

Figura: Clase 7: Grandes desvíos y simulación del paseo al azar

# Probabilidad - Clase 7

## Teorema de Grandes desvíos y simulación de un paseo al azar

Ernesto Mordecki

Facultad de Ciencias, Universidad de la República. Montevideo, Uruguay

Curso de la Licenciatura en Matemática - 2020

# Contenidos

Desigualdad de grandes desvíos para el Esquema de Bernoulli  
Cota superior de la probabilidad en grandes desvíos

Cota inferior de la probabilidad en grandes desvíos

Fórmula de Stirling

Comparación de las cotas de Bernoulli y de grandes desvíos

Paseo al azar

# Teorema de Bernoulli. Ley de los grandes números

## Teorema

Dados  $\varepsilon > 0$ ,  $\eta > 0$ , números arbitrariamente pequeños, existe un número natural  $n_0$  tal que, para todo  $n \geq n_0$ ,

$$\mathbf{P} \left( \left| \frac{\mu}{n} - p \right| < \varepsilon \right) \geq 1 - \eta.$$

Mas precisamente

$$\mathbf{P} \left( \left| \frac{\mu}{n} - p \right| > \varepsilon \right) \leq \frac{pq}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2} \rightarrow_n 0.$$

porque  $pq = p(1-p) \leq 1/4$ :

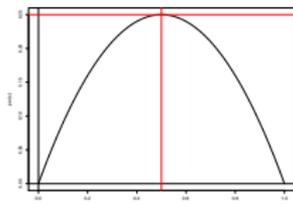


Figura: El máximo se da en  $p = 1/2$  y vale  $1/4$ .

# Desigualdad de grandes desvíos para el Esquema de Bernoulli

- ▶ Consideremos un esquema de Bernoulli con  $n$  experimentos, probabilidad de éxito  $p$ ,  $0 < p < 1$ , y  $q = 1 - p$ .
- ▶ Esperamos que  $\mu/n \sim p$ .
- ▶ Nos interesa encontrar una **cota** para la probabilidad de que la frecuencia  $\mu/n$  se desvíe de el valor que esperamos  $p$  en una cierta cantidad.
- ▶ Elegimos  $\delta = p + \varepsilon > p$  y acotamos

$$\mathbf{P} \left( \frac{\mu}{n} \geq \delta \right).$$

# Cota superior de la probabilidad en grandes desvíos

Veamos Para acotar elegimos  $\theta > 0$  auxiliar. Haremos una cuenta parecida a la demostración del Teorema de Bernoulli (el truco del **1**):

$$1 \leq \frac{e^{\theta k}}{e^{\theta \delta n}}$$

para  $k \geq n\delta$ . Tenemos:

$$\begin{aligned} \mathbf{P}\left(\frac{\mu}{n} \geq \delta\right) &= \sum_{k \geq n\delta} \mathbf{P}(\mu = k) = \sum_{k \geq n\delta} 1 \binom{n}{k} p^k q^{n-k} \\ &\leq \sum_{k \geq n\delta} \frac{e^{\theta k}}{e^{\theta \delta n}} \binom{n}{k} p^k q^{n-k} \\ &\leq e^{-\theta \delta n} \sum_{k=1}^n e^{\theta k} \binom{n}{k} p^k q^{n-k} \end{aligned}$$

$$\begin{aligned}\mathbf{P}\left(\frac{\mu}{n} \geq \delta\right) &\leq e^{-\theta\delta n} \sum_{k=1}^n e^{\theta k} \binom{n}{k} p^k q^{n-k} \\ &= e^{-\theta\delta n} \sum_{k=1}^n \binom{n}{k} (pe^\theta)^k q^{n-k} \\ &= e^{-\theta\delta n} (pe^\theta + q)^n \quad (\text{Binomio de Newton}) \\ &= e^{-\theta\delta n} e^{n \log(pe^\theta + q)} \\ &= \exp\left(-n \left[\theta\delta - \log(pe^\theta + q)\right]\right).\end{aligned}$$

Consideramos entonces la función (auxiliar también):

$$f(\theta) = \theta\delta - \log(pe^\theta + q),$$

Lo que obtuvimos es

$$\mathbf{P}\left(\frac{\mu}{n} \geq \delta\right) \leq \exp(-nf(\theta)).$$

y elegimos  $\theta$  para maximizarla, lo que nos va a dar la mejor cota con esta metodología.

Derivamos

$$f'(\theta) = \delta - \frac{pe^\theta}{pe^\theta + q},$$

de donde el máximo se da en

$$\theta^* = \log(\delta q / (1 - \delta)p),$$

y el valor máximo  $f'(\theta^*)$  (después de algunas cuentas) es

$$H(\delta) = \delta \log \frac{\delta}{p} + (1 - \delta) \log \frac{1 - \delta}{q}.$$

Observemos que  $H(p) = 0$ , la cota nos da uno (es una cota no informativa). Pero para  $\delta > p$  obtenemos una acotación útil

# Cota de Grandes desvíos

**Teorema.** Para un esquema de Bernoulli, dado  $\delta > p$  se tiene la *desigualdad de grandes desvíos*:

$$\mathbf{P} \left( \frac{\mu}{n} \geq \delta \right) \leq e^{-nH(\delta)},$$

Observemos que de la desigualdad obtenida, si tomamos logaritmo

$$\log \mathbf{P} \left( \frac{\mu}{n} \geq \delta \right) \leq -nH(\delta),$$

y si dividimos por  $n$  y tomamos límite superior

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P} \left( \frac{\mu}{n} \geq \delta \right) \leq -H(\delta).$$

# Cota inferior del intervalo

- ▶ Queremos acotar la frecuencia por valores menores:

$$\gamma = p - \varepsilon \leq \frac{\mu}{n}$$

- ▶ Así tenemos acotada la probabilidad de que

$$\left| \frac{\mu}{n} - p \right| \geq \varepsilon.$$

- ▶ Para  $\gamma < p$ , queremos acotar

$$\mathbf{P}\left(\frac{\mu}{n} \leq \gamma\right).$$

- ▶ **Estrategia:** ¡convertir los fracasos en éxitos!,
- ▶ Contamos

$$\tilde{\mu} = n - \mu.$$

- ▶ tenemos un esquema con probabilidad de éxito  $q$ .

Tenemos

$$\begin{aligned}\mathbf{P}\left(\frac{\mu}{n} \leq \gamma\right) &= \mathbf{P}\left(1 - \frac{\mu}{n} \geq 1 - \gamma\right) \\ &= \mathbf{P}\left(\frac{n - \mu}{n} \geq 1 - \gamma\right) = \mathbf{P}\left(\frac{\tilde{\mu}}{n} \geq 1 - \gamma\right) \leq e^{-n\tilde{H}(1-\gamma)}.\end{aligned}$$

donde  $\tilde{H}$  se obtiene de  $H$  cambiando  $p$  por  $q$ , es decir

$$\tilde{H}(\delta) = \delta \log \frac{\delta}{q} + (1 - \delta) \log \frac{1 - \delta}{p} = H(1 - \delta),$$

y la cota es

$$\mathbf{P}\left(\frac{\mu}{n} \leq \gamma\right) \leq \exp(-nH(\gamma)).$$

# Teorema de Grandes Desvíos

**Teorema.** Para un esquema de Bernoulli con probabilidad  $p$  de éxito, se tiene

$$\begin{aligned}\mathbf{P}\left(\left|\frac{\mu}{n} - p\right| \geq \varepsilon\right) &= \mathbf{P}\left(\frac{\mu}{n} \geq p + \varepsilon\right) + \mathbf{P}\left(\frac{\mu}{n} \leq p - \varepsilon\right) \\ &\leq \exp(-nH(p + \varepsilon)) + \exp(-nH(p - \varepsilon)).\end{aligned}$$

En el caso simétrico, obtenemos

$$\mathbf{P}\left(\left|\frac{\mu}{n} - \frac{1}{2}\right| \geq \varepsilon\right) \leq 2 \exp(-nH(\varepsilon))$$

donde

$$H(\varepsilon) = (1/2 - \varepsilon) \log(1 - 2\varepsilon) + (1/2 + \varepsilon) \log(1 + 2\varepsilon)$$

# Cota inferior de la probabilidad en grandes desvíos

Obtuvimos

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P} \left( \frac{\mu}{n} \geq \delta \right) \leq -H(\delta).$$

Veamos que este límite superior es efectivamente un límite.  
Acotamos inferiormente con  $k = [n\delta] + 1$ :

$$\mathbf{P} \left( \frac{\mu}{n} \geq \delta \right) \geq \mathbf{P} (\mu = k) = C_k^n p^k (1-p)^{n-k}.$$

de donde

$$\frac{1}{n} \log \mathbf{P} \left( \frac{\mu}{n} \geq \delta \right) \geq \frac{1}{n} \log \left( C_k^n p^k (1-p)^{n-k} \right).$$

# Fórmula de Stirling

## Teorema

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} (n/e)^n} = 1.$$

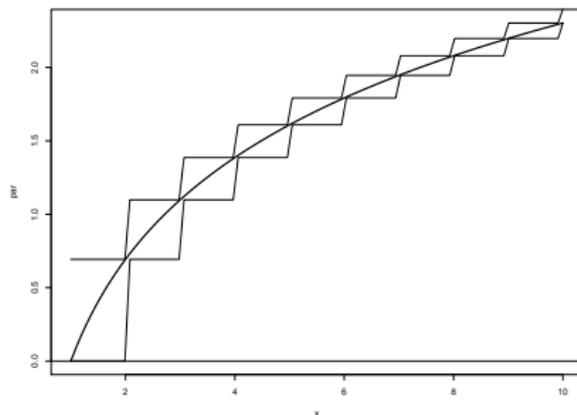
O en forma equivalente

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n, \quad \text{cuando } n \rightarrow \infty$$

# Idea de la demostración

$$\log n! = \log 1 + \log 2 + \cdots + \log n = \sum_{k=2}^n \log k.$$

Tenemos además (recordar la prueba del criterio integral)



De aquí

$$\log 2 + \dots + \log(n-1) \leq \int_1^n \log x dx \leq \log 2 + \dots + \log n + \log n$$

Es decir

$$\int_1^n \log x dx \leq \log 2 + \dots + \log n \leq \int_1^{n+1} \log x dx$$

Además

$$\int_1^n \log x dx = n \log n - n + 1$$

Calculemos

$$e^{\int_1^n \log x dx} = e^{n \log n - n + 1} = e^{n \log n} e^{-n} e = n^n e^{-n} e = e \left( \frac{n}{e} \right)^n$$

Tomando exponencial en la desigualdad de arriba, obtenemos

$$\begin{aligned} e \left(\frac{n}{e}\right)^n &\leq n! \leq e \left(\frac{n+1}{e}\right)^{n+1} = e \left(\frac{n+1}{e}\right)^n \left(\frac{n+1}{e}\right) \\ &= e \left(\frac{n+1}{e}\right)^n \left(\frac{n+1}{e}\right) = (n+1) \left(\frac{n}{e}\right)^n \left(\frac{n+1}{n}\right)^n \\ &\leq (n+1) e \left(\frac{n}{e}\right)^n. \end{aligned}$$

basado en la desigualdad

$$\left(1 + \frac{1}{n}\right)^n = \left(\frac{n+1}{n}\right)^n \leq e \leq \left(\frac{n+1}{n}\right)^{n+1}$$

Esto no prueba la equivalencia pero da una idea del crecimiento. El crecimiento exacto, que está entre  $e$  y  $e(n+1)$  es

$$\sqrt{2\pi n}.$$

Usemos entonces la fórmula de Stirling para el factorial:

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

Habíamos establecido, con  $k = [n\delta] + 1$ , que

$$\frac{1}{n} \log \mathbf{P} \left( \frac{\mu}{n} \geq \delta \right) \geq \frac{1}{n} \log \left( C_k^n p^k (1-p)^{n-k} \right).$$

La cuenta es:

$$\begin{aligned} C_k^n p^k (1-p)^{n-k} &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &\sim \frac{\sqrt{2\pi n}}{\sqrt{2\pi k} \sqrt{2\pi(n-k)}} \left(\frac{n}{e}\right)^n \left(\frac{e}{k}\right)^k \left(\frac{e}{n-k}\right)^{n-k} \\ &= \sqrt{\frac{n}{2\pi k(n-k)}} \left(\frac{n}{k}\right)^k \left(\frac{n}{n-k}\right)^{n-k} p^k (1-p)^{n-k} \\ &= \sqrt{\frac{n}{2\pi k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{n(1-p)}{n-k}\right)^{n-k} \end{aligned}$$

Primero vemos que

$$\frac{k}{n} = \frac{[n\delta] + 1}{n} \rightarrow \delta.$$

La cuenta, tomando logaritmo es:

$$\begin{aligned} & \frac{1}{n} \log \left( C_k^n p^k (1-p)^{n-k} \right) \\ & \sim \frac{1}{n} \log \left( \sqrt{\frac{n}{2\pi k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{n(1-p)}{n-k}\right)^{n-k} \right) \\ & = \frac{1}{2n} \log \left( \frac{n}{2\pi k(n-k)} \right) - \frac{k}{n} \log \left( \frac{k/n}{p} \right) - \frac{n-k}{n} \log \left( \frac{1-k/n}{1-p} \right) \\ & \rightarrow -\delta \log \left( \frac{\delta}{p} \right) - (1-\delta) \log \left( \frac{1-\delta}{1-p} \right) = -H(\delta), \end{aligned}$$

En vista de (1), obtenemos la velocidad de los grandes desvíos en el esquema de Bernoulli:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P} \left( \frac{\mu}{n} \geq \delta \right) = -H(\delta). \quad (1)$$

# Comparación de las cotas de Bernuolli y de grandes desvíos

La idea central de las acotaciones es acotar el error de la estimación

$$\frac{\mu}{n} \sim p.$$

Para eso se asume un error  $\varepsilon$  tolerable y se estudia la probabilidad de cometer un error mayor, es decir la probabilidad

$$\mathbf{P} \left( \left| \frac{\mu}{n} - p \right| \geq \varepsilon \right).$$

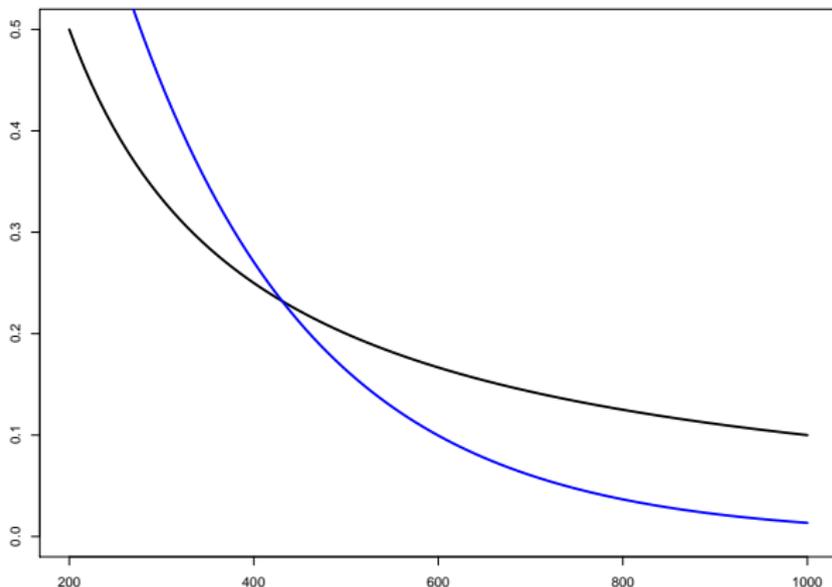
La cota de Bernoulli fué obtenida en 1713, mientras que el teorema de grandes desvíos, en una forma más general que la presentada, fue obtenido por Cramér en 1936. En la Figura 24 vemos el error de ambas fórmulas como función de  $n$  para  $\varepsilon = 0,05$ .

Tenemos

$$H(1/2+0,05) = H(0,55) = 0,45 \log 0,9 + 0,55 \log 1,1 = 0.005.$$

Las cotas de Bernuolli y de grandes desvíos son entonces

$$B(n) = \frac{1}{4(0,05)^2 n} = \frac{100}{n}, \quad GD(n) = 2 \exp(-0.005n).$$



# Simulación de un paseo al azar

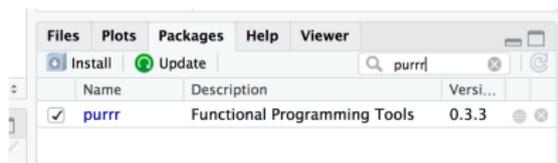
Habíamos definido un paso al azar como el movimiento de una partícula en los enteros, correspondiente a un esquema de Bernolli:

- ▶ Empezamos en  $x = 0$ .
- ▶ Si el primer experimento es un éxito, movemos una unidad a la derecha,
- ▶ Si el primer experimento es un fracaso, movemos una unidad a la izquierda,
- ▶ Si así llegamos a  $x$ , pasamos a  $x + 1$  con probabilidad  $p$  y a  $x - 1$  con probabilidad  $q$

# Simulación

Hay muchas formas de simular un paseo al azar. Nosotros vamos a hacer la más parecida al teórico Para eso

- ▶ Cargamos el paquete *purrr* escribiendo  
`>library(purrr)`
- ▶ o podemos hacerlo en la ventana de paquetes del R studio



- ▶ ahora podemos simular un vector con los resultados de  $n$  experimentos de Bernoulli con probabilidad de éxito  $p$ , escribiendo
- ▶  $n=100$   
 $p=0.5$   
`experimentos=rbernoulli(n,p)`

- ▶ Si vemos que hay en `experimentos`, obtenemos

```
> experimentos
```

```
[1] TRUE TRUE FALSE FALSE FALSE TRUE TRUE TRUE TRUE TRUE FALSE TRUE FALSE F
[15] TRUE TRUE FALSE TRUE TRUE FALSE TRUE TRUE TRUE FALSE TRUE FALSE TRUE F
[29] TRUE TRUE FALSE FALSE FALSE TRUE FALSE TRUE TRUE TRUE FALSE TRUE
[43] TRUE TRUE TRUE FALSE FALSE TRUE FALSE TRUE FALSE TRUE TRUE TRUE F
[57] TRUE FALSE TRUE TRUE FALSE FALSE TRUE FALSE TRUE FALSE FALSE TRUE
[71] FALSE FALSE FALSE TRUE FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE F
[85] TRUE FALSE TRUE TRUE FALSE FALSE FALSE TRUE FALSE TRUE FALSE TRUE F
[99] FALSE FALSE
```

- ▶ En realidad, TRUE es un uno, y FALSE es un cero. Por ejemplo

```
> sum(experimentos)
```

```
[1] 51
```

- ▶ Ahora tenemos que pasar de ceros y unos a  $\pm 1$ . La fórmula es

$$y = 2x - 1, \quad x = 0, 1.$$

- ▶ Entonces definimos

```
pasos<-2*experimentos-1
```

- ▶ Ahora tenemos que acumular los pasos. Recordemos que empezamos desde  $x = 0$ . Para eso se utiliza el comando *cumsum*, que aplicado a un vector produce las sumas acumuladas. Si fuera una serie, produce las reducidas. Además agregamos el cero. Tenemos

```
paseo<-c(0, cumsum(pasos))
```

- ▶ En total, escribimos

```
> n=100
```

```
> p=0.5
```

```
> experimentos<-rbernoulli(n, p)
```

```
> pasos<-2*experimentos-1
```

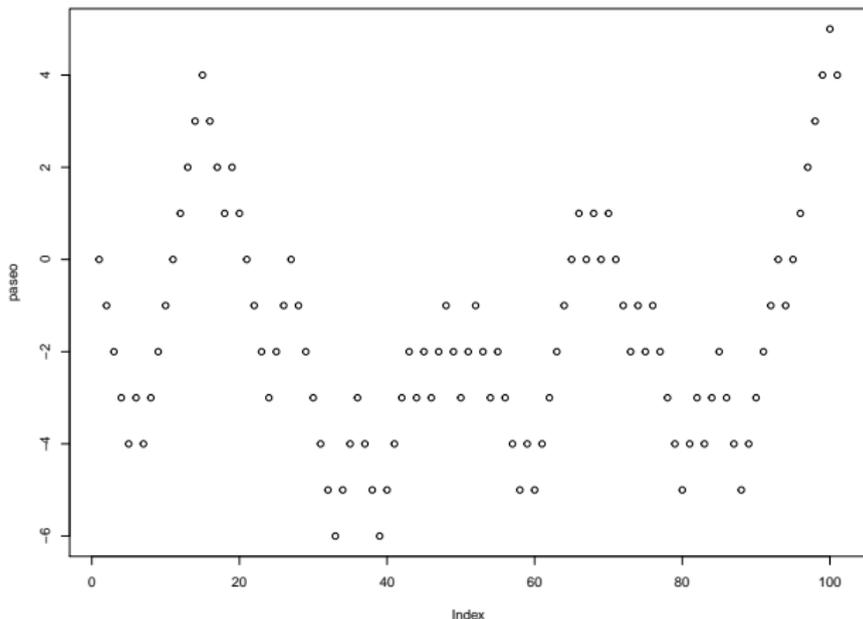
```
> paseo<-c(0, cumsum(pasos))
```

```
> paseo
```

```
[1] 0 -1 -2 -3 -4 -3 -4 -3 -2 -1 0 1 2 3 4 3 2 1 2 1 0 -1 -2 -3  
[30] -3 -4 -5 -6 -5 -4 -3 -4 -5 -6 -5 -4 -3 -2 -3 -2 -3 -2 -1 -2 -3 -2 -1 -2  
[59] -4 -5 -4 -3 -2 -1 0 1 0 1 0 1 0 -1 -2 -1 -2 -1 -2 -3 -4 -5 -4 -3  
[88] -5 -4 -3 -2 -1 0 -1 0 1 2 3 4 5 4
```

```
>
```

- ▶ Ahora graficamos el paseo al azar mediante `plot(paseo)`:



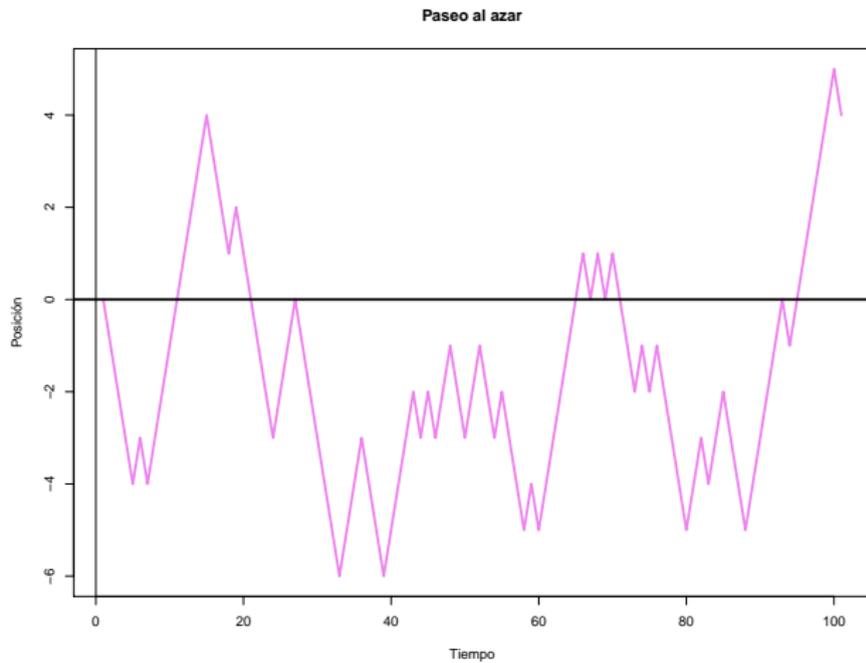
- ▶ Pero queremos embellecer nuestro dibujo

► **Escribimos entonces**

```
plot(paseo,type="l",lwd=3,  
col="violet",xlab = "Tiempo", ylab =  
"Posición",main = "Paseo al azar")  
abline(h=0,lwd=3)  
abline(v=0)
```

- `type="l"` define las líneas continuas en vez de los puntos, `lwd=3` aumenta tres veces el grosor del dibujo, `col` el color del gráfico, `xlab`, `ylab` son las etiquetas que van en los ejes, `main` es el nombre del gráfico, `abline(h=0)` agrega el eje `x`, `abline(v=0)` agrega el eje `y`.

# El resultado:



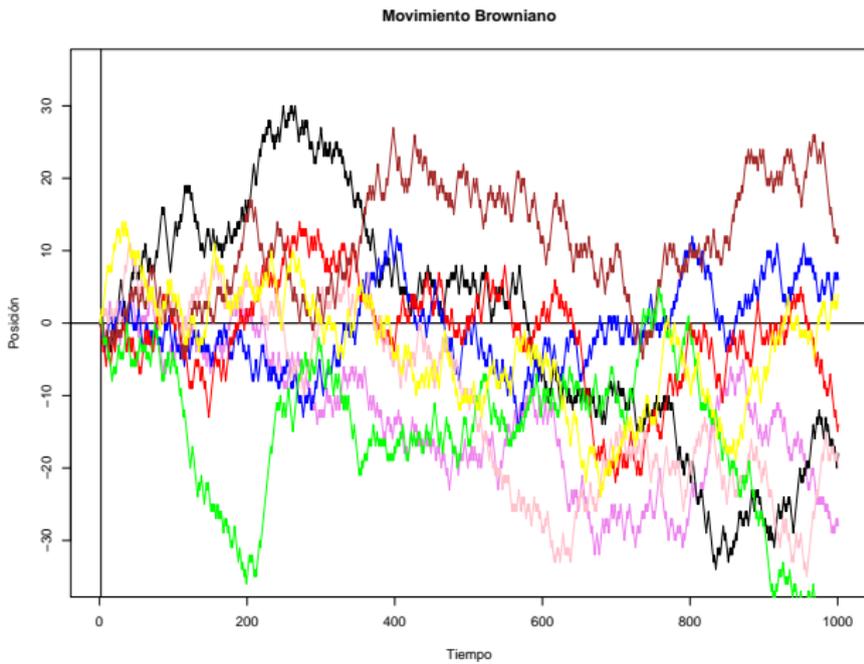


Figura: ¡Buen fin de semana!