Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Temporada Académica de Verano 2015

Curso	**:**	**Probabilidad y Estadística**
Sigla	**:**	**EYP1113**
Pauta	**:**	**I3**
Profesores	**:**	**Ricardo Aravena C. y Ricardo Olea O.**

Problema 1

El consumo de agua potable (en m^3) en un hogar puede ser modelado mediante una distribución Weibull(η, β), la cual tiene función de probabilidad acumulada

$$F_X(x) = 1 - \exp\left[-\left(\frac{x}{\eta}\right)^\beta\right], \quad x \geq 0, \quad \eta > 0, \quad \beta > 0$$

Considere $X_1, \ldots, X_n$ como el consumo en m^3 de agua potable de una muestra aleatoria de hogares, los cuales se comportan de manera independientes. Si k_p es el percentil $p \times 100\,\%$ del consumo mínimo de agua potable en los hogares perteneciente a la muestra, muestre que

$$\ln(k_p) = \ln(\eta) + \frac{1}{\beta}\left[\ln\left(-\ln\left(1-p\right)\right) - \ln\left(n\right)\right]$$

Solución

Alternativa 1: Sea $Y = \text{mín}\{X_1, \ldots, X_n\}$, del formulario se tiene que

$$\begin{aligned}
f_Y(y) &= n \cdot \left\{\exp\left[-\left(\frac{y}{\eta}\right)^\beta\right]\right\}^{n-1} \cdot \frac{\beta}{\eta} \cdot \left(\frac{y}{\eta}\right)^{\beta-1} \cdot \exp\left[-\left(\frac{y}{\eta}\right)^\beta\right] \quad \textbf{[0.8 Ptos.]} \\
&= \frac{n\,\beta}{\eta} \cdot \left(\frac{y}{\eta}\right)^{\beta-1} \cdot \exp\left[-n\left(\frac{y}{\eta}\right)^\beta\right] \quad \textbf{[0.8 Ptos.]} \\
&= \frac{n^{1/\beta}\,\beta}{\eta} \cdot \left(n^{1/\beta}\,\frac{y}{\eta}\right)^{\beta-1} \cdot \exp\left[-\left(n^{1/\beta}\,\frac{y}{\eta}\right)^\beta\right], \quad y \geq 0, \quad \eta > 0, \quad \beta > 0, \quad n \in \mathbb{N} \quad \textbf{[0.8 Ptos.]}
\end{aligned}$$

Es decir, $Y \sim \text{Weibull}\left(\frac{\eta}{n^{1/\beta}}, \beta\right)$**[0.8 Ptos.]**, por lo tanto

$$F_Y(y) = 1 - \exp\left[-\left(\frac{y}{\eta/n^{1/\beta}}\right)^\beta\right], \quad y \geq 0, \quad \eta > 0, \quad \beta > 0, \quad n \in \mathbb{N} \quad \textbf{[0.8 Ptos.]}$$

Si k_p es el percentil $p \times 100\,\%$, entonces $F_Y(k_p) = p$. **[0.4 Ptos.]**

Reemplazando se tiene que

$$1 - \exp\left[-\left(\frac{y}{\eta/n^{1/\beta}}\right)^\beta\right] = p \quad \textbf{[0.4 Ptos.]}$$

$$\Rightarrow \exp\left[-\left(\frac{y}{\eta/n^{1/\beta}}\right)^{\beta}\right] = 1 - p \quad \textbf{[0.2 Ptos.]}$$

$$\Rightarrow \left[-\left(\frac{y}{\eta/n^{1/\beta}}\right)^{\beta}\right] = \ln(1-p) \quad \textbf{[0.2 Ptos.]}$$

$$\Rightarrow \left(\frac{y}{\eta/n^{1/\beta}}\right)^{\beta} = -\ln(1-p) \quad \textbf{[0.2 Ptos.]}$$

$$\Rightarrow \beta\,\ln(k_p) - \beta\,\ln(\eta) + \ln(n) = \ln(-\ln(1-p)) \quad \textbf{[0.2 Ptos.]}$$

$$\Rightarrow \ln(k_p) = \ln(\eta) - \frac{1}{\beta}\,\ln(n) + \frac{1}{\beta}\,\ln(-\ln(1-p)) \quad \textbf{[0.2 Ptos.]}$$

$$\Rightarrow \ln(k_p) = \ln(\eta) + \frac{1}{\beta}\left[\ln\left(-\ln\left(1-p\right)\right) - \ln\left(n\right)\right] \quad \textbf{[0.2 Ptos.]}$$

Alternativa 2: Sea $Y = \text{mín}\{X_1, \ldots, X_n\}$ con función de probabilidad acumulada dada por

$$\begin{aligned}
F_Y(y) &= 1 - P(Y > y) \quad \textbf{[0.8 Ptos.]}\\
&= 1 - P(X_1 > y, \ldots, X_n > y) \quad \textbf{[0.8 Ptos.]}\\
&= 1 - \prod_{i=1}^{n} P(X_i > y), \quad \text{por independencia} \quad \textbf{[0.8 Ptos.]}\\
&= 1 - [1 - F_X(y)]^n, \quad \text{por idéntica distribución} \quad \textbf{[0.8 Ptos.]}\\
&= 1 - \exp\left[-n\left(\frac{y}{\eta}\right)^{\beta}\right], \quad y \geq 0, \quad \eta > 0, \quad \beta > 0, \quad n \in \mathbb{N} \quad \textbf{[0.8 Ptos.]}
\end{aligned}$$

Si k_p es el percentil $p \times 100\,\%$, entonces $F_Y(k_p) = p$. **[0.4 Ptos.]**

Reemplazando se tiene que

$$1 - \exp\left[-n\left(\frac{y}{\eta}\right)^{\beta}\right] = p \quad \textbf{[0.4 Ptos.]}$$

$$\Rightarrow \exp\left[-n\left(\frac{y}{\eta}\right)^{\beta}\right] = 1 - p \quad \textbf{[0.2 Ptos.]}$$

$$\Rightarrow \left[-n\left(\frac{y}{\eta}\right)^{\beta}\right] = \ln(1-p) \quad \textbf{[0.2 Ptos.]}$$

$$\Rightarrow n\left(\frac{y}{\eta}\right)^{\beta} = -\ln(1-p) \quad \textbf{[0.2 Ptos.]}$$

$$\Rightarrow \beta\,\ln(k_p) - \beta\,\ln(\eta) + \ln(n) = \ln(-\ln(1-p)) \quad \textbf{[0.2 Ptos.]}$$

$$\Rightarrow \ln(k_p) = \ln(\eta) - \frac{1}{\beta}\,\ln(n) + \frac{1}{\beta}\,\ln(-\ln(1-p)) \quad \textbf{[0.2 Ptos.]}$$

$$\Rightarrow \ln(k_p) = \ln(\eta) + \frac{1}{\beta}\left[\ln\left(-\ln\left(1-p\right)\right) - \ln\left(n\right)\right] \quad \textbf{[0.2 Ptos.]}$$

+ 1 Punto Base

Problema 2

La autoridad ha informado que el retraso en publicar los resultados de la CASEN se debe a las dificultades encontradas en la nueva forma de construir el indicador.

Suponga que se han definido dos fórmulas para medir la pobreza a partir de tres dimensiones - digamos indicador 1 e indicador 2 - las cuales son obtenidas a partir de diferentes variables de la encuesta CASEN. Cada dimension se sintetiza en una variable. Estas variables son construidas de tal forma que se comporten como una distribución normal de media cero y varianza 1. Además, las variables al interior de cada indicador son independientes entre si. PERO hay correlación entre las variables de la misma dimensión. La correlación entre estas variables y las ponderaciones utilizada por cada indicador se presentan en la siguiente tabla.

Dimensión	IND1 - variables	PESO	IND2 - variables	PESO	Correlación
Educación	años de estudio	0.1	último nivel aprobado	0.2	+0.6
Ingresos	salarios, otros ingresos	0.5	salarios, otros beneficios	0.5	+0.8
Calidad de Vida	vivenda, acceso salud	0.4	entorno, gastos y cobertura salud	0.3	-0.5

El indicador final se obtiene como la suma ponderada (pesos) de cada una de las variables a cuyo resultado se le suma la pobreza del último periodo (es decir 0,15 ya que alcanzó a un 15 %).

Determine la probabilidad que el indicador 1 muestre un nivel de pobreza inferior al 90 % nivel de pobreza obtenido con el indicador 2.

Solución

Definamos las siguientes variables:

- X_1: años de estudio.
- X_2: salarios, otros ingresos.
- X_3: vivenda, acceso salud.
- Y_1: último nivel aprobado.
- Y_2: salarios, otros beneficios.
- Y_3: entorno, gastos y cobertura salud.
- U: Indicador 1.
- V: Indicador 2.

donde

$$U = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 \qquad \text{y} \qquad V = b_0 + b_1 Y_1 + b_2 Y_2 + b_3 Y_3 \quad \textbf{[0.5 Ptos.]}$$

Se pide

$$P(U < 0.9 \cdot V) = P(W < 0) \quad \textbf{[0.5 Ptos.]}$$

con

$$W = (a_0 - 0.9\, b0) + a_1 X_1 + a_2 X_2 + a_3 X_3 - 0.9\, b_1 Y_1 - 0.9\, b_2 Y_2 - 0.9\, b_3 Y_3 \quad \textbf{[0.5 Ptos.]}$$

Como W es una combinanción lineal de variables aleatorias con distribución Normal(0,1), entonces

$$W \sim \text{Normal}(\mu, \sigma) \quad \textbf{[0.5 Ptos.]}$$

con

$$\mu = (a_0 - 0.9\, b0) = 0.015 \quad \textbf{[0.5 Ptos.]}$$

y

$$
\begin{aligned}
\sigma^2 = {} & a_1^2 \cdot \mathrm{Cov}(X_1,\, X_1) + a_2^2 \cdot \mathrm{Cov}(X_2,\, X_2) + a_3^2 \cdot \mathrm{Cov}(X_3,\, X_3) \\
& + 0.9^2\, b_1^2 \cdot \mathrm{Cov}(Y_1,\, Y_1) + 0.9^2\, b_1^2 \cdot \mathrm{Cov}(Y_2,\, Y_2) + 0.9^2\, b_1^2 \cdot \mathrm{Cov}(Y_3,\, Y_3) \\
& - 2 \cdot 0.9 \cdot a_1 \cdot b_1 \cdot \mathrm{Cov}(X_1,\, Y_1) - 2 \cdot 0.9 \cdot a_2 \cdot b_2 \cdot \mathrm{Cov}(X_2,\, Y_2) - 2 \cdot 0.9 \cdot a_2 \cdot b_2 \cdot \mathrm{Cov}(X_3,\, Y_3) \quad \textbf{[1.5 Ptos.]}
\end{aligned}
$$

ya que solo hay dependencia por dimensión.

Dado que todas las variables aleatorias tiene varianza unitaria, entonces

$\mathrm{Cov}(X_1,\, Y_1) = \mathrm{Corr}(X_1,\, Y_1)$, $\quad \mathrm{Cov}(X_2,\, Y_2) = \mathrm{Corr}(X_2,\, Y_2)$, $\quad \mathrm{Cov}(X_3,\, Y_3) = \mathrm{Corr}(X_3,\, Y_3)$ **[0.5 Ptos.]**

Reemplazando se tiene que

$$W \sim \text{Normal}(0.015,\, 0.6739) \quad \textbf{[1.0 Ptos.]}$$

Por tanto

$$P(W < 0) = \Phi\left(\frac{0.000 - 0.015}{0.6739}\right) = \Phi(-0.022) = 1 - \Phi(0.022) = 1 - 0.5080 = 0.492 \quad \textbf{[0.5 Ptos.]}$$

+ 1 Punto Base

Problema 3

La distribución Gamma(k, ν) es muy usada para modelar tiempos por ejemplo de espera, vida y falla, entre otros. Usualmente el parámetro k es conocido, cuando se quiere modelar tiempos asociados a eventos que ocurren dentro de un proceso de Poisson. Considere entonces una muestra aleatoria $X_1, \ldots, X_n$ proveniente de una distribución Gamma(k, ν) relacionada a eventos que ocurren según un proceso de Poisson con k conocido.

(a) **[2.0 Ptos.]** Obtenga paso a paso el estimador máximo verosímil de ν.

(b) **[2.0 Ptos.]** Obtenga paso a paso la cota de Cramer-Rao para un estimador de ν.

(c) **[2.0 Ptos.]** Calcule mediante aproximación de primer orden la varianza del estimador de ν obtenido en (a) y compare con la varianza asintótica del estimador máximo verosímil de ν.

Solución

(a) La función de verosimilitud por la independencia de la muestra y su idéntica distribución está dada por

$$L(\nu) = \frac{\nu^{n\,k}}{[\Gamma(k)]^n} \cdot \left(\prod_{i=1}^n X_i\right)^{k-1} \cdot \exp\left(-\nu \sum_{i=1}^n X_i\right) \quad \textbf{[0.5 Ptos.]}$$

mientras que su logaritmo natural queda como

$$\ln\,L(\nu) = n\,k\,\ln(\nu) - n\,\ln\Gamma(k) + (k-1)\sum_{i=1}^n \ln(X_i) - -\nu \sum_{i=1}^n X_i \quad \textbf{[0.5 Ptos.]} \tag{1}$$

derivando parcialmente (**??**) con respecto a ν y luego igualando a cero se tiene que

$$\textbf{[0.5 Ptos.]} \quad \frac{\partial}{\partial\,\nu}\ln\,L(\nu) = \frac{n\,k}{\nu} - \sum_{i=1}^n X_i = 0 \Rightarrow \hat{\nu} = \frac{n\,k}{\sum_{i=1}^n X_i} \quad \textbf{[0.5 Ptos.]}$$

(b) Derivando parcialmente doce veces (**??**) con respecto a ν se tiene que

$$\textbf{[0.5 Ptos.]} \quad \frac{\partial^2}{\partial\,\nu^2}\ln\,L(\nu) = -\frac{n\,k}{\nu^2} \Rightarrow I(\nu) = \frac{n\,k}{\nu^2} \quad \textbf{[1.0 Ptos.]}$$

Luego

$$\text{CCR} = \frac{\nu^2}{n\,k} \quad \textbf{[0.5 Ptos.]}$$

(c) Sea $Y = \frac{1}{n\,k}\sum_{i=1}^n X_i \sim \text{Gamma}(n\,k,\, n\,k\,\nu)$ **[0.5 Ptos.]**, lo que implica que $\mu_Y = \frac{1}{\nu}$ **[0.2 Ptos.]** y $\sigma_Y^2 = \frac{1}{n\,k\,\nu^2}$. **[0.2 Ptos.]**

Tenemos que

$$\hat{\nu} = g(Y) \approx g(\mu_Y) + (Y - \mu_Y)\,g'(\mu_Y) \quad \textbf{[0.3 Ptos.]}$$

con $g(Y) = \frac{1}{Y}$. **[0.2 Ptos.]**

Se pide

$$\text{Var}(\hat{\nu}) \approx \text{Var}\left[g(\mu_Y) + (Y - \mu_Y)\,g'(\mu_Y)\right] = \left[g'(\mu_Y)\right]^2\,\sigma_Y^2 = \frac{1}{1/\nu^4} \cdot \frac{1}{n\,k\,\nu^2} = \frac{\nu^2}{n\,k} \quad \textbf{[0.5 Ptos.]}$$

Vemos que coincide con la cota de Cramer-Rao, que corresponde a la varianza asintótica. **[0.1 Ptos.]**

+ 1 Punto Base

Problema 4

Como usted debe saber, las vigas de madera laminadas deben cumplir normas estrictas respecto a la resistencia. Suponga que se ha establecido que las vigas deben presentar una resistencia media de al menos 3.0 kN/m^2, con una variabilidad (en términos de desviación estándar) no mayor a 0.2 kN/m^2. Suponga que se recibe una partida de un millar de vigas y se seleccionan 50 vigas las cuales son sometidas a diferentes cargas, observando:

- 16 de ellas no resisten la carga máxima.
- Entre las que no resisten la carga máxima, se observa una resistencia media 2.9 kN/m^2 con una desviación estándar de 0.25 kN/m^2.
- Entre las que resisten la carga máxima se tiene una resistencia media 3.2 kN/m^2 con una desviación estándar de 0.35 kN/m^2.

(a) **[2.0 Ptos.]** ¿Existe evidencia estadística que permita afirmar que más de un cuarto de las vigas no resisten la carga máxima? Use $\alpha = 10\,\%$.

(b) **[4.0 Ptos.]** Asumiendo normalidad, ¿existe evidencia estadística que permita indicar que, entre las vigas que no resisten la carga máxima, no satisfacen: (i) la normativa respecto a la resistencia media; (ii) la normativa respecto a la variabilidad? Para ambos casos use $\alpha = 5\,\%$.

(a) Definamos como p a la probabilidad que una viga NO resista la carga máxima.

Interesa realizar una prueba de hipótesis sobre el parámetro p, con $n = 50$.

- Hipótesis
$$\text{H}_0 : p = 0.25 \quad \text{vs.} \quad \text{H}_a : p > 0.25 \quad \textbf{[0.8 Ptos.]}$$
- Test
$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0}{n}}} = \frac{0.32 - 0.25}{\sqrt{\frac{0.25 \cdot 0.75}{50}}} = 1.143 \quad \textbf{[0.7 Ptos.]}$$
- Decisión

Alternativa 1: Como $Z_0 = 1.143$ no es mayor que $k_{0.90} = 1.28$, no existe suficiente evidencia contra H_0, es decir no se puede afirmar que más de un cuarto de las vigas no resisten la carga máxima **[0.5 Ptos.]**

Alternativa 2: Tenemos que el valor-p $= P(Z > 1.143) = 1 - 0.873 = 0.127$, como este no es menor a un $10\,\%$, no existe suficiente evidencia para rechaza H_0, es decir no se puede afirmar que más de un cuarto de las vigas no resisten la carga máxima **[0.5 Ptos.]**

(b) (i) Test sobre μ con σ desconocido, distribución normal con $n = 16$

- Hipótesis
$$\text{H}_0 : \mu = 3.0 \qquad \text{vs.} \qquad \text{H}_a : \mu < 3 \quad \textbf{[0.8 Ptos.]}$$
- Test
$$T_0 = \frac{\overline{X}_n - \mu_0}{S/\sqrt{n}} = \frac{2.9 - 3.0}{0.25/\sqrt{16}} = -1.6 \quad \textbf{[0.7 Ptos.]}$$
- Decisión

Alternativa 1: Como $T_0 = -1.6$ no es menor que $t_{0.05}(15) = -1.753$ no existe evidencia suficiente contra H_0, es decir, la normativa respecto a la resistencia media se satisface. **[0.5 Ptos.]**

Alternativa 2: Tenemos que el valor-p $= P(T < -1.6) \to 0.05 <$ valor-p < 0.10. Como valor-p no es menor que un 5 %, no existe evidencia suficiente contra H_0, es decir, la normativa respecto a la resistencia media se satisface. **[0.5 Ptos.]**

(ii) Test sobre σ, distribución normal y $n = 16$.

- Hipótesis

$$\mathrm{H}_0 : \sigma = 0.2 \qquad \text{vs.} \qquad \mathrm{H}_0 : \sigma > 0.2 \quad \textbf{[0.8 Ptos.]}$$

- Test

$$C_0 = \frac{(16-1)0.25^2}{0.2^2} = 23.4375 \quad \textbf{[0.7 Ptos.]}$$

- Decisión

Como $C_0 = 23.7375$ no es mayor $c_{0.95}(15) = 25$, no existe suficiente evidencia contra H_0. Es decir, la normativa respecto a la variabilidad se satisface. **[0.5 Ptos.]**

+ 1 Punto Base

Formulario

Propiedades función $\Gamma(\cdot)$

$$(1) \quad \Gamma(k) = \int_0^\infty u^{k-1}\, e^{-u}\, du; \quad (2) \quad \Gamma(a+1) = a\,\Gamma(a);$$

$$(3) \quad \Gamma(n+1) = n!, \quad \text{si } n \in \mathbb{N}_0; \quad (4) \quad \Gamma(1/2) = \sqrt{\pi}$$

Propiedades función $B(\cdot,\, \cdot)$

$$(1) \quad B(q,\, r) = \int_0^1 x^{q-1}\, (1-x)^{r-1}\, dx; \quad (2) \quad B(q,\, r) = \frac{\Gamma(q)\,\Gamma(r)}{\Gamma(q+r)}$$

Igualdades

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k\, b^{n-k}, \qquad \sum_{k=x}^{\infty} \phi^k = \frac{\phi^x}{1-\phi} \quad \text{si } |\phi| < 1, \qquad \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \exp(\lambda)$$

Propiedad distribución Gamma

$$\text{Si } T \sim \text{Gamma}(k,\, \nu), \text{ con } k \in \mathbb{N} \longrightarrow F_T(t) = 1 - \sum_{x=0}^{k-1} \frac{(\nu\, t)^x\, e^{-\nu\, t}}{x!}$$

Transformación

Sea $Y = g(X)$ una función cualquiera, con k raíces:

$$f_Y(y) = \sum_{i=1}^{k} f_X\left(g_i^{-1}(y)\right) \cdot \left| \frac{d}{dy}\, g_i^{-1}(y) \right|$$

$$p_Y(y) = \sum_{i=1}^{k} p_X\left(g_i^{-1}(y)\right)$$

Sea $Z = g(X, Y)$ una función cualquiera:

$$p_Z(z) = \sum_{g(x,y)=z} p_{X,Y}(x, y)$$

Sea $Z = g(X, Y)$ una función invertible para X o Y fijo:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{X,Y}(g^{-1}, y) \left| \frac{\partial}{\partial z} g^{-1} \right| dy \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x, g^{-1}) \left| \frac{\partial}{\partial z} g^{-1} \right| dx \end{aligned}$$

Esperanza y Varianza Condicional

$$\text{E}(Y) = \text{E}[\text{E}(Y \mid X)] \quad \text{y} \quad \text{Var}(Y) = \text{Var}[\text{E}(Y \mid X)] + \text{E}[\text{Var}(Y \mid X)]$$

Teorema del Límite Central

Sean $X_1,\, \ldots,\, X_n$ variables aleatorias independientes e idénticamente distribuidas, entonces

$$Z_n = \frac{\displaystyle\sum_{i=1}^{n} X_i - n \cdot \mu}{\sqrt{n}\, \sigma} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \longrightarrow Z \sim \text{Normal}(0, 1),$$

cuando $n \to \infty$, $\text{E}(X_i) = \mu$ y $\text{Var}(X_i) = \sigma^2$.

Propiedades Esperanza, Varianza y Covarianza

Sean $X_1,\, X_2,\, \ldots,\, X_n,\, Y_1,\, Y_2,\, \ldots,\, Y_m$ variables aleatorias y $a_0,\, a_1,\, \ldots,\, a_n,\, b_0,\, b_1,\, \ldots,\, b_m$ constantes conocidas.

- $\mathrm{E}\left(a_0 + \sum_{i=1}^{n} a_i \cdot X_i\right) = a_0 + \sum_{i=1}^{n} a_i \cdot \mathrm{E}(X_i)$.
- $\mathrm{Cov}\left(a_0 + \sum_{i=1}^{n} a_i \cdot X_i,\, b_0 + \sum_{j=1}^{m} b_j \cdot Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i \cdot b_j \cdot \mathrm{Cov}\,(X_i, Y_j)$.
- $\mathrm{Var}\left(a_0 + \sum_{i=1}^{n} a_i \cdot X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \cdot a_j \cdot \mathrm{Cov}\,(X_i, X_j)$.
- Si $X_1, \ldots, X_n$ son variables aleatorias independientes, entonces

$$\mathrm{Var}\left(a_0 + \sum_{i=1}^{n} a_i \cdot X_i\right) = \sum_{i=1}^{n} a_i^2 \cdot \mathrm{Var}\,(X_i)$$

Mínimo y Máximo

Sean $X_1, \ldots, X_n$ variables aleatorias continuas independientes con idéntica distribución (f_X y F_X), entonces para:

$$Y_1 = \text{mín}\{X_1, \ldots, X_n\} \longrightarrow f_{Y_1} = n\,[1 - F_X(y)]^{n-1}\, f_X(y)$$
$$Y_n = \text{máx}\{X_1, \ldots, X_n\} \longrightarrow f_{Y_n} = n\,[F_X(y)]^{n-1}\, f_X(y)$$

Mientras que la distribución conjunta entre Y_1 e Y_n está dada por:

$$f_{Y_1,Y_n}(u,v) = n\,(n-1)\,[F_X(v) - F_X(u)]^{n-2}\, f_X(v)\, f_X(u), \qquad u \leq v$$

Aproximación de Momentos

Sean $X_1, \ldots, X_n$ variables aleatorias con valores esperados $\mu_{X_1}, \ldots, \mu_{X_n}$ y varianzas $\sigma^2_{X_1}, \ldots, \sigma^2_{X_n}$ e Y una función de ellas. La aproximación de primer orden está dada por

$$Y \approx g(\mu_{X_1}, \ldots, \mu_{X_n}) + \sum_{i=1}^{n} (X_i - \mu_{X_i})\,\frac{\partial}{\partial X_i} g(\mu_{X_1}, \ldots, \mu_{X_n})$$

Estimador Máximo Verosímil

Sea $X_1,\ldots,X_n$ una muestra aleatoria independiente e idénticamente distribuida con función de probabilidad p_X o de densidad f_X, determinada por un parámetro θ. Si $\hat{\theta}$ es el estimador máximo verosímil del parámetro θ, entonces:

- $\mathrm{E}(\hat{\theta}) \to \theta$, cuando $n \to \infty$.
- $\mathrm{Var}(\hat{\theta}) = \dfrac{1}{I_n(\theta)}$, con $I_n(\theta) = -\mathrm{E}\left[\dfrac{\partial^2}{\partial\,\theta^2}\,\ln L(\theta)\right]$.
- $\hat{\theta} \mathrel{\dot\sim} \text{Normal}\left(\theta,\, \sqrt{\dfrac{1}{I_n(\theta)}}\right)$, cuando $n \to \infty$.
- El estimador máximo verosímil de $g(\theta)$ es $g(\hat{\theta})$, cuya varianza está dada por: $\mathrm{Var}[g(\hat{\theta})] = \dfrac{[g'(\theta)]^2}{I_n(\theta)}$.

Error Cuadrático Medio

El error cuadrático medio de un estimador $\hat{\theta}$ de θ se define como:

$$\mathrm{ECM}(\hat{\theta}) = \mathrm{E}\left\{\left[\hat{\theta} - \mathrm{E}(\hat{\theta})\right]^2\right\} = \mathrm{Var}(\hat{\theta}) + \text{Sesgo}^2$$

Distribuciones Muestrales

Sean $X_1,\ldots,X_n$ variables aleatorias independientes e idénticamente distribuidas Normal(μ, σ), entonces

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}(0,1), \quad \frac{\overline{X}_n - \mu}{s/\sqrt{n}} \sim \text{t-student}(n-1), \quad \frac{s^2\,(n-1)}{\sigma^2} \sim \chi^2_{n-1}$$

con $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$.

Distribución	Densidad de Probabilidad	Θ_X	Parámetros	Esperanza y Varianza
Binomial	$\binom{n}{x}\, p^x\, (1-p)^{n-x}$	$x = 0, \ldots, n$	$n,\, p$	$\mu_X = n\,p$ $\sigma^2_X = n\,p\,(1-p)$ $M(t) = [p\,e^t + (1-p)]^n, \quad t \in \mathbb{R}$
Geométrica	$p\,(1-p)^{x-1}$	$x = 1, 2, \ldots$	p	$\mu_X = 1/p$ $\sigma^2_X = (1-p)/p^2$ $M(t) = p\,e^t/[1-(1-p)\,e^t], \quad t < -\ln(1-p)$
Binomial-Negativa	$\binom{x-1}{r-1}\, p^r\,(1-p)^{x-r}$	$x = r, r+1, \ldots$	$r,\, p$	$\mu_X = r/p$ $\sigma^2_X = r\,(1-p)/p^2$ $M(t) = \left\{p\,e^t/[1-(1-p)\,e^t]\right\}^r, \quad t < -\ln(1-p)$
Poisson	$\dfrac{(\nu\,t)^x\,e^{-\nu\,t}}{x!}$	$x = 0, 1, \ldots$	ν	$\mu_X = \nu\,t$ $\sigma^2_X = \nu\,t$ $M(t) = \exp\left[\lambda\left(e^t - 1\right)\right], \quad t \in \mathbb{R}$
Exponencial	$\nu\,e^{-\nu\,x}$	$x \geq 0$	ν	$\mu_X = 1/\nu$ $\sigma^2_X = 1/\nu^2$ $M(t) = \nu/(\nu - t), \quad t < \nu$
Gamma	$\dfrac{\nu^k}{\Gamma(k)}\, x^{k-1}\,e^{-\nu\,x}$	$x \geq 0$	$k,\, \nu$	$\mu_X = k/\nu$ $\sigma^2_X = k/\nu^2$ $M(t) = [\nu/(\nu - t)]^k, \quad t < \nu$
Normal	$\dfrac{1}{\sqrt{2\,\pi}\,\sigma}\, \exp\left[-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right]$	$-\infty < x < \infty$	$\mu,\, \sigma$	$\mu_X = \mu$ $\sigma^2_X = \sigma^2$ $M(t) = \exp(\mu\,t + \sigma^2\,t^2/2), \quad t \in \mathbb{R}$
Log-Normal	$\dfrac{1}{\sqrt{2\,\pi}\,(\zeta\,x)}\, \exp\left[-\dfrac{1}{2}\left(\dfrac{\ln x - \lambda}{\zeta}\right)^2\right]$	$x \geq 0$	$\lambda,\, \zeta$	$\mu_X = \exp\left(\lambda + \dfrac{1}{2}\,\zeta^2\right)$ $\sigma^2_X = \mu^2_X\left(e^{\zeta^2} - 1\right)$ $E(X^r) = e^{r\,\lambda}\,M_Z(r\,\zeta)$, con $Z \sim$Normal(0,1)
Uniforme	$\dfrac{1}{(b-a)}$	$a \leq x \leq b$	$a,\, b$	$\mu_X = (a+b)/2$ $\sigma^2_X = (b-a)^2/12$ $M(t) = [e^{t\,b} - e^{t\,a}]/[t\,(b-a)], \quad t \in \mathbb{R}$
Beta	$\dfrac{1}{B(q,\, r)}\, \dfrac{(x-a)^{q-1}\,(b-x)^{r-1}}{(b-a)^{q+r-1}}$	$a \leq x \leq b$	$q,\, r$	$\mu_X = a + \frac{q}{q+r}\,(b-a)$ $\sigma^2_X = \dfrac{q\,r\,(b-a)^2}{(q+r)^2\,(q+r+1)}$
Hipergeométrica	$\dfrac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$	$\text{máx}\{0, n+m-N\} \leq x \leq \text{mín}\{n, m\}$	$N,\, m,\, n$	$\mu_X = n\,\frac{m}{N}$ $\sigma^2_X = \left(\frac{N-n}{N-1}\right)\, n\,\frac{m}{N}\,\left(1 - \frac{m}{N}\right)$

Tablas de Percentiles p

Distribución Normal Estándar k_p

k_p	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Distribución t-student $t_p(\nu)$

ν	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$
1	3.078	6.314	12.706	31.821
2	1.886	2.920	4.303	6.965
3	1.638	2.353	3.182	4.541
4	1.533	2.132	2.776	3.747
5	1.476	2.015	2.571	3.365
6	1.440	1.943	2.447	3.143
7	1.415	1.895	2.365	2.998
8	1.397	1.860	2.306	2.896
9	1.383	1.833	2.262	2.821
10	1.372	1.812	2.228	2.764
11	1.363	1.796	2.201	2.718
12	1.356	1.782	2.179	2.681
13	1.350	1.771	2.160	2.650
14	1.345	1.761	2.145	2.624
15	1.341	1.753	2.131	2.602
16	1.337	1.746	2.120	2.583
17	1.333	1.740	2.110	2.567
18	1.330	1.734	2.101	2.552
19	1.328	1.729	2.093	2.539
20	1.325	1.725	2.086	2.528
21	1.323	1.721	2.080	2.518
22	1.321	1.717	2.074	2.508
23	1.319	1.714	2.069	2.500
24	1.318	1.711	2.064	2.492
25	1.316	1.708	2.060	2.485
26	1.315	1.706	2.056	2.479
27	1.314	1.703	2.052	2.473
28	1.313	1.701	2.048	2.467
29	1.311	1.699	2.045	2.462
30	1.310	1.697	2.042	2.457
∞	1.282	1.645	1.960	2.326

Distribución Chi-Cuadrado $c_p(\nu)$

ν	$c_{0.025}$	$c_{0.05}$	$c_{0.10}$	$c_{0.90}$	$c_{0.95}$	$c_{0.975}$	$c_{0.99}$	$c_{0.995}$
1	0.00	0.00	0.02	2.71	3.84	5.02	6.63	7.88
2	0.05	0.10	0.21	4.61	5.99	7.38	9.21	10.60
3	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84
4	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93