EJERCICIOS DE INFERENCIA
GRUPO M1

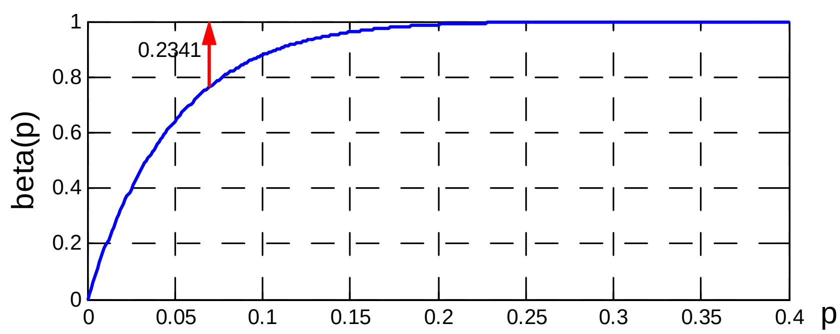
1. El Ministerio de Defensa está considerando un nuevo sistema de apoyo para el lanzamiento de misiles de corto alcance. El sistema existente tiene errores en el 7% de los lanzamientos y se desea comprobar si el nuevo sistema tiene una probabilidad de fallo menor. El ensayo va a consistir en realizar 20 lanzamientos y se concluirá que el nuevo sistema es mejor si no se produce ningún fallo. Llamando p a la probabilidad de fallo del sistema nuevo y aceptando independencia entre los resultados del lanzamiento, obtenga y represente gráficamente la probabilidad de *error de tipo II* del contraste

$$\begin{aligned} H_0 &: p = 0.07 \\ H_1 &: p < 0.07 \end{aligned}$$

Obtenga la probabilidad de error de tipo I. Interprete el resultado y valore si el método de decisión es adecuado. (Febrero 2002).

SOLUCIÓN: *La variable aleatoria* X = *"número de fallos en los 20 lanzamientos", tiene distribución binomial con* $n = 20$*, y* p igual a la probabilidad de fallo del sistema. Según el enunciado se rechaza H_0 si $X = 0$ y se acepta si $X \geq 1$. *La probabilidad de error tipo II, aceptar* H_0 *cuando* H_0 *es falso, depende de* p y se obtiene como

$$\beta(p) = P(X \geq 1|p) = 1 - (1-p)^{20}$$



La probabilidad de error tipo I, es la probabilidad de rechazar H_0 *cuando* H_0 *es cierta,*

$$\alpha = P(X = 0|p = 0.07) = (1 - 0.07)^{20} = 0.2341$$

El método de decisión no parece muy adecuado. Tiene una probabilidad muy alta de recharzar H_0

cuando es cierta. Por otra parte, si el sistema nuevo tien una probabilidad de fallo de $p = 0.05$*, la probabilidad de equivoquarse y aceptar* H_0 *es* $1 - (1 - 0.05)^{20} = 0.6415$

■

2. Estime por máxima verosimilitud α en el caso de una variable aleatoria con función de densidad

$$f(x) = \frac{2x}{\alpha}\exp\{-\frac{x^2}{\alpha}\}, \quad x > 0, \alpha > 0$$

Teniendo en cuenta que $\int_0^\infty \frac{2x^3}{\alpha}\exp\{-\frac{x^2}{\alpha}\}dx = \alpha$, demuestre que el estimador es centrado. (Febrero 2002)

SOLUCIÓN: *La función de verosimilitud es*

$$\begin{aligned} l(\alpha) &= \frac{2^n x_1 x_2 \cdots x_n}{\alpha^n}\exp(-\frac{1}{\alpha}\sum x_i^2); \\ L(\alpha) &= \log l(\alpha) \\ &= -n\log(\alpha) - \frac{1}{\alpha}\sum x_i^2 + k \end{aligned}$$

Derivando e igualando a cero,

$$\begin{aligned}
\frac{\partial L}{\partial \alpha} &= -\frac{n}{\hat{\alpha}} + \frac{1}{\hat{\alpha}^2}\sum x_i^2 = 0; \\
\widehat{\alpha} &= \frac{\sum x_i^2}{n}
\end{aligned}$$

Como $E[X^2] = \int_0^\infty \frac{2x^3}{\alpha}\exp\{-\frac{x^2}{\alpha}\}dx = \alpha$, *es directo comprobar que es un estimador centrado:*

$$E[\widehat{\alpha}] = E[\frac{\sum X_i^2}{n}] = \frac{\sum E[X_i^2]}{n} = \alpha.$$

■

3. La distribución gamma
$$f_Y(y) = \frac{1}{\alpha^\beta \Gamma(\beta)} y^{\beta-1} e^{-y/\alpha}, \quad y > 0;$$
es utilizada con mucha frecuencia como modelo para describir los niveles de precipitación, donde α y β son parámetros a estimar. Se dispone de las precipitaciones máximas en un día (medidas en pulgadas) para un total de 36 tormentas observadas durante los últimos veranos en un sistema montañosos peninsular, siendo la media $\overline{y} = 7.29$, y la varianza (sin corregir) $s_y^2 = 32.45$.

   Estima α y β por el método de los momentos, teniendo en cuenta que
$$E[Y] = \alpha\beta, \quad E[Y^2] = \alpha^2\beta(\beta+1).$$
(Septiembre 2002)

   SOLUCIÓN:

   $Var(Y) = E[Y^2] - E[Y]^2 = \alpha^2\beta(\beta+1) - \alpha^2\beta^2 = \alpha^2\beta$
$$\begin{cases} \hat{\alpha}\hat{\beta} = \bar{y}, & \hat{\alpha} = \dfrac{s_y^2}{\bar{y}} = 4.45 \\ \hat{\alpha}^2\hat{\beta} = s_y^2, & \hat{\beta} = \dfrac{\bar{y}^2}{s_y^2} = 1.63 \end{cases}$$
■

4. Se ha tomado una muestra al azar de 1011 vehículos que salían de Madrid por la carretera de La Coruña en una de las horas punta y se ha contado el número de ocupantes

Ocupantes	1	2	3	4	5	6
Frecuencia	678	227	56	28	8	14

   Utiliza el contraste de la χ^2 para decidir si los datos se ajustan a la distribución geométrica de probabilidad:
$$P(X = x) = (1-p)^{x-1}p, \quad x = 1, 2, ...$$
(Septiembre 2002)

   SOLUCIÓN: *La función de verosimilitud es*
$$\begin{aligned}
l(p) &= \prod_{i=1}^n P(X = x_i) = \prod_{i=1}^n (1-p)^{x_i - 1}p \\
&= (1-p)^{\sum_{i=1}^n (x_i - 1)} p^n \\
l(p) &= (1-p)^{s-n}p^n
\end{aligned}$$
donde $s = \sum_{i=1}^n x_i = 678 + 2\times 227 + 3\times 56 + 4\times 28 + 5\times 8 + 6\times 14 = 1536$ *y* $n = 1011$. *La función soporte es*
$$L(p) = (s-n)\log(1-p) + n\log p$$

Derivando e igualando a cero se obtiene $\hat{p} = \frac{n}{s} = 0.6582$

Ocupantes	**1**	**2**	**3**	**4**	**5**	**6 o más**	**SUMA**
$\mathbf{P(X = x)}$	.6582	.2249	.0769	.0263	.0090	.0047	**1**
Esperados	665.4	227.4	77.7	26.6	9.1	4.7	**1011**
Observados	678.0	227.0	56.0	28.0	8.0	14.0	**1011**
$\frac{(O_i - E_i)^2}{E_i}$	.2369	.0008	6.079	.076	.1289	18.27	**24.8**

Como $24.8 > \chi^2_{4;0.05} = 9.48$ *se rechaza la hipótesis de que los datos tienen distribución geométrica.*

■

5. Una teoría genética sostiene que los descendientes de cierta especie corresponden a uno de los cuatro grupos A, B, C y D con probabilidades:

A	**B**	**C**	**D**
$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$

En una muestra aleatoria de 288 especímenes se han observado los siguientes resultados

A	**B**	**C**	**D**
150	62	51	25

¿Es aceptable la teoría genética? (Utilice nivel de significación $\alpha = 0.01$) (Feb. 03)

SOLUCIÓN: *Multiplicando* $n = 288$ *por cada una de las probabilidades de la tabla obtenemos los valores esperados*

	A	**B**	**C**	**D**
O_i	150	62	51	25
E_i	162	54	54	18

$\sum_{i=1}^4 \frac{(O_i - E_i)^2}{E_i} = 4.96$. *No se ha estimado ningún parámetro con los datos, por lo tanto el valor límite de aceptación es* $\chi^2_{3;0.01} = 11.34$, *y como es mayor que el obtenido se acepta la teoría genética.*

■

6. El tiempo de vida en días de un componente de un sistema eléctrico sigue una distribución de Rayleigh con función de densidad:

$$f(x) = \frac{2x}{\mu} e^{-(x^2/\mu)}, \quad x \geq 0, \mu > 0.$$

Se han puesto a funcionar 10 componentes simultáneamente, el primero de los cuales falló el día 3, el segundo el día 10 y el tercero el día 15. Después de 20 días el resto de los componentes seguían funcionando. Con esta información, se pide estimar el parámetro μ. (Feb. 03)

SOLUCIÓN: *Los datos son* $x_1 = 3, x_2 = 10$, $x_3 = 15$ *y* $x_i > T = 20$, *para el resto. Teniendo en cuenta que*

$$P(X > T) = \int_T^\infty \frac{2x}{\mu} e^{-(x^2/\mu)} dx = \exp(-\frac{1}{\mu} T^2),$$

la función de verosimilitud es

$$\begin{aligned} l(\mu) &= f(x_1) f(x_2) f(x_3) P(X > T)^7 \\ &= \frac{2^3 x_1 x_2 x_3}{\mu^3} \exp[-\frac{1}{\mu}(x_1^2 + x_2^2 + x_3^2)] \exp(-\frac{7}{\mu} T^2) \end{aligned}$$

La función soporte

$$L(\mu) = -3\log\mu - \frac{1}{\mu}(x_1^2 + x_2^2 + x_3^2 + 7T^2)$$

Derivando con respecto a μ, igualando a cero y despejando

$$\hat{\mu} = \frac{x_1^2 + x_2^2 + x_3^2 + 7T^2}{3} = 1044.6$$

■

7. La prueba de signos se usa para contrastar hipótesis acerca de la mediana η de una distribución continua. Suponga que se dispone de $X_1, X_2, ..., X_n$ una muestra aleatoria simple con la misma función de densidad y se desea hacer el siguiente contraste respecto a su mediana η :

$$\begin{aligned} H_0 &: \eta = \eta_0 \\ H_1 &: \eta < \eta_0 \end{aligned}$$

El procedimiento de signos es sencillo de describir. Si H_0 es cierto, cualquier diferencia $D_i = X_i - \eta_0$ tiene la misma probabilidad de ser positiva o negativa. Se define como estadístico de contraste S^+ el número de diferencias positivas. (Si la variable X_i es continua la probabilidad de que la diferencia D_i sea igual a cero es nula y por tanto no se considera esta posibilidad). Indique la distribución de probabilidad de la variable aleatoria S^+ si H_0 es cierto. Se ha realizado un ensayo para medir el esfuerzo cortante necesario para separar una unión soldada. De las 20 pruebas realizadas 14 han dado lugar a un esfuerzo cortante inferior a 2000 psi. Contraste mediante el método anterior si la mediana del esfuerzo cortante es 2000 psi o menor con $\alpha = 0.05$, calculando el nivel crítico (p valor). (Feb. 03)

SOLUCIÓN: *Es fácil observar que S^+ tiene distribución binomial de parámetros $n = 20$ y si H_0 es cierto $p = \frac{1}{2}$. El contraste planteado es equivalente a*

H_0 : $p = 1/2$ (La mediana es $\eta = \eta_0$, la probabilidad de que $X_i - \eta_0$ sea positiva es $1/2$).
H_1 : $p < 1/2$ (La mediana es $\eta < \eta_0$, la prob. de que $X_i - \eta_0$ sea positiva es menor de $1/2$).

Utilizando la aproximación normal,

$$\begin{aligned} S^+ &\rightsquigarrow N\left(20p, \sqrt{20p(1-p)}\right) \\ \frac{S^+ - 20p}{\sqrt{20p(1-p)}} &\rightsquigarrow N(0,1) \end{aligned}$$

Sustituyendo la hipótesis nula y el valor observado

$$Z_0 = \frac{6-10}{\sqrt{5}} = -2.23$$

El contraste es unilateral, teniendo en cuenta que si $Z \rightsquigarrow N(0,1)$ entonces $P(Z < -1.64) = 0.05$, como $Z_0 < -1.64$, se rechaza la hipótesis nula y se concluye que la mediana es significativamente menor que 2000. El p-valor es $P(Z < -2.23) = 0.0129$.

■

8. En una línea de alta tensión de 500 km de longitud se producen una media de 80 averías al año. Una compañía eléctrica ha construido otra línea de características similares de 1400 kilómetros de longitud. En el primer año el número de averías registradas fue de 270. Con este dato se puede concluir que esta línea tiene un número medio de averías superior al esperado. Realiza el contraste con $\alpha = 0.05$, suponiendo que el número de averías anuales en un tramo de longitud fija es una variable aleatoria con distribución de Poisson. (Feb. 04)

SOLUCIÓN: *Teniendo en cuenta las propiedades de la distribución de Poisson, el número medio (esperado) de averías en la línea de 1400 km es* $\lambda_0 = 1400 \times \frac{80}{500} = 224$. *El número de averías X en una año tiene distribución de Poisson de parámetro* λ, *se desea realizar el contraste*

$$\begin{aligned} H_0 &: \lambda = 224 \\ H_1 &: \lambda > 224 \end{aligned}$$

Utilizando la aproximación normal, $X \to N(224, \sqrt{224})$,

$$Z = \frac{X - 244}{\sqrt{224}} \to N(0,1)$$

El contraste es unilateral, $z_{0.05} = 1.64$, *y el estadístico decontraste para* $x = 270$

$$z_0 = \frac{270 - 244}{\sqrt{224}} = 1.7372 > 1.64$$

por lo tanto se rechaza H_0, *el número de averías observado es significativamente mayor que el esperado.*

■

9. Una encuesta de opinión sobre la Constitución Europea en el Reino Unido indica que 220 de 500 entrevistados son favorables a ella. Si p es la proporción real de ciudadanos a favor de la Constitución, calcula el nivel crítico (p - valor) del contraste

$$\begin{aligned} H_0 &: p = 0.5 \\ H_1 &: p < 0.5 \end{aligned}$$

(Jun 04)

SOLUCIÓN: *Utilizando la aproximación*

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \rightsquigarrow N(0,1)$$

Estimamos la proporción de voto favorable $\hat{p} = \frac{220}{500} = 0,44$. *Sustituyendo este valor y la hipótesis nula se tiene*

$$z = \frac{0.44 - 0.5}{\sqrt{\frac{0.5 \times (1-0.5)}{500}}} = -2.68$$

Si $Z \rightsquigarrow N(0,1)$ el p-valor es $P(Z < -2.68) = 0.0037$. Se rechaza H_0 para los niveles de significación habituales 0.05 y 0.01.

■

10. El Ministerio de Hacienda ha tomado una muestra de ingresos mensuales (euros) de 10000 familias de una comunidad autónoma. La figura representa el histograma de los ingresos y el histograma de los datos transformados en logaritmos.

	Ingresos	log(Ingresos)
media	1706.8	7.32
desv. típica	914.5	0.502

Calcula los intervalos de confianza para los ingresos medios mensuales (en euros) por familia con los datos transformados y sin transforma. Indica cuál es el más adecuado y explica a qué se debe las diferencias. (Es importante que los intervalos de confianza se proporcionen en las mismas unidades (euros) para que pueda interpretarse) (Jun 04)

SOLUCIÓN: *Sea X la variable ingresos mensuales con media μ_X y $Y = \log X$ la variable transformada con media μ_Y. Utilizando el teorema central del límite se obtiene un intervalo de confianza para los datos sin transformar*

$$\begin{aligned} \mu_X &\in 1706.8 \pm 1.96 \frac{914.5}{\sqrt{10000}} \\ 1688.8 &\leq \mu_X \leq 1724.8 \end{aligned}$$

Con los datos transfomados el cálculo del intervalo de confianza es más adecuado pues el histograma se parece más a la campana de Gauss que el histograma de los datos originales que es claramente asimétrico.

$$\begin{aligned} \mu_Y &\in 7.32 \pm 1.96 \frac{0.502}{\sqrt{10000}} \\ 7.3102 &\leq \mu_Y \leq 7.3298 \end{aligned}$$

Calculando la exponencial en los tres miembros de la desigualdad se obtiene

$$1495.5 \leq e^{\mu_Y} \leq 1525.1$$

Obviamente $\mu_X \neq e^{\mu_Y}$ y por tanto los intervalos de confianza no tienen por qué coincidir.

NOTA.- *Se puede demostrar que $\mu'_X = e^{\mu_Y}$ es la media grométrica de X y se verifica que $\mu'_X < \mu_X$, lo que concuerda con los resultados de los intervalos obtenidos. Si Y tiene distribución normal con media μ_Y y desviación típica σ_Y , se cumple que $\mu_X = e^{\mu_Y + \sigma_Y^2/2}$ lo que se corresponde con los datos del ejercicio. $e^{\bar{y} + \hat{s}_Y^2/2} = 1696.0 \simeq \bar{x}$.*

■

11. El número medio de errores de impreta en el periódico "Cibeles" durante los últimos 10 años ha sido de 18 por ejemplar. Cada ejemplar tiene 30 páginas y tiene periodicidad diaria. El periódico ha ampliado el número de secciones lo que implica aumentar el número de páginas a 40, aprovechando el cambio se ha introducido un nuevo sistema informático de corrección de erratas. La primera semana de publicación los ejemplares han tenidos los siguientes números de errores: 16 (lunes), 20 (martes), 12 (miércoles), 10 (jueves) y 23 (viernes). Aceptando que el número de errores sigue una distribución de Poisson, contrastar con $\alpha = 0.05$ que se ha producido una reducción en el *número de errores medio.* Calcula el nivel crítico (*p-valor*) del contraste. (Feb 05)

SOLUCIÓN: *Si se mantiene el número medio de erratas por página, el número medio (esperado) de errores en el periódico de 40 páginas es $\lambda_0 = 18 \times \frac{40}{30} = 24$. Se trata de hacer el contraste*

$$\begin{aligned} H_0 &: \lambda = 24 \\ H_1 &: \lambda < 24 \end{aligned}$$

Utilizando la aproximación normal, la media de cinco días cumple $\bar{X} \to N(24, \sqrt{\frac{24}{5}})$,

$$Z = \frac{\bar{X} - 24}{\sqrt{\frac{24}{5}}} \to N(0,1)$$

El contraste es unilateral, $z_{0.05} = 1.64$, y el estadístico decontraste para $\bar{x} = \frac{16+20+12+10+23}{5} = 16.2$

$$z_0 = \frac{16.2 - 24}{\sqrt{\frac{24}{5}}} = -3.56 < -1.64$$

por lo tanto se rechaza H_0, el número de errores observado es significativamente menor que el existente antes del cambio. El p-valor es $P(Z < -3.56) = 0.0001855$.

■

12. Una variable aleatoria X tiene como función de densidad

$$f(x) = \frac{2}{\beta} x \exp\left(-\frac{x^2}{\beta}\right), \quad x > 0,\ \beta > 0.$$

Obtén el estimador máximo verosímil de β. Calcula el sesgo teniendo en cuenta que $E[X^2] = \beta$. (Feb. 05)

SOLUCIÓN: *(Ver solución del ejercicio 2).*

■

13. Una encuesta sobre la Constitución Europea en un país tuvo los siguientes resultados

A favor	810
En contra	520
Indecisos	250
Abstenciones	120

Llamando p al porcentaje de indecisos en el conjunto del país, contrasta la hipótesis

$$\begin{aligned} H_0 &: \ p = 10\% \\ H_1 &: \ p > 10\% \end{aligned}$$

Calcula el nivel crítico del contraste. (Jun. 05)

SOLUCIÓN: *Estimamos la proporción de indecisos frente al total*

$$\hat{p} = \frac{250}{1700} = 0,1471$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

Sustituyendo los valores observados y el valor que se desea contrastar (hipótesis nula), se tiene

$$z = \frac{0.1471 - 0.10}{\sqrt{\frac{0.1\times(1-0.9)}{1700}}} = 6.47$$

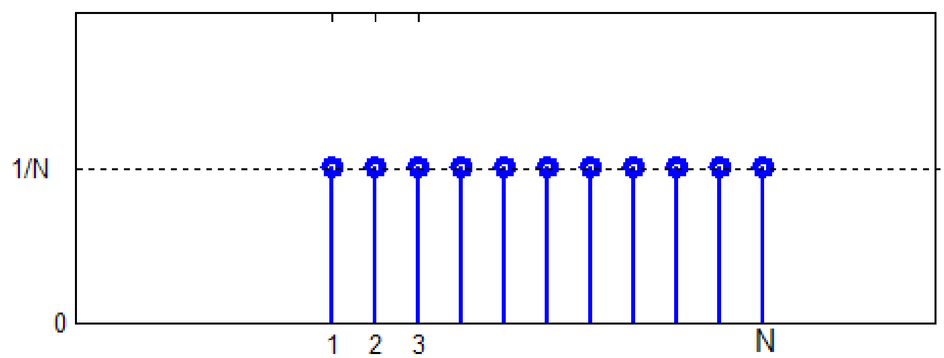
Valor muy superior a los límite habituales y se rechaza H_0. El p-valor es prácticamente 0, pues $P(Z > 6.47) \simeq 0$.

■

14. En un cesta hay bolas numeradas desde el 1 hasta N, siendo N desconocido. Se extraen dos bolas al azar (una detrás de otra con reposición) y se observan los números X e Y. Estudia si $\hat{N} = X + Y$ es un estimador centrado de N. (Jun 05)

SOLUCIÓN: *X resultado de la primera bola e Y el resultado de la segunda bola.Las dos variables aleatorias tienen la misma distribución de probabilidad. Como cualquier bola tiene la misma probabilidad de salir en cada extracción:*

$$P(X = x) = \frac{1}{N}, \quad x = 1, 2, 3, ..., N$$



Se puede comprobar que $E[X] = \frac{N+1}{2}$ *y por tanto*

$$E[\hat{N}] = E[X+Y] = E[X] + E[Y] = N + 1$$

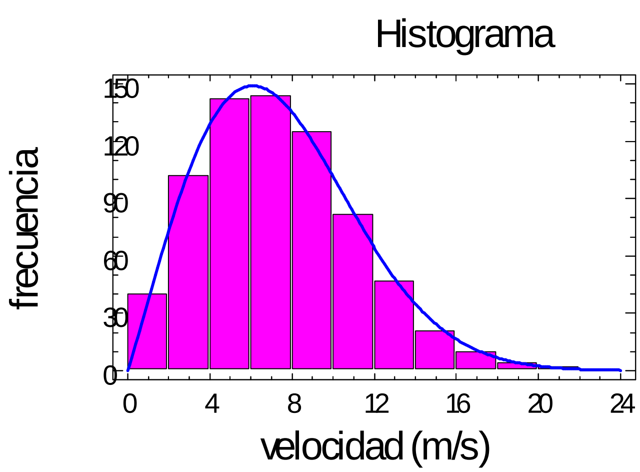
No es un estimador centrado.

■

15. Para los parques generadores de energía eléctrica es muy importante conocer la distribución de probabilidad de la velocidad del viento en un emplazamiento. Una distribución de probabilidad habitualmente empleada es la siguiente distribución de Weibull ,

$$F(x) = 1 - \exp\{-(\frac{x}{\alpha})^2\}, \qquad x \geq 0, \alpha > 0.$$

En la figura se muestra el histograma de 720 medidas de velocidad media del viento en un parque situado en Castilla-La Mancha, junto con la distribución de Weibull con $\hat{\alpha} = 8.4$, estimado por el método de los momentos. Cada una de las 720 medidas corresponde a la velocidad media del viento durante una hora. En todo el problema se supondrá que es una muestra de observaciones independientes.



Clases	Límite Inferior	Límite Superior	Frecuencia Observada
1	0	2	40
2	2	4	102
3	4	6	142
4	6	8	144
5	8	10	125
6	10	12	82
7	12	14	47
8	14	16	21
9	16	18	10
10	18	∞	7
Total			720

(a) Calcula la frecuencia esperada en cada una de las clases de la tabla anterior utilizando la distribución de Weibull estimada.
SOLUCIÓN: *La probabilidad de caer en la clase i es*

$$\begin{aligned} P(c_{i-1} &\leq X \leq c_i) = F(c_i) - F(c_{i-1}) \\ &= \exp\{-(\frac{c_{i-1}}{\alpha})^2\} - \exp\{-(\frac{c_i}{\alpha})^2\} \end{aligned}$$

Por ejemplo para la clase 5,

$$\begin{aligned} P(8 &\leq X \leq 10) = F(10) - F(8) \\ &= \exp\{-(\frac{8}{8.4})^2\} - \exp\{-(\frac{10}{8.4})^2\} \\ &= 0.161 \end{aligned}$$

Los valores se muestran en la tabla del apartado b.

(b) Realiza el contraste χ^2 de bondad de ajuste con nivel de significación 0.05
SOLUCIÓN:

Clases	Límite Inferior	Límite Superior	Probabilidad $P(c_{i-1} \leq X \leq c_i)$	Frecuencia Esperada	Frecuencia Observada	$\frac{(O_i - E_i)^2}{E_i}$
1	0	2	.055	39.7	40	.003
2	2	4	.147	106.4	102	.182
3	4	6	.196	141.4	142	.001
4	6	8	.196	141.6	144	.041
5	8	10	.161	116.2	125	.672
6	10	12	.112	81.0	82	.013
7	12	14	.067	48.8	47	.065
8	14	16	.036	25.6	21	.840
9	16	18	.016	11.8	10	.283
10	18	∞	.001	7.3	7	.012
Total			1	720	720	2.11

Como $2.11 < \chi^2_{8;0.05} = 15.50$ *se acepta la distribución de Weibull.*

(c) Una variable aleatoria X con la distribución anterior cumple

$$E[X] = \frac{\sqrt{\pi}}{2}\alpha, \quad E[X^2] = \alpha^2$$

Proporciona un intervalo con 95% de confianza para el parámetro α.(Utiliza la estimación por momentos, $\hat{\alpha} = 8.4$ y la distribución aproximada del estimador basada en el teorema central del límite)

SOLUCIÓN: *La varianza de X es* $Var[X] = E[X^2] - E[X]^2 = \alpha^2 - \frac{\pi}{4}\alpha^2 = \alpha^2(1-\frac{\pi}{4})$. *El estimador de α por el método de los momentos es* $\hat{\alpha} = \frac{2}{\sqrt{\pi}}\bar{X}$ *que es centrado y tiene varianza*

$$\begin{aligned} Var(\hat{\alpha}) &= \frac{4}{\pi}Var(X)/n \\ &= \frac{\alpha^2}{n}(\frac{4-\pi}{\pi}). \end{aligned}$$

Por el teorema central del límite

$$\hat{\alpha} \sim N\left(\alpha, \alpha\sqrt{\frac{4-\pi}{\pi n}}\right)$$

$$\frac{\hat{\alpha}-\alpha}{\alpha\sqrt{\frac{4-\pi}{\pi n}}} \sim N(0,1)$$

de donde se obtiene el intervalo de confianza

$$\begin{aligned} \alpha &\in \hat{\alpha} \pm 1.96\hat{\alpha}\sqrt{\frac{4-\pi}{\pi n}} \\ \alpha &\in 8.4 \pm 0.32 \end{aligned}$$

■

(d) Una velocidad del viento excesiva produce deterioro del rotor del aerogenerador. La empresa fabricante garantiza su funcionamiento durante 3 años (26280 horas).Si se toma como umbral de velocidades 24 m/s. ¿Cuál es la probabilidad de que se supere este umbral más de tres veces durante el periodo de garantía? (Proporciona un resultado numérico, utilizando la aproximación que consideres más conveniente. Se supone que las velocidas medias del viento en cada hora son variables aleatorias independientes).
(Jun. 05)

SOLUCIÓN: $P(X > 24) = 1 - F(24) = \exp\left\{-\left(\frac{24}{8.4}\right)^2\right\} = 2.85 \times 10^{-4}$. *Llamando Y la variable aleatoria número de horas con velocodades superiores a 24 km/h, se pide calcular* $P(Y > 3)$. *Utilizando la aproximación de Poisson con*

$$\lambda = 26280 \times 2.85 \times 10^{-4} = 7.48 \text{ fallos/año}$$

se tiene

$$\begin{aligned} P(Y &> 3) = 1 - P(Y \le 3) \\ &= 1 - e^{-7.48}(1 + 7.48 + \frac{7.48^2}{2} + \frac{7.48^3}{6}) \\ &= 1 - .0595 \\ &= 0.9405 \end{aligned}$$

■

16. Durante 3 años consecutivos (1095 días) el número de nacimientos en un hospital provincial han sido 5558, de los cuales 2981 fueron niños y 2577 niñas. En la tabla se muestra el número de días en los que se han producido X nacimientos

X	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Días	6	32	100	129	196	210	161	102	78	39	22	13	4	3

(a) ¿Se puede aceptar la hipótesis de que niño y niña son sucesos equiprobables? Indica el p-valor del contraste.

SOLUCIÓN:

$$\begin{aligned} H_0 &: p = 0.5 \\ H_1 &: p \neq 0.5 \end{aligned}$$

Utilizando la aproximación

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \rightsquigarrow N(0,1)$$

Estimamos la proporción de MUJERES $\hat{p} = \dfrac{2577}{5558} = 0,4637$. *Sustituyendo este valor y la hipótesis nula se tiene*

$$z_0 = \frac{0.4637 - 0.5}{\sqrt{\frac{0.5 \times (1-0.5)}{5558}}} = -5.41$$

El p-valor es $P(|Z| > 5.41) = 6.3 \times 10^{-8} \approx 0$. *Se rechaza* H_0 .

■

(b) Contrasta que "el número de nacimientos en un día" sigue la distribución de Poisson.

SOLUCIÓN: *El parámetro es el cociente entre el número total de nacimientos y el número de días*

$$\hat{\lambda} = \frac{5558}{1095} = 5.076 \text{ nacimientos/día}$$

Para obtener las probabilidades y los valores esperados de manera ágil es útil tener en cuenta las siguiente relaciones:

$$\begin{aligned} P(X &= k) = e^{-\lambda}\frac{\lambda^k}{k!} \\ &= e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} \times \frac{\lambda}{k} \\ &= P(X = k-1) \times \frac{\lambda}{k} \end{aligned}$$

$$\begin{aligned} E_k &= n \times P(X = k) \\ &= n \times P(X = k-1) \times \frac{\lambda}{k} \\ &= E_{k-1} \times \frac{\lambda}{k} \end{aligned}$$

Aplicando la recurrencia anterior, empezando por $E_0 = e^{-\lambda} \times n = 6.84$, *se tiene* $E_1 = E_0 \times \lambda, E_2 = E_1 \times \lambda/2, E_3 = E_2 \times \lambda/3, ...$*Las dos últimas clases se agrupan para que el valor esperado sea mayor que 5.*

$x \rightarrow$	0	1	2	3	4	5	6	7	8	9	10	11	12 o más
$P(X = x)$	.006	.032	.080	.136	.173	.175	.148	.108	.068	.038	.020	.009	.006
E_i	6.84	34.7	88.1	149.1	189.2	192.0	162.4	117.8	74.7	42.2	21.4	9.9	6.7
O_i	6	32	100	129	196	210	161	102	78	39	22	13	7
$\frac{(O_i - E_i)^2}{E_i}$	.10	.21	1.61	2.7	.25	1.68	.01	2.12	.14	.24	.02	.99	.01

$$\sum_{i=0}^{12} \frac{(O_i - E_i)^2}{E_i} = 10.08$$

Se acepta que los datos tienen distribución de Poisson pues el límite de la región de aceptación es $\chi^2_{11,0.05} = 19.68$

■

(c) Proporciona un intervalo de confianza para el número medio de nacimientos en una semana. Proporciona un intervalo de predicción con el 95% de probabilidad para el número de nacimientos en una semana.

SOLUCIÓN: *Si la tasa de nacimientos es de 5.076 al día, el número de nacimientos medio a la semana es* $\hat{\lambda}_S = 7 \times 5.076 = 35.5$. *Lógicamente, coincide con el cociente entre el número de nacimientos 5558 y el número de semanas* $N_S = 156.4$ *(1095/7*≈ 156.4*).Nos piden dar un intervalo de confianza para* λ_S*, teniendo en cuenta que*

$$\hat{\lambda}_S \to N(\lambda_S, \sqrt{\frac{\lambda_S}{N_S}}$$

$$\begin{aligned} \lambda_S &\in \hat{\lambda}_S \pm 1.96\sqrt{\frac{\hat{\lambda}_S}{N_S}} \\ \lambda_S &\in 35.5 \pm 0.94 \end{aligned}$$

La segunda parte del apartado nos piden un intervalo de predicción para Y = *"número de nacimientos en una semana".* Y *es una variable aleatoria con distribución de Poisson, aproximando por la normal*

$$Y \to N(\lambda_S, \sqrt{\lambda_S})$$

Con probabilidad del .95 el número de nacimientos está en el intervalo

$$Y \in \lambda_S \pm 1.96\sqrt{\lambda_S}$$

Sustituyendo la estimación $\hat{\lambda}_S$ *(no se tiene en cuenta el error en la estiamción de* λ_S*) :*

$$Y \in 35.5 \pm 11.7$$

(d) Aceptando que el proceso de nacimientos es de Poisson, ¿cuál es el tiempo medio transcurrido entre dos nacimientos?, ¿cuál es la probabilidad de que pasen más de 4 horas entre un nacimiento y otro?

(Feb 06)

SOLUCIÓN: *Sea* T *el tiempo transcurrido entre un nacimiento y otro. Si el proceso de nacimientos es de Poisson,* T *tiene distribución exponencial con media*

$$\begin{aligned} E[T] &= \frac{1}{\lambda} = \frac{1}{5.076}\text{días/nacimiento} \\ &= 4.72 \text{ horas/nacimiento} \end{aligned}$$

La función de densidad de T *es* $f_T(t) = \frac{1}{4.72}\, e^{-\frac{t}{4.72}}, \quad t > 0$ *y*

$$P(T > 4 \text{ horas}) = e^{-\frac{4}{4.72}} = 0.4291$$

■

17. La Unión Europea ha realizado una encuesta para relanzar la Constitución Europea. Ha preguntado a 25000 personas de los distintos países de la Unión, el resultado ha sido

	Total
A favor	10350
En contra	8425
Abstención	4250
No sabe o No contesta	1975

La aprobación se realiza si los votos a favor superan a los votos en contra, es decir no se tienen encuenta las abstenciones o los votos nulos. Suponiendo que los que se encuentran en las categorías "Abstención" y "No sabe o No contesta" no van a votar, se puede concluir que la proporción del voto favorable es significativamente mayor del 50%. Realiza un contraste con nivel de significación igual a 0.05 . (Jun. 06)

SOLUCIÓN: *Estimamos la proporción de voto favorable frente a votos válidos* $(n = 18775)$ *como:*

$$\hat{p} = \frac{10350}{18775} = 0,551$$

Se desea contrastar

$$\begin{aligned} H_0 &: p \leq 0,5 \\ H_1 &: p > 0,5 \end{aligned}$$

Teniendo en cuenta que

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

tiene distribución normal de media 0 y desviación 1. Sustituyendo los valores observados y el valor que se desea contrastar (hipótesis nula), se tiene

$$z = \frac{0.551 - 0.5}{\sqrt{\frac{0.5 \times (1-0.5)}{18775}}} = 14.04$$

Valor muy superior a límite de al región de aceptación de H_0,que es 1.64. Por tanto se rechaza H_0 y se puede afirmar con un 95% de confianza que el voto favorable es significativamente mayor que el 50%.

■

18. Un profesional de golf está midiendo la precisón de un nuevo juego de palos. Ha lanzado diez bolas con uno de ellos y la distancia (en metros) a la bandera conseguida ha sido la siguiente

0,7 1,2 1,9 2,2 4,2 6,6 6,7 8,1 12,7 20,6

Utilizando estos datos estima por máxima verosimilitud el parámetro α de la siguiente función de densidad

$$f(x) = \frac{2x}{\alpha^2} \exp(-\frac{x^2}{\alpha^2}), \quad \alpha > 0, x \geq 0$$

(Jun 06)

SOLUCIÓN:

$$\begin{aligned} l(\alpha) &= 2^{20} x_1 x_2 \cdots x_n \alpha^{-20} \exp(-\frac{1}{\alpha^2} \sum x_i^2); \\ L(\alpha) &= -20 \log(\alpha) - \frac{1}{\alpha^2} \sum x_i^2 + k; \\ \frac{\partial L}{\partial \alpha} &= -\frac{20}{\alpha} + \frac{2}{\alpha^3} \sum x_i^2 = 0; \end{aligned}$$

$$\widehat{\alpha} = \sqrt{\frac{\sum x_i^2}{10}} = 8,76$$

■

19. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria simple de una variable aleatoria con distribución de Poisson de media λ, ¿es el estimador

$$T = \frac{X_1 + 2X_2 + 2X_3 + \cdots + 2X_{n-1} + X_n}{2(n-1)}$$

un estimador insesgado de λ? Calcula el sesgo y la varianza de T. Propón un estimador centrado y con menor varianza.(Sep. 06)

SOLUCIÓN:

$$\begin{aligned} E[T] &= \frac{E[X_1] + 2E[X_2] + 2E[X_3] + \cdots + 2E[X_{n-1}] + E[X_n]}{2(n-1)} \\ &= \frac{\lambda + 2\lambda + 2\lambda + \cdots + 2\lambda + \lambda}{2(n-1)} = \lambda \end{aligned}$$

Es un estimador centrado y su sesgo es cero. La varianza se obtiene de la siguiente forma

$$\begin{aligned} Var(T) &= \frac{Var(X_1) + 4Var(X_2) + 4Var(X_3) + \cdots + 4Var(X_{n-1}) + Var(X_n)}{4(n-1)^2} \\ &= \frac{\lambda + 4\lambda + 4\lambda + \cdots + 4\lambda + \lambda}{4(n-1)^2} \\ &= \frac{4n-6}{4(n-1)^2}\lambda \\ &= \frac{2n-3}{2(n-1)^2}\lambda \end{aligned}$$

Un estimador centrado y con menor varianza es la media aritmética $\bar{X} = \sum X_i/n$ *que tiene varianza* λ/n *que es menor como se demuestra a continuación:*

$$\frac{2n-3}{2(n-1)^2} = \frac{2(n-1)-1}{2(n-1)^2} = \frac{1}{n-1} - \frac{1}{2(n-1)^2} > \frac{1}{n-1} > \frac{1}{n}$$

■

20. Durante el año 1982, los 25670 nacidos en el País Vasco se distribuyeron en 13333 niños y 12337 niñas ¿Es aceptable la hipótesis de que la probabilidad de niño es la misma que la de niña ? (Sept 06)

SOLUCIÓN: *Estimamos la proporción de voto favorable frente a votos válidos* $(n = 18775)$ *como:*

$$\hat{p} = \frac{12337}{25670} = 0,4806$$

Se desea contrastar

$$\begin{aligned} H_0 &: p = 0,5 \\ H_1 &: p \neq 0,5 \end{aligned}$$

Teniendo en cuenta que

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

tiene distribución normal de media 0 y desviación 1. Sustituyendo los valores observados y el valor que se desea contrastar (hipótesis nula), se tiene

$$z = \frac{0.4806 - 0.5}{\sqrt{\frac{0.5\times(1-0.5)}{25670}}} = -6.21$$

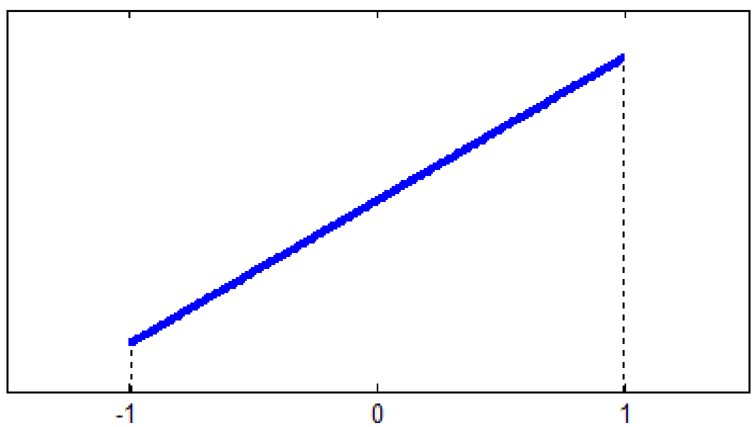
El contraste es bilateral, $P(|Z| < 1.96) = 0.95$*; como* $|z| > 1.96$ *se rechazar* H_0.

21. El coseno del ángulo con el que se emiten ciertas partículas de un proceso radiactivo es una variable aleatoria con densidad de probabilidad

$$f_X(x) = \frac{1+\theta x}{2}, \qquad -1 \leq x \leq 1.$$

Estima θ por el método de los momentos. (Feb. 07)

SOLUCIÓN:



$$E[X] = \int_{-1}^{1} x(\frac{1+\theta x}{2})dx = \frac{\theta}{3}$$

$$\frac{\hat{\theta}}{3} = \bar{X} \Rightarrow \hat{\theta} = 3\bar{X}$$

■

22. En los último 27 años, el número medio de aves protegidas muertas en una línea de álta tensión que cruza un parque natural ha sido de 7 al mes. El departamento de Medio Ambiente de la comunidad autonoma responsable ha colocado sistemas espantapájaros a lo largo de la línea y en el año de funcionamiento se han contabilizado un total 48 aves muertas. Utilizando la distribución de Poisson, propón un método para contrastar si el sistema instalado ha sido efectivo y realiza el contraste. (Utiliza nivel de significación 0.05) (Jun 07)

SOLUCIÓN: *El número de aves muertas* X *en un mes tiene distribución de Poisson de parámetro* λ*, se desea realizar el contraste*

$$\begin{aligned} H_0 &: \lambda = 7 \\ H_1 &: \lambda < 7 \end{aligned}$$

Utilizando la aproximación normal, $\bar{X} \to N(7, \sqrt{\frac{7}{12}})$,

$$Z = \frac{\bar{X} - 7}{\sqrt{\frac{7}{12}}} \to N(0,1)$$

El contraste es unilateral, $z_{0.05} = 1.64$*, y la media es* $\bar{x} = \frac{48}{12} = 4$,

$$z_0 = \frac{4-7}{\sqrt{\frac{7}{12}}} = -3.92 < -1.64$$

por lo tanto se rechaza H_0*, el número de aves muertas se ha reducido significativamente.*

■

23. En las últimas 100 semanas una compañía eléctrica ha tenido 992 interrupciones en sus líneas de distribución.. En la tabla siguiente se muestra la distribución de fallos

Fallos	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Nº Semanas	2	0	5	4	10	11	17	11	10	11	8	5	3	1	1	1

¿Se puede aceptar la hipótesis de que el número de fallos sigue la distribución de Poisson ($\alpha = 0.05$)? (Sep. 07)

SOLUCIÓN: *Aplicamos el método explicado en el ejercicio 16 (b) para obtener los valores esperados,* $E_k = E_{k-1}\dfrac{\lambda}{k}$*, utilizando* $\hat{\lambda} = 992/100 = 9.92$.

$$\begin{aligned} E_0 &= 100e^{-9.92} = 4.9181 \times 10^{-3} \\ E_1 &= 4.9181 \times 10^{-3} \times 9.92 = 4.8788 \times 10^{-2} \\ E_2 &= 4.8788 \times 10^{-2} \times 9.92/2 = 0.24199 \\ & \ldots \end{aligned}$$

Agrupamos las clases que tienen valores esperados menor de 5. Los resultados se muestran en la tabla:

Fallos	5 o menos	6	7	8	9	10	11	12	13	14	15 o más	SUMA
O_i	7	4	10	11	17	11	10	11	8	5	3	100
E_i	6.2	6.5	9.2	11.4	12.6	12.5	11.3	9.6	7.1	5.0	7.7	100
$\frac{(O_i-E_i)^2}{E_i}$	0.1	.97	.07	.02	1.53	.18	.15	.22	.11	.0	.38	3.71

Como $\chi^2_{9,0.05} = 16.92 > 3.71$*, se acepta la hipótesis de que los datos tienen distribución de Poisson.*

■

24. En una urna hay N bolas numeradas de 1 hasta N, siendo N desconocido. Se obtiene una muestra con reemplazamiento de 10 bolas resultando

$$703, 785, 363, 454, 1050, 986, 798, 1338, 693, 646$$

Estimar N por el método de los momentos. (Sep. 07)

SOLUCIÓN: *Ver el ejercicio 14 donde se muestra que* $E[X] = \dfrac{N+1}{2}$*, por tanto*

$$\begin{aligned} \frac{\hat{N}+1}{2} &= \bar{X} \quad \Rightarrow \quad \hat{N} = 2\bar{X} - 1 \\ \hat{N} &= 1562 \end{aligned}$$

■