

Curso:
Métodos de Monte Carlo
Unidad 5, Sesión 13: Métodos para mejorar la
eficiencia

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

dictado semestre 1 - 2016

Contenido:

1. Introducción.
2. Muestreo según importancia.

Introducción

Como ya discutimos en la sesión 6, en su forma más general los métodos de Monte Carlo se aplican para calcular la integral de Lebesgue-Stieltjes

$$\zeta = \int_{\mathcal{Z}} \kappa(\mathbf{z}) dF(\mathbf{z}),$$

donde $F()$ es una función de distribución de un vector aleatorio en la región m -dimensional $\mathcal{Z} \subseteq \mathbb{R}^m$, y $\kappa()$ una función arbitraria en ese dominio (que llamamos kernel).

En la sesión 11 indicamos que existen muchas funciones de distribución y kernels alternativos que satisfacen la misma expresión, y que por lo tanto pueden ser utilizados como base de un método Monte Carlo para calcular ζ .

Vamos a llamar *plan de muestreo* a cualquier pareja $\{F(\mathbf{z}), \kappa(\mathbf{z}); \mathbf{z} \in \mathcal{Z}\}$, ya que efectivamente son la base de un muestreo que lleva a una estimación insesgada de ζ .

En esta sesión (y las siguientes), el foco está en discutir distintas opciones de planes de muestreo, de manera de mejorar la eficiencia computacional.

Cuando se formula un problema de integración multi-variable que puede ser resuelto por Monte Carlo, suele existir un plan de muestreo (dado explícita o implícitamente) que es el más natural teniendo en cuenta el contexto en el que el problema surge y es definido. En muchos casos, es posible aplicar directamente este plan de muestreo, si los recursos computacionales existentes son suficientes para obtener con el mismo la precisión deseada (lo que teniendo en cuenta las reducciones de costo y aumentos de potencia de las computadoras, se da en un número cada vez más elevado de casos).

Sin embargo, existen casos en los que la precisión deseada no es alcanzable en un tiempo razonable con el plan de muestreo inicial, y es entonces necesario estudiar que alternativas existen.

Todo plan de muestreo conlleva cuatro costos (en tiempo). Tenemos el costo de inicialización T_1 , en el cual se inicializa el modelo y calculan las

variables y estado inicial. Luego, para cada replicación, tendremos la generación de las muestras a partir de la distribución $F(\mathbf{z})$, de tiempo T_2 por muestra, y la evaluación de la función $\kappa(\mathbf{z})$, con un tiempo T_3 por evaluación. Finalmente, existe una etapa de finalización, que calcula e imprime los valores relevantes, con tiempo T_4 . Por lo tanto, n replicaciones independientes llevarán un tiempo $T_1 + n(T_2 + T_3) + T_4$.

Si tenemos entonces dos planes de muestreo l y m , y queremos alcanzar una cierta precisión (ϵ, δ) , necesitaremos n_l y n_m replicaciones independientes respectivamente, con tiempos de ejecución $T_{1l} + n_l(T_{2l} + T_{3l}) + T_{4l}$ y $T_{1m} + n_m(T_{2m} + T_{3m}) + T_{4m}$. Preferiremos el plan cuyo tiempo sea entonces el menor para el problema específico que estamos tratando.

Como hemos visto, en general el plan que induce el menor valor para la varianza $\text{Var}(\kappa(\mathbf{z}))$ es el que requiere menor número de replicaciones

Esta observación ha motivado la búsqueda de planes de muestreo que disminuyan la varianza, por lo que el tema de mejora de eficiencia computacional en Monte Carlo ha recibido tradicionalmente el nombre de

reducción de la varianza (y con este nombre se encuentra en la mayoría de los libros y material publicado sobre el tema).

Sin embargo, esta terminología es engañosa, ya que puede perfectamente pasar que un plan de muestreo l tenga una varianza mucho menor que otro plan m , que por lo tanto n_l sea menor que n_m , pero que los costos asociados al muestreo y cálculo del kernel para l sean mucho mayores, y por lo tanto que el costo total $T_{1l} + n_l(T_{2l} + T_{3l}) + T_{4l}$ sea mayor que $T_{1m} + n_m(T_{2m} + T_{3m}) + T_{4m}$.

Esta sesión y las siguientes describen los métodos generales más empleados para aumentar la eficiencia. Estos son:

- Muestreo según importancia, que funciona alterando la distribución $F(\mathbf{z})$.
- Variables de control, que mantiene la misma distribución, pero modifica el kernel $\kappa(\mathbf{z})$.

- Muestreo correlacionado (que incluye muestreo antitético y muestreo con re-uso de secuencias de números aleatorios).
- Muestreo estratificado, que particiona la región \mathcal{Z} y realiza un muestreo por región.

Existen otros métodos generales, así como métodos específicos derivados para problemas particulares; pero los mencionados son los de mayor rango de aplicabilidad, por lo que nos concentraremos en los mismos.

Dado que los tiempos de ejecución dependen de la implementación concreta, el lenguaje empleado, y el tamaño de la instancia, resulta imposible en una discusión tan general como la que daremos tenerlos en cuenta de manera detallada, por lo que efectivamente la base de discusión y comparación será la varianza y las reducciones que pueden alcanzarse de la misma; queremos sin embargo enfatizar lo ya dicho, respecto a la importancia de los tiempos de generación de variables y evaluación de funciones, para determinar de manera completa la eficiencia de un método.

Muestreo según importancia

Consideremos dos planes $\{F(\mathbf{z}), \kappa(\mathbf{z}); \mathbf{z} \in \mathcal{Z}\}$ y $\{F^*(\mathbf{z}), \kappa^*(\mathbf{z}); \mathbf{z} \in \mathcal{Z}\}$ para calcular ζ , diferenciables en todo $z \in \mathcal{Z}$ con densidades de probabilidad f y f^* respectivamente, y tales que $f^*(\mathbf{z}) > 0$ para todo $z \in \mathcal{Z}$ (el requerimiento de diferenciability se toma para conveniencia de exposición, pero el método vale también para distribuciones discretas).

Definimos

$$\kappa^*(\mathbf{z}) = \kappa(\mathbf{z})R(\mathbf{z}),$$

donde $R(\mathbf{z})$ es la *función de importancia*

$$R(\mathbf{z}) = f(\mathbf{z})/f^*(\mathbf{z}), \mathbf{z} \in \mathcal{Z}.$$

Si \mathbf{Z} y \mathbf{Z}^* son v.a. con distribuciones F y F^* respectivamente, entonces

$$E_F(\kappa(\mathbf{Z})) = E_{F^*}(\kappa^*(\mathbf{Z}^*)) = \zeta,$$

$$\begin{aligned}\text{Var}_F(\kappa(\mathbf{Z})) &= \int_{\mathcal{Z}} \kappa^2(\mathbf{z}) f(\mathbf{z}) d\mathbf{z} - \zeta^2, \\ \text{Var}_{F^*}(\kappa^*(\mathbf{Z}^*)) &= \int_{\mathcal{Z}} \kappa^2(\mathbf{z}) R(\mathbf{z}) f(\mathbf{z}) d\mathbf{z} - \zeta^2.\end{aligned}$$

Por lo tanto, podemos calcular la diferencia de las varianzas de los dos planes como

$$\text{Var}_F(\kappa(\mathbf{Z})) - \text{Var}_{F^*}(\kappa^*(\mathbf{Z}^*)) = \int_{\mathcal{Z}} \kappa^2(\mathbf{z}) (1 - R(\mathbf{z})) d\mathbf{z}.$$

Si $\kappa(\mathbf{z}) > 0$, podemos elegir $f^*(\mathbf{z}) = \kappa(\mathbf{z}) f(\mathbf{z}) / \zeta$, que resulta en que $\text{Var}_{F^*}(\kappa^*(\mathbf{Z}^*)) = 0$, siendo por lo tanto el plan de menor varianza posible, y por lo tanto el óptimo.

Es claro que para esto necesitamos conocer ζ , que es la medida a estimar, por lo tanto en la práctica no es posible alcanzar dicho plan. Sin embargo, conocer la forma del mismo resulta de gran interés. En particular, es

posible ver que si se elige $f^*(\mathbf{z})$ de manera que sea mayor que $f(\mathbf{z})$ cuando $\kappa(\mathbf{z})f(\mathbf{z})$ es grande, y que sea menor cuando $\kappa(\mathbf{z})f(\mathbf{z})$ es pequeño, lograremos reducir la varianza del nuevo plan. Dado que esto corresponde a redistribuir la densidad de probabilidad de acuerdo a la importancia relativa medida por $\kappa(\mathbf{z})f(\mathbf{z})$, de allí surge el nombre de *muestreo según importancia* (*importance sampling*).

Veremos a continuación un par de ejemplos donde es aplicable el muestreo según importancia.

Conversión de un integrando no acotado en uno acotado

El muestreo según importancia ofrece importantes beneficios cuando $\{\kappa(\mathbf{z}); \mathbf{z} \in \mathcal{Z}\}$ no es acotado, y se toma como alternativa $\{\kappa^*(\mathbf{z}); \mathbf{z} \in \mathcal{Z}\}$ acotado.

Para dar un ejemplo muy sencillo (que puede ser resuelto por métodos numéricos determinísticos), supongamos que queremos estimar

$$\zeta = \int_0^1 z^{\alpha-1} e^{-z} dz$$

con $1/2 < \alpha \leq 1$.

El plan de muestreo obvio es tomar la distribución uniforme, con $f(z) = 1$ y el kernel $\kappa(z) = z^{\alpha-1} e^{-z}$, para $0 \leq z \leq 1$; donde $\kappa(z)$ tiende a infinito cuando z tiende a 0. Entonces, considerando Z una v.a. uniforme $(0, 1)$, tenemos que

$$E_F(Z) = \zeta$$

y

$$\text{Var}_F(\kappa(Z)) = \int_0^1 z^{2(\alpha-1)} e^{-2z} dz - \zeta^2.$$

Alternativamente, podemos tomar $f^*(z) = \alpha z^{\alpha-1}$ (que corresponde a la distribución Beta con parámetros α y 1), y kernel $\kappa^*(z) = (1/\alpha)e^{-z}$, con $0 \leq z \leq 1$, de tal forma que $\kappa^*(z)$ es ahora acotado.

Entonces, considerando ahora Z^* una v.a. con distribución Beta($\alpha, 1$), tenemos que

$$\mathbb{E}_F^*(Z^*) = \zeta$$

y

$$\text{Var}_{F^*}(\kappa^*(Z^*)) = 1/\alpha \int_0^1 z^{\alpha-1} e^{-2z} dz - \zeta^2.$$

Si comparamos numéricamente las dos varianzas a través del coeficiente de varianzas, $\text{Var}_F(\kappa(Z))/\text{Var}_{F^*}(\kappa^*(Z^*))$, veremos que cuanto más cerca de 5 esté el valor de α , mayor es la reducción de varianza que podemos lograr

con esta nueva formulación (este cociente supera el valor 10 para $\alpha = 0.65$; y el valor 100 para $\alpha = 0.525$).

Si bien el muestreo de una v.a. de distribución Beta toma más tiempo que el de una uniforme, las ganancias en varianza son muy importantes, especialmente cuando α está cercano a $1/2$.

Muestreo según importancia para un problema de conteo

Recordemos el problema de conteo de una unión de conjuntos visto en la sesión 7, en el cual tenemos un conjunto base $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_r$ de cardinal r , y una familia $\mathcal{F} = \{\mathcal{S}_1, \dots, \mathcal{S}_k\}$ de subconjuntos de \mathcal{X} , para la que queremos evaluar $\zeta = |\bigcup_{i=1}^k \mathcal{S}_i|$ = cantidad de elementos de \mathcal{X} que pertenecen al menos a un subconjunto de \mathcal{F} .

El muestreo natural y más directo es el que describimos en la sesión 7, por el cual sorteamos de manera uniforme en \mathcal{X} , es decir asignamos a cada elemento de \mathcal{X} probabilidad $1/r$ de ser elegido, y tomamos $\kappa(\mathbf{x}) = r$ si $\mathbf{x} \in \mathcal{X}'$, con $\mathcal{X}' = \bigcup_{i=1}^k \mathcal{S}_i$, y 0 en otro caso.

Con esta elección, tenemos que

$$\mathbb{E}(\bar{\zeta}_n) = \zeta$$

y que

$$\mathbb{E}(V(\bar{\zeta}_n)) = \text{Var}(\bar{\zeta}_n) = (r - \zeta)\zeta/n.$$

Supongamos que conocemos las cardinalidades de los conjuntos $\mathcal{S}_1, \dots, \mathcal{S}_k$. Entonces es posible utilizar esta información en un esquema de muestreo según importancia para mejorar la eficiencia del método Monte Carlo.

Definamos $w(\mathbf{x}) = |\{i : \mathbf{x} \in \mathcal{S}_i\}|$. Entonces podemos definir una nueva distribución de probabilidad X^* tal que $\text{Prob}(X^* = \mathbf{x}) = w(\mathbf{x})/w$ si $\mathbf{x} \in \mathcal{X}'$, y $\text{Prob}(X^* = \mathbf{x}) = 0$ en otro caso, y donde $w = \sum_{\mathbf{x} \in \mathcal{X}} w(\mathbf{x})$.

Si empleamos el kernel $\kappa^*(\mathbf{x}) = \kappa(\mathbf{x})f(\mathbf{x})/f^*(\mathbf{x})$ (definiendo $0/0 = 0$, entonces

$$\sum_{\mathbf{x} \in \mathcal{X}} \kappa^*(\mathbf{x})f^*(\mathbf{x}) = |\mathcal{X}'| = \zeta.$$

Aunque no conocemos explícitamente $w(\mathbf{x})$, podemos generar elementos con distribución f^* de la siguiente forma:

1. En el primer paso, se selecciona al azar un conjunto J con distribución de probabilidades $\text{Prob}(J = j) = |\mathcal{S}_j| / \sum_{i=1}^k |\mathcal{S}_i|$.

2. Elegir al azar \mathbf{X} de \mathcal{S}_J con probabilidad uniforme en el conjunto (i.e, para todo $\mathbf{x} \in \mathcal{S}_J$, su probabilidad de ser sorteado es $1/|\mathcal{S}_J|$).

Este esquema de muestreo tiene varianza

$$\text{Var}_{F^*}(\kappa^*(\mathbf{X})) \leq \zeta^2(k-1)/2,$$

que no depende de r como el esquema directo.

Preguntas para auto-estudio

- ¿Qué es un plan de muestreo? ¿Qué costos en tiempo de cálculo tiene asociados?
- ¿Cuáles son los métodos generales más empleados para aumentar la eficiencia de un plan de muestreo?
- ¿Cómo es el esquema de muestreo según importancia? ¿Qué ventajas es posible obtener con este esquema?