Estadística Matemática

Cuaderno de Ejercicios

Ernesto Barrios

17 de enero de 2024

Versión 0.47

Índice

Prefacio

Los ejercicios en este documento consisten en problemas teóricos y de cálculo de probabilidades y estadística matemática seleccionados de varios textos para apoyar mi curso de Estadística Matemática impartido en ITAM.

La mayoría de los problemas tienen la referencia a los libros de donde fueron tomados. La ficha bibliográfica completa se incluye al final del documento. Aquellos que no tienen referencia es porque no recordamos de dónde fueron extraídos o bien son de nuestra autoría.

Con el apoyo de Álvaro López Zamora y Fernando Pérez Millán hemos iniciado la colección de las respuestas a los problemas. Estamos trabajando en ello. Si de las respuestas publicadas al final del documento no coincide con la suya agradeceremos nos lo haga nota para corregirlas.

Cualquier comentario y/o sugerencia será bienvenido. Diríjalo a Ernesto Barrios `<ebarrios at itam.mx>`.

Ciudad de México, 17 de enero de 2024

1. Introducción

1. En el libro de Chihara and Hesterberg (2019) encontrará, por ejemplo, la liga `https://www.bts.gov/topics/airlines-and-airports/quick-links-popular-air-carrier-statistics` donde encontrará estadísticas (información). ¿Cómo describir de manera eficiente la información incluida en ese sitio? ¿Gráficas, tablas, *dashboards*? Su respuesta es tema de investigación y desarrollo de la **estadística descriptiva** o el **análisis exploratorio** de datos. Vea por ejemplo Tukey (2020).

2. Transformaciones, estadísticos de orden y Método Delta

Transformaciones

1. Sea Z una variable aleatoria distribuida normal estándar. Utilice el teorema de transformación para mostrar que $X = \mu + \sigma Z$ se distribuye normalmente con media μ y varianza σ^2.
2. Sea $X \sim \text{Gamma}(\alpha, \beta)$, con α el parámetro de forma y β el parámetro de escala de la distribución Gamma. Sea $k > 0$, verifique que $Y = kX \sim \text{Gamma}(\alpha, k\beta)$.
3. Sea Z distribuida normal estándar. Calcule la *f. g. m.* (*f. c.*) de Z^2. Utilice el teorema de unicidad para mostrar que $Y = Z^2$ sigue una distribución $\text{Gamma}(\alpha = 1/2, \beta = 2)$, conocida como distribución Ji-cuadrada con 1 grado de libertad y denotada por χ^2_1.
4. Sea $Y \sim \chi^2_k$ y denote $\chi^2(p, k)$ su p-ésimo cuantil, esto es, $p = \mathbb{P}(Y \leq \chi^2(p, k))$. Sean $r > 0$, $W = rY$ y $0 < q < 1$. Determine w de manera que $q = \mathbb{P}(W \leq w)$ y expréselo como función de los cuantiles de la distribución χ^2.

Estadísticos de orden

5. Sea $\boldsymbol{U} = (U_1, \ldots, U_n)$ una *m. a.* de $U \sim \text{Unif}(a, b)$. Encuentre la distribución de $U_{(n)} = \text{máx}\{U_1, \ldots, U_n\}$.
6. Sea $\boldsymbol{T} = (T_1, \ldots, T_n)$ una *m. a.* de $T \sim \text{Exp}(\theta)$. Encuentre la distribución de $T_{(1)} = \text{mín}\{T_1, \ldots, T_n\}$.
7. Sea $\boldsymbol{U} = (U_1, \ldots, U_n)$ una *m. a.* de $U \sim \text{Unif}(0, 1)$. Muestre que sus estadísticos de orden siguen distribuciones beta. A saber

$$U_{(k)} \sim \text{Beta}(k, n + 1 - k)$$

Método Delta

8. (Dudewicz and Mishra 1988) Empleando el *Método Delta* encuentre expresiones para aproximar la media y la varianza de $Y = g(X)$, para:
   a) $g(x) = \sqrt{x}$.
   b) $g(x) = \log x$.
   c) $g(x) = \sin^{-1} x$.
9. Considere $U \sim \text{Unif}(0, 1)$ y $V \sim \text{Unif}(1, 2)$.
   a) Sea $X = \sqrt{U}$. Encuentre $\mu_X = \mathbb{E}[X]$ y $\sigma^2_X = \text{var}(X)$. Determine aproximaciones de μ_X y σ^2_X utilizando el método delta.

b) Repita el inciso anterior para $Y = \sqrt{V}$.

c) Determine los errores de aproximación en ambos casos y discuta porque su diferencia.

10. Encuentre una aproximación para la media y la varianza de $Y = \sqrt{X}$, donde $X \sim \text{Po}(\lambda)$.

11. **Método Delta**: Sean X y Y *v. a.*'s con medias y varianzas μ_X, μ_Y, y σ_X^2, σ_Y^2, respectivamente. Sea h una función diferenciable y tal que $Z = h(X, Y)$ es también una *v. a.*. Entonces, por la aproximación por series de Taylor, la media y la varianza de Z pueden aproximarse por:

$$\begin{aligned}\mathbb{E}[Z] &\approx h(\mu) + \frac{1}{2}\sigma_X^2 \frac{\partial^2 h(\mu)}{\partial x^2} + \frac{1}{2}\sigma_Y^2 \frac{\partial^2 h(\mu)}{\partial y^2} + \sigma_{XY} \frac{\partial^2 h(\mu)}{\partial x \partial y} \\ \text{var}(Z) &\approx \sigma_X^2 \left(\frac{h(\mu)}{\partial x}\right)^2 + \sigma_Y^2 \left(\frac{h(\mu)}{\partial y}\right)^2 + 2\sigma_{XY} \left(\frac{h(\mu)}{\partial x}\right)\left(\frac{h(\mu)}{\partial y}\right)\end{aligned}$$

donde $h(\mu) = h(\mu_X, \mu_Y)$ y $\sigma_{XY} = \text{cov}(X, Y)$.

Los lados aleatorios X y Y de un triángulo rectángulo tienen media x_0 y y_0 respectivamente, la misma varianza σ^2 y covarianza σ_{XY}. El ángulo entre los lados del triángulo es entonces

$$\Theta = \tan^{-1}\left(\frac{Y}{X}\right)$$

Encuentre una aproximación a la media y varianza de Θ.

3. Distribuciones Muestrales

Distribuciones de algunas sumas de variables aleatorias

Utilice la función característica o la función generadora de momentos y el teorema de unicidad para mostrar las siguientes proposiciones.

1. Sean $X_i \sim \text{Bin}(n_i, p)$, $i = 1, 2$, *v. a.* independientes. Entonces $X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p)$.

2. Sean $Y_i \sim \text{BinNeg}(r_i, p)$, $i = 1, 2$, *v. a.* independientes. Entonces $Y_1 + Y_2 \sim \text{BinNeg}(r_1 + r_2, p)$.

3. Sean $N_i \sim \text{Po}(\lambda_i)$, $i = 1, 2$, *v. a.* independientes. Entonces $N_1 + N_2 \sim \text{Po}(\lambda_1 + \lambda_2)$.

4. Sean $X_i \sim \text{Ber}(p)$, $i = 1, \ldots, n$, *v.a.i.i.d.*. Entonces $S_n \equiv X_1 + \cdots + X_n \sim \text{Bin}(n, p)$.

5. Sean $Y_i \sim \text{Geom}(p)$, $i = 1, \ldots, r$, *v.a.i.i.d.*. Entonces $Y_1 + \cdots + Y_r \sim \text{BinNeg}(r, p)$.

6. Sean $X_i \sim \text{N}(\mu, \sigma^2)$, $i = 1, \ldots, n$, *v.a.i.i.d.*. Entonces $\bar{X} \equiv \frac{1}{n}\sum_1^n X_i \sim \text{N}(\mu, \sigma^2/n)$.

7. Sean $W_i \sim \text{Gamma}(\alpha_i, \lambda)$, $i = 1, 2$, *v. a.* independientes. Entonces $W_1 + W_2 \sim \text{Gamma}(\alpha_1 + \alpha_2, \lambda)$.
   (Con esta parametrización, α es el parámetro de forma y λ el parámetro tasa, así $W \sim \text{Gamma}(\alpha, \lambda)$, entonces $\mathbb{E}[X] = \alpha/\lambda$. O bien, si el **parámetro de escala** es β, entones $\beta = 1/\lambda$ y $\mathbb{E}[W] = \alpha\beta$.)

8. Sean $T_i \sim \text{Exp}(\lambda)$, $i = 1, \ldots, m$, *v.a.i.i.d.*. Entonces $E_m \equiv T_1 + \cdots T_m \sim \text{Gamma}(m, \lambda)$, conocida también como distribución m-Erlang de parámetro λ.

9. Sean $U_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, $i = 1, 2$, *v. a.* independientes. Entonces $U_1 + U_2 \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

10. Sean $Z_i \sim \mathrm{N}(0, 1)$, $i = 1, \dots, n$, *v.a.i.i.d.*. Entonces $S_n \equiv \sum_1^n Z_i \sim \mathrm{N}(0, n)$.

11. Sea $Y \sim \mathrm{Gamma}(\alpha = \nu/2, \beta = 2)$. Y se dice que sigue una distribución Ji-cuadrada con ν grados de libertad ($\nu = 1, 2, 3, \dots$) y es denotada por $Y \sim \chi^2_\nu$. Muestre que si $Y_i \sim \chi^2_1$, para $i = 1, \dots, k$ *v.a.i.i.d.*, entonces $S_k = \sum_1^k Y_i \sim \chi^2_k$.

12. Sean $V_i \sim \chi^2_{\nu_i}$, $i = 1, \dots, m$, *v. a.* independientes, entonces $\sum_1^m V_i \sim \chi^2_\nu$, con $\nu = \sum_1^m \nu_i$.

13. Sean $X_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, \dots n$, *v.a.i.i.d.* Entonces

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n$$

Distribuciones de algunas cocientes de variables aleatorias

14. ((Wackerly, Mendenhall III, and Scheaffer 2008) Ej. 7.7) Sean $X_1, \dots, X_m$ y $Y_1, \dots, Y_n$ dos muestras aleatorias (*m. a.*) independientes donde $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ y $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$. Entonces la diferencia de las medias muestrales $\bar{X} - \bar{Y}$ es una combinación lineal de *v. a.*'s normales independientes y por lo tanto se distribuye normalmente.
    a) Encuentre $\mathbb{E}[\bar{X} - \bar{Y}]$.
    b) Encuentre $\mathrm{var}(\bar{X} - \bar{Y})$.
    c) Suponga que $\sigma_1^2 = 2.0$ y $\sigma_2^2 = 2.5$, y que $m = n$. Encuentre los tamaño de muestras para que $(\bar{X} - \bar{Y})$ esté a no más de una unidad de $(\mu_1 - \mu_2)$ con una probabilidad de al menos 0.95.

15. ((Wackerly, Mendenhall III, and Scheaffer 2008) Ej. 7.8) Con referencia al problema anterior, suponga que $\sigma_1^2 = 0.4$ y $\sigma_2^2 = 0.8$. Si las medias son iguales, i.e. $\mu_1 = \mu_2$ y $m = 10 = n$, encuentre la probabilidad de que la media muestral $\bar{X}$ exceda la media muestral $\bar{Y}$ por al menos una unidad.

16. ((Wackerly, Mendenhall III, and Scheaffer 2008) Ej. 7.14) Sea Z una *v. a.* distribuida normal estándar independiente de Y distribuida χ^2_ν. Entonces

$$T = \frac{Z}{\sqrt{Y/\nu}}$$

sigue una distribución t con ν grados de libertad.
    a) Determine $\mathbb{E}[Z]$ y $\mathbb{E}[Z^2]$. (*Sugerencia*: recuerde que para toda *v. a.*, $\mathbb{E}[Z^2] = \mathrm{var}(Z) + \mathbb{E}[Z]^2$.)
    b) Muestre que si $Y \sim \chi^2_\nu$, entonces

$$\mathbb{E}[Y^r] = \frac{\Gamma(\nu/2 + r)}{\Gamma(\nu/2)} 2^r, \qquad \text{para } r > -\nu/2$$

    c) Use los incisos anteriores para mostrar que
        1) $\mathbb{E}[T] = 0$, si $\nu > 1$.
        2) $\mathrm{var}(T) = \nu/(\nu - 2)$, si $\nu > 2$.

17. ((Rice 2007) Ej. 6.7) Muestre que si $T \sim t_1$, entonces $T \sim \mathrm{Cauchy}$.

18. Muestre que si $T_n \sim t_n$, entonces $T_n \xrightarrow{D} Z$, donde $Z \sim \mathrm{N}(0,1)$.

    Sugerencia: Utilice la aproximación de Stirling a la función gamma. A saber,

    $$\Gamma(z) = \sqrt{\frac{2\pi}{z}}\left(\frac{z}{e}\right)^z \Big(1 + O(z^{-1})\Big)$$

    Esto es, para z grande,

    $$\Gamma(z) \approx \sqrt{\frac{2\pi}{z}}\left(\frac{z}{e}\right)^z$$

    que da lugar la fórmula de Stirling: $n! \approx \sqrt{2\pi n} n^n e^{-n}$.

19. ((Rice 2007) Ej. 6.5) Muestre que si $X \sim F_{m,n}$, entonces $X^{-1} \sim F_{n,m}$.

20. ((Rice 2007) Ej. 6.6) Muestre que si $T \sim t_n$, entonces $T^2 \sim F_{1,n}$.

21. ((Wackerly, Mendenhall III, and Scheaffer 2008) Ej. 7.16) Sean W_1 y W_2 *v. a.*'s independientes con $W_i \sim \chi^2_{n_i}$. Entonces

    $$F = \frac{W_1/n_1}{W_2/n_2}$$

    sigue una distribución F con n_1 y n_2 grados de libertad.

    a) Muestre que $\mathbb{E}[F] = n_2/(n_2 - 2)$, si $n_2 > 2$.

    b) Muestre que $\mathrm{var}(F) = \left[2n_2^2(n_1 + n_2 - 2)\right] / \left[n_1(n_2-2)^2(n_2-4)\right]$, si $n_2 > 4$.

22. Sea $W \sim F_{m,n}$, $0 < p < 1$ y denote $F(p;m,n)$ al p-ésimo percentil de la distribución de W. Esto es, $p = \mathbb{P}(W \le F(p;m,n))$. Muestre entonces que

    $$F(p;m,n) = \frac{1}{F(1-p;n,m)}$$

23. ((Rice 2007) Ej. 6.9) Sea $X_1, \ldots, X_n$ una *m. a.* donde $X \sim \mathrm{N}(\mu, \sigma^2)$ y sea S^2 la varianza muestral Esto es,

    $$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

    Encuentre la media y la varianza de S^2.

24. ((Wackerly, Mendenhall III, and Scheaffer 2008) Ej. 7.11) Considere $X_1, \ldots, X_{20}$, una *m. a.* de una población $\mathrm{N}(\mu, 1.4)$ y sea S^2 la varianza muestral de las 20 mediciones.

    a) Encuentre algún b tal que $\mathbb{P}(S^2 \le b) = 0.975$.

    b) Encuentre algún a tal que $\mathbb{P}(a \le S^2) = 0.975$.

    c) Si a y b son de los incisos anteriores, ¿cuánto es $\mathbb{P}(a \le S^2 \le b)$?

25. ((Wackerly, Mendenhall III, and Scheaffer 2008) Ej. 7.18) Considere las varianzas muestrales S_1^2 y S_2^2 del problema 14 con $n_1 = 10$ y $n_2 = 8$.

    a) Encuentre algún b tal que $\mathbb{P}(\frac{S_1^2}{S_2^2} \le b) = 0.95$.

    b) Encuentre algún a tal que $\mathbb{P}(a \le \frac{S_1^2}{S_2^2}) = 0.95$.

    c) Si a y b son de los incisos anteriores, ¿cuánto es $\mathbb{P}(a \le \frac{S_1^2}{S_2^2} \le b)$?

26. ((Wackerly, Mendenhall III, and Scheaffer 2008) Ej. 7.19) Sean $X_1, \ldots, X_5$ una *m. a.* de tamaño 5 de una distribución normal estándar. $\bar{X} = \frac{1}{5}\sum_{i=1}^{5} X_i$, y X_6 de la misma población e independiente de las anteriores.
    a) ¿Cuál es la distribución de $W = \sum_{i=1}^{5} X_i^2$? Justifique.
    b) ¿Cuál es la distribución de $U = \sum_{i=1}^{5}(X_i - \bar{X})^2$? Justifique.
    c) ¿Cuál es la distribución de $W = \sum_{i=1}^{5}(X_i - \bar{X})^2 + X_6^2$? Justifique.
27. ((Wackerly, Mendenhall III, and Scheaffer 2008) Ej. 7.38) Sean $X's$, $\bar{X}$, W y U como en la pregunta anterior.
    a) ¿Cuál es la distribución de $\sqrt{5}X_6/\sqrt{W}$? Justifique.
    b) ¿Cuál es la distribución de $2X_6/\sqrt{U}$? Justifique.
    c) ¿Cuál es la distribución de $2(5\bar{X}^2 + X_6^2)/U$? Justifique.
28. ((Wackerly, Mendenhall III, and Scheaffer 2008) Ej. 7.69) Sea $X_1, \ldots, X_{10}$ una *m. a.* de una población $\mathrm{N}(0, \sigma^2)$.
    a) Determine la distribución de $(10)\bar{X}^2/S^2$.
    b) Determine la distribución de $S^2/[(10)\bar{X}^2]$.
    c) Determine la constante c tal que
    $$\mathbb{P}\left(-c \leq \frac{S}{\bar{X}} \leq c\right) = 0.95$$
    y donde el estadístico $S/\bar{X}$ es el coeficiente de variación muestral.
29. Sea U y V dos *v.a.i.i.d.* exponencialmente con media 1. Encuentre
    a) $\mathbb{P}(U/V \leq 1)$.
    b) La constante c tal que $\mathbb{P}(U/V \leq c) = 0.95$.

4. Teoremas Límite

1. ((Ross 2006) Ej. 8.4) Sea $X_1, \ldots, X_{20}$ variables aleatorias independientes con *f. m. p.* distribuida Poisson con media 1.
    a) Utilice la desigualdad de Markov para obtener una cota a $\mathbb{P}\left(\sum_1^{20} X_i > 15\right)$
    b) Aplique el teorema central del límite para aproximar $\mathbb{P}\left(\sum_1^{20} X_i > 15\right)$
2. ((Ross 2006) Ej. 8.7) Una persona tiene 100 bombillas cuyo tiempo de vida se distribuye exponencialmente y de manera independiente con media 5 horas. Si las bombillas son usadas una por una, reemplazando inmediatamente una bombilla cada que otra falla, aproxime la probabilidad de que al menos una bombilla siga funcionando después de 525 horas.
3. ((Ross 2006) Ej. 8.9) Si X es una variable aleatoria distribuida Gamma con parámetros $(n, 1)$, aproximadamente que tan grande necesita ser n para que se cumpla que $\mathbb{P}\left(\left|\frac{X}{n} - 1\right| > .01\right) < .01$.
4. ((Ross 2006) Ej. 8.13) Las calificaciones de los exámenes de cierto profesor tienen media 74 y desviación estándar 14. Si el profesor va a entregar dos exámenes, uno a un grupo con 25 alumnos y otro a una con 64.

a) Aproxime la probabilidad de que el promedio de las calificaciones del grupo de 25 sea mayor a 80.

b) Aproxime la probabilidad de que el promedio de las calificaciones del grupo grande sea mayor que las del grupo chico por más de 2.2 puntos.

5. ((Ross 2006) Self Test 8.11) En una colección de 40 pilas cada pila es igualmente probable de ser del tipo A o del tipo B. El tipo A tienen una duración media es de 50 y desviación estándar de 15, el tipo B tiene una duración media es de 30 y desviación estándar de 6.

   a) Aproxime la probabilidad de que la vida total de las pilas sea mayor a 1700.

   b) Suponga que se sabe que 20 de las pilas son del tipo A y 20 del tipo B. Ahora aproxime la probabilidad de que la vida total de las pilas sea mayor a 1700.

6. ((Hoel, Port, and Stone 1971) EJ. 7.14) Se toma una muestra de tamaño n para determinar el porcentaje de personas que planean votar por un partido en la siguientes elecciones. Sea $X_k = 1$ si la k-ésima persona de la muestra que planea votar por ese cierto partido y $X_k = 0$ de cualquier otra manera. Suponga que $X_1, \dots, X_n$ son *v.a.i.i.d.* tal que $\mathbb{P}(X_i = 1) = p$, y $\mathbb{P}(X_i = 0) = 1 - p$. Entonces $\mu = p$ y $\sigma^2 = p(1-p)$, suponga también que p se acerca lo suficiente a 0.5 de tal forma que $\sigma = \sqrt{p(1-p)}$ puede ser aproximada de manera correcta por $\sigma \approx 1/2$. La variable aleatoria S_n/n denota la fracción de las personas en la muestra que planean votar por ese partido y puede usarse para estimar la verdadera probabilidad p. Utilice la aproximación normal para resolver:

   a) Suponga que $n = 900$. Encuentre la probabilidad de que $\left|\frac{S_n}{n} - p\right| \geq .025$

   b) Suponga que $n = 900$. Encuentre la c tal que $\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq c\right) = .01$

   c) Encuentre n tal que $\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq 0.025\right) = .01$

7. ((Hoel, Port, and Stone 1971) Ej. 7.38) Una moneda justa se lanza hasta que caen 100 soles. Encuentre la probabilidad de que al menos 226 tiros sean necesarios.

8. ((Harris 1966) Ej. 7.29) Considere $\{X_n\}$ una sucesión de ensayos Bernoulli con probabilidad de éxito p. Esto es, $\mathbb{P}(X_i = 1) = p$ y $\mathbb{P}(X_i = 0) = 1 - p$. Se define la proporción empírica $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Encuentre el menor entero para el cual $\mathbb{P}(|\hat{p}_n - p| \leq 0.01) \geq 0.95$. (Use la aproximación normal.)

9. ((Harris 1966) Ej. 7.49) Considere la proporción empírica $\hat{p}_n$. Si $p = 0.5$, utilice la desigualdad de Chebyshev para que determinar el mínimo entero n para el cual $\mathbb{P}(0.4 < \hat{p}_n < 0.6) > 0.9$.

10. ((Harris 1966) Ej. 7.50) Suponga que $p = 0.2$. Encuentre n tal que $\mathbb{P}(|\hat{p}_n - 0.2| > 0.01) < 0.05$, usando la desigualdad de Chebyshev y compárela con aquella obtenida mediante el teorema central de límite.

11. ((Harris 1966) Ej. 7.51) Encuentre n tal que $\mathbb{P}(|\hat{p}_n - p| > 0.01) < 0.05$, usando la desigualdad de Chebyshev y compárela con aquella obtenida mediante el teorema central de límite. Compare el resultado con el problema anterior.

12. ((Harris 1966) Ej. 7.52) Sea $X_1, \dots, X_n$ una *m. a.* de X con *f. d. p.* dada por $f(x) = xe^{-x}\mathbb{1}_{(0,\infty)}(x)$. Usando la desigualdad de Chebyshev y el teorema central de límite encuentre n tal que $\mathbb{P}(|\bar{X} - \mu| < 0.10) \geq 0.95$.

13. ((Harris 1966) Ej. 7.1) Usando funciones características muestre que la sucesión de distribuciones binomiales con parámetros n y p_n, tal que $np_n \to \lambda$ mientras que $n \to \infty$, converge en distribución a la distribución Poisson parámetro λ.

14. ((Harris 1966) Ej. 7.6) Sea $X_1, \dots, X_n$ una muestra aleatoria de una distribución uniforme en $[0, a)$. Sea $Y_n = \min\{X_1, \dots, X_n\}$, y sea $W_n = nY_n$. Muestre que Y_n y W_n convergen en distribución y encuentre las distribuciones límite.

15. ((Rice 2007) Ej. 5.28) Sea $\{f_n\}$ una sucesión de *f. m. p.* definidas por $f_n(x) = \frac{1}{2}\mathbb{1}_{\{\pm 1/2^n\}}(x)$, para $n = 1, 2, \dots$. Muestre que $\lim_{n\to\infty} f_n(x) = \frac{1}{2}\mathbb{1}_{\{0\}}(x)$, lo que significa que las *f. m. p.* f_n no convergen a una *f. m. p.*, pero que existe una *f. p. a.* F tal que $\lim_{x\to\infty} F_n(x) = F(x)$.

16. ((Rice 2007) Ej. 5.29) Sean $U_1, U_2, \dots$ *v.a.i.i.d.* uniformemente en $[0, 1]$. Sea $U_{(n)} = \max\{U_1, \dots, U_n\}$. Encuentre la *f. p. a.* de $U_{(n)}$. Sea ahora $V_n = n(1 - U_{(n)})$ y encuentre la correspondiente distribución límite.

17. ((Harris 1966) Ej. 7.16) Sea $X_1, \dots, X_n$ una muestra aleatoria de una distribución Cauchy. Muestre que $\bar{X}_n$ no tiene una distribución asintótica normal. Encuentre la distribución límite. (Considere la función característica de la distribución Cauchy(0,1) dada por $\varphi(t) = e^{-|t|}$.)

18. ((Knight 2000) Ej. 3.3) Sean $X_1, X_2, \dots$ *v.a.i.i.d.* exponencialmente con parámetro tasa λ y sea $X_{(n)} = \max\{X_1, \dots, X_n\}$. Muestre que $X_{(n)} - \ln(n)/\lambda \xrightarrow[n\to\infty]{D} V$, donde V es una *v. a.* tal que para todo v se tiene que

$$\mathbb{P}(V \leq v) = \exp\{-e^{-\lambda v}\}$$

19. ((Knight 2000) Ej. 3.4) Sean $Y_1, Y_2, \dots$ *v.a.i.i.d.* no negativas con *f. p. a.* F. Sea $W_n = \min\{Y_1, \dots, Y_n\}$.

    a) Suponga que $F(y)/y \xrightarrow[x\to 0]{} \lambda$. Muestre que $nW_n \to Y \sim \text{Exp}(\lambda)$.

    b) Suponga que $F(y)/y^\alpha \xrightarrow[x\to 0]{} \lambda$, para algún $\alpha >$. Encuentre la distribución límite de $n^{1/\alpha}W_n$.

5. Estimadores Puntuales

1. ((Rice 2007) Ex. 8.17) Suponga que $X_1, \dots, X_n$ es una *m. a.* de variables en distribuidas en $[0, 1]$, con la siguiente *f. d. p.*:

$$f(x; \alpha) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}\left[x(1-x)\right]^{\alpha-1}$$

donde $\alpha > 1$ es el parámetro a ser determinado de la muestra. (*Note* que la distribución es una Beta de parámetros $\theta_1 = \alpha = \theta_2$.) Se tiene entonces que

$$\begin{aligned} \mathbb{E}[X] &= 1/2 \\ \text{var}(X) &= \frac{1}{4(2\alpha+1)} \end{aligned}$$

   a) ¿Cómo se podría utilizar el método de momentos para estimar el parámetro α?

   b) ¿Cuál sería la ecuación que satisface $\hat{\alpha}$, EMV?

   c) Encuentre la varianza asintótica del EMV.

  d) Encuentre un estadístico suficiente para α.

2. Suponga $X_1, \ldots, X_n$, una *m. a.* de una población modelada mediante la *f. d. p.* dada por

$$f(x;\theta) = e^{-(x-\theta)}, \quad x \geq \theta$$

  a) Encuentre $\tilde{\theta}$, EMM.

  b) Encuentre $\hat{\theta}$, EMV.

  c) Encuentre un estadístico suficiente para θ.

3. Una compañía ha manufacturado ciertos artículos y les ha imprimido su número de serie. Los números de serie comienzan en 1 y terminan en N, donde N es el número de artículos manufacturados. Uno de estos artículos es elegido al azar y éste tiene el número de serie 888. ¿Cuál es el estimador de momentos? ¿Cuál es la estimación de máxima verosimilitud?

4. Suponga que cierta componente electrónica tiene un tiempo de vida T descrito por la distribución exponencial con *f. d. p.* con un tiempo medio de vida τ. Cinco nuevas componentes se ponen a prueba y la primera en fallar lo hace a los 100 días. No se registran más observaciones.

  a) ¿Cuál es la función de verosimilitud para τ?.

  b) ¿Cuál es el EMV de τ?

  c) ¿Cuál es la distribución muestral del EMV?

  d) ¿Cuál es el error estándar del estimador EMV?

  Sugerencia: ¿Cómo se distribuye $T_{(1)} = \min\{T_i\}$?.

5. ((Rice 2007) Ex. 8.26) En un esfuerzo para determinar el tamaño de una población animal, 100 animales fueron capturados y marcados. Tiempo después otros 50 animales fueron capturados de los cuales 20 de ellos ya estaban marcados. ¿Cómo estimaría N, el tamaño de la población? ¿Qué supuestos sobre el proceso de captura/recaptura debe considerar?

  (Para la estimación de N, el tamaño de la población, considere tanto el método de momentos como el de máxima verosimilitud. Para éste último note que el parámetro N es un entero positivo, luego analice el cociente $L(N)/L(N-1), N \in \mathbb{N}_0$, para determinar el máximo.)

6. La distribución de Pareto es usada en actuaría para modelar la *severidad* de los siniestros. A saber,

$$f(x;x_0,\theta) = \theta x_0^{\theta} x^{-(\theta+1)}, \qquad x \geq x_0;\ \theta > 1$$

  Suponga que $x_0 > 0$ y que $X_1, \ldots, X_n$ es una *m. a.* de tal distribución.

  a) Encuentre el estimador de momentos (EMM) de θ.

  b) Encuentre el estimador de máxima verosimilitud (EMV) de θ.

  c) Encuentre la varianza asintótica del EMV.

  d) Encuentre un estadístico suficiente para θ.

7. Considere el siguiente método para estimar λ de una distribución de Poisson.

Observe que $p_0 = P(X = 0) = e^{-\lambda}$. Luego, denotando por Y_n el número de ceros en una muestra aleatoria de tamaño n, λ podría estimarse mediante

$$\tilde{\lambda} = -\log\left(\frac{Y_n}{n}\right)$$

Use el *método de propagación del error* (*método delta*) para obtener aproximaciones a la varianza y sesgo del estimador. Compare la varianza de $\tilde{\lambda}$ con la varianza del estimador $\hat{\lambda}$ EMV, calculando la *eficiencia relativa* para varios valores de λ. Note que $Y_n \sim \text{Bin}(n, p_0)$.

8. Considere la *distribución doble exponencial o Laplaciana* con parámetro θ. Su *f. d. p.* está dada por
$$f(x;\theta) = \frac{1}{2}e^{-|x-\theta|}, \qquad -\infty < x < \infty$$
Para una *m. a.* de tamaño $n = 2m + 1$, muestre que el EMV de θ es la mediana de la muestra, $\tilde{X} = \hat{\theta} = X_{(m+1)}$, el $(m + 1)$-ésimo estadístico de orden.

9. Sea $X_1, \ldots, X_n$ una *m. a.* de una distribución uniforme en $[0, \theta]$.
   a) Encuentre $\tilde{\theta}$, el EMM de θ, su media y su varianza.
   b) Encuentre $\hat{\theta}$, el EMV de θ.
   c) Encuentre la *f. d. p.* de $\hat{\theta}$. Calcule su media y varianza.
   d) Compare las varianzas, sesgo y ECM de los estimadores EMM y EMV.

10. Considere el problema de estimación de la varianza de una distribución normal con media desconocida a partir de una *m. a.* $X_1, \ldots, X_n$.
    a) ¿Cuál de los siguientes estimadores es insesgado?
    $$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad\qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$
    b) ¿Cuál de los estimadores tiene el menor error cuadrático medio (ECM)?
    c) ¿Para qué valor de ρ, el estimador $\rho\sum_{i=1}^{n}(X_i - \bar{X})^2$ tiene un ECM mínimo?

11. Sea $X_1, \ldots, X_n$ una *m. a.* de una población X distribuida exponencialmente con $\mathbb{E}[X] = \tau$.
    a) Encuentre $\hat{\tau}$, el EMV de τ.
    b) ¿Cuál es la distribución exacta de $\hat{\tau}$?
    c) Use el *TCL* para encontrar una aproximación normal a la distribución muestral de $\hat{\tau}$.
    d) Muestre que $\hat{\tau}$ es insesgado y encuentre su varianza exacta.
    e) ¿Hay algún otro estimador insesgado con menor varianza?

12. Sea $X = (X_1, \ldots, X_n)'$ una *m. a.* de una población Poisson con media λ, y sea $T_n = \sum_{i=1}^{n} X_i$.
    a) Muestre que la distribución conjunta de X dado T_n es independiente de λ, y por lo tanto T_n es un estadístico suficiente para λ.
    b) Muestre que X_1 no es un estadístico suficiente.

c) Utilice el *teorema de factorización* para mostrar que T_n es un estadístico suficiente. Identifique las funciones g y h.

13. Utilice el teorema de factorización para encontrar un estadístico suficiente de la distribución exponencial.

14. Sea $X = (X_1, \ldots, X_n)'$ una *m. a.* de una población con *f. d. p.*

$$f(x;\theta) = \frac{\theta}{(1+x)^{\theta+1}}, \qquad 0 < \theta < \infty,\ 0 \le x < \infty$$

Encuentre un estadístico suficiente para θ.

15. Muestre que $\prod_{i=1}^n X_i$, y $\sum_{i=1}^n X_i$, son estadísticos suficientes conjuntos para la distribución gamma.

16. Encuentre un estadístico suficiente para la *densidad Rayleigh* con parámetro θ,

$$f(x;\theta) = \frac{x}{\theta^2}\, e^{-x^2/2\theta^2}, \qquad x \ge 0$$

17. Muestre que las distribuciones Poisson, geométrica y binomial negativa son miembros de la familia exponencial de un parámetro.

18. Muestre que las distribuciones $\mathrm{N}(\mu, \sigma^2)$ y $\mathrm{Ga}(\alpha, \beta)$ son miembros de la familia exponencial de 2 parámetros y encuentre correspondientes estadísticos suficientes.

19. Considere $\boldsymbol{X}_n = (X_1, \ldots, X_n)$, una *m. a.* de $X \sim \mathrm{N})\theta, \theta^2)$, $\theta \in \Theta = \mathbb{R}^+$. Muestre que $f(x;\theta)$ es miembro de la familia exponencial de dos parámetros a pesar de que el espacio parametral Θ es unidimensional.

6. Estimación por Intervalos

1. Considere X y Y poblaciones independientes y sean $\boldsymbol{X}_{n_1} = (X_1, \ldots, X_{n_1})$ una muestra aleatoria de $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ y $\boldsymbol{Y}_{n_2} = (Y_1, \ldots, Y_{n_2})$ una muestra aleatoria de $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$ con las σ_i desconocidas. Suponga que $\sigma_2 = k\sigma_1$, con $k > 0$ conocida. Construya un intervalo del $100(1-\alpha)\,\%$ de confianza para la diferencia de medias $\theta = \mu_1 - \mu_2$.

2. ((Hogg and Craig 1978), Ex. 6.21) Sean $X_1, X_2, \ldots, X_n, X_{n+1}$ una muestra aleatoria de tamaño $n+1, n > 1$, de una distribución $\mathrm{N}(\mu, \sigma^2)$. Sea $\bar{X} = \sum_1^n X_i/n$ y $S^2 = \sum_1^n (X_i - \bar{X})^2/n$. Encuentre la constante c tal que el estadístico $c\left(\bar{X} - X_{n+1}\right)/S$ tenga una distribución t-Student.

   Si $n = 8$, determine k tal que $\mathbb{P}\left(\bar{X} - kS < X_9 < \bar{X} + kS\right) = 0.80$. Al intervalo observado $(\bar{x} - ks, \bar{x} + ks)$ se le llama un intervalo de predicción del 80 por ciento para X_9.

3. ((Hogg and Craig 1978), Ex. 6.25) Sea $\bar{X}$ la media de una muestra aleatoria de tamaño 25 de una distribución Gamma con $\alpha = 4$ y $\beta > 0$. Use el teorema central del límite para encontrar un intervalo de confianza aproximado de 0.954 para μ (la media de la distribución Gamma). *Sugerencia*: Construya su intervalo de confianza a partir de la variable aleatoria

$$\frac{\bar{X} - 4\beta}{\sqrt{4\beta^2/25}} = \frac{5\bar{X}}{2\beta} - 10$$

4. Discuta el problema de encontrar un intervalo de confianza para $\mu_1 - \mu_2$, que representan las medias de dos distribuciones normales independientes, con varianzas σ_1^2 y σ_2^2 conocidas aunque no necesariamente iguales.

5. Discuta el ejercicio anterior cuando asumimos que las varianzas son desconocidas, pero no son iguales. Este problema es realmente complejo y la discusión debe ir orientada a explicar dónde radica la dificultad del ejercicio (Problema de Behrens-Fisher). Por otra parte, si las varianzas son desconocidas, pero σ_1^2/σ_2^2 es una constante k conocida, entonces una estadística que sea una variable aleatoria distribuida t-Student puede ser usada. ¿Por qué?

6. ((Hogg and Craig 1978), Ex. 6.33) Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución $\mathrm{N}(\mu, \sigma^2)$. Sean $0 < a < b$. Muestre que la esperanza matemática de la longitud del intervalo aleatorio

$$\left(\sum_1^n \frac{(X_i - \mu)\,2}{b}, \sum_1^n \frac{(X_i - \mu)\,2}{a} \right)$$

es $n\sigma^2(b-a)/ab$.

7. ((Hogg and Craig 1978), Ex. 6.35) Sea S^2 la varianza de una muestra aleatoria de tamaño n tomada de una distribución $\mathrm{N}(\mu, \sigma^2)$, con μ y σ^2 desconocidas. Sea $g(z)$ la *f. d. p.* de $Z = nS^2/\sigma^2$, cuya distribución es $\chi^2(n-1)$. Sean a y b tales que el intervalo observado $\left(ns^2/b, ns^2/a\right)$ es un intervalo de confianza de 95 % para σ^2. Si su longitud $ns^2(b-a)/ab$ queremos que sea mínimo, muestre que a y b deben satisfacer la condición de $a^2 g(a) = b^2 g(b)$. [*Sugerencia*: Si $G(z)$ es la función de distribución de Z, entonces derive $G(b) - G(a) = 0.95$ y $(b-a)/ab$ con respecto a b y por la primer ecuación, a debe ser función de b. Posteriormente iguale la derivada a cero. Vea Mood et al. (1974) Capítulo VIII, sección 3.2]

8. ((Rice 2007) 8.60) Sea $X_1, X_2, \ldots, X_n$ una muestra *i.i.d.* de una distribución exponencial con *f. d. p.*:

$$f(x;\tau) = \frac{1}{\tau} e^{-x/\tau}, \quad 0 \le x < \infty$$

   a) Encuentre el EMV de τ.

   b) ¿Cuál es la distribución muestral de dicho EMV?

   c) Use el teorema central del límite para encontrar una aproximación normal a esa distribución muestral.

   d) Muestre que el EMV es insesgado y encuentre su varianza explícitamente.

   e) ¿Existe algún otro estimador insesgado con una varianza menor?

   f) Encuentre la forma de un intervalo de confianza aproximado para τ.

   g) Encuentre la forma de un intervalo de confianza exacto para τ.

9. Sean $X_1, \ldots, X_m$ una *m. a.* de $X \sim \mathrm{N}(\mu_X, \sigma_X^2)$ independiente de $Y_1 \ldots, Y_n$, *m. a.* de $Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$. Construya un intervalo de confianza para el parámetro $\theta = \sigma_Y^2/\sigma_X^2$ y para el parámetro $\tau = \sqrt{\theta}$.

10. Sean $X_1 \sim \mathrm{Bin}(n_1, p_1)$ y $X_2 \sim \mathrm{Bin}(n_2, p_2)$ independientes. Construya un intervalo de aproximadamente $100(1-\alpha)$ % de confianza para $\theta = p_2 - p_1$ suponiendo n_1 y n_2 conocidas.

7. Pruebas de Hipótesis

Problemas seleccionados de Hogg and Craig (1978), Mood, Graybill, and Boes (1974), Rice (2007).

1. Sea $X \sim f(x:\theta) = \frac{1}{\theta} e^{-x/\theta} \mathbb{1}_{(0,\infty)}(x)$. Considere la hipótesis nula $\mathrm{H}_0 : \theta = \theta_0 = 2$ y la hipótesis alternativa $\mathrm{H}_1 : \theta = \theta_1 = 4$. Sea X_1, X_2 una *m. a.* de tamaño 2 de esta distribución. Muestre que la mejor prueba de H_0 contra H_1 se puede llevar a cabo mediante el estadístico $X_1 + X_2$. Muestre que la significancia de la prueba basada en la región critica $\mathcal{R}_\alpha = \{X_1 + X_2 > 9.5\}$ es de $\alpha = 0.05$.

2. Repita el problema anterior para $\mathrm{H}_1 : \theta = \theta_1 = 6$. Generalize el resultado para todo $\mathrm{H}_1 : \theta = \theta_1 > 2$.

3. Sea $X_1 \ldots, X_n$ una *m. a.* de una distribución con *f. d. p.* $f(x;\delta) = \delta x^{\delta-1} \mathbb{1}_{(0,1)}(x)$. Muestre que una mejor región crítica para probar $\mathrm{H}_0 : \delta = 1$ *vs.* $\mathrm{H}_1 : \delta = 2$ es $\mathcal{R} = \{(x_1, \ldots, x_n) : c \leq \prod_{i=1}^n x_i\}$.

4. Sea $X_1, \ldots, X_{10}$ *m. a.* de $\mathrm{N}(\mu, \sigma^2)$. Encuentre la mejor prueba para contrastar la hipótesis nula simple $\mathrm{H}_0 : \mu = 0,\ \sigma^2 = 1$, contra la hipótesis alternativa $\mathrm{H}_1 : \mu = 1,\ \sigma^2 = 4$.

5. Sea $\boldsymbol{X} = (X_1, \ldots, X_n)$ *m. a.* de $\mathrm{N}(\mu, 100)$. Muestre que $\mathcal{R} = \{(X_1, \ldots, X_n) : c \leq \bar{X}\}$ es una mejor región crítica para probar $\mathrm{H}_0 : \mu = 75$ *vs.* $\mathrm{H}_1 : \mu = 78$. Encuentre n y c tal que $\mathbb{P}(\boldsymbol{X} \in \mathcal{R} | H_0) = 0.05$ y $\mathbb{P}(\boldsymbol{X} \in \mathcal{R} | H_a) \approx 0.90$.

6. Sea $X_1, \ldots, X_n$ *m. a.* de una población con *f. d. p.* dada por $f(x;p) = p^x(1-p)^{1-x} \mathbb{1}_{\{0,1\}}(x)$. Muestre que la región crítica para probar $\mathrm{H}_0 : p = 1/2$ *vs.* $\mathrm{H}_1 : p = 1/3$. Utilice el teorema central del límite para encontrar n y c de manera que aproximadamente se tenga que $\mathbb{P}(\sum X_i \leq c | H_0) = 0.10$ y $\mathbb{P}(\sum X_i \leq c | H_1) = 0.80$.

7. Sea $X \sim \mathrm{Unif}(0, \theta)$, $\theta > 0$. Sean $Y_1 < Y_2 < Y_3 < Y_4$ los estadísticos de orden de una *m. a.* de tamaño 4 de X. Sea y_4 la observación de Y_4. Se rechaza $\mathrm{H}_0 : \theta = 1$ en favor de $\mathrm{H}_1 : \theta \neq 1$ si $y_4 < 1/2$, o bien, $y_4 \geq 1$. Encuentre la potencia de la prueba $K(\theta)$, para $0 < \theta$.

8. Considere dos distribuciones independientes $\mathrm{N}(\mu_1, 400)$ y $\mathrm{N}(\mu_2, 225)$. Sea $\theta = \mu_1 - \mu_2$. Sea $\bar{x}$ y $\bar{y}$ las medias observadas de dos *m. a.*'s independientes, ambas de tamaño n, de estas dos distribuciones respectivamente. Se rechaza $\mathrm{H}_0 : \theta = 0$ en favor de $\mathrm{H}_1 : \theta > 0$ si y solo si $\bar{x} - \bar{y} \geq c$. Si $K(\theta)$ es la potencia de la prueba. Encuentre n y c de manera que $K(0) = 0.05$ y $K(10) = 0.90$ aproximadamente.

9. Sea $X \sim \mathrm{N}(\mu, \sigma^2)$. Se desea probar $\mathrm{H}_0 : \mu = 0,\ \sigma^2 > 0$ *vs.* $\mathrm{H}_1 : \mu \neq 0,\ \sigma^2 > 0$. Considere la razón de verosimilitud generalizada. Si $n = 10$ y los valores experimentales llevan a que $\bar{x} = 0.6$ y $\sum_1^{10}(x_i - \bar{x}) = 3.6$. ¿Rechazaría o no H_0 a un nivel de significancia de 5 %?

10. Sean $X_1, \ldots, X_n$ y $Y_1, \ldots, Y_m$ *m. a.*'s de distribuciones independientes $\mathrm{N}(\mu_X, \sigma_X^2)$ y $\mathrm{N}(\mu_Y, \sigma_Y^2)$, respectivamente.

    a) Muestre que la razón de verosimilitud para probar $\mathrm{H}_0 : \mu_X = \mu_Y, \sigma_X^2 = \sigma_Y^2$ contra todas las alternativas está dado por

    $$\mathrm{CV} = \frac{\left[\sum_1^n (x_i - \bar{x})^2/n\right]^{n/2} \left[\sum_1^m (y_i - \bar{y})^2/m\right]^{m/2}}{\left\{\frac{1}{n+m}\left[\sum_1^n (x_i - u)^2 + \sum_1^m (y_i - u)^2\right]\right\}^{(n+m)/2}}$$

    donde $u = (n\bar{x} + m\bar{y})/(n+m)$.

b) Muestre que la prueba de razón de verosimilitud para contrastar $\mathrm{H}_0 : \sigma_X^2 = \sigma_Y^2$ contra $\mathrm{H}_1 : \sigma_X^2 \neq \sigma_Y^2$ sin especificar μ_X y μ_Y se puede basar en el estadístico

$$F = \frac{\sum_1^n (X_i - \bar{X})^2/(n-1)}{\sum_1^m (Y_i - \bar{Y})^2/(m-1)}$$

11. Considere n ensayos independientes tales que $X_1, \ldots, X_k$ son el número de veces que el experimento terminó en las clases mutuamente exclusivas y exhaustivas $A_1, \ldots, A_k$ respectivamente. Si $p_i = \mathbb{P}(A_i)$ es constante a través de los n ensayos, entonces la probabilidad de una secuencia particular $L = p_1^{x_1} \cdots p_k^{x_k}$.

a) Recordando que $p_1 + \cdots + p_k = 1$, muestre que la razón de verosimilitud para probar que $\mathrm{H}_0 : p_i = p_{i0} > 0$, $i = 1, \ldots, k$ contra todas las alternativas está dada por:

$$\Lambda = \prod_{i=1}^{k} \left(\frac{np_{i0}}{x_i} \right)^{x_i}.$$

b) Muestre que

$$-2 \ln \Lambda = \sum_{i=1}^{k} \frac{x_i(x_i - np_{i0})^2}{(np_i')^2}$$

donde p_i' está entre p_{i0} y x_i/n. (*Sugerencia*: Expanda el $\ln p_{i0}$ en Series de Taylor con el término remanente involucrando $(p_{i0} - x_i/n)^2$.

c) Para n grande, argumente que $x_i/(np_i')^2$ es aproximado por $1/(np_{i0})$ y por lo tanto

$$-2 \ln \Lambda \approx \sum_{i=1}^{k} \frac{(x_i - np_{i0})^2}{np_{i0}}$$

12. Sea X *una* observación que viene de la *f. d. p.* $f(x; \theta) = (2\theta x + 1 - \theta)\mathbb{1}_{(0,1)}(x)$, con $-1 \leq \theta \leq 1$.

a) Encuentre la prueba más potente de tamaño α para probar $\mathrm{H}_0 : \theta = 0$, *vs.* $\mathrm{H}_1 : \theta = 1$. (Su prueba debe quedar en términos de α.)

b) Para probar $\mathrm{H}_0 : \theta \leq 0$ *vs.* $\mathrm{H}_1 : \theta > 0$, se usó el siguiente criterio: Rechace H_0 si $X > 1/2$. Determine la significancia y la potencia de la prueba.

c) ¿Hay alguna prueba uniformemente más potente para $\mathrm{H}_0 : \theta \leq 0$ *vs.* $\mathrm{H}_1 : \theta > 0$? Si es así, ¿cuál es ésta?

d) ¿Cuál es la prueba de razón de verosimilitud generalizada para probar $\mathrm{H}_0 : \theta = 0$ *vs.* $\mathrm{H}_1 : \theta \neq 0$?

e) Entre todas la posibles pruebas de razón de verosimilitud simples para probar $\mathrm{H}_0 : \theta = 0$ *vs.* $\mathrm{H}_1 : \theta = 1$?, encuentre aquella que minimiza $\alpha + \beta$, donde α y β son los errores de tipo I y II respectivamente.

13. Sea $X_1, \ldots, X_n$ una *m. a.* de una distribución Poisson con *f. d. p.*

$$f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \mathbb{1}_{\{0,1,\ldots\}}(x)$$

a) Encuentre la prueba uniformemente más potente para probar $\mathrm{H}_0 : \lambda = \lambda_0$ *vs.* $\mathrm{H}_1 : \lambda > \lambda_0$, y bosqueje la función potencia de la prueba para $\lambda_0 = 1$ y $n = 25$. Utilice el teorema central del límite y considere $\alpha = 0.05$

b) Pruebe $\mathrm{H}_0 : \lambda = \lambda_0$ *vs.* $\mathrm{H}_1 : \lambda \neq \lambda_0$. Encuentre la forma de la región crítica después de usar el principio de razón de verosimilitud. (Recuerde, la región debe quedar en términos del estadístico $\sum X_i$.)

c) Una prueba razonable para $\mathrm{H}_0 : \lambda = \lambda_0$ *vs.* $\mathrm{H}_1 : \lambda \neq \lambda_0$ debe ser la siguiente: Rechace H_0 si $|\bar{X} - \lambda_0| \geq k$. Para $\alpha = 0.05$, encuentre k tal que $\mathbb{P}(\text{Rechazar}\mathrm{H}_0|\mathrm{H}_0) = 0.05$. (Suponga n suficientemente grande de manera que el teorema central de limite se puede utilizar para aproximar el valor de k.)

14. Sea $X_1, \ldots, X_n$ *m. a.* de una distribución uniforme sobre $(\theta, \theta + 1)$. Para probar $\mathrm{H}_0 : \theta = 0$ *vs.* $\mathrm{H}_1 : \theta > 0$ se usó la siguiente prueba: Rechace H_0 si y solo si $X_{(1)} \geq k$ o $X_{(n)} \geq 1$ y donde k es una constante.

    a) Determine k de manera que la prueba sea de tamaño α.

    b) Encuentre la potencia de la prueba que encontró el inciso anterior.

    c) Confirme o contradiga lo siguiente: si se elige k de manera de que la prueba tenga significancia α, entonces la prueba misma es UMP de tamaño α.

15. Suponga que $X_1 \ldots, X_n$ es una *m. a.* de la densidad $f(x; \theta)$, para la cual, T es un estadístico suficiente para el parámetro θ. Muestre que la razón de verosimilitud para la prueba de $\mathrm{H}_0 : \theta = \theta_0$ *vs.* $\mathrm{H}_1 : \theta = \theta_1$, es función del estadístico T. Explique cómo, si la distribución de T bajo H_0 es conocida, la región de rechazo se puede elegir de manera que la prueba sea de tamaño α.

16. Sea $X \sim \mathrm{N}(0, \sigma^2)$, y considere la hipótesis $\mathrm{H}_0 : \sigma = \sigma_0$ *vs.* $\mathrm{H}_1 : \sigma = \sigma_1$. Ambos σ's son fijos.

    a) ¿Cuál es la razón de verosimilitud como función de x? ¿Qué valores favorecen H_0? ¿Cuál es la región de rechazo para un tamaño de la prueba α?

    b) Si $X_1, \ldots, X_n$ es una *m. a.* de la distribución anterior. Repita el ejercicio del inciso anterior.

    c) ¿Es la prueba del inciso anterior UMP para probar $\mathrm{H}_0 : \sigma = \sigma_0$ *vs.* $\mathrm{H}_1 : \sigma > \sigma_0$.

17. Sea $X_1, \ldots, X_n$ *m. a.* de una distribución doble exponencial con *f. d. p.* dada por: $f(x; \lambda) = \frac{1}{2}\lambda e^{-\lambda|x|}$. Derive la prueba de razón de verosimilitud para contrastar $\mathrm{H}_0 : \lambda = \lambda_0$ *v. a.* $\mathrm{H}_1 : \lambda = \lambda_1$, donde $\lambda_1 > \lambda_0$, valores especificados. ¿Es la prueba UMP en contra de la hipótesis alternativa $\mathrm{H}_1 : \lambda > \lambda_0$?

18. Considere dos funciones de densidad sobre $[0, 1]$: $f_0(x) = 1$ y $f_1(x) = 2x$. Dentro de todas la pruebas de significancia α de la hipótesis nula $\mathrm{H}_0 : X \sim f_0$ *vs.* $\mathrm{H}_1 : X \sim f_1$, ¿qué tan grande puede ser la potencia?

Referencias

Chihara, L. and T. Hesterberg (2019). *Mathematical Statistics with R* (2 ed.). Hoboken, NJ: Wiley.

Dudewicz, E. J. and S. N. Mishra (1988). *Modern Mathematical Statistics.* New York, N.Y.: Wiley.

Harris, B. (1966). *Theory of Probability.* Reading, MA.: Addison-Wedsley.

Hoel, P. G., S. C. Port, and C. J. Stone (1971). *Introduction to Probability Theory*. Boston: Houghton Miffling Company.

Hogg, R. V. and A. T. Craig (1978). *Introduction to Mathematical Statistics* (4 ed.). New York: Macmillan Publishing Co., Inc.

Knight, K. (2000). *Mathematical Statistics*. Boca Raton, Florida: Chapman & Hall/CRC.

Mood, A. M., F. A. Graybill, and D. C. Boes (1974). *Introduction to the Theory of Statistics* (3rd ed.). Singapore: McGraw-Hill.

Rice, J. S. (2007). *Mathematical Statistics and Data Analysis* (3rd ed.). Belmont, California: Brooks/Cole: Cengage Learning.

Ross, S. (2006). *A First Course in Probability* (7th ed.). Upper Saddle River, N.J.: Prentice Hall.

Tukey, J. (2020). *Exploratory Data Analysis* (1 ed.). Hoboken, N.J.: Pearson Education.

Wackerly, D. D., W. Mendenhall III, and R. L. Scheaffer (2008). *Mathematical Statistics with Applications* (7 ed.). Australia: Thomson.

Respuestas

1. Introducción

1:1. —

2. Suma de Variables Aleatorias y Método Delta

2:1. —

2:2. —

2:3. —

2:4. —

2:5. —

2:6. —

2:7. —

2:8. —

2:9. —

2:10. —

2:11. —

2:12. —

2:13. —

2:14. —

2:15. *a*) $\mathbb{E}[Y] \approx \sqrt{\mu_X}\left(1 - \frac{\sigma_X^2}{8\mu_X^2}\right)$ y $\mathrm{var}(Y) \approx \frac{\sigma_X^2}{4\mu_X}$.

b) $\mathbb{E}[Y] \approx \log \mu_X - \frac{\sigma_X^2}{2\mu_X^2}$ y $\mathrm{var}(Y) \approx \frac{\sigma_X^2}{\mu_X^2}$.

c) $\mathbb{E}[Y] \approx \sin^{-1} \mu_X + \frac{\mu_X \sigma_X^2}{2\left(1-\mu_X^2\right)^{\frac{3}{2}}}$ y $\mathrm{var}(Y) \approx \frac{\sigma_X^2}{1-\mu_X^2}$.

2:16. *a*) $\mu_X = \frac{2}{3}$ y $\sigma_X^2 = \frac{1}{18}$
$\mathbb{E}[X] \approx 0.67$ y $\mathrm{var}(X) \approx 0.0417$

b) $\mu_Y = 1.2189$ y $\sigma_Y^2 = 0.0142$
$\mathbb{E}[Y] \approx 1.2191$ y $\mathrm{var}(Y) \approx 0.0139$

c) —

2:17. $\mathbb{E}[Y] \approx \sqrt{\lambda} - \frac{1}{8\sqrt{\lambda}}$ y $\mathrm{var}(Y) \approx \frac{1}{4}$

2:18. $\mathbb{E}[\Theta] \approx \tan^{-1} \frac{y_0}{x_0} + \sigma_{XY} \frac{y_0^2 - x_0^2}{\left(y_0^2 + x_0^2\right)^2}$
$\mathrm{var}(\Theta) \approx \frac{\sigma^2\left(y_0^2+x_0^2\right) - 2\sigma_{XY} y_0 x_0}{y_0^2 + x_0^2}$

3. Distribuciones Muestrales

3:1. *a*) $\mathbb{E}[\bar{X} - \bar{Y}] = \mu_1 - \mu_2$

b) $\mathrm{Var}(\bar{X} - \bar{Y}) = \frac{\sigma_1}{m} + frac\sigma_2 n$

c) $n \geq 18$

3:2. 0.0038

3:3. *a*) $\mathbb{E}[Z^2]$

b) —

c) —

3:4. —

3:5. —

3:6. —

3:7. —

3:8. —

3:9. —

3:10. $\mathbb{E}[S^2] = \sigma^2$, $\mathrm{Var}(S^2) = \frac{2\sigma^4}{n-1}$

3:11. *a*) $b = 2.421$

b) $a = 0.656$

c) 0.95

3:12. *a*) $b = 7.36$

b) $a = 0.608$

c) 0.9

3:13. *a*) $W \sim \chi^2_{(5)}$

b) $U \sim \chi^2_{(4)}$

c) $W \sim \chi^2_{(5)}$

3:14. *a*) $t_{(5)}$

b) $t_{(4)}$

c) $F_{(2,4)}$

3:15. *a*) $F_{(1,9)}$

b) $F_{(9,1)}$

c) $c = 49.04$

3:16. *a*) 1/2

b) $c = 19$

4. Teoremas Límite

4:1. *a*) $\frac{20}{15}$

b) 0.8682

4:2. 0.3085

4:3. $n > 66348$

4:4. *a*) 0.016

b) 0.252

4:5. *a*) 0.3068

b) 0.0831

4:6. *a*) $\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq 0.25\right) = 0.1336$

b) $c = 0.043$

c) $n = 2654$

4:7. 0.033

4:8. $n = 38416pq$

4:9. $n = 250$

4:10. Chebyshev: $n = 32000$ y TCL: $n = 6146$

4:11. Chebyshev: $n = 200000pq$ y TCL: $n = 38416pq$

4:12. Chebyshev: $n = 4000$ y TCL: $n = 768$

4:13. —

4:14. —

4:15. —

4:16. —

4:17. —

4:18. —

4:19. —

5. Estimadores Puntuales

5:1. —

5:2. $\tilde{\theta} = \frac{(n+1)\hat{\theta}}{n}$

5:3. —

5:4. *a*) $\tilde{\theta}_{\text{EMM}} = \frac{5}{12}$

b) $\text{ee}(\tilde{\theta}_{\text{EMM}}) = 0.1728$

c) $\hat{\theta}_{\text{EMV}} = \frac{1}{2}$

d) $\text{ee}(\hat{\theta}_{\text{EMV}}) = 0.1581$

5:5. —

5:6. *a*) $\tilde{p} = \frac{1}{\bar{Y}}$

b) $\hat{p} = \frac{1}{\bar{Y}}$

c) $\operatorname{var}(\hat{p}) = \frac{p^2(1-p)}{n}$

5:7. —

5:8. *a*) $\tilde{\sigma} = \sqrt{\frac{1}{2}\mu_2}$

b) $\hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n}|X_i|$

c) $\operatorname{var}(\hat{\sigma}) = \frac{1}{n}\sigma^2$

d) $\sum_{i=1}^{n}|X_i|$

5:9. *a*) $\tilde{\mu} = \mu_1$ y $\tilde{\sigma} = \sqrt{m_2 - m_1^2}$

b) $\hat{\mu} = \frac{1}{2}\left(X_{(n)} + X_{(1)}\right)$ y $\hat{\sigma} = \frac{1}{2\sqrt{3}}\left(X_{(n)} - X_{(1)}\right)$

5:10. *a*) $\tilde{\alpha} = \frac{2}{16\mu_2 - 1} - \frac{1}{2}$

b) $\frac{2n\Gamma'(2\alpha)}{\Gamma(2\alpha)} - \frac{2n\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum_{i=1}^{n}\log\left(X_i(1-X_i)\right) = 0$

c) $\operatorname{var}(\hat{\alpha}) = \frac{1}{n\left(\frac{\partial^2 l}{\partial\alpha^2}\right)}$

d) $\sum_{i=1}^{n}\log\left(X_i(1-X_i)\right)$

5:11. *a*) $\tilde{\theta} = \bar{X} - 1$

b) $\hat{\theta} = X_{(1)}$

c) $X_{(1)}$

5:12. $\tilde{N} = 1775$ y $\hat{N} = 888$

5:13. *a*) $\frac{1}{\tau}e^{-\frac{500}{\tau}}$

b) $\hat{\tau} = 500$

c) $\hat{\tau} \sim \operatorname{Exp}(\tau)$

d) τ

5:14. —

5:15. *a*) $\tilde{\theta} = \frac{\bar{X}}{\bar{X} - x_0}$

b) $\hat{\theta} = \frac{n}{\sum_{i=1}^{n}\log(X_i) - n\log(x_0)}$

c) $\operatorname{var}(\hat{\theta}) = \frac{\theta^2}{n}$

d) $\sum_{i=1}^{n}\log(X_i)$

5:16. —

5:17. —

5:18. *a*) $\tilde{\theta} = 2\bar{X}$, $\mathbb{E}[\tilde{\theta}] = \theta$ y $\operatorname{var}(\tilde{\theta}) = \frac{\theta^2}{12n}$

b) $\hat{\theta} = X_{(n)}$

c) $f_{\hat{\theta}}(u) = \frac{n}{\theta^n}u^{n-1}\mathbb{I}_{[0,\theta]}$, $\mathbb{E}[\hat{\theta}] = \frac{n}{n+1}\theta$ y $\operatorname{var}(\hat{\theta}) = \frac{n\theta^2}{(n+2)(n+1)^2}$

d) —

5:19. —

5:20. *a*) $\hat{\tau} = \bar{X}$

b) $\hat{\tau} \sim \mathrm{Ga}(n, \frac{\tau}{n})$

c) $\hat{\tau} \sim \mathrm{N}(\tau, \frac{\tau^2}{n})$

d) $\mathrm{var}(\hat{\tau}) = \frac{1}{n\tau^2}$

e) $\hat{\tau}$ es UMVUE

5:21. —

5:22. $\sum_{i=1}^{n} X_i$

5:23. $\sum_{i=1}^{n}(1 + X_i)$

5:24. —

5:25. $\sum_{i=1}^{n} X_i^2$

5:26. —

5:27. —

5:28. —

6. Estimación por intervalos

6:1. $(\bar{X} - \bar{Y} \pm t_{(1-\frac{\alpha}{2};\ n_1+n_2-2)} S_p \sqrt{\frac{1}{n_1} + \frac{k^2}{n_2}})$ con $S_p = \frac{1}{n_1+n_2-2} \frac{k^2(n_1-1)S_1^2+(n_2-1)S_2^2}{k^2}$

6:2. $c = \sqrt{\frac{n}{n+1}}$, $k = 1.5008$

6:3. $(\frac{5}{6}\bar{X}, \frac{5}{4}\bar{X})$

6:4. —

6:5. —

6:6. —

6:7. —

6:8. *a*) $\hat{\tau} = \bar{X}$

b) $\hat{\tau} \sim \mathrm{Ga}(n, \frac{\tau}{n})$

c) $\hat{\tau} \sim \mathrm{N}(\tau, \frac{\tau^2}{n})$

d) $\mathrm{Var}(\hat{\tau}) = \frac{1}{n\tau^2}$

e) $\hat{\tau}$ es UMVUE

f) $(\frac{n\bar{X}}{Z_{1-\frac{\alpha}{2}}+n}, \frac{n\bar{X}}{Z_{\frac{\alpha}{2}}+n})$

g) $(\frac{2n\bar{X}}{\chi_{1-\frac{\alpha}{2};\ 2n}}, \frac{2n\bar{X}}{\chi_{\frac{\alpha}{2};\ 2n}})$

6:9. $(F_{(\frac{\alpha}{2};\ m-1,\ n-1)} \frac{S_Y^2}{S_X^2}, F_{(1-\frac{\alpha}{2};\ m-1,\ n-1)} \frac{S_Y^2}{S_X^2})$

6:10. —

7. Contraste de Hipótesis

7:1. —

7:2. $\mathcal{R}_\alpha = \{X_1 + X_2 > \lambda\}$ es la región de rechazo para la PUMP.

7:3. —

7:4. $\mathcal{R} = \{\underline{X} \in \mathbb{R}^n | 3\sum X_i^2 + 2\sum X_i > c\}$

7:5. $n = 94.74$ y $c = 76.68$

7:6. $n = 38.64$ y $c = 15.34$

7:7. $K(\theta) = \begin{cases} 1 & \text{si } \theta \leq 1/2, \\ \frac{1}{16\theta^4} & \text{si } 1/2 < \theta \leq 1, \\ 1 - \frac{15}{16\theta^4} & \text{si } \theta > 1 \end{cases}$

7:8. $n = 53.29$ y $c = 5.616$

7:9. No se rechaza la hipótesis nula.

7:10. —

7:11. —

7:12. *a*) La mejor prueba es aquella basada en $\mathcal{R}_\alpha = \{X > 1 - \alpha\}$

b) $\alpha = 1/2$, $K(\theta) = \frac{1}{2} + \frac{\theta}{4}$

c) La PUMP se basa en la región $\mathcal{R}_\alpha = \{X > 1 - \alpha\}$

d) $\mathcal{R} = \{|X - \frac{1}{2}| > c\}$

e) $\mathcal{R} = \{X > 1/2\}$

7:13. *a*) La PUMP se basa en el región $\mathcal{R} = \{\sum X_i > c\}$ y $K(\lambda) = 1 - \Phi(\frac{33.2 - 25\lambda}{5\sqrt{\lambda}})$

b) $\mathcal{R} = \{\sum X_i(\ln(n\lambda_0) - \ln(\sum X_i) + 1) < c\}$

c) $k = \frac{1.96\sqrt{\lambda_0}}{\sqrt{n}}$

7:14. *a*) $k = 1 - \alpha^{\frac{1}{n}}$

b) $K(\theta) = \begin{cases} \alpha + 1 - (1-\theta)^n & \text{si } 0 \leq \theta \leq 1, \\ 1 & \text{si } \theta \geq 1, \end{cases}$

c) —

7:15. —

7:16. $\mathcal{R}_\alpha = \{\sum X_i^2 > (\sigma_0^2)\chi^2_{(1-\alpha;\ n)}\}$ y la prueba es UMP.

7:17. $\mathcal{R} = \{\sum |X_i| < c\}$ y la prueba es UMP.

7:18. Fijando α la potencia máxima está dada por $K(\theta) = 2\alpha - \alpha^2$