Práctico 8
Ley Fuerte y Estimación Paramétrica

LFGN y Montecarlo

Ejercicio 1
Este ejercicio describe el *método de Montecarlo* para el cálculo de integrales.

1. Sean $(U_i)_{i\in\mathbb{N}} \sim \mathcal{U}[a,b]$ *iid* y $f \in R[a,b]$ (f es integrable Riemann en $[a,b]$), mostrar que:

$$\frac{1}{n}\sum_{i=1}^{n} f\left(U_i\right) \xrightarrow[n]{c.s.} \frac{1}{b-a}\int_a^b f\left(x\right)dx.$$

2. Sea D una región arbitraria de $[0,1]\times[0,1]$ y sean $U_1,\ U_2,\ldots,U_n$ variables aleatorias independientes e idénticamente distribuidas con distribución $\mathcal{U}\left([0,1]\times[0,1]\right)$, es decir que se cumple que $\mathbf{P}\left(U\in A\right) = \text{área}(A\cap[0,1]\times[0,1])$.
   Si $a_n = \dfrac{\#\left\{i : 1 \leqslant i \leqslant n \ y \ U_i \in D\right\}}{n}$ probar que $a_n \xrightarrow[n]{c.s.} \text{área}(D)$.

Ejercicio 2
Utilizando algún software de generación de números aleatorios y el método de Montecarlo para resolver integrales, obtenga una estimación de $\displaystyle\int_0^1 \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx$, mediante la generación de 100 números aleatorios. Comparar el resultado obtenido con el provisto por las tablas de la distribución $N(0,1)$.

Ejercicio 3
Calcular el volumen de la región n-dimensional siguiente:

$$R_n = \left\{(x_1,x_2,\ldots,x_n)\in\mathbb{R}^n : 0 \le x_1 \le x_2 \ldots \le x_n \le 1\right\}$$

Ejercicio 4
Demostrar que si $X_n \xrightarrow[n]{c.s.} a$ y $g:\mathbb{R}\to\mathbb{R}$ continua entonces $g(X_n) \xrightarrow[n]{c.s.} g(a)$.

Ejercicio 5
Considérese una red con n terminales comunicados (con $C_2^n = \frac{n(n-1)}{2}$ enlaces). Cada enlace funciona independientemente, con probabilidad p (o falla con probabilidad $1-p$).
Se desea estimar la probabilidad $P(n)$ de que la red permanezca conexa, es decir, que cada par de nodos se pueda alcanzar entre sí mediante enlaces.

1. Sorteando una cantidad N suficientemente grande de tales redes, utilice el método de Montecarlo para hallar un estimador $\overline{P_N}$ para $P(n)$
2. Aplique la técnica anterior para estimar la probabilidad de que una red completa de 50 nodos resulte conexa, cuando $p = 0,99$ y eligiendo el tamaño de muestras N. ¿Encuentra alguna dificultad? Calcular la varianza de $\overline{P_N}$.

Ejercicio 6 Examen diciembre de 2004
Considere una variable aleatoria X absolutamente continua con densidad de probabilidad

$$f_X(x) = \begin{cases} \frac{2a}{x^2} & \text{si } 1 \le x \le b \\ 0 & \text{en caso contrario,} \end{cases} \quad \text{donde } a > \tfrac{1}{2} \text{ y } b > 1. \tag{1}$$

1. Hallar b en función de a.
2. Hallar$E(X)$, $Var(X)$ y m_X, la mediana de X.

Considere ahora una muestra, $X_1, X_2, \ldots, X_{100}$, de variables aleatorias *i.i.d.* con densidad $f_X(x)$.

3. Construya un estimador consistente para el parámetro b. Justifique su respuesta.
4. Suponiendo $a = 1$, estime la probabilidad de que el promedio de la muestra esté en el intervalo $[1{,}8 \log 2, \;\; 2{,}1 \log 2]$.

Estimación Puntual

Definición: Sean $\{X_n\}_{n\in\mathbb{N}}$ *iid* con distribución F_θ donde θ es un parámetro. Se considera la familia $\{T_n(X_1,\ldots,X_n)\}_{n\in\mathbb{N}}$ donde T_n es una función de los n datos (que cumple ciertas hipótesis). $T_n(X_1,\ldots,X_n)$ se llama un estimador de θ. Un estimador se dice *consistente* si $T_n(X_1,\ldots,X_n) \xrightarrow[n]{c.s.} \theta$.

Ejercicio 7
Sean $(X_n)_{n\in\mathbb{N}}$ *iid* tales que $\mathbf{E}(X_1) = \mu$ y $\mathbf{Var}(X_1) = \sigma^2 < \infty$ $(\sigma > 0)$.

1. Demostrar que $\overline{X}_n$ es un estimador consistente de μ, esto es que $\overline{X}_n \xrightarrow[n]{c.s.} \mu$.
2. Demostrar que si $\sigma_n^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$ y $s_n^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$ entonces

$$\sigma_n^2 \xrightarrow[n]{c.s.} \sigma^2 \qquad s_n^2 \xrightarrow[n]{c.s.} \sigma^2 \qquad \sigma_n \xrightarrow[n]{c.s.} \sigma \qquad s_n \xrightarrow[n]{c.s.} \sigma$$

(σ_n^2 y s_n^2 son estimadores consistentes de σ^2 y σ_n y s_n son estimadores consistentes de σ.)

Sugerencia: $\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i)^2 - \left(\overline{X}_n\right)^2$ *y usar los siguientes resultados:*

Si $X_n \xrightarrow[n]{c.s.} X$ *y* $g : \mathbb{R} \to \mathbb{R}$ *es continua entonces* $g(X_n) \xrightarrow[n]{c.s.} g(X)$

Si $X_n \xrightarrow[n]{c.s.} X$ *e* $Y_n \xrightarrow[n]{c.s.} Y$ *y* $g : \mathbb{R}^2 \to \mathbb{R}$ *es continua entonces* $g(X_n, Y_n) \xrightarrow[n]{c.s.} g(X, Y)$

Ejercicio 8
Definimos el *coeficiente de correlación* de dos variables aleatorias X e Y tales que $0 < \mathbf{Var}(X), \mathbf{Var}(Y) < \infty$ mediante

$$\rho = \frac{Cov(X,Y)}{\sqrt{\mathbf{Var}(X)}\sqrt{\mathbf{Var}(Y)}}.$$

Sean $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$ *iid*, tales que $0 < \mathbf{Var}(X_n) = \mathbf{Var}(X) = \sigma^2(X) < \infty$, $0 < \mathbf{Var}(Y_n) = \mathbf{Var}(Y) = \sigma^2(Y) < \infty$. Sea

$$\rho_n = \frac{1}{n}\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)\left(Y_i - \overline{Y}_n\right)}{\sigma_n(X)\,\sigma_n(Y)}$$

con $\sigma_n^2(X) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$ y análogamente con $\sigma_n^2(Y)$. Demostrar que $\rho_n \xrightarrow[n]{c.s.} \rho$.

Sugerencia: $\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)\left(Y_i - \overline{Y}_n\right) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_iY_i - \overline{X}_n\overline{Y}_n$

Ejercicio 9
Sea X una variable con recorrido $R_X = \{1, 2, 3\}$ y probabilidades puntuales $p(X = 1) = a$ y $P(X = 2) = b$. Disponemos de la siguiente muestra de tal variable, que se asume i.i.d.:

$$1, 2, 1, 2, 3, 1, 3, 1.$$

Hallar estimadores consistentes $\overline{a}$ y $\overline{b}$, utilizando el método de los momentos.

Ejercicio 10
Sean $X_1, X_2, \ldots X_n iid \sim F$ Encontrar los estimadores de máxima verosimilitud para los siguientes parámetros:

1. p si la distribución es Ber(p)
2. λ si la distribución es $\mathcal{P}(\lambda)$
3. p si la distribución es Geo(p)
4. μ y σ^2 si la distribución es $N(\mu, \sigma^2)$
5. a y b si la distribución es $\mathcal{U}[a, b]$.

Ejercicio 11
Consideremos una variable aleatoria X con densidad $f_X(t) = \frac{1}{2\sqrt{a(t-1)}}$, si $1 < t < 1+a$, y $f_X(t) = 0$ en caso contrario. Dada la muestra $X_1, \ldots . X_n$ i.i.d. de X, calcular el estimador por máxima verosimilitud para a.

Sesgo de un estimador

Definición: Sean $\{X_n : n \in \mathbb{N}\}$ *iid* con distribución F_θ. Se define el *sesgo* de un estimador de θ, $T_n = T_n(X_1, \ldots, X_n)$ como $\mathbf{E}(T_n - \theta)$. Un estimador T_n se dice *insesgado* si su sesgo es cero, es decir $\mathbf{E}(T_n) = 0 \; \forall \, n \in \mathbb{N}$. Decimos que es *asintóticamente insesgado* si $\mathbf{E}(T_n - \theta) \underset{n}{\rightarrow} 0$.

Ejercicio 12
Sean $(X_n)_{n \in \mathbb{N}}$ *iid* tales que $\mathbf{E}(X_1) = \mu$ y $\mathbf{Var}(X_1) = \sigma^2 < \infty$ $(\sigma > 0)$.

1. Mostrar que $\overline{X}_n$ es insesgado como estimador de μ, que σ_n^2 no es insesgado como estimador de σ^2 y que s_n^2 es insesgado para σ^2.
2. Mostrar que σ_n y s_n no son insesgados como estimadores de σ.
3. Sea $X_1, \ldots, X_n \sim \exp(\lambda)$ *iid*, entonces $\dfrac{1}{\overline{X}_n}$ es consistente como estimador de λ pero no insesgado.

Sugerencia: para las dos últimas partes usar la desigualdad de Jensen[1].

Ejercicio 13
Se considera una muestra $X_1, X_2, ..., X_n$ *iid* con media $\mathbf{E}(X) = \mu$ y varianza $\mathbf{Var}(X) = \sigma^2$. Se considera el estimador $\widehat{\mu} = \sum_{i=1}^{n} a_i X_i$ (una combinación lineal de las observaciones).

1. Hallar la relación que tienen que cumplir los coeficientes a_i para que $\widehat{\mu}$ sea un estimador insesgado de la media μ.
2. Entre todos los estimadores lineales e insesgados de la media μ hallar el de varianza mínima.
   Sugerencia: usar la desigualdad de Cauchy-Schwarz para vectores en $\mathbb{R}^n$.

[1] Desigualdad de Jensen: si $\varphi : \mathbb{R} \to \mathbb{R}$ es una función con $\varphi(x)'' \geqslant 0 \; \forall x$, entonces $\varphi(\mathbf{E}(X)) \leqslant \mathbf{E}(\varphi(X))$. Además el igual se cumple si $\varphi(x)$ es lineal. El enunciado anterior es válido más en general para cualquier φ convexa

Ejercicio 14
Una forma de evaluar la eficiencia de un estimador es a través del *error cuadrático medio*, que se define como $ECM(T_n) = \mathbf{E}\left((T_n - \theta)^2\right)$. Demostrar que si el estimador T_n tiene sesgo a_n entonces $ECM(T_n) = \mathbf{Var}\,(T_n) + a_n^2$.

Ejercicio 15
Sean $X_1, \ldots, X_n$ $iid \sim \mathcal{P}(\lambda)$.

1. Estimar λ por el método de los momentos. Observar que es insesgado y calcular el error cuadrático medio.
2. Probar que s_n^2 también es un estimador insesgado para λ.
3. Encontrar el estimador de máxima verosimilitud para λ y observar que coincide con el estimador obtenido por el método de los momentos.

Ejercicio 16
Sean $X_1, \ldots, X_n$ $iid \sim \mathcal{U}[0, \theta]$. Interesa estimar el valor de θ.

1. Hallar el estimador de θ por el método de los momentos.
2. Estudiar su sesgo, varianza y error cuadrático medio.
3. Demostrar que el estimador de máxima verosimilitud de θ es X_n^*, el máximo de los valores muestrales.

Ejercicio 17 Primer parcial, mayo de 1999
Supongamos que un bolillero contiene N bolillas que pueden ser rojas, blancas o azules. Supongamos que hay r bolillas rojas y b blancas (y por lo tanto $N - (r + b)$ azules), con $r \geqslant 1,\ b \geqslant 1$. Tomamos una muestra de n bolillas elegidas al azar con reposición.

Definimos

X = número de bolillas rojas observadas en la muestra,

Y = número de bolillas blancas observadas en la muestra,

Z = número de bolillas azules observadas en al muestra.

1. Calcular las funciones de probabilidad de X, de Y y de Z
2. Calcular $\mathbf{E}\,(X)\,,\ \mathbf{Var}\,(X)\,,\ \mathbf{E}\,(Y)\,,\ \mathbf{Var}\,(Y)$.
3. Hallar la función de probabilidad de $X + Y$. Calcular $\mathbf{E}\,(X + Y)\,,\ \mathbf{Var}\,(X + Y)$.
4. ¿Son X e Y independientes? Justifique la respuesta.
5. Supongamos que $N = 2000$. Si se repite 2500 veces de manera independiente la experiencia de muestrear 10 bolillas con reposición y se obtiene una cantidad promedio de 3,2 bolillas rojas y 4,6 bolillas blancas, ¿le parece creíble que en el bolillero haya más de 1250 bolillas azules?

Ejercicio 18 Examen diciembre de 2003

Para modelar la cantidad X de productos defectuosos que se encuentran en una línea de producción en determinado período de tiempo se utiliza un modelo de dos parámetros llamado “Poisson con Ceros Forzados” (PCF). Decimos que la variable X tiene distribución de Poisson con ceros Forzados, y se nota $X \sim PCF(\lambda, p)$, si X puede escribirse como $X = YZ$ con Y, Z independientes tales que:

Y tiene distribución de Poisson de parámetro λ ($Y \sim \mathcal{P}\,(\lambda)$)

Z tiene distribución de Bernoulli de parámetro p ($Z \sim \mathrm{Ber}\,(p)$)

El modelo representa que si la producción tiene una muy baja tasa de errores, la probabilidad de que no haya defectuosos es más alta que en un modelo de Poisson tradicional.

1. Hallar $\mathbf{P}(X = k)$ $\forall k \geq 0$.
2. Decimos que $X = 0$ es un "cero forzado" cuando ocurre $Z = 0$. Dado que se observa $X = 0$ hallar la probabilidad de que no sea un "cero forzado" (esto es $Z = 1$).
3. Hallar $\mathbf{E}(X)$, $\mathbf{E}\left(X^2\right)$ y $\mathbf{Var}(X)$.
4. Dada una muestra $X_1, \ldots, X_n$ *iid* $\sim PCF(\lambda, p)$, halle estimadores $\hat{\lambda}$ y $\hat{p}$ de los parámetros por el método de los momentos.
5. Para la siguiente lista de 10 observaciones: 0, 0, 0, 2, 0, 0, 4, 1, 0, 0, estime λ y p por el método de los momentos.