

# Análisis Estadístico con R

*true*

*1 de agosto de 2018*

## Contents

<b>Simulación Monte Carlo</b>	<b>1</b>
Aplicaciones: estadística aplicada . . . . .	1
En Finanzas . . . . .	2
<b>Referencias</b>	<b>4</b>

## Simulación Monte Carlo

Los métodos de Monte Carlo (o los experimentos de Monte Carlo) son una amplia clase de algoritmos computacionales que se basan en el muestreo aleatorio repetido para obtener resultados numéricos.

Su idea esencial es usar la aleatoriedad para resolver problemas que podrían ser determinísticos en principio.

A menudo se utilizan en problemas físicos y matemáticos y son más útiles cuando es difícil o imposible utilizar otros enfoques.

Los métodos de Monte Carlo se usan principalmente en tres clases de problemas:

- optimización,
- integración numérica y
- generación de realizaciones a partir de una distribución de probabilidad.

El método tiene diferentes variantes, pero su estructura general es:

1. Definir un dominio de posibles entradas
2. Generar entradas al azar de una distribución de probabilidad sobre el dominio
3. Realice un cálculo determinista en las entradas
4. Agregue los resultados

### Ejemplo

Por ejemplo, considere un cuadrante inscrito en un cuadrado unitario. Dado que la relación de sus áreas es  $\pi/4$ , el valor de  $\pi$  se puede aproximar usando un método de Monte Carlo:

1. Dibuja un cuadrado, luego inscribe un cuadrante dentro de él
2. Distribuir uniformemente un número determinado de puntos sobre el cuadrado
3. Cuente el número de puntos dentro del cuadrante, es decir, teniendo una distancia desde el origen menor que 1
4. La relación del recuento interno y el recuento total de la muestra es una estimación de la relación de las dos áreas,  $\pi/4$ . Multiplica el resultado por 4 para estimar  $\pi$

### Aplicaciones: estadística aplicada

1. **Comparar estadísticos para muestras pequeñas en condiciones de datos reales.** Aunque pueden calcularse las propiedades de estadística y error de tipo I para datos extraídos de distribuciones teóricas clásicas (p. Ej., Curva normal, distribución de Cauchy) para condiciones asintóticas (es decir,

tamaño de muestra infinito y efecto de tratamiento infinitesimalmente pequeño), los datos reales a menudo sí no tener tales distribuciones.

2. **Para proporcionar implementaciones de pruebas de hipótesis que son más eficientes que las pruebas exactas**, como las pruebas de permutación (que a menudo son imposibles de calcular) mientras que son más precisas que los valores críticos para las distribuciones asintóticas.
3. **Proporcionar una muestra aleatoria de la distribución posterior en la inferencia bayesiana.** Esta muestra luego aproxima y resume todas las características esenciales de la distribución posterior.

### Ejemplo

¿Cuál es la desviación estándar de la probabilidad de que la probabilidad de la suma de dos dados sea  $\geq 10$ ?

1. Simular la salida de dos dados.
2. para un  $n$  dado, calcular el porcentaje que cumple la condición.
3. repetir el proceso  $N$  veces y así obtener la probabilidad.

```
t.dice <- function()
{
  sample(1:6,size = 2,replace = TRUE)
}
```

### En Finanzas

Veamos la siguiente referencia (Seynaeve and Verbeke (2017))<sup>1</sup>

### Ejemplo

Tomado de Seila (1995): Simulación Monte Carlo para Problemas Actuariales.

La simulación de Monte Carlo es una colección de técnicas para extraer información de un modelo estocástico. Tenemos un modelo estocástico y queremos usar ese modelo para que nos diga algo sobre el sistema que representa.

La simulación proporciona algunas técnicas, basadas en el muestreo del comportamiento del modelo, para proporcionar esta información.

El análisis matemático es otro método para extraer información de un modelo estocástico.

Entonces, la simulación y el análisis matemático funcionan para lograr los mismos objetivos, pero sus enfoques difieren. Vamos a ilustrar esta idea usando un ejemplo.

Considere un modelo de reclamos agregado simple, donde  $Y$  es el total de los reclamos,  $N$  es el número de reclamos recibidos en un mes y  $X_1, X_2$ , (y así sucesivamente) son los montos de los reclamos individuales. Además, supongamos que  $N$  tiene una distribución de Poisson con una media de 10.0. Entonces, la cantidad media de reclamos por mes es diez y la varianza también es diez

$$Y = \sum_{i=1}^N X_i$$

Además, suponga que las cantidades de reclamo tienen una distribución lognormal con parámetros  $\mu = 0.7$  y  $\sigma^2 = 1.80$ . Los montos de los reclamos se expresan en unidades de miles de dólares. Luego, la media, la varianza y la desviación estándar de los montos de los reclamos individuales son 4.95, 123.88 y 11.13, respectivamente. El siguiente ejemplo muestra cómo se calcularon estos.

$$E(X) = e^{\mu + \frac{\sigma^2}{2}} = 4.95$$

$$Var(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1) = 123.88$$

$$\sigma(X) = \sqrt{VAR(X)} = 11.13$$

Suponga que queremos la siguiente información del modelo:

- Media del total de reclamos ( $E(Y)$ )
- Varianza del total de reclamos ( $Var(Y)$ )
- Percentil 90 del total de reclamos ( $\phi_{0.9}(Y)$ )

### Enfoque matemático:

Usa teoría de probabilidad para calcular los parámetros

$$E(Y) = E(N)E(X) = (10)(4.95) = 49.53$$

$$Var(Y) = Var(N)[E(X)]^2 + E(N)Var(n) = (10)(4.95)^2 + (10)(123.88) = 1484.12$$

$$\sigma(Y) = \sqrt{1484.12} = 38.52$$

Pero...

$$\phi_{0.9}(Y) = ?$$

### Enfoque Simulación

El enfoque de simulación muestrea las observaciones de  $Y$  y calcula las estimaciones de los parámetros que queremos conocer utilizando estas observaciones.

1. Muestree  $N$  muestras del número de reclamos mensuales, de una distribución de Poisson con media 10.
2. Muestree cada uno de los  $N$  montos reclamados de una distribución lognormal con parámetros  $\mu = 0.7$  y  $\sigma^2 = 1.80$ .
3. Agregar el número de reclamos para tener el total del mes.
4. Repetir el proceso 2000 veces.

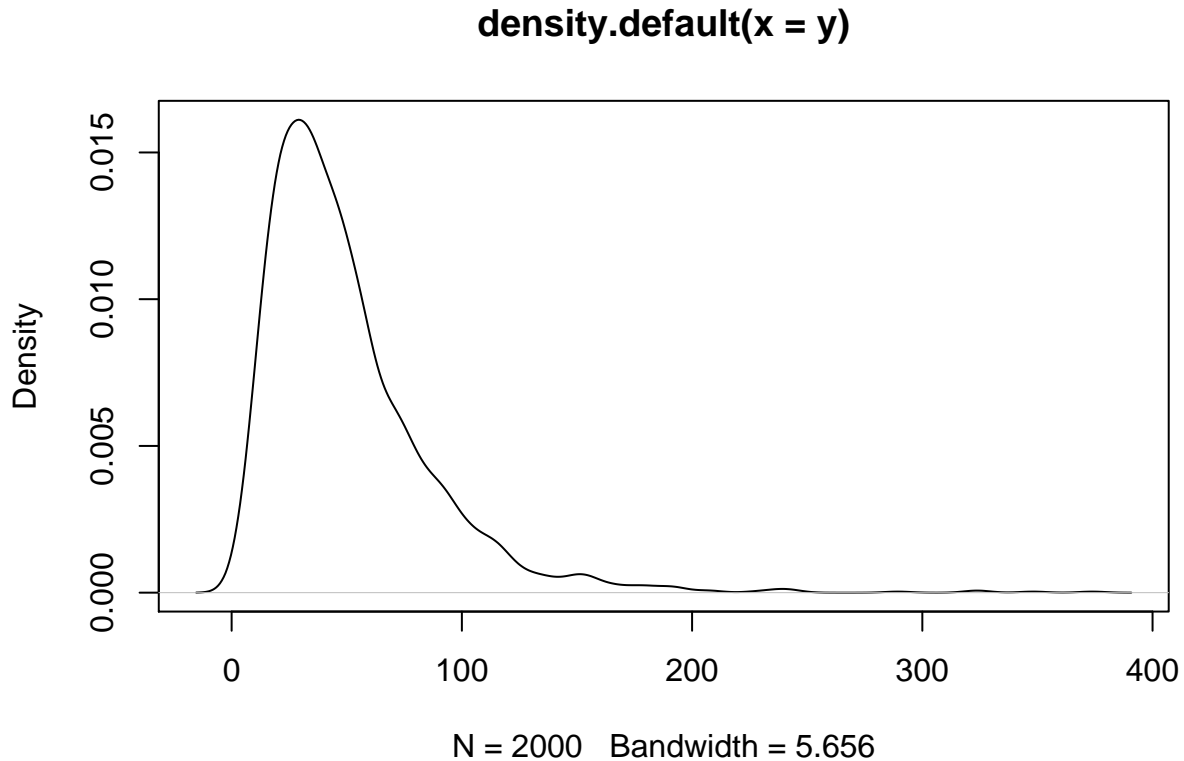
```
#### Parámetros:
N <- 10 # número de reclamos
mu.c <- 10 # media del numero de reclamos
mu.l <- 0.7 # media de la cantidad de reclamos
var.l <- 1.8 # varianza de la cantidad de reclamos

# Para un mes:
set.seed(1)
# numero de reclamos en el mes:
n.claims <- rpois(n = 1, lambda = mu.c)
# monto del numero de reclamos:
m.claims <- rlnorm(n.claims, meanlog = mu.l, sdlog = sqrt(var.l))
# Suma del monto
sum(m.claims)

## [1] 37.81192
```

Simulemos el proceso anterior 5000 veces, deberíamos obtener valores similares a estos:

```
plot(density(y))
```



```
mean(y)
```

```
## [1] 50.41593
```

```
var(y)
```

```
## [1] 1340.117
```

```
sd(y)
```

```
## [1] 36.60761
```

```
quantile(y,probs = 0.9)
```

```
##      90%
```

```
## 93.90308
```

## Referencias

Seila, Andrew F. 1995. "Monte Carlo Solution for Actuarial Problems." *Record of Society of Actuaries* 21 (3A).

Seynaeve, Daan, and Tobias Verbeke. 2017. "Software for Monte Carlo Simulation of a Simple Risk Assessment." *EFSA Supporting Publications* 14 (11). Wiley Online Library: 1316E.