Ley de los Grandes Números

Mariela Sued, Jemina García y Ana M. Bianco

03 junio 2020

Un caso especial: muestra aleatoria (m.a.)

Un caso especial: muestra aleatoria (m.a.)

1. Repetimos un experimento muchas veces en idénticas condiciones y de manera independiente.
2. X_i: resultado de la i-ésima repetición.
3. $X_i \sim F$: tienen todas la misma distribución porque repetimos en idénticas condiciones.
4. $X_1, \ldots, X_n$ son independientes (estadísticamente) por construcción: las repeticiones se realizan de manera independiente (en sentido de la vida).

Un caso especial: muestra aleatoria (m.a.)

1. Repetimos un experimento muchas veces en idénticas condiciones y de manera independiente.
2. X_i: resultado de la i-ésima repetición.
3. $X_i \sim F$: tienen todas la misma distribución porque repetimos en idénticas condiciones.
4. $X_1, \ldots, X_n$ son independientes (estadísticamente) por construcción: las repeticiones se realizan de manera independiente (en sentido de la vida).

Luego, $X_i \sim F$ para todo i, y por consiguiente:

- $\mathbb{P}(X_i \leq t) = \mathbb{P}(X_j \leq t) = \mathbb{P}(X_1 \leq t)$
- $\mathbb{E}[X_i] = \mathbb{E}[X_j] = \mathbb{E}[X_1]$
- $\mathbb{V}(X_i) = \mathbb{V}(X_j) = \mathbb{V}(X_1)$

- Dicho de otra forma:
  mongo(X_i)=mongo(X_j)=mongo(X_1).

Un caso especial: muestra aleatoria (m.a.)

1. Repetimos un experimento muchas veces en idénticas condiciones y de manera independiente.
2. X_i: resultado de la i-ésima repetición.
3. $X_i \sim F$: tienen todas la misma distribución porque repetimos en idénticas condiciones.
4. $X_1, \ldots, X_n$ son independientes (estadísticamente) por construcción: las repeticiones se realizan de manera independiente (en sentido de la vida).

Luego, $X_i \sim F$ para todo i, y por consiguiente:

- $\mathbb{P}(X_i \leq t) = \mathbb{P}(X_j \leq t) = \mathbb{P}(X_1 \leq t)$
- $\mathbb{E}[X_i] = \mathbb{E}[X_j] = \mathbb{E}[X_1]$
- $\mathbb{V}(X_i) = \mathbb{V}(X_j) = \mathbb{V}(X_1)$

- Dicho de otra forma: mongo(X_i)=mongo(X_j)=mongo(X_1).

Muestra (aleatoria) o variables i.i.d.

$X_1, \dots, X_n$ son una muestra aleatoria si son v. a. independientes, idénticamente distribuídas.

$$\boxed{X_1, \dots, X_n \text{ , i.i.d.}}$$

En tal caso, $X_i \sim F$ para todo i, y por consiguiente,

- mongo(X_i)=mongo(X_j)=mongo(X_1).

Muestra (aleatoria) o variables i.i.d.

¿Qué relación guarda la muestra con nuestros datos?

Una noche en el casino: Ruleta

¿Apostamos una fichita?

- Si acierto: me devuelven 36 fichas (35+la mía)
- Si pierdo: se llevan mi ficha

¿Apostamos una fichita?

- Si acierto: me devuelven 36 fichas (35+la mía)
- Si pierdo: se llevan mi ficha

Dicho de otro modo:

- Si acierto: Gano 35
- Si pierdo: Gano -1

Una larga noche . . .

[1] – 1 – 1 35 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1
[20] – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 35 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1
[39] – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1
[58] – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1
[77] – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1
[96] – 1 – 1 – 1 – 1 35 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1
[115] – 1 – 1 – 1 – 1 – 1 – 1 35 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1
[134] – 1 – 1 – 1 – 1 – 1 – 1 – 1 35 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1
[153] – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1 – 1

estos son los resultados...
...ésta es la historia de lo que pasó

. . . pensemos. . .

Podríamos decir que los datos que les doy son

- el esultado o lo que observamos en cada jugada:

$$x_1 = -1, x_2 = -1, x_3 = 35, \dots, x_{27} = -1, \dots, x_{28} = 35, x_{29} = -1$$

Estos datos, ¿son los resultados o las observaciones de qué?

. . . pensemos. . .

- $x_1 = -1$ es el valor observado de la *ganancia obtenida en la primera jugada* o sea de la v.a. X_1

. . . pensemos. . .

- $x_1 = -1$ es el valor observado de la *ganancia obtenida en la primera jugada* o sea de la v.a. X_1
- $x_2 = -1$ es el valor observado de la *ganancia obtenida en la segunda jugada* o de sea la v.a. X_2

. . . pensemos. . .

- $x_1 = -1$ es el valor observado de la *ganancia obtenida en la primera jugada* o sea de la v.a. X_1
- $x_2 = -1$ es el valor observado de la *ganancia obtenida en la segunda jugada* o de sea la v.a. X_2
- $x_3 = 35$ es el valor observado de la *ganancia obtenida en la tercera jugada* o de sea la v.a. X_3
- .....
- $x_i = -1$ es el valor observado de la *ganancia obtenida en la i–ésima jugada* o sea de la v.a. X_i
- ....
- $X_1, X_2, \ldots, X_i \ldots, X_n$ son variables aleatorias idénticamente distribuidas

. . . pensemos. . .

- $x_1 = -1$ es el valor observado de la *ganancia obtenida en la primera jugada* o sea de la v.a. X_1
- $x_2 = -1$ es el valor observado de la *ganancia obtenida en la segunda jugada* o de sea la v.a. X_2
- $x_3 = 35$ es el valor observado de la *ganancia obtenida en la tercera jugada* o de sea la v.a. X_3
- .....
- $x_i = -1$ es el valor observado de la *ganancia obtenida en la i–ésima jugada* o sea de la v.a. X_i
- ....
- $X_1, X_2, \ldots, X_i \ldots, X_n$ son variables aleatorias idénticamente distribuidas
- Por las condiciones de *este experimento* las X_i resultan independientes.

$$X_1, X_2, \ldots, X_i \ldots, X_n \text{ i.i.d.}$$

Para pensar: ¿Siempre es así?

Supongamos que en una caja tenemos 100 bolitas: 90 blancas y 10 rojas. Consideremos las siguientes dos situaciones:

- extraemos 5 bolitas **con reposición**. Para cada extracción definimos $1 \leq i \leq 5$:

$$X_i = \begin{cases} 1 & \text{si la i-ésima bolita extraída es roja} \\ 0 & \text{si la i-ésima bolita extraída es blanca} \end{cases}$$

¿Son las X_i v.a. i.i.d.? ....

Para pensar: ¿Siempre es así?

Supongamos que en una caja tenemos 100 bolitas: 90 blancas y 10 rojas. Consideremos las siguientes dos situaciones:

- extraemos 5 bolitas **con reposición**. Para cada extracción definimos $1 \leq i \leq 5$:

$$X_i = \begin{cases} 1 & \text{si la i-ésima bolita extraída es roja} \\ 0 & \text{si la i-ésima bolita extraída es blanca} \end{cases}$$

¿Son las X_i v.a. i.i.d.? ....

- extraemos 5 bolitas **sin reposición**. Para cada extracción definimos $1 \leq i \leq 5$:

$$Y_i = \begin{cases} 1 & \text{si la i-ésima bolita extraída es roja} \\ 0 & \text{si la i-ésima bolita extraída es blanca} \end{cases}$$

¿Son las Y_i v.a. independientes? ....

Para seguir pensando. . .

¿Qué relación guarda todo esto con nuestras simulaciones?

Para seguir pensando...

¿Qué relación guarda todo esto con nuestras simulaciones?

¿Qué estamos haciendo con el comando `runif(10,0,1)`?

Volviendo a la muestra

$X_1, \ldots, X_n$ i.i.d., con $\mathbb{E}[X_i] = \mu$ y $\mathbb{V}[X_i] = \sigma^2$, para todo i.

- $S_n = \sum_{i=1}^{n} X_i \quad \Longrightarrow \mathbb{E}[S_n] = n\,\mu$ y $\mathbb{V}[S_n] = n\,\sigma^2$

- $\overline{X} = \overline{X}_n = \dfrac{\sum_{i=1}^{n} X_i}{n} \quad \Longrightarrow \mathbb{E}[\overline{X}_n] = \mu$ y $\mathbb{V}[\overline{X}_n] = \dfrac{\sigma^2}{n}$

Caso Particular: Suma y Promedio de normales

- $(X_i)_{1\leq i\leq n}$ i.i.d., $X_i \sim \mathcal{N}(\mu, \sigma^2)$
- Suma: $S_n = X_1 + \ldots + X_n \sim ?$
- Pomedio: $\overline{X}_n = \frac{1}{n}(X_1 + \ldots + X_n) \sim ?$

Vayamos a la lista de tareas

Simulemos:

Estudiaremos la distribución de S_n y de $\overline{X}_n$ de variables $X_1, \ldots, X_n$ i.i.d. $\mathcal{N}(\mu, \sigma^2)$.

Para ello, fijado n, generaremos datos correspondientes a una muestra $X_1, \ldots, X_n$ i.i.d. de variables aleatorias distribuidas como X, con una distribución $\mathcal{N}(5, 4)$ y luego calcularemos la suma y el promedio de cada conjunto de datos.

Repetimos este procedimiento $Nrep = 1000$ veces. A partir de las $Nrep = 1000$ replicaciones realizaremos un histograma con las sumas y los promedios generados, para obtener una aproximación de la densidad de S_n y de $\bar{X}_n$.

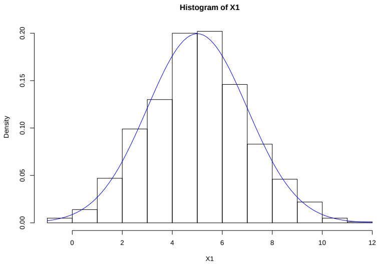
Para ello, fijado n, generaremos datos correspondientes a una muestra $X_1, \ldots, X_n$ i.i.d. de variables aleatorias distribuidas como $X \sim \mathcal{N}(5, 4)$, para $n = 1, 2, 5, 10, 25$. Luego, haremos un histograma para luego tratar de responder a qué densidad se parece el histograma obtenido y superponerle una densidad adecuada.

$n = 1$

- Consideramos $n = 1$ en cuyo caso la variable coincide con la suma y el promedio.

Generamos entonces Nrep $=$1000 datos correspondientes a $X_1 \sim \mathcal{N}(5, 4)$ y luego hacemos un histograma.

```
Nrep<- 1000
set.seed(123)
X1<-rnorm(Nrep,mean=5,sd=2)    # Genero normales
hist(X1,freq=F)
curve(dnorm(x, mean=5,sd=2), add=T, col="blue")
```



$n = 2$

- Consideramos $n = 2$ y las variables

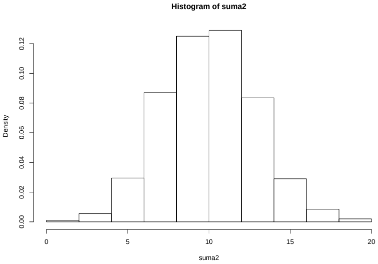
$$S_2 = X_1 + X_2 \qquad \text{y} \qquad \bar{X}_2 = \frac{X_1 + X_2}{2}\,.$$

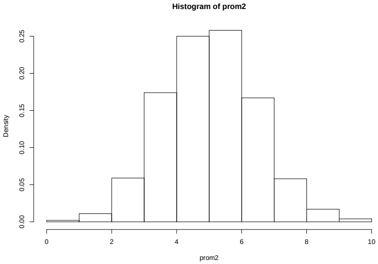
Generamos $n = 2$ datos (independientes) correspondientes a variables aleatorias con distribución $\mathcal{N}(5, 4)$ y computamos la suma y el promedio. Replicamos $Nrep = 1000$ veces y realizamos los dos histograma con los $Nrep = 1000$ promedios obtenidos.

```
Nrep<- 1000
X1<-rnorm(Nrep,mean=5,sd=2)   # Genero normales
X2<-rnorm(Nrep,mean=5,sd=2)   # Genero normales
suma2<- X1+X2
prom2<- suma2/2
```

Graficamos

```r
hist(suma2,freq=F)
hist(prom2,freq=F)
```





Otra forma

```
ene=2
matriz2<-matrix(rnorm(1000*ene,mean=5,sd=2),nrow=ene, ncol=1000,byrow=T)
suma2<-apply(matriz2,2,sum)
prom2<-apply(matriz2,2,mean)
```

Grafico

```
hist(suma2,freq=F)
hist(prom2,freq=F)
```