Universidad de Chile
Facultad de Ciencias Físicas y Matemáticas
Departamento de Ingeniería Matemática
MA3402 Estadística

Guía de Ejercicios 2

Profesor: Joaquín Fontbona.
Auxiliares: Catalina Lizana, Fabián Ulloa y Claudio López.

P1. [Hipoteseando] Sean $\beta \in]0,1[$ y $\theta_0 > 0$ dos constantes que suponemos conocidas. Sean $X_1, \ldots, X_n$ variables aleatorias reales i.i.d. cuya distribución tiene la densidad

$$f_\theta(x) = \frac{1}{(1+\beta)\theta} \mathbf{1}_{[-\beta\theta;\theta]}(x),$$

donde $\theta > 0$ es un parámetro desconocido. Queremos probar la hipótesis nula $H0 : \theta \in]0, \theta_0]$ contra la hipótesis alternativa $H1 : \theta > \theta_0$ a un nivel $\alpha \in]0,1[$. El objetivo del ejercicio es construir y estudiar tres tests para estas hipótesis y comparar su potencia.

a) **[Preliminar]** Sean $U_1, \ldots, U_n$ variables uniformes i.i.d. en un intervalo $[a,b]$. Demuestre que la función de distribución F de $Z = \max(U_1, \ldots, U_n)$ se escribe:

$$F(z) = \begin{cases} 0 & \text{si } z < a, \\ \left(\frac{z-a}{b-a}\right)^n & \text{si } a \leq z \leq b, \\ 1 & \text{si } z > b. \end{cases}$$

El primer test está definido por su región crítica:

$$W_1 = \{(x_1, \ldots, x_n) : \max_{1\leq i\leq n} x_i > c_1\}.$$

b) Suponemos $c_1 \in]0, \theta_0[$. Calcule

$$E_1(\theta) = P_\theta\left(\max_{1\leq i\leq n} X_i > c_1\right)$$

para todo $\theta > 0$ y demuestre que $\theta \mapsto E_1(\theta)$ es creciente.

Sea $\gamma = (1+\beta)(1-\alpha)^{1/n}$. Las valores de β, α y n son tales que $\gamma > 1$.

c) Dé la expresión del valor crítico c_1 en función de β, γ y θ_0 y verifique que $c_1 \in]0, \theta_0[$. (Recordar que queremos una prueba de nivel α.)

d) Para $\theta > \theta_0$, sea $\pi_1(\theta)$ la potencia del test 1. Dé la expresión de $\pi_1(\theta)$ en función de θ, β, n y c_1, luego en función de θ, β, n, γ y θ_0.

e) Inspirándose en el test 1, proponga un segundo test utilizando como estadístico de prueba $\min_{1\leq i\leq n} X_i$. Rehaga las tres preguntas anteriores para este segundo test e indique:

- la región crítica del test, W_2
- la expresión del valor crítico c_2 en función de γ y θ_0,
- la expresión de la potencia $\pi_2(\theta)$ en función de θ, β, n, γ y θ_0.

f) ¿Es alguno de estos tests uniformemente más potente que el otro?

Para el tercer test se utiliza el estadístico de test dado por:

$$T_3 = \max\left(\max_{1\leq i\leq n} X_i, -\frac{1}{\beta}\min_{1\leq i\leq n} X_i\right).$$

g) Sea, para $1 \leq i \leq n$, $Y_i = \text{máx}(X_i, -\frac{1}{\beta}X_i)$. Exprese T_3 en función de los Y_i.

h) Suponemos que el valor crítico del test, denotado c_3, vive en $]0, \theta_0[$. Para $\theta > 0$, exprese $E_3(\theta) = P(T_3 > c_3)$ en función de la función de distribución F_θ asociada a la densidad f_θ.

i) Dé la expresión del valor crítico c_3. Verifique que $c_3 \in]0, \theta_0[$.

j) Calcule $\pi_3(\theta)$ en función de θ, β, n, γ y θ_0. Compare los tres tests.

P2. ($\star$) Recordamos la definición de una Ley multinomial de orden K. Sea $N \in \mathbb{N}^*$, $p \in]0, 1[^K$ tal que $\sum_{i=1}^K p_i = 1$. Llamamos ley multinomial de parámetros (N, p) a la ley de probabilidad en $\{0, 1, \ldots, N\}^K$ definida por la función de masa:

$$P(x_1, \ldots, x_K) = \begin{cases} \frac{N!}{\prod_{i=1}^K x_i!} \prod_{i=1}^K p_i^{x_i}, & \text{si } (x_1, \ldots, x_K) \in 0, 1, \ldots, N^K \text{ tal que } \sum_{i=1}^K x_i = N, \\ 0, & \text{en otro caso.} \end{cases}$$

Denotamos $X \sim M(N, p)$.

Recordamos también la definición de una ley de Dirichlet de orden K. Sea $a = (a_1, \ldots, a_K) \in \mathbb{R}_+^K$. Llamamos ley de Dirichlet de parámetro a a la ley de probabilidad con soporte $S = \{x \in [0, 1]^K : \sum_{i=1}^K x_i = 1\}$, definida por la densidad:

$$p(x_1, \ldots, x_K) = \begin{cases} \frac{1}{\beta(a)} \prod_{i=1}^K x_i^{a_i - 1}, & \text{si } (x_1, \ldots, x_K) \in S, \\ 0, & \text{en otro caso.} \end{cases}$$

Denotamos esta ley como $\text{Dir}(a)$. La ley de Dirichlet de orden 2 es la ley Beta, $\text{Dir}(a_1, a_2) = \text{Beta}(a_1, a_2)$.

a) Sin realizar el cálculo, indicar cómo determinar la función $\beta(a)$. Se supone conocida para el resto del ejercicio.

b) Sea Y una variable aleatoria que sigue una ley multinomial de orden K, $K \geq 3$, y de parámetros (N, θ), donde N es conocido e $\theta = (\theta_1, \ldots, \theta_K)$ es desconocido. Sea $y = (y_1, \ldots, y_K)$ una observación de la variable Y. Nos encontramos en el marco de la estimación bayesiana para θ. Suponemos la distribución a priori $\pi = \text{Dir}(a)$, con $a = (a_1, \ldots, a_K) \in \mathbb{R}_+^K$. Determinar la ley de la distribución a posteriori $p(\theta|y)$.

c) 1) Demuestre que si $(X_1, \ldots, X_{K-1}, X_K)$ sigue una ley de Dirichlet de parámetros $(a_1, \ldots, a_{K-1}, a_K)$, entonces $(X_1, \ldots, X_{K-2}, X_{K-1} + X_K)$ sigue una ley de Dirichlet de parámetros $(a_1, \ldots, a_{K-2}, a_{K-1} + a_K)$.

2) Denotamos $a_r = \sum_{i=3}^K a_i$ y $y_r = \sum_{i=3}^K y_i$. Deducir de la pregunta anterior que $p(\theta_1, \theta_2|y) \propto \theta_1^{a_1 + y_1 - 1} \theta_2^{a_2 + y_2 - 1} (1 - \theta_1 - \theta_2)^{a_r + y_r - 1}$.

P3. ($\star$) **[¿Y si no puedo invertir?]** Sea X una variable aleatoria con valores en $\mathbb{R}^p$ e Y una variable aleatoria con valores en $\mathbb{R}$. Suponemos que la ley de X es conocida. Sean N pares independientes de observaciones $(x_i, y_i)_{1 \leq i \leq N}$. Nos interesamos en la regresión lineal de Y dado X:

$$Y = \beta_0 + \beta^T X + \xi, \;\; \textit{donde } \; \xi \sim \mathcal{N}(0, \sigma^2) \;\; \textit{y es independiente de } \; X$$

Suponemos que $\overline{x} = \frac{1}{n} \sum_{i=1}^n x_i = 0$ e $\overline{y} = \frac{1}{n} \sum_{i=1}^n y_i = 0$. ¿Podemos ponernos siempre bajo este supuesto?

Denotamos x_{ij} como la j-ésima componente de las variables explicativas de la muestra i, x_i. Definimos la *matriz de design* del problema de regresión y el vector de *labels* como:

$$X = \begin{pmatrix} x_{11} & \ldots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{N1} & \ldots & x_{Np} \end{pmatrix} \;\; y \;\; Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$$

a) Recuerde el problema de estimación de (β_0, β) en este contexto. Demuestre que bajo nuestros supuestos, necesariamente $\beta_0 = 0$. Podemos asumir esto para los items siguientes.

b) Si $N > p$ recuerde cómo es la solución explícita al problema de los mínimos cuadrados.

c) Sea $p > N$. En este caso, la regresión lineal clásica no es posible (¿por qué?). Para solucionar este problema consideraremos el problema de *regresión **ridge*** (también se le dice *con penalización* $L2$). Es decir, para $\lambda > 0$, buscaremos:

$$\hat{\beta}_{\text{ridge}} \in \arg\min_{\gamma} ||X\gamma - Y||_2^2 + \lambda||\gamma||_2^2$$

Explique por qué esto nos permite resolver el problema y encuentre la solución explícita (demostrando, de paso, que es única).

P4. [**¿Regresión Lineal?**] Queremos explicar la relación entre una v.a. Y que toma valores en $\{0, 1\}$ (e.g. para una *clasificación binaria*) y un vector de covariables X a valores en $\mathbb{R}^{p+1}$ (con la primera coordenada estando fija en 1). Aplicar un modelo de regresión lineal clásico, $Y_i = \beta^T X_i + \varepsilon_i$ (con ε_i centrados), NO parece ser la mejor idea, pues *el lado derecho es, en principio, **no acotado***.

Lo que sí podemos hacer es *recurrir a los **modelos lineales generalizados***. Por ejemplo, consideremos la función **logit**: $]0, 1[\to \mathbb{R} : \pi \mapsto \log\left(\frac{\pi}{1-\pi}\right)$ (cuya inversa es la función *sigmoide*); y modelaremos la relación entre los Y y los X como: $\pi_i := \mathbb{P}(Y_i = 1|X_i) = \textbf{logit}^{-1}(\beta_0 + \beta^T X_i)$. Es decir: $\textbf{logit}(\mathbb{E}[Y|X]) = \beta_0 + \beta^T X$ (es *casi* un modelo lineal).

Sea $(X_1, Y_1), \ldots, (X_n, Y_n)$ una muestra i.i.d. de estas v.a.s. Asumimos que los $(X_i)_i$ son todos fijos:

a) Demuestre que la log-verosimilitud (condicional) de la muestra en función de β se escribe como:

$$\ell(\beta|(X_i, Y_i)_i) = \sum_{i=1}^{n} \left[Y_i X_i^T \beta - \log\left(1 + \exp\left(X_i^T \beta\right)\right)\right]$$

b) No existe una expresión analítica para el EMV. Se debe utilizar un algoritmo de optimización (*Newton-Raphson*), para aproximar el óptimo mediante una secuencia $(\beta^{(k)})_{k\in\mathbb{N}}$ dada por:

$$\beta^{(k+1)} = \beta^{(k)} - [H(\beta^{(k)})]^{-1} U(\beta^{(k)})$$

donde $S(\beta^{(k)}) = \frac{\partial \ell}{\partial \beta}$ y $H(\beta^{(k)})$ es tal que $H(\beta^{(k)})_{i,j} = \frac{\partial \ell^2}{\partial \beta_i \partial \beta_j}$.

Dé una expresión explícita de $U(\beta^{(k)})$ y $H(\beta^{(k)})$ en el contexto del modelo logístico.

P5. Consideremos $(x_1, Y_1), \ldots, (x_n, Y_n) \in \mathbb{R}^p \times \mathbb{R}$ observaciones (ejemplos) en las que se supone que siguen un modelo lineal de la forma:

$$Y_i = \beta_0 + \beta^T x_i + \epsilon_i, \quad (1)$$

donde $\epsilon_1, \ldots, \epsilon_n \sim \mathrm{N}(0, \sigma^2)$. Se supone que los x_i son deterministas, $\beta_0 \in \mathbb{R}$, $\beta \in \mathbb{R}^p$ son parámetros desconocidos, y $\sigma^2 > 0$ se supone conocido. Para simplificar las notaciones, se incluirá un "1" al principio del vector de covariables, para poder escribir el modelo de la forma:

$$Y_i = \beta^T x_i + \epsilon_i, \quad (2)$$

donde el vector β ahora tiene tamaño $p + 1$ (y β_0 corresponde a su primera coordenada). Definimos la *matriz de design* A como:

$$A = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix} \in \mathbb{R}^{n\times(p+1)}, \quad (3)$$

y suponemos que A tiene rango $p+1$.

a) Demuestre que A^TA es definida positiva. Recuerde la expresión del estimador de mínimos cuadrados, que se denotará como $\hat{\beta}$, y justifique que está bien definido.

b) Sea R_n el riesgo empírico, que está dado por:

$$\hat{R}_n = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{\beta}^T x_i\right)^2$$

Demuestre que

$$\hat{R}_n = \frac{1}{n} Y^T Q Y$$

donde Q es una matriz de un proyección ortogonal que usted debe determinar (en particular, debe explicitar su imagen y su rango).

c) Demuestre que

$$\mathbb{E}[\hat{R}_n] = \sigma^2\left(1 - \frac{p+1}{n}\right)$$

P6. ($\star$) [**Triangulando**] Se miden los ángulos $\theta_1, \theta_2, \theta_3$ de un triángulo, obteniéndose las observaciones y_1, y_2, y_3. Las mediciones tienen errores normales, centrados, de varianza común σ^2 e independientes. Esto permite plantear el modelo lineal $y_i = \theta_i + \varepsilon_i$ $(i = 1,2,3)$ sujeto a la restricción $\theta_1 + \theta_2 + \theta_3 = \pi$.

a) Muestre que el estimador de mínimos cuadrados de θ_i, tomando en cuenta la restricción, está dada por:

$$\hat{\theta}_i = y_i - \left(\bar{y} - \frac{\pi}{3}\right)$$

b) Se sospecha que el triángulo tiene dos posibles combinaciones de ángulos: $(\pi/3, \pi/4, 5\pi/12)$ y $(\pi/2, \pi/3, \pi/6)$. Escriba en detalle el modelo paramétrico del problema y plantee, en términos de los parámetros del problema, y las hipótesis H_0 y H_1.

c) Suponga σ conocido y desarrolle el Test de Neyman Pearson de nivel α, explicitando la región crítica.

P7. ($\star$) [**Testeando Cosillas**] Considere un modelo en el que

$$Y_i = \begin{cases} x_i\beta_1 + \varepsilon_i, & \text{si } i = 1,\dots,n_1, \\ x_i\beta_2 + \varepsilon_i, & \text{si } i = n_1+1,\dots,n_1+n_2 = n, \end{cases}$$

con $\varepsilon_1,\dots,\varepsilon_n$ i.i.d. de $N(0,\sigma^2)$ y variables independientes $x_1,\dots,x_n$ tomadas como constantes conocidas.

a) Escriba este modelo como un modelo lineal.

b) Encuentre estimadores insesgados para β_1 y β_2.

c) Considere la cantidad $\xi = X^T\beta$, siendo X alguna matriz de diseño y $\beta = (\beta_1, \beta_2)^T$ y

$$s^2 = \frac{||Y - \xi||^2}{n-2}$$

¿qué distribución sigue s^2?

Ahora se decide llevar a cabo el siguiente test de nivel $1-\alpha$:

$$H_0\text{: } \beta_1 = \beta_2 \text{ v/s } H_1 : \beta_1 \neq \beta_2$$

Para ello asumamos H_0, denotando a β_0 el parámetro que tiene igual valor que β_1 y β_2 y siga el siguiente esquema:

d) Sea X_0 la matriz de diseño bajo H_0 y $\xi = X_0^T\beta_0$, calcule $||\xi - \xi_0||^2$. ¿Qué interpretación le da a esta cantidad? ¿Qué pasaría con esta cantidad si H_0 fuera cierta?

e) Con s^2 y $||\xi - \xi_0||^2$ construya un estadístico que defina el p-value del test y su región crítica de rechazo, explicite ambas. *Indicación: Si $S_1 \sim \chi^2_{d_1}$, $S_2 \sim \chi^2_{d_2}$, entonces*

$$\frac{S_1/d_1}{S_2/d_2} \sim F_{d_1,d_2}$$

donde F_{d_1,d_2} es ley Fisher de grados de libertad d_1, d_2, asuma conocida esta distribución para todo d_1, d_2.

P8. ($\star$) [**Bayesiando**] Sea $X_n = (X_1, \ldots, X_n)$ una muestra de n observaciones independientes e idénticamente distribuidas según la ley de Poisson con parámetro $\theta > 0$.

a) Demuestre que el estimador de máxima verosimilitud de θ, denotado como $\hat{\theta}_n^{\text{EMV}}$, está bien definido y dé su expresión.

b) Verifique que el vector de *score* está centrado, calcule la cota de Cramér-Rao y demuestre que $\hat{\theta}_n^{\text{EMV}}$ es un estimador insesgado eficiente.

Ahora, el parámetro desconocido θ se distribuye a priori como Gamma(α_0, β_0), donde α_0 y β_0 son números reales estrictamente positivos dados:

$$\pi(\theta) \propto (\beta_0\theta)^{\alpha_0-1}e^{-\beta_0\theta}\mathbf{1}_{\{\theta>0\}}.$$

c) Demuestre que la familia de leyes gamma es conjugada para el parámetro θ.

d) Determine el estimador bayesiano óptimo para la pérdida cuadrática, denotado como $\hat{\theta}_n^{\text{Bayes}}$. (Se puede utilizar sin demostración el resultado general visto en clase).

Volviendo a un enfoque "frecuentista" para las cuatro últimas preguntas, lo que significa que nos interesamos en las propiedades de los estimadores $\hat{\theta}_n^{\text{EMV}}$ y $\hat{\theta}_n^{\text{Bayes}}$ para un $\theta > 0$ fijo.

e) Demuestre que el estimador bayesiano $\hat{\theta}_n^{\text{Bayes}}$ es sesgado independientemente de la elección de los parámetros α_0 y β_0, pero asintóticamente insesgado.

f) Calcule el riesgo cuadrático de cada uno de los dos estimadores, $\hat{\theta}_n^{\text{EMV}}$ y $\hat{\theta}_n^{\text{Bayes}}$.

g) Demuestre que ninguno de los dos estimadores, $\hat{\theta}_n^{\text{EMV}}$ y $\hat{\theta}_n^{\text{Bayes}}$, es preferible al otro en términos de riesgo cuadrático.

h) Demuestre que el estimador bayesiano $\hat{\theta}_n^{\text{Bayes}}$ es fuertemente consistente y asintóticamente gaussiano, especificando su varianza asintótica. Compare con las propiedades asintóticas de $\hat{\theta}_n^{\text{EMV}}$. (Indicación: observe que $\hat{\theta}_n^{\text{Bayes}}$ tiene la forma $a_n + b_nX_n$, con $a_n = O(1/n)$ y $b_n = 1 + O(1/n)$).

P9. [Esto NO es Lineal] En este ejercicio, suponemos que observamos n variables aleatorias:

$$Y_i = f(x_i) + \varepsilon_i, \quad 1 \leq i \leq n, \quad (1)$$

donde los puntos x_i son deterministas, espaciados regularmente entre $1/n$ y 1:

$$x_i = \frac{i}{n}, \quad (2)$$

los ε_i son i.i.d. $N(0, \sigma^2)$ con varianza $\sigma^2 > 0$ conocida, y $f : [0,1] \to \mathbb{R}$ tiene la forma:

$$f(x) = a_0 + \sum_{k=1}^{K} a_k \cos(2\pi k x), \quad (3)$$

con $a_0, a_1, \ldots, a_K$ siendo parámetros reales desconocidos.

a) Bajo estas hipótesis, escriba el modelo (1)–(3) como un modelo lineal gaussiano $Y_i = (X\beta)_i + \varepsilon_i$ especificando X y β.

b) Suponemos ahora que $K+1 \leq n$. Verifique que el estimador de mínimos cuadrados $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_2, \ldots, \hat{\beta}_K)$ de β está bien definido y demuestre que es igual a:

$$\begin{cases} \hat{\beta}_0 = \frac{1}{n} \sum_{i=1}^{n} Y_i = \bar{Y}_n, \\ \hat{\beta}_k = \frac{2}{n} \sum_{i=1}^{n} Y_i \cos\left(\frac{2\pi k i}{n}\right). \end{cases}$$

c) De ello, obtenga un estimador $\hat{\mu}$ de $\mu = (f(i/n))_{1 \leq i \leq n}$ y proponga finalmente un estimador $\hat{f}$ de la función f.

d) Calcule la esperanza de la suma de los residuos al cuadrado normalizados $r_n = \frac{1}{n}\mathbb{E}\left[\sum_k (Y - X\hat{\beta})_k^2\right]$. ¿Qué sucede cuando K está fijo y n tiende a infinito?

e) Suponemos ahora que $n = K+1$. Dé el valor de r_n. ¿Qué se puede decir de la función $\hat{f}$ en los puntos $x_i = i/n, 1 \leq i \leq n$?