

EST-24107: Simulación

Profesor: Alfredo Garbuno Iñigo — Primavera, 2022 — Integración Monte Carlo.

Objetivo. Estudiar integración numérica en un contexto probabilístico. Estudiar, en particular, el método Monte Carlo y entender sus bondades y limitaciones como un método de aproximación de integrales.

Lectura recomendada: Una lectura mas técnica sobre reglas de cuadratura se puede encontrar en la sección 3.1 de Reich and Cotter [1]. Y una buena referencia (técnica) sobre el método Monte Carlo lo encuentran en Sanz-Alonso et al. [3]. En el curso estamos explorando mas allá de lo que ofrece el capítulo 3 de Robert and Casella [2].

1. INTRODUCCIÓN

En muchas aplicaciones nos interesa poder resolver integrales de manera numérica. Estas pueden ser de cualquier forma. Por ejemplo, nos puede interesar resolver

$$\int_{\Theta} h(\theta) d\theta, \quad (1)$$

que bien puede ser reexpresada como una integral bajo una medida de probabilidad.

Por ejemplo,

$$\int_{\Theta} f(\theta) \pi(\theta) d\theta, \quad (2)$$

de tal forma que podemos pensar en la ecuación de arriba como un valor esperado de una variable $\theta \sim \pi(\cdot)$ y calcular mediante un método numérico.

- La pregunta clave (I) es: ¿qué distribución podemos utilizar?
- La pregunta clave (II) es: ¿con qué método numérico resuelvo la integral?
- La pregunta clave (III) es: ¿y si no hay método numérico?

2. INTEGRACIÓN NUMÉRICA

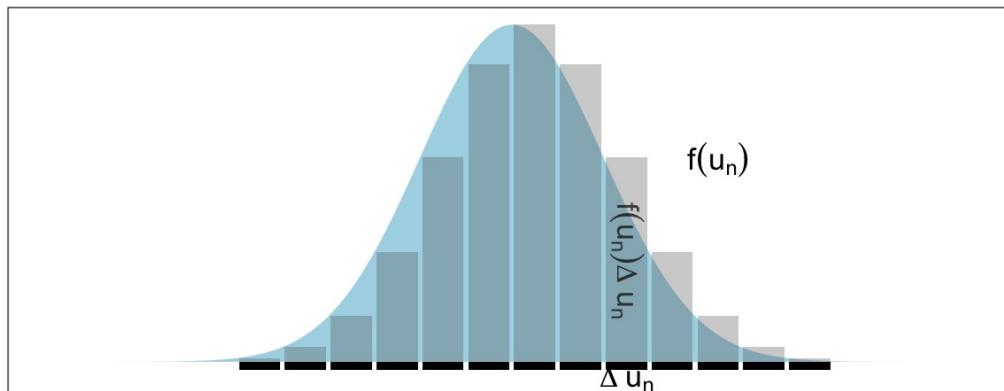
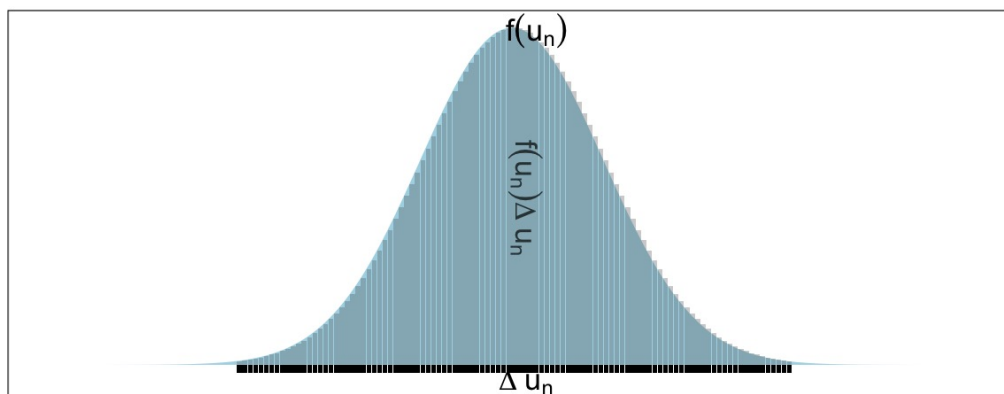
Recordemos la definición de integrales Reimann:

$$\int f(x) dx \approx \sum_{n=1}^N f(u_n) \Delta u_n =: \hat{\pi}_N^R(f).$$

La aproximación utilizando una malla (cuadrícula) de N puntos sería:

$$\sum_{n=1}^N f(u_n) \Delta u_n.$$

El método es útil cuando las integrales se realizan cuando tenemos pocos parámetros. Es decir, cuando el dominio de integración es $\mathcal{X} \subseteq \mathbb{R}^p$ con p pequeña.

FIGURA 1. Integral por medio de discretización con $N = 11$.FIGURA 2. Integral por medio de una malla fina, $N = 101$.

2.1. Análisis de error

El concepto de **integrabilidad de Darboux** nos puede ayudar a acotar el error cometido por nuestra estrategia de integración. Por ejemplo, para una partición ρ_N (malla) del intervalo tenemos que

$$L_{f,\rho_N} \leq \hat{\pi}_N^R(f) \leq U_{f,\rho_N} . \quad (3)$$

Lo que recordarán de sus cursos de cálculo es que

$$\lim_{N \rightarrow \infty} |U_{f,\rho_N} - L_{f,\rho_N}| = 0 , \quad (4)$$

y que además se satisface

$$\int f(x)dx = \lim_{N \rightarrow \infty} U_{f,\rho_N} = \lim_{N \rightarrow \infty} L_{f,\rho_N} . \quad (5)$$

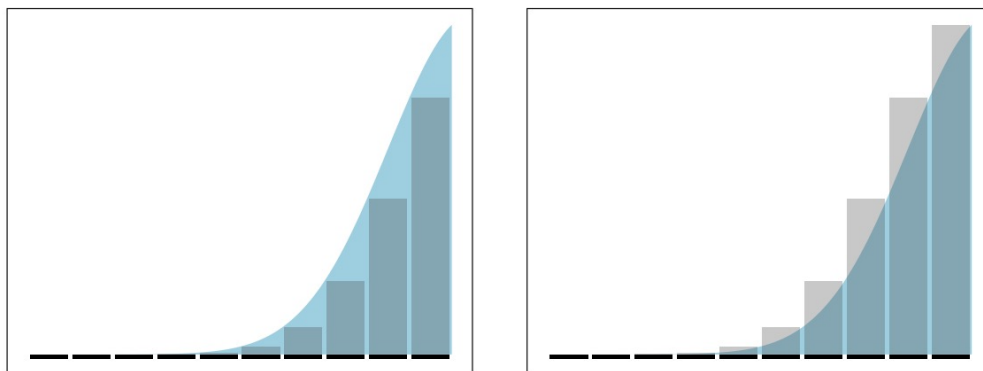


FIGURA 3. Integrales y cotas de Darboux.

2.2. Más de un parámetro

Consideramos ahora un espacio con $\theta \in \mathbb{R}^p$. Si conservamos N puntos por cada dimensión, ¿cuántos puntos en la malla necesitaríamos? Lo que tenemos son recursos computacionales limitados y hay que buscar hacer el mejor uso de ellos. En el ejemplo, hay zonas donde no habrá contribución en la integral.

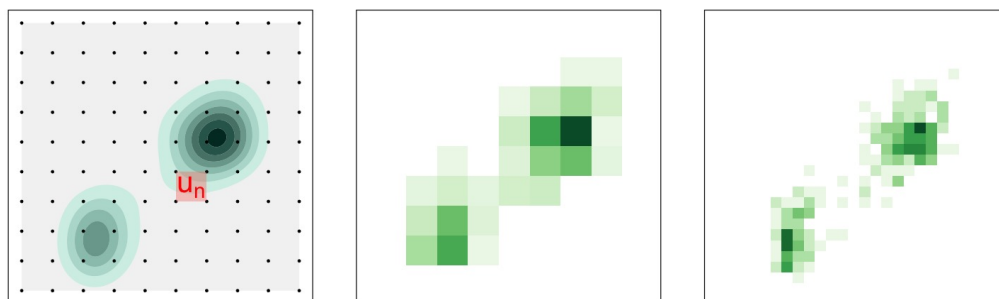


FIGURA 4. Integral multivariada por método de malla.

2.3. Reglas de cuadratura

Por el momento hemos escogido aproximar las integrales por medio de una aproximación con una malla uniforme. Sin embargo, se pueden utilizar aproximaciones

$$\int f(x)dx \approx \sum_{n=1}^N f(\xi_n) \omega_n.$$

Estas aproximaciones usualmente se realizan para integrales en intervalos cerrados $[a, b]$. La regla de cuadratura determina los pesos ω_n y los centros ξ_n pues se escogen de acuerdo a ciertos criterios de convergencia.

Por ejemplo, se consideran polinomios que aproximen con cierto grado de precisión el integrando. Los pesos y los centros se escogen de acuerdo a la familia de polinomios. Pues para cada familia se tienen identificadas las mallas que optimizan la aproximación. Ver sección 3.1 de Reich and Cotter [1].

3. INTEGRACIÓN MONTE CARLO

$$\pi(f) = \mathbb{E}_\pi[f] = \int f(x)\pi(x)dx,$$

$$\pi_N^{\text{MC}}(f) = \frac{1}{N} \sum_{n=1}^N f(x^{(n)}), \quad \text{donde } x^{(n)} \stackrel{\text{iid}}{\sim} \pi, \quad \text{con } n = 1, \dots, N,$$

$$\pi_N^{\text{MC}}(f) \approx \pi(f).$$

3.1. Ejemplo: Dardos

Consideremos el experimento de lanzar dardos uniformemente en un cuadrado de tamaño 2, el cual contiene un círculo de radio 1.

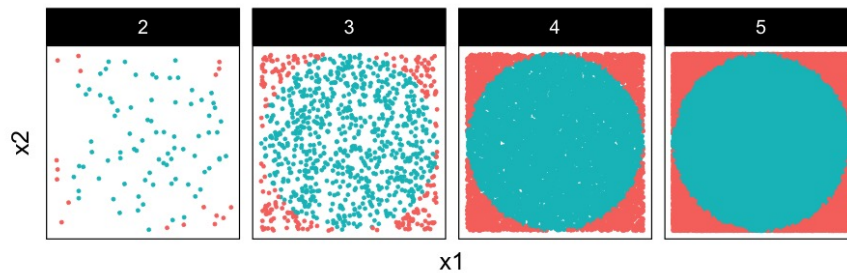


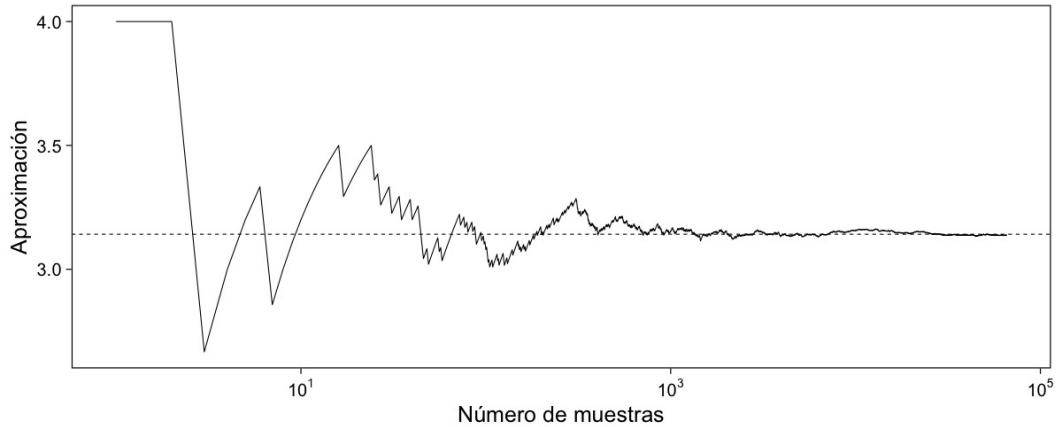
FIGURA 5. Integración Monte Carlo para aproximar π .

Si escogemos N suficientemente grande entonces nuestro promedio converge a la integral. En Fig. 6 se muestra para cada n en el eje horizontal cómo cambia nuestra estimación $\hat{\pi}_n^{\text{MC}}(f)$.

También podemos replicar el experimento unas M veces y observar cómo cambiaría nuestra estimación con distintas semillas. Por ejemplo, podemos replicar el experimento 10 veces. En R y python lo usual es utilizar arreglos multidimensionales para poder guardar muestras bajo distintas replicas.

```
1 set.seed(108)
2 nsamples <- 10**4; nexp <- 100
3 U <- runif(nexp * 2 * nsamples)
4 U <- array(U, dim = c(nexp, 2, nsamples))
5 apply(U[1:5,,], 1, str)
```

```
1 num [1:2, 1:10000] 0.455 0.164 0.529 0.415 0.474 ...
2 num [1:2, 1:10000] 0.404 0.9282 0.0883 0.3126 0.4386 ...
3 num [1:2, 1:10000] 0.351 0.449 0.369 0.814 0.695 ...
4 num [1:2, 1:10000] 0.664 0.627 0.185 0.882 0.113 ...
5 num [1:2, 1:10000] 0.4635 0.0115 0.0117 0.5086 0.9384 ...
6 NULL
```

FIGURA 6. Estimación $\pi_N^{MC}(f)$ con $N \rightarrow \infty$.

```

1 resultados ← apply(U, 1, function(x){
2   dardos ← apply(x**2, 2, sum)
3   exitos ← ifelse(dardos ≤ 1, 1, 0)
4   prop ← cummean(exitos)
5   4 * prop
6 })

```

Lo cual nos permite realizar distintos escenarios posibles.

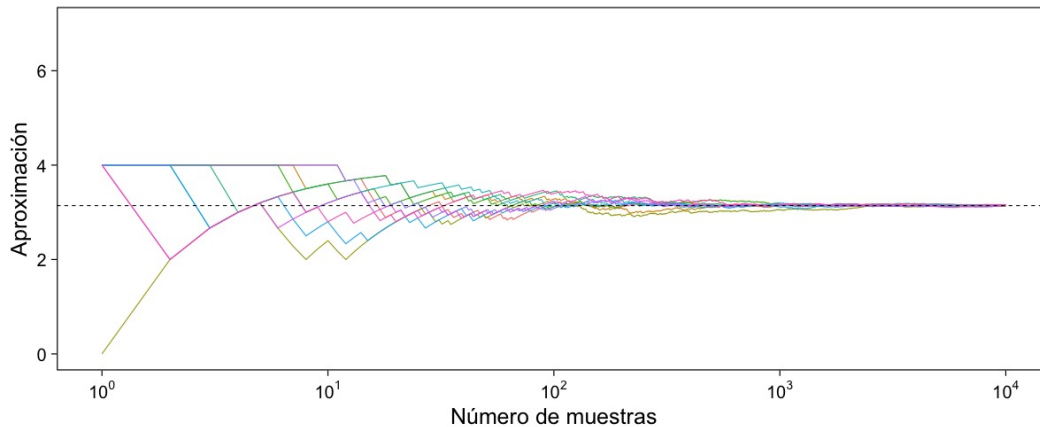


FIGURA 7. Réplica de las trayectorias de diversas realizaciones de la aproximación de la integral.

Bajo ciertas consideraciones teóricas podemos esperar un buen comportamiento de nuestro estimador de la integral. E incluso podríamos (si el número de simulaciones lo permite) aproximar dicho comportamiento utilizando distribuciones asintóticas, (TLC).

Podemos explicar la reducción de los intervalos de confianza por medio de la varianza de la estimación de la integral en las distintas réplicas que tenemos. Mas adelante explicaremos de dónde viene la línea punteada. Vemos cómo, aunque captura bien la reducción en varianza, puede sub- o sobre-estimarla.

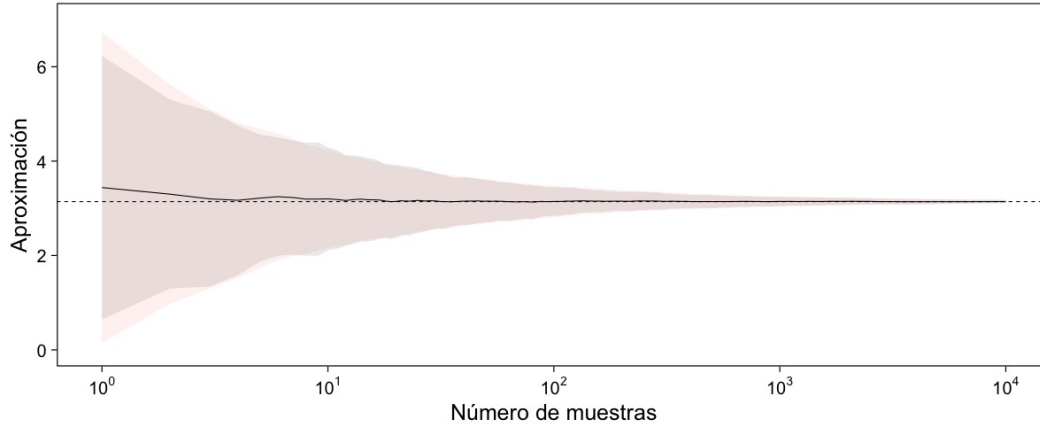


FIGURA 8. Comportamiento promedio e intervalos de confianza.

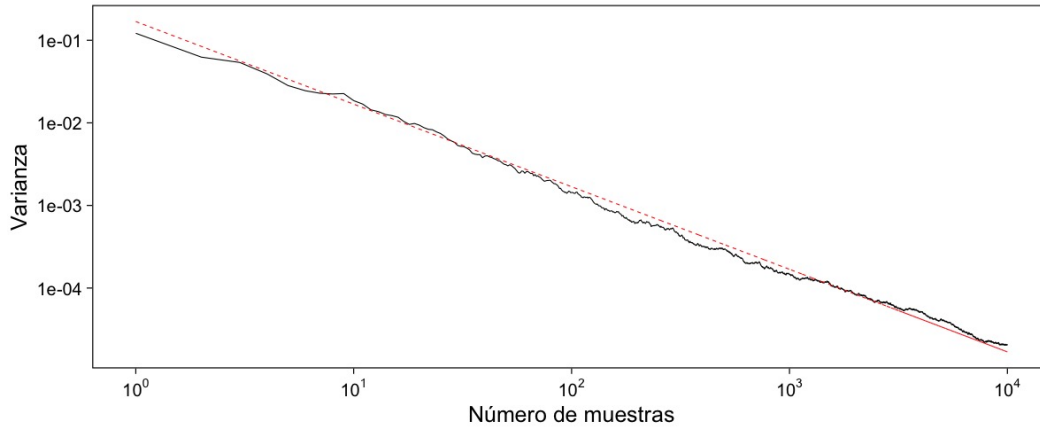


FIGURA 9. Comportamiento promedio e intervalos de confianza.

3.2. Propiedades

A continuación enunciaremos algunas propiedades clave del método Monte Carlo. Poco a poco las iremos explicando y en particular discutiremos mas a fondo algunas de ellas.

3.2.1. Teorema [Error Monte Carlo] Sea $f : \mathbb{R}^p \rightarrow \mathbb{R}$ cualquier función bien comportada[†]. Entonces, el estimador Monte Carlo es **insesgado**. Es decir, se satisface

$$\mathbb{E} \left[\hat{\pi}_N^{\text{MC}}(f) - \pi(f) \right] = 0, \quad (6)$$

para cualquier N . Usualmente estudiamos el error en un escenario pesimista donde medimos el **error cuadrático medio** en el peor escenario

$$\sup_{f \in \mathcal{F}} \mathbb{E} \left[\left(\hat{\pi}_N^{\text{MC}}(f) - \pi(f) \right)^2 \right] \leq \frac{1}{N}.$$

Esta desigualdad nos muestra una de las propiedades que usualmente se celebran de los métodos Monte Carlo. La integral y nuestra aproximación de ella por medio de simulaciones tiene un error acotado proporcionalmente por el número de simulaciones.

En particular, la varianza del estimador (**error estándar**) satisface la igualdad

$$\text{ee}^2 \left(\hat{\pi}_N^{\text{MC}}(f) \right) = \frac{\mathbb{V}_\pi(f)}{N}.$$

Esta igualdad, aunque consistente con nuestra desigualdad anterior, nos dice algo más. El error de nuestra aproximación **depende** de la varianza de f bajo la distribución π .

3.2.2. Teorema [TLC para estimadores Monte Carlo] Sea f una función **bien comportada** ^{††}, entonces bajo una N suficientemente grande tenemos

$$\sqrt{N} \left(\hat{\pi}_N^{\text{MC}}(f) - \pi(f) \right) \sim \mathbf{N}(0, \mathbb{V}_\pi(f)). \quad (7)$$

3.2.3. Nota: El estimador Monte Carlo del que hablamos, $\hat{\pi}_N^{\text{MC}}(f)$, es una estimación con una **muestra finita de simulaciones**. En ese sentido podemos pensar que tenemos un *mapeo* de muestras a estimador

$$(x^{(1)}, \dots, x^{(N)}) \mapsto \hat{\pi}_N^{\text{MC}}(f), \quad (8)$$

con $x^{(i)} \stackrel{\text{iid}}{\sim} \pi$.

De lo cual es natural pensar: ¿y si hubiéramos observado otro conjunto de simulaciones? Nuestro proceso de estimación es el mismo pero la muestra puede cambiar.

En este sentido nos preguntamos por el **comportamiento promedio** bajo distintas muestras observadas

$$\mathbb{E}[\hat{\pi}_N^{\text{MC}}(f)] = \mathbb{E}_{x_1, \dots, x_N}[\hat{\pi}_N^{\text{MC}}(f)]. \quad (9)$$

De la misma manera nos podemos preguntar sobre la **dispersión alrededor de dicho promedio** (varianza)

$$\mathbb{V}[\hat{\pi}_N^{\text{MC}}(f)] = \mathbb{V}_{x_1, \dots, x_N}[\hat{\pi}_N^{\text{MC}}(f)]. \quad (10)$$

Al ser un ejercicio de **estimación** la desviación estándar del estimador recibe el nombre de **error estándar**. Lo cual denotamos por

$$\text{ee}[\hat{\pi}_N^{\text{MC}}(f)] = \left(\mathbb{V}[\hat{\pi}_N^{\text{MC}}(f)] \right)^{1/2} = \left(\frac{\mathbb{V}_\pi(f)}{N} \right)^{1/2}. \quad (11)$$

3.2.4. Nota: Para algunos estimadores la fórmula del error estándar se puede obtener de manera analítica (curso de **Inferencia Matemática**). Para otro tipo, tenemos que utilizar propiedades asintóticas (p.e. cota de Cramer-Rao).

Hay casos en los que no existe una fórmula asintótica o resultado analítico, pero podemos usar simulación [8] para cuantificar dicha dispersión (lo veremos en otra sección del curso).

3.2.5. Nota: Hay situaciones en las que la distribución normal asintótica no tiene sentido. Para este tipo de situaciones también veremos cómo podemos utilizar simulación para cuantificar dicha dispersión.

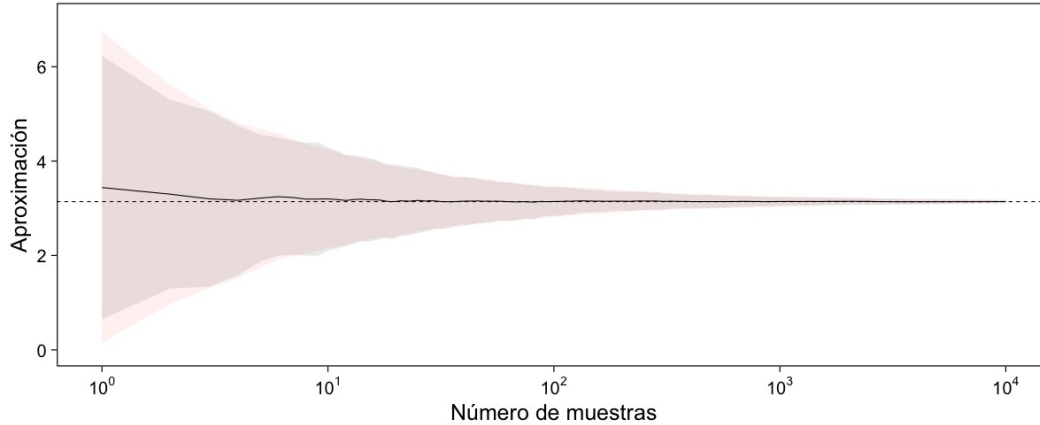


FIGURA 10. Comportamiento promedio e intervalos de confianza con aproximación asintótica.

3.3. Estimación de una proporción

El lanzamiento de dados que vimos es un ejemplo de una situación muy usual en estimación de integrales. Queremos estimar la tasa de éxito a partir de ver el éxito o fracaso de experimentos Bernoulli.

Si denotamos por θ la tasa de éxito. Entonces nuestro experimento (lanzar dados dentro del círculo) determina que $S_n \sim \text{Binomial}(N, \theta)$ y que \bar{X}_n es un **estimador** de θ . Por lo tanto,

$$\hat{\theta}_n := \bar{X}_n \approx \theta \quad (12)$$

3.3.1. *Ejercicio:* ¿Cuál es la fórmula del error estándar para este estimador?

3.3.2. *Pregunta:* ¿Cuántas muestras necesitamos para tener una *buena* aproximación?

4. DESIGUALDADES DE CONCENTRACIÓN

En muchas situaciones nos interesa establecer cuántas simulaciones necesitamos para poder aproximar nuestras integrales hasta cierto orden. Por ejemplo, la tabla en ?? muestra la aproximación conforme aumenta N .

	N	estimado	dif.abs
1			
2	1	10	1.0000 0.5354
3	2	100	0.7800 0.5904
4	3	1000	0.7910 0.5876
5	4	10000	0.7879 0.5884

LISTING 1. Aproximación de la proporción de dados dentro de la diana.

4.1. Desigualdad de Chebyshev

Lo que queremos es encontrar una N tal que con una alta probabilidad nuestro estimador sea cercano al parámetro que está ajustando. Esto lo escribimos como

$$\text{Prob}(|\hat{\theta}_N - \theta| < \epsilon) \geq 1 - \delta. \quad (13)$$

4.1.1. **Teorema [Desigualdad de Chebyshev]:** Sea X una variable aleatoria con media y varianza finita denotadas por μ y σ^2 respectivamente. Entonces para cualquier constante positiva $k \in \mathbb{R}$, tenemos que

$$\text{Prob}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}. \quad (14)$$

Lo cual podemos utilizar para encontrar una cota inferior para N .

4.1.2. **Ejercicio:** Calcula la desigualdad y obtén el número de simulaciones necesarios para encontrar un estimador con nivel de precisión ϵ con una probabilidad α .

4.1.3. **Discusión:** El resultado anterior nos permite escribir que con alta probabilidad (al menos $1 - \delta$) tendremos que

$$\hat{\theta}_N = \theta \pm \epsilon. \quad (15)$$

4.2. Desigualdad de Hoeffding

La desigualdad de Chebyshev nos permite encontrar cotas para prácticamente cualquier situación[†]. Sin embargo, el precio es la calidad de la estimación.

Una alternativa es utilizar la desigualdad de Hoeffding que nos permite establecer cotas desviaciones de variables aleatorias acotadas.

Aunque la discusión es a nivel variable aleatoria, lo que estamos discutiendo es relevante en integración Monte Carlo. Pues, si $f : X \rightarrow [a, b]$ podemos pensar en $f(X)$ como una variable aleatoria acotada en $[a, b]$ y nuestra discusión procede.

4.2.1. **Teorema [Desigualdad de Hoeffding]:** Sea X_1, \dots, X_n una muestra iid de variables aleatorias con valores en $[a, b]$. Entonces para cualquier $t \geq 0$ y usando $S_n = X_1 + \dots + X_n$ tenemos que

$$\text{Prob}(|S_n - \mathbb{E}[S_n]| \geq t) \leq 2 \exp\left(-\frac{2t^2}{n(b-a)^2}\right). \quad (16)$$

4.2.2. **Ejercicio:** Establece el tamaño de muestra necesario, N , para garantizar con probabilidad al menos $1 - \delta$ que nuestro estimador será ϵ preciso.

5. JUSTIFICACIÓN DEL MÉTODO MONTE CARLO

Lo que hemos discutido hasta ahora nos permite ver que el método Monte Carlo —aproxima integrales con promedios— tiene buenas propiedades. El tiro de gracia es el siguiente resultado.

5.0.1. **Teorema [Ley (fuerte) de los grandes números]:** Sea X_1, \dots, X_n una muestra de variables iid y sea X una variable con la misma distribución. Si utilizamos una $f : \mathbb{R} \rightarrow \mathbb{R}$ acotada, entonces $h(X_1), \dots, h(X_n)$ son variables independientes y acotadas con media finita. De tal forma que se satisface que

$$\text{Prob}\left(\lim_{N \rightarrow \infty} \hat{\pi}_N^{\text{MC}}(h) = \pi(h)\right) = 1. \quad (17)$$

5.0.2. **Teorema [Ley (débil) de los grandes números]:** De los resultados anteriores ya sabíamos que

$$\lim_{N \rightarrow \infty} \text{Prob}\left(\left|\hat{\pi}_N^{\text{MC}}(h) - \pi(h)\right| < \epsilon\right) = 1. \quad (18)$$

6. CONSIDERACIONES

Supongamos que queremos resolver la integral

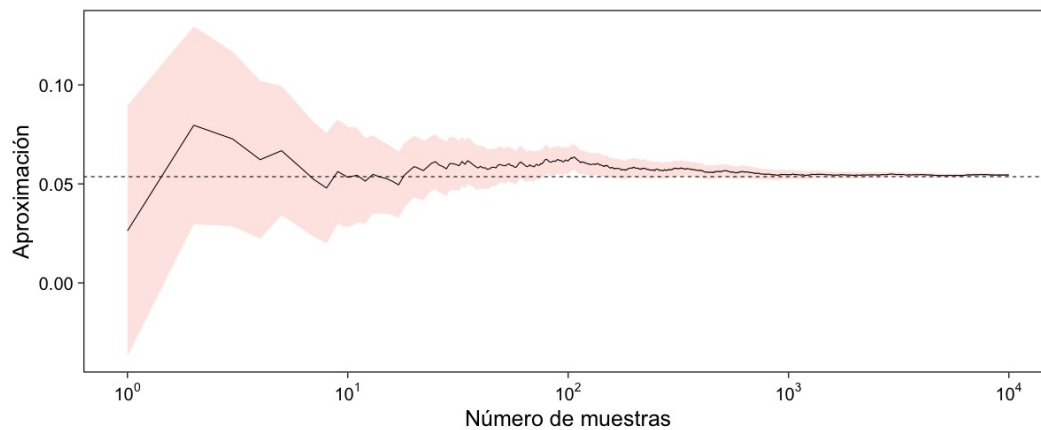
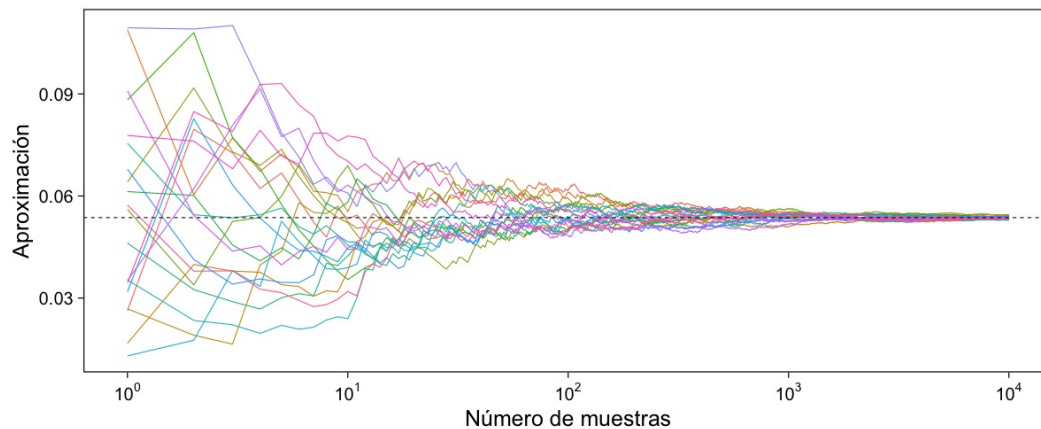
$$\int_2^3 e^{-\frac{x^2}{2}} dx. \quad (19)$$

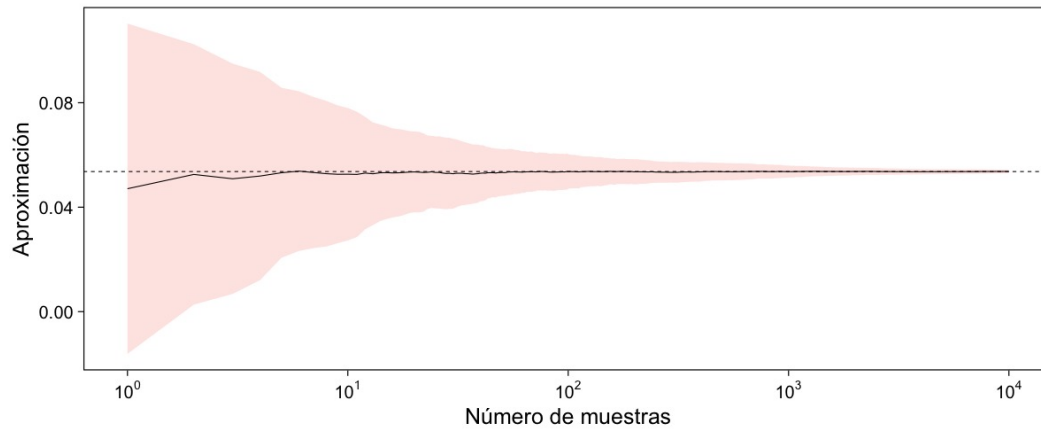
6.1. Primera estrategia

Al tener un intervalo acotado podemos pensar en una $U(2, 3)$.

```
1 set.seed(108); nsamples <- 10**4; nexp <- 100
2 h <- function(x){ exp(-x**2/2) }
3 u <- runif(nexp * nsamples, min = 2, max = 3)
4 x <- array(u, c(nexp, nsamples))
5 h_x <- h(x)
```

```
1 estimador_mc <- apply(h_x, 1, cummean)      # ojo, transpone el resultado
2 media_mc <- apply(estimador_mc, 1, mean)
3 error_mc <- apply(estimador_mc, 1, sd)
```



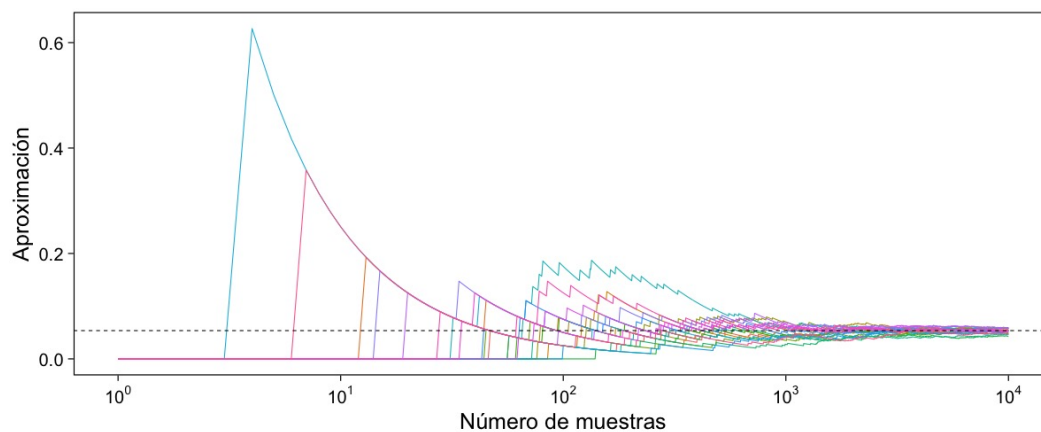


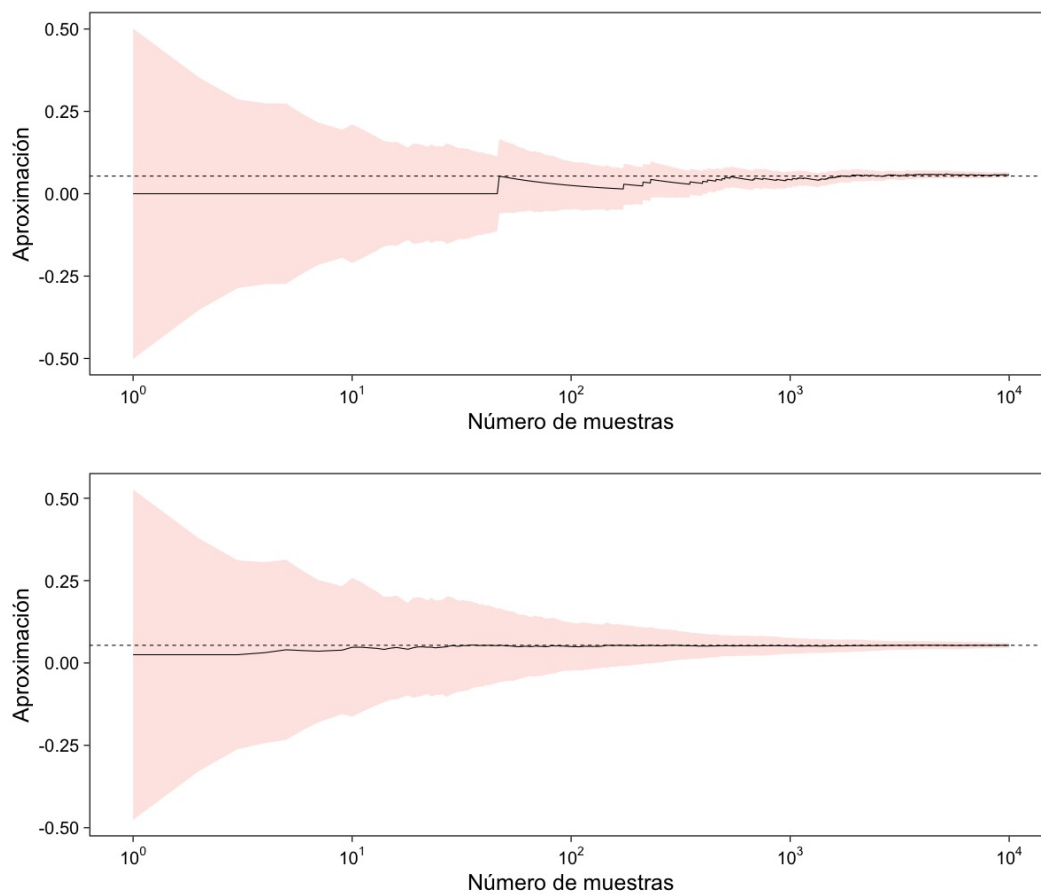
6.2. Segunda estrategia

Por la forma que tiene la integral podemos considerar también una integral bajo una distribución normal.

```
1 set.seed(108); nsamples <- 10**4; nexp <- 100
2 f <- function(x){ ifelse(x >= 2 & x <= 3, sqrt(2 * pi), 0) }
3 u <- rnorm(nexp * nsamples)
4 x <- array(u, c(nexp, nsamples))
5 f_x <- f(x)
```

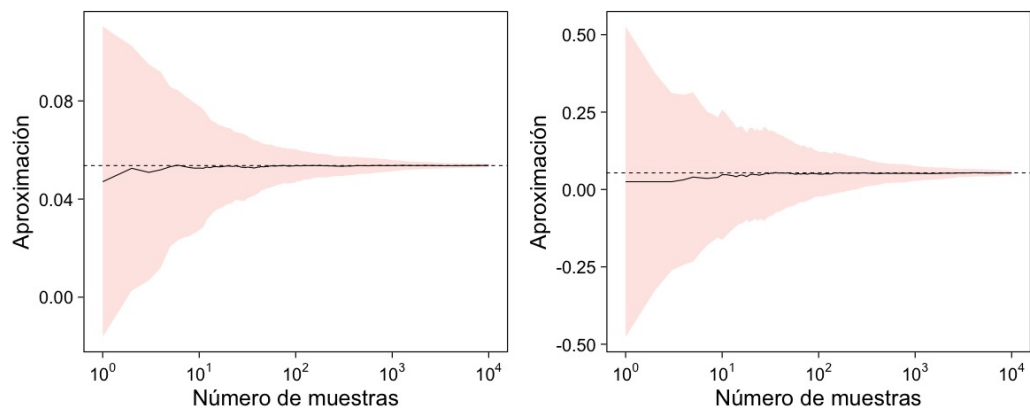
```
1 estimador_mc <- apply(f_x, 1, cummean)      # ojo, transpone el resultado
2 media_mc <- apply(estimador_mc, 1, mean)
3 error_mc <- apply(estimador_mc, 1, sd)
```

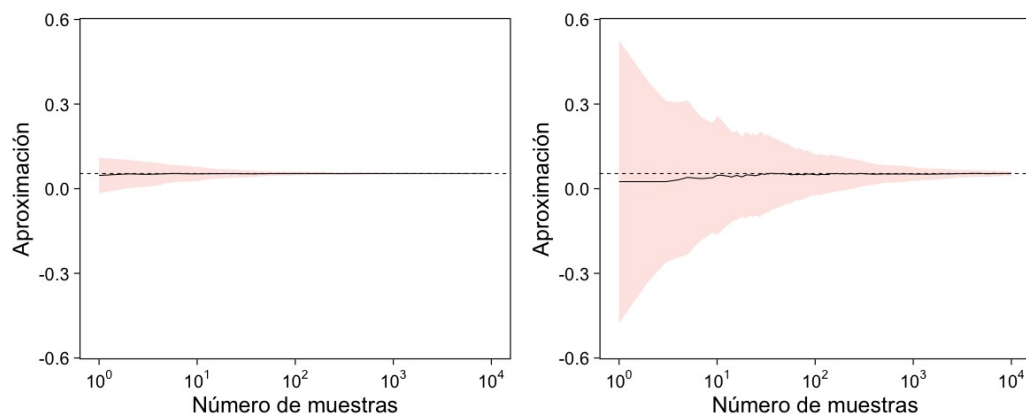




6.3. Comparación

¿Cuál método preferimos?





REFERENCIAS

- [1] S. Reich and C. Cotter. *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge University Press, Cambridge, 2015. ISBN 978-1-107-06939-8 978-1-107-66391-6. [1](#), [4](#)
- [2] C. Robert and G. Casella. *Introducing Monte Carlo Methods with R*. Springer New York, New York, NY, 2010. [1](#)
- [3] D. Sanz-Alonso, A. M. Stuart, and A. Taeb. Inverse Problems and Data Assimilation. *arXiv:1810.06191 [stat]*, jul 2019. [1](#)