Facultad de Ingeniería **UBA**

Probabilidad y Estadística

Guía de Ejercicios (2^a Parte)

Primer Cuatrimestre del 2010

Versión 1.0

Guía 9

9.1 Los valores muestrales 0.69, 0.21, 0.89, 0.79 y 0.46 se obtuvieron de una variable aleatoria cuya densidad es $f_{X|\theta=n}(x) = nx^{n-1}\,\mathbf{1}\{0 < x < 1\}$. Suponiendo que la distribución *a priori* del parámetro n es uniforme sobre el conjunto $\{1, 2, 3, 4, 5, 6\}$, hallar la función de probabilidad *a posteriori* de n y calcular su media y su moda.

9.2 La longitud (en cm.) de ciertas varillas es una variable aleatoria con distribución normal de media μ y desvío estándar 2. La media μ es una variable aleatoria discreta cuya función de probabilidad *a priori* es $\mathbf{P}(\mu = 10) = 0.25$ y $\mathbf{P}(\mu = 14) = 0.75$. Se observa que la longitud de una varilla es 12.1 cm.

(**a**) Hallar la función de probabilidad *a posteriori* de μ.

(**b**) En virtud de la información muestral estimar la probabilidad de que al observar otra varilla, su longitud sea mayor que 13 cm.

9.3 En una urna hay 6 bolas. *A priori* se supone que la cantidad de bolas blancas en la urna se distribuye uniformemente entre 0 y 6 (incluidos los extremos). Se extraen dos bolas de la urna: una es blanca, la otra es negra.

(**a**) Hallar la distribución a posteriori de la cantidad de bolas blancas que había inicialmente en la urna.

(**b**) Estimar la cantidad de bolas blancas que había inicialmente en la urna.

9.4 El peso, W (en gramos), de ciertos objetos es una variable aleatoria cuya función densidad es de la forma $f_{W|\theta}(w) = \frac{1}{2}(w - \theta)\,\mathbf{1}\{\theta \leqslant w \leqslant \theta + 2\}$. *A priori*, θ tiene distribución uniforme sobre el intervalo $[0, 10]$. Se observa que el peso de un objeto es 5 gramos. Hallar la distribución a posteriori de θ.

9.5 Sea X una variable aleatoria con densidad $f_{X|\theta=a}(x) = ax^2\,\mathbf{1}\{0 < x < b\}$.

(**a**) Encontrar la relación entre a y b.

(**b**) *A priori*, se supone que los valores de a están distribuidos uniformemente sobre el intervalo $(0, 2)$. Si se obtuvieron los valores muestrales 0.2, 0.8 y 3 hallar la función de distribución a posteriori de a y calcular su media y su moda.

9.6 Se consideran dos monedas, M_1 y M_2, con las siguientes características: $\mathbf{P}(cara|M_1) = 0.6$ y $\mathbf{P}(cara|M_2) = 0.4$. Se elige una moneda al azar y se la tira repetidas veces. Dado que los primeros dos tiros resultaron ceca, calcular la media de la cantidad adicional de tiros hasta que aparezca la cara.

9.7 Un proceso de producción, produce con una calidad del $100\,p\,\%$ de artículos defectuosos. *A priori* se supone que la proporción p de artículos defectuosos se distribuye uniformemente sobre el intervalo $(0, 1)$. De una partida se examina una muestra de 6 artículos y se encuentran 2 defectuosos. En una caja se ponen otros 4 artículos de la misma partida. En base a la información muestral, estimar la probabilidad de que la caja contenga un solo artículo defectuoso.

9.8 Sea X el número de caras en n tiros de una moneda cuya probabilidad de salir cara es p. *A priori* se asigna una distribución Beta(ν_1, ν_2) para p y se observan x caras en n tiros.

(**a**) Mostrar que la media *a posteriori* de p se encuentra entre la media *a priori*, $\frac{\nu_1}{\nu_1+\nu_2}$, y la frecuencia relativa con que se observa cara, $\frac{x}{n}$.

(**b**) Mostrar que la varianza *a posteriori* de p se comporta asintóticamente de la misma forma que $\frac{x(n-x)}{n^3}$.

(**c**) Mostrar que si la distribución *a priori* para p tiene parámetros $\nu_1 = \nu_2 = 1$, la varianza *a posteriori* de p es menor que su varianza *a priori*.

(**d**) Dar un ejemplo de parámetros ν_1, ν_2 para la distribución *a priori* y datos x y n, para los que la varianza *a posteriori* de p es mayor que la varianza *a priori*.

(**e**) Calcular la probabilidad *a posteriori* de que en el próximo tiro de la moneda salga cara.

(**f**) ¿Cómo deben ser los parámetros ν_1 y ν_2 de la distribución *a priori* para p si se quiere expresar que la moneda "parece ser honesta"?

(**g**) Si la moneda parece ser honesta y como resultado de los n tiros se observaron 1 ceca y $n-1$ caras, ¿cómo debe ser n para que en el siguiente tiro las probabilidades sean 2 a 1 a favor de cara?

9.9 El Gobierno de la Ciudad Autónoma de Buenos Aires contrata a una empresa encuestadora para "determinar" la proporción, p, de la población de la Ciudad irritada por las polémicas declaraciones de uno de sus flamantes ministros. Para analizar los datos la encuestadora cuenta con dos estadísticos bayesianos, el primero opina que la densidad *a priori* para p es $f_1(p) = 10(1-p)^9\,\mathbf{1}\{0 < p < 1\}$ pero el segundo opina que sería más apropiado usar una *a priori* de la forma $f_2(p) = 10p^9\,\mathbf{1}\{0 < p < 1\}$.

(**a**) ¿Analizar qué significado tienen esas dos opiniones y qué consecuencias tienen sobre el análisis de los datos si se decide que para estimar p se utilizará la media de la distribución *a posteriori* basada en el resultado de una encuesta realizada a 10 vecinos de la Ciudad?

(**b**) ¿Qué cantidad de vecinos de la Ciudad deben encuestarse si se quiere garantizar que las medias de las dos distribuciones *a posteriori* de p difieran en menos de 0.001?

9.10 Sea θ una variable aleatoria con densidad $f_\theta(t) \propto t^{\nu-1}e^{-\lambda t}\,\mathbf{1}\{t > 0\}$, donde ν es un número natural y λ es un número real positivo.

(**a**) Calcular la media y la moda de θ.

(**b**) Sean $X_1, \dots, X_n$ variables aleatorias que condicionadas a $\theta = t$ son independientes y cada una tiene distribución de Poisson de media t. Hallar la forma de la densidad *a posteriori* de θ si los valores observados de $X_1, \dots, X_n$ son $x_1, \dots, x_n$. Calcular la media *a posteriori* y mostrar que se puede expresar en la forma de un promedio ponderado $\gamma_n \bar{x} + (1-\gamma_n)\mu_0$, donde $\bar{x} = \frac{1}{n}\sum_{k=1}^n x_k$, $\mu_0 = \mathbf{E}[\theta]$ y $\gamma_n \to 1$ cuando $n \to \infty$.

(**c**) Sean $X_1, \dots, X_n$ variables aleatorias que condicionadas a $\theta = t$ son independientes y cada una tiene distribución exponencial de intensidad t. Hallar la moda de la distribución *a posteriori* de θ si los valores observados de $X_1, \dots, X_n$ son $x_1, \dots, x_n$.

9.11 El número de accidentes que ocurren diariamente en una planta industrial tiene una distribución Poisson de media λ desconocida. Sobre la base de experiencias previas en plantas industriales similares un estadístico afirma que los posibles valores de λ se distribuyen *a priori* como una variable exponencial de media $1/2$. Durante los primeros 9 días de funcionamiento de la planta industrial ocurrieron un total de 45 accidentes.

(**a**) Hallar la distribución *a posteriori* de λ y en base a la información suministrada calcular la probabilidad de que durante el décimo día ocurran dos o más accidentes.

(**b**) Hallar un intervalo de confianza de nivel 0.95 para estimar λ.

9.12 Se toma una muestra de tamaño 4 de una distribución uniforme en el intervalo $[0, b]$. La distribución a priori de b es $f(b) = 6/b^2\, \mathbf{1}\{b \geqslant 6\}$. Hallar en cada caso la distribución *a posteriori*, su media y su moda, si los resultados muestrales fueron:

(**a**) 2, 5, 6, 9.

(**b**) 1, 3, 4, 5.

9.13 Se toma una muestra al azar de n estudiantes de una población y se miden sus alturas. La altura promedio resulta ser 1.77 m. Se supone que las alturas de la población están distribuidas como una normal de media μ desconocida con desvío estándar 5 cm. Se considera que la distribución *a priori* de μ es una normal de media 1.73 m. y desvío estándar 2.5 cm.

(**a**) Hallar la distribución *a posteriori* de μ.

(**b**) Mostrar que con los datos del problema, la media *a posteriori* puede expresarse como un promedio ponderado de la forma $\gamma_n 1.77 + (1 - \gamma_n)1.73$ y analizar el comportamiento de γ_n cuando $n \to \infty$.

(**c**) Se muestrea al azar un nuevo estudiante de la población y mide x m. Hallar la distribución predictiva para x.

(**d**) Para $n = 10$, dar un intervalo de confianza de nivel 0.95 para μ y un intervalo predictivo de nivel 0.95 para x.

(**e**) Hacer lo mismo para $n = 100$.

Ejercicios Complementarios

9.14 El tiempo de funcionamiento en años de cada chip de computadoras, producido por una firma de semiconductores china, se distribuye como una exponencial de media $1/\lambda$. La distribución a priori de λ es una Gamma con función de densidad

$$f(\lambda) = \frac{1}{2}\lambda^2 e^{-\lambda}\, \mathbf{1}\{\lambda > 0\}.$$

Sabiendo que el promedio del tiempo de funcionamiento de 20 chips examinados es de 4.5 años, hallar la densidad *a posteriori* de λ y estimar puntualmente λ utilizando la moda.

9.15 En un maratón participan N personas, numeradas de 1 a N. Se elige un maratonista al azar y tiene el número 203. Se quiere estimar N.

(**a**) Si la distribución *a priori* de N es una geométrica de media 100, ¿cuál es la distribución *a posteriori* de N?

(**b**) Calcular la media y el desvío estándar de N.

9.16 Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución uniforme sobre el intervalo $[\theta - 1/2, \theta + 1/2]$. Se asume que θ se distribuye a priori uniformemente en el intervalo $(10, 20)$.

(**a**) Si $n = 1$ y se observó el valor 12, hallar la distribución a posteriori de θ.

(**b**) Si $n = 6$ y se observaron los valores 11, 11.5, 11.7, 11.1, 11.4, y 10.9, hallar la distribución a posteriori de θ.

(**c**) En cada caso, obtener estimadores puntuales.

9.17 Sea X una variable aleatoria con distribución uniforme en el intervalo $[\theta - 1, \theta]$. El parámetro θ posee una distribución *a priori* de la forma $f(\theta) = e^{-2|\theta|}$. Si se observa un único valor de dicha variable y resulta ser 0, estimar la probabilidad de que la variable X sea en módulo menor a 1/4.

Guía 10

10.1 Sea X_1, X_2, X_3, X_4 una muestra aleatoria de una variable aleatoria tal que $\mathbf{E}_\theta[X] = \theta$ y $\mathbf{var}_\theta[X] = 1$. Comparar los siguientes estimadores para θ:

$$\overline{X} = \frac{1}{4}\sum_{i=1}^{4} X_i, \qquad \frac{X_1 + 2X_2 + 3X_3 + 4X_4}{10}, \qquad \frac{X_1 + X_2 + X_3}{3}.$$

10.2 Sea X una variable aleatoria con distribución Bernoulli de parámetro θ. Sean $\hat{\theta}_1 = X$ y $\hat{\theta}_2 = 1/2$ dos estimadores para θ.

(**a**) Analizar si $\hat{\theta}_1$ y $\hat{\theta}_2$ son estimadores insesgados para θ.

(**b**) Comparar los errores cuadráticos medios (ECMs) de los dos estimadores y graficarlos como funciones de θ.

10.3 En una serie de m experiencias Bernoulli se registran X éxitos, y en una serie posterior de n experiencias Bernoulli se registran Y éxitos ($n \neq m$). Se proponen los siguientes estimadores del parámetro p:

$$\hat{p}_1 = \frac{1}{2}\left(\frac{X}{m} + \frac{Y}{n}\right) \qquad \text{y} \qquad \hat{p}_2 = \frac{X+Y}{m+n}.$$

¿Cuál es preferible? Justificar la respuesta.

10.4 Sea $X_1, \dots, X_n$ una muestra aleatoria de una variable aleatoria $X \sim U(0, \theta)$. Se considera $X_{(n)} = \text{máx}(X_1, \dots, X_n)$ como estimador para θ.

(**a**) Hallar la función densidad de $X_{(n)}$ y mostrar que

$$\mathbf{E}_\theta[X_{(n)}] = \frac{n}{n+1}\,\theta \qquad \text{y} \qquad \mathbf{var}_\theta[X_{(n)}] = \frac{n\theta^2}{(n+1)^2(n+2)}.$$

(**b**) Calcular el sesgo del estimador $X_{(n)}$ y demostrar que se trata de un estimador asintóticamente insesgado para θ.

(**c**) Usando la desigualdad de Chebyshev demostrar $X_{(n)}$ es un estimador débilmente consistente.

(**d**) Mostrar que el estimador $2\bar{X}$ para θ tiene mayor ECM que el estimador $X_{(n)}$.

10.5 Sea $X_1, X_2, \dots, X_n$ una muestra de una variable con distribución $N(0, 1)$. Se definen $h(x) = \mathbf{1}\{x < 0.44\}$, $Y = h(X)$ y $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Si $n = 25$, calcular

(**a**) $\mathbf{P}(\overline{X} < 0.25)$.

(**b**) $\mathbf{P}(S^2 > 0.577)$.

(**c**) $\mathbf{P}(0.576 \leqslant \hat{p} \leqslant 0.764)$.

(**d**) Repita el inciso anterior para $n = 4$.

10.6 Hallar los estimadores de máxima verosimilitud (MV) para muestras aleatorias $X_1, \dots, X_n$ de las siguientes distribuciones: $N(\mu, 1)$, Bernoulli(p), $U(0, \theta)$.

10.7 Una urna contiene bolas rojas y bolas negras. Se sabe que la proporción p de bolas rojas en la urna es $2/5$ ó $4/5$. Usando el método de MV estimar p si en una muestra aleatoria de $n = 5$ bolas (con reposición) se observaron 3 bolas rojas y 2 negras.

10.8 Una planta de ensamblaje recibe un lote de 100 piezas de precisión de las cuales una cantidad desconocida M son defectuosas. Para controlar el lote se elige al azar una muestra (sin reposición) de 10 piezas, se examinan y resulta que hay exactamente 2 defectuosas. Usar el método de MV para estimar M.

10.9 Sea $X_1, \dots, X_n$ una muestra aleatoria de tamaño n de una variable aleatoria X perteneciente a una *familia exponencial uniparamétrica* con función de probabilidad o función densidad de la forma

$$f_{X|\theta}(x) = c(x)\, e^{a(\theta)T(x)+b(\theta)}$$

donde a y b son funciones a valores reales de $\theta \in \Theta$, y T y c son funciones a valores reales de x que no dependen de θ.

(**a**) Mostrar que el EMV (estimador de MV) de θ basado en la muestra $X_1, \dots, X_n$ es solución de la ecuación

$$\frac{-b'(\theta)}{a'(\theta)} = \frac{1}{n}\sum_{i=1}^{n} T(X_i)$$

siempre y cuando la solución pertenezca al espacio paramétrico, Θ, correspondiente al parámetro θ.

(**b**) Mostrar que las siguientes familias de distribuciones son familias exponenciales uniparamétricas y usando el resultado del **Inciso** (**a**) hallar el EMV del parámetro θ basado en una muestra aleatoria de tamaño n, $X_1, \dots, X_n$: Bernoulli(θ), binomial(n, θ), geométrica(θ), Pascal(n, θ), Poisson(θt), Beta(θ, β), Beta(α, θ), exponencial(θ), Gamma(ν, θ), Gamma(θ, λ), normal(θ, σ^2), normal(μ, θ).

10.10 Hallar los estimadores de máxima verosimilitud de los parámetros de las siguientes distribuciones:

(**a**) $N(\mu, \sigma^2)$, donde $\mu \in \mathbb{R}$ y $\sigma^2 > 0$.

(**b**) Pareto:

$$f(x|\alpha, \beta) = \left(\frac{\alpha}{\beta}\right)\left(\frac{x}{\beta}\right)^{-(\alpha+1)} \mathbf{1}\{x \geqslant \beta\}, \text{ donde } \alpha > 0,\ \beta > 0.$$

10.11 En una urna hay r bolas rojas y 5 bolas negras, donde $0 \leqslant r \leqslant 6$. Se extraen 2 bolas al azar sin reposición, se las examina y se las repone en la urna. Sabiendo que las dos bolas examinadas resultaron rojas, estimar por MV la probabilidad de que al extraer nuevamente dos bolas al azar sin reposición una sea roja y la otra negra.

10.12 Un buzo debe realizar una tarea en el océano que le insumirá 45 minutos. La duración en minutos de cada tanque de oxígeno es una variable aleatoria X con distribución normal con desvío $\sigma = 2$ y media desconocida. En una muestra aleatoria $X_1, \dots, X_9$ se observaron las siguientes duraciones para los tanques:

$$37.447,\ 51.101,\ 34.258,\ 38.401,\ 33.288,\ 45.971,\ 47.348,\ 36.241,\ 41.585$$

Estimar por MV la probabilidad de que el buzo pueda terminar su tarea si lleva un solo tanque de oxígeno, utilizando la siguiente información muestral:

(**a**) los valores observados $X_1, \dots, X_9$.

(**b**) los valores $Y_1, \dots, Y_9$, donde $Y_i = \mathbf{1}\{X_i > 45\}$.

(**c**) Mostrar mediante simulaciones que aunque la estimación del **Inciso** (**a**) pueda ser sesgada, presenta menor ECM que la estimación del **Inciso** (**b**). (*Sugerencia*: elegir un valor μ para la media de la normal y generar m muestras de tamaño $n = 9$ para obtener m valores de cada estimador, evaluar el sesgo y el ECM; repetir el procedimiento para distintos valores de μ en el intervalo $45 \pm 4\,\sigma$ y graficar el sesgo y el ECM en función de μ.)

10.13 Una máquina produce objetos de diámetro aleatorio con distribución normal de media 100 cm. y varianza desconocida. Los objetos de diámetro inferior a 99 cm. se consideran defectuosos. Calcular el EMV para la varianza sabiendo que en una muestra aleatoria de 10 objetos se encontraron exactamente

(**a**) 2 defectuosos.

(**b**) 7 defectuosos.

10.14 Los siguientes datos corresponden a los tiempos de funcionamiento (en días) hasta que ocurre la primer falla de una muestra de 10 lavarropas industriales:

$$359,\ 317,\ 2379,\ 53,\ 554,\ 536,\ 414,\ 122,\ 345,\ 739.$$

Suponiendo que los datos obedecen a una distribución exponencial hallar los EMVs para su media, varianza y mediana.

10.15 La duración (en horas) de ciertas baterías es una variable aleatoria T con distribución exponencial. Se pusieron a prueba 10 baterías y solamente 7 superaron dos horas de duración. Estimar por MV la probabilidad de que una batería de ese tipo dure más de 2.5 horas.

10.16 Se tienen 3 observaciones normales con la misma media μ y desviaciones $1, 2, 3$. Calcular la varianza del EMV de μ y compararla con la del promedio $\bar{X}$.

10.17 Sean $X_1, \dots, X_n$ e $Y_1, \dots, Y_m$ dos muestras aleatorias independientes de dos distribuciones normales de la misma media μ y varianzas 4 y 9, respectivamente. Hallar el EMV para μ.

10.18 Durante un horario pico el tiempo (en segundos) entre vehículos que atraviesan una cabina de peaje es una variable aleatoria T con distribución exponencial de intensidad λ. Hallar el EMV de λ sabiendo que $\mathbf{E}[T] \leqslant 15$ y que 35 vehículos demoraron un total de 10 minutos en atravesar la cabina de peaje.

10.19 [*Degroot, pág. 355*] Lucas y Monk deben estimar un parámetro $\theta > 0$. Lucas puede observar el valor de una variable aleatoria X con distribución Gamma de parámetros $r = 3$ y $\lambda = \theta$. Monk puede observar el valor de una variable aleatoria Y con distribución Poisson de media 2θ. Lucas observa el valor 2 y Monk el valor 3. Mostrar que las funciones de verosimilitud obtenidas Lucas y Monk serán proporcionales y encontrar el EMV de θ obtenido por cada uno.

10.20 [*Degroot, pág. 379*] Se realiza una encuesta sobre una cuestión "sensible" utilizando el siguiente método: *Se elige una muestra representativa de n sujetos de la población. A cada sujeto se le hace al azar y en forma independiente una pregunta "estándar" o la pregunta "sensible". Ambas preguntas pueden tener respuesta positiva o negativa (por ej., le gusta o no le gusta, está o no de acuerdo, etc.). La probabilidad de recibir una respuesta positiva para la pregunta "estándar" es 1/2, pero para la pregunta "sensible" es un valor p desconocido.*
Los n sujetos encuestados dieron un total de r respuestas positivas. Hallar el EMV de p.

Ejercicios Complementarios

10.21 Sea $X_1, \dots, X_n$ una muestra aleatoria de una variable aleatoria X.

(**a**) Calcular $\bar{X}$ y S para una muestra de volumen $n = 10$ cuyos valores son: $1, 2, \dots, 10$.

(**b**) Suponer que por un error de tipeo, el "10" se transcribe como "100" ¿Cómo se modifican $\bar{X}$ y S?.

(**c**) Hacer lo mismo para la media podada $\bar{X}_{0.25}$ definida por

$$\bar{X}_{0.25} = \frac{1}{n - 2m} \sum_{i=m+1}^{n-m} X_{(i)},$$

donde $X_{(1)} \leqslant \cdots \leqslant X_{(n)}$ son las observaciones ordenadas y $m = [(0.25)n]$.

10.22 Sea $X_1, \dots, X_n$ una muestra de una variable aleatoria X tal que $\mathbf{E}_\theta[X] = \mu(\theta)$ y $\mathbf{var}_\theta[X] = \sigma^2$, donde σ^2 es conocido. Se consideran los siguientes estimadores para $\mu(\theta)$

$$\hat{\mu}_{\mathbf{w}} = \sum_{i=1}^{n} w_i X_i,$$

donde $w_i \geqslant 0$, $i = 1, \dots, n$ son constantes conocidas y $\mathbf{w} = (w_1, \dots, w_n)$.

(**a**) Hallar una condición sobre las constantes w_i, necesaria y suficiente, para que los estimadores $\hat{\mu}_{\mathbf{w}}$ resulten insesgados para $\mu(\theta)$.

(**b**) Entre todos los estimadores $\hat{\mu}_{\mathbf{w}}$ que resulten insesgados para $\mu(\theta)$ hallar el que minimiza el ECM. (*Sugerencia*: utilizar multiplicadores de Lagrange.)

10.23 Sea $X_1, \dots, X_n$ una muestra aleatoria de tamaño n de una variable aleatoria X cuya función densidad es de la forma

$$f_\theta(x) = e^{-(x-\theta)}\, \mathbf{1}\{x > \theta\}, \qquad \theta > 0.$$

Se proponen dos estimadores para θ: $\hat{\theta}_1 = \bar{X}$ y $\hat{\theta}_2 = X_{(1)}$. Analizar el sesgo de los estimadores propuestos; comparar sus ECMs y graficarlos como funciones de θ.

10.24 [*ver* **Ejercicio 3.18**] Sea $X_1, \dots, X_n$ una muestra de una variable aleatoria X tal que $\mathbf{E}[X^4] < \infty$.

(**a**) Mostrar que cuando la media, μ, de X es conocida el estimador para la varianza de X, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$, es insesgado y débilmente consistente.

(**b**) Mostrar que cuando la media de X es desconocida, el estimador para la varianza de X, $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$, es insesgado. (También es débilmente consistente pero no es tan fácil de mostrar.)

10.25 Se tienen n observaciones normales con la misma media μ desconocida y desviaciones estándar $\sigma_1, \dots, \sigma_n$ conocidas pero no necesariamente iguales. Calcular la varianza del EMV de μ y compararla con la del promedio simple $\bar{X}$. (*Sugerencia*: utilizar la desigualdad de Cauchy-Schwartz)

10.26 Sea $X_1, \dots, X_n$ una muestra aleatoria de una distribución uniforme sobre el intervalo $[\theta - 1/2, \theta + 1/2]$. Hallar el EMV para θ.

10.27 En una fábrica hay dos máquinas: $M1$ y $M2$. El 20 % de las piezas elaboradas por $M1$ y el 40 % de las elaboradas por $M2$ son defectuosas. De un lote de 1000 piezas cuya procedencia se ignora, aunque se sabe que son todas de la misma máquina, se extraen 10 piezas al azar y se encuentran 3 defectuosas. Estimar por MV la varianza de la cantidad de piezas defectuosas en el lote.

10.28 Jorge ingresa en un banco a las 11:30 para cobrar un cheque. Recibe el número 68 de la fila de espera y observa que está siendo atendido el cliente número 61. El tiempo de atención (en minutos) para cada cliente se distribuye como una variable exponencial de intensidad λ. A las 11:45 comienza a ser atendido el cliente con el número 66 y Jorge decide salir del banco a fumar un cigarrillo. Suponiendo que Jorge demora 6 minutos en volver a la fila de espera, estimar por MV la probabilidad de que haya perdido su turno en la fila.

10.29 Sea T una variable aleatoria con distribución Gamma de parámetros $\nu \in \mathbb{R}^+$ y $\lambda \in \mathbb{R}^+$ cuya función densidad es

$$f(t|\nu, \lambda) = \frac{\lambda^\nu t^{\alpha-1}}{\Gamma(\nu)} e^{-\lambda t}\, \mathbf{1}\{t > 0\},$$

donde $\Gamma(\nu) = \int_0^\infty t^{\nu-1} e^{-t} dt$. En particular, $\Gamma(\nu) = (\nu - 1)!$ para todo $\nu \in \mathbb{N}$

(**a**) Suponiendo que λ es conocido y $\nu \in \mathbb{N}$, hallar el EMV para ν basado en una muestra aleatoria de tamaño n.

(**b**) Suponiendo que $\lambda = 1$ y $\nu \in \mathbb{R}^+$, hallar la ecuación de verosimilitud para ν. ¿Es posible obtener una expresión explícita del EMV para ν? Si la respuesta es negativa, realizar un programa que permita calcularlo usando una computadora.

(**c**) Se proponen los siguientes procedimientos para encontrar los EMVs para ν y λ en el caso en que $\nu \in \mathbb{N}$ y $\lambda \in \mathbb{R}^+$:

1. Calcular la función de verosimilitud (función de $\mathbb{N} \times \mathbb{R}^+ \to \mathbb{R}$) y optimizarla gráficamente o con algún algoritmo.
2. Para cada $\nu \in \mathbb{N}$, hallar una expresión del EMV de λ; reemplazarla en la función de verosimilitud; analizar la monotonía de la función de ν así obtenida y encontrar una expresión (o condición) que permita computar el EMV para ν.

En cada caso, elegir valores para lo parámetros desconocidos, simular datos y comprobar que los procedimientos propuestos para hallar los estimadores sean satisfactorios.

10.30 ⌨ Sea $T_1, \dots, T_n$ un muestra aleatoria del tiempo de duración sin fallas de una máquina cuya función intensidad de fallas es $\lambda(t) = \alpha\beta t^{\beta-1}\,\mathbf{1}\{t > 0\}$.

(**a**) Suponer que β es conocido y hallar el EMV para α.

(**b**) Suponer ahora que los dos parámetros β y α son desconocidos. Hallar las ecuaciones de verosimilitud para los dos parámetros. Proponer un procedimiento para hallar los EMVs de los dos parámetros e implementarlo en un computador.

(**c**) Generar una muestra con $n = 10$ valores de la distribución de T suponiendo que $\beta = \alpha = 1$ y usar el procedimiento propuesto en el **Inciso** (**b**) para calcular los EMVs para β y α. Comparar las estimaciones con los valores usados para generar la muestra.

(**d**) Basándose en la siguiente muestra de valores:

$$2.30,\ 2.00,\ 2.87,\ 2.71,\ 5.48,\ 5.85,\ 1.58,\ 3.62,\ 1.90,\ 2.43,$$

hallar el EMV de $\mathbf{P}(T > 2.5)$.

Guía 11

11.1 Sea X una variable aleatoria cuya función densidad es de la forma

$$f(x|\theta) = \frac{2x}{\theta^2}\mathbf{1}\{0 \leqslant x \leqslant \theta\}, \qquad \theta > 0.$$

Para cada $0 < \alpha < 1$, hallar un cota inferior de nivel $1-\alpha$ para θ basada en una muestra de X de tamaño 1.

11.2 Sea $X_1, \dots, X_n$ una muestra aleatoria de tamaño n de una distribución $U[0, \theta]$.

(**a**) Construir un pivote para θ a partir de su EMV.

(**b**) Para cada $0 < \alpha < 1$, hallar un intervalo de confianza de nivel $1-\alpha$ para θ.

(**c**) [*Soong, pág. 312*] Hallar el valor de k para que el intervalo $(0, k(X_1+X_2))$ tenga nivel de confianza 0.95.

11.3 Un emisor transmite una señal de valor μ. Pero el canal de comunicación corrompe la transmisión con un ruido normal aditivo $N \sim N(0,1)$ y el receptor recibe una señal de valor $X = \mu + N$. Como resultado de la transmisión sucesiva e independiente de la misma señal de valor μ se recibieron 9 señales de valores:

$$8.016, 8.488,\ 7.395,\ 9.011,\ 7.532,\ 7.841,\ 8.651,\ 6.917,\ 8.490.$$

(**a**) Hallar un intervalo de confianza de nivel 0.95 para μ.

(**b**) ¿Cuántas veces debe repetirse la transmisión de la señal μ si se pretende que, con una confianza del 95 %, el receptor pueda decodificarla con un error $\leqslant 0.01$?

11.4 En la siguiente tabla se muestran las mediciones de la velocidad de la luz realizadas por el físico Albert Michelson entre el 5 de junio y el 5 de julio de 1879. Los valores dados + 299.000 son las mediciones de Michelson en km/s.

850	740	900	1070	930	850	950	980	980	880
1000	980	930	650	760	810	1000	1000	960	960
960	940	960	940	880	800	850	880	900	840
830	790	810	880	880	830	800	790	760	800
880	880	880	860	720	720	620	860	970	950
880	910	850	870	840	840	850	840	840	840
890	810	810	820	800	770	760	740	750	760
910	920	890	860	880	720	840	850	850	780
890	840	780	810	760	810	790	810	820	850
870	870	810	740	810	940	950	800	810	870

Suponiendo que las mediciones de la velocidad de la luz obedecen a una distribución normal de media μ y desvío estándar σ desconocidos,

(**a**) construir un intervalo de confianza del 95 % para μ. ¿Se encuentra en ese intervalo el "verdadero" valor de la velocidad de la luz?

(**b**) construir un intervalo de confianza del 95 % para σ.

(**c**) Construir un histograma y explorar informalmente la hipótesis de que la distribución de las mediciones es normal.

11.5 [*Predicción*] Sean $X_1, \dots, X_n, X_{n+1}$ una muestra aleatoria de una población normal de media μ y varianza σ^2. Sean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ y $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$.

(**a**) Hallar la distribución de $X_{n+1} - \bar{X}_n$.

(**b**) Sabiendo que $\sigma^2 = 1$ y que se observó que $\bar{X}_5 = 4$, encontrar un intervalo de tolerancia del 90 % para X_6.

(**c**) Si la varianza σ^2 es desconocida y se observó que $\bar{X}_5 = 4$ y $S_5^2 = 1$, encontrar un intervalo de tolerancia del 90 % para X_6.

11.6 Se testea una muestra de 300 chips de memoria y resultan 30 defectuosos.

(**a**) Construir un intervalo de confianza de nivel 90 % para la proporción de chips defectuosos.

(**b**) Dar una cota inferior de nivel 90 % para la proporción de chips defectuosos.

(**c**) Repetir los incisos anteriores si al testear los 300 chips resulta 1 defectuoso.

11.7 Una empresa farmacéutica quiere estimar la proporción de personas que sufrirán reacciones adversas a cierta droga. La empresa quiere un intervalo bilateral cuyo margen de error sea de 0.01 con 95 % de confianza. ¿De qué tamaño debe ser la muestra?

11.8 Sea X una variable con distribución exponencial de intensidad $\lambda > 0$. Hallar un intervalo de confianza de nivel 0.90 para λ si

(**a**) en una muestra aleatoria de tamaño 10 se observó que $\sum_{i=1}^{10} X_i = 2$,

(**b**) en una muestra aleatoria de tamaño 100 se observó que $\sum_{i=1}^{100} X_i = 25$.

11.9 Los siguientes valores son las duraciones (en horas) de una muestra de 16 lámparas:

$$2632,\ 1742,\ 667,\ 1184,\ 3896,\ 809,\ 1412,\ 1789,$$
$$2396,\ 3358,\ 9901,\ 1925,\ 2981,\ 1843,\ 8315,\ 1299.$$

Suponiendo que los datos obedecen a una distribución exponencial de intensidad λ, calcular una cota inferior de nivel 0.95 para λ.

11.10 Se puede considerar que la emisión de partículas alfa obedece a un proceso de Poisson de intensidad λ partículas por segundo.

(**a**) Se observa una sustancia radiactiva durante 10 segundos y se registran 4 emisiones. Hallar un intervalo de confianza para λ de nivel 0.95.

(**b**) Se observa la misma sustancia hasta que se emita la cuarta partícula, lo que sucede a los 10 segundos. Hallar un intervalo de confianza para λ de nivel 0.95.

11.11 Los fabricantes A y B producen el mismo tipo de alambre de cobre. Los valores de resistencia a la tensión de dos muestras de cable (en libras) son

$$A: \ 5112, 5114, 5115, 5110, 5108, 5101, 5107, 5100, 5106, 5105,$$
$$B: \ 5075, 5072, 5083, 5087, 5074; 5091, 5074.$$

Suponiendo que las mediciones son normales $N(\mu_A, \sigma_A^2)$ y $N(\mu_B, \sigma_B^2)$, respectivamente y que las muestras son independientes, hallar un intervalo de confianza del 99 % para $\mu_A - \mu_B$ si se sabe que

(**a**) $\sigma_A^2 = 289$ y $\sigma_B^2 = 625$.

(**b**) $\sigma_A^2 = 289$ y $\sigma_B^2 = 289$.

(**c**) $\sigma_A^2 = \sigma_B^2$.

11.12 Lucas necesita un enlace satelital y está considerando alquilárselo al proveedor 1. Durante una prueba del enlace el 1 % de 200 mensajes transmitidos registraron error. Monk le presenta a Lucas al proveedor 2, que le ofrece otro enlace a un 20 % menos de costo. Al probar este enlace el 3 % de 250 mensajes transmitidos registraron error. Si p_i es la tasa de error del enlace ofrecido por el proveedor i, hallar un intervalo de confianza del 95 % para la diferencia $p_1 - p_2$. En el lugar de Lucas, ¿a quién le alquila el enlace?

Ejercicios Complementarios

11.13 Sea $X_1, \dots, X_{10}$ una muestra aleatoria de una distribución $U(\theta, \theta + 1)$. Usando el estadístico $M = \min(X_1, \dots, X_{10})$ construir un intervalo de confianza del 95 % para θ.

11.14 La distribución de las concentraciones de SO_4^{-2} para una estación de contaminación es normal con un desvío estándar de 4 p.p.m. (partes por millón). Captando 16 muestras de una hora de duración cada una, seleccionadas al azar en un período de estudio, se obtuvo una media de 9 p.p.m.

(**a**) ¿Cuál es el intervalo de confianza del 95 % para estimar la concentración media?

(**b**) ¿Cuál debe ser el tamaño de la muestra si se desea que el error sea igual o inferior a 0.25 p.p.m., con una confianza del 90 %?

(**c**) ¿Cuál es el intervalo de confianza del 95 % para estimar la concentración media si el desvío fue estimado a partir de los datos de la muestra, dando como resultado 4 p.p.m.?

11.15 Sea X una variable aleatoria de varianza 20 y con distribución uniforme sobre el intervalo (a, b). Sea Y una variable aleatoria de varianza 10 con la misma distribución que la suma de dos variables aleatorias independientes cada una con distribución uniforme sobre el intervalo (c, d). En una muestra aleatoria de X se obtuvieron los valores: 14.88, 16.29, 4.22, 16.40, 12.05, 3.77, 6.57, 10.73, 17.09, 17.2; y en una muestra aleatoria de Y se obtuvieron los valores: 6.61, 15.07, 12.38, 10.05, 7.66. Hallar un intervalo de confianza del 95 % para la diferencia entre la media de X y la media de Y.

11.16 Se toman muestras de aire en dos sitios diferentes de una ciudad. Las muestras del sitio 1 arrojaron los siguientes valores: 0.46, 0.62, 0.37, 0.4, 0.44, 0.58, 0.48, 0.53 y las muestras del sitio 2: 0.82, 0.61, 0.89, 0.51, 0.33, 0.48, 0.23, 0.25, 0.67, 0.88. Suponiendo que las mediciones son normales $N(\mu_1, \sigma_1^2)$ y $N(\mu_2, \sigma_2^2)$, respectivamente y que las muestras son independientes, hallar un intervalo de confianza del 95 % de la forma $\sigma_1^2/\sigma_2^2 \geqslant k$ en cada uno de los siguientes casos

(**a**) $\mu_1 = 0.7$ y $\mu_2 = 0.6$,

(**b**) $\mu_1 = 0.485$ y $\mu_2 = 0.567$,

(**c**) μ_1 y μ_2 son desconocidas.

Guía 12

12.1 Se sabe que una caja contiene 3 bolas rojas y 4 negras o bien 4 bolas rojas y 3 negras. Se tomará la decisión relativa al contenido de la caja extrayendo 3 bolas. Si H_0 corresponde a las 3 bolas rojas y 4 negras y si H_0 será rechazada si se observan 3 rojas, ¿cuánto valen los errores de tipo I y tipo II?

12.2 Sea X una variable uniforme sobre el intervalo $(0, \theta)$. Se quiere verificar $H_0 : \theta = 3$ frente a $H_1 : \theta = 2$ por medio de un solo valor observado de X.

(**a**) Calcular las probabilidades de los errores de tipo I y de tipo II si se elige el intervalo $x < 1$ como la región crítica.

(**b**) Calcular las probabilidades de dichos errores si se elige el intervalo $1 < x < 2$ como región crítica.

(**c**) Si la región crítica es un intervalo, ¿cómo debe ser el intervalo para que con $\alpha = 0.5$ se minimice β?

12.3 Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución uniforme sobre el intervalo $(0, \theta)$, donde $\theta > 0$. Se propone el siguiente criterio para testear las hipótesis $H_0 : \theta \in [3, 4]$ contra $H_1 : \theta \notin (3, 4)$:
"*Rechazar la hipótesis* H_0 *cuando y sólo cuando* $\text{máx}(X_1, \ldots, X_n) \notin [2.9\,, 4]$".

(**a**) Hallar y graficar la función de potencia del test.

(**b**) Determinar el nivel de significación del test.

(**c**) Encontrar el mínimo $n \geqslant 1$ tal que el nivel de significación del test sea $\alpha = 0.1$.

12.4 Se dispone de una muestra de tamaño 1 para decidir entre dos hipótesis sobre la media μ de una distribución exponencial. La hipótesis H afirma que $\mu \leqslant 1$ y la hipótesis K afirma que $\mu > 1$. Diseñar una regla de decisión tal que la máxima probabilidad de equivocarse al rechazar la hipótesis H cuando H es verdadera sea 0.1. Calcular la probabilidad de decidir erróneamente cuando el verdadero valor de la media es $\mu = 1.1$.

12.5 Testear $H_0 : \mu \leqslant 5$ contra $H_1 : \mu > 5$ sobre la base de una muestra aleatoria de tamaño 4, extraída de una población que tiene una distribución de Poisson. Usar un nivel de significación del 5 %. Si dicha muestra aleatoria arrojó los siguientes valores: 3, 7, 2, y 4, ¿hay evidencia significativa para rechazar H_0?

12.6 Verdadero o falso:

(**a**) El nivel de significación de un test de hipótesis es igual a la probabilidad de que la hipótesis nula sea verdadera.

(**b**) La probabilidad de que la hipótesis nula sea rechazada es igual a la potencia del test.

(**c**) Cuando se disminuye el nivel de significación de un test, es esperable que la potencia aumente.

(**d**) Si no se rechaza la hipótesis nula, debe decidirse aceptar la hipótesis alternativa como cierta.

(**e**) Un error de tipo I ocurre cuando el estadístico del test cae en la región de rechazo del test.

(**f**) Un error de tipo II es más grave que un error de tipo I.

(**g**) La región crítica puede depender de los valores muestrales.

(**h**) El valor de los errores asociados a un test no puede depender de los valores muestrales.

12.7 Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución normal de media $\mu \in \mathbb{R}$ (desconocida) y desvío estándar σ conocido.

(**a**) Diseñar un test de hipótesis de nivel 0.05 para $H_0 : \mu \leqslant \mu_0$ contra $H_1 : \mu > \mu_0$.

(**b**) Determinar la función de potencia del test y analizar sus propiedades.

12.8 En un proceso químico es necesario que una solución que se usa como reactivo tenga pH 8.21. Se dispone de un método para determinar el pH que produce mediciones normalmente distribuidas con media igual al verdadero valor del pH y desvío 0.02. Diseñar un test de hipótesis de manera que: si el pH es 8.21, se obtenga esa conclusión con probabilidad 0.95, y si el pH difiere de 8.21 en 0.03 (en cualquiera de las direcciones), se descarta la solución con probabilidad no inferior a 0.95.

(**a**) ¿Qué se puede concluir si la media de las mediciones observadas fuese 8.32?

(**b**) Si el pH es 8.33, hallar la probabilidad de concluir que el pH no es 8.21.

12.9 Usando los datos de Michelson (ver **Ejercicio 11.4**) testear si la velocidad de la luz es mayor que 790+299000 km/s. al 0.05 de significación.

12.10 En la siguiente tabla se muestran las mediciones de la densidad de la Tierra relativa a la densidad del agua obtenidas por Henry Cavendish en 1789 mediante el experimento de la balanza de torsión

5.50	5.61	4.88	5.07	5.26	5.55	5.36	5.29	5.58	5.65
5.57	5.53	5.62	5.29	5.44	5.34	5.79	5.10	5.27	5.39
5.42	5.47	5.63	5.34	5.46	5.30	5.75	5.68	5.85	

Asumir que las densidades de la Tierra medidas tienen distribución normal de media μ y varianza σ^2 ambas desconocidas y testear al 0.05 de significación si la densidad de la Tierra es menor que 5.5 veces la densidad del agua.

12.11 Sea $X_1, \ldots, X_{10}$ una muestra de aleatoria de una variable aleatoria X con distribución uniforme sobre el intervalo $[0, \theta]$. Al nivel de significación 0.05 diseñar un test para $H_0 : \mathbf{E}[X] = 2$ contra $H_1 : \mathbf{E}[X] \neq 2$.

12.12 El diámetro de las piezas de una máquina debe ser 10 cm. Debido a imperfecciones del proceso de producción de dichas piezas, el diámetro es una variable aleatoria. El desvío estándar depende de factores intrínsecos del proceso de producción y se mantiene constante. Por experiencia histórica se sabe que el desvío estándar es 0.3. Por otra parte, la media μ depende del ajuste de varios parámetros involucrados en el proceso y su valor es desconocido.

(**a**) En una muestra de 100 piezas se obtuvo un diámetro promedio de 10.1. Diseñar un test del 0.1 de nivel de significación.

(**b**) Calcular la potencia del test diseñado en **Inciso** (**a**) en $\mu = 10.05$. .

(**c**) Calcular aproximadamente el tamaño muestral necesario para un nivel de significación del 0.1 y una potencia de 0.8 cuando $\mu = 10.05$.

12.13 Las cajas de leche en polvo de la marca *Spiky Milk* anuncian un peso neto de un kilo. En rigor, el peso neto (en gramos) es una variable aleatoria X de media μ y desvío estándar σ (ambos desconocidos). En una muestra aleatoria de 75 cajas, se observaron los valores $\bar{X} = 992$ y $S = 10$. Al 0.05 de nivel de significación, realizar los siguientes tests:

(**a**) $H_0 : \mu \geqslant 1000$ contra $H_1 : \mu < 1000$.

(**b**) $H_0 : \sigma \geqslant 14$ contra $H_1 : \sigma < 14$.

(**c**) Si los datos muestrales diesen lugar al rechazo de la H_0 en los dos incisos anteriores, ¿es verdad que con 0.05 de significación puede rechazarse la hipótesis $\mu \geqslant 1000$ y $\sigma \geqslant 14$?

12.14 Una fábrica envió catálogos publicitarios a los posibles clientes. La experiencia mostró que la probabilidad de que la organización receptora del catálogo pida el artículo ofrecido es igual a 0.08. La fábrica envió 1000 catálogos nuevos de forma mejorada y recibió 100 pedidos. ¿Se puede considerar que la nueva forma de publicidad es significativamente más eficaz que la primera?

12.15 Usando los resultados del **Ejercicio 11.6 Inciso** (**c**), mostrar que al 10 % de significación

(**a**) no puede rechazarse la hipótesis $p = 0.01$.

(**b**) se rechaza la hipótesis $p \leqslant 0.0009$.

12.16 [Maronna p.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 de 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 se podría concluir?

(**b**) El comité de investigaciones de 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 $\leqslant 0.01$; (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 $\geqslant 0.95$. ¿Cuántas tostadas hay que arrojar para que se cumplan estas condiciones?

12.17 Usando los resultados del **Ejercicio 11.9** decidir al 0.05 de significación si $\lambda \leqslant 0.00019$.

12.18 En vista de los resultados del **Ejercicio 11.11**, ¿puede decirse en cada caso que, con un nivel de significación del 10 %, las resistencias medias de los cables producidos por ambos fabricantes son iguales?

12.19 Considérense las siguientes longitudes de cables (en cm.) obtenidas al muestrear los lotes independientes A y B:

```
Lote A: 134  146  105  119  124  161  107  83  113  129  97  123.
Lote B: 70      118     101     85      107     132     94.
```

Suponer que las longitudes de los cables en dichos lotes tienen una distribución normal con el mismo desvío. ¿Puede afirmarse al 1 % que las medias de los lotes son significativamente diferentes?

12.20 La siguiente tabla muestra para cada uno de 11 individuos, la proporción (como porcentaje) de plaquetas sanguíneas aglutinadas antes y después de fumar un cigarrillo.

Antes	25	25	27	44	30	67	53	53	52	60	28
Después	27	29	37	56	46	82	57	80	61	59	43

Las plaquetas tienen un rol importante en la formación de coágulos. Suponiendo normalidad de los datos ¿qué puede concluirse sobre los efectos del cigarrillo?

12.21 En un examen rindieron 100 alumnos del curso A y 80 del curso B. Del curso A aprobaron 20 y del curso B lo hicieron 5. ¿Considera que los datos muestran, con un nivel de significación del 10 %, que la probabilidad de aprobación depende del curso?

12.22 En una sucesión de 100 lanzamientos independientes de una moneda se observaron 59 caras y 41 cecas ¿Estos datos son conciliables con la hipótesis de que la moneda es honesta?

12.23 Considerar los primeros 1000 decimales del número π:

```
3.14159265358979323846264338327950288419716939937510582097494459230781640
  06286208998628034825342117067982148086513282306647093844609550582231725
  ...
```

Contar la cantidad de veces en que aparece cada dígito y aplicar el test chi-cuadrado al 0.05 de significación para ver si esas frecuencias son compatibles con la hipótesis de que los dígitos son equiprobables.

12.24 Conseguir una computadora que tenga instalado el programa Excel y usando ese programa simular una muestra aleatoria de volumen 100000 de una población normal $N(0,1)$. Aplicar el test chi-cuadrado para verificar si la muestra obtenida es compatible con la hipótesis de que proviene de una distribución normal $N(0,1)$.

12.25 La siguiente tabla muestra la distribución de frecuencias de la duración de ciertas baterías en años:

Clase	1.45-1.95	1.95-2.45	2.45-2.95	2.95-3.45	3.45-3.95	3.95-4.45	4.45-4.95
Frecuencia	2	1	4	15	10	5	3

¿Puede afirmarse a un nivel del 5 % que la duración de las baterías se ajusta a una distribución normal con media 3.5 años y desvío estándar 0.7 años?

12.26 Durante 100 intervalos de un segundo se registraron las cantidades de emisiones de una pieza de material radioactivo. La distribución de frecuencias se muestra en la siguiente tabla:

Cantidad de emisiones	0	1	2	3	4	5	6	7	8	9	10	11
Frecuencia	3	1	2	12	23	13	20	9	6	6	4	1

(**a**) Testear a un nivel de 5 % si los datos provienen de una distribución Poisson de media 5.

(**b**) Testear a un nivel de 5 % si los datos provienen de una distribución Poisson. ¿Cuál es la conclusión?

12.27 [*ver* **Ejercicio 11.4**] ¿A un nivel de 5 %, puede afirmarse que las mediciones de la velocidad de la luz de Michelson se ajustan a una distribución normal?

Ejercicios Complementarios

12.28 Se construye una máquina que controla automáticamente la cantidad de cinta sobre un carrete. La máquina se considera eficaz si el desvío estándar σ de la cantidad de cinta sobre el carrete no supera 0.15 cm. Si en una muestra de 20 carretes se obtiene una estimación de varianza $S^2 = 0.025$, ¿estamos en condiciones de concluir que la máquina es ineficaz?

12.29 Un productor de una nueva línea de neumáticos afirma que su utilidad media es mayor que 40000 km. Para verificar dicha afirmación se sometieron a prueba 12 neumáticos y sus utilidades (en miles de km.) resultaron las siguientes:

```
40.5   40.6   41.3   41.6   42.0   39.8
40.7   40.9   40.2   39.9   41.0   40.6
```

Suponiendo que la utilidad de los neumáticos tiene distribución normal, verificar la afirmación del productor a un 5 % de nivel de significación.

12.30 La duración (en segundos) de las llamadas telefónicas para solicitar un generador de números aleatorios a la empresa *"Random numbers? Llame Ya!"* es una variable aleatoria, X, de media μ y desvío estándar σ (ambos desconocidos). En una muestra aleatoria de la duración de 50 llamadas se observaron los valores $\bar{X} = 310$ y $S = 25$. Al 0.1 de nivel de significación, se puede concluir que

(a) $\mu > 300$?

(b) $\sigma > 20$?

12.31 Se debe comprar un producto cuyo peso obedece a una distribución normal y se requiere que la varianza sea inferior a 4. Si el producto no cumple con el requisito se desea cometer un error en la decisión que no supere el 5 %. Por otro lado, si la varianza fuese igual a 3, se desea que el error en la decisión sea del 5 %. Diseñar un test de hipótesis para decidir si el producto cumple la especificación.

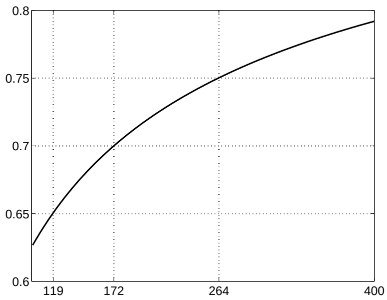

FIGURA 1. Gráfico del cociente entre los fractiles $\chi^2_{p=0.05}$ y $\chi^2_{p=0.95}$ en función de los grados de libertad.

(*Sugerencia*: Utilizar la información que aparece en la Figura 1 y recordar que si Q tiene distribución χ^2 con n grados de libertad, para valores grandes de n se cumple que $W = \sqrt{2Q}$ se distribuye aproximadamente como una normal de media $\sqrt{2n-1}$ y desvío estándar 1.)

12.32 Conseguir una moneda de 5 centavos. Lanzarla 100 veces y aplicar el test chi cuadrado para saber si la moneda es honesta.

12.33 De cierto generador de números aleatorios se afirma que los produce de acuerdo con la distribución $U[0,1]$. Para verificar esa hipótesis se producen 10000 números con el mencionado generador. Para economizar espacio se registra la cantidad de números de la forma $0.d...$, donde $d = 0, 1, \dots, 9$. Se obtuvieron los resultados siguientes:

d	0	1	2	3	4	5	6	7	8	9
$\#\{0.d...\}$	977	1045	1017	1024	901	928	1045	984	1034	1045

¿Estos datos son conciliables con la afirmación?

12.34 Paenza (Página 12, 3 de mayo de 2009, Contratapa) afirma que la distribución del primer dígito significativo, D, de los precios de los productos que se ofrecen en un supermercado se rige por la ley de Benford: $P(D = d) = \log_{10}\left(\frac{d+1}{d}\right)$, $d = 1, \ldots, 9$. Se registraron los precios de 2500 productos ofrecidos por un supermercado y se obtuvieron las siguientes frecuencias:

d	1	2	3	4	5	6	7	8	9
n_d	730	441	304	247	226	160	147	114	131

¿Estos datos son conciliables con la afirmación de Paenza?