Simulación a Eventos Discretos

Tema 9: Reducción de varianza

Respuestas de interés, estimaciones de parámetros

Un programa de simulación es un sistema de muestreo estadístico basado en computadora. En base a la muestra obtenida se estiman medidas sobre toda la población.

En la mayoría de las simulaciones, los experimentos tienen por objetivo obtener valores medios de los resultados que se muestrean (de sus distribuciones), además de saber en qué medida los valores se alejan de las medias.

La dispersión de valores alrededor de la media se cuantifica en base a un intervalo de confianza, que es proporcional a la varianza.

La varianza se estima en base a varias replicaciones y su valor disminuye según se realicen más replicaciones (ver tema 5).

Métodos de reducción de varianza

En algunos contextos no es viable aumentar de forma arbitraria la cantidad de replicaciones independientes, dado que la ejecución de cada una es computacionalmente costosa.

Entonces, cuando no hay margen para mejorar la eficiencia computacional del modelo, buscamos eficiencia estadística en el diseño de experimentos y su análisis.

Lo anterior constituye la motivación para el surgimiento y desarrollo de los métodos de reducción de varianza.

El método a utilizar depende del problema, no hay garantías de que funcione (incluso los resultados podrían ser peor). Antes de aplicar un método se debe realizar una prueba preliminar para determinar si realmente contribuye a reducir la varianza.

Método de torrentes comunes

Aplica a simulaciones comparativas.

Idea básica: las configuraciones alternativas se deben comparar bajo las mismas condiciones experimentales; las diferencias serán debidas a las configuraciones más que a las fluctuaciones de las condiciones.

Condiciones experimentales en SED: muestras aleatorias generadas.

Torrentes comunes: justificación teórica

Sean X_{1j} y X_{2j} observaciones de las configuraciones 1 y 2, respectivamente, correspondientes a la replicación independiente $j = 1, 2, \ldots n$.

Buscamos estimar $\xi = \mu_1 - \mu_2 = X_{1j} - X_{2j}$, en base a $Z_j = X_{1j} - X_{2j}$ para n replicaciones independientes de las dos configuraciones.

Entonces $E(Z_j) = \xi$ y $\bar{Z}(n) = \sum_{j=1}^{n} Z_j/n$ es un estimador de ξ.

Dado que las Z_j son i.i.d., se cumple que

$$Var[\bar{Z}(n)] = \frac{Var(Z_j)}{n} = \frac{Var(X_{1j}) + Var(X_{2j}) - 2Cov(X_{1j}, X_{2j})}{n}$$

Torrentes comunes: justificación teórica

Si las simulaciones de las diferentes configuraciones son independientes (es decir, con diferentes torrentes para una misma distribución presente en las dos configuraciones), la covarianza $Cov(X_{1j}, X_{2j})$ será nula.

Si es posible ejecutar las diferentes configuraciones de forma que X_{1j} y X_{2j} estén positivamente correlacionadas, entonces $Cov(X_{1j}, X_{2j}) > 0$ y por lo tanto la varianza del estimador $\bar{Z}(n)$ se reduce.

Esto es posible debido a la naturaleza determinística y reproducible de los generadores de números seudoaleatorios.

Torrentes comunes: consideraciones prácticas

Se debe dedicar el mismo torrente de números seudoaleatorios (misma semilla del generador) a la misma distribución en las dos configuraciones.

Se deben dedicar diferentes torrentes a las diferentes distribuciones del modelo y a las diferentes replicaciones independientes de las dos configuraciones.

Entonces, el número total de torrentes necesarios es $d \times n$, donde d es la cantidad de distribuciones distintas en el modelo y n es la cantidad de replicaciones.

Notar que no es inmediato disponer de diferentes torrentes independientes en un generador de números seudoaleatorios. La selección arbitraria de dos semillas no garantiza obtener torrentes independientes entre sí. Algunas implementaciones de generadores incluyen una funcionalidad para sobreponerse a esta dificultad.

Torrentes comunes: consideraciones prácticas

La media de la diferencia se estima como $\bar{Z}(n)$ (diapositiva 5) y su varianza con la fórmula de $S^2(n)/n$. Ver Tema 5: Análisis de resultados (estimación de medias y varianzas, intervalos de confianza), diapositiva 4.

Para que el método tenga el efecto deseado, el uso de los torrentes debe estar sincronizado. Por ejemplo, en el hospital simple el paciente i-ésimo debe arribar exactamente en el mismo instante de tiempo en ambas configuraciones.

La sincronización total puede requerir mucho trabajo de programación. Alternativamente, una sincronización parcial (de algunas distribuciones) puede ser aceptable.

Si el método no se implementa debidamente, la varianza puede aumentar en lugar de disminuir.

Torrentes comunes: consideraciones prácticas

El método logra el efecto deseado si $S_Z^2(n) < S_1^2(n) + S_2^2(n)$, donde:

- $S_Z^2(n)$ se calcula en base a las $Z_j = X_{1j} - X_{2j}$.
- $S_1^2(n)$ se calcula en base a las X_{1j}.
- $S_2^2(n)$ se calcula en base a las X_{2j}.

Método de variables antitéticas

Aplica a simulaciones de un sistema único.

Al igual que en torrentes comunes, la idea básica es introducir correlación, pero en este caso negativa.

Se basa en pares de corridas: si en una se obtuvieron valores muy altos de las respuestas de interés, se trata de generar otra corrida que produzca valores muy bajos.

Esta correlación negativa contribuye a que en el promedio, los resultados estén más cercanos a la media de interés.

Variables antitéticas: implementación en SED

Se utilizan números seudoaletorios complementarios.

Si el número U_k uniformemente distribuido en $(0, 1)$ se usó para un propósito particular (por ejemplo, para generar un tiempo de estadía), entonces en otra replicación (complementaria) se usa $1 - U_k$.

El efecto deseado puede lograrse si se utiliza el método de la transformación inversa para obtener una muestra de una varibable aleatoria a partir de un número seudoaletorio en el intervalo $(0, 1)$.

Variables antitéticas: justificación teórica

Se ejecutan n pares de replicaciones, que resultan en pares de respuestas $(X_1^{(1)}, X_1^{(2)}), \ldots (X_n^{(1)}, X_n^{(2)})$, donde $X_j^{(1)}$ es la respuesta del j-ésimo par que usa U_k y $X_j^{(2)}$ es la que usa $1 - U_k$, denominada antitética.

Ambas respuestas son válidas, en particular $E(X_j^{(1)}) = E(X_j^{(2)}) = \mu$, cada par j es independiente de otro j' y la cantidad total de replicaciones es $2n$.

Para $j = 1, 2, \ldots n$, sea $X_j = (X_j^{(1)} + X_j^{(2)})/2$ y el promedio de las X_j, $\bar{X}(n)$ el estimador de $\mu = E(X_j^{(l)}) = E(X_j) = E[\bar{X}(n)]$, con $l = 1, 2$.

Dado que las X_j son i.i.d. se cumple que

$$Var[\bar{X}(n)] = \frac{Var(X_j)}{n} = \frac{Var(X_j^{(1)}) + Var(X_j^{(2)}) + 2Cov(X_j^{(1)}, X_j^{(2)})}{4n}$$

Si se induce correlación negativa entre $X_j^{(1)}$ y $X_j^{(2)}$ entonces $Cov(X_j^{(1)}, X_j^{(2)}) < 0$, lo que reduce la varianza de la estimación.

Variables antitéticas: consideraciones prácticas

La aplicación del método conlleva la realización de $2n$ replicaciones.

Se calculan los n valores X_j y luego se estiman la media y la varianza mediante las fórmulas habituales $\bar{X}(n)$ y $S^2(n)$, respectivamente.

Para las replicaciones j se utilizan diferentes torrentes de números seudoaleatorios (se necesitan n); para las replicaciones 1 y 2 se utilizan torrentes antitéticos.

Para comprobar la magnitud de la reducción de la varianza, se hace una serie de $2n$ replicaciones independientes sin torrentes antitéticos, se necesitan $2n$ torrentes. Luego se estiman $\bar{X}(n)$ y $S^2(n)$ de la forma habitual.

Se debe garantizar que torrentes antitéticos generan respuestas opuestas. Esto puede asegurarse en buena medida utilizando el método de la transformación inversa, que es una función monótona de la entrada.

Método de variables de control

También trata de aprovechar la correlación entre variables.

Sea X la variable aleatoria que representa un resultado y supongamos que queremos estimar $E(X) = \mu$.

Supongamos también que en la simulación hay otra variable Y que está correlacionada con X y que conocemos $E(Y) = \nu$.

Utilizaremos nuestro conocimiento sobre Y, en el sentido de que acercará X a su media μ, reduciendo su variabilidad.

Por eso llamamos a Y *variable de control* de X; la utilizaremos para ajustar X, es decir para controlarla parcialmente.

Variables de control: ejemplo

En el hospital simple, si nuestra respuesta de interés es el largo de la cola X (con valor medio desconocido μ), podemos utilizar el tiempo entre arribos Y (con valor medio conocido ν) como variable de control. El valor exacto ν lo conocemos, dado que es un parámetro de entrada de la simulación (por ejemplo, la media de una distribución exponencial negativa, estimada a partir de datos de entrada). En cada corrida de la simulación, el tiempo medio entre arribos observado, presentará una discrepancia con respecto al valor esperado ν. Esta discrepancia se utiliza para corregir la estimación de la respuesta de interés, en el método de variables de control.

Variables de control: justificación teórica

Sea a una constante a determinar, que tiene el mismo signo que la correlación entre X e Y.

$(Y - \nu)$ es la desviación de Y con respecto de su media ν.

Calculamos el valor controlado de X como $X_c = X - a(Y - \nu)$.

Si X e Y están correlacionados positivamente, entonces $a > 0$, por lo que ajustaremos X hacia abajo si $Y > \nu$ y hacia arriba si $Y < \nu$.

Si X e Y están correlacionados negativamente haremos lo opuesto.

Variables de control: justificación teórica

Teniendo en cuenta que $X_c = X - a(Y - \nu)$, $E(X) = \mu$ y $E(Y) = \nu$, entonces para cualquier a se cumple

- $E(X_c) = \mu$
- $Var(X_c) = Var(X) + a^2 Var(Y) - 2aCov(X, Y)$,

por lo tanto la dispersion de X_c es menor que la de X si y solo si $2aCov(X, Y) > a^2 Var(Y)$.

Lo importante de este método es elegir el mejor valor de a, de forma de minimizar el valor de $Var(X_c)$. Para eso, consideramos la expresión dependiente de a, derivamos y obtenemos el mejor valor $a^* = Cov(X, Y)/Var(Y)$.

Variables de control: justificación teórica

Considerando las expresiones de $Var(X_c)$ y a^* llegamos a

$$Var(X_c) = Var(X) - \frac{[Cov(X,Y)]^2}{Var(Y)}.$$

Sin embargo, no siempre conocemos $Var(Y)$ (dependiendo de su naturaleza) y generalmente desconocemos $Cov(X,Y)$, por lo tanto es imposible calcular el valor exacto a^*. Entonces, utilizamos un método alternativo para estimarlo, usando los datos de la propia simulación.

Variables de control: consideraciones prácticas

Supongamos que realizamos n corridas independientes para obtener las observaciones $X_1, X_2, \ldots X_n$ de X cuyo promedio es $\bar{X}(n)$, así como $Y_1, Y_2, \ldots Y_n$ de Y cuyo promedio es $\bar{Y}(n)$.

Sea $S_Y^2(n) = \sum_j [X_j - \bar{X}(n)]^2/(n-1)$ la varianza estimada de Y en esas n observaciones.

Estimamos la constante $a^* = \hat{C}_{XY}(n)/S_Y^2$, donde la covarianza es estimada como

$$\hat{C}_{XY}(n) = \frac{\sum_{j=1}^{n}[X_j - \bar{X}(n)][Y_j - \bar{Y}(n)]}{n-1}$$

Luego calculamos el valor corregido $\bar{X}_c(n) = \bar{X}(n) - a^*(n)[\bar{Y}(n) - \nu]$.

Variables de control: consideraciones prácticas

Este método es útil para estimar de forma más precisa la media puntual μ, pero si queremos calcular intervalos de confianza necesitamos varias muestras independientes de la media corregida.

Si la media corregida se calcula a partir n replicaciones y quiero obtener m muestras independientes, en total necesito $n \times m$ replicaciones.

Por lo tanto, al aplicar este método se deben tener en cuenta simultáneamente los aspectos de eficiencia computacional y eficiencia estadística de la simulación.

Otros métodos

Estimación indirecta, condicionamiento.

Se basan en la propiedad de que una parte de la estimación puede realizarse a partir de un valor fijo conocido.