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

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

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.