

# Método Monte-Carlo

Alberto Benavides

11 de septiembre de 2017

## Hipótesis

Se pueden pronosticar sucesos reales registrados de manera estadística a partir de la implementación del método Monte-Carlo en una simulación computacional.

## Simulación y resultados

Para esta simulación se utilizó una computadora portátil con sistema operativo Windows 10 Home Single Language con un procesador Intel(R) Core(TM) i7-7500U CPU @ 2.70GHz, 2904 MHz de dos procesadores principales y cuatro procesadores lógicos. Las operaciones en paralelo que se ejecutaron en cada experimento utilizan las librerías `parallel` y `doParallel` del lenguaje R.

El primer punto de interés en esta investigación era comprobar que el uso del método Monte-Carlo proporciona resultados cercanos a los esperados en experimentos y operaciones de valores conocidos, es por ello que se eligieron dos simulaciones que cumplen con estas características. A continuación se detallará cada una de ellas.

## Predicción del valor de una integral

En esta etapa se generan números pseudoaleatorios mediante el uso de la librería `distr` de R para aproximar un resultado a la  $\int_3^7 f(x)$  para

$$f(x) = \frac{1}{e^x + e^{-x}}.$$

Estas aproximaciones se obtuvieron de la repetición del siguiente experimento:

```
g <- function(x) { return((2 / pi) * f(x)) }  
generador <- r(AbscontDistribution(d = g))  
muestra <- generador(100)
```

```

desde <- 3
hasta <- 7
cuantos <- 500
parte <- function() {
  valores <- generador(pedazo)
  return(sum(valores >= desde & valores <= hasta))
}

```

Éste se replicó en secuencias de  $i$  a  $10 \times i$  en pasos de  $i$ , donde  $i = \{10^2, 10^3, 10^4, 10^5, 10^6\}$ , de modo que la primera de estas secuencias sería como sigue:

100, 200, 300, 400, 500, 600, 700, 800, 900, 1000.

Además, se registró el tiempo que tomaron estos experimentos agrupados por las secuencias descritas. Los experimentos de cien hasta mil repeticiones, siempre agrupados en las secuencias arriba detalladas, representan tiempos despreciables y que no guardan una relación representativa. Sus gráficas se muestran en la figura 1 (p. 2).

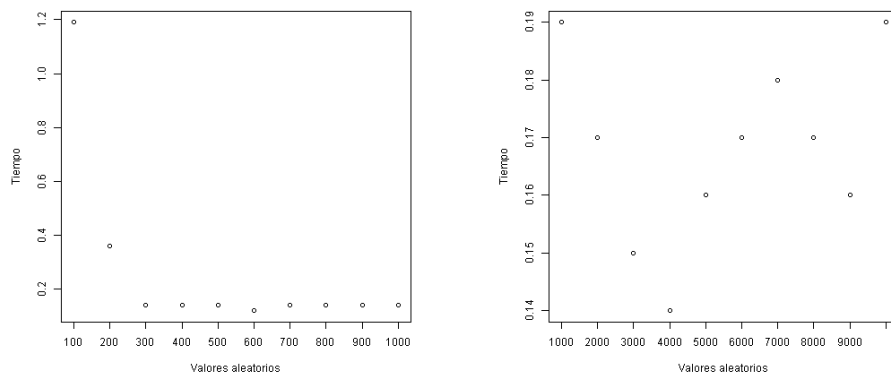


Figura 1: Tiempos que requirió el experimento en correrse con las repeticiones dadas por la secuencia de  $i = 10^2$  (izquierda) y de  $i = 10^3$  (derecha).

Más representativos fueron los tiempos del resto de las secuencias, en donde se puede apreciar el desarrollo de una función muy cercana a una función lineal para los tiempos con relación a los tamaños de repeticiones. Estas observaciones se pueden corroborar en la figura 2 (p. 3).

Ahora bien, el valor de la integral antes definida es de 0.048834, calculado en Wolfram Alpha. Para conocer la cercanía que las aproximaciones obtenidas tuvieron con este resultado, se restaba a cada una de esas aproximaciones dicho valor y al resultado se le sacaba su valor absoluto, para así tener una cifra positiva

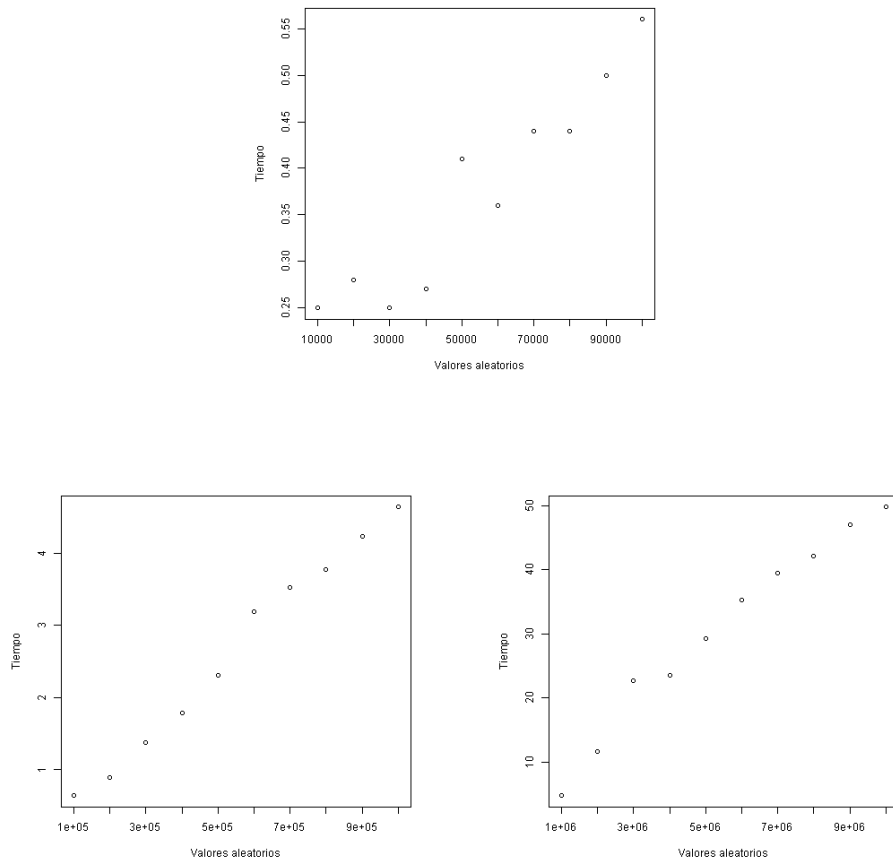


Figura 2: Tiempos que tomó en realizarse el experimento en repeticiones de  $i = 10^4$  (arriba),  $i = 10^5$  (inferior izquierda) e  $i = 10^6$  (inferior derecha).

que representara la distancia de cada aproximación con el valor calculado. De entre todas las diferencias calculadas en el experimento, se muestran en la figura 3 (p. 4) las menores de todas las secuencias.

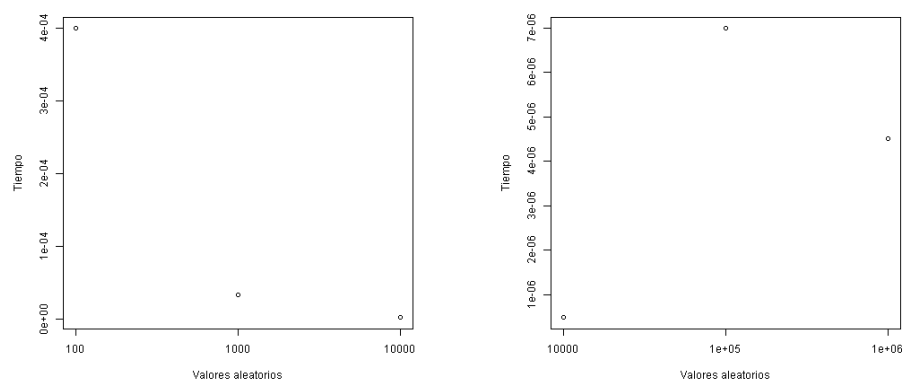


Figura 3: Diferencias mínimas de las secuencias de  $i = 10^2, 10^3, 10^4$  (izquierda) e  $i = 10^4, 10^5, 10^6$  (derecha).

Esta selección de diferencias mínimas muestra que en la secuencia de  $i = 10^4$  se registró la diferencia mínima de entre todas las obtenidas, por lo que será precisamente la gráfica de esa secuencia la que se muestre (figura 4, p. 5), por ser la de nuestro interés. Finalmente, en esta secuencia se puede observar que la diferencia con el valor esperado es despreciable.

## Aproximación al valor de $\pi$

El segundo experimento consistió en utilizar la función `runif` de R para obtener valores al azar dentro del rango  $[-0.5, 0.5]$  para  $x$  y  $y$  que a su vez representan puntos en los ejes horizontal y vertical del plano cartesiano. Estos puntos, distribuidos aleatoriamente, fueron pasados por una función de circunferencia que los distinguía entre dentro de una circunferencia o fuera de ella mediante la asignación de valores lógicos de *Verdadero* y *Falso*, que luego fueron operados de tal manera que su resultado fuera aproximado a  $\pi$ . El código de esta explicación, en R, es como sigue:

```
points <- 100000
circle <- function(){
  xs <- runif(points, min=-0.5, max=0.5)
  ys <- runif(points, min=-0.5, max=0.5)
  in.circle <- xs^2 + ys^2 <= 0.5^2
  mc.pi <- (sum(in.circle)/points)*4
```

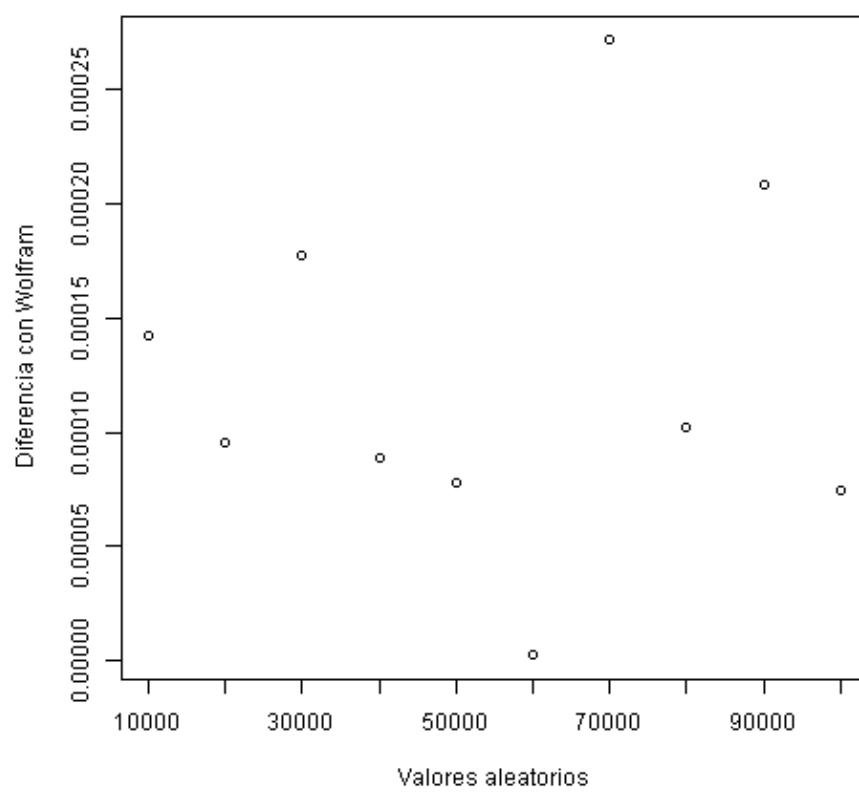


Figura 4: Diferencias de la secuencia donde  $i = 10^4$ .

```

    return (mc.pi)
}

```

Este experimento se realizó en una secuencia de  $i = 10^4$  como la descrita en el experimento anterior y se eligió esta secuencia por haber sido la que mejores resultados dio en aquel experimento. Sus tiempos, aunque mucho mayores que los requeridos en el experimento anterior, reflejan una clara función lineal respecto a la cantidad de repeticiones, misma que se muestra en la figura 5 (p. 6).

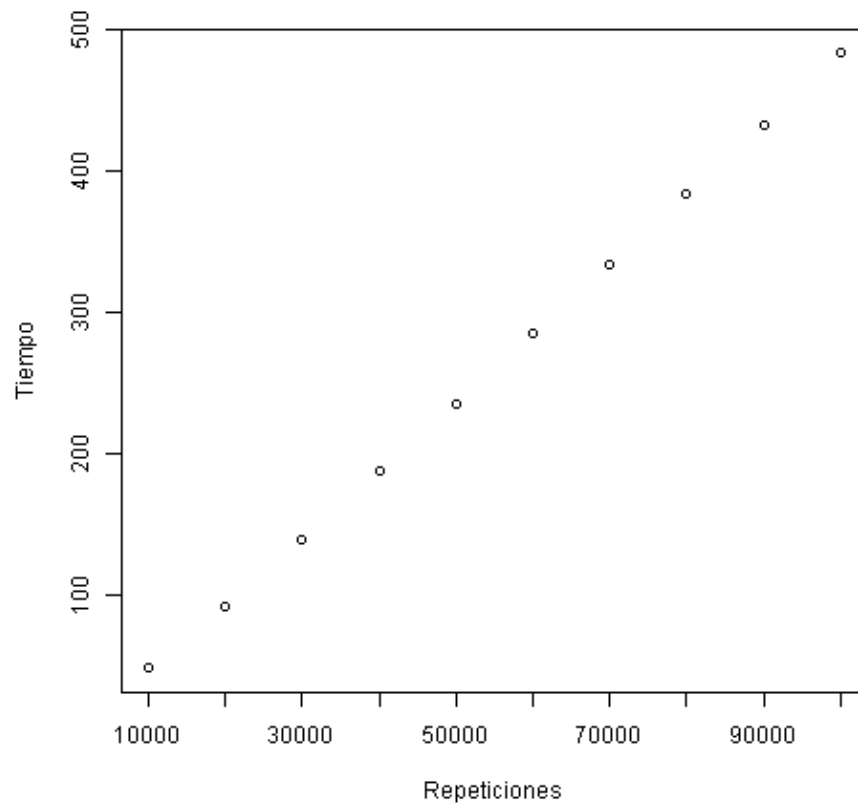


Figura 5: Tiempo en segundos que requiere el segundo experimento para la secuencia descrita con  $i = 10^4$ .

A las aproximaciones de esta secuencia para este experimento se les aplicaron las mismas operaciones de diferencia y valor absoluto, ahora respecto a  $\pi$ , y en la figura 6 (p. 7) se comprueba que en su mayoría son cercanas a cero.

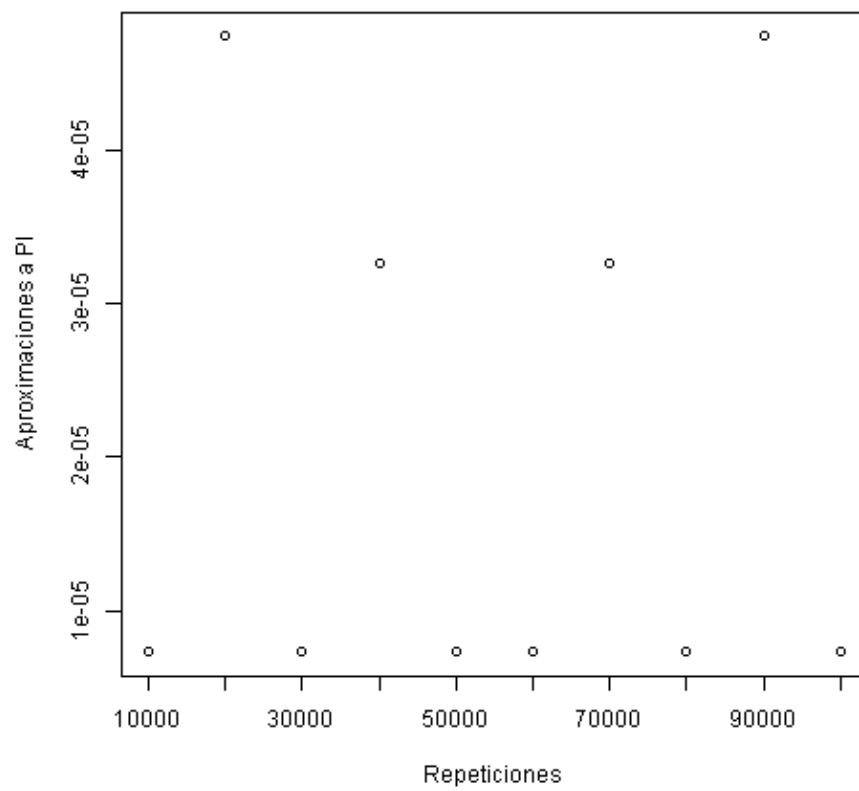


Figura 6: Diferencia respecto a  $\pi$  de las aproximaciones del segundo experimento en la secuencia donde  $i = 10^4$ .

## Pronóstico de casos de Zika a partir de datos estadísticos reales

Una vez comprobado que la precisión mediante el método Monte-Carlo es suficientemente satisfactoria, se puede proceder a su aplicación para pronosticar sucesos con base en datos conocidos, como se realizó exitosamente en los dos casos anteriores.

Ahora se analizarán los casos de Zika registrados desde la primera semana hasta la semana 34 en el estado de Oaxaca durante el año 2016<sup>1</sup>. La función elegida para aplicar el método Monte-Carlo fue `rtruncnorm` contenida en la librería `truncnorm` de R. Esta función toma como parámetros (en paréntesis los usados en este experimento) una cantidad de datos a obtener (68), un valor mínimo (0), uno máximo (infinito) y una media y una desviación estándar (obtenidas estas últimas de los datos estadísticos reales). Con base en estos parámetros, la función genera la cantidad especificada de muestras aleatorias. Este experimento se repitió 500000 veces, a cada una de estas repeticiones se les calculó la media y la mediana y se seleccionaron dos muestras, particularmente las que obtuvieron una diferencia de cero respecto a la media y mediana de los datos base para este experimento. Ambas se graficaron junto a los datos observados en la figura 7 (p. 8). Esta comparación hace patente que si bien no coinciden exactamente los casos, sí hay puntos de incidencia y encuentro a lo largo de las semanas observadas.

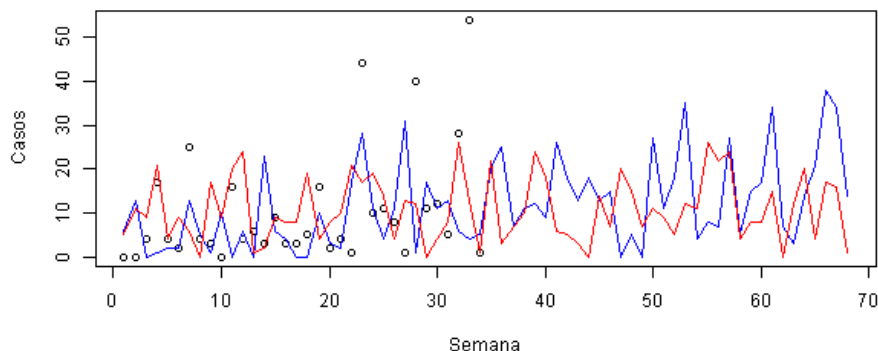


Figura 7: Comparación entre los datos observados (puntos) y los obtenidos por igualdad de media (rojos) y mediana (azules) respecto a la media y mediana de los datos observados.

<sup>1</sup><http://www.gob.mx/salud/acciones-y-programas/boletinepidemiologico-sistema-nacional-de-vigilancia-epidemiologica-sistema-unico-de-informacion-90794>



## Conclusiones

1. El método Monte-Carlo en un número de repeticiones superior a las 10000 provee aproximaciones muy precisas respecto a los valores esperados.
2. Los tiempos de ejecución de experimentos con repeticiones en paralelo incrementan de manera lineal al superar las 1000 repeticiones.
3. Con el método utilizado no se logra una identidad de muestras respecto a las registradas, pero estadísticamente se puede hablar de un pronóstico por partir de la media y desviación estándar de los datos registrados y seleccionar los resultados que guardan una relación de igualdad en las medias y medianas respecto a los datos registrados.
4. Falta profundizar, en futuras investigaciones, en otro método que se aproxime más precisamente a los datos registrados para poder hablar propiamente de un pronóstico.