Simulación II

Prof. José Niño Mora

Investigación Operativa, Grado en Estadística y Empresa, 2011/12

Universidad Carlos III de Madrid
www.uc3m.es

Esquema

- Determinación del tamaño de la muestra

- Generación de variables discretas

- Generación de variables continuas

El problema del tamaño de la muestra

- En simulación, queremos **estimar la media** $\boxed{\mu}$ de una v.a. $\boxed{Y}$ con varianza $\boxed{\sigma^2}$ a partir de la media de una muestra i.i.d. $\boxed{Y_1, Y_2, \ldots, Y_n}$. Ej: $\boxed{Y = g(X), \mu = \mathsf{E}[Y] = \mathsf{E}[g(X)]}$

- La **media muestral** es: $\boxed{\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i}$ y la **varianza muestral** es

$$\boxed{\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_n)^2}$$

- ¿Cómo de grande ha de ser el **tamaño de la muestra** $\boxed{n}$ para que, dados $\boxed{\epsilon > 0}$, $\boxed{0 < \alpha < 1}$,

$$\boxed{\mathsf{P}\{|\widehat{\mu}_n - \mu| > \epsilon\} \le \alpha,}$$

es decir, para que la **probabilidad de que el error de estimación sea mayor que** $\boxed{\epsilon}$ no exceda $\boxed{\alpha}$?

Intervalos de confianza

- Aplicaremos **Intervalos de Confianza**. Por el **Teorema Central del Límite (TCL)**: para $\boxed{n}$ grande,

$$\boxed{\sqrt{n}\,\frac{\widehat{\mu}_n-\mu}{\widehat{\sigma}_n} \overset{\mathrm{d}}{\approx} \mathrm{Normal}(0,1)}$$

- Por tanto, $\boxed{\mathsf{P}\left\{\sqrt{n}\left|\frac{\widehat{\mu}_n-\mu}{\widehat{\sigma}_n}\right| > z_{\alpha/2}\right\} \approx \alpha,}$ donde $\boxed{z_{\alpha/2}}$ es el cuantil a nivel $\boxed{\alpha/2}$ de $\boxed{z \sim \mathrm{Normal}(0,1)}$: $\boxed{\mathsf{P}\{z > z_{\alpha/2}\} = \alpha/2}$

- **Intervalo de confianza para la media con nivel de significación** $\boxed{\alpha}$ (ej: $\boxed{\alpha = 0.05}$, $\boxed{z_{0.025} = 1.96}$):

$$\boxed{\mathsf{P}\left\{\widehat{\mu}_n - \frac{\widehat{\sigma}_n}{\sqrt{n}} z_{\alpha/2} < \mu < \widehat{\mu}_n + \frac{\widehat{\sigma}_n}{\sqrt{n}} z_{\alpha/2}\right\} \approx 1-\alpha}$$

Determinación del tamaño de la muestra

- ¿Cómo de grande ha de ser el **tamaño de la muestra** $\boxed{n}$ para que, dados $\boxed{\epsilon > 0}$, $\boxed{0 < \alpha < 1}$,

$$\boxed{\mathrm{P}\left\{|\widehat{\mu}_n - \mu| > \epsilon\right\} \leq \alpha,}$$

es decir, para que la **probabilidad de que el error de estimación sea mayor que** $\boxed{\epsilon}$ no exceda $\boxed{\alpha}$?

- Como $\boxed{\mathrm{P}\left\{|\widehat{\mu}_n - \mu| > \frac{\widehat{\sigma}_n z_{\alpha/2}}{\sqrt{n}}\right\} \approx \alpha,}$ obtenemos:

$$\boxed{\epsilon \approx \frac{\widehat{\sigma}_n z_{\alpha/2}}{\sqrt{n}} \Longrightarrow n \approx \left\{\frac{\widehat{\sigma}_n z_{\alpha/2}}{\epsilon}\right\}^2}$$

- **El error de estimación es inversamente proporcional a la raíz cuadrada del tamaño de la muestra**

Determinación del tamaño de la muestra

- ¿Cómo de grande ha de ser el **tamaño de la muestra** $\boxed{n}$ para que, dados $\boxed{\epsilon > 0}$, $\boxed{0 < \alpha < 1}$,

$$\boxed{\mathsf{P}\left\{|\widehat{\mu}_n - \mu| > \epsilon\right\} \leq \alpha,}$$

es decir, para que la **probabilidad de que el error de estimación sea mayor que** $\boxed{\epsilon}$ no exceda $\boxed{\alpha}$?

- Respuesta: si $\boxed{\widehat{\sigma}_n \approx \widehat{\sigma}}$, tomamos $\boxed{n \approx \left\{\frac{\widehat{\sigma} z_{\alpha/2}}{\epsilon}\right\}^2}$

- Ej: Si $\boxed{\widehat{\sigma} = 2, \epsilon = 0.01, \alpha = 0.05 \quad (z_{\alpha/2} = 1.96)}$:

$$\boxed{n \approx \left\{\frac{\widehat{\sigma} z_{\alpha/2}}{\epsilon}\right\}^2 = \left\{\frac{2 \times 1.96}{0.01}\right\}^2 = 153664}$$

Determinación del tamaño de la muestra

- Ej: Si $\boxed{\widehat{\sigma} = 2, \epsilon = 0.01, \alpha = 0.05 \quad (z_{\alpha/2} = 1.96)}$:

$$\boxed{n \approx \left\{ \frac{\widehat{\sigma} z_{\alpha/2}}{\epsilon} \right\}^2 = \left\{ \frac{2 \times 1.96}{0.01} \right\}^2 = 153664}$$

- ¿Y si queremos tener una precisión de un decimal más? Tomamos $\boxed{\epsilon = 0.001}$, y obtenemos

$$\boxed{n \approx \left\{ \frac{\widehat{\sigma} z_{\alpha/2}}{\epsilon} \right\}^2 = \left\{ \frac{2 \times 1.96}{0.001} \right\}^2 = 15366400}$$

- Aumentar la precisión de la estimación en una cifra decimal requiere aumentar en 100 veces el tamaño de la muestra

- En simulación, la precisión es muy costosa computacionalmente

Generación de números pseudo-aleatorios

Caso discreto

- Hemos visto colmo generar una sucesión $\boxed{U_0, U_1, U_2, \ldots}$ de **números pseudo-aleatorios**, que **se comporta como una muestra i.i.d. de v.a.** $\boxed{\sim U(0,1)}$

- Supongamos que nos interesa generar números pseudo-aleatorios que se comporten como una muestra de una distribución discreta: p.ej., Binomial, Geométrica, etc.

- Veremos que podemos utilizar para ello el generador de números pseudo-uniformes

Generación de v.a. discretas

- Queremos simular una **v.a. discreta** $\boxed{X}$ con distribución

$$\boxed{\mathsf{P}\{X = k\} = p_k, \quad k = 0, 1, \ldots}$$

- Generamos una v.a. $\boxed{U \sim U(0,1)}$
- Definimos $\boxed{X}$ por:

$$\boxed{X \triangleq \begin{cases} 0 & \text{si } U \le p_0 \\ 1 & \text{si } p_0 < U \le p_0 + p_1 \\ \ldots & \ldots \\ k & \text{si } p_0 + \cdots + p_{k-1} < U \le p_0 + \cdots + p_k \\ \ldots & \ldots \end{cases}}$$

- Entonces, $\boxed{X}$ tiene la distribución requerida. ¿Por qué?

Generación de v.a. discretas

- Entonces, $\boxed{X}$ tiene la distribución requerida. ¿Por qué?

$$\boxed{\begin{aligned} \mathsf{P}\{X = k\} &= \mathsf{P}\{p_0 + \cdots + p_{k-1} < U \leq p_0 + \cdots + p_k\} \\ &= p_k \end{aligned}}$$

Generación de la distribución Binomial

- Queremos generar $\boxed{X \sim \mathrm{B}(p)}$ ($\boxed{p}$: probabilidad de éxito):

$$\boxed{\mathsf{P}\{X = 1\} = p, \quad \mathsf{P}\{X = 0\} = 1 - p}$$

- Dada una v.a. $\boxed{U \sim U(0,1)}$, construimos $\boxed{X}$:

$$\boxed{X \triangleq 1_{\{U > 1-p\}} = \begin{cases} 0 & \text{si } U \leq 1 - p \\ 1 & \text{si } U > 1 - p \end{cases}}$$

- Entonces, $\boxed{X \sim \mathrm{B}(p)}$

Generación de la distribución Geométrica

- Queremos generar $\boxed{X \sim \mathrm{Geom}(p)}$ (que representa, p. ej., el # de fracasos antes del primer éxito, en pruebas Binomiales i.i.d. con probabilidad de éxito $\boxed{p}$):

$$\boxed{\mathsf{P}\{X = k\} = p(1-p)^k, \quad k = 0, 1, 2, \ldots}$$

- Su media y varianza son:

$$\boxed{\mathsf{E}\left[X\right] = \frac{1-p}{p}, \quad \mathrm{Var}\left[X\right] = \frac{1-p}{p^2}}$$

Generación de la distribución Geométrica

- Dada una v.a. $\boxed{U \sim U(0,1)}$, construimos $\boxed{X}$:

$$\boxed{X \triangleq \begin{cases} 0 & \text{si } U \leq p \\ k & \text{si } \displaystyle\sum_{l=0}^{k-1} p(1-p)^l < U \leq \sum_{l=0}^{k} p(1-p)^l, k \geq 1 \end{cases}}$$

Generación de la distribución Geométrica

- Como

$$\boxed{\sum_{l=0}^{k} p(1-p)^l = p\sum_{l=0}^{k}(1-p)^l = p\frac{1-(1-p)^{k+1}}{1-(1-p)} = 1-(1-p)^{k+1}}$$

podemos simplificar la expresión

$$\boxed{\sum_{l=0}^{k-1} p(1-p)^l < U \leq \sum_{l=0}^{k} p(1-p)^l}$$

como

$$\boxed{1-(1-p)^k < U \leq 1-(1-p)^{k+1}}$$

ó

$$\boxed{(1-p)^{k+1} \leq 1-U < (1-p)^k}$$

Generación de la distribución Geométrica

- Como la función logaritmo $\boxed{t \mapsto \ln(t)}$ es creciente, podemos simplificar aún más la expresión

$$\boxed{(1-p)^{k+1} \leq 1 - U < (1-p)^k}$$

como

$$\boxed{(k+1)\ln(1-p) \leq \ln(1-U) < k\ln(1-p)}$$

- Dividiendo por $\boxed{\ln(1-p) < 0}$, simplificamos más:

$$\boxed{k < \frac{\ln(1-U)}{\ln(1-p)} \leq k+1,}$$

ó

$$\boxed{\frac{\ln(1-U)}{\ln(1-p)} - 1 \leq k < \frac{\ln(1-U)}{\ln(1-p)} \iff k = \left\lceil \frac{\ln(1-U)}{\ln(1-p)} - 1 \right\rceil}$$

Generación de la distribución Geométrica

- Generamos $\boxed{U \sim U(0,1)}$

- Construimos

$$\boxed{X \triangleq \left\lceil \frac{\ln(1-U)}{\ln(1-p)} - 1 \right\rceil}$$

- Entonces, $\boxed{X \sim \mathrm{Geom}(p)}$

Generación de v.a. continuas

Método de la transformada inversa

- Queremos generar una **v.a. continua** $\boxed{X}$ con **función de distribución** creciente dada

$$\boxed{F(x) \triangleq \mathsf{P}\{X \leq x\}}$$

- Generamos una v.a. $\boxed{U \sim U(0,1)}$

- Construimos $\boxed{X \triangleq F^{-1}(U)}$ (es decir: $\boxed{F(X) = U}$)

- Entonces, $\boxed{X}$ tiene distribución $\boxed{F(\cdot)}$. ¿Por qué?

Generación de v.a. continuas

- Construimos $\boxed{X \triangleq F^{-1}(U)}$ (es decir: $\boxed{F(X) = U}$)

- Entonces, $\boxed{X}$ tiene distribución $\boxed{F(\cdot)}$. ¿Por qué?

- Como la función de distribución es creciente:

$$\boxed{\begin{aligned} \mathsf{P}\{X \leq x\} &= \mathsf{P}\left\{F\left(F^{-1}(U)\right) \leq F(x)\right\} \\ &= \mathsf{P}\{U \leq F(x)\} \\ &= F(x) \end{aligned}}$$

Generación de la distribución $\boxed{U[a,b]}$

- Queremos generar $\boxed{X \sim U(a,b)}$: su función de distribución es

$$\boxed{F(x) \triangleq \mathsf{P}\{X \leq x\} = \frac{x-a}{b-a}, \qquad x \in [a,b]}$$

- Generamos una v.a. $\boxed{U \sim U[0,1]}$

- Calculamos $\boxed{X}$ tal que $\boxed{F(X) = U}$:

$$\boxed{\begin{aligned} F(X) = U &\iff \frac{X-a}{b-a} = U \\ &\iff X = a + (b-a)U \end{aligned}}$$

Generación de la distribución Exponencial

- Queremos generar $\boxed{X \sim \mathrm{Exp}(\lambda)}$ (que representa, p. ej., el tiempo entre llegadas consecutivas en un sistema de colas): su función de distribución es

$$\boxed{F(x) \triangleq \mathsf{P}\{X \leq x\} = 1 - e^{-\lambda x}, \quad x \geq 0}$$

- Su media y varianza son:

$$\boxed{\mathsf{E}\left[X\right] = \frac{1}{\lambda}, \quad \mathrm{Var}\left[X\right] = \frac{1}{\lambda}}$$

Generación de la distribución Exponencial

- Generamos una v.a. $\boxed{U \sim U(0,1)}$

- Calculamos $\boxed{X}$ tal que $\boxed{F(X) = U}$:

$$\boxed{\begin{aligned} F(X) = U &\Longleftrightarrow 1 - e^{-\lambda X} = U \\ &\Longleftrightarrow e^{-\lambda X} = 1 - U \\ &\Longleftrightarrow -\lambda X = \ln(1 - U) \\ &\Longleftrightarrow X = -\frac{\ln(1 - U)}{\lambda} \end{aligned}}$$

Generación de la distribución Weibull

- Queremos generar $\boxed{X \sim \text{Weibull}(\alpha, \beta)}$ (que representa, p. ej., el tiempo de vida de una componente electrónica): su función de distribución es

$$\boxed{F(x) \triangleq \mathsf{P}\{X \leq x\} = 1 - e^{-(x/\beta)^\alpha}, \quad x \geq 0}$$

y su función de densidad es

$$\boxed{f(x) \triangleq \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha}, \quad x \geq 0}$$

- Su media y varianza son:

$$\boxed{\mathsf{E}[X] = \beta\Gamma(1 + 1/\alpha), \quad \text{Var}[X] = \beta^2 \left\{\Gamma(1 + 2/\alpha) - \Gamma^2(1 + 1/\alpha)\right\}}$$

Generación de la distribución Weibull

- Generamos una v.a. $\boxed{U \sim U(0,1)}$

- Calculamos $\boxed{X}$ tal que $\boxed{F(X) = U}$:

$$\boxed{\begin{aligned} F(X) = U &\Longleftrightarrow 1 - e^{-(X/\beta)^\alpha} = U \\ &\Longleftrightarrow e^{-(X/\beta)^\alpha} = 1 - U \\ &\Longleftrightarrow (X/\beta)^\alpha = -\ln(1-U) \\ &\Longleftrightarrow X/\beta = \{-\ln(1-U)\}^{1/\alpha} \\ &\Longleftrightarrow X = \beta\,\{-\ln(1-U)\}^{1/\alpha} \end{aligned}}$$