# Curso:
# Métodos de Monte Carlo Unidad 2, Sesión 5: Cálculo de intervalos de confianza

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

1er semestre - 2021

# Intervalos de confianza para un tamaño de muestra dado

En la sesión anterior hemos discutido métodos para, dada una especificación de error, fijar el tamaño de muestra $n$ que garantiza alcanzar la misma.

Sin embargo, en muchos casos se fija el tamaño de muestra $n$ a través de otras consideraciones (tiempo de cálculo disponible, por ejemplo), o no hay una especificación de error claramente predeterminada.

En estos casos, en general se plantea el problema inverso: conociendo el tamaño de muestra $n$ prefijado, dado un nivel de confianza deseado, y conociendo los resultados de las replicaciones simuladas, ¿qué precisión podemos adjudicar al resultado obtenido?

Esto equivale al cálculo del intervalo de confianza $[I_1(S, n, \delta), I_2(S, n, \delta)]$, donde los valores conocidos son $n$, el tamaño de muestra, $(1 - \delta)$, el nivel de confianza deseado para el intervalo, y $S$, el número acumulado de replicaciones que cayeron dentro del volumen a estimar (para decirlo de manera más general, la suma de los valores $\phi(\mathbf{X}^{(j)})$, $1 \leq j \leq n$).

En esta sesión discutiremos diversos métodos para calcular $I_1$ e $I_2$, de manera que la probabilidad de que $\lambda(\mathcal{R}) \in [I_1(S, n, \delta), I_2(S, n, \delta)]$ sea mayor o igual a $1 - \delta$.

# Distribución de $S$

En primer término, vamos a estudiar la distribución de la variable aleatoria $S = \sum_{j=1}^{n} \phi(\mathbf{X}^{(j)})$, que acumula los valores calculados a lo largo de la simulación.

Como ya habíamos mencionado, $\phi(\mathbf{X}^{(j)})$ es una variable aleatoria con distribución Bernoulli y parámetro $\lambda(\mathcal{R})$, $Ber(\lambda)$, cuya distribución de probabilidad es la siguiente:

- Prob $\left(\phi(\mathbf{X}^{(j)}) = 1\right) = \lambda$.
- Prob $\left(\phi(\mathbf{X}^{(j)}) = 0\right) = 1 - \lambda$.

La v.a. $S$ es la suma de $n$ v.a. de Bernoulli independientes, y su distribución se conoce con el nombre de distribución Binomial, de parámetros $n$ y $\lambda$, $Bi(n, \lambda)$:

- Prob $(S = k) = C_k^n \lambda^k (1 - \lambda)^{n-k}$, para $k = 0, \ldots, n$.

Es fácil derivar esta distribución, teniendo en cuenta que la probabilidad de que $S$ sea a $k$ es igual a la probabilidad de que $k$ de los valores $\phi(\mathbf{X}^{(j)})$ sean iguales a la unidad, y $n-k$ sean nulos, y esto puede darse de $C_k^n$ (combinaciones de $n$ tomadas de a $k$) formas distintas.

La función de distribución verifica

- $F_{Bi(n,\lambda)}(i) = \mathsf{Prob}\,(S \leq i) = \sum_{k=0}^{i} C_k^n \lambda^k (1-\lambda)^{n-k}$, $i = 1, \ldots, n$.

La esperanza de esta distribución es $\mathsf{E}\,(S) = n\lambda$, y su varianza (que habíamos calculado precedentemente) es $\mathsf{Var}\,(S) = n\lambda(1-\lambda)$.

# Cálculo de intervalo de confianza basado en la distribución binomial

Supongamos que tenemos valores $\alpha_1$ y $\alpha_2$ tales que $0 < \alpha_1 < \alpha_2 < 1$. Definimos

$$\theta_1(s, n, \alpha_1) = [\theta : 1 - F_{Bi(n,\theta)}(s-1) = \alpha_1]$$
$$\theta_2(s, n, \alpha_2) = [\theta : F_{Bi(n,\theta)}(s) = \alpha_2]$$

que existen y están bien definidos (son únicos), y cumplen $0 \leq \theta_1(s, n, \alpha_1) < \theta_2(s, n, \alpha_2) \leq 1$.

Si ocurren exactamente $S$ "éxitos" (valores 1) en la secuencia de $n$ muestras Bernoulli independientes con probabilidad de éxito $\lambda$, entonces para cualquier par de valores $\alpha_1$ y $\alpha_2$ que satisfagan $1 + \alpha_1 - \alpha_2 = \delta$, el intervalo $[\theta_1(S, n, \alpha_1), \theta_2(S, n, \alpha_2)]$ cubre $\lambda$ con coeficiente de confianza (es decir, con probabilidad) igual a
$\text{Prob}\,(\theta_1(S, n, \alpha_1) < \lambda < \theta_2(S, n, \alpha_2)) \geq 1 - \delta$.

La justificación de este resultado se deriva de considerar el testeo de hipótesis simples de la forma $\lambda = \theta$ para cada $\theta \in (0, 1)$. Para un $\theta$

específico, cada test consiste en determinar una región de aceptación $(S_1(\theta, n), S_2(\theta, n))$ con probabilidades de error $\text{Prob}\,(S \geq S_1(\theta, n)) \leq \alpha_1$ y $\text{Prob}\,(S \leq S_2(\theta, n)) \leq 1 - \alpha_2$.

Entonces para un valor de $S$ observado, el intervalo de confianza$[\theta_1(S, n, \alpha_1), \theta_2(S, n, \alpha_2)]$ para $\lambda$ consiste en todos los valores de $\theta$ para los cuales la hipótesis $\lambda = \theta$ sería aceptada. Esto es,

$$[\theta_1(S, n, \alpha_1), \theta_2(S, n, \alpha_2)] = \{\theta \in (0, 1) : S_1(\theta, n) \leq S \leq S_2(\theta, n)\}.$$

Dado que hay una infinidad de valores $\alpha_1 < \alpha_2$ que pueden ser utilizados, es necesario fijar un criterio para elegir. El más común es el de elegirlos de manera que las probabilidades de las colas de la distribución sean iguales, es decir tomar $\theta_1(S, n, \alpha_1)$ y $\theta_2(S, n, \alpha_2)$ con $\alpha_1 = \delta/2$ y $\alpha_2 = 1 - \delta/2$.

Para calcular $\theta_1(S, n, \alpha_1)$ y $\theta_2(S, n, \alpha_2)$, se puede aprovechar la siguiente relación entre la distribución binomial y la distribución Beta:

$$1-F_{Bi(n,\theta)}(i-1) = F_{Beta(i,n-i+1)}(\theta) = \frac{1}{B(i,n-i+1)}\int_0^{\theta} z^{i-1}(1-z)^{n-i}dz,$$

donde $B(,)$ es la función Beta (ver cualquier texto básico de probabilidades o las referencias `http://mathworld.wolfram.com/BetaDistribution.html` y `http://en.wikipedia.org/wiki/Beta_distribution` por más información sobre la distribución Beta y la funcion Beta).

Dado que muchos paquetes estadísticos y varias bibliotecas de distribuciones proveen funciones para calcular la distribución Beta y su inversa, es posible emplear esta relación para determinar los valores de
$\theta_1(S,n,\alpha_1) = F^{-1}_{Beta(S,n-S+1)}(\delta/2)$ y
$\theta_2(S,n,\alpha_2) = F^{-1}_{Beta(S+1,n-S)}(1-\delta/2)$.

Sin embargo, el cálculo de la distribución Beta sufre de inestabilidades numéricas, especialmente cuando $n$ es grande y $S$ es o cercano a $n$ o cercano a $0$. Por lo tanto, es usual emplear otros intervalos de confianza que discutiremos a continuación.

# Cálculo de intervalo de confianza de Wald

Este intervalo de confianza se basa en el empleo directo de la aproximación normal, y es a veces llamado el intervalo de confianza estándar, si bien su validez para distribuciones distintas de la normal no siempre está bien justificada.

Lo presentamos por ser el de cálculo más sencillo, y el más presentado en textos introductorios.

Sea $\bar{\lambda}(\mathcal{R}) = S/n$ el estimador de $\lambda(\mathcal{R})$ obtenido por Monte Carlo, y $V[\bar{\lambda}(\mathcal{R})]$ la estimación de su varianza. El intervalo de Wald se define por

$$\begin{aligned} I_1(S, n, \delta) &= \bar{\lambda}(\mathcal{R}) - \phi^{-1}(1 - \delta/2)\sqrt{V[\bar{\lambda}(\mathcal{R})]} \\ &= \bar{\lambda}(\mathcal{R}) - \phi^{-1}(1 - \delta/2)\sqrt{\bar{\lambda}(\mathcal{R})(1 - \bar{\lambda}(\mathcal{R}))/n}, \end{aligned}$$

$$I_2(S, n, \delta) = \bar{\lambda}(\mathcal{R}) + \phi^{-1}(1 - \delta/2)\sqrt{V[\bar{\lambda}(\mathcal{R})]}$$

$$= \quad \bar{\lambda}(\mathcal{R}) + \phi^{-1}(1-\delta/2)\sqrt{\bar{\lambda}(\mathcal{R})(1-\bar{\lambda}(\mathcal{R}))/n}.$$

Este intervalo se justifica por la observación de que, aplicando el T. del Límite Central, sabemos que $S$ tiende a una distribución normal de media $n\lambda$ y desviación $\sqrt{n\lambda(1-\lambda)}$; por lo tanto, en el límite,

$$\text{Prob}\left(\begin{array}{c} n\lambda - \phi^{-1}(1-\delta/2)\sqrt{n\lambda(1-\lambda)} \\ \leq S \leq \\ n\lambda + \phi^{-1}(1-\delta/2)\sqrt{n\lambda(1-\lambda)} \end{array}\right) =$$

$$= \phi(\phi^{-1}(1-\delta/2)) - \phi(\phi^{-1}(\delta/2)) = 1-\delta.$$

Sin embargo, como mencionamos, el intervalo no resulta completamente satisfactorio en particular cuando $S$ es una v.a. binomial, como en el caso que nos ocupa. Una discusión de este punto se desarrolla en las lecturas complementarias indicadas luego de presentar otros intervalos de confianza alternativos.

# Cálculo de intervalo de confianza empleando la desigualdad de Chebyshev

El siguiente teorema puede ser usado como base para calcular un intervalo de confianza:

**Teorema 1.** *Sea $S$ una v.a. de media $n\lambda$ y varianza $n\lambda(1-\lambda)$, y definamos*

$$\omega_1(z,n,\beta) = \frac{z + \beta^2/2 - \beta\sqrt{\beta^2/4 + z(n-z)/n}}{n+\beta^2}$$

$$\omega_2(z,n,\beta) = \frac{z + \beta^2/2 + \beta\sqrt{\beta^2/4 + z(n-z)/n}}{n+\beta^2}$$

*con $\beta > 0$, $n > 0$ entero, $0 \leq z \leq n$. Entonces existe $\beta = \beta(\delta)$ tal que el intervalo abierto $(\omega_1(z,n,\beta), \omega_2(z,n,\beta))$ cubre $\lambda$ con probabilidad mayor o igual a $1-\delta$.*

**Prueba.** Recordamos la desigualdad de Chebyshev: dada $Z$ una v.a. con distribución $F_Z$ definida en $(-\infty, \infty)$, $\mathsf{E}(Z) = 0$ y $\sigma^2 = \mathsf{Var}(Z) = \mathsf{E}\left(Z^2\right) < \infty$, entonces, para todo $\beta > 0$,

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

Sustituyendo $Z$ por $S/n - \lambda$, y tomando el complemento a 1 de la desigualdad, tenemos que

$$\mathsf{Prob}\left(\frac{|S/n - \lambda|}{\sqrt{\lambda(1-\lambda)/n}} < \beta\right) \geq 1 - \frac{1}{\beta^2},$$

pero

$$\mathsf{Prob}\left(\frac{|S/n - \lambda|}{\sqrt{\lambda(1-\lambda)/n}} < \beta\right) = \mathsf{Prob}\left((S/n - \lambda)^2 < \beta^2 \lambda(1-\lambda)/n\right).$$

Entonces tenemos que

$$\text{Prob}\left((S/n-\lambda)^2 < \beta^2\lambda(1-\lambda)/n\right) \geq 1 - \frac{1}{\beta^2}.$$

Tomamos un valor de $\beta = \delta^{-1/2}$, y reordenando en potencias de $\lambda$ queda una expresión cuadrática que puede factorizarse como un producto de dos monomios, que coinciden con los valores de $\omega_1(S, n, \delta^{-1/2})$ y $\omega_2(S, n, \delta^{-1/2})$, por lo que

$$\text{Prob}\left(\left(\omega_1(S,n,\delta^{-1/2}) - \lambda\right)\left(\omega_2(S,n,\delta^{-1/2}) - \lambda\right) < 0\right) \geq 1 - \frac{1}{\beta^2}.$$

Teniendo en cuenta que el evento en el cual el producto es negativo es el evento en el cual $\lambda$ pertenece al intervalo $(\omega_1(S,n,\delta^{-1/2}), \omega_2(S,n,\delta^{-1/2}))$, hemos encontrado entonces un valor $\beta = \beta(\delta) = \delta^{-1/2}$ que verifica el teorema.

De este teorema, vemos que el intervalo de confianza $(\omega_1(S, n, \delta^{-1/2}), \omega_2(S, n, \delta^{-1/2}))$ efectivamente garantiza un nivel de cobertura de al menos $1 - \delta$.

Sin embargo, este intervalo puede ser demasiado ancho, es decir dar una estimación del error cometido que es demasiado pesimista, para una cantidad de replicaciones $n$ dada.

En particular, si en las ecuaciones precedentes tomáramos $\beta = \phi^{-1}(1 - \delta/2)$ el inverso de la distribución normal evaluado en $1 - \delta/2$, tenemos un intervalo de confianza alternativo, cuyo coeficiente de confianza converge a $1 - \delta$ cuando $n$ tiende a infinito. Como el cociente asintótico de los anchos de los intervalos de Chebyshev y normales tiende a $\delta^{-1/2}\phi^{-1}(1 - \delta/2)$, esto muestra que el intervalo de Chebyshev es conservativo en esa situación.

Sin embargo, el intervalo normal basado en $\beta = \phi^{-1}(1 - \delta/2)$ no garantiza el nivel de confianza $1 - \delta$ para $n$ finito, sino que hay un error de aproximación que puede ser mayor o peor dependiendo no sólo de $n$, sino también de $\lambda$ y $S$.

Una discusión detallada de estos problemas y de intervalos de confianza alternativos se encuentra en el artículo “Interval estimation for a binomial proportion”, de Brown, L.D., Cai, T., DasGupta, A., publicado en Statistical Science 16, 101-133 (2001).

Dicho material es considerado parte de esta sesión como lectura adicional obligatoria (en particular, allí se define el intervalo de Agresti-Coull, pedido en los ejercicios) y está accesible en `http://web.archive.org/web/20170222133234/http://www-stat.wharton.upenn.edu/~tcai/paper/Binomial-StatSci.pdf` (último acceso: 2020-03-31).

## Material adicional de referencia

- Material adicional sobre intervalos de confianza, en un contexto más general: Capítulo 9 de las notas de curso "A Course in Statistical Inference" por David Olive (Preprint M-02-006), disponible en `http://lagrange.math.siu.edu/Olive/infbook.htm` (último acceso: 2020-03-31).

## Preguntas para auto-estudio

- ¿Cuál es la distribución de la v.a. $S$, suma de $n$ v.a. de Bernoulli independientes de parámetro $\lambda$? ¿Cómo se puede calcular un intervalo de confianza basado en dicha distribución? ¿Qué dificultades hacen que estos intervalos no sean muy utilizados en la práctica?

- ¿Cuál es la forma del intervalo de confianza de Wald? ¿Qué problemas pueden encontrarse al utilizar este intervalo? ¿Qué intervalos alternativos pueden emplearse? ¿Qué ventajas tiene el intervalo de Agresti-Coull frente al intervalo de Wald?

- ¿Cuál es la forma del intervalo de confianza de Chebyshev?

# Ejercicio

Entrega 3: Ejercicio 5.1: [tarea en grupo]

Problema (enunciado en sesión 3, ejercicio 3.1): se desea estimar el volumen de una región $\mathcal{R}$ de $[0,1]^6$ definida por todos los puntos de la hiper-esfera de centro $(0.45, 0.5, 0.6, 0.6, 0.5, 0.45)$ y radio $0.35$, que además cumplan las restricciones siguientes: $3x_1 + 7x_4 \leq 5$; $x_3 + x_4 \leq 1$; $x_1 - x_2 - x_5 + x_6 \geq 0$.

- Parte a: Compartir en el grupo los códigos desarrollados para la parte a, validarlos revisando los códigos, y verificando si las salidas para tamaños de muestra de $10^6$ son consistentes. Indicar si se detectaron errores en los mismos, y en ese caso dar los códigos corregidos. Elegir uno de los códigos para las partes siguientes, explicar los motivos de la selección.

- Parte b: calcular la cantidad de replicaciones a realizar para garantizar un error menor a $1.0 \times 10^{-4}$ con probabilidad $0.95$, utilizando el criterio de peor caso de Hoeffding.

- Parte c: utilizando el código elegido en la parte a, y la cantidad de replicaciones definida en el punto anterior, calcular el intervalo de confianza de nivel $0.95$ utilizando el criterio de Chebyshev, y el criterio de Agresti-Coull. Comparar el ancho de estos intervalos entre sí y con el criterio de error manejado en el punto previo.

Fecha entrega: ver calendario.