PROBABILIDAD y ESTADÍSTICA (61.06 - 81.09)

Evaluación Integradora — Segundo cuatrimestre – 2017

Duración: 4 horas. — 15/II/18 – 9:00 hs.

---

1. Una cueva será iluminada por dos lámparas L_1 y L_2 cuyas duraciones (en horas), T_1 y T_2, son independientes y tienen distribuciones Weibull de parámetros $(2,7)$ y $(2,9)$, cuyas densidades son

$$f_{T_1}(t_1) = \frac{2}{7}\left(\frac{t_1}{7}\right)e^{-(t_1/7)^2}\mathbf{1}\{t_1 > 0\}, \quad f_{T_2}(t_2) = \frac{2}{9}\left(\frac{t_2}{9}\right)e^{-(t_2/9)^2}\mathbf{1}\{t_2 > 0\},$$

respectivamente. Las lámparas se encenderán simultáneamente. Calcular la probabilidad de que la lámpara L_1 sea la primera en apagarse.

R. [*Referencia*: **Ejercicio 2.17 (a)** y **Ejercicio 4.14 (b)**] Queremos calcular $\mathbf{P}(T_1 < T_2)$, la probabilidad de que T_1 sea menor que T_2. Para calcular esta probabilidad, podemos integrar la densidad conjunta $f_{T_1,T_2}(t_1,t_2)$ sobre la región $\mathcal{B} = \left\{(t_1,t_2) \in \mathbb{R}^2 : 0 < t_1 < t_2\right\}$, teniendo en cuenta que las variables T_1 y T_2 son independientes y que $\int_0^t \frac{k}{\alpha}\left(\frac{s}{\alpha}\right)^{k-1} e^{-(s/\alpha)^k} ds = 1 - e^{-(t/\alpha)^k}$, tenemos

$$\begin{aligned}
\mathbf{P}(T_1 < T_2) &= \iint_{\mathcal{B}} f_{T_1,T_2}(t_1,t_2)dt_1dt_2 = \iint_{\mathcal{B}} f_{T_1}(t_1)f_{T_2}(t_2)dt_2 \\
&= \int_0^\infty f_{T_2}(t_2)\left(\int_0^{t_2} f_{T_1}(t_1)dt_1\right)dt_2 = \int_0^\infty \frac{2}{9}\left(\frac{t_2}{9}\right)e^{-(t_2/9)^2}\left(1 - e^{-(t_2/7)^2}\right)dt_2 \\
&= 1 - \int_0^\infty \frac{2}{9}\left(\frac{t_2}{9}\right)e^{-(t_2/9)^2}e^{-(t_2/7)^2}dt_2 = 1 - \int_0^\infty \frac{2}{9}\left(\frac{t_2}{9}\right)e^{-(\sqrt{130}\,t_2/63)^2}dt_2 \\
&= 1 - \frac{49}{130}\int_0^\infty \frac{2\sqrt{130}}{63}\left(\frac{\sqrt{130}t_2}{63}\right)e^{-(\sqrt{130}\,t_2/63)^2}dt_2 = 1 - \frac{49}{130} = \frac{81}{130} \approx 0.623.
\end{aligned}$$

□

Otro modo. Notar que para $T \sim \text{Wei}(k,\alpha)$ se tiene que $T^k \sim \mathcal{E}\left((1/\alpha)^k\right)$, tener en cuenta que la función $t \mapsto t^2$ es creciente sobre la semirrecta positiva, y usar el resultado del **Ejercicio 4.14 (b)**: $\mathbf{P}(T_1 < T_2) = \mathbf{P}\left(T_1^2 < T_2^2\right) = \frac{(1/7)^2}{(1/7)^2+(1/9)^2} = \frac{81}{130}$. □

---

2. El panadero llega a la panadería a las 3:00+T, donde T es una variable aleatoria con distribución exponencial de media 5 minutos. La producción del día (en kg) es una variable aleatoria con distribución normal: de media 100 y desvío 10 cuando llega antes de las 3:15, y de media 80 y desvío 10 en caso contrario. La producción del día fue 90 kg, calcular la probabilidad de que el panadero haya llegado a la panadería antes de las 3:15.

R. [*Referencia*: **Ejercicio 5.9**] Sea W la producción del día (en kg). La distribución de W depende de los valores de T: $W|T < 15 \sim \mathcal{N}(100, 10^2)$ y $W|T \geq 15 \sim \mathcal{N}(80, 10^2)$, donde $T \sim \mathcal{E}(1/5)$. Queremos calcular $\mathbf{P}(T < 15|W = 90)$.

Aplicando la *regla de Bayes para mezclas de variables aleatorias continuas* tenemos

$$\mathbf{P}(T < 15|W = 90) = \frac{f_{W|T<15}(90)\mathbf{P}(T < 15)}{f_W(90)},$$

siendo

$$
\begin{aligned}
f_W(90) &= f_{W|T<15}(90)\cdot \mathbf{P}(T<15) + f_{W|T\geq 15}(90)\cdot \mathbf{P}(T\geq 15)\\
&= \frac{1}{\sqrt{2\pi}\,10}e^{-\frac{1}{2}\left(\frac{90-100}{10}\right)^2}\cdot\left(1-e^{-\frac{15}{5}}\right) + \frac{1}{\sqrt{2\pi}\,10}e^{-\frac{1}{2}\left(\frac{90-80}{10}\right)^2}\cdot e^{-\frac{15}{5}}\\
&= \frac{1}{\sqrt{2\pi}\,10}e^{-\frac{1}{2}}\cdot\left(1-e^{-3}\right) + \frac{1}{\sqrt{2\pi}\,10}e^{-\frac{1}{2}}\cdot e^{-3} = \frac{1}{\sqrt{2\pi}\,10}e^{-\frac{1}{2}}.
\end{aligned}
$$

Por lo tanto,

$$\mathbf{P}(T<15|W=90) = 1-e^{-3} \approx 0.9502.$$

$\square$

---

3. Un dado equilibrado tiene sus caras pintadas de la siguiente manera: una cara blanca, dos rojas y tres negras. Se arroja el dado hasta obtener la tercera cara negra. Calcular la varianza de la cantidad de caras blancas observadas.

R. [*Referencia*: **Ejercicio 6.20**] Sea N_b la cantidad de caras blancas observadas hasta que se observa la tercera cara negra. Queremos calcular $\mathbf{var}(N_b)$.

Sea, por otra parte, N la cantidad de veces que se arroja el dado hasta obtener la tercera cara negra. $N \sim \text{Pascal}(3, 1/2)$, y la distribución de N_b depende de los eventos $N=n$, $n\geq 3$. Teniendo en cuenta que sabiendo que no salió la cara negra, la probabilidad de que salga la cara blanca es $\frac{1/6}{1/6+2/6} = 1/3$, tenemos

$$N_b|N=n \sim \mathcal{B}(n-3, 1/3).$$

En consecuencia, $\mathbf{E}[N_b|N] = \frac{1}{3}\cdot(N-3)$ y $\mathbf{var}(N_b|N) = \frac{1}{3}\cdot\frac{2}{3}\cdot(N-3) = \frac{2}{9}\cdot(N-3)$. Teniendo en cuenta las propiedades de la esperanza y de la varianza, el *Teorema de Pitágoras* da

$$\mathbf{var}(N_b) = \mathbf{var}(\mathbf{E}[N_b|N]) + \mathbf{E}[\mathbf{var}(N_b|N)] = \frac{1}{9}\cdot\mathbf{var}(N) + \frac{2}{9}\cdot(\mathbf{E}[N]-3)$$

y como $\mathbf{E}[N] = 3\cdot\frac{1}{1/2} = 6$ y $\mathbf{var}(N) = 3\cdot\frac{1-(1/2)}{(1/2)^2} = 6$, resulta $\mathbf{var}(N_b) = \frac{12}{9} = \frac{4}{3}$. $\square$

---

4. A partir de las 0:00 arriban mensajes a una dirección de correo electrónico de acuerdo con un proceso Poisson de intensidad 4 por hora. Independientemente de los arribos, cada uno de los mensajes es spam con probabilidad 0.2. Calcular la media del tiempo que transcurrirá desde las 0:00 hasta que por primera vez arriben dos mensajes spam consecutivos.

R. [*Referencia*: **Ejercicio 5.13** y **Ejercicio 7.1**] Sea T el tiempo (en minutos) que transcurrirá desde las 0:00 hasta que por primera vez arriben dos mensajes spam consecutivos. Queremos calcular $\mathbf{E}[T]$.

Indicando, para cada $i=1,2$, mediante S_i al evento de que el i-ésimo mensaje que arriba a la casilla de correos después de las 0:00 es spam, se tiene

$$T|S_1^c \sim T_1+T, \quad T|S_1S_2^c \sim T_1+T_2+T, \quad T|S_1S_2 \sim T_1+T_2,$$

donde T_1 y T_2 son variables aleatorias independientes con distribución exponencial de media 15. Aplicando la *fórmula de probabilidades totales* tenemos

$$
\begin{aligned}
\mathbf{E}[T] &= \mathbf{E}[T|S_1^c]\mathbf{P}(S_1^c) + \mathbf{E}[T|S_1S_2^c]\mathbf{P}(S_1S_2^c) + \mathbf{E}[T|S_1S_2]\mathbf{P}(S_1S_2)\\
&= (15+\mathbf{E}[T])(0.8) + (30+\mathbf{E}[T])(0.2)(0.8) + 30(0.2)^2\\
&= 12 + (0.8)\mathbf{E}[T] + 4.8 + (0.16)\mathbf{E}[T] + 1.2 = 18 + (0.96)\mathbf{E}[T],
\end{aligned}
$$

y por lo tanto, $\mathbf{E}[T] = \frac{18}{0.04} = 450$. $\square$

Otro modo. Consideramos $T_1, T_2, \dots$ la sucesión de tiempos entre arribos de mensajes a la dirección de correo electrónico desde las 0:00 y N la cantidad de mensajes hasta que por primera vez arriban dos mensajes spam consecutivos. Con esa nomenclatura el tiempo, T, que transcurre desde las 0:00 hasta que por primera vez arriban dos mensajes spam consecutivos es $T = \sum_{i=1}^{N} T_i$. Teniendo en cuenta que las variables $N, T_1, T_2, \dots$ son independientes y que, medida en minutos, $T_i \sim \mathcal{E}(1/15)$, tenemos

$$\mathbf{E}[T] = \mathbf{E}[T_1] \cdot \mathbf{E}[N] = 15 \cdot \mathbf{E}[N].$$

Para calcular $\mathbf{E}[N]$ aplicamos la misma técnica que utilizamos en la resolución anterior. Observamos que

$$N|S_1^c \sim 1 + N, \quad N|S_1S_2^c \sim 2 + N, \quad T|S_1S_2 \sim 2,$$

y aplicamos la *fórmula de probabilidades totales* para obtener la ecuación en $\mathbf{E}[N]$

$$\begin{aligned}\mathbf{E}[N] &= \mathbf{E}[N|S_1^c]\mathbf{P}(S_1^c) + \mathbf{E}[N|S_1S_2^c]\mathbf{P}(S_1S_2^c) + \mathbf{E}[N|S_1S_2]\mathbf{P}(S_1S_2) \\ &= (1 + \mathbf{E}[N])(0.8) + (2 + \mathbf{E}[N])(0.2)(0.8) + 2(0.2)^2 = 1.2 + (0.96)\mathbf{E}[N],\end{aligned}$$

cuya solución es $\mathbf{E}[N] = \frac{1.2}{0.04} = 30$. Finalmente, queda $\mathbf{E}[T] = 15 \cdot 30 = 450$. $\square$

---

5. Sean $U_1, \dots, U_{100}$ variables aleatorias independientes cada una con distribución $\mathcal{U}(0, 2e)$. Calcular aproximadamente

$$\mathbf{P}\left(\log\left(\frac{1}{100}\sum_{i=1}^{100} U_i\right) > \frac{31}{30}\right).$$

R. [*Referencia*: **Ejercicio 8.16**] Queremos calcular aproximadamente la probabilidad

$$p := \mathbf{P}\left(\log\left(\frac{1}{100}\sum_{i=1}^{100} U_i\right) > \frac{31}{30}\right) = \mathbf{P}\left(\sum_{i=1}^{100} U_i > 100 \cdot e^{\frac{31}{30}}\right).$$

Teniendo en cuenta que las variables aleatorias $U_1, \dots, U_{100}$ son independientes e idénticamente distribuidas, y que para la variable $U \sim \mathcal{U}(0, 2e)$ se tiene que $\mathbf{E}[U] = e$ y $\mathbf{var}(U) = e^2/3$, aplicamos el *Teorema central del límite* y queda

$$\mathbf{P}\left(\frac{\sum_{i=1}^{100} U_i - 100 \cdot e}{\sqrt{100 \cdot (e^2/3)}} > z\right) \approx 1 - \Phi(z).$$

Poniendo $z = \frac{x - 100 \cdot e}{\sqrt{100 \cdot (e^2/3)}}$ -y simplificando términos dentro de la probabilidad ubicada del lado derecho de la igualdad- resulta $\mathbf{P}\left(\sum_{i=1}^{100} U_i > x\right) \approx 1 - \Phi\left(\frac{x - 100 \cdot e}{\sqrt{100 \cdot (e^2/3)}}\right)$. En particular, para $x = 100 \cdot e^{\frac{31}{30}}$ resulta que

$$p \approx 1 - \Phi\left(\frac{100 \cdot e^{\frac{31}{30}} - 100 \cdot e}{\sqrt{100 \cdot (e^2/3)}}\right) = 1 - \Phi(0.587) = 1 - 0.7214 = 0.2786.$$

$\square$

PROBABILIDAD y ESTADÍSTICA (61.09 - 81.04)

Evaluación Integradora — Segundo cuatrimestre – 2017

Duración: 4 horas. — 15/II/18 – 9:00 hs.

---

1. Una cueva será iluminada por dos lámparas L_1 y L_2 cuyas duraciones (en horas), T_1 y T_2, son independientes y tienen distribuciones Weibull de parámetros $(2,7)$ y $(2,9)$, cuyas densidades son

$$f_{T_1}(t_1) = \frac{2}{7}\left(\frac{t_1}{7}\right)e^{-(t_1/7)^2}\mathbf{1}\{t_1>0\}, \quad f_{T_2}(t_2) = \frac{2}{9}\left(\frac{t_2}{9}\right)e^{-(t_2/9)^2}\mathbf{1}\{t_2>0\},$$

respectivamente. Las lámparas se encenderán simultáneamente. Calcular la probabilidad de que la lámpara L_1 sea la primera en apagarse.

R. [*Referencia*: **Ejercicio 2.17 (a)** y **Ejercicio 4.14 (b)**] Queremos calcular $\mathbf{P}(T_1 < T_2)$, la probabilidad de que T_1 sea menor que T_2. Para calcular esta probabilidad, podemos integrar la densidad conjunta $f_{T_1,T_2}(t_1,t_2)$ sobre la región $\mathcal{B} = \left\{(t_1,t_2)\in\mathbb{R}^2 : 0 < t_1 < t_2\right\}$, teniendo en cuenta que las variables T_1 y T_2 son independientes y que $\int_0^t \frac{k}{\alpha}\left(\frac{s}{\alpha}\right)^{k-1}e^{-(s/\alpha)^k}ds = 1 - e^{-(t/\alpha)^k}$, tenemos

$$\begin{aligned}
\mathbf{P}(T_1<T_2) &= \iint_{\mathcal{B}} f_{T_1,T_2}(t_1,t_2)dt_1dt_2 = \iint_{\mathcal{B}} f_{T_1}(t_1)f_{T_2}(t_2)dt_2 \\
&= \int_0^\infty f_{T_2}(t_2)\left(\int_0^{t_2} f_{T_1}(t_1)dt_1\right)dt_2 = \int_0^\infty \frac{2}{9}\left(\frac{t_2}{9}\right)e^{-(t_2/9)^2}\left(1-e^{-(t_2/7)^2}\right)dt_2 \\
&= 1 - \int_0^\infty \frac{2}{9}\left(\frac{t_2}{9}\right)e^{-(t_2/9)^2}e^{-(t_2/7)^2}dt_2 = 1 - \int_0^\infty \frac{2}{9}\left(\frac{t_2}{9}\right)e^{-(\sqrt{130}\,t_2/63)^2}dt_2 \\
&= 1 - \frac{49}{130}\int_0^\infty \frac{2\sqrt{130}}{63}\left(\frac{\sqrt{130}t_2}{63}\right)e^{-(\sqrt{130}\,t_2/63)^2}dt_2 = 1 - \frac{49}{130} = \frac{81}{130} \approx 0.623.
\end{aligned}$$

$\square$

Otro modo. Notar que para $T \sim \text{Wei}(k,\alpha)$ se tiene que $T^k \sim \mathcal{E}\left((1/\alpha)^k\right)$, tener en cuenta que la función $t \mapsto t^2$ es creciente sobre la semirrecta positiva, y usar el resultado del **Ejercicio 4.14 (b)**: $\mathbf{P}(T_1<T_2) = \mathbf{P}\left(T_1^2 < T_2^2\right) = \frac{(1/7)^2}{(1/7)^2+(1/9)^2} = \frac{81}{130}$. $\square$

---

2. El panadero llega a la panadería a las 3:00+T, donde T es una variable aleatoria con distribución exponencial de media 5 minutos. La producción del día (en kg) es una variable aleatoria con distribución normal: de media 100 y desvío 10 cuando llega antes de las 3:15, y de media 80 y desvío 10 en caso contrario. La producción del día fue 90 kg, calcular la probabilidad de que el panadero haya llegado a la panadería antes de las 3:15.

R. [*Referencia*: **Ejercicio 5.9**] Sea W la producción del día (en kg). La distribución de W depende de los valores de T: $W|T<15 \sim \mathcal{N}(100,10^2)$ y $W|T\geq 15 \sim \mathcal{N}(80,10^2)$, donde $T \sim \mathcal{E}(1/5)$. Queremos calcular $\mathbf{P}(T<15|W=90)$.

Aplicando la *regla de Bayes para mezclas de variables aleatorias continuas* tenemos

$$\mathbf{P}(T<15|W=90) = \frac{f_{W|T<15}(90)\mathbf{P}(T<15)}{f_W(90)},$$

siendo

$$
\begin{aligned}
f_W(90) &= f_{W|T<15}(90)\cdot \mathbf{P}(T<15) + f_{W|T\geq 15}(90)\cdot \mathbf{P}(T\geq 15)\\
&= \frac{1}{\sqrt{2\pi}\,10}e^{-\frac{1}{2}\left(\frac{90-100}{10}\right)^2}\cdot\left(1-e^{-\frac{15}{5}}\right) + \frac{1}{\sqrt{2\pi}\,10}e^{-\frac{1}{2}\left(\frac{90-80}{10}\right)^2}\cdot e^{-\frac{15}{5}}\\
&= \frac{1}{\sqrt{2\pi}\,10}e^{-\frac{1}{2}}\cdot\left(1-e^{-3}\right) + \frac{1}{\sqrt{2\pi}\,10}e^{-\frac{1}{2}}\cdot e^{-3} = \frac{1}{\sqrt{2\pi}\,10}e^{-\frac{1}{2}}.
\end{aligned}
$$

Por lo tanto,

$$\mathbf{P}(T<15|W=90) = 1-e^{-3}\approx 0.9502.$$

□

---

3. A partir de las 0:00 arriban mensajes a una dirección de correo electrónico de acuerdo con un proceso Poisson de intensidad 4 por hora. Independientemente de los arribos, cada uno de los mensajes es spam con probabilidad 0.2. Calcular la media del tiempo que transcurrirá desde las 0:00 hasta que por primera vez arriben dos mensajes spam consecutivos.

R. [*Referencia*: **Ejercicio 5.13** y **Ejercicio 7.1**] Sea T el tiempo (en minutos) que transcurrirá desde las 0:00 hasta que por primera vez arriben dos mensajes spam consecutivos. Queremos calcular $\mathbf{E}[T]$.

Indicando, para cada $i=1,2$, mediante S_i al evento de que el i-ésimo mensaje que arriba a la casilla de correos después de las 0:00 es spam, se tiene

$$T|S_1^c\sim T_1+T,\quad T|S_1S_2^c\sim T_1+T_2+T,\quad T|S_1S_2\sim T_1+T_2,$$

donde T_1 y T_2 son variables aleatorias independientes con distribución exponencial de media 15. Aplicando la *fórmula de probabilidades totales* tenemos

$$
\begin{aligned}
\mathbf{E}[T] &= \mathbf{E}[T|S_1^c]\mathbf{P}(S_1^c) + \mathbf{E}[T|S_1S_2^c]\mathbf{P}(S_1S_2^c) + \mathbf{E}[T|S_1S_2]\mathbf{P}(S_1S_2)\\
&= (15+\mathbf{E}[T])(0.8) + (30+\mathbf{E}[T])(0.2)(0.8) + 30(0.2)^2\\
&= 12 + (0.8)\mathbf{E}[T] + 4.8 + (0.16)\mathbf{E}[T] + 1.2 = 18 + (0.96)\mathbf{E}[T],
\end{aligned}
$$

y por lo tanto, $\mathbf{E}[T]=\frac{18}{0.04}=450$. □

Otro modo. Consideramos $T_1, T_2, \dots$ la sucesión de tiempos entre arribos de mensajes a la dirección de correo electrónico desde las 0:00 y N la cantidad de mensajes hasta que por primera vez arriban dos mensajes spam consecutivos. Con esa nomenclatura el tiempo, T, que transcurre desde las 0:00 hasta que por primera vez arriban dos mensajes spam consecutivos es $T=\sum_{i=1}^N T_i$. Teniendo en cuenta que las variables $N, T_1, T_2, \dots$ son independientes y que, medida en minutos, $T_i\sim\mathcal{E}(1/15)$, tenemos

$$\mathbf{E}[T]=\mathbf{E}[T_1]\cdot\mathbf{E}[N] = 15\cdot\mathbf{E}[N].$$

Para calcular $\mathbf{E}[N]$ aplicamos la misma técnica que utilizamos en la resolución anterior. Observamos que

$$N|S_1^c\sim 1+N,\quad N|S_1S_2^c\sim 2+N,\quad T|S_1S_2\sim 2,$$

y aplicamos la *fórmula de probabilidades totales* para obtener la ecuación en $\mathbf{E}[N]$

$$\begin{aligned}\mathbf{E}[N] &= \mathbf{E}[N|S_1^c]\mathbf{P}(S_1^c) + \mathbf{E}[N|S_1S_2^c]\mathbf{P}(S_1S_2^c) + \mathbf{E}[N|S_1S_2]\mathbf{P}(S_1S_2) \\ &= (1+\mathbf{E}[N])(0.8) + (2+\mathbf{E}[N])(0.2)(0.8) + 2(0.2)^2 = 1.2 + (0.96)\mathbf{E}[N],\end{aligned}$$

cuya solución es $\mathbf{E}[N] = \frac{1.2}{0.04} = 30$. Finalmente, queda $\mathbf{E}[T] = 15 \cdot 30 = 450$. □

---

4. Los capacitores de la marca *Ching-Pum* pueden estar fallados o no de manera independiente unos de otros. Se examinan tres lotes de 500 capacitores cada uno y se observan las siguientes cantidades de capacitores fallados: 3, 8, 1. En base a esa información muestral estimar por máxima verosimilitud la esperanza de la cantidad de capacitores fallados que habrá en otro lote de 500 capacitores de la marca *Ching-Pum*.

R. [*Referencia*: **Ejercicio 9.6 (b)**] Sea X la cantidad de capacitores fallados en un lote de 500. Tenemos $X \sim \mathcal{B}(500, p)$ y queremos estimar por máxima verosimilitud la función de p, $g(p) := \mathbf{E}_p[X] = 500 \cdot p$ basados en los valores observados (x_1, x_2, x_3) de una muestra aleatoria (X_1, X_2, X_3).

Teniendo en cuenta el *principio de invariancia* del estimador de máxima verosimilitud se tiene

$$\widehat{g(p)}_{mv} = g(\hat{p}_{mv}) = 500 \cdot \hat{p}_{mv},$$

y el problema se reduce a estimar p.

Como la familia de distribuciones $\{\mathcal{B}(500, p) : p \in (0,1)\}$ es regular, el estimador máximo verosímil de p, basado en (x_1, x_2, x_3), satisface la ecuación de verosimilitud

$$\sum_{i=1}^{3} \frac{d}{dp} \log f_p(x_i) = 0,$$

donde $f_p(x) = \binom{500}{x} p^x (1-p)^{500-x}$, $x \in \{0, 1, \dots, 500\}$, es la función de probabilidad de la distribución $\mathcal{B}(500, p)$.

Derivando, respecto de p, $\log f_p(x) = \log \binom{500}{x} + x \log p + (500 - x) \log(1-p)$ resulta

$$\frac{d}{dp} \log f_p(x) = \frac{x}{p} - \frac{500 - x}{1 - p} = \frac{(1-p)x - p(500 - x)}{p(1-p)} = \frac{x - 500p}{p(1-p)}.$$

En consecuencia, la solución de la ecuación de verosimilitud $\frac{x_1 + x_2 + x_3 - 1500 \cdot p}{p(1-p)} = 0$ es

$$p = \frac{x_1 + x_2 + x_3}{1500}.$$

Teniendo en cuenta la información muestral $(x_1, x_2, x_3) = (3, 8, 1)$ resulta $\hat{p}_{mv} = \frac{12}{1500} = \frac{4}{500}$, y el estimador máximo verosímil de la esperanza de la cantidad de capacitores fallados que habrá en otro lote de 500 capacitores es 4. □

---

5. Los rulemanes de cierta marca *premium* tienen una duración aleatoria R (en millones de vueltas) con función de distribución:

$$F_R(r) = \left(1 - \frac{10}{r^2 + 10}\right) \mathbf{1}\{r > 0\}.$$

Para abaratar costos se propone comprar rulemanes *alternativos*. La siguiente tabla muestra la distribución de frecuencias de la duración de 100 rulemanes *alternativos*:

Duración	0-2	2-4	4-8	8-12	12-20
Frecuencia	22	37	26	11	4

En base a esos datos analizar, con un nivel de significación asintótico de 0.05, si se puede rechazar la hipótesis de que la duración de los rulemanes *alternativos* tiene la misma distribución que la duración de los *premium*.

R. [*Referencia*: **Ejercicio 10.32**] Sea X la duración aleatoria (en millones de vueltas) de los rulemanes *alternativos*. Queremos testear la hipótesis $H_0 : F_X = F_R$ contra la hipótesis $H_1 : F_X \neq F_R$, al nivel de significación asintótico $\alpha = 0.05$, basados en datos obtenidos sobre la base de una muestra aleatoria de tamaño $n = 100$, $\mathbf{X} = (X_1, \dots, X_{100})$, de la variable X. Aplicaremos un test de bondad de ajuste χ^2 basado en una tabla de frecuencias.

Como los 100 datos observados aparecen resumidos en una tabla de frecuencias organizada sobre la base de 5 clases C_i, $i = 1, \dots, 5$, en primer lugar, analizamos si las mismas satisfacen las *condiciones de Fischer* sobre la esperanza de las frecuencias $n_i := \sum_{j=1}^{100} \mathbf{1}\{X_j \in C_i\}$ bajo la hipótesis H_0:

$$e_i \geq 5 \text{ para todo } i = 1, \dots, 5,$$

donde $e_i := 100 \cdot \mathbf{P}(R \in C_i)$, porque $n = 100$.

Como por definición, $C_i := (r_{i-1}, r_i]$, $i = 1, \dots, 5$, tenemos

$$p_i := \mathbf{P}(R \in C_i) = F_R(r_i) - F_R(r_{i-1}) = \frac{10}{r_{i-1}^2 + 10} - \frac{10}{r_i^2 + 10},$$

y como $(r_0, r_1, r_2, r_3, r_4, r_5) = (0, 2, 4, 8, 12, \infty)$, se tiene: $p_1 = 2/7 \approx 0.28$, $p_2 = 30/91 \approx 0.32$, $p_3 = 120/481 \approx 0.24$, $p_4 = 200/2849 \approx 0.07$, y $p_5 = 5/77 \approx 0.06$. Teniendo en cuenta que las frecuencias esperadas $e_i = 100 \cdot p_i$ en las clases C_i satisfacen las condiciones de Fischer, podemos aplicar el test de hipótesis χ^2

$$\delta(\mathbf{X}) = \mathbf{1}\left\{ \sum_{i=1}^{5} \frac{(n_i - e_i)^2}{e_i} > \chi^2_{4,0.95} \right\},$$

donde $\chi^2_{4,0.95} = 9.4877$ es el cuantil 0.95 de la distribución χ^2_4.

Basados en la tabla de frecuencias tenemos $n_1 = 22$, $n_2 = 37$, $n_3 = 26$, $n_4 = 11$ y $n_5 = 4$. Reemplazando esos valores y los de las frecuencias esperadas en el estadístico $\sum_{i=1}^{5} \frac{(n_i - e_i)^2}{e_i}$, resulta

$$\sum_{i=1}^{5} \frac{(n_i - e_i)^2}{e_i} = 5.2631,$$

y como $5.2631 < 9.4877$, se concluye que NO HAY EVIDENCIA PARA RECHAZAR la hipótesis de que la duración de los rulemanes *alternativos* tiene la misma distribución que la duración de los *premium*. □