Probabilidad - Clase 24
Simulación - Estadística

Ernesto Mordecki

Facultad de Ciencias, Universidad de la República. Montevideo, Uruguay

Curso de la Licenciatura en Matemática - 2020

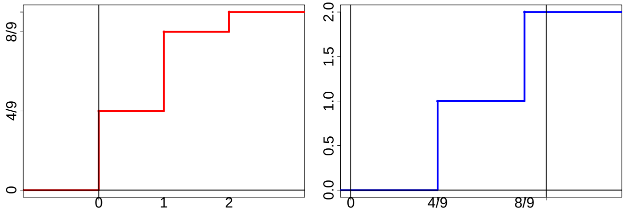
Contenidos

Inversa generalizada de una distribución

Para $F : \mathbb{R} \to [0, 1]$ función de distribución, se define su inversa generalizada

$$F^{-1}(p) = \inf\{x \in \mathbb{R} : F(x) \geq p\}, \quad \text{para } 0 < p < 1.$$

$$F^{-1}(p) = \begin{cases} 0, & \text{cuando } 0 < p \leq 4/9 \\ 1, & \text{cuando } 4/9 < p \leq 8/9 \\ 2, & \text{cuando } 8/9 < p \leq 1 \end{cases}$$



Simulación de un vectores gaussianos bidimensional (Box-Muller)

Como la función

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt,$$

no tiene fórmula elemental, no se puede aplicar el método de la transformación inversa.

Teorema (Método de Box-Muller) Sean U, V dos variables aleatorias uniformes independientes. Entonces

$$(X, Y) = \left(\sqrt{-2\log(U)}\sin(2\pi V), \sqrt{-2\log(U)}\cos(2\pi V)\right)$$

es un vector normal bivariado con[1]

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

[1] μ es el *a* de antes.

Demostración

Sea $A \subset \mathbb{R}^2$. Se tiene que

$$P((X,Y) \in A) = \iint_{(\sqrt{-2\log(u)}\sin(2\pi v), \sqrt{-2\log(u)}\cos(2\pi v)) \in A} dudv$$

Hacemos el cambio de variable

$$g(u,v) = (\sqrt{-2\log(u)}\sin(2\pi v), \sqrt{-2\log(u)}\cos(2\pi v)) = (x,y)$$

Despejando en términos de u y v tenemos que

$$g^{-1}(x,y) = \left(e^{\frac{-x^2-y^2}{2}}, \frac{1}{2\pi}\arctan(y/x)\right)$$

Calculemos el jacobiano: $\mathbb{J}_{g^{-1}}(x,y)$

$$\mathbb{J}_{g^{-1}}(x,y) = \begin{vmatrix} e^{\frac{-x^2-y^2}{2}}(-x) & e^{\frac{-x^2-y^2}{2}}(-y) \\ \frac{1}{2\pi(1+(y/x)^2)}(-y/x^2) & \frac{1}{2\pi(1+(y/x)^2)}(1/x^2) \end{vmatrix}$$

$$= \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$$

Luego,

$$\begin{aligned} P((X,Y)\in A) &= \iint_{(\sqrt{-2\log(u)}\sin(2\pi v),\sqrt{-2\log(u)}\cos(2\pi v))\in A} dudv \\ &= \frac{1}{2\pi}\iint_{(x,y)\in A} e^{-\frac{1}{2}(x^2+y^2)} dxdy \end{aligned}$$

v.a. i.i.d - m.a.s

- Un concepto central en probabilidad y estadística lo conforman un conjunto

$$X_1, X_2, \ldots, X_n$$

de variables aleatorias independientes con distribución común.

- En probabilidad, decimos que se trata de n variables aleatorias independientes con idéntica distribución, y se abrevia v.a. i.i.d.

- En estadística, decimos que se trata de una muestra aleatoria simple: y se abrevia m.a.s.

- Sea $F(x)$ la distribución común de las variables, $p(x)$ su densidad, p_k la probabilidad en el caso discreto.
- La distribución del vector $X = (X_1, X_2, \ldots, X_n)$ viene dada por
$$F(x_1, \ldots, x_n) = F(x_1) \ldots F(x_n)$$
- En caso de ser continuas, la densidad conjunta es
$$p(x_1, \ldots, x_n) = p(x_1) \ldots p(x_n)$$
- En el caso discreto, la probabilidad conjunta es
$$P(X_1 = k_1, \ldots, X_n = k_n) = p_{k_1} \ldots p_{k_n}.$$

Estadística: Estimación de un parámetro

Supongamos la siguiente situación estadística:

- Observamos el resultado numérico de n experimentos que se realizan en forma independiente y bajo las mismas condiciones.
- Es válido suponer que lo que observamos es una muestra aleatoria simple $X_1, \ldots, X_n$
- Dichas variables dependen de un parámetro θ que desconocemos
- Queremos a partir de la observación tener una idea del valor verdadero de θ

Estimador de un parámetro

- Se trata de calcular una aproximación de $\theta \in \mathbb{R}$ (valor teórico desconocido), que denotamos $\hat{\theta}$, a partir de $X_1, \ldots, X_n$.
- Más formalmente escribimos
$$\hat{\theta}_n = f(X_1, \ldots, X_n).$$
es decir, un *estimador* es una función $f \colon \mathbb{R}^n \to \mathbb{R}$.
- Si bien no es una condición de la definición, se espera que $\hat{\theta}_n$ sea cercano a θ, o converga a θ en algún sentido.
- A veces distingimos con la notación $x_1, \ldots, x_n$ (dato observado) de $X_1, \ldots, X_n$ (variable aleatoria a observar)

Estimador de máxima verosimilitud

Vamos a usar la misma estrategia que antes:

- escribimos la densidad (o la probabilidad en el caso discreto) como función de la muestra, y obtenemos al variar θ el máximo valor posible de esa densidad.
- definimos entonces la *verosimilitud*

$$L(\theta, x_1, \ldots, x_n) := p(x_1, \theta) \ldots p(x_n, \theta)$$

- definimos

$$\hat{\theta} = \arg\max_{\theta \in \mathbb{R}} L(\theta, x_1, \ldots, x_n)$$

- Muchas veces es útil definir la *log-verosimilitud*:

$$\begin{aligned}\ell(\theta, x_1, \ldots, x_n) :&= \log p(x_1, \theta) \ldots p(x_n, \theta)\\ &= \sum_{k=1}^{n} \log p(x_k, \theta).\end{aligned}$$

- Como el logaritmo es una función monótona, tenemos

$$\hat{\theta}_n = \arg\max_{\theta \in \mathbb{R}} \ell(\theta, x_1, \ldots, x_n)$$

- El mismo valor que maximiza L es el que maximiza ℓ.

Ejemplo: estimación del *a* en el caso normal

- Supongamos que sabemos que los datos siguen una distribución normal, conocemos el parámetro de escala σ^2 pero desconocemos el parámetro de posición a.
- Eso puede corresponder a mediciones con un aparato que introduce un error σ^2 conocido.
- Nuestro parámetro de interés es entonces a.
- Veamos entonces si obtenemos f para tener una estimación

La densidad normal es

$$p(x,a) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^2}$$

La verosimilitud es entonces

$$\begin{aligned} L(a, x_1, \ldots, x_n) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_1-a}{\sigma}\right)^2} \ldots \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_n-a}{\sigma}\right)^2} \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{k=1}^n \left(\frac{x_k-a}{\sigma}\right)^2} \end{aligned}$$

Por eso la log-verosimilitud es

$$\begin{aligned} \ell(a, x_1, \ldots, x_n) &= \log\left[\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{k=1}^n \left(\frac{x_k-a}{\sigma}\right)^2}\right] \\ &= n\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\sum_{k=1}^n \left(\frac{x_k-a}{\sigma}\right)^2 \end{aligned}$$

El primer sumando no depende de a. Tenemos entonces que minimizar

$$h(a) = \sum_{k=1}^{n} \left(\frac{x_k - a}{\sigma} \right)^2 .$$

Derivando con respecto a a:

$$h'(a) = \frac{2}{\sigma} \sum_{k=1}^{n} \frac{x_k - a}{\sigma} = \frac{2}{\sigma} \left(\sum_{k=1}^{n} x_k - na \right) .$$

Entonces, el valor donde se obtiene el máximo de L es

$$\hat{a} = \frac{\sum_{k=1}^{n} x_k}{n} =: \bar{x}_n .$$

Luego, el estimador de máxima verosimilitud de a es el promedio(!)

Caso vectorial para θ

- Podemos generalizar cuando θ es un vector en vez de un parámetro.
- Además el conjunto de búsqueda no tiene necesariamente que ser $\mathbb{R}^n$. El conjunto se denomina Θ
- Entonces, el estimador de máxima verosimilitud es

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} \ell(\theta, x_1, \ldots, x_n)$$

- Por ejemplo, si queremos estimar a y σ en el caso gaussiano, tenemos

$$(a, \sigma) \in \Theta = \mathbb{R} \times [0, \infty)$$

- La cuentas que hicimos son válidas, pensando que ℓ es función también de σ
- Resulta

$$(\hat{a}, \hat{\sigma}) = \arg\max_{\Theta} \left[n \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2} \sum_{k=1}^{n} \left(\frac{x_k - a}{\sigma}\right)^2 \right]$$

- Las cuentas van en un ejercicio del práctico 5.