

Curso:
Métodos de Monte Carlo
**Unidad 5, Sesión 14: Métodos para aumentar
la eficiencia (2)**

Departamento de Investigación Operativa
Instituto de Computación, Facultad de Ingeniería
Universidad de la República, Montevideo, Uruguay

dictado semestre 1 - 2024

Contenido:

1. Variables de control.
2. Muestreo estratificado.
3. Inducción de correlaciones.

Muestreo utilizando variables de control (control variates)

Mientras que el muestreo según importancia (visto en la sesión previa) altera la distribución de muestreo de \mathbf{Z} para lograr una mejora en la eficiencia computacional, el método de variables de control mantiene la distribución de muestreo, y mejora la eficiencia explotando la covarianza entre $\kappa(\mathbf{z})$ y una función auxiliar $v(\mathbf{z})$ para la cual el valor de la integral $\theta = \int_{\mathbf{Z}} v(\mathbf{z})dF(\mathbf{z})$ es conocido.

El teorema siguiente describe las ideas principales del método.

Teorema 1. Sean X e Y dos variables aleatorias con medias respectivas μ_X y μ_Y , varianzas σ_X^2, σ_Y^2 , y covarianza σ_{XY} . Entonces:

1.

$$W(\alpha) = Y - \alpha(X - \mu_X)$$

es un estimador insesgado de μ_Y para todo $\alpha \in (-\infty, \infty)$, con

$$\text{Var}(W(\alpha)) = \sigma_Y^2 - 2\alpha\sigma_{XY} + \alpha^2\sigma_X^2,$$

y

2. $\alpha^* = \sigma_{XY}/\sigma_X^2$ minimiza la varianza de $W(\alpha)$, con

$$\text{Var}(W(\alpha^*)) = \sigma_Y^2(1 - \rho_{XY}^2) \leq \text{Var}(W(0)) = \text{Var}(Y) = \sigma_Y^2,$$

donde $\rho_{XY} = \sigma_{XY}/\sqrt{\sigma_X^2\sigma_Y^2}$ es la correlación entre X e Y .

Prueba. La primer parte se prueba de manera inmediata, utilizando la linealidad de la esperanza, y las propiedades básicas de la varianza.

Para la segunda parte, alcanza con ver que la expresión de la varianza es una función convexa en α , y anular la derivada para encontrar el valor de α que minimiza esta función.

Si observamos el cociente de las varianzas

$\text{Var}(Y) / \text{Var}(W(\alpha^*)) = 1/(1 - \rho_{XY}^2)$, vemos que cuanto mayor sea la correlación entre X e Y , mayor será el cociente, y por lo tanto se alcanza una mejora de precisión más importante utilizando W en lugar de Y .

Dado que la correlación figura elevada al cuadrado, no importa su signo, sino su valor absoluto.

Por ejemplo, para alcanzar una reducción de varianza de $1/(1 - \rho_{XY}^2) = 2$, es preciso que $|\rho_{XY}| = \sqrt{2}/2 = 0.7071\dots$

Por lo tanto, es preciso buscar variables que tengan un alto nivel de correlación con la variable de interés.

Aunque el valor óptimo de α , α^* , es muy raras veces conocido de manera exacta, es posible emplear datos generados en corridas preliminares de simulación para obtener estimaciones de σ_{XY} y de σ_Y^2 que permitan a su vez estimar α^* .

Volviendo a la expresión de la varianza de W , observamos que

$$\text{Var}(W(\alpha)) = \sigma_Y^2(1 - \rho_{XY}^2 + (\alpha - \alpha^*)^2\sigma_X^2)$$

y por lo tanto podemos reescribir el cociente de varianzas como

$$\text{Var}(Y) / \text{Var}(W(\alpha)) = \text{Var}(W(0)) / \text{Var}(W(\alpha^*)) = \frac{1}{1 - \rho_{XY}^2 + (\alpha - \alpha^*)^2\sigma_X^2 / \sigma_Y^2}.$$

Entonces podemos ver que $\text{Var}(Y) / \text{Var}(W(\alpha)) \geq 1$ si y solo si $\alpha \in [0, 2\alpha^*]$ para $\sigma_{XY} \geq 0$ ó $\alpha \in [2\alpha^*, 0]$ para $\sigma_{XY} \leq 0$. Por lo tanto, mientras que empleemos un valor de α que esté en ese rango, tendremos una mejora de precisión, que a su vez llevará a una mayor eficiencia (fuera de ese rango en cambio estaremos empeorando el desempeño del método).

En particular, supongamos que $\hat{\alpha}$ es un estimador insesgado de α^* , obtenido por ejemplo a través de corridas preliminares. Aplicando la desigualdad de Chebyshev tenemos que

$\text{Prob}(|\hat{\alpha} - \alpha^*| < |\alpha^*|) \geq 1 - \text{Var}(\hat{\alpha}) / (\alpha^*)^2$, por lo tanto, con probabilidad no menor que $1 - \text{Var}(\hat{\alpha}) / (\alpha^*)^2$, $\hat{\alpha}$ pertenecerá al intervalo deseado para lograr una mejora computacional.

Muestreo estratificado

Supongamos que podemos particionar el dominio de integración \mathcal{Z} en r conjuntos disjuntos $\mathcal{Z}_1, \dots, \mathcal{Z}_r$ (por lo tanto, $\mathcal{Z}_i \cap \mathcal{Z}_j = \emptyset$ para $i \neq j$, y $\bigcup_{i=1}^r \mathcal{Z}_i = \mathcal{Z}$).

Podemos entonces escribir $\zeta = \int_{\mathcal{Z}} \kappa(\mathbf{z}) dF(\mathbf{z})$ de la siguiente manera alternativa:

$$\zeta = \sum_{i=1}^r p_i \zeta_i$$

$$\zeta_i = \int_{\mathcal{Z}_i} \kappa(\mathbf{z}) dF_{(i)}(\mathbf{z})$$

$$p_i = \int_{\mathcal{Z}_i} dF(\mathbf{z})$$

$$dF_{(i)} = p_i^{-1} dF(\mathbf{z}), \mathbf{z} \in \mathcal{Z}_i,$$

y

$$p_1 + \dots + p_r = 1.$$

Los conjuntos \mathcal{Z}_i son llamados estratos.

Supongamos que generamos $\mathbf{Z}^{(1,i)}, \dots, \mathbf{Z}^{(n_i,i)}$, n_i muestras independientes del estrato i , para todos los valores $i = 1, \dots, r$. Entonces

$$\bar{\zeta}(n_1, \dots, n_r) = \sum_{i=1}^r p_i/n_i \sum_{j=1}^{n_i} \kappa(\mathbf{Z}^{(j,i)})$$

es un estimador insesgado de ζ , cuya varianza verifica

$$\text{Var}(\bar{\zeta}(n_1, \dots, n_r)) = \sum_{i=1}^r \sigma_i^2 p_i^2 / n_i,$$

donde

$$\sigma_i^2 = \text{E} \left((\kappa(\mathbf{Z}^{(1,i)}) - \zeta_i)^2 \right) = \int_{\mathcal{Z}_i} (\kappa(\mathbf{z}) - \zeta_i)^2 dF_i(\mathbf{z}) = 1/p_i \int_{\mathcal{Z}_i} (\kappa(\mathbf{z}) - \zeta_i)^2 dF(\mathbf{z}).$$

Un método de muestreo estratificado es cualquier plan de muestreo que particiona \mathcal{Z} de esta manera, eligiendo de acuerdo a un cierto criterio el número de muestras que deben ser tomadas en cada estrato; dicho criterio debería ser elegido de tal manera que garantizara una mejora en la varianza del nuevo estimador respecto al estándar.

Si suponemos ya elegida la partición, la pregunta a resolver es entonces como fijar el número de muestras en cada estrato. El siguiente teorema muestra una forma de elegir estos valores que garantiza una reducción de la varianza

Teorema 2. *Para un valor n fijo, la asignación $n_i = np_i$, $1 \leq i \leq r$ asegura que $Var(\bar{\zeta}(n_1, \dots, n_r)) \leq Var(\bar{\zeta}_n)$.*

Prueba. Observemos primero que la asignación verifica que $n = n_1 + \dots + n_r$, dado que los valores p_i son una distribución de probabilidad y por lo tanto su suma es 1.

Volviendo a la fórmula de la varianza presentada recientemente,

$$\begin{aligned}
\text{Var}(\bar{\zeta}(n_1, \dots, n_r)) &= \sum_{i=1}^r \sigma_i^2 p_i^2 / n_i \\
&= \sum_{i=1}^r \sigma_i^2 p_i / n \\
&= 1/n \sum_{i=1}^r \int_{\mathcal{Z}_i} (\kappa(\mathbf{z}) - \zeta_i)^2 dF_{(i)}(\mathbf{z}) \\
&= 1/n \sum_{i=1}^r \int_{\mathcal{Z}_i} (\kappa(\mathbf{z}) - \zeta + \zeta - \zeta_i)^2 dF_{(i)}(\mathbf{z}) \\
&= \text{Var}(\bar{\zeta}_n) - 1/n \sum_{i=1}^r p_i (\zeta - \zeta_i)^2 \\
&\leq \text{Var}(\bar{\zeta}_n).
\end{aligned}$$

Estudiando estas fórmulas, vemos que esta asignación de valores para las muestras en cada estrato nunca empeorarán la varianza, y que en realidad, a menos que el valor de ζ_i en cada estrato i sea idéntico a ζ , la varianza del muestreo estratificado será estrictamente menor que la varianza del método clásico.

Por otro lado, hay una indicación útil para la elección de los estratos, ya que de esa misma fórmula resulta que cuanto más dispersos sean los valores ζ_i alrededor de ζ , mayor será la reducción de la varianza alcanzada, por lo tanto es interesante buscar estratos que difieran lo más posible en este sentido.

Estimación de la varianza y construcción de intervalos de confianza

Recordemos que denominamos $\mathbf{Z}^{(j,i)}$, con $j \in 1, \dots, n_i$, a las muestras independientes del estrato i , para todos los valores $i = 1, \dots, r$.

Definimos

$$\hat{\zeta}_i = \sum_{j=1}^{n_i} \kappa(Z^{(j,i)}) / n_i,$$

y

$$\hat{\sigma}_i^2 = \sum_{j=1}^{n_i} [\kappa(Z^{(j,i)}) - \hat{\zeta}_i]^2 / (n_i - 1).$$

Entonces el estimador puntual estratificado es

$$\bar{\zeta}(n_1, \dots, n_r) = \sum_{i=1}^r p_i \hat{\zeta}_i,$$

y

$$\text{Var}(\zeta(n_1, \dots, n_r)) = \sum_{i=1}^r p_i^2 \hat{\sigma}_i^2 / n_i$$

es un estimador insesgado de su varianza, que puede utilizarse para construir el intervalo de confianza usual basado en la aproximación normal.

Inducción de correlaciones

Los métodos que vimos hasta ahora están todos basados en el principio de emplear replicaciones independientes. Sin embargo, si se tiene información suficiente sobre la estructura del problema, es posible utilizar replicaciones con dependencias, y aprovechar las mismas para inducir una mejora de la eficiencia.

Supongamos que tenemos dos variables aleatorias X e Y con medias respectivas μ_X y μ_Y , varianzas σ_X^2, σ_Y^2 , y covarianza σ_{XY} .

Sea $W = X + Y$ y $Z = X - Y$. Si X e Y son independientes, entonces

$$\sigma_{XY} = 0,$$

$$\text{Var}(W) = \sigma_X^2 + \sigma_Y^2,$$

$$\text{Var}(Z) = \sigma_X^2 + \sigma_Y^2.$$

En cambio, si X e Y son dependientes, entonces $\sigma_{XY} \neq 0$, y

$$\text{Var}(W) = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY},$$

$$\text{Var}(Z) = \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}.$$

Si X e Y son independientes e idénticamente distribuidas, entonces $W/2$ es un estimador insesgado de su media común $\mu_X = \mu_Y$, con varianza $\sigma^2/2$, con $\sigma^2 = \sigma_X^2 = \sigma_Y^2$.

Sin embargo, si son idénticamente distribuidas pero dependientes, entonces $W/2$ es un estimador insesgado de su media común $\mu_X = \mu_Y$, pero su varianza será $(\sigma^2 + \sigma_{XY})/2$. Por lo tanto, si logramos utilizar un plan de muestreo tal que $\sigma_{XY} < 0$, obtendremos una reducción de la varianza, sin incurrir costos adicionales a los de efectuar replicaciones independientes.

En otras circunstancias, lo que interesa es estimar la diferencia $\mu_X - \mu_Y$. En este caso, si X e Y son independientes, entonces Z es un estimador insesgado de $\mu_X - \mu_Y$, con varianza $\sigma_X^2 + \sigma_Y^2$.

Si en cambio X e Y son dependientes, entonces Z es un estimador insesgado de $\mu_X - \mu_Y$, pero su varianza es con varianza $\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$. Esto indica que si encontramos un plan de muestreo que induce $\sigma_{XY} > 0$, obtendremos una reducción de la varianza.

No es trivial encontrar planes de muestreo que verifiquen estas condiciones. En el caso unidimensional, y cuando se emplea un método de generación de distribuciones a partir de números aleatorios que garantiza monotonía (en particular, cuando se utiliza el método de la transformada inversa), es posible demostrar que el máximo de σ_{XY} se alcanza cuando para generar X e Y se utiliza el mismo número uniforme U , método conocido como *torrentes comunes* (reuso de la misma secuencia de números pseudoaleatorios), y que el mínimo de σ_{XY} se alcanza cuando para generar X se emplea un número U y para generar Y se emplea su complemento a 1, $1 - U$ (que será también un número aleatorio, de distribución uniforme), método conocido como *muestreo antitético*.

Por lo tanto, el muestreo antitético será útil para estimar una media $\mu = \mu_X$, basado en generar valores de $W/2 = (X + Y)/2$; y el método de

secuencias aleatorias comunes (o compartidas) para estimar una diferencia $\mu_X - \mu_Y$.

Es posible generalizar estas demostraciones para ciertos casos multi-dimensionales, aunque requiere hipótesis de monotonía más difíciles de verificar en este caso. s También es posible generalizar las fórmulas a casos en los que se induce dependencia entre $n > 2$ replicaciones.

Damos a continuación el esquema de Monte Carlo antitético para el caso en que la distribución de las variables es uniforme en el hipercubo, el mismo es aplicable en el caso general, sustituyendo la generación en el hipercubo por la generación de acuerdo a las distribuciones F_X correspondientes.

Procedimiento MonteCarloAntitetico (función φ , entero n , real $1 - \delta$)

Entrada: función a integrar φ , n tamaño de la muestra, $1 - \delta$ nivel de confianza

Salida: $\bar{\zeta}_n$ estimación de la integral, $V[\bar{\zeta}_n]$ estimación de la varianza

1. $S = 0; T = 0; /*$ Inicialización $*/$
2. For $j = 1, \dots, n$ do
 - 2.1 Sortear $\mathbf{X}^{(j)}$ con distribución uniforme en \mathcal{J}^m ;
 - 2.2 Calcular $\mathbf{Y}^{(j)} = 1 - \mathbf{X}^{(j)}$ (entendiendo esta notación como el cálculo componente a componente del complemento a 1)
 - 2.4 $W = (\varphi(\mathbf{X}^{(j)}) + \varphi(\mathbf{Y}^{(j)}))/2$;
 - 2.2 $S = S + W; /*$ Acumular $*/$
 - 2.3 If $j > 1$ then $T = T + (1 - 1/j) (W - S/(j - 1))^2$.
3. $\bar{\zeta}_n = S/n; /*$ Estimador puntual de $\lambda(\mathcal{R})$ $*/$
4. $\hat{\sigma}_n^2 = T/(n - 1); /*$ Estimador puntual de la varianza $*/$
5. $V[\bar{\zeta}_n] = \hat{\sigma}_n^2/n; /*$ Estimador puntual de la varianza de $\bar{\zeta}_n$ $*/$
6. Calcular $[I_1(S, n, \delta), I_2(S, n, \delta)] /*$ un intervalo de confianza de nivel $(1 - \delta)$ para $\zeta(\mathcal{R})$ $*/$

Preguntas para auto-estudio

- ¿Cómo funciona el muestreo utilizando variables de control? ¿Porqué influye la correlación entre las variables en la reducción de varianza obtenida?
- ¿Qué es el muestreo estratificado? ¿Cuál sería la manera más eficiente de definir los estratos?
- ¿Qué es el muestreo antitético y qué es el muestreo utilizando secuencias aleatorias comunes?

Obligatorio 7

Ejercicio 14.1 (grupal): Partiendo de uno de los códigos elaborados para resolver el ejercicio 6.2, utilizar el método de muestreo estratificado para calcular la integral de la función $x_1x_2x_3x_4x_5$ sobre el hipercubo \mathcal{J}^m de dimensión $m = 5$ en base a 10^6 iteraciones. Calcular media, desviación estándar y un intervalo de confianza de nivel 95%.

Comparar con los resultados obtenidos con el código del ejercicio 6.2.

Sugerencia: definir 5 estratos, en función del valor de x_5 , tomando los siguientes intervalos:

$[0, 0.72)$, $[0.72, 0.83)$, $[0.83, 0.90)$, $[0.90, 0.95)$, $[0.95, 1]$. Hacer dos experimentos, uno tomando $10^6/5$ iteraciones en cada estrato, otro tomando una cantidad de iteraciones proporcional a la probabilidad de cada estrato.

Fecha entrega: Ver cronograma del curso.