Simulación a Eventos Discretos

Tema 5: Análisis de resultados (estimación de medias y varianzas, intervalos de confianza)

Resultados de simulación y procesos estocásticos

Si las entradas de un modelo de simulación son muestras de variables aleatorias, también lo son las salidas.

Por lo tanto se debe tener cuidado al momento de formular conclusiones sobre algunas características del modelo (por ejemplo, tiempos de espera, largos de colas, utilización de recursos).

Un *proceso estocástico* se puede definir como una colección de v.a. similares ordenadas en el tiempo, definidas en un espacio común de muestras (*espacio de estados*).

El espacio de estados puede ser discreto ($X_1 \ldots X_n$, por ejempo, tiempos de espera de clientes) o continuo ($X(t), t \geq 0$, por ejemplo, número de clientes en una cola).

Hipótesis para análisis estadístico

Para inferir sobre el proceso estocástico subyacente de un conjunto de salidas de simulación, se deben asumir ciertas hipótesis para que los análisis estadísticos sean válidos.

Por ejemplo, es común asumir que un proceso discreto $X_1 \ldots X_n$ es de covarianza estacionaria. Esto significa que $\mu_i = \mu$, $\sigma_i^2 = \sigma^2$ para todo i y la covarianza de (X_i, X_{i+j}) es independiente de i (depende de la separación j y no del momento i).

Un proceso estocástico comenzando en el instante de tiempo 0 en general no cumple esta propiedad; sin embargo, a partir de cierto k puede cumplirla.

Estimación de media y varianza

Si tenemos $X_1, X_2 \ldots X_n$ muestras de variables aleatorias independientes y con idéntica distribución, con media μ y varianza σ^2, la *media de la muestra*

$$\bar{X}(n) = \frac{\sum_{i=1}^{n} X_i}{n} = \hat{\mu}$$

es un estimador de μ, es decir $E[\bar{X}(n)] = \mu$.

También se cumple que

$$S^2(n) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} = \hat{\sigma}^2$$

es un estimador de σ^2, es decir $E[S^2(n)] = \sigma^2$.

Estimación de media y varianza

Si utilizamos solamente $\bar{X}(n)$ como estimador de μ no es suficiente, pues es una variable aleatoria que tiene varianza, por lo tanto en una muestra puede estar cerca de μ y en otra no.

Por lo tanto es necesario construir un *intervalo de confianza*, para lo cual necesitamos calcular $Var[\bar{X}(n)]$.

$$Var[\bar{X}(n)] = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

Remplazando σ^2 por su estimador, tenemos que

$$V\hat{a}r[\bar{X}(n)] = \frac{S^2(n)}{n} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n(n-1)}$$

Pero dado que las diferentes X_i están correlacionadas, para calcular las medidas de interés se debe agrupar salidas en la forma de nuevas observaciones que sí sean independientes (siguiente clase).

Intervalo de confianza

Sea la variable aleatoria $Z_n = [\bar{X}(n) - \mu]/\sqrt{\frac{\sigma^2}{n}}$.

El *teorema central del límite* dice que la distribución de Z_n tiende a una distribución normal estándar ($\mu = 0$ y $\sigma^2 = 1$) cuando n tiende a infinito (muchas observaciones).

Sustituimos σ^2 por su estimador $S^2(n)$ y se cumple que

$$P(-z_{1-\alpha/2} \leq \frac{\bar{X}(n)-\mu}{\sqrt{S^2(n)/n}} \leq z_{1-\alpha/2}) =$$
$$P[\bar{X}(n) - z_{1-\alpha/2}\sqrt{S^2(n)/n} \leq \mu \leq \bar{X}(n) + z_{1-\alpha/2}\sqrt{S^2(n)/n}] \approx 1 - \alpha$$

donde $z_{1-\alpha/2}$ es el $1 - \alpha$ punto crítico de la distribución normal estándar (valor tabulado).

Intervalo de confianza

Para n suficientemente grande el intervalo de confianza para μ es

$$\bar{X}(n) \pm z_{1-\alpha/2}\sqrt{S^2(n)/n}$$

Si no es posible tener un n grande (por ejemplo, si cada ejecución de la simulación es costosa), el intervalo se calcula como

$$\bar{X}(n) \pm t_{n-1,1-\alpha/2}\sqrt{S^2(n)/n}$$

donde $t_{n-1,1-\alpha/2}$ es el $1-\alpha$ punto crítico de la distribución t-student con $n-1$ grados de libertad (valor tabulado). Es una estimación más conservadora, el intervalo es exacto (no una tendencia) y vale para sumas de variables aleatorias con distribución normal.

En ambos casos hay errores de aproximación que deben tenerse presentes.

Ley de los grandes números

Si tenemos $X_1, X_2 \ldots X_n$ variables aleatorias independientes y con idéntica distribución, con media μ, entonces $\bar{X}(n)$ tiende a μ con probabilidad 1 cuando n tiende a infinito.

Implicancia en simulación: si realizamos un número infinito de experimentos, donde cada uno resulta en un $\bar{X}(n)$, y n es suficientemente grande, entonces $\bar{X}(n)$ estará arbitrariamente cerca de μ en casi todos los experimentos.