Entrega ejercicio Temas 2 y 3

Jose Antonio Lorencio Abril

Curso 21/22

Ejercicio 1. Sea $(X_1, ..., X_n)$ una m.a.s. de una población $X \sim N(\mu, \sigma^2)$. Demostrar que $\overline{X}$ y $(X_1 - \overline{X}, ..., X_n - \overline{X})$ son independientes.
Demostrar a partir de este resultado que $\overline{X}$ y S^2 son independientes.
Primero, notamos que, para todo $j = 1, ..., n$, es

$$\overline{X} - X_j = \sum_{i=1}^{n} \frac{X_i}{n} - X_j = \frac{X_1 + ... + X_{j-1} + X_{j+1} + ... + X_n}{n} - \frac{n-1}{n} X_j$$

donde el primero de los sumandos es $N\left(\frac{n-1}{n}\mu, \frac{n-1}{n^2}\sigma^2\right)$ y el segundo es $N\left(\frac{n-1}{n}\mu, \left(\frac{n-1}{n}\right)^2 \sigma^2\right)$. Por tanto,

$$\overline{X} - X_j \sim N\left(0, \left(\frac{n-1}{n^2} + \frac{(n-1)^2}{n^2}\right)\sigma^2\right) = N\left(0, \frac{n-1}{n}\sigma^2\right)$$

para todo $j = 1, ..., n$.
Por otro lado

$$Cov\left(X_j - \overline{X}, \overline{X}\right) = Cov\left(X_j, \overline{X}\right) - Cov\left(\overline{X}, \overline{X}\right) = E\left(X_j \overline{X}\right) - E\left(X_j\right) E\left(\overline{X}\right) - Var\left(\overline{X}\right) =$$

$$= E\left(X_j \frac{1}{n} \sum_{1}^{n} X_i\right) - \mu^2 - \frac{\sigma^2}{n} = \frac{1}{n} \sum_{1}^{n} E\left(X_i X_j\right) - \mu^2 - \frac{\sigma^2}{n} = \frac{1}{n} \sum_{i \neq j} E\left(X_i\right) E\left(X_j\right) + \frac{1}{n} E\left(X_j^2\right) - \mu^2 - \frac{\sigma^2}{n} =$$

$$= \frac{(n-1)\mu^2 + Var\left(X_j^2\right) + E\left(X_j\right)^2 - n\mu^2 - \sigma^2}{n} = \frac{-\mu^2 + \sigma^2 + \mu^2 - \sigma^2}{n} = 0$$

por lo tanto, tenemos que $\left(\overline{X}, X_1 - \overline{X}, ..., X_n - \overline{X}\right)$ es una distribución normal multidimensional de la forma

$$N\left((\mu, 0, 0, ..., 0), \begin{pmatrix} \frac{\sigma^2}{n} & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \Sigma & \\ 0 & & & \end{pmatrix}\right)$$

para alguna matriz de covarianzas Σ, que no necesitamos determinar.
Por tanto, dado que en la normal multivariante, la incorrelación implica independencia, tenemos que las variables aleatorias $\overline{X}$ y $\left(X_1 - \overline{X}, ..., X_n - \overline{X}\right)$ son independientes, como queríamos ver.
Ahora, para la última parte, notamos que la función

$$\begin{array}{cccc} g: & \mathbb{R}^n & \to & \mathbb{R} \\ & (x_1, ..., x_n) & \mapsto & \frac{\sum x_i^2}{n} \end{array}$$

es medible Borel, y que

$$S^2 = \frac{\sum_{i=1}^{n}\left(X_j - \overline{X}\right)^2}{n} = g\left(X_1 - \overline{X}, ..., X_n - \overline{X}\right)$$

por lo que S^2 es una transformación de $\left(X_1 - \overline{X}, ..., X_n - \overline{X}\right)$ mediante una función medible Borel, y como esta variable aleatoria es independiente de $\overline{X}$, entonces también lo es S^2.

Ejercicio 2. Sea X una variable aleatoria con distribución de Poisson, de parámetro λ ($X \sim P(\lambda)$) y sea $(X_1, ..., X_n)$ una m.a.s. de X.

a) Obtener las condiciones de regularidad de Cramer-Rao para el estadístico $\overline{X}$

i) $\Psi = sop(\boldsymbol{X}) = \{0, 1, ...\}^n$ independiente de λ

ii) Calculamos la función de verosimilitud

$$L(\boldsymbol{x}, \lambda) = \prod_1^n p(x_i, \lambda) = \prod_1^n \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod x_i!}$$

y esta función es claramente derivable respecto de $\lambda, \forall\lambda > 0, \forall\boldsymbol{x} \in sop(\boldsymbol{X})$, pues es el producto de una exponencial por una potenciación.

iii) Primero, definamos $Y = \sum_1^n X_i$, de forma que

$$\varphi_Y(t) = \prod_1^n \varphi_{X_i}(t) = \prod_1^n e^{\lambda\left(e^{it}-1\right)} = e^{n\lambda\left(e^{it}-1\right)}$$

y entonces $Y \sim P(n\lambda)$, por lo tanto, $\overline{X} = \frac{Y}{n} \sim P(\lambda)$. Por tanto, el problema se simplifica y podemos trabajar con la función puntual de probabilidad de la Poisson, en lugar de con la de verosimilitud. Así, tenemos

$$E\left[\overline{X}\right] = \sum_0^\infty i \cdot p_{\overline{X}}(i, \lambda) = \sum_0^\infty i\frac{e^{-\lambda}\lambda^i}{i!}$$

y para que $\sum_n g_n(x)$ sea derivable bajo el signo de la serie, basta que $\sum_n g_n'(x)$ converja uniformemente. Ahora bien, la derivada del interior es

$$\frac{\partial}{\partial\lambda}g_i(\lambda) = i\frac{-e^{-\lambda}\lambda^i}{i!} + i\frac{e^{-\lambda}i\lambda^{i-1}}{i!} = -i\frac{e^{-\lambda}\lambda^i}{i!} + i^2\frac{e^{-\lambda}\lambda^{i-1}}{i!}$$

y entonces

$$\sum_0^\infty \frac{\partial}{\partial\lambda}g_i(\lambda) = \sum_0^\infty -ip_{\overline{X}}(i, \lambda) + \frac{i^2}{\lambda}p_{\overline{X}}(i, \lambda) = -E\left[\overline{X}\right] + \frac{1}{\lambda}E\left[\overline{X}^2\right]$$

con la convergencia uniforme pues

$$\sum_{i=\lambda}^\infty \left|\frac{\partial}{\partial\lambda}g_i(\lambda)\right| = \sum_{i=\lambda}^\infty \left|i\frac{e^{-\lambda}\lambda^{i-1}}{i!}\left[-\lambda + i\right]\right| \leq \sum_{i=\lambda}^\infty \left|i^2\frac{e^{-\lambda}\lambda^{i-1}}{i!}\right| = e^{-\lambda}\sum_{i=\lambda}^\infty \left|i\frac{\lambda^{i-1}}{(i-1)!}\right| \overset{K_\lambda > \lambda}{\leq} e^{-\lambda}\sum_{i=\lambda}^\infty \frac{iK_\lambda^{i-1}}{(i-1)!}$$

Ahora bien:

$$\sum_{i=2}^\infty \frac{iK^{i-1}}{(i-1)!} = \sum_{i=1}^\infty \frac{(i+1)K^i}{i!} = \sum_{i=1}^\infty \frac{K^i}{(i-1)!} + \sum_{i=1}^\infty \frac{K^i}{i!} =$$

$$= K\sum_{i=1}^\infty \frac{K^{i-1}}{(i-1)!} + e^K - 1 = K\sum_{i=0}^\infty \frac{K^i}{i!} + e^K - 1 = Ke^K + e^K - 1$$

entonces la serie anterior converge independientemente del K_λ que tomemos, pues aunque empezamos en $i = \lambda$, la cantidad finita de términos iniciales no nos importa. El criterio M de Weierstrass nos proporciona la convergencia uniforme de las derivadas buscadas (Suponiendo que el dominio de λ está acotado, para poder tomar un K que acote a todos los λ buscados.

Esto es razonable, de todas formas, ya que la cota podemos tomarla arbitrariamente alta y, en consecuencia, nos servirá para cualquier λ) y tenemos la derivabilidad bajo el signo de la serie de $E\left[\overline{X}\right]$.

Falta ver que $\sum ... \sum_{\Psi} L(\boldsymbol{x}, \lambda)$ es derivable bajo el signo de la serie, pero para esto basta ver que $\sum_x p(x, \lambda) = \sum_i \frac{e^{-\lambda}\lambda^i}{i!}$ lo es, por el siguiente criterio de convergencia:

Si $f_n \overset{unif}{\to} f$ y $h_n \overset{unif}{\to} h$ y existen constantes con $|f_n(x)| \leq M_1, |h_n(x)| \leq M_2, \forall x$, entonces $f_n h_n \overset{unif}{\to} fh$.

Para verlo, calculamos

$$\frac{\partial}{\partial\lambda} p_i(\lambda) = \frac{-e^{-\lambda}\lambda^i + e^{-\lambda} i \lambda^{i-1}}{i!} = \frac{e^{-\lambda}\lambda^{i-1}}{i!}\left[-\lambda + i\right]$$

entonces, razonando como antes, obtenemos

$$\left|\frac{\partial}{\partial\lambda} g_i(\lambda)\right| \leq \frac{iK_\lambda^{i-1}}{i!} = \frac{K_\lambda^{i-1}}{(i-1)!}$$

Esta serie da e^{K_λ}. Así, la serie de las derivadas converge uniformemente por el criterio M de Weierstrass (sobre acotados en λ).

Ahora aplicamos el criterio mencionado tomando $f_n = h_n = \sum_{i=1}^n \frac{\partial}{\partial\lambda} p_i(\lambda)$ con la cota e^K. Haciendo esto n veces obtenemos la convergencia uniforme de la derivada de la función de verosimilitud (sobre acotados en λ).

iv)

$$E\left[\left(\frac{\partial \log L(\boldsymbol{X}, \lambda)}{\partial\lambda}\right)^2\right] < \infty$$

$$\log L(\boldsymbol{X}, \lambda) = -n\lambda + \sum x_i \log\lambda - \sum \log x_i!$$

$$\frac{\partial}{\partial\lambda} L(\boldsymbol{x}, \lambda) = -n + \frac{\sum x_i}{\lambda}$$

$$\left(\frac{\partial}{\partial\lambda} L(\boldsymbol{x}, \lambda)\right)^2 = n^2 + \frac{\left(\sum x_i\right)^2}{\lambda^2} - 2\frac{n}{\lambda}\sum x_i = n^2 + \frac{\sum x_i^2 + \sum_{i\neq j} x_i x_j}{\lambda^2} - 2\frac{n}{\lambda}\sum x_i$$

$$E\left[\left(\frac{\partial \log L(\boldsymbol{X}, \lambda)}{\partial\lambda}\right)^2\right] = n^2 + \frac{1}{\lambda^2}\sum E\left[X^2\right] + \frac{1}{\lambda^2}\sum_{i\neq j} E\left[X_i X_j\right] - 2\frac{n}{\lambda}\sum E\left[X\right] =$$

$$= n^2 + \frac{\sum\left(Var\left[X\right] + E\left[X\right]^2\right)}{\lambda^2} + \frac{\sum_{i\neq j} E\left[X_i\right] E\left[X_j\right]}{\lambda^2} - \frac{2n \cdot n\lambda}{\lambda} =$$

$$= \frac{n\lambda + n\lambda^2 + \left(n^2 - n\right)\lambda^2 - n^2\lambda^2}{\lambda^2} = \frac{n\lambda + \cancel{n\lambda^2} + \cancel{n^2\lambda^2} - \cancel{n\lambda^2} - \cancel{n^2\lambda^2}}{\lambda^2} = \frac{n}{\lambda}$$

que es finito.

b) Obtener la cantidad de información de Fisher

Ya lo hemos hecho, pero lo hacemos de otra forma:

$$\log p(x, \lambda) = -\lambda + x\log\lambda - \log x!$$

$$\frac{\partial}{\partial\lambda}\log p(x, \lambda) = -1 + \frac{x}{\lambda}$$

$$E\left[\left(\frac{\partial}{\partial\lambda}\log p\left(X,\lambda\right)\right)^2\right] = E\left[1+\frac{X^2}{\lambda^2}-2\frac{X}{\lambda}\right] = 1+\frac{1}{\lambda^2}E\left[X^2\right]-\frac{2}{\lambda}E\left[X\right] =$$

$$= 1+\frac{Var\left[X\right]+E\left[X\right]^2}{\lambda^2}-\frac{2}{\lambda}\lambda = \frac{\lambda+\lambda^2}{\lambda^2}-1 = \frac{\lambda+\lambda^2-\lambda^2}{\lambda^2} = \frac{1}{\lambda}$$

Y, como X verifica las condiciones de regularidad de Cramer-Rao, entonces es

$$I_n\left(\lambda\right) = nE\left[\left(\frac{\partial}{\partial\lambda}\log p\left(X,\lambda\right)\right)^2\right] = \frac{n}{\lambda}$$

c) Estudiar si el estadístico $\overline{X}$ es EIMV de λ

Primero, como $E\left[\overline{X}\right] = E\left[X\right] = \lambda$, sabemos que es un estimador insesgado. Pero, además, como

$$Var\left[\overline{X}\right] = \frac{Var\left[X\right]}{n} = \frac{\lambda}{n} = \frac{1}{I_n\left(\lambda\right)}$$

Y entonces el estimador alcanza la cota de Cramer-Rao, por lo que es mínima varianza. En conclusión, $\overline{X}$ es EIMV de λ.

Ejercicio 3. Sea $\boldsymbol{P} = \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T$, entonces se verifica:

1. $\boldsymbol{P}, \boldsymbol{I}_n - \boldsymbol{P}$ son simétricas e idempotentes
2. $r\left(\boldsymbol{I}_n - \boldsymbol{P}\right) = tr\left(\boldsymbol{I}_n - \boldsymbol{P}\right) = n - k$
3. $\left(\boldsymbol{I}_n - \boldsymbol{P}\right)X = 0$

Para ver 1., primero vamos a ver que si A es invertible, entonces $A^{T^{-1}} = A^{-1^T}$:

$$A^T A^{-1^T} = \left(A^{-1}A\right)^T = I^T = I$$

$$A^{-1^T} A^T = \left(AA^{-1}\right)^T = I^T = I$$

como afirmábamos. Ahora:

$$P^T = \left(X\left(X^TX\right)^{-1}X^T\right)^T = X^{T^T}\left(X^TX\right)^{-1^T}X^T = X\left(X^TX^{T^T}\right)^{-1}X^T = X\left(X^TX\right)^{-1}X^T = P$$

por lo que P es simétrica.
$I_n - P$ es simétrica, por ser suma de simétricas.
Para ver la idempotencia:

$$P^2 = \left(X\left(X^TX\right)^{-1}X^T\right)\left(X\left(X^TX\right)^{-1}X^T\right) = X\cancel{\left(X^TX\right)^{-1}\left(X^TX\right)}\left(X^TX\right)^{-1}X^T = X\left(X^TX\right)^{-1}X^T$$

$$\left(I_n - P\right)^2 = I_n^2 - I_nP - PI_n + P^2 = I_n - 2P + P = I_n - P$$

2. Las matrices simétricas son diagonalizables, o sea, si A es simétrica entonces existe Q ortogonal con $A = QDQ^T$ donde D es una diagonal de valores propios. Además, las matrices idempotentes tienen todos su valores propios 0 o 1, puesto que $\left(QDQ^T\right)^2 = QDQ^TQDQ^T = QD^2Q^T = A^2 = A = QDQ^T \implies D^2 = D$ y los valores propios son 0 o 1. Así, si vemos que $I_n - P$ tiene rango $n-k$, tendremos el resultado por la invarianza de la traza. Ahora bien $I_n = QI_nQ^T$, por lo que $I_n - P = Q\left(I_n - D\right)Q^T$ y entonces la traza será $n - m$, donde m es la cantidad de 1 que hay en D. Por tanto, si vemos que P tiene traza k, o equivalentemente que P tiene rango k, tendremos el resultado.

$$P = X\left(X^TX\right)^{-1}X^T$$

Nótese que $P \in \mathcal{M}_{n\times n}$ y $X^T \in \mathcal{M}_{k\times n}$, veamos que $Ker\left(X^T\right) = Ker\left(P\right)$ y que esto implica $r\left(P\right) = k$, como queremos.
$\left[Ker\left(X^T\right) = Ker\left(P\right)\right]$:
($\subset$) Si $X^Ta = 0$, entonces $Pa = X\left(X^TX\right)^{-1}X^Ta = 0$.
($\supset$) Si $Pa = 0$, $\implies X^TPa = \cancelto{1}{X^TX\left(X^TX\right)^{-1}}X^Ta = 0$.
Luego $Ker\left(X^T\right) = Ker\left(P\right) = K$.
En general, se tiene que si $A \in \mathcal{M}_{m\times n}\left(\mathbb{R}\right)$, entonces $n = r\left(A\right) + dim\left(Ker\left(A\right)\right)$.
Por tanto, en nuestro caso tenemos

$$n = r\left(X^T\right) + dim\left(K\right) = k + dim\left(K\right) \implies dim\left(K\right) = n - k$$

y por otro lado

$$n = r\left(P\right) + dim\left(K\right) = r\left(P\right) + n - k \implies r\left(P\right) = k$$

como queríamos ver.
3.

$$\left(I_n - P\right)X = I_nX - X\cancelto{0}{\left(X^TX\right)^{-1}}X^TX = X - X = 0$$

Ejercicio 4: Demostrar que.

$$\sum_{i=1}^{n}\left(y_i-\sum_{j=1}^{k}x_{ij}\beta_j\right)^2=\left(\boldsymbol{y}-\boldsymbol{X}\hat{\beta}\right)^T\left(\boldsymbol{y}-\boldsymbol{X}\hat{\beta}\right)+\left(\hat{\beta}-\beta\right)^T\boldsymbol{X}^T\boldsymbol{X}\left(\hat{\beta}-\beta\right)$$

Veámoslo, aunque voy a usar la notación $\boldsymbol{y}=Y$ y $\boldsymbol{X}=X$, para facilitar la escritura del ejercicio.

$$\sum_{i=1}^{n}\left(y_i-\sum_{j=1}^{k}x_{ij}\beta_j\right)^2=\sum_{i=1}^{n}\left(y_i^2-2y_i\sum_{j=1}^{k}x_{ij}\beta_j+\left(\sum_{j=1}^{k}x_{ij}\beta_j\right)^2\right)=(*)$$

Notemos en este punto que

$$\sum_{i=1}^{n}y_i^2=Y^TY$$

$$\sum_{i=1}^{n}y_i\sum_{j=1}^{k}x_{ij}\beta_j=\sum_{i=1}^{n}y_iX_i\beta=Y^TX\beta$$

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{k}x_{ij}\beta_j\right)^2=\sum_{i=1}^{n}(X_i\beta)^2=(X\beta)^T(X\beta)$$

Además, dado que $Y^TX\beta$ es un escalar (matriz de dimensión 1×1), se tiene que

$$Y^TX\beta=\left(Y^TX\beta\right)^T=(X\beta)^TY \tag{1}$$

por lo que

$$(*)=Y^TY-2Y^TX\beta+(X\beta)^T(X\beta)\overset{(1)}{=}Y^TY-Y^TX\beta-(X\beta)^TY+(X\beta)^T(X\beta)=$$

$$=Y^T(Y-X\beta)-(X\beta)^T(T-X\beta)=\left(Y^T-(X\beta)^T\right)(Y-X\beta)=(Y-X\beta)^T(Y-X\beta)=$$

$$=\left(Y-X\hat{\beta}+X\hat{\beta}-X\beta\right)^T\left(Y-X\hat{\beta}+X\hat{\beta}-X\beta\right)=$$

$$\left(Y-X\hat{\beta}\right)^T\left(Y-X\hat{\beta}\right)+\left(Y-X\hat{\beta}\right)^T\left(X\hat{\beta}-X\beta\right)+\left(X\hat{\beta}-X\beta\right)^T\left(Y-X\hat{\beta}\right)+\left(X\hat{\beta}-X\beta\right)^T\left(X\hat{\beta}-X\beta\right)=(**)$$

Aquí tenemos ya el primero término. Además, el último es:

$$\left(X\hat{\beta}-X\beta\right)^T\left(X\hat{\beta}-X\beta\right)=\left(X\left(\hat{\beta}-\beta\right)\right)^TX\left(\hat{\beta}-\beta\right)=\left(\hat{\beta}-\beta\right)^TX^TX\left(\hat{\beta}-\beta\right)$$

que es el otro sumando buscado. Por lo que solo resta ver que los otros dos sumandos dan 0. Aquí vamos a utilizar que $Y^TX\beta$ tiene dimensión 1, por lo que coincide con su traspuesta.

$$\left(Y-X\hat{\beta}\right)^T\left(X\hat{\beta}-X\beta\right)+\left(X\hat{\beta}-X\beta\right)^T\left(Y-X\hat{\beta}\right)=$$

$$=Y^TX\hat{\beta}-Y^TX\beta-\hat{\beta}^TX^TX\hat{\beta}+\hat{\beta}^TX^TX\beta+\hat{\beta}^TX^TY-\hat{\beta}^TX^TX\hat{\beta}-\hat{\beta}^TX^TY+\beta^TX^TX\hat{\beta}=$$

$$=2Y^TX\hat{\beta}-2Y^TX\beta-2\hat{\beta}^TX^TX\hat{\beta}+2\hat{\beta}^TX^TX\beta=2\left(Y^TX\hat{\beta}-Y^TX\beta-\hat{\beta}^TX^TX\hat{\beta}+\hat{\beta}^TX^TX\beta\right)$$

Esto dará 0 si, y solo si, da 0 al dividir por 2. Continuamos con esa expresión y usamos que $\hat{\beta}=\left(X^TX\right)^{-1}X^TY$

$$Y^TX\hat{\beta}-Y^TX\beta-\hat{\beta}^TX^TX\hat{\beta}+\hat{\beta}^TX^TX\beta=$$

$$= Y^T \left(X\hat{\beta} - X\beta\right) - Y^T X \cancel{\left(X^T X\right)^{-1} X^T X} \left(X^T X\right)^{-1} X^T Y + Y^T X \cancel{\left(X^T X\right)^{-1} X^T} X\beta =$$

$$= Y^T \left(X\hat{\beta} - X\beta\right) - Y^T X\hat{\beta} + Y^T X\beta = 0$$

Tal y como queríamos ver, y ahora sabemos que

$$(**) = \left(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}\right)^T \left(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}\right) + \left(\hat{\beta} - \beta\right)^T \boldsymbol{X}^T \boldsymbol{X} \left(\hat{\beta} - \beta\right)$$

y tenemos el resultado buscado.