\xi^*(x) = p(1-p)^{x-1} \; x = 1, 2, ...$$

siendo p cierto parámetro del intervalo $(0, 1]$. Se pide

(a) Calcular un estimador de p mediante el método de los momentos (ayuda: $E[\xi] = 1/p$).

(b) Determinar un estimador de p mediante el método de máxima verosimilitud.

Universidad de Castilla-La Mancha
Facultad de Ciencias del Medio Ambiente
Tercer curso de CC Ambientales
Estadística, 13-12-05. Examen tipo A

1. Elegido un individuo de una determinada especie arborea, de un determinado bosque, y observado en laboratorio, se observó que estaba afectado por un el hongo saza. La probabilidad de que en la población de la que se eligió el individuo, uno de ellos esté afectado por saza es de 0.01. Se sabe además que la probabilidad de que el aparato de laboratorio encargado de examinar la infección, detecte que un individuo está afectado estándolo es 0.97, y no estándolo es 0.001. ¿Qué podemos decir acerca de la probabilidad de acierto del laboratorio para el individuo que hemos elegido? (2.5 puntos)

2. Se sabe que la página impresa de un libro contiene 3000 espacios (que pueden estar en blanco u ocupados con algún símbolo). Se supone que el linotipista comete 2 errores cada 6000 espacios. Calcular

   (a) La probabilidad de que una página no contenga errores. (1.25 puntos)

   (b) La probabilidad de que un capítulo de 16 páginas no contenga errores. (1.25 puntos)

3. Sea $X \sim N(3, 0.5)$. Calcular las siguientes probabilidades: $P(X \leq 3.32)$, $P(2.15 \leq X \leq 3.35)$ y $P(X^2 \geq 4)$. (Usar los datos: $F(-0.64) = 1-0.7389$, $F(0.7) = 0.758$, $F(1.7) = 0.9954$, $F(-2) = 0.02$, $F(-5) = 0$, donde F es la función de distribución de una $N(0,1)$) (2.5 puntos)

4. La vida de un virus tiene una duración modelada por una siguiente variable aleatoria ξ cuya función de densidad es

$$f_\xi(x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{k}{x^4} & \text{si } x \geq 1 \end{cases}$$

   (a) Determinar k para que f_ξ sea efectivamente una función de densidad de probabilidad. (0.5 puntos)

   (b) Hallar el valor esperado de vida del virus (0.75 puntos)

(c) Hallar la mediana (0.75 puntos)

(d) Entre las duraciones calculadas en los dos aprtados anteriores, ¿cuál de las dos sería más representativa? Justifica la respuesta. (0.5 puntos)

Universidad de Castilla-La Mancha
Facultad de Ciencias del Medio Ambiente
Tercer curso de CC Ambientales
Estadística, 13-12-05. Examen tipo B

1. Una persona tiene dos negocios en funcionamiento, A y B. El primero puede producir mayor beneficio, pero en el 25% de las operaciones arroja pérdida, mientras que en el segundo, donde la perspectiva es menor, arroja pérdida sólo el 5% de los casos. Se supone que el conjunto de operaciones es análogo en ambos negocios. Si, analizando el resultado económico de una de las operaciones, arrojase pérdida, ¿cuál sería la probabilidad de que dicha operación correspondiese al negocio B?

   (2.5 puntos)

2. Se sabe que la probabilidad de que una planta reaccione desfavorablemente tras la realización de un injerto es de 0.002.

   (a) Determinar la probabilidad de que en un grupo de 2000 plantas injertadas haya como mucho 3 que reaccionen desfavorablemente. (1.25 puntos)

   (b) ¿Cual sería la probabilidad de que el número de reacciones desfavorables esté comprendido entre 1 y 6 si sabemos que hay a lo sumo 2 que han reaccionado desfavorablemente? (1.25 puntos)

3. Sea $X \sim N(3, 0.5)$. Calcular las siguientes probabilidades: $P(X \leq 3.32)$, $P(2.15 \leq X \leq 3.35)$ y $P(X^2 \geq 4)$. (Usar los datos: $F(-0.64) = 1-0.7389$, $F(0.7) = 0.758$, $F(1.7) = 0.9954$, $F(-2) = 0.02$, $F(-5) = 0$, donde F es la función de distribución de una $N(0, 1)$)

   (2.5 puntos)

4. La longitud en milímetros ξ que se puede estirar un filamento de nylon sin ruptura sigue una distribución de probabilidad cuya densidad de probabilidad es

$$f_\xi(x) = \begin{cases} 5e^{-5x} & \text{si } x \geq 0 \\ 0 & \text{en el resto} \end{cases}$$

   (a) Un filamento se dice que pasa el control de calidad si se estira entre 0.18 y 0.22 milímetros. Calcular la probabildad de que un filamento pasa dicho control.

   (0.75 puntos)

(b) Determinar la longitud esperada de estiramiento para un filamento de nylon.

(0.75 puntos)

(c) Se examinan 100 filamentos. Determinar la probabilidad de que al menos dos de ellos superen el control de calidad.

(1 puntos)

Solución del examen TIPO A:

1. Probabilidad de infección es 0.01, i.e. $P(\text{infec.}) = 0.01$. el resto de la información se escribe
$$P\left(DI|\,\text{infec.}\right) = 0.97,\ \ P\left(DI|\,\overline{\text{infec.}}\right) = 0.001.$$
Para analizar el éxito del laboratorio en el caso dado hemos de calcular
$$\begin{aligned}
P\left(\text{infec.}|\,DI\right) &= \frac{P\left(DI|\,\text{infec.}\right)P\left(\text{infec.}\right)}{P\left(DI\right)} \\
&= \frac{P\left(DI|\,\text{infec.}\right)P\left(\text{infec.}\right)}{P\left(DI|\,\text{infec.}\right)P\left(\text{infec.}\right) + P\left(DI|\,\overline{\text{infec.}}\right)P\left(\overline{\text{infec.}}\right)} \\
&= \frac{(0.97)\,(0.01)}{(0.97)\,(0.01) + (0.001)\,(1-0.01)} \approx 0.907
\end{aligned}$$
lo que se traduce en alta fiabiliada para la situación descrita.

2. Podemos modelizar la v.a. $\xi =$ número de erratas/pág. con un binomial $B(3000, \frac{2}{6000})$. Aproximamos la distribución de esta v.a. por la de una Poisson de parámetro $\lambda = 3000\frac{2}{6000} = 1$.

   (a) La probabilidad de que una página no contenga errores coincide con
   $$P_{\xi}^{*}\left(0\right) = e^{-1}\frac{1^0}{0!} = e^{-1}.$$

   (b) Puesto que hay independencia en las erratas entre páginas tnedremos que la probabilidad de que un capítulo de 16 páginas no contenga errores es $\left(e^{-1}\right)^{16}$.

3. Como $X \sim N(3, 0.5)$ entonce $\psi = \frac{X-3}{0.5} \sim N(0,1)$. Así
$$\begin{aligned}
P(X &\leq 3.32) = P\left(\psi = \frac{X-3}{0.5} \leq \frac{3.32-3}{0.5} = 0.64\right) = F_{\psi}(0.64) \\
&= 1 - F_{\psi}(-0.64) = 0.7389
\end{aligned}$$
$$\begin{aligned}
P(2.15 &\leq X \leq 3.35) = \\
P(-1.7 &= \frac{2.15-3}{0.5} \leq \frac{X-3}{0.5} \leq \frac{3.35-3}{0.5} = 0.7) \\
&= F_{\psi}(0.7) - F_{\psi}(-1.7) \\
&= 0.758 - (1 - F_{\psi}(1.7)) = 0.758 - (1 - 0.9954) \\
&= 0.753\,4
\end{aligned}$$

$$\begin{aligned} P(X^2 &\geq 4) = P(X \geq 2) + P(X \leq -2) \\ &= P(\frac{X-3}{0.5} \geq -2) + P(\frac{X-3}{0.5} \leq \frac{-2-3}{0.5} = -10) \\ &= (1 - F_\psi(-2)) + F_\psi(-10) \\ &= (1 - 0.02) + 0 = 0.98 \end{aligned}$$

ya que $F_\psi(-10) \leq F_\psi(-5) = 0.\ :\ 2.275013 \times 10^{-2}$

4. La vida de un virus tiene una duración modelada por una siguiente variable aleatoria ξ cuya función de densidad es

$$f_\xi(x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{k}{x^4} & \text{si } x \geq 1 \end{cases}$$

(a) Las condiciones que ha de cumplir la densidad son $f_\xi(x) \geq 0$ y $\int_{\mathbb{R}} f_\xi(x)\, dx = 1$. La primera condicón establece que $k \geq 0$ y la segunda

$$1 = \int_{\mathbb{R}} f_\xi(x)\, dx = \int_1^\infty \frac{k}{x^4} dx = \frac{1}{3} k.$$

De resultas $k = 3$

(b) Hallar el valor esperado de vida del virus es la esperanza o primer momento:

$$\alpha_1 = \int_{\mathbb{R}} x f_\xi(x)\, dx = \int_1^\infty \frac{3x}{x^4} dx = \frac{3}{2}$$

(c) La mediana es aquel valor y tal que

$$0.5 = \int_1^y \frac{3}{x^4} dx,$$

y así, después de integrar, da como resultado $y = 1.26$.

(d) La varianza es $\sigma^2 = \alpha_2 - \alpha_1^2$; como

$$\alpha_2 = \int_1^\infty \frac{3x^2}{x^4} dx = 3,$$

entonces $\sigma^2 = \alpha_2 - \alpha_1^2 = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}$. La desviación típica será $\sigma = \sqrt[2]{\frac{3}{4}} = 0.8660254$

La desviación absoluta con respecto a la mediana es

$$D = \int_1^\infty \frac{3\,|x - Me|}{x^4} dx$$

siendo

$$Me = y = 1.26$$

Así $D = \int_1^\infty \frac{3|x-1.26|}{x^4}dx = 0.389\,881\,6$, que como es menor que σ, diremos que la mediana es mejor representante en elsentido de que la dispersión asociada es menor.

Solución del examen TIPO B

1. La información de qu disponemos es la siguiente:

$$\begin{aligned} P(A) &= P(B) = 0.5 \\ P\left(Perd|\, A\right) &= 0.25,\ P\left(Perd|\, B\right) = 0.05 \end{aligned}$$

   Nos están pidiendo $P\left(B|\, Perd\right)$ y esto, en virtud de la fórmula de la probabilidad condicionada es

$$\begin{aligned} P\left(B|\, Perd\right) &= \frac{P\left(Perd|\, B\right)P(B)}{P\left(Perd\right)} \\ &= \frac{P\left(Perd|\, B\right)P(B)}{P\left(Perd|\, B\right)P(B) + P\left(Perd|\, A\right)P(A)} \\ &= \frac{(0.05)\,(0.5)}{(0.05)\,(0.5) + (0.25)\,(0.5)} \approx 0.16 \end{aligned}$$

2. Se sabe que la probabilidad de que una planta reaccione desfavorablemente tras la realización de un injerto es de 0.002, i.e. $P(Rd) = 0.002$. Se inspeccionan 2000 plantas, de forma que si definimos ξ, la v.a. que da el número de reacciones desfavorables, entonces $\xi \sim B(2000, 0.002)$. La distribución de esta v.a. se puede aproximar por una Poisson de parámentro $\lambda = np = 4$. Hecho esto contestamos a las cuestiones formuladas

   (a) Calculamos

$$\begin{aligned} P\left(\xi \leq 3\right) &= P\left(\xi = 0\right) + P\left(\xi = 1\right) + P\left(\xi = 2\right) + P\left(\xi = 3\right) \\ &= e^{-4}\left(1 + \frac{4}{1} + \frac{16}{2} + \frac{64}{6}\right) = 0.43. \end{aligned}$$

   (b)

$$\begin{aligned} P\left(\xi \in [1,6]|\, \xi \leq 2\right) &= \frac{P\left(\xi \in [1,2]\right)}{P\left(\xi \leq 2\right)} \\ &= \frac{P\left(\xi = 1\right) + P\left(\xi = 2\right)}{P\left(\xi = 0\right) + P\left(\xi = 1\right) + P\left(\xi = 2\right)} \\ &= \frac{\left(\frac{4}{1} + \frac{16}{2}\right)}{\left(1 + \frac{4}{1} + \frac{16}{2}\right)} = 0.92 \end{aligned}$$

3. Como $X \sim N(3, 0.5)$ entonce $\psi = \frac{X-3}{0.5} \sim N(0,1)$. Así

$$\begin{aligned} P(X &\leq 3.32) = P\left(\psi = \frac{X-3}{0.5} \leq \frac{3.32-3}{0.5} = 0.64\right) = F_\psi(0.64) \\ &= 1 - F_\psi(-0.64) = 0.7389 \end{aligned}$$

$$\begin{aligned} P(2.15 &\leq X \leq 3.35) = \\ P(-1.7 &= \frac{2.15-3}{0.5} \leq \frac{X-3}{0.5} \leq \frac{3.35-3}{0.5} = 0.7) \\ &= F_\psi(0.7) - F_\psi(-1.7) \\ &= 0.758 - (1 - F_\psi(1.7)) = 0.758 - (1 - 0.9954) \\ &= 0.7534 \end{aligned}$$

$$\begin{aligned} P(X^2 &\geq 4) = P(X \geq 2) + P(X \leq -2) \\ &= P(\frac{X-3}{0.5} \geq -2) + P(\frac{X-3}{0.5} \leq \frac{-2-3}{0.5} = -10) \\ &= F_\psi(-2) + F_\psi(-10) \\ &= 0.027 + 0 = 0.027 \end{aligned}$$

ya que $F_\psi(-10) \leq F_\psi(-5) = 0$.

4. La longitud en milímetros ξ que se puede estirar un filamento de nylon sin ruptura sigue una distribución de probabilidad cuya densidad de probabilidad es

$$f_\xi(x) = \begin{cases} 5e^{-5x} & \text{si } x \geq 0 \\ 0 & \text{en el resto} \end{cases}$$

   (a) La probabildad de que un filamento se estire entre 0.18 y 0.22 milímetros es

$$\begin{aligned} & P\left(.18 < \xi < .22\right) \\ = & \int_{0.18}^{0.22} 5e^{-5x} dx = -e^{-5x}\Big|_{0.18}^{0.22} \\ = & e^{-5(0.18)} - e^{-5(0.22)} \approx 0.074 \end{aligned}$$

   (b) La longitud esperada en milímetros es

$$E[\xi] = \int_0^\infty 5x^2 e^{-5x} dx = 1/5.$$

   ($\alpha_2 = \frac{2}{25}$ y $\sigma^2 = \frac{2}{25} - \frac{1}{25} = \frac{1}{25}$).

(c) La probabilidad de que un filamento supere el control ha sidoe calculado en el primer apartado. Lo que hacemos ahora es inspeccionar 100 filamentos, de modo que si ψ es la v.a. que da el número de filamentos que superan el control, entonces lo que se pide es

$$\begin{aligned} P(\psi \geq 2) &= 1 - P(\psi < 1) \\ &= 1 - P(\psi = 0) + P(\psi = 1) \\ &= 1 - (0.074)^0 (1 - 0.074)^{100} - (100)(0.074)^1 (1 - 0.074)^{99} \\ &= 0.99 \end{aligned}$$

Universidad de Castilla-La Mancha
Facultad CC del Medio Ambiente
Examen de Estadística. Tercer curso de CC Ambientales
Convocatoria extraordinaria de Diciembre 2005

1. Un laboratorio ha desarrollado una prueba para el diagnóstico de la hepatitis de tipo C. Esta prueba tiene un 95 % de exactitud tanto en los que tienen hepatitis C como entre los que no la tienen. Si el 0'5 % de la población realmente tiene hepatitis C, calcular la probabilidad de que un determinado individuo tenga tal enfermedad, si la prueba dice que la tiene. [2 puntos]

2. Emplear el Teorema Central de Límite para resolver las siguientes cuestiones:

   (a) Una compañía aérea sabe por experiencia que el 12% de las reservas telefónicas de plazas no se llevan a efecto, de modo que reserva más plazas de las que dispone. Si en un vuelo hay 150 plazas, ¿cuántas reservas puede hacer la compañía para que la probabilidad de cubrir al menos 145 plazas sea del 99%? [1 punto]

   (b) Si la compañía reserva 160 plazas, ¿cuál es la probabilidad de que, al menos un pasajero no tenga plaza disponible a la hora de embarcar? [1 punto]

3. La distancia ξ entre un árbol cualquiera y el árbol más próximo a él en un bosque sigue una distribución con función de densidad de probabilidad

$$f_\xi(x) = \begin{cases} 2\theta x \exp\left(-\theta x^2\right) & \text{si } x \geq 0 \\ 0 & \text{en el resto} \end{cases}$$

   ($\theta > 0$). Obtener el estimador de máxima verosimilitud de θ, supuesto que se realiza una m. a. s. de tamaño n. Realiza la estimación de máxima verosimilitud del parámetro θ si $n = 10$ y la realización de la muestra es

$$\left\{ \begin{array}{c} x_1 = 2,\ x_2 = 2.1,\ x_3 = 3.1,\ x_4 = 2.8,\ x_5 = 1.9, \\ x_6 = 1.45,\ x_7 = 1.55,\ x_8 = 1.77,\ x_9 = 2,\ x_{10} = 2.6 \end{array} \right\}$$

   [2 puntos]

4. El coseno ξ del ángulo con el que se emiten los electrones en un proceso radiactivo es una variable aleatoria con densidad de probabilidad

$$f_{\xi}(x) = \begin{cases} \frac{1+\theta x}{2} & \text{si } x \in [-1,1] \\ 0 & \text{en el resto} \end{cases}$$

donde $\theta \in [-1,1]$. Dada una m.a.s. $\xi_1, ..., \xi_n$, encontrar el estimador de θ por el método de los momentos. [2 puntos]

5. ¿Cuál es el intervalo de confianza para la varianza de una población que se distribuye normalmente y donde se supone que la esperanza es desconocida? Aplica este intervalo al siguiente ejemplo: un metalúrgico ha hecho 4 determinaciones del punto de fusión del manganeso, 1269, 1271, 1263 y 1265 grados C. Tras valorar los resultados y los posibles errores decidió que la varianza habría de ser 1 ó 2. ¿ Es posible esta afirmación con los datos experimentales obtenidos, suponiendo normalidad y a un nivel de confianza del 95 % ?. [2 puntos]

Resolución

1. Según los datos del problema tenemos que, denotando

   - $\mathtt{PrH}$ = la prueba da resultado positivo de hepatitis
   - $\mathtt{PrNH}$ = la prueba da resultado negativo de hepatitis
   - $\mathtt{PeH}$ = la persona tiene hepatitis
   - $\mathtt{PeNH}$ = la persona no tiene hepatitis,

$$\begin{aligned} P(\mathtt{PeH}) &= 0.005, \\ P\left(\mathtt{PrH}|\,\mathtt{PeH}\right) &= P\left(\mathtt{PrNH}|\,\mathtt{PeNH}\right) = 0.95, \\ P\left(\mathtt{PrNH}|\,\mathtt{PeH}\right) &= P\left(\mathtt{PrH}|\,\mathtt{PeNH}\right) = 0.05. \end{aligned}$$

   Por otra parte sabemos que en virtud del Teorema de Bayes

$$\begin{aligned} P\left(\mathtt{PeH}|\,\mathtt{PrH}\right) &= \frac{P\left(\mathtt{PeH} \cap \mathtt{PrH}\right)}{P\left(\mathtt{PrH}\right)} \\ &= \frac{P\left(\mathtt{PrH}|\,\mathtt{PeH}\right) P(\mathtt{PeH})}{P\left(\mathtt{PrH}|\,\mathtt{PeH}\right) P(\mathtt{PeH}) + P\left(\mathtt{PrH}|\,\mathtt{PeNH}\right) P(\mathtt{PeNH})}. \end{aligned}$$

   Sustituimos ahora los datos de arriba y obtenemos la solución:

$$\boxed{P\left(\mathtt{PeH}|\,\mathtt{PrH}\right) = \frac{0.95\ 0.005}{0.95\ 0.005 + 0.05\ 0.995} \approx 0.0871}$$

2.
$$P\left\{\sum_{i=1}^{n} \xi_i \geq 145\right\} = P\left\{\eta = \frac{\sum_{i=1}^{n} \xi_i - n\,(0.88)}{\sqrt{n\,(0.88)\,(0.12)}} > \frac{145 - n\,(0.88)}{\sqrt{n\,(0.88)\,(0.12)}}\right\} = 0.99.$$

   Empleando el TCL tendremos la aporximación

$$P\left\{\sum_{i=1}^{n} \xi_i \geq 151\right\} \approx P\left\{N(0,1) > c\right\} = 0.99$$

   con $c = \frac{145 - n(0.88)}{\sqrt{n(0.88)(0.12)}}$. Pero esta estimación implica (mirar tablas de la normal)

$$c = \frac{145 - n\,(0.88)}{\sqrt{n\,(0.88)\,(0.12)}} = -2.325,$$

cuya solución aproximada es $n = 177$. Basta que el número de reservas sea mayor o igual que 177 para cubrir 145 de las plazas.
La segunda cuestión se resuelve de forma parecida:

$$\begin{aligned}
P\left\{\sum_{i=1}^{160} \xi_i \geq 151\right\} &= P\left\{\eta \geq \frac{151-(0.88)(160)}{\sqrt{0.1056\,(160)}} = 2.481\,5\right\} \\
&\approx 1 - P\{\eta < 2.481\,5\} = 1 - 0.9934 \\
&= 0.0066
\end{aligned}$$

3. Sea la función de verosimilitud

$$\begin{aligned}
L &= f_\xi(x_1)\ldots f_\xi(x_n) = 2\theta x_1 \exp\left(-\theta x_1^2\right)\ldots 2\theta x_n \exp\left(-\theta x_n^2\right) \\
&= \left\{\begin{array}{cc} (2\theta)^n \exp\left(-\theta \sum_{i=1}^n x_i^2\right) & \text{si } x_i \geq 0, \text{ para todo } i \\ 0 & \text{en el resto.} \end{array}\right.
\end{aligned}$$

Obviamente esta función alcanza el máximo cuando L toma cuando todos los x_i son positivos. Esto es equivalente a maximizar la función

$$l = \log L = n \log(2\theta) - \theta \sum_{i=1}^n x_i^2, \ (\theta > 0).$$

$$l' = \frac{n}{\theta} - \sum_{i=1}^n x_i^2 = 0$$

equivale a

$$\theta = \frac{n}{\sum_{i=1}^n x_i^2};$$

como $l'' = \frac{-n}{\theta^2} < 0$ entonces la función es estrictamente cóncava. Esto implica que para $\theta = \frac{n}{\sum_{i=1}^n x_i^2}$ se alcanza el máximo y por tanto

$$\theta^* = \frac{n}{\sum_{i=1}^n \xi_i^2}$$

es el estimador de máxima verosimilitud. La estimación con las datos dados es

$$\theta^* = \frac{10}{\sum_{i=1}^{10} (x_i)^2} = 0.208\,908\,3$$

4. Según este método debemos resolver

$$\alpha_1 = \overline{\xi}$$

i.e.

$$\int_{-1}^{1} \frac{1+\theta x}{2} x dx = \frac{1}{3}\theta = \overline{\xi};$$

por tanto el estimador pedido es

$$\theta^* = 3\overline{\xi}.$$

5. Para estimar la varianza se emplea la cantidad pivotal

$$Q = \frac{n-1}{\sigma^2} S^2 \sim \varkappa^2_{n-1}.$$

Al buscar puntos $q_1 < q_2$ tales que

$$P\left(q_1 \leq Q \leq q_2\right) = 1 - \alpha$$

se toman

$$q_1 = \varkappa^2_{n-1,1-\frac{\alpha}{2}} < q_2 = \varkappa^2_{n-1,\frac{\alpha}{2}}.$$

Llevamos esta información a la doble desigualdad $q_1 \leq Q \leq q_2$. Operando en esta desigualdad se llega a

$$1 - \alpha = P\left(q_1 \leq Q \leq q_2\right) = P\left(\frac{n-1}{\varkappa^2_{n-1,\frac{\alpha}{2}}} S^2 \leq Q \leq \frac{n-1}{\varkappa^2_{n-1,1-\frac{\alpha}{2}}} S^2\right);$$

esto significa que el intervalo pedido es

$$I = \left[\frac{n-1}{\varkappa^2_{n-1,\frac{\alpha}{2}}} S^2, \frac{n-1}{\varkappa^2_{n-1,1-\frac{\alpha}{2}}} S^2\right].$$

Aplicamos esto al ejemplo expuesto: realizamos el estadístico S^2 con ayuda de la muestra dada

$$\begin{aligned} S^2 &= \frac{1}{n-1} \sum_{i=1}^{4} \left(\xi_i - \overline{\xi}\right)^2 \\ &= \frac{1}{3}\left[2^2 + 4^2 + (-4)^2 + (-2)^2\right] \\ &= \frac{40}{3} \end{aligned}$$

(nótese que $\overline{\xi} = 1267$).

$$\begin{aligned} q_1 &= \chi^2_{3,0.025} = 9.35, \\ q_2 &= \chi^2_{3,0.975} = 0.216. \end{aligned}$$

Por tanto el intervalo para este ejemplo es

$$I = \left[\frac{40}{9.35}, \frac{40}{0.216} \right] = [4.278, 181.18] \, .$$

Como $1{,}2 \notin I$ entonces no podremos afirmar que σ^2 es 1 ó 2 a un nivel de confianza 0'95.

Universidad de Castilla-La Mancha
Facultad Ciencias del Medio Ambiente
Estadística. Tercer Curso
Examen de convocatoria ordinaria, 25-01-2005

INSTRUCCIONES

- Escribe claramente tus apellidos y nombre, en todas y cada una de las hojas de examen.

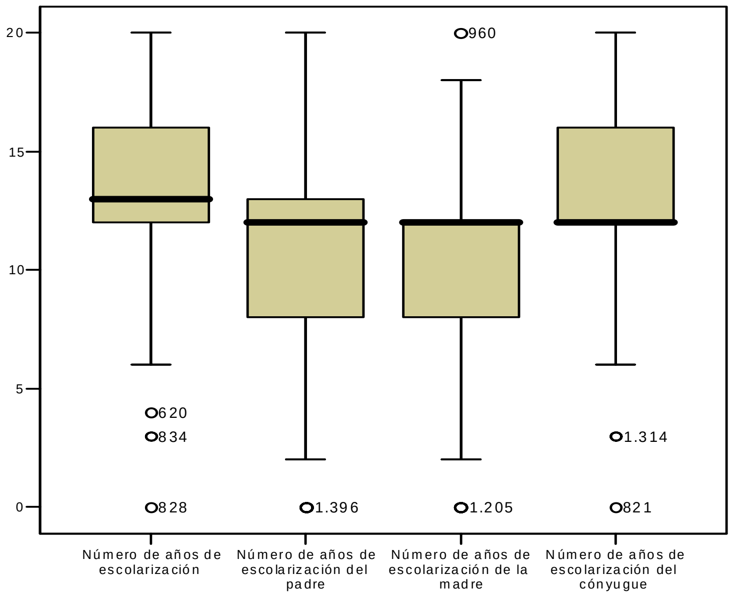
- No se admitirá más de una hoja por problema. No escribas partes de problemas diferentes en una misma hoja. Los problemas se han de entregar por separado.

- En función de si uno tiene el examen parcial de evaluación suspenso, la duración, los ejercicios y su puntuación, quedan distribuidos del siguiente modo:

	Ejercicios	Puntuación	Duración
Con parcial aprobado	3,4,5,7 (8 opcional)	2'5 ptos cada uno (hasta 1 pto)	3h.
Sin parcial aprobado	1,2,3,4,6,7,8	1,1'5,1'5,2,1,1'5,1'5 ptos. resp.	3h.

EJERCICIOS

1. En una encuesta realizada sobre la población de los EE.UU. de Norte America, se han estudiado entre otras, cuatro variables de interés relacionadas con el nivel de estudios. Éstas son: Número de años de escolarización de la persona encuestada, de su padre, de su madre y de su cónyuge. Se han obtenido los diagramas de cajas correspondientes a las cuatro variables así como unos estadísticos descriptivos con el fin de caracterizar estas variables, y se han obtenido los siguientes resultados (téngase en cuenta que no todos los encuestados respondieron a todas las preguntas):



Estadísticos descriptivos

	Número de años de escolarización	Número de años de escolarización del padre	Número de años de escolarización del cónyugue	Número de años de escolarización de la madre	N válido (según lista)
N	1510	1069	790	1233	535
Rango	20	20	20	20	
Mínimo	0	0	0	0	
Máximo	20	20	20	20	
Suma	19455	11632	10184	13298	
Media	12.88	10.88	12.89	10.79	
Desv. típ.	2.984	4.129	3.059	3.463	
Varianza	8.904	17.045	9.359	11.990	
Asimetría	-.168	-.184	-.237	-.646	
	.063	.075	.087	.070	
Curtosis	.710	-.094	1.098	.956	
	.126	.149	.174	.139	

A la vista de los resultados responda a las siguientes preguntas:

(a) Valore las diferencias estadísticas más significativas entre los niveles educativos de la generación anterior y la que respondía a la encuesta.

(b) Dónde se produce mayor cambio generacional en los parámetros de posición o en los de dispersión. Justifique la respuesta.

(c) Valore la simetría de las distribuciones. Justifique el resultado

(d) Valore el apuntamiento de las distribuciones. Justifique la respuesta.

(e) Hay alguna información en el rango. ¿Por qué?¿Hay alguna información en el rango intercuartílico? ¿Por qué?

(f) ¿Diría que las distribuciones son normales? Justifique la respuesta.

2. En tres grandes poblaciones A, B y C la proporción de individuos infectados por un determinado virus es del 30, 60 y 10% respectivamente. Se toma al azar una de las 3 poblaciones (las tres son igual de probables), y de ella elegimos a 10 individuos al azar, resultando que 2 de ellos están infectados. ¿A qué población es más probable que pertenezcan?

3. El número de partículas en suspensión que había en Madrid el día 22 de Febrero estaba distribuido de manera uniforme entre 20000 y 30000 particulas por decímetro cúbico (dm^3). Sabiendo que en cada inspiración introducimos 0'5 dm^3 de aire en nuestros pulmones y que realizamos 10 inspiraciones por minuto, calcular la probabilidad de que el número de partículas inspirado en una hora sea superior a 7510000 (Utilizar el Teorema Central del Límite).

4. Un investigador está estudiando un parámetro θ que representa la duración media (en días) de vida de una rata desde que se le suministra cierto tipo de veneno. Se sabe que esta duración viene dada por una variable aleatoria de distribución $N(\theta, 0.2)$. Les suministra el veneno a 100 ratas resultando que la suma de los tiempos que sobreviven dichas ratas es de 30 días. En base a esto construir un intervalo de confianza del 95% para el parámetro θ. Interpretar convenientemente el resultado obtenido.

5. Una población sigue una distribución de probabilidad cuya función de densidad de masa viene dada por

$$P_{\xi}^{*}(x) = p(1-p)^{x-1} \; x = 1, 2, ...$$

siendo p cierto parámetro del intervalo $(0, 1]$. Se pide determinar el estimador de p mediante el método de máxima verosimilitud ¿Cuál sería la estimación correspondiente si la realización de $\overline{\xi}$ es $\overline{x} = 6.5$?

6. Lanzamos un dardo sobre el plano $\mathbb{R}^2$ al objeto de acertar en una diana de radio $\sqrt[2]{1'39}$ y centro en el punto de coordenadas $(3, 2)$. Se supone que los lanzamientos tienen carácter aleatorio y que cada impacto (ξ_1, ξ_2) es un vector aleatorio, de suerte que las componentes ξ_1 y ξ_2 son independientes y además $\xi_1 \sim N(3, 1)$ y $\xi_2 \sim N(2, 1)$. Calcular la probabilidad

de que el dardo impacte dentro de la diana (Utilizar la distribución de una $\mathcal{X}^2$).

7. Sean, ξ_1 la v.a. que representa a los gastos por impagados en millones de euros de una determinada empresa, y ξ_2 la v.a. de los ingresos netos. Se sabe que la función de densidad del vector (ξ_1, ξ_2) es

$$f_{(\xi_1,\xi_2)}(x,y) = \begin{cases} 5(1+x)e^{20(2-y)} & \text{si } x \in (0,2) \text{ e } y > 2 \\ 0 & \text{en el resto} \end{cases}$$

   (a) Calcular la probabilidad de que los gastos por impagados sean menores a un millón de euros.

   (b) ¿Son independientes los gastos por impagados y los ingresos netos?

   (c) ¿Qué cantidad de ingresos netos se esperan obtener si los impagados ascendieron hasta 1 millón de euros?

8. Una empresa desea determinar la proporción de clientes dispuestos a adquirir uno de sus productos. Estima que dicha proporción es 0.45 ó 0.5. Decidir en base al Principio de Máxima Verosimilitud, una estimación de dicha proporción si después de realizar una muestra aleatoria simple de tamaño 15 entre sus clientes potenciales, 6 de ellos afirmaron estar dispuestos a la adquirir y los 9 restantes no estaban dispuestos a optar por dicho producto.

Universidad de Castilla-La Mancha
Estadística. Tercer Curso de Ciencias del Medio Ambiente
Resolución del examen convocatoria de 25-01-2005

1.

2. Del enunuciado del problema se desprende lo siguiente: si Inf denota estar infectado y A, B y C haber elegido a dichas poblaciones, entonces

$$P(Inf|\,A) = 3/10,\ P(Inf|\,B) = 6/10,\ P(Inf|\,C) = 1/10,$$

y

$$P(A) = P(B) = P(C) = \frac{1}{3}.$$

Se sabe que S el suceso consistente en la realización de 10 obsevaciones con resultado de 2 infecciones ha tenido lugar. La pregunta es saber cuál de las siguientes cantidades es mayor:

$$P(A|\,S),\ P(B|\,S),\ P(C|\,S).$$

Calculemos: se sabe que

$$P(A|\,S) = \frac{P(A\cap S)}{P(S)} = \frac{P(S|\,A)P(A)}{P(S|\,A)P(A) + P(S|\,B)P(B) + P(S|\,C)P(C)},$$

ahora bien, $P(S|\,A)$ es la probabilidad de que hayan tenido lugar 2 infecciones de 10 individuos sabiendo que la probabilidad de infección en A es $3/10$. Esta probabilidad no es otra cosa que el valor de la masa de un binomial de parámetros $p = 3/10$ y $n = 10$; análogamente para las otras probabilidades condicionadas. Por tanto

$$P(S|\,A) = \binom{10}{2}\left(\frac{3}{10}\right)^2\left(\frac{7}{10}\right)^8$$

y así

$$P(A|\,S) = \frac{\binom{10}{2}\left(\frac{3}{10}\right)^2\left(\frac{7}{10}\right)^8}{\binom{10}{2}\left(\frac{3}{10}\right)^2\left(\frac{7}{10}\right)^8 + \binom{10}{2}\left(\frac{6}{10}\right)^2\left(\frac{4}{10}\right)^8 + \binom{10}{2}\left(\frac{1}{10}\right)^2\left(\frac{9}{10}\right)^8}.$$

Puesto que el denominador que aparece en el cálculo de $P(A|\,S)$, $P(B|\,S)$ y $P(C|\,S)$ es el mismo, basta con comparar

$$\left(\frac{3}{10}\right)^2\left(\frac{7}{10}\right)^8,\left(\frac{6}{10}\right)^2\left(\frac{4}{10}\right)^8,\left(\frac{1}{10}\right)^2\left(\frac{9}{10}\right)^8.$$

El valor más grande es el correspondiente a $P(A|\, S)$ [2]. La población con más probabilidad de haber sido elegida es A.

3. Supongamos que X es la v.a. que proporciona el número de partículas en suspensión (expresada en miles de unidades por dm^3). Entonces $X \sim U(20,30)$ y por ende
$$E[X] = 25 \text{ y } V[X] = \frac{25}{3}.$$
Realizamos 600 inspiraciones y en cada una introducimos 0'5 dm^3 de aire, esto es, nos llevamos a los pulmones 300 dm^3. Sean $X_1, X_2, ..., X_{300}$ las v.a. que dan el número de partículas por cada una de las 300 veces que almacenamos 1 dm^3. Nos están pidiendo calcular $p = P(\sum_{i=1}^{300} X_i > 7510)$:
$$p = P(\frac{\frac{\sum_{i=1}^{300} X_i}{300} - 25}{\sqrt{\frac{1}{300}\frac{25}{3}}} > \frac{\frac{7510}{300} - 25}{\sqrt{\frac{1}{300}\frac{25}{3}}})$$
pero en virtud del TCL podemos aproximar p por $P(\phi > 0.2)$ con $\phi \sim N(0,1)$, con lo cual acudimos a las tablas de la normal para aproximar la probabilidad pedida como
$$p \approx P(\phi > 0.2) = 0.5793$$

4. Sea la variable aleatoria X que da el tiempo de vida después de ingerir el veneno: $X \sim N(\theta, 0.2)$. Sea la muestra $X_1, X_2, ..., X_{100}$. Si pretendemos la cosntrucción de un intervalo tomaremos como cantidad pivotal a
$$Q = \frac{\frac{\sum_{i=1}^{100} X_i}{100} - \theta}{\frac{0.2}{\sqrt{100}}}$$
Sabemos que Q está distribuido como una $N(0,1)$. Desarrollamos y llegamos a que el intervalo es
$$I = \left[\frac{\sum_{i=1}^{100} X_i}{100} - Z_{0.025}\left(\frac{0.2}{\sqrt{100}}\right), \frac{\sum_{i=1}^{100} X_i}{100} + Z_{0.025}\left(\frac{0.2}{\sqrt{100}}\right)\right]$$
Usandolas tablas y los datos del investigador llegamos a que $Z_{0.025} = 1.96$ y a que el intervalo obteniendo:
$$I = [0.260\,8, 0.339\,2]\,.$$

[2]Para cerciorarse de ello basta con esbozar la gráfica de la función $\psi(p) = p^2(1-p)^8$ y notar que el máximo se da para $p = 0.2$. Los valores en 0.6 y 0.1 están más lejos de este máximo que 0.3.

5. El EMV es el que maximiza la función de verosimilitud

$$\begin{aligned} l &= \log(p(1-p)^{x_1-1}p(1-p)^{x_2-1}...p(1-p)^{x_n-1}) \\ &= n\log p - n\log(1-p) + \left(\sum_{i=1}^{n} x_i\right)\log(1-p) \end{aligned}$$

Derivamos con respecto a p e igualamos a cero:

$$l' = \frac{n}{p} + \frac{n - \left(\sum_{i=1}^{n} x_i\right)}{1-p} = 0$$

La solución es $p = \frac{n}{\left(\sum_{i=1}^{n} x_i\right)} = \frac{1}{\overline{\xi}}$. Puesto que $l'' = \frac{-n}{p^2} + \frac{\left(n - \left(\sum_{i=1}^{n} x_i\right)\right)}{(1-p)^2} < 0$ (¿por qué?) el estimador máximo verosimil es $p^* = \frac{1}{\overline{\xi}}$. La estimación que correspondería a p, dado que la realización de $\overline{\xi}$ es $\overline{x} = 6.5$, sería $p \approx \frac{1}{6.5}$.

6. Si usamos el Teorema de Pitágoras nos daremos cuenta de que nos están pidiendo que evaluemos la probabilidad

$$p = P\left\{(\xi_1 - 3)^2 + (\xi_2 - 2)^2 \leq 1.39\right\},$$

y puesto que $(\xi_1 - 3)^2$ y $(\xi_2 - 2)^2$ son cuadrados de normales $(0,1)$, entonces $\Psi = (\xi_1 - 3)^2 + (\xi_2 - 2)^2$ se distribuye como una $\mathcal{X}_2^2$. Buscamos y obtenemos

$$p = P\left(\Psi \leq 1.39\right) = 0.5$$

7.

(a) Se puede calcular la marginal de ξ_1 y evaluar $P(\xi_1 \leq 1) = \int_{-\infty}^{1} f_{\xi_1}(x)dx$, o directamente:

$$\begin{aligned} P(\xi_1 &\leq 1, \xi_2 \text{ cualquiera}) = P(\xi_1 \leq 1, -\infty < \xi_2 < \infty) \\ &= \int_{-\infty}^{1}\left(\int_{-\infty}^{+\infty} f_{(\xi_1,\xi_2)}(x,y)\,dy\right)dx \\ &= \int_{0}^{1} 5(1+x)\left(\int_{2}^{+\infty} e^{40-20y}dy\right)dx \\ &= \frac{3}{8} \end{aligned}$$

(b) Las densidades marginales son

$$\int_{0}^{2} 5(1+x)\left(\int_{2}^{+\infty} e^{40-20y}dy\right)dx$$

$$
\begin{aligned}
f_{\xi_1}(x) &= \int_{\mathbb{R}} f_{(\xi_1,\xi_2)}(x,y)\,dy \\
&= \begin{cases} \int_2^{+\infty} 5(1+x)e^{40-20y}dy & \text{si } x\in(0,2) \\ 0 & \text{si } x\notin(0,2) \end{cases} \\
&= \begin{cases} \frac{1}{4}(1+x) & \text{si } x\in(0,2) \\ 0 & \text{si } x\notin(0,2) \end{cases}
\end{aligned}
$$

y

$$
\begin{aligned}
f_{\xi_2}(y) &= \int_{\mathbb{R}} f_{(\xi_1,\xi_2)}(x,y)\,dx \\
&= \begin{cases} \int_0^{2} 5(1+x)e^{40-20y}dx & \text{si } y\in(2,\infty) \\ 0 & \text{si } y\notin(2,\infty) \end{cases} \\
&= \begin{cases} 20e^{40-20y} & \text{si } y\in(2,\infty) \\ 0 & \text{si } y\notin(2,\infty) \end{cases}
\end{aligned}
$$

Puesto que $f_{(\xi_1,\xi_2)}(x,y) = f_{\xi_1}(x)\, f_{\xi_2}(y)$ las variables son independientes.

(c) Nos piden $E[\xi_2|\,\xi_1 = 1]$ pero dado que hay independencia

$$
\begin{aligned}
E[\xi_2|\,\xi_1=1] &= E[\xi_2] \\
&= \int_2^{+\infty} 20ye^{40-20y}dy \\
&= \frac{41}{20}
\end{aligned}
$$

8. θ es 0.45 ó 0.5. Habremos de evaluar $P^*_\xi(6)$ cuando $\xi \sim B(15,\theta)$:

	$P^*_\xi(6)$
$\theta = 0.45$	$\binom{15}{6}(0.45)^6(0.65)^9$
$\theta = 0.5$	$\binom{15}{6}(0.5)^6(0.5)^9$

El valor mayor se corresponde con $\theta = 0.45$, luego esta sería la estimación máximo verosimil.

Universidad de Castilla-La Mancha
Facultad de CC Medioambientales
Tercer curso de Ciencias Ambientales
Estadística. Examen ordinario 26-01-06
EXAMEN CON PARCIAL APROBADO

1. Una compañía aérea sabe por experiencia que el 12% de las reservas telefónicas de plazas no se llevan a efecto, de modo que reserva más plazas de las que dispone. Si en un vuelo hay 150 plazas, ¿cuántas reservas puede hacer la compañía para que la probabilidad de cubrir al menos 145 plazas sea del 99%? Si la compañía reserva 160 plazas, ¿cuál es la probabilidad de que, al menos un pasajero no tenga plaza disponible a la hora de embarcar? (Ayuda: utilizar el Teorema Central del Límite)

2. Queremos saber, con un nivel de confianza del 95%, entre qué valores, expresados en segundos, estará comprendida la reacción de una célula ante la presencia de un determinado agente externo. Para ello tomamos una muestra de 300 células y anotamos el número de segúndos que tarda en reaccionar cada una ellas. Calculada la media y la cuasivarianza de la muestra se obtuvieron los resultados 8 y $(12)^2$ respectivamente. Hallar el intervalo de confianza del 95% para la media si se supone que la duración de reacción sigue una ley normal.

3. Sea una población modelada por una normal $N(\mu, \sigma)$ donde se sabe que $\mu = 1.8$. Se realiza una m.a.s. de tamaño 10 resultando los datos

   $$1.8 \quad 1.78 \quad 1.77 \quad 1.8 \quad 1.78 \quad 1.8 \quad 1.82 \quad 1.81 \quad 1.8 \quad 1.79$$

   Encontrar el estimador de máxima verosimilitud para σ, y con los datos de arriba realizar una estimación de máxima verosimilitud de σ.

4. En un instituto de investigación se mide el número de horas que tarda en desaparecer una afección después de la aplicación de un medicamento. Se supone que la desviación típica es de 99 horas y se pretende encontrar un intervalo para la media poblacional μ. Determinar el tamaño de la muestra necesario para estimar μ con un error menor a 5 horas y a un nivel de confianza del 95%. Se supone que el tiempo que tarda en desaparecer la afección sigue una ley normal $N(\mu, 99)$.

Universidad de Castilla-La Mancha
Facultad de CC Medioambientales
Tercer curso de Ciencias Ambientales, Estadística. Examen ordinario
26-01-06
EXAMEN SIN PARCIAL APROBADO

1. Un laboratorio de química orgánica sintetiza tres tipos diferentes de moléculas A, B y C en porcentajes respectivos 40%, 40% y 20%. Las moléculas se clasifican en activas y no activas, de acuerdo a un criterio preestablecido, obervándose que son activas el 30% de las de tipo A, el 50% de las del tipo B y el 80% de las del tipo C.¿Qué porcentaje de moléculas son activas? ¿Cómo se distribuyen por tipos las moléculas activas? (esto es, decir cuáles son las frecuencias de A, B y C cuando éstas pertenecen a la población de las activas)

2. Se sabe que la v.a X que representa los coeficientes intelectuales de los alumnos de un colegio sigue una normal. Conocemos $P(X \geq 1.4) = 0.1056)$ y $P(X \geq 1) = 0.4013$. Determinar la esperanza μ y la desviación típica σ de la v.a. X.

3. Se sospecha que la proporción p de individuos de una población cierta característica es muy baja. Si se observan 4 individuos, ¿cuál es, en función de p, la probabilidad de no detectar dicha caraterística? Si suponemos que $p = 0.02$, ¿cuál es el número mínimo de individuos que se deben observar para detectar la característica con una probabilidad mayor o igual que 0.95?

4. La altura que alcanza una determinada planta viene dada mediante una v.a. ξ cuya densidad de probabilidad es

$$f_\xi(x) = \begin{cases} \frac{1}{4}e^{-\frac{1}{4}x} & \text{si } x \geq 0 \\ 0 & \text{en el resto} \end{cases}$$

Se toma una muestra aleatoria simple de tamaño 50, $\xi_1, ..., \xi_{50}$, de la v.a. ξ. Calcular, con ayuda del Teorema Central de Límite, la probabilidad de que la media muestral esté comprendida entre 3 y 5.

5. En la tabla adjunta se representan los valores del pH, de una solución en 10 determinaciones diferentes:

6.8	6.78	6.77	6.8	6.78	6.8	6.82	6.81	6.8	6.79

suponiendo normal la distribución de la población de todas las determinaciones del pH de esa solución, encontrar un intervalo de confianza al 95% para la varianza poblacional. ¿Con cuántos decimales exactos podemos dar una aproximación de la varianza?

6. Suponemos que el tiempo medio de vida de un determinado gusano está dado mediante la v.a. X cuya densidad es

$$f_X(x) = \begin{cases} \theta e^{-\theta x} & \text{si } x \geq 0 \\ 0 & \text{en el resto} \end{cases}$$

(donde $\theta > 0$). Obtener un estimador de máxima verosimilitud para θ

SOLUCION EXAMEN CON PARCIAL APROBADO

1.

$$P\left\{\sum_{i=1}^{n}\xi_i \geq 145\right\} = P\left\{\eta = \frac{\sum_{i=1}^{n}\xi_i - n\,(0.88)}{\sqrt{n\,(0.88)\,(0.12)}} > \frac{145 - n\,(0.88)}{\sqrt{n\,(0.88)\,(0.12)}}\right\} = 0.99.$$

Empleando el TCL tendremos la aporximación

$$P\left\{\sum_{i=1}^{n}\xi_i \geq 145\right\} \approx P\left\{N(0,1) > c\right\} = 0.99$$

con $c = \frac{145 - n(0.88)}{\sqrt{n(0.88)(0.12)}}$. Pero esta estimación implica (mirar tablas de la normal)

$$c = \frac{145 - n\,(0.88)}{\sqrt{n\,(0.88)\,(0.12)}} = -2.325$$

cuya solución aproximada es $n = 177$. Basta que el número de reservas sea mayor o igual que 177 para cubrir 145 de las plazas.
La segunda cuestión se resuelve de forma parecida:

$$\begin{aligned} P\left\{\sum_{i=1}^{160}\xi_i \geq 151\right\} &= P\left\{\eta \geq \frac{151 - (0.88)\,(160)}{\sqrt{0.1056\,(160)}} = 2.481\,5\right\} \\ &\approx 1 - P\left\{\eta < 2.481\,5\right\} = 1 - 0.9934 \\ &= 0.0066 \end{aligned}$$

2. $\alpha = 0.05$, $n = 300$, $\overline{x} = 8$ $s^2 = (12)^2$. El intervalo para la media poblacional, supuesta la desviación típica desconocida es

$$\begin{aligned} I &= \left[\overline{x} \pm \frac{s}{\sqrt[2]{n}}\mathrm{t}_{299,0.025}\right] \\ &= \left[8 \pm \frac{12}{\sqrt[2]{300}}\mathrm{t}_{299,0.025}\right] \end{aligned}$$

que de manera aproximada -usando una t de 200 grados de libertad resulta que el intervalo pedido es

$$I == \left[8 \pm \frac{12}{\sqrt[2]{300}}1.972\right] = [6.\,63,\ 9.36].$$

3. Si $\xi \sim N(\mu, \sigma)$, con $\theta = \sigma$ parámetro desconocido, $\xi_1, ..., \xi_n$ es una m.a.s. y $x_1, ..., x_n$ es una realización de la misma, la función de verosimilitud es

$$\begin{aligned} f_\theta(x_1, ..., x_n) &= \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x_1-\mu)^2}{2\sigma^2}} \cdots \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x_n-\mu)^2}{2\sigma^2}} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(\frac{-\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}\right). \end{aligned}$$

En lugar de maximizar esta función lo hacemos con $l = \log(f_\theta(x_1, ..., x_n))$ i.e. con

$$l_\theta(x_1, ..., x_n) = \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n - \frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}.$$

Para calcular el máximo de esta función de la variables $\sigma \in \mathbb{R}^+$ determinamos sus puntos críticos. Los puntos críticos los obtenemos resolviendo $\dfrac{\partial l_\theta}{\partial \sigma} = 0$, es decir

$$-n\frac{\sigma^2}{\sigma^3} + \frac{\sum_{i=1}^n (x_i-\mu)^2}{\sigma^3} = 0.$$

La solución, el EMV es,

$$\sigma^* = \sqrt[2]{\frac{\sum_{i=1}^n (x_i-\mu)^2}{n}}$$

Con los datos

$$x = [1.8, 1.78, 1.77, 1.8, 1.78, 1.8, 1.82, 1.81, 1.8, 1.79]$$

la estimación pedida es

$$\sigma = \sqrt[2]{\frac{\sum_{i=1}^{10} (x_i - 1.8)^2}{10}} = \sqrt[2]{\frac{0.002\,3}{10}} = 0.015$$

4. $X =$ número de horas que tarda $\sim N(\mu, 99)$. El intervalo de confianza es

$$I = \left[\overline{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt[2]{n}}\right]$$

de manera que el error máxima es

$$e = long(I) = 2Z_{0.025}\frac{99}{\sqrt[2]{n}} \leq 5.$$

Por tanto basta con tomar n tal que

$$2\,(1.96)\,\frac{99}{\sqrt[2]{n}} \leq 5$$

esto es, basta con $n \geq 6025$.

SOLUCIÓN DEL EXAMEN SIN PARCIAL APROBADO

1. Los datos son

$$P\left(Act|\,A\right)=0.3,\;\; P\left(Act|\,B\right)=0.5,\;\; P\left(Act|\,C\right)=0.8$$

El Teorema de la Probabilidad Total establece que

$$\begin{aligned} P(Act) &= P(Act\cap A)+P(Act\cap B)+P(Act\cap C) \\ &= P\left(Act|\,A\right)P(A)+P\left(Act|\,B\right)P(B)+P\left(Act|\,C\right)P\left(C\right) \\ &= (0.3)\,(0.4)+(0.5)\,(0.4)+(0.8)\,(0.2)=0.48 \end{aligned}$$

Las probabilidades o frecuencias de las moléculas se calculan como sigue:

$$\begin{aligned} P\left(A|\,Act\right) &= \frac{P(Act\cap A)}{P(Act)}=\frac{P\left(Act|\,A\right)P\left(A\right)}{P(Act)}=\frac{(0.3)\,(0.4)}{0.48}=0.25 \\ P\left(B|\,Act\right) &= \frac{P(Act\cap B)}{P(Act)}=\frac{P\left(Act|\,B\right)P\left(B\right)}{P(Act)}=\frac{(0.5)\,(0.4)}{0.48}=0.416 \\ P\left(C|\,Act\right) &= \frac{P(Act\cap C)}{P(Act)}=\frac{P\left(Act|\,C\right)P\left(C\right)}{P(Act)}=\frac{(0.8)\,(0.2)}{0.48}=0.33 \end{aligned}$$

2. Se sabe que la v.a $X\sim N\left(\mu,\sigma\right)$ y que los datos son $P\left(X\geq 1.4\right)=0.1056)$ y $P\left(X\geq 1\right)=0.4013$. Estandarizamos la variable y buscamos en las tablas:

$$\begin{aligned} P\left(X\geq 1.4)\right) &= P\left(\frac{X-\mu}{\sigma}\geq\frac{1.4-\mu}{\sigma}\right)=0.1056, \\ P\left(X\geq 1\right) &= P\left(\frac{X-\mu}{\sigma}\geq\frac{1-\mu}{\sigma}\right)=0.4013. \end{aligned}$$

Puesto que la v.a. $\frac{X-\mu}{\sigma}$ es necesariamente una $N(0,1)$, basándonos en las tablas de ésta tendremos que

$$\begin{aligned} \frac{1.4-\mu}{\sigma} &= 1.59 \\ \frac{1-\mu}{\sigma} &= 0.25 \end{aligned}$$

Esto da lugar a que $\sigma=0.29$ y $\mu=0.92$.

3. Sea X_i la v.a. definida como cero si el inidividuo i posee la característica y cero en toro caso. $X_i \sim B(1,p)$. Al observarse 4 consideramos una v.a. $Y = \sum_{i=1}^{4} X_i$ la cual se distribuye como una $B(4,p)$. La probabilidad de no detectar la característica es

$$p = P(Y=0) = \binom{4}{0} p^0 (1-p)^4 = (1-p)^4$$

Si $p = 0.02$ y definimos $Y = \sum_{i=1}^{4} X_i$, entonces $Y \sim B(n, 0.02)$ y la probabilidad establecida es

$$\begin{aligned} 0.95 &\leq P(Y \geq 1) = 1 - P(Y=0) \\ &= 1 - (1-(0.02))^n \end{aligned}$$

i.e.

$$(1-(0.02))^n \leq 0.05$$

y tomando logaritmos queda

$$n \geq \frac{\ln(0.05)}{\ln(0.98)} \simeq 149$$

4. Calculamos la esperanza y varianza: :

$$\begin{aligned} \mu &= \int_{-\infty}^{\infty} x f_\xi(x)\, dx = \int_0^{+\infty} \frac{x}{4} \exp\left(\frac{-x}{4}\right) dx = 4 \\ \alpha_2 &= \int_{-\infty}^{\infty} x^2 f_\xi(x)\, dx = \int_0^{+\infty} \frac{x^2}{4} \exp\left(\frac{-x}{4}\right) dx = 32 \\ \sigma^2 &= 32 - 4^2 = 16 \end{aligned}$$

La probabilidad que piden es

$$\begin{aligned} p &= P\left(3 \leq \overline{\xi} \leq 5\right) \\ &= P\left(-0.44 = \frac{3-4}{16/\sqrt{50}} \leq \overline{\xi} \leq \frac{5-4}{16/\sqrt{50}} = 0.44\right) \end{aligned}$$

que por el TCL puede aproximarse por

$$\begin{aligned} p &= P(-0.44 \leq N(0,1) \leq 0.44) \\ &= F(0.44) - F(-0.44) \\ &= 0.67 - (1 - F(0.44)) \\ &= 2\,(0.67) - 1 = 0.34 \end{aligned}$$

5. La varianza poblacional se estima por intervalo del siguiente modo:
$$I=\left[\frac{(n-1)\,S^2}{\mathcal{X}^2_{n-1,\frac{\alpha}{2}}},\frac{(n-1)\,S^2}{\mathcal{X}^2_{n-1,1-\frac{\alpha}{2}}}\right]$$
con un nivel de confianza $1-\alpha$. Para nuestro caso $\alpha=0.05$, $n=10$ y los datos de la muestra son los del vector
$$x=[6.8,6.78,6.77,6.8,6.78,6.8,6.82,6.81,6.8,6.79]\,.$$
Así pues
$$\begin{aligned}\overline{x} &= \frac{\sum_{j=1}^{10}x_j}{10}=6.\,795\\ s^2 &= \frac{\sum_{j=1}^{10}(x_j-\overline{x})^2}{9}=\frac{\sum_{j=1}^{10}(x_j-6.795)^2}{9}=2.\,277\,778\times10^{-4}\end{aligned}$$
Además
$$\mathcal{X}^2_{9,0.025}=19.023,\ \mathcal{X}^2_{n-1,0.975}=2.7$$
Sustituimos en el intervalo
$$\begin{aligned}I &= \left[\frac{9\,(2.\,277\,778\times10^{-4})}{19.023},\frac{9\,(2.\,277\,778\times10^{-4})}{2.7}\right]\\ &= [0.000107,0.000759]\end{aligned}$$
resultando que la precisión es de tres decimales.

6. La verosimilitud es
$$L=\begin{cases}\theta e^{-\theta x_1}\theta e^{-\theta x_2}...\theta e^{-\theta x_n} & \text{si todos los } x_i\geq 0\\ 0 & \text{en otro caso}\end{cases}$$
Por consiguiente el máximo se alcanza cuando todos los x_i son mayores o iguales que cero. Maximizamos la expresión
$$L=\theta^n e^{-\theta\sum_{i=1}^n x_i}$$
lo cual equivale a hacerlo con su logaritmo, i.e. con
$$l=\log L=n\log\theta-\theta\left(\sum_{i=1}^n x_i\right)$$
Al hacer $l'=0$ queda
$$n/\theta-\left(\sum_{i=1}^n x_i\right)=0$$

por lo que

$$\theta^* = \frac{n}{(\sum_{i=1}^{n} x_i)} = \frac{1}{\overline{x}}$$

es el EMV.