Probabilidad y Estadística A (61.06 No Ind - 81.03)

Guía de ejercicios - Estadística

Facultad de ingeniería, UBA rev01b - junio de 2023

Glosario de símbolos

en : Alto. Estos ejercicios son importantes, pudiendo ser o no difíciles de resolver. Se recomienda fuertemente resolverlos.

: Curva peligrosa. Estos ejercicios pueden ser más difíciles de lo que parecen a simple vista, o por el contrario, si se miran bien resultan más fáciles de lo que parecen. Ante la duda, consulte a los docentes del curso o en clases de consultas.

 \triangle : En blanco. Se deja así para compatibilizar la numeración de esta guía con la de otras materias.

②: Optativo. Algunos requieren ver temas que no entran estrictamente en el programa (se aclaran entre corchetes); otros tienen un grado excesivo de dificultad. No es necesario hacerlos a fines de aprobar la materia.

🕏 : Ejercicios muy difíciles.

🔂 : Ayuda para resolver el ejercicio.

Notas

Basada en: Probabilidad y Estadística, Guía de Trabajos Prácticos 2ª Parte (Estadística), Primer Cuatrimestre del 2020, Versión 1.3 v01b - Se corrige ejercicio 9.10 inciso a y se agrega ayuda.

9. Guía 9

9.1.

9.2. P [ver teorema de factorización para suficiencia] Sea $\underline{X}^{(n)}$ una muestra aleatoria de tamaño n de una distribución con densidad

$$f_{\theta}(x) = e^{-(x-\theta)} \mathbf{1} \{x > \theta\}.$$

Verificar que $T = \min(X_1, \dots, X_n)$ es un estadístico suficiente para θ .

9.3. \triangle

9.4. P [ver definición de familia exponencial] Mostrar que las siguientes familias de distribuciones son familias exponenciales a 1 parámetro: (a) Bernoulli(p); (b) Pascal(4, p); (c) Poisson(μ); (d) Exponencial(λ).

9.5. Una moneda tiene una probabilidad de cara $p, p \in \{2/5, 4/5\}$. En 10 lanzamientos de la moneda se observaron exactamente 3 caras. Estimar por máxima verosimilitud la probabilidad de que en otros tres lanzamientos se observe exactamente una cara.

9.6. e La cantidad de accidentes de tránsito por semana en la intersección de Paseo Colón y Estados Unidos tiene una distribución de Poisson de media l.

Hallar el estimador de máxima verosimilitud de l basado en una muestra aleatoria de la cantidad de accidentes durante n semanas. Mostrar que se trata de un estimador insesgado para l y hallar la expresión de su error cuadrático medio.

En una muestra de 100 semanas se observaron las frecuencias:

Cantidad de accidentes	0	1	2	3	4	5
Frecuencia	10	29	25	17	13	6

En virtud de la información muestral calcular el estimador de máxima verosimilitud de l y estimar la probabilidad de que en la semana del 24 de diciembre de 2017 no ocurra ningún accidente en la mencionada esquina.

9.7. \bigcirc Sea $\underline{X}^{(n)}$ una muestra aleatoria de tamaño n de la distribución uniforme sobre el intervalo $[0,\theta].$

- (a) Hallar un estadístico suficiente para θ basado en $\underline{X}^{(n)}$.
- (b) Hallar el estimador de máxima verosimilitud de θ basado en $\underline{X}^{(n)}.$
- (c) P Sea $\hat{\theta}_n$ el estimador de máxima verosimilitud de θ hallado en el inciso anterior. Mostrar que $\mathbf{E}_{\theta}[\hat{\theta}_n] = \frac{n}{n+1}\theta$ y $\mathbf{var}_{\theta}(\hat{\theta}_n) = \frac{n\theta^2}{(n+1)^2(n+2)}$ para concluir que $\hat{\theta}_n$ converge en media cuadrática a θ cuando $n \to \infty$.

2

$$f_{\theta}(x) = 3\theta^3 x^{-4} \mathbf{1} \{ x \ge \theta \}, \quad \theta > 0.$$

- (a) Hallar el estimador de máxima verosimilitud de θ basado en una muestra aleatoria de los tamaños de n archivos.
- (b) \bullet Hallar expresiones para la esperanza y la varianza del estimador de máxima verosimilitud de θ .
- (c) \bullet Mostrar que el estimador de máxima verosimilitud de θ converge en media cuadrática al verdadero valor de θ .

9.9. \nearrow La duración, X, en años de ciertos discos rígidos tiene la distribución Pareto con densidad

$$f_{\theta}(x) = \theta x^{-(\theta+1)} \mathbf{1} \{x > 1\}, \quad \theta > 1.$$

- (a) \bigcirc [factorización para suficiencia] Usar el criterio de factorización de Neyman-Fisher para hallar un estadístico suficiente para θ , basado en una muestra aleatoria de la duración de n discos.
- (b) (b) [familia exponencial] Mostrar que las distribuciones f_{θ} , $\theta > 1$, pertenecen a una familia exponencial y usar esa propiedad para hallar un estadístico suficiente para θ . ¿Cuál es su distribución? (c):notar que log $X \sim \text{Exponencial}(\theta)$.
- (c) Hallar el estimador de máxima verosimilitud de θ basado en una muestra aleatoria de la duración de n discos rígidos. Mostrar que se trata de un estimador asintóticamente insesgado y cuya varianza tiende a 0 cuando $n \to \infty$.
- (d) \bigcirc [distribución asintótica de estadísticos máximo verosímil, número de información de Fisher] Hallar la distribución asintótica del estimador de máxima verosimilitud de θ . 5: es fácil ver que $I(\theta) = \theta^{-2}$.
- (a) $m{\Theta}$ [factorización para suficiencia] Hallar un estadístico suficiente para θ basado en una muestra aleatoria de la duración de n dispositivos.
- (b) Hallar el estimador de máxima verosimilitud de θ basado en una muestra aleatoria de la duración de n dispositivos.

(c) $m{P}$ [simulación] Usando los números aleatorios

0.186, 0.178, 0.488, 0.255, 0.392, 0.234, 0.597, 0.205, 0.611, 0.651.

simular 10 valores de X cuando $\theta=1$ y en base a esa información muestral calcular el estimador de máxima verosimilitud de θ .

(d) Se pusieron a prueba 10 de esas máquinas y se obtuvieron los siguientes tiempos:

2.00, 5.48, 2.43, 1.90, 5.85, 1.58, 2.30, 2.87, 3.62, 2.71.

Basándose en la información muestral calcular el estimador de máxima verosimilitud de la probabilidad de que una máquina del mismo tipo funcione sin fallas más de dos años y medio.

A partir de la definición de función intensidad de fallas, se llega a la función de densidad $f_{\theta}(x) = (3/\theta)(x/\theta)^2 \exp(-(x/\theta)^3) \mathbf{1} \{x > 0\}$. Puede verificar el estimador que obtenga en (b) buscando los estimadores de una distribución Weibull en bibliografía.

- **9.11.** En una mesa electoral votaron 129 ciudadanos. Se extrajeron (sin reposición) 7 sobres al azar de la urna, se examinaron y resultó que el candidato verde obtuvo exactamente 3 votos. Estimar por máxima verosimilitud la cantidad de votos por el candidato verde que había en la urna.
- **9.12.** P [familia exponencial, suficiencia] Mostrar que la familia de distribuciones $\Gamma(\nu,\lambda)$ es una familia exponencial a 2 parámetros. Hallar un estadístico suficiente para (ν,λ) basado en una muestra aleatoria de tamaño n
- 9.13. P [familia exponencial, suficiencia]
- (a) Mostrar que la familia de distribuciones $\mathcal{N}(\mu, \sigma^2)$ puede expresarse en la forma

$$f_{\theta}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2\right),$$

donde $\theta = (\mu, \sigma^2)$.

- (b) Sea X_1, \ldots, X_n una muestra aleatoria de tamaño n de la distribución $\mathcal{N}(\mu, \sigma^2)$. Hallar la expresión de la densidad conjunta y mostrar que $T = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$ es un estadístico suficiente para θ .
- (c) Sea $\bar{X}=\frac{1}{n}\sum_{i=1}^n X_i$. Mostrar que $\sum_{i=1}^n (X_i-\bar{X})^2=\sum_{i=1}^n X_i^2-n\bar{X}^2$ y deducir que $T'=(\bar{X},S^2)$, donde

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

es un estadístico suficiente para θ .

- (d) Hallar el estimador de máxima verosimilitud para θ basado en la muestra aleatoria X_1, \ldots, X_n .

10. Guía 10

- 10.1. Las políticas de fomento del empleo están determinadas por la tasa de desocupación y tienen por objetivo mantenerla por debajo de un nivel que se considera aceptable, por ejemplo un 4%. Si se quiere diseñar un test acorde con esa situación,
- (a) ¿cuál debe ser la hipótesis nula y cuál la alternativa?;
- (b) ¿qué significan los errores de tipo I y los errores de tipo II?;
- (c) ¿qué valores considera apropiados para el nivel de significación del test?
- (a) Calcular el nivel de significación del test.
- (b) Calcular la probabilidad de cometer errores de tipo II para todas las situaciones posibles. ¿Cuál es la máxima probabilidad de cometer un error de tipo II?
- (c) Tabular y graficar la función de potencia del test.
- (d) Repetir los incisos anteriores en el caso de que las dos bolas se hubiesen extraído sin reposición.
- **10.3.** P [conceptual] Se observará un único valor de una variable aleatoria X cuya función de probabilidad p(x) puede ser $p_0(x)$ o $p_1(x)$, donde $p_0(x)$ y $p_1(x)$ están definidas en la siguiente tabla:

x	0	1	2	3	4	5
$p_0(x)$	0.02	0.03	0.05	0.05	0.35	0.50
$p_1(x)$	0.04	0.05	0.08	0.12	0.41	0.30

- (a) Hallar todos los test de nivel $\alpha = 0.05$ de la hipótesis $H_0: p(x) = p_0(x)$ contra $H_1: p(x) = p_1(x)$.
- (b) Calcular β para cada uno de los test hallados en (a). ¿Cuál es el mejor test de todos?
- 10.4. La porota vende dos variedades de soja. El rinde (en toneladas) por hectárea de la variedad 1 es una variable aleatoria con distribución normal de media 6.2 y desvío 0.45, y el de la variedad 2 es una variable aleatoria con distribución normal de media 7.0 (mayor que las de variedad 1) y desvío 0.45. Vivaldo compró semillas de la variedad 2 y antes de seguir comprando, quiere asegurarse de que las semillas que le enviaron son de la variedad 2.

- (a) Diseñar un test de hipótesis que le garantice a Vivaldo que la probabilidad de seguir comprando semillas a *La porota* cuando le hayan enviado de la variedad 1 sea 0.05.
- (b) Calcular β para $\mu = 7.0$.
- (c) ¿Cuántas hectáreas deben cultivarse para que $\beta(7.0) \leq 0.1$?
- (d) Vivaldo cultivó 10 hectáreas con la semillas que le enviaron y obtuvo los siguientes rindes:

```
7.36, 7.62, 7.02, 6.99, 6.66, 6.74, 6.25, 6.41, 6.91, 7.11.
```

Basándose en esa información: calcular el p-valor del test, y determinar qué debe hacerse.

- 10.5. Un productor afirma que la media del voltaje de ruptura de ciertos capacitores es mayor que 200. El voltaje de ruptura de dichos capacitores obedece a una distribución normal de varianza 25. Usando una muestra aleatoria de tamaño 10 diseñar un test de hipótesis de nivel de significación $\alpha=0.05$ para decidir si la afirmación del productor es verdadera y calcular la probabilidad de decidir erróneamente cuando el verdadero valor de la media del voltaje de ruptura es 210.
- 10.6. Una máquina produce varillas cuya longitud (en cm) es una variable aleatoria con distribución normal de varianza 25. Se examina una muestra aleatoria de 36 varillas producidas por esa máquina y se registra una longitud promedio de 51.74 cm. Con un nivel de significación de 0.05, ¿se puede garantizar que la longitud media de las varillas producidas por esa máquina supera los 50 cm? Hallar el p-valor.
- 10.7. En 1761 James Short midió 53 veces la paralaje solar. El promedio de las mediciones resultó ser 8.616 segundos de grado. Suponiendo que la distribución de las mediciones es una normal de media μ y desvío $\sigma=0.75$, decidir a un nivel de significación $\alpha=0.05$ si hay suficiente evidencia para rechazar la hipótesis $\mu=8.789$. Hallar el p-valor.
- 10.8. Para ir de su casa al trabajo, Aparicio va por un camino por el que tarda, en media, 40 minutos en llegar. Juan le sugiere otro camino para reducir ese tiempo. Aparicio lo probó 10 veces y tardó en llegar los siguientes tiempos:

Suponiendo que los tiempos de viaje obedecen a una distribución normal, ¿se puede asegurar, con un nivel de significación de 0.1, que el camino sugerido por Juan es más rápido?

10.9. En la *Burbuja feliz* se acaba de instalar una máquina para llenar sifones de soda. La máquina es eficaz cuando el desvío estándar de la cantidad de soda en los sifones no supera 25 mililitros. En una muestra de 10 sifones se observaron las siguientes cantidades (en litros) de soda:

1.029, 0.943, 1.071, 0.986, 0.962, 0.995, 0.991, 1.002, 1.003, 0.978.

Suponiendo que la cantidad de soda en los sifones obedece a una distribución normal. ¿se puede asegurar, con un nivel de significación de 0.05, que la máquina no es eficaz?

- (a) Asegurarse que el peso medio es menor a 1
- (b) Asegurarse que el desvío del peso es menor a 0.14

10.11. △

10.12. De Los siguientes datos son las duraciones (en horas) de una muestra de 6 lámparas: 61, 1905, 1076, 623, 33, 167. Suponiendo que los datos obedecen a una distribución exponencial de intensidad λ determinar si ellos permiten con un nivel de significación $\alpha=0.05$ refutar la hipótesis de que $\lambda \leq 0.0005$. Hallar el p-valor.

10.13. △

10.14. Basándose en una muestra aleatoria de una población con distribución uniforme sobre el intervalo $[0,\theta]$, $\theta > 0$, diseñar un test de hipótesis para $H_0: \theta \leq 1$ cuyo nivel de significación sea $\alpha = 0.05$ y tal que el valor de la función de potencia en $\theta = 1.1$ sea 0.9.

10.15. La longitud en metros de cada rollo de alambre en un lote es una variable aleatoria con distribución uniforme sobre el intervalo $[15,15+\theta]$. Se examinaron 4 rollos y la máxima longitud observada resulto ser 25 metros. En base a la información muestral, y con nivel de significación de 0.01 ¿se puede afirmar, que la longitud media de los rollos del lote es menor que 20 metros?

10.16. \bigcirc Sea X una variable aleatoria cuya función de densidad es

$$f_{\theta}(x) = \frac{3x^2}{\theta^3} \mathbf{1} \{ 0 \le x \le \theta \}, \qquad \theta > 0.$$

Diseñar un test de hipótesis de nivel 0.1 para testear la hipótesis de que la media de X es mayor que 1.2 basado en una muestra de X de tamaño 1. Graficar la función de potencia del test. \circlearrowleft : Usar pivote $Q = \max(X)/\theta$

10.17. En una elección se presentarán dos candidatos: el amarillo y el azul. Basándose en los resultados de una encuesta que se realizará sobre un conjunto de 100 ciudadanos a los que se les preguntará a cuál de los dos votará, diseñar un test de hipótesis para verificar si la intención de votos por el candidato azul supera el 35 %, con un nivel de significación asintótico del 5 %.

10.18. P [Maronna pp. 143-144] Una de las más celebres "Leyes de Murphy" establece que "si se deja caer al suelo una tostada untada con dulce, la probabilidad de que caiga del lado del dulce es mayor que la de que caiga del lado del pan".

- (a) Para verificarla, se realizó un experimento en la University of Southwestern Louisana, en el que se dejaron caer 1000 tostadas untadas con mermelada de grosellas, de las cuales cayeron 540 del lado del dulce. ¿Qué conclusión puede sacar?
- (b) El comité de investigaciones de la University of Southwestern Louisana decreta que, para que el experimento sea considerado concluyente, deberá cumplir con (i) si la Ley de Murphy es falsa, la probabilidad de que el test la confirme debe ser ≤ 0.01 ; (ii) si la Ley es cierta, y la probabilidad de caer del lado del dulce es > 0.6, entonces la probabilidad de confirmarla debe ser ≥ 0.95 . ¿Cuántas tostadas hay que arrojar para que se cumplan estas condiciones?

10.19. Un fabricante asegura que produce con una calidad del $5\,\%$ de artículos defectuosos. Un comprador de grandes cantidades de esos artículos observa una muestra de 100 artículos y descubre 10 defectuosos. Realizar un test de hipótesis para determinar con un nivel de significación asintótico de 0.05 si existen motivos para dudar de la afirmación del fabricante.

10.20. P En una urna hay n bolas negras. Eusebio afirma que $n \geq 24$. Se agregan 10 bolas rojas. Luego se realizan 100 extracciones de una bola con reposición y se observan 28 rojas. Se puede rechazar la afirmación de Eusebio con nivel de significación asintótico $\alpha = 0.1$?

10.21. Una fuente radiactiva emite partículas alfa de acuerdo con un proceso de Poisson de intensidad λ por segundo. Se la observó durante 3 horas y se registraron 5029 emisiones. ¿Al 0.01 de significación asintótica, se puede rechazar la hipótesis $\lambda=0.5$? Hallar el p-valor aproximado.

10.22. La cantidad de accidentes de tránsito por semana en la intersección de Paseo Colón y Estados Unidos obedece a una distribución de Poisson. En una muestra de 80 semanas se observaron las frecuencias:

Cantidad de accidentes	0	1	2	3	4
Frecuencia	28	25	20	5	2

Al $5\,\%$ de significación asintótica, ¿se puede afirmar que la media de la cantidad de accidentes por semana en la intersección de Paseo Colón y Estados Unidos es menor que 2? Hallar el p-valor aproximado.

11. Guía 11

11.1. Un emisor transmite una señal de valor μ . El canal de comunicación corrompe la transmisión con un ruido normal aditivo $N \sim \mathcal{N}(0,1)$. El receptor recibe una señal de valor $X = \mu + N$. El emisor transmitió 9 veces la señal y el receptor recibió los siguientes valores:

```
8.016, 8.488, 7.395, 9.011, 7.532, 7.841, 8.651, 6.917, 8.490.
```

En base a esa información muestral construir un intervalo de confianza de nivel 0.95 para el valor de la señal transmitida.

- 11.2.
 © Un emisor transmite una señal de valor μ . El canal de comunicación corrompe la transmisión con un ruido normal aditivo $N \sim \mathcal{N}(0, \sigma^2)$, donde $\sigma^2 = 1/100$. El receptor recibe una señal de valor $X = \mu + N$. Para que el receptor pueda decodificar la señal con cierta precisión el emisor repite la transmisión n veces. El receptor decodifica la señal promediando los valores recibidos. Hallar el mínimo valor de n tal que, con un nivel de confianza de 0.99, el receptor pueda decodificar la señal con un error ≤ 0.01 .
- 11.3. Para calcular la distancia entre la Tierra y el Sol, James Short realizó en 1761 varias mediciones de la paralaje solar (ángulo bajo el que se ve el radio ecuatorial de la tierra desde el centro del sol). Los datos siguientes son algunas de las mediciones, en segundos de grado, obtenidas por Short
 - $9.11,\ 8.66,\ 8.34,\ 8.60,\ 7.99,\ 8.58,\ 8.34,\ 7.33,\ 8.64,\ 9.27,\ 9.06,\ 9.25.$

Suponiendo que las mediciones tienen distribución normal $\mathcal{N}(\mu, \sigma^2)$, usar esas observaciones muestrales para construir un intervalo de confianza de nivel 0.95 para μ .

11.4. El voltaje de ruptura de ciertos capacitores obedece a una distribución normal. Se pusieron a prueba 10 capacitores y se obtuvieron los voltajes de ruptura

196.73, 204.37, 201.57, 197.58, 205.89, 199.03, 201.75, 206.53, 199.31, 202.27.

En base a la información muestral construir una cota inferior de confianza de nivel 0.95 para la media del voltaje de ruptura de dichos capacitores.

- 11.5. Sea T una variable aleatoria con distribución exponencial de intensidad λ . Hallar un intervalo de confianza de nivel 0.90 para λ suponiendo que
- (a) en una muestra aleatoria de tamaño 10 se observó que $\sum_{i=1}^{10} t_i = 29.51,$
- (b) en una muestra aleatoria de tamaño 100 se observó que $\sum_{i=1}^{100} t_i = 223.21$.

11.6. Clientes arriban a un banco de acuerdo con proceso de Poisson de intensidad λ por minuto. El banco abre sus puertas a las 10:00; los primeros clientes arribaron a las 10:01, 10:03, 10:11, 10:12, 10:13, 10:16. En base a estos datos construir una cota inferior de confianza de nivel 0.9 para la intensidad λ .

11.7. △

11.8. \bigcirc La longitud en metros de cada rollo de tela en un lote es una variable aleatoria con distribución uniforme sobre el intervalo [15, 15+ θ]. Se examinaron 4 rollos y la máxima longitud observada resultó ser 25 metros. En base a la información muestral construir una cota superior de confianza de nivel 0.99 para θ .

11.9. △

- 11.10. Se recibe un lote de artículos provenientes de un fabricante que asegura que el porcentaje de artículos defectuosos es como máximo 2%. Al observar una muestra de 200 artículos se descubren 11 defectuosos.
- (a) Construir un intervalo de confianza de nivel asintótico 0.9 para la proporción de artículos defectuosos.
- (b) Hallar una cota inferior de nivel asintótico 0.95 para la proporción de artículos defectuosos y en base a ese resultado evaluar la afirmación del fabricante.

11.11. \triangle

- 11.12. En una elección se presentarán dos candidatos: el amarillo y el azul. Se realizó una encuesta a 100 ciudadanos y exactamente 44 respondieron que votarán al candidato azul. Usando esa información construir una cota inferior de confianza de nivel asintótico 0.95 para la proporción de votantes a favor del candidato azul.
- 11.13. Una fuente de polonio emite partículas alfa de acuerdo con un proceso de Poisson de intensidad λ por segundo. Se la observó durante 4 horas y se registraron 11150 emisiones. En base a esta información muestral construir un intervalo de confianza de nivel asintótico 0.99 para la intensidad del proceso.
- 11.14. Los terremotos ocurren en una región con riesgo sísmico de acuerdo con un proceso de Poisson de intensidad λ por año. En los últimos 200 años se registraron 7 terremotos en esa región. En base a esa información, hallar una cota inferior de confianza de nivel asintótico 0.95 para λ .
- 11.15. Una sustancia radiactiva emite partículas alfa de acuerdo con un proceso de Poisson de intensidad λ por segundo.

Se la observó durante 10 segundos y se registraron 4 emisiones. En base a esta información muestral construir un intervalo de confianza de nivel asintótico 0.95 para λ .

Se volvió a observar la misma sustancia radiactiva hasta que emitió la cuarta partícula alfa, lo que sucedió a los 10 segundos. En base a esta información muestral construir un intervalo de confianza de nivel asintótico 0.95 para λ .

12. Guía 12



Figura 1: Inferencia bayesiana is coming

Referencias

- [1] Grynberg, S. Borradores, Curso 23. Buenos Aires: [digital], marzo a junio de 2013.
- [2] Maronna, R. Probabilidad y Estadística Elementales para Estudiantes de Ciencia. 1ra ed. La Plata: [digital], 1995.
- [3] Feller, W. An Introduction to Probability Theory and Its Applications, Vol. I. 2da ed. New York: John Wiley & Sons, 1957.
- [4] Feller, W. An Introduction to Probability Theory and Its Applications, Vol. II. 2da ed. New York: John Wiley & Sons, 1971.
- [5] Grimmet, G., Stirzaker, D. *Probability and Random Processes*. 3ra. ed. Gran Bretaña: Oxford University Press, 2001.
- [6] DeGroot, M. H. *Probability and Statistics*. 2nd. ed. EE.UU.: Addison-Wesley Publishing Company, 1989.
- [7] Boente, G.; Yohai, V. Notas de Estadística. Buenos Aires: [digital], 2012
- [8] Bickel, P. J.; Doksum, K. A. Mathematical Statistics, Basic Ideas and Selected Topics, Vol. 1. 2nd. ed. EE.UU.: Prentice Hall, 2001