Simulación III

Dpto. Ingeniería Industrial, Universidad de Chile

IN47B, Ingeniería de Operaciones

29 de marzo de 2007

Contenidos

Definiciones

- $\mathbb{E}(X) = \int_{Dom(X)} x \cdot f(x) \cdot dx.$
- $\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}((X - \mathbb{E}(X))^2).$
- $\mathbb{C}\text{ov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$
- Si X, Y son independientes $\mathbb{C}\text{ov}(X, Y) = 0$.
- La reciproca no es cierta.
- $\mathbb{C}\text{or}(X, Y) = \dfrac{\mathbb{C}\text{ov}(X, Y)}{\sqrt{\mathbb{V}\text{ar}(X)\mathbb{V}\text{ar}(Y)}}.$

Estimadores

- Consideramos $X_i : i = 1, \ldots, n$ una muestra de X.
- $\overline{X}(n) = \dfrac{\sum_{i=1}^{n} X_i}{n}$
- $\overline{X}(n)$ es un estimador no sesgado de $\mathbb{E}(X)$.
- $S^2(n) = \dfrac{\sum_{i=1}^{n} (X_i - \overline{X}(n))^2}{n-1}$.
- $S^2(n)$ es un estimador no sesgado de $\mathbb{V}\mathrm{ar}(X)$.
- $\mathbb{V}\mathrm{ar}(\overline{X}(n)) = \dfrac{\mathbb{V}\mathrm{ar}(X)}{n}$.

Teorema Central del Límite

- Consideramos $\{X_i\}_{i\in\mathbb{N}}$ variables aleatoreas iid con $\mathbb{E}(X_i) = \mu$ y $\mathbb{V}\mathrm{ar}(X_i) = \sigma^2$.
- Definimos $Z_i = \dfrac{\overline{X}(n) - \mu}{\sqrt{\frac{\sigma^2}{n}}}$.
- Llamamos $F_n(z) = \mathbb{P}(Z_n \leq x)$

Teorema Central del Límite

$$\lim_{n\to\infty} F_n(z) = \Phi(z)$$

Donde $\Phi(z) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{z} e^{-y^2/2} dy$.

Teorema Central del Límite

- Básicamente el TCL dice que $Z_n \sim \mathcal{N}(0,1)$ cuando n es grande.
- Otro problema es que Z_n esta definido en términos de σ^2.
- Definimos $t_n = \dfrac{\overline{X}(n) - \mu}{\sqrt{\frac{S^2(n)}{n}}}$.
- Se puede demostrar que t_n tambien converge a una $\mathcal{N}(0,1)$.
- De ahi podemos decir que $\mathbb{P}(-z_{1-\alpha/2} \leq t_n \leq z_{1-\alpha/2}) \approx 1 - \alpha$ para n suficientemente grande.

Intervalos de Confianza

- De lo anterior, podemos concluir que

$$\mathbb{P}(l(n) \leq \mu \leq u(n)) = 1 - \alpha$$

para n suficientemente grande, donde

$$l(n) = \overline{X}(n) - z_{1-\alpha/2}\sqrt{\frac{S^2(n)}{n}}$$

y

$$u(n) = \overline{X}(n) + z_{1-\alpha/2}\sqrt{\frac{S^2(n)}{n}}$$

Intervalos de Confianza

- ¿Qué pasa si $X_i \sim \mathcal{N}(0,1)$?
- En ese caso $t_n \sim$ T-student de $n-1$ grados de libertad.
- Intervalo de confianza exacto esta dado por $\overline{X}(n) \pm t_{n-1,1-\alpha/2}\sqrt{S^2(n)/n}$.
- Se tiene que estos intervalos son mayores a considerar que $t_n \sim \mathcal{N}(0,1)$.
- En general, debemos preguntarnos ¿Qué significa n suficientemente grande?.

Cobertura Real

- Consideramos intervalos de confianza derivados de la T-student.
- Distinto número de muestras $n = 5, 10, 20,$ y 40.
- Consideramos X_i iid con distintas distribuciones.
- Comparamos cobertura real del intervalo estimado a 90 % sobre 500 repeticiones.

Dist	ν	n=5	n=10	n=20	n=40
Normal	0.00	0.910	0.902	0.898	0.900
Exponencial	2.00	0.854	0.878	0.870	0.890
Chi^2	2.83	0.810	0.830	0.848	0.890
Lognormal	6.18	0.758	0.768	0.842	0.852
Hiper-exp	6.43	0.584	0.586	0.682	0.774

Midiendo Simetría de Distribuciones

- ¿Qué es ν?
- $\nu = \dfrac{\mathbb{E}((X-\mu)^3)}{\sigma^3}$.
- ν es una medida de simetria de la distribución.
- Simetría de una distribuición es un factor importante al momento de determinar cuando *n es suficientemente grande* en el contexto del TCL.
- En general, No deberíamos mirar solamente a μ, si no que también a σ^2 cuando describimos una distribuición.

¿Cómo obtenemos variables iid?

- Consideremos un sistema de simulación donde hay sólo una medida de desempeño, que es reportada en distintos puntos J durante la simulación.
- Suponemos además que ejecutamos n corridas independientes de la simulación, esto define $X_{i,j}$ con $i = 1, \ldots, n$ y $j \in J$.
- Asumiendo buenos números aleatoreaos, podemos considerar $\{X_{i,j}\}_{i=1}^{n}$ como variables iid.
- Desafortunadamente $\{X_{i,j}\}_{j \in J}$ en la práctica no son independientes, de hecho, usualmente, tienen correlación positiva.

Algunos Ejemplos prácticos

- Consideramos un modelo M/M/1 con taza de ocupación $\rho = ,9$.
- Tratamos de estimar promedio de lso 25 primeros atrazos.
- Computamos 500 intervalos de confianza basados en 5,10,20 y 40 replicaciones.
- Comparamos proporción de intervalos *correctos* y su ancho medio.

M/M/1, estimando d_{25}

n	cobertura	intervalo 90 %	medio ancho
5	0.880	± 0.024	0.67
10	0.864	± 0.025	0.44
20	0.886	± 0.023	0.30
40	0.914	± 0.021	0.21

Tiempo medio a Falla

- Sistema con tres componentes.
- Sistema funciona mientras componente 1 funcione y componente 2 o componente 3 funcionen.
- Tiempo falla $G = \text{mín}\{G_1, \text{máx}\{G_2, G_3\}\}$, G_i es Weibull(0.5,1).

n	cobertura	intervalo 90 %	medio ancho
5	0.708	± 0.033	1.16
10	0.750	± 0.032	0.82
20	0.800	± 0.029	0.60
40	0.840	± 0.027	0.44