

Curso:
Métodos de Monte Carlo.
Unidad 2, Sesión 4: Consideraciones sobre el error y el tamaño de la muestra

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

dictado semestre 1 - 2025

Contenido:

1. Error e intervalos de confianza.
2. Determinación de tamaño de muestra.
3. Ejercicio.

Consideraciones sobre el error cometido por un método de Monte Carlo

Tanto en los métodos determinísticos como en los métodos de Monte Carlo, el error disminuye a medida que el número de puntos considerados en la evaluación n aumenta. Sin embargo, la forma de estudiar esta dependencia es distinta en ambos casos. En los métodos de Monte Carlo, debemos enfocarnos en la variación inherente al muestreo estadístico, en lugar de en un estudio de error matemático de la aproximación.

En particular, se sabe que si $\bar{\lambda}(\mathcal{R})$ es la estimación por Monte Carlo de $\lambda(\mathcal{R})$, entonces

$$\text{Prob} \left(\lim_{n \rightarrow \infty} \bar{\lambda}(\mathcal{R}) = \lambda(\mathcal{R}) \right) = 1,$$

lo que se expresa también de manera equivalente como

$$\lim_{n \rightarrow \infty} \bar{\lambda}(\mathcal{R}) = \lambda(\mathcal{R}) \text{ c.p.1 (con probabilidad 1),}$$

$$\lim_{n \rightarrow \infty} \bar{\lambda}(\mathcal{R}) = \lambda(\mathcal{R}) \text{ c.s. (casi seguramente),}$$

$$\lim_{n \rightarrow \infty} \bar{\lambda}(\mathcal{R}) = \lambda(\mathcal{R}) \text{ casi en todas partes.}$$

Esta propiedad corresponde a la forma más fuerte de convergencia probabilística y cuando se verifica para un estimador, como en este caso para $\bar{\lambda}(\mathcal{R})$, se dice que $\bar{\lambda}(\mathcal{R})$ es un estimador fuertemente consistente de $\lambda(\mathcal{R})$.

Sin embargo, aunque esta es una propiedad asintótica muy deseable, la convergencia con probabilidad 1 del estimador no da información sobre el problema de mayor interés práctico, consistente en estimar el error para un valor de n finito.

En contraste con los métodos determinísticos, en el caso de los métodos de Monte Carlo la misma información obtenida al realizar las muestras independientes $\phi(\mathbf{X}^{(j)})$ puede ser utilizada para evaluar la precisión del estimador puntual $\bar{\lambda}(\mathcal{R})$.

En la sesión anterior vimos que $V[\bar{\lambda}(\mathcal{R})] = (S/n)(1 - S/n)/(n - 1)$ (con $S = \sum_{j=1}^n \phi(\mathbf{X}^{(j)})$) era un estimador insesgado de $\text{Var}(\bar{\lambda}(\mathcal{R}))$.

Se emplea muchas veces su raíz cuadrada, $\sqrt{V[\bar{\lambda}(\mathcal{R})]}$, a veces denotada como *error estándar*, como una medida (si bien grosera) del error estadístico cometido en $\bar{\lambda}(\mathcal{R})$. El error estándar es una estimación de la desviación estándar de esta variable aleatoria, y como tal tiene sus mismas unidades, por lo que es directamente comparable con los valores numéricos de la estimación; por esto se suele preferir su uso al de $V[\bar{\lambda}(\mathcal{R})]$ en la presentación de resultados, si bien claramente la información presentada es exactamente la misma, simplemente teniendo en cuenta la operación adicional de tomar raíz cuadrada o elevar al cuadrado.

Sin embargo, es posible calcular de manera bastante directa que $\bar{\lambda}(\mathcal{R})$ y $V[\bar{\lambda}(\mathcal{R})]$ están fuertemente correlacionados, con el siguiente resultado asintótico:

$$\lim_{n \rightarrow \infty} \text{Corr}(\bar{\lambda}(\mathcal{R}), V[\bar{\lambda}(\mathcal{R})]) = \begin{cases} 1 & \text{si } 0 \leq \lambda(\mathcal{R}) < 1/2 \\ 0 & \text{si } \lambda(\mathcal{R}) = 1/2 \\ -1 & \text{si } 1/2 < \lambda(\mathcal{R}) \leq 1 \end{cases}$$

Esto implica que, si $0 \leq \lambda(\mathcal{R}) < 1/2$, una observación $\bar{\lambda}(\mathcal{R})$ menor (mayor) que $\lambda(\mathcal{R})$ tiende a estar acompañada de una observación $V[\bar{\lambda}(\mathcal{R})]$ menor (respectivamente mayor) que $\text{Var}(\bar{\lambda}(\mathcal{R}))$, por lo que se introducen imprecisiones en el conocimiento de la precisión del estimador. En forma similar (si bien simétrica) pasa si $1/2 < \lambda(\mathcal{R}) \leq 1$.

Una consideración de mayor gravedad es que este fenómeno no desaparece o se atenúa cuando $n \rightarrow \infty$, sino por el contrario, se agudiza. Todo esto hace que sólo se pueda considerar $V[\bar{\lambda}(\mathcal{R})]$ como un indicador muy grosero del error del método.

Es conveniente entonces agregar un nuevo paso en el pseudocódigo visto en la unidad anterior para Monte Carlo, específicamente luego del cálculo de $\bar{\lambda}(\mathcal{R})$ y $V[\bar{\lambda}(\mathcal{R})]$ en los pasos 3 y 4, agregar el paso 5 siguiente:

5. Calcular $[I_1(S, n, \delta), I_2(S, n, \delta)]$ /*un intervalo de confianza de nivel $(1 - \delta)$ para $\lambda(\mathcal{R})$ */

Aquí $0 \leq \delta \leq 1$ es un nuevo parámetro de entrada para el método de Monte Carlo. Recordamos la definición de un intervalo de confianza: $[I_1, I_2]$ es un intervalo de confianza de nivel $(1 - \delta)$ para la medida λ (muchas veces expresado porcentualmente, $100(1 - \delta)\%$) ssi I_1 e I_2 son dos variables aleatorias tales que $\text{Prob}(I_1 \leq \lambda \leq I_2) \geq 1 - \delta$.

Dicho de otra manera, si tenemos $[I_1, I_2]$ un intervalo de confianza del 95% para λ , entonces la probabilidad de que λ pertenezca a dicho intervalo es de al menos 0.95 (y la probabilidad de que esté fuera del intervalo es a lo sumo 0.05).

Asociado a la idea de intervalo de confianza, está la de nivel de cobertura del mismo. Se dice que el intervalo *cubre* al parámetro estimado cuando este último pertenece al intervalo. Si repetimos K veces el experimento global, tendremos K valores numéricos para el intervalo de confianza (que pueden verse como muestras del mismo). El *nivel de cobertura* del intervalo es la proporción de veces que la medida exacta perteneció al intervalo, en relación al número total de experimentos globales. Si el intervalo está definido correctamente, esperamos que el nivel de cobertura

sea igual (o mayor) a $1 - \delta$, es decir que si especificamos un intervalo de confianza del 95% y se realizan 1000 experimentos globales, en promedio al menos en 950 de estos λ efectivamente deberían pertenecer al intervalo $[I_1, I_2]$.

Dado un conjunto de experimentos concreto, el nivel de cobertura empírico es el porcentaje de estos experimentos en los que el intervalo de confianza calculado cubrió el valor exacto. En el caso de ejemplo, si se hicieron 1000 experimentos y de los mismos en 965 casos el intervalo de confianza incluyó el valor exacto, observamos una cobertura empírica de 96,5% (por lo tanto, levemente mayor que la especificada). Es claro que la cobertura empírica puede verse también como una variable aleatoria, por lo tanto si hiciéramos otros 1000 experimentos con otras semillas la misma podría cambiar; de todas formas, es un indicador importante para ver si la calidad de la estimación de los intervalos de confianza es buena. Lamentablemente sólo es posible calcular la cobertura empírica si se conoce el valor exacto, esto limita su utilidad a casos muy particulares.

Existen varias fórmulas alternativas para obtener intervalos de confianza

aplicables a las estimaciones realizadas por Monte Carlo, cada una basada en distintas propiedades estadísticas (y eventualmente distintas hipótesis). La misma teoría que las fundamenta se puede utilizar para determinar el tamaño de la muestra necesario para obtener una estimación que cumpla con un cierto criterio de error.

Dado que es algo más sencillo presentar las consideraciones de tamaño de muestra, en esta sesión nos concentraremos en las mismas, y en la próxima sesión nos enfocaremos en el problema de calcular los intervalos de confianza asociados.

Tamaño muestral

Como hemos señalado, es posible derivar por diversos métodos un tamaño de muestra que permita mantener el error dentro de los límites deseados. Estos métodos varían en su generalidad, algunos sólo son válidos bajo ciertas hipótesis que pueden ajustarse más o menos a la realidad. También varían en su grado de conservadurismo; dado un idéntico nivel de error al que aspiramos, los métodos más conservadores indicarán un mayor tamaño de muestra (resultando en un mayor costo computacional). En algunos casos las derivaciones del tamaño de muestra dependen del propio valor λ a estimar, lo que puede ser al mismo tiempo un problema (ya que es un valor desconocido), o una oportunidad (pues puede emplearse información sobre cotas inferiores y superiores, si las hubiera, para obtener el tamaño de muestra apropiado).

Veremos a continuación tres derivaciones:

1. Basada en la desigualdad de Chebyshev.

2. Basada en el Teorema del Límite Central.
3. Basada en un teorema demostrado por Hoeffding en 1963.

Cálculo de n empleando la desigualdad de Chebyshev

Teorema 1. (desigualdad de Chebyshev). Sea Z una v.a. con distribución F_Z definida en $(-\infty, \infty)$, $E(Z) = 0$ y $\sigma^2 = \text{Var}(Z) = E(Z^2) < \infty$. Entonces, para todo $\beta > 0$,

$$\text{Prob} \left(\frac{|Z|}{\sigma} \geq \beta \right) \leq \frac{1}{\beta^2}.$$

En el caso presente, nos interesa calcular el n necesario para garantizar una precisión ϵ , es decir, encontrar el n tal que

$$\text{Prob} (|S/n - \lambda| < \epsilon).$$

Notamos $\bar{\lambda}_n = S/n$, y sabemos que $\sigma^2 = \lambda(1 - \lambda)/n$. Partiendo de la desigualdad de Chebyshev, sustituyendo $Z = \bar{\lambda}_n - \lambda$ y operando, se llega a que

$$\text{Prob} (|\bar{\lambda}_n - \lambda| < \epsilon) \geq 1 - \lambda(1 - \lambda)/(n\epsilon^2).$$

como σ positivo

De esta expresión, podemos evaluar los requerimientos de tamaño muestral. Para esto, es necesario tener como dado conocido los requerimientos de error, dados por una cota del error ϵ y un nivel de confianza $1 - \delta$ con el cual se desea garantizar ese error. Si queremos entonces que

$$\text{Prob} (|\bar{\lambda}_n - \lambda| < \epsilon) \geq 1 - \delta,$$

comparando con la ecuación anterior vemos que es necesario que

$$\delta \geq \lambda(1 - \lambda)/(n\epsilon^2),$$

de aquí despejamos el valor de n requerido, que notaremos $n_C(\epsilon, \delta, \lambda)$, y es igual a

$$n_C(\epsilon, \delta, \lambda) = \lceil \lambda(1 - \lambda)/(\delta\epsilon^2) \rceil.$$

Por el teorema previo, cualquier muestra de tamaño mayor o igual a $n_C(\epsilon, \delta, \lambda)$ garantizará la especificación de error que mencionamos,

$\text{Prob} (|\bar{\lambda}_n - \lambda| < \epsilon) \geq 1 - \delta$, y que notaremos *criterio de error absoluto* (ϵ, δ) .

Queda por resolver el problema de que λ es desconocido. Sin embargo, sabemos que $0 \leq \lambda \leq 1$ (por ser el volumen de una región incluida en el hipercubo \mathcal{J}^m), por lo tanto $\lambda(1 - \lambda) \leq 1/4$, lo que nos lleva al tamaño de muestra en el peor caso (o pesimista) de

$$n_C(\epsilon, \delta) = \lceil 1/(4\delta\epsilon^2) \rceil.$$

para todo $\lambda \in [0, 1]$.

Cálculo de n empleando el Teorema Central del Límite

Como hemos visto, la distribución de probabilidad de $(S - n\lambda)/\sqrt{(n\lambda(1 - \lambda))}$ converge cuando $n \rightarrow \infty$ a la distribución normal,

$$F_N(z) = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-y^2/2} dy,$$

para todo z real. Esta distribución tiene media 0 y varianza 1, y la denotamos por $N(0, 1)$.

Supongamos que tomamos ahora

$$n_N(\epsilon, \delta, \lambda) = \lceil \lambda(1 - \lambda) (\Phi^{-1}(1 - \delta/2)/\epsilon)^2 \rceil$$

donde Φ^{-1} es la inversa de la distribución de la normal (0,1) Entonces, cuando $\epsilon \rightarrow 0$, si tomamos como tamaño de muestra el dado por

$n_N(\epsilon, \delta, \lambda)$, llegamos a la siguiente propiedad:

$$\lim_{\epsilon \rightarrow 0} \text{Prob} (|S/n_N(\epsilon, \delta, \lambda) - \lambda|/\epsilon \leq 1) = 1 - \delta.$$

El tamaño muestral en el peor caso correspondiente a esta fórmula está dado por

$$n_N(\epsilon, \delta) = \lceil (\Phi^{-1}(1 - \delta/2)/2\epsilon)^2 \rceil.$$

Para un mismo criterio de error (ϵ, δ) , el tamaño de muestra preconizado por $n_N(\epsilon, \delta)$ es en general substancialmente menor que el que surge de aplicar el criterio de Chebyshev, $n_C(\epsilon, \delta)$, por lo que ha sido muy común que se emplee el mismo.

Sin embargo, hay varias fuentes importantes de error, por un lado el que surge de aproximar una distribución discreta (la de S) por una distribución continua, por otro el hecho que el tamaño muestral es finito (mientras que los resultados del T.L.C son asintóticos en n), y finalmente el hecho que la

aproximación que hemos visto es también asintótica en ϵ , y si este último valor no es suficientemente pequeño no será válida.

Algunas de estas fuentes de error no son tan significativas cuando λ es cercano a $1/2$, y en cambio pueden generar problemas muy grandes cuando λ se acerca a 0 o a 1 , pero en todos los casos el tamaño de muestra sugerido por n_N puede ser optimista y no garantizar el nivel de cobertura estipulado.

Cálculo de n empleando el Teorema de Hoeffding

Teorema 2. (Hoeffding 1963). Sea Z_1, \dots, Z_n un conjunto de v.a. independientes, tales que $\mu_i = E(Z_i) \in (0, 1)$ y que $\text{Prob}(0 \leq Z_i \leq 1) = 1$, para $i = 1, \dots, n$. Sea $\mu = (\mu_1 + \dots + \mu_n)/n$. Entonces, para $0 < \epsilon \leq 1 - \mu$ y $\bar{Z}_n = (Z_1 + \dots + Z_n)/n$,

$$\text{Prob}(\bar{Z}_n - \mu \geq \epsilon) \leq e^{-nw(\epsilon, \mu)} \leq e^{-ng(\mu)\epsilon^2} \leq e^{-2n\epsilon^2},$$

donde

$$w(\epsilon, \mu) = (\mu + \epsilon)\ln(\mu/(\mu + \epsilon)) + (1 - \mu - \epsilon)\ln((1 - \mu)/(1 - \mu - \epsilon))$$

y

$$g(\mu) = \frac{1}{1 - 2\mu} \ln\left(\frac{1 - \mu}{\mu}\right)$$

si $0 < \mu < 1/2$, y

$$g(\mu) = \frac{1}{2\mu(1 - \mu)}$$

si $1/2 \leq \mu < 1$.

De este teorema, se deriva el siguiente criterio de peor caso para el tamaño de muestra:

$$n_H(\epsilon, \delta) = \lceil \ln(2/\delta)/(2\epsilon^2) \rceil = \lceil 2\ln(2/\delta)/(4\epsilon^2) \rceil.$$

Preguntas para auto-estudio

- ¿Qué es el nivel de cobertura de un intervalo de confianza?
- ¿Qué derivaciones para calcular tamaños de muestras se vieron en esta sesión? ¿Cuáles resultan en valores más conservadores y cuáles en valores más optimistas?

Ejercicio

Entrega 2 - Ejercicio 4.1 (individual):

1. Comparar y discutir la dependencia de los criterios de peor caso n_C , n_N , n_H frente a los parámetros ϵ y δ .
2. Calcular n_C , n_N , n_H para $\epsilon = 0.01$, $\delta = 0.001, 0.01, 0.05$.

Fecha entrega: ver el calendario en el EVA.