Matemáticas Especiales II

Año 2014

Prof: T. S. Grigera — JTP: V. Fernández — AD: S. Franchino

Práctica 12 — Estadística

Ley de grandes números, teorema central del límite y estimadores

Esta práctica abarca los siguientes temas:

a) Propiedades de combinaciones lineales de varias variables aleatorias. Convergencia de sucesiones de variables aleatorias. Lema de Tchebishev y ley de los grandes números: formas fuerte y débil, interpretación. Teorema central de límite: casos de distribuciones con varianza finita e infinita.

b) Definición de muestra aleatoria. Estimación de parámetros: Estimadores sesgados y no-sesgados. Estimadores eficientes.

c) Estimación por máxima verosimilitud. Carácter consistente y asintóticamente eficiente del estimador de m.v. Aplicación a poblaciones gaussianas. Rectas de regresión: cuadrados mínimos.

Bibliografía: Marinari y Parisi (2002, cap. 3), Cramer (1946, caps. 20, 32–34), Meyer (1961, caps. 12–14).

Convergencia de distribuciones. La afirmación de que una sucesión de variables aleatorias x_n tiende a un límite X puede tener distintos sentidos, según la noción de convergencia empleada. Tres nociones de convergencia empleadas en teoría de probabilidades son:

- **Convergencia como distribución:** La más débil de las tres, significa que la distribución de las x_n se acerca a la distribución de X:
$$\lim_{n\to\infty} f_n(x_n) = f(X). \tag{12.1}$$
- **Convergencia en probabilidad:** Quiere decir que la probabilidad de que x_n esté lejos de X disminuye al aumentar n. Con más precisión,
$$x_n \xrightarrow{P} X \iff P(|x_n - X| > \epsilon) \underset{n}{\to} 0 \quad \forall\epsilon. \tag{12.2}$$
Esto implica convergencia como distribución.
- **Convergencia fuerte.** Se dice que la sucesión converge *fuertemente* o *con probabilidad 1* o *casi con seguridad* si el conjunto de eventos tales que efectivamente $x_n \to X$ tiene probabilidad 1. Este criterio implica la convergencia en los dos sentidos anteriores.

Problema 1. Ley de los grandes números. Consideremos N variables aleatorias x_i, independientes y con idéntica distribución (iid). Típicamente se trata de N realizaciones del mismo experimento. Sean $\mu = \langle x\rangle$, $\sigma^2 = \langle x^2\rangle - \mu^2$, $z = \sum_i x_i$, $y_N = z/N$ y supongamos que σ^2 existe (es finita).

a) **Lema de Tchebishev.** Demuestre que
$$P(\{|x-\mu| > t\}) \le (\sigma/t)^2, \qquad \textit{(lema de Tchebishev).} \tag{12.3}$$
Ayuda: Escriba la definición de σ^2 y observe que una cota inferior es $\int_{|x-\mu|>t} P(x)$.

b) **Ley de los grandes números.** Ahora, con ayuda del lema anterior trataremos de probar que
$$y_N = \frac{1}{N}\sum_i x_i \xrightarrow[N]{P} \langle x\rangle \equiv \mu, \qquad \textit{(forma débil de la ley de los grandes números).} \tag{12.4}$$
Observe que $\mathrm{Var}[z] = N\sigma^2$ y utilice el lema de Tchebishev para probar que $P(|z - N\mu| > t) < N\sigma^2/t$. Eligiendo apropiadamente t demuestre que
$$P(|y_N - \mu| > \epsilon) < \sigma^2/N\epsilon^2, \tag{12.5}$$
de donde se sigue (12.4). *Nota:* la presente demostración requiere que la distribución tenga varianza finita, pero la condición puede relajarse y la ley de los grandes números es válida si $\int |x|p(x)\,dx$ es finita.

c) **Forma fuerte de la ley de los grandes números.** Se puede demostrar también el límite en el sentido de convergencia fuerte, o con probabilidad 1. Esto equivale a decir que $|y_N - \mu| < \epsilon \; \forall N > m(\epsilon)$ con probabilidad 1.

Problema 2. Sean X_k $k = 1, \cdots, n$, variables aleatorias que toman los valores ± 1 con probabilidad 1/2. Muestre que la función característica de la variable a aleatoria $(\sum_k X_k)/\sqrt{n}$ es $[cos(\omega/\sqrt{n})]^n$. Demuestre que cuando $n \to \infty$ esta función característica tiende a $e^{-\omega^2/2}$. Interpretar el resultado.

Problema 3. Teorema central del límite. La ley de los grandes números asegura que $y_N = z/N$ tiene una distribución con varianza cada vez más pequeña alrededor de $\langle x \rangle$. El TCL dice *qué forma* tiene la distribución para N grande: si la distribución de las x_i tiene varianza finita, la distribución de z será Gaussiana con $\langle z \rangle = N\mu$, $\langle z^2 \rangle - \langle z \rangle^2 = N\sigma^2$. El problema anterior es un ejemplo particular del resultado de este teorema.

Con más precisión, podemos decir que si $y_N = z/N$, la ley de los grandes números garantiza que $P(y_N) \to \delta(y - \mu)$. El TCL dice que si $w_N = z/\sqrt{N}$, entonces $P(w_N)$ tiende para $N \to \infty$ a una Gaussiana con varianza σ^2 y media μ.

Siga los siguientes pasos para demostrar el TCL.

a) Sea $z = \sum_i^N x_i$, $p_x(x)$ la densidad de probabilidad de las x_i, que además son independientes entre sí, $\mu = 0$ y σ^2 los dos primeros momentos centrados de $p_x(x)$ y $\tilde{p}_x(q)$ la función característica. Demuestre que

$$\tilde{p}_z(q) = [\tilde{p}_x(q)]^N . \tag{12.6}$$

b) Demuestre que la característica de w es

$$\tilde{p}_w(q) = \tilde{p}_z(q/\sqrt{N}), \tag{12.7}$$

y por lo tanto

$$\tilde{p}_w(q) = \exp[N \log \tilde{p}_x(q/\sqrt{N})]. \tag{12.8}$$

c) Observe que cuando $N \to \infty$, $\tilde{p}_w(q)$ se anulará a menos que $\tilde{p}_x(q) = 1$. Como ese valor se toma para $q = 0$, desarrolle $\tilde{p}_x(q)$ en torno a 0 y obtenga

$$\tilde{p}_w(q) \approx e^{-q^2\sigma^2/2} + O\Big(1/\sqrt{N}\Big). \tag{12.9}$$

d) El *teorema de continuidad de Paul-Levy* afirma que si una suceción de caracterísiticas converge puntalmente a una función $\phi(q)$ continua en 0, entonces las respectivas distribuciones tienden a una distribución cuya característica es $\phi(q)$. Aceptando esto, concluya que

$$\lim_{N \to \infty} p_{w_N}(w) = \frac{e^{-w^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}. \tag{12.10}$$

Mediante el cambio de variable $x \longrightarrow x - \mu$ se puede demostrar la afirmación equivalente para el caso general de una distribución con media arbitraria.

Note que para que el teorema valga *la varianza debe ser finita.* Si la varianza es divergente, existen otros teoremas (TCLs generalizados) que muestran que la distribución límite de w_N tiene decaimiento a potencias ($w^{-\alpha}$ con $1 < \alpha < 3$).

Problema 4. Estimadores de la media y la varianza. Considere los estimadores $\bar{x}$ y S^2 de la media y la varianza respectivamente, definidos por

$$\bar{x} = \frac{1}{N} \sum_i x_i, \tag{12.11}$$

$$S^2 = \frac{1}{N} \sum (x_i - \bar{x})^2. \tag{12.12}$$

a) Demuestre que $\bar{x}$ es un estimador *no sesgado* de la media. Calcule la varianza de $\bar{x}$.

b) Muestre que S^2 es un estimador sesgado. Construya un estimador no sesgado de la varianza y calcule la varianza del mismo.

Intervalo y nivel de confianza. Puesto que los estimadores de parámetros de una distribución son a su vez variables aleatorias (dado que dependen de una muestra, es decir de un conjunto de valores experimentales), al comunicar el valor numérico obtenido para un dado estimador es necesario también caracterizar las variaciones esperables del valor obtenido en el caso de repetir la medición. La forma más general de hacer esto es indicar un *intervalo de confianza* $[t_1, t_2]$ y un valor o *nivel de confianza* η tales que uno pueda afirmar que, si el parámetro que se estimó es θ, el valor del estimador estará dentro del intervalo de confianza con probabilidad mayor que η:

$$P\{t_1 \leq \theta \leq t_2\} = \eta. \tag{12.13}$$

Si la distribución de θ es simétrica, se puede expresar también como $\theta = \hat{\theta} \pm \epsilon$ con $P \geq \eta$. El intervalo de confianza es $[\hat{\theta} - \epsilon, \hat{\theta} + \epsilon]$. Si la distribución de θ es además Gaussiana (como sucederá si el TCL es aplicable al estimador θ), niveles típicos de confianza son $\eta = 0{,}68$ (con lo que ϵ resulta igual σ_θ, el denominado *error estándard*) o $\eta = 0{,}95$, que corresponde a $\epsilon = 2\sigma_\theta$.

Problema 5. Se encuesta a un número N de votantes para determinar el porcentaje a favor de cierto candidato. Si una proporción p desconocida de votantes lo favorece, cuántas personas deben ser encuestadas para predecir el valor de p con un margen de error 0.045 a un nivel de confianza del 95 %? (suponga que los votantes son independientes).

Estimación por el método de máxima verosimilitud. Dada una variable aleatoria x con distribución descripta por un conjunto de parámetros $\{\theta_i\}$ (cuya forma se conoce o se supone), se trata de estimar el valor de estos parámetros a partir de un conjunto de resultados $\{x_i\}$ de N realizaciones independientes del experimento que mide x. El *método de m.v.* consiste en calcular $P(\{x_i\}|\{\theta_i\})$ como función de los parámetros y elegir como estimadores el conjunto $\{\hat{\theta}_i\}$ de valores que maximizan la probabilidad.

Problema 6. Muestre que el estimador de máxima verosimilitud para el parámetro λ de la distribución exponencial $p(x) = \lambda e^{-\lambda x}$ es $\hat{\lambda}(\{x_i\}) = N/\sum_i x_i$. Calcule el valor medio de $\hat{\lambda}$ y observe que para N finito es un estimador sesgado. *Sugerencia:* No maximice $P(\{x_i\}|\lambda)$ sino su logaritmo.

Problema 7. Verifique que para el problema 5, el estimador de máxima verosimilitud de p, habiéndose encuestado a N personas, de las cuales k votaron a favor, es $p = \frac{k}{n}$.

Estimador eficiente. Se dice que un estimador es *eficiente* si es no sesgado y de varianza mínima.

Problema 8. Demuestre que el estimador de máxima verosimilitud es *asintóticamente eficiente.*

a) Demuestre primero que entre todos los estimadores eficientes existe uno con varianza mínima. Sea $\{x_i\}$ el conjunto de los resultados experimentales y $\hat{\theta}(\{x_i\})$ un estimador eficiente ($\langle\hat{\theta}\rangle = \theta$). Llamemos $F(\{x_i\}, \theta) = \prod_i f(x_i, \theta)$ la densidad de probabilidad conjunta de las x_i (v.a. independientes).

   I) Definamos $v = d \ln F(x_i, \theta)/d\theta \equiv \sum_i v_i$. Muestre que

$$\langle v_i \rangle = 0, \qquad \text{y} \qquad \operatorname{Cov} \hat{\theta}, v = \langle \hat{\theta} v \rangle = 1. \tag{12.14}$$

   II) Utilizando la desigualdad de Schwartz ($\|x\|^2 \|y\|^2 \geq (x, y)^2$) y el resultado anterior muestre que $\operatorname{Var} \hat{\theta} \operatorname{Var} v \geq 1$. Reescribiendo adecuadamente $\operatorname{Var} v$ concluya que

$$\operatorname{Var} \hat{\theta} \geq \frac{1}{NI(\theta)}, \qquad I(\theta) = -\frac{d^2}{d\theta^2}\langle \ln f \rangle. \tag{12.15}$$

   Ha encontrado entonces una cota inferior para la varianza de todos los posibles estimadores no sesgados de θ,

b) Ahora puede argumentar (sin rigurosidad) que el estimador de m.v. es asintóticamente eficiente.

   I) Veamos primero que es *consistente* (asintóticamente no sesgado). Escriba la probabilidad $F(\{x_i\}, \hat{\theta})$ como función del valor estimado. Si $\hat{\theta}$ es el estimador de m.v., se debe cumplir

$$\left.\frac{d \ln F}{d\hat{\theta}}\right|_{\theta = \hat{\theta}} = 0. \tag{12.16}$$

Desarrollando $\ln F$ alrededor del valor real θ hasta orden cúbico en $(\hat{\theta} - \theta)$ muestre que

$$\hat{\theta} - \theta = \frac{-\frac{1}{N} \sum_i v_i}{\frac{1}{N} \sum_i \frac{d^2}{d\theta^2} \ln f(x_i) + \frac{a}{N}(\hat{\theta} - \theta)} \xrightarrow[N\to\infty]{P} \frac{\langle v \rangle}{I(\theta)} = 0, \tag{12.17}$$

donde el límite está garantizado por la ley de los grandes números. El estimador de m.v. es pues asintóticamente no sesgado.

II) Para demostrar la eficiencia asintótica hace falta mostrar que la varianza del estimador es mínima. Escribiendo

$$-\left(\frac{1}{\sqrt{N}} \sum_i \frac{d^2}{d\theta^2} \ln f(x_i) \right) (\hat{\theta} - \theta) = \frac{\frac{1}{\sqrt{N}} \sum_i v_i}{1 + \frac{a}{\sqrt{N}}(\hat{\theta} - \theta)}, \tag{12.18}$$

y argumentando que el prefactor del m.i. se acerca a $\sqrt{N} I(\theta)$ para $N \to \infty$ se tendrá que la varianza del m.i. se acerca a $N I^2(\theta) \operatorname{Var}(\hat{\theta} - \theta)$. Concluya entonces la plausibilidad de

$$\operatorname{Var} \hat{\theta} - \theta = \frac{1}{I(\theta)}, \tag{12.19}$$

es decir de que la varianza del estimador sea igual a la cota inferior encontrada antes. La varianza del estimador de m.v. es por lo tanto mínima y el mismo es *asintóticamente eficiente.*

Problema 9. Doce piezas de cierta marca de pan se analizan para determinar su contenido de carbohidratos, con los siguientes resultados: 76.93; 76.88; 77.07; 76.68; 76.39; 75.09; 76.88; 77.67; 78.15; 76.50; 77.16; 76.42.

a) Suponiendo que la distribución de contenido de carbohidrato es normal, estime el verdadero contenido medio de carbohidrato de esta marca.

b) ¿Cuál es el error estándard estimado del estimador que usó en a)?.

Problema 10. Muestre que el método de mínimos cuadrados para ajustar una curva $y = f(x; \theta_i)$ a un conjunto de puntos experimentales (x_i, y_i) (es decir estimar los parámetros θ_i) resulta de construir un estimador de máxima verosimilitud suponiendo que la distribución de probabilidad de los valores y_i es Gaussiana con media $f(x_i)$ y varianza arbitraria pero idéntica para todos los puntos.