Probabilidad y Estadística
Segundo del grado en Telecomunicaciones, UAM, 2014-2015

Breve solucionario del examen parcial, 17-12-2014

1. (2 puntos) Tenemos una muestra de tamaño 9 de una variable X que sigue una $\mathcal{N}(\mu, \sigma)$. La media muestral es $\overline{x} = 2.0$, y la cuasidesviación típica muestral, $s = 0.5$. Queremos contrastar la hipótesis $H_0 : \mu \leq 1.5$. Al nivel de significación $\alpha = 5\,\%$, ¿rechazaremos o aceptaremos la hipótesis? Calcula (aproximadamente) el p-valor.

Solución. Basta aplicar el contraste correspondiente: rechazaremos la hipótesis si

$$\overline{x} > \mu_0 + t_{\{n-1;\alpha\}} \frac{s}{\sqrt{n}}.$$

En este caso, con $n = 9$, $\alpha = 5\,\%$, $\mu_0 = 1.5$ y $s = 0.5$, el número de la derecha es

$$1.5 + 1.86\,\frac{0.5}{\sqrt{9}} = 1.5 + 1.86\,\frac{1}{6} = 1.5 + 0.31 = 1.81.$$

Como $\overline{x} = 2$, rechazamos la hipótesis.
Para el p-valor, buscamos el α para el que

$$t_{\{8;\alpha\}} = (\overline{x} - \mu_0)\,\frac{\sqrt{n}}{s} = (1 - 1.5)\,\frac{\sqrt{9}}{0.5} = 3.$$

Mirando la tabla de percentiles de la t con 8 grados de libertad, vemos que el p-valor está entre $1\,\%$ y $0.5\,\%$ (y más cerca del $1\,\%$, en realidad).

2. (2 puntos) En un concurso se utiliza un bombo en el que nos aseguran que hay 500 bolas: 50 blancas, 150 azules, 250 rojas y 50 negras. Como nos nos fiamos mucho, decidimos observar los resultados de 100 extracciones del bombo, en las que han salido 7 blancas, 33 azules, 40 rojas y 20 negras. (Tras cada extracción, la bola se devuelve al bombo). ¿Podemos rechazar, con nivel de significación $5\,\%$, la hipótesis de que la composición del bombo es la que nos garantizaban?

Solución. Aplicamos un text χ^2 de bondad de ajuste. Hay cuatro clases. Los datos observados están en el enunciado. Los esperados (en la muestra de 100) se obtienen aplicando las proporciones del modelo ($10\,\%$ de blancas, $30\,\%$ de azules, $50\,\%$ de rojas y $10\,\%$ de negras). formamos la tabla habitual:

	O	E	$(O-E)^2/E$
blancas	7	10	$3^2/10$
azules	33	30	$3^2/30$
rojas	40	50	$10^2/50$
negras	20	10	$10^2/10$

que nos da

$$B = \frac{9}{10} + \frac{9}{30} + \frac{100}{50} + \frac{100}{10} = \frac{12}{10} + 2 + 10 = 13.2.$$

El percentil con el que hay que comparar es $\chi^2_{\{3,5\,\%\}} = 7.81$. Se rechaza la hipótesis.

3. (2 puntos) El número de erratas por página en un libro sigue una Poisson. Queremos contrastar la hipótesis de que el parámetro de la Poisson es $\lambda = 5$. Tras revisar las 20 primeras páginas, hemos encontrado lo siguiente:

número de erratas	0	1	2	3	4	5	6	7	8	9	10
# de páginas	0	1	0	0	3	7	4	2	2	0	1

(por ejemplo, se han encontrado 7 páginas con exactamente 5 erratas). Fijado un nivel de significación $\alpha = 5\,\%$, ¿qué podemos concluir?

SOLUCIÓN. Aplicamos el contraste paramétrico para la Poisson: rechazamos la hipótesis $H_0 : \lambda = 5$ si

$$|\overline{x} - 5| > z_{\alpha/2} \sqrt{\frac{5}{20}} = 1.96 \cdot \frac{1}{2} = 0.98.$$

La media muestral es

$$\overline{x} = \frac{1 \cdot 1 + 3 \cdot 4 + 7 \cdot 5 + 4 \cdot 6 + 2 \cdot 7 + 2 \cdot 8 + 1 \cdot 10}{20} = \frac{112}{20} = 5.6.$$

Como 0.6 es menor que 0.98, aceptamos la hipótesis.

4. (2 puntos) Tenemos dos poblaciones de insectos. En la primera hay una proporción p_1 de insectos portadores de un cierto virus; p_2 es la correspondiente proporción en la segunda población. Estas proporciones p_1 y p_2 son, claro, desconocidas. Para obtener un intervalo de confianza para la diferencia $p_1 - p_2$, tomaremos una muestra de tamaño n en la población 1 y otra muestra (del mismo tamaño) en la población 2. Determina n para que la longitud de ese intervalo de confianza al $95\,\%$ sea, como mucho, de un $2\,\%$. (Nota: observa que se pide el valor de n. Estamos diseñando el procedimiento, así que las muestras todavía *no están disponibles*).

SOLUCIÓN. Según las tablas, el intervalo de confianza para la diferencia de proporciones $p_1 - p_2$ estaría centrado en $\overline{x}_1 - \overline{x_2}$ y su longitud sería

$$\text{longitud IC} = 2\, z_{\alpha/2} \sqrt{\tfrac{\overline{x}_1(1-\overline{x}_1)}{n_1} + \tfrac{\overline{x}_2(1-\overline{x}_2)}{n_2}} = \frac{2\, z_{\alpha/2}}{\sqrt{n}} \sqrt{\overline{x}_1(1-\overline{x}_1) + \overline{x}_2(1-\overline{x}_2)},$$

usando que en este caso $n_1 = n_2 = n$. Pero, ¡atención!, esta expresión solo tiene sentido si disponemos de la muestra (y por tanto de su tamaño y de las medias muestrales). No es nuestro caso. Aquí justamente es n lo que queremos determinar. Pero aunque desconozcamos las medias muestrales, podemos ponernos en el "peor caso", es decir, sustituir esas medias muestrales por $1/2$ (pues la función $x(1-x)$ tiene su máximo en $x = 1/2$ cuando $x \in [0,1]$). Haciendo eso tendríamos que

$$\text{longitud IC} \le \frac{2\, z_{\alpha/2}}{\sqrt{n}} \sqrt{\frac{1}{4} + \frac{1}{4}}\,.$$

Exigiendo ahora que esa longitud sea menor que el $2\,\%$ y tomando $\alpha = 5\,\%$, obtenemos

$$\frac{2\, z_{2.5\,\%}}{\sqrt{n}} \sqrt{\frac{1}{2}} \le \frac{2}{100} \quad \Longrightarrow \quad \sqrt{n} \ge \frac{196}{100}\,\frac{100}{2}\,\frac{1}{\sqrt{2}} = \frac{98}{\sqrt{2}},$$

lo que da finalmente que n debe ser mayor o igual que $2 \cdot 98^2$.

5. (2 puntos) La variable aleatoria X toma los valores 0, 1 y 2 con probabilidades θ, $1/2$ y $1/2 - \theta$, respectivamente. Aquí θ es un parámetro, $\theta \in [0, 1/2]$.

a) Dada una muestra $x_1, \ldots, x_n$ de la variable X, obtén un estimador de θ por el método de los momentos.

b) Obtén el estimador de θ por máxima verosimilitud.

Solución. a) Llamamos $\overline{x}$ a la media muestral:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Ahora basta calcular la esperanza de X:

$$\mathbf{E}(X) = 0\cdot\theta + 1\cdot\frac{1}{2} + 2\cdot\Big(\frac{1}{2}-\theta\Big) = \frac{3}{2} - 2\theta$$

(que es función de θ), igualar a $\overline{x}$ y despejar θ:

$$\frac{3}{2} - 2\theta = \overline{x} \quad\Longrightarrow\quad \widehat{\theta} = \frac{3-2\overline{x}}{4}.$$

b) Dada la muestra $(x_1,\dots,x_n)$, la función de verosimilitud es

$$V(\theta; x_1,\dots,x_n) = \theta^a\Big(\frac{1}{2}\Big)^b\Big(\frac{1}{2}-\theta\Big)^c,$$

donde a es el número de veces que aparece el 0 en la muestra, b el número de unos, y c el número de doses (de forma que $a+b+c=n$). Con el procedimiento habitual para localizar puntos críticos (tomar logaritmos, derivar con respecto a θ e igualar a 0) se llega a la ecuación

$$\frac{a}{\theta} = \frac{c}{1/2-\theta},$$

cuya solución es

$$\widehat{\theta} = \frac{1}{2}\,\frac{a}{a+c}.$$

Compruébese si la estimación es razonable en los casos en los que $a=0$ ó $c=0$ (que habría que haber tratado por separado en el análisis anterior).

Percentiles de la χ^2 con n grados de libertad ($n=1,\dots,12$), $\alpha = 5\,\%$:

n	1	2	3	4	5	6	7	8	9	10	11	12
$\chi^2_{\{n;5\,\%\}}$	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31	19.68	21.03

Algunos percentiles de la t de Student con n grados de libertad ($n=1,\dots,12$):

n	1	2	3	4	5	6	7	8	9	10	11	12
$t_{\{n;0.5\,\%\}}$	63.66	9.92	5.84	4.60	4.03	3.71	3.50	3.36	3.25	3.17	3.11	3.05
$t_{\{n;1.0\,\%\}}$	31.82	6.96	4.54	3.75	3.36	3.14	3.00	2.90	2.82	2.76	2.72	2.68
$t_{\{n;2.5\,\%\}}$	12.71	4.30	3.18	2.78	2.57	2.45	2.36	2.31	2.26	2.23	2.20	2.18
$t_{\{n;5\,\%\}}$	6.31	2.92	2.35	2.13	2.02	1.94	1.89	1.86	1.83	1.81	1.80	1.78

Algunos valores de percentiles de la normal estándar:

α	5 %	4.5 %	4.0 %	3.5 %	3.0 %	2.5 %	2.0 %	1.5 %	1.0 %	0.5 %
z_α	1.645	1.695	1.751	1.812	1.881	1.960	2.054	2.170	2.326	2.576

Variable Poisson. Una variable aleatoria X sigue una POISSON(λ), con $\lambda > 0$, si toma los valores 0, 1, 2,… con probabilidades

$$\mathbf{P}(X=k) = \lambda^k\,\frac{e^{-\lambda}}{k!} \qquad \text{para } k = 0,1,2,\dots$$