Grupo MAT ESI UCLM

Sesión 6: Distribuciones de Probabilidad

EUSEBIO ANGULO SÁNCHEZ-HERRERA

CURSO 2021/2022

ESCUELA SUPERIOR DE INFORMÁTICA CIUDAD REAL

Contenidos Principales

- ¿Qué es una variable aleatoria?
- Variables aleatorias discretas
  - Binomial
  - Poisson
- Variables aleatorias continuas
  - Uniforme
  - Normal
  - Exponencial
- Teorema del Límite Central: Importancia de N(0,1)

Variables Aleatorias

- **Variables numéricas**.
  - <u>Cuantitativas</u>: temperatura, tiempo…
  - <u>Cualitativas</u>: control de calidad, colores,…
  - Cuyos valores se determinan aleatoriamente.

- Una variable aleatoria X es definida especificando:
  - x_i, valores posibles, y
  - *pdf->Probability density function*, función de densidad
  - *cdf->Cumulative distribution function*, función de distribución
    - Función de probabilidad (discreta)
    - Función de densidad (continua)

Variables Aleatorias Discretas

- Una variable aleatoria discreta es definida:
  - x_i finita o numerable
  - *pdf* función de probabilidad: $P(X = x_i)$
  - *cdf* función de distribución: $F(x_i) = P(X \leq x_i) = \sum_{j \leq i} P(X = x_j)$

>?distributions

dbinom (binomial), dpois (Poisson), dgeom (geométrica), dnbinom (binomial negativa), dhyper (hipergeométrica)..

Variables Aleatorias Discretas

- X = Número de sucesos en *n* **Bernuilli experiments.**

{pasa o no pasa} ⟷ *{0,1}*

(Control de calidad de un producto)

Distribución de Bernoulli

Tenemos un experimento de Bernoulli si al realizarlo sólo son posibles dos resultados:

- X=1 (**éxito**, con probabilidad **p**)
- X=0 (**fracaso**, con probabilidad **q=1-p**)

Ejemplos

- Lanzar una moneda y que salga cara.
- Elegir una persona de una población de 1000 y que esté enfermo.

Variables Aleatorias Discretas

Distribución Binomial

Si se repite un número fijo de veces, n, un experimento de Bernoulli con parámetro p, la variable aleatoria **X = " número de éxitos en esos n ensayos"** sigue una distribución binomial de parámetros (n, p).

Decimos que: **X~ Bin (n, p)**

Ejemplos

- Lanzar una moneda 10 veces y contar las caras. **Bin(n=10, p=1/2)**

- Valores posibles para X ~ B(**n**, **p**)

Función de Densidad de Probablilidad *pdf* : $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$

p=probabilidad de suceso en cada experimento

Función de Distribución Acumulada cdf : $P(X \le k) = \sum_{j \le k} P(X = j)$

Distribución Binomial

- Ejemplo: Número de productos que pasan un control de calidad
  - μ(media): *np*
  - σ^2 (varianza): *np(1-p)*

>**d**binom(x,size=n,prob=p) # *pdf*

>**p**binom(x,size=n,prob=p) # *cdf*

>**r**binom(x,size=n,prob=p) # *Muestra aleatoria (simulación)*

- Ejemplo: *B(10,1/2), n=10 Bernoulli exponencial con p=1/2*
- Probabilidad de 5 sucesos P(X=5):

  >dbinom(5,size=10,prob=1/2)

  [1] 0.2460938

Distribución Binomial

- Ejemplo: *B(10,1/2), n=10 Bernoulli exponencial con p=1/2*
- Probabilidad de hasta 5 sucesos P(X≤5):

  >pbinom(5,size=10,prob=1/2)

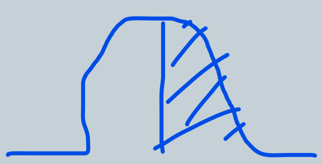
  [1] 0.6230469
- Probabilidad de más de 5 sucesos P(X>5):

  >1-pbinom(5,size=10,prob=1/2) #lower.tail!

  [1] 0.3769531
- Quantile: 0.8: v | P(X≤v)=0.8.

  >qbinom(0.8, size=10,prob=1/2)

probabilidad mayor que
si te preguntan por el valor k



Distribución Binomial

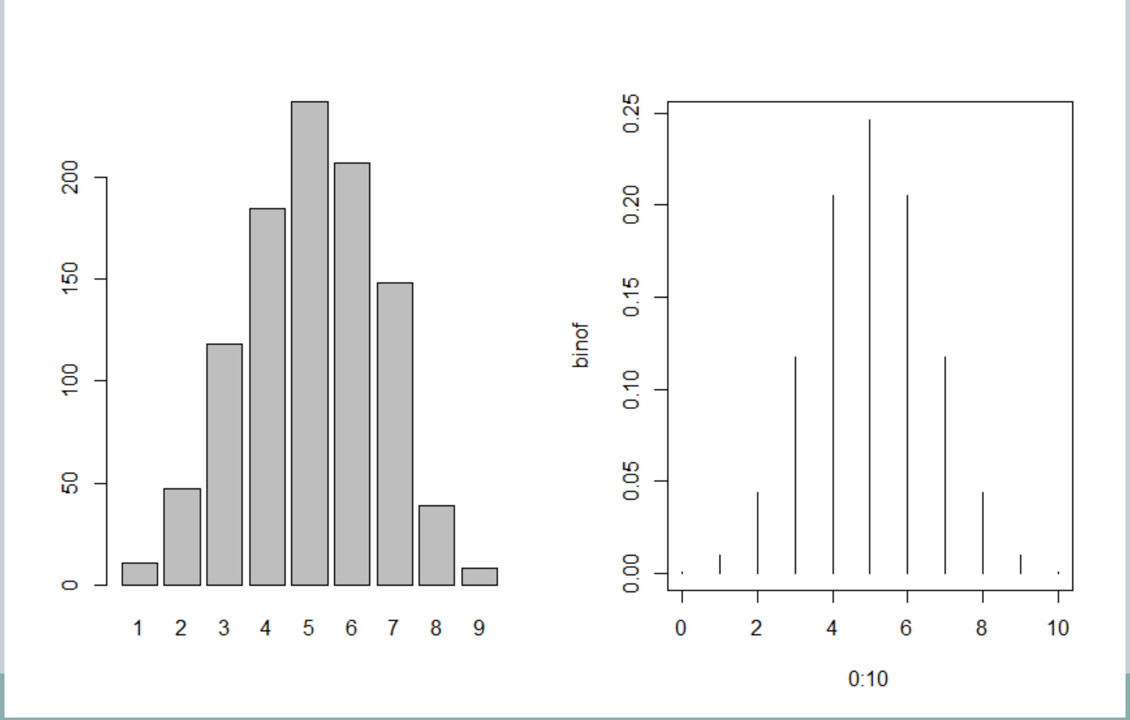
- Resultado de 1000 valores simulados a partir de *B(10,1/2), n=10 con p=1/2:*

  >binoR<-rbinom(1000,size=10,prob=1/2); binoR

```
  [1] 6 5 6 4 5 6 2 6 7 2 7 5 6 5 4 6 6 3 6 5 5 5 7 8 4 3 5 7 6 5 7 3 4 7 7 4 9 7 5 2 1 4 6 4 1 5 4 6 5 7 7 6 4 7 2 7 6 5
 [59] 7 5 4 7 4 5 6 6 7 6 6 3 3 7 4 4 5 7 5 1 6 3 5 9 6 8 7 3 4 7 4 4 5 2 6 5 6 4 5 3 3 3 5 5 6 6 2 4 5 6 7 4 5 6 5 4 6 5
[117] 6 3 5 6 3 7 4 4 5 5 3 5 5 7 6 2 7 7 6 6 3 5 4 7 4 5 4 4 6 6 7 6 7 5 5 2 6 3 4 4 5 5 3 5 5 7 4 7 5 4 5 5 7 6 3 7 6 5
[175] 7 5 4 6 4 5 5 5 5 8 2 3 6 7 5 7 5 7 7 6 3 4 7 7 6 5 4 5 7 8 6 6 7 7 5 9 7 7 7 4 5 4 5 5 5 8 5 5 3 4 8 5 7 6 5 5 5 2
[233] 7 6 6 4 4 4 6 5 3 6 5 7 6 7 3 6 5 7 4 5 5 3 6 5 4 5 5 3 3 7 5 7 6 4 4 7 4 5 4 5 6 8 8 4 6 3 8 5 4 6 6 7 6 5 5 7 5 3
[291] 5 7 2 4 5 4 5 8 3 4 5 6 4 5 5 5 5 7 3 5 4 5 5 7 7 8 6 8 4 5 5 4 4 6 5 6 6 6 6 3 4 1 4 2 3 5 8 6 4 7 5 4 6 4 6 5 4 5
[349] 5 3 4 5 4 5 5 3 7 5 5 5 3 6 4 7 6 6 5 4 3 7 5 3 4 8 3 4 5 4 4 4 2 4 3 5 2 5 6 6 8 6 5 4 2 6 6 3 3 8 3 3 6 4 6 7 3 6
[407] 4 6 2 4 4 6 2 5 4 6 4 7 3 5 4 7 7 7 5 7 7 3 6 2 7 5 6 3 8 7 6 4 5 7 7 7 3 5 7 6 4 6 7 5 4 5 6 3 5 2 6 4 4 6 8 8 9 5
[465] 6 4 6 5 7 2 7 4 4 5 4 4 7 4 6 4 4 6 5 6 5 7 5 2 7 6 6 5 3 1 5 6 7 6 4 3 7 7 6 6 8 5 6 5 5 3 6 5 4 6 7 3 7 2 6 4 2 1
[523] 7 6 7 8 1 7 6 5 7 7 5 6 8 6 5 7 5 6 3 3 5 6 6 4 6 5 2 4 5 6 4 2 6 7 5 3 6 5 6 4 4 4 6 8 5 1 4 8 4 6 2 2 4 2 7 7 8 5
[581] 4 7 5 6 4 6 6 7 3 5 6 3 6 5 4 4 6 3 3 7 6 7 6 7 5 5 8 6 4 4 3 6 5 7 5 7 7 6 4 5 6 3 3 5 8 4 2 6 7 5 4 3 5 6 7 7 5 7
[639] 7 5 5 7 6 7 6 4 4 5 6 7 5 7 5 2 6 5 4 6 4 3 9 4 6 4 6 6 3 7 6 4 1 3 5 3 6 5 5 3 3 6 8 6 2 3 4 4 5 5 3 6 4 4 3 4 5 5
[697] 7 8 5 4 9 5 4 6 6 4 5 5 8 3 5 4 2 6 6 6 4 7 6 4 6 5 6 5 4 5 5 5 6 6 4 4 5 4 4 5 5 4 3 5 7 3 6 6 5 4 6 5 5 5 3 5 1 6
[755] 3 7 2 7 5 5 4 6 7 3 5 7 7 3 8 2 6 3 3 4 8 3 2 4 3 4 5 4 5 7 2 3 5 4 5 6 6 5 6 6 6 8 3 9 4 2 7 5 3 7 6 1 8 6 4 6 6 4
[813] 4 3 6 3 6 7 5 4 4 7 7 6 3 7 3 3 4 4 5 5 7 5 4 6 3 4 3 5 5 6 3 3 6 6 5 4 4 3 4 3 5 5 5 4 6 3 5 6 4 6 5 4 7 2 6 5 4 4
[871] 5 6 8 4 3 3 4 4 2 5 6 7 5 3 3 3 5 4 6 4 7 7 5 2 4 5 2 4 2 5 5 4 6 6 7 6 3 4 4 4 3 6 5 2 6 4 5 7 6 6 4 4 3 4 3 6 2 3
[929] 9 4 7 3 6 8 7 2 8 7 7 7 5 5 5 5 6 3 6 5 3 6 5 6 5 6 7 3 7 6 7 4 5 5 5 3 8 4 7 5 4 4 6 6 6 5 5 7 2 5 5 6 4 6 2 5 6 3
[987] 8 7 6 6 3 7 4 5 3 3 7 3 5 5
```

Distribución Binomial


```
>par(mfrow=c(1,2))        # divide la pantalla gráfica
>barplot(table(binoR))   # diagrama barras datos generados antes
>binof<-dbinom(0:10,size=10,prob=1/2)  # función de densidad
>plot(0:10,binof,type="h");
```



Distribución Poisson

- También se denomina sucesos raros.

Poison n grande y p pequeña

- Se obtiene como aproximación de una distribución binomial con la misma media, para '**n grande**' (n>30) y '**p pequeño**' (p<0,1)
- La variable aleatoria **X="número de éxitos"**, decimos que:

$$X \sim \textbf{Poiss}(\lambda)$$

$$pdf: \quad P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

$$cdf: \quad P(X \leq k) = \sum_{k=1}^{x} \frac{e^{-\lambda}\lambda^k}{k!}$$

- Ejemplos:
  - Tiempo de inactividad no planificado en una línea de embalaje durante 1 día.
  - Número de defectos por m² en placas de acero.

Distribución Poisson

>**d**pois(x, lambda=λ)   # *pdf*

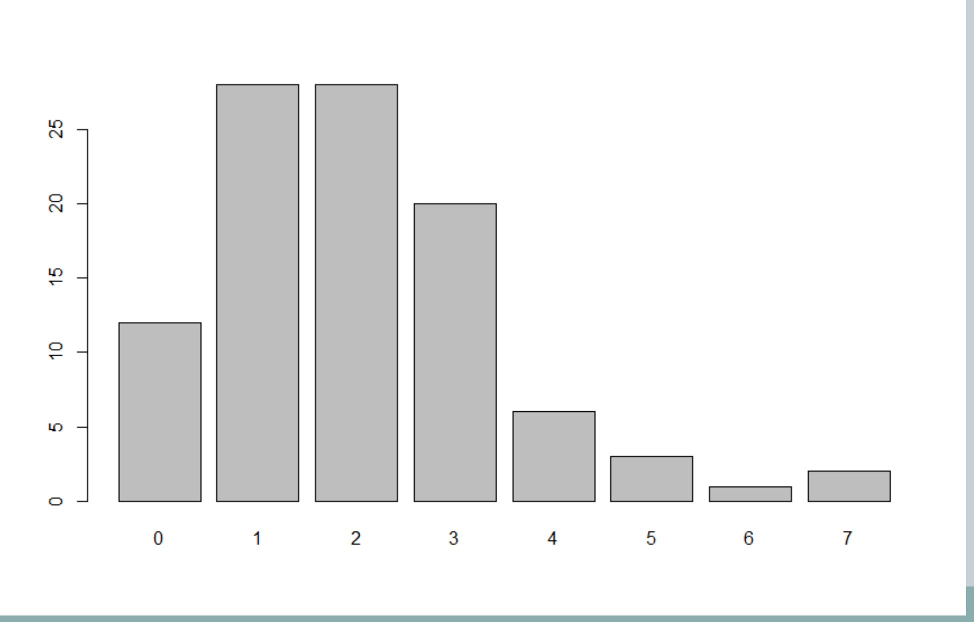
>**p**pois(x, lambda=λ)   # *cdf*

>**r**pois(x, lambda=λ)   # *Muestra aleatoria (simulación)*

> poissonR<-rpois(100,lambda=2);poissonR

[1] 2 1 3 1 6 2 3 4 4 3 1 2 0 3 1 4 2 1 …

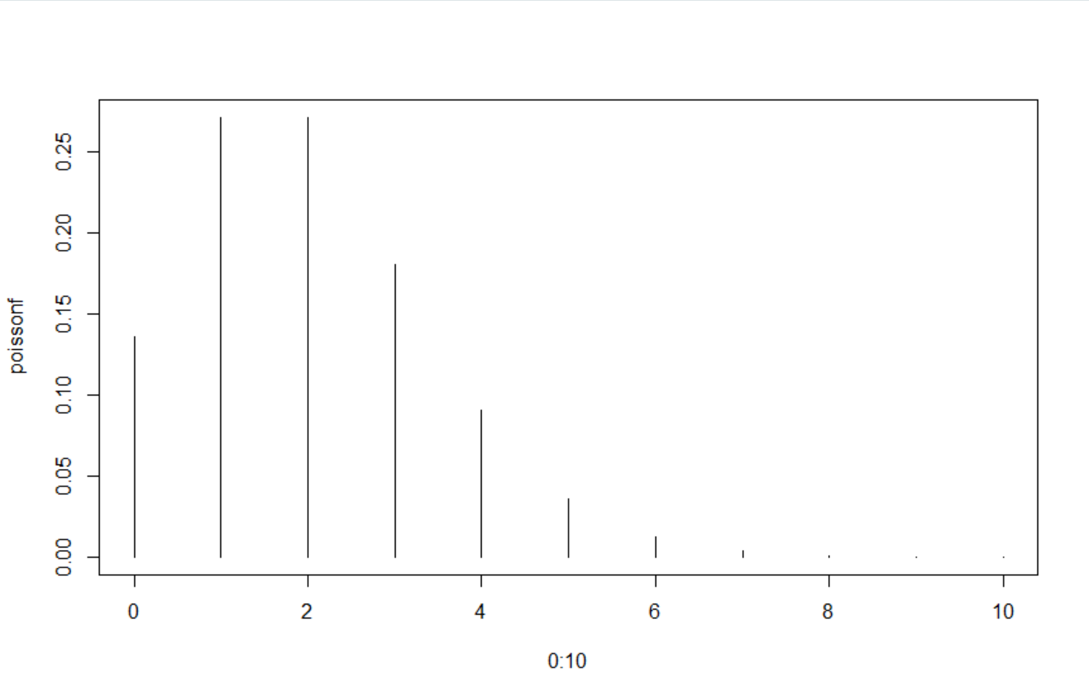
>barplot(table(poissonR))



Distribución Poisson


```
>poissonf<-dpois(0:10,lambda=2)
>plot(0:10,poissonf,type="h")
```

h = como valores discretos



Variables Aleatorias Continuas

- Una variable aleatoria **continua** X es definida por:

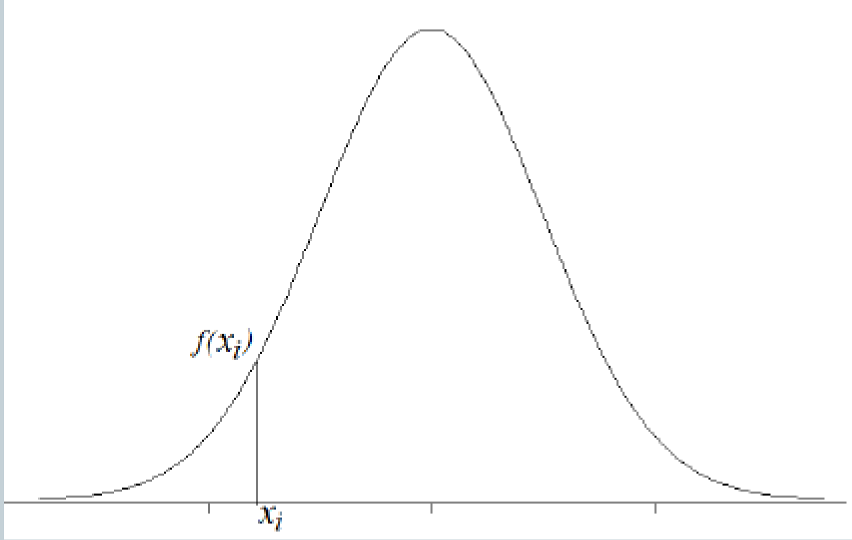
  1) $x_i \in$ Intervalo

  2) *pdf* Función de densidad: $f(x_i) \neq P(X=x_i)$

  3) *cdf* : Función de distribución: $F(x_i) = P(X \leq x_i) = \int_{-\infty}^{x_i} f(x)dx$

$P(X=x_i)=0$

$f(x_i) \neq 0$



>?distributions

dunif (distribución uniforme),

dnorm (distribución normal),

dexp (distribución exponen.),

dchisq, dt, df, dgamma…

Distribución Uniforme

- Valores posibles: [a,b]

$$pdf : f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{en otro caso} \end{cases}$$

$$cdf : F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a < x < b \\ 1 & x > b \end{cases}$$

Distribución Uniforme

- Ejemplo:

  Distribución de grosor de piezas de trabajo.

$$E[X] = \mu : \frac{a+b}{2}$$

$$Var[X] = \sigma^2 : \frac{(b-a)^2}{12}$$

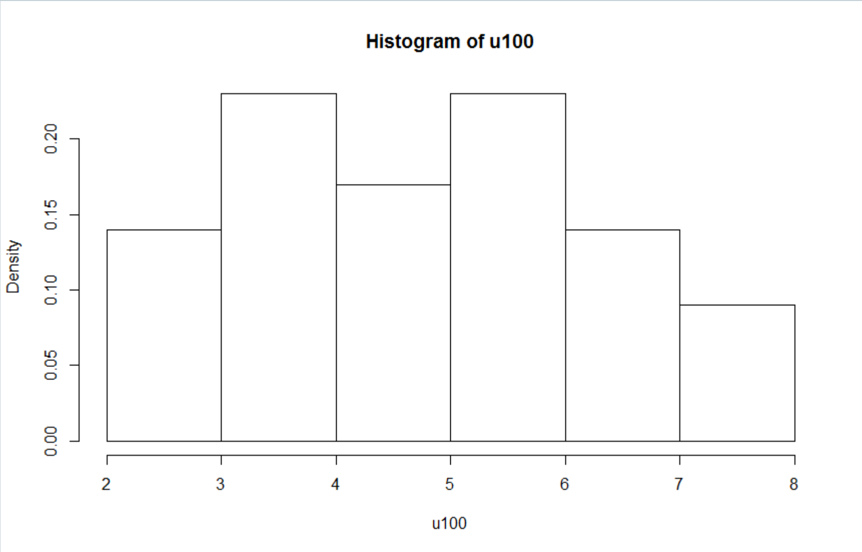
>**d**unif(x,min=a,max=b) # *pdf*

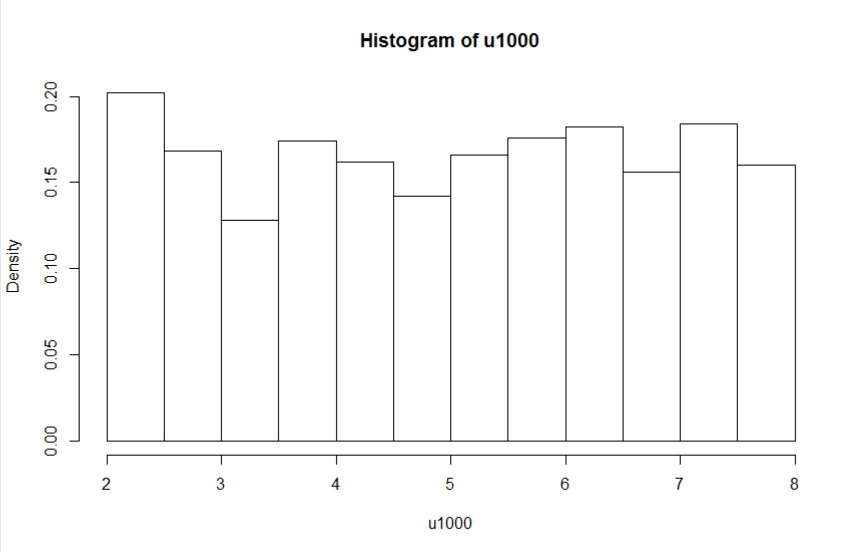
>**p**unif(x,min=a,max=b) # *cdf*

>**r**unif(k,min=a,max=b) # *Muestra aleatoria (simulación)*

Distribución Uniforme


```
>u10<-runif(10,min=2,max=8)      # k = 10 values
>u100<-runif(100,2,8)            # k = 100 values
>u1000<-runif(1000,2,8)          # k = 1000 values
>hist(u100,prob=TRUE)
>hist(u1000,prob=TRUE)
```





Distribución Normal

- Es la más común en las distribuciones continuas

$$pdf : f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$cdf : F(x) = \int_{-\infty}^{x} f(x)dx$$

$$\text{Si no existe primitiva} \Rightarrow \text{consultar tabla}$$

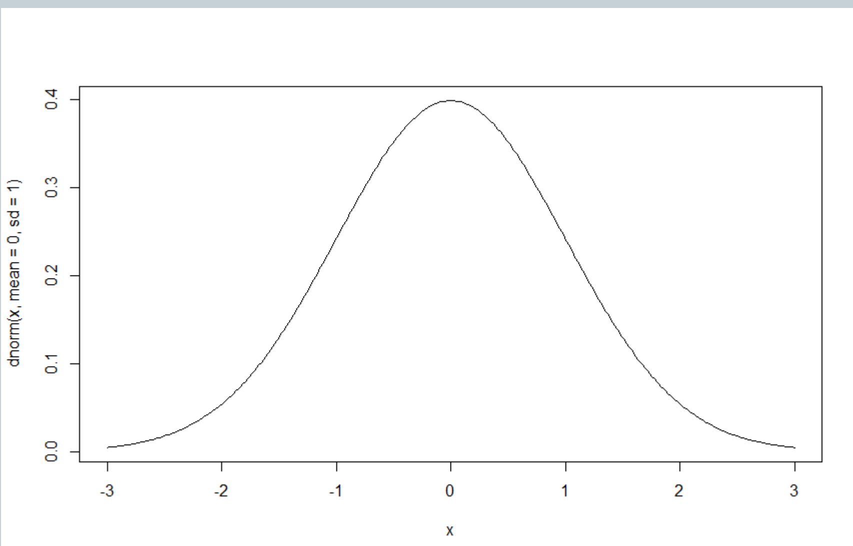
>dnorm(x,mean=μ,sd= σ) # *pdf*

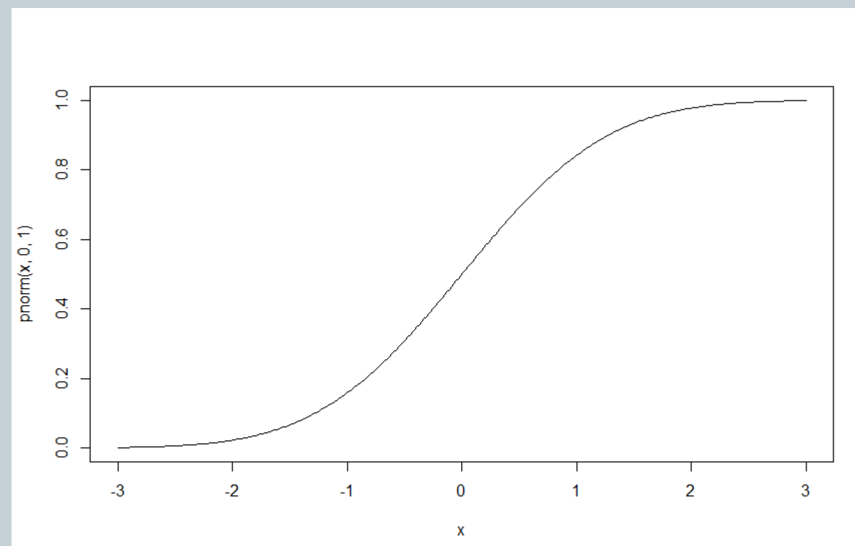
>pnorm(x,mean=μ,sd= σ) # *cdf*

>rnorm(k,mean=μ,sd= σ) # *Muestra aleatoria (simulación)*

Distribución Normal

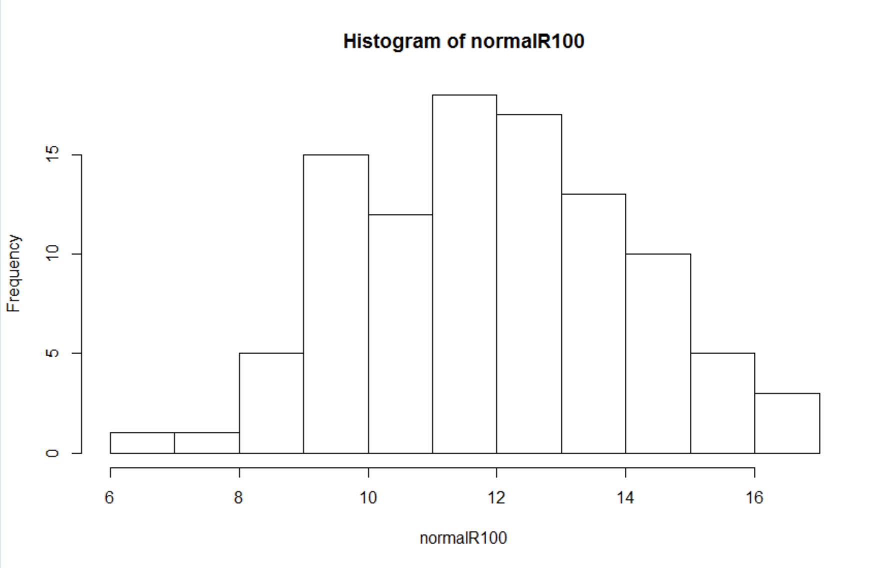

```
>x<-seq(-3,3,length=300)
>plot(x,dnorm(x,mean=0,sd=1),type="l")
>plot(x,pnorm(x,0,1),type="l")
>dnorm(-1,0,1);pnorm(-1,0,1)
```

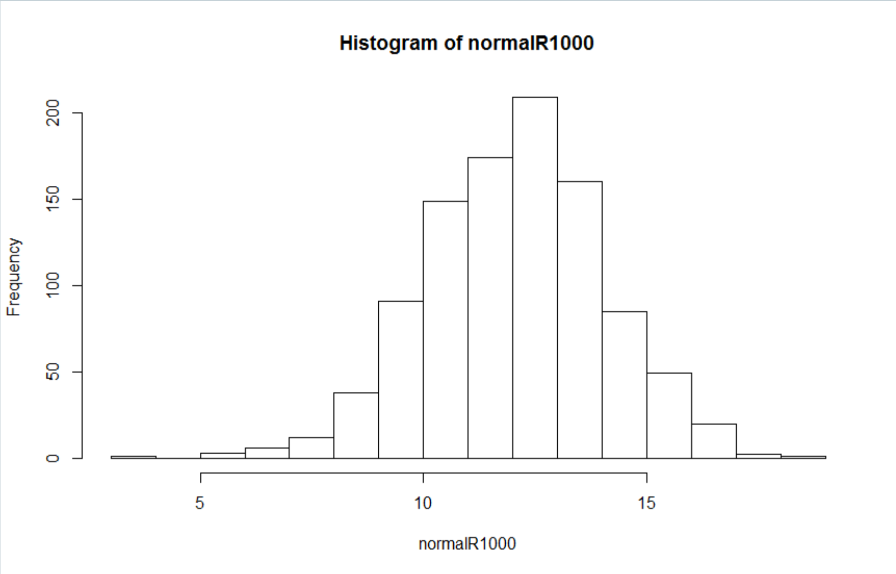




Distribución Normal


```
>normalR100<-rnorm(100,mean=12,sd=2) # k=100 values
>normalR1000<-rnorm(1000,12,2)          #k=1000 values
>hist(normalR100)
>hist(normalR1000)
```





Distribución Exponencial

- Tiempos de vida de individuos, componentes, tiempos de servicio…

$$pdf\ :\ f(x) = \lambda e^{-\lambda x}$$

$$cdf\ : F(x) = 1 - e^{-\lambda x}$$

- Tiempo de vida de un aparato electrónico.

$$\mu : 1/\lambda$$

$$\sigma^2 : 1/\lambda^2$$

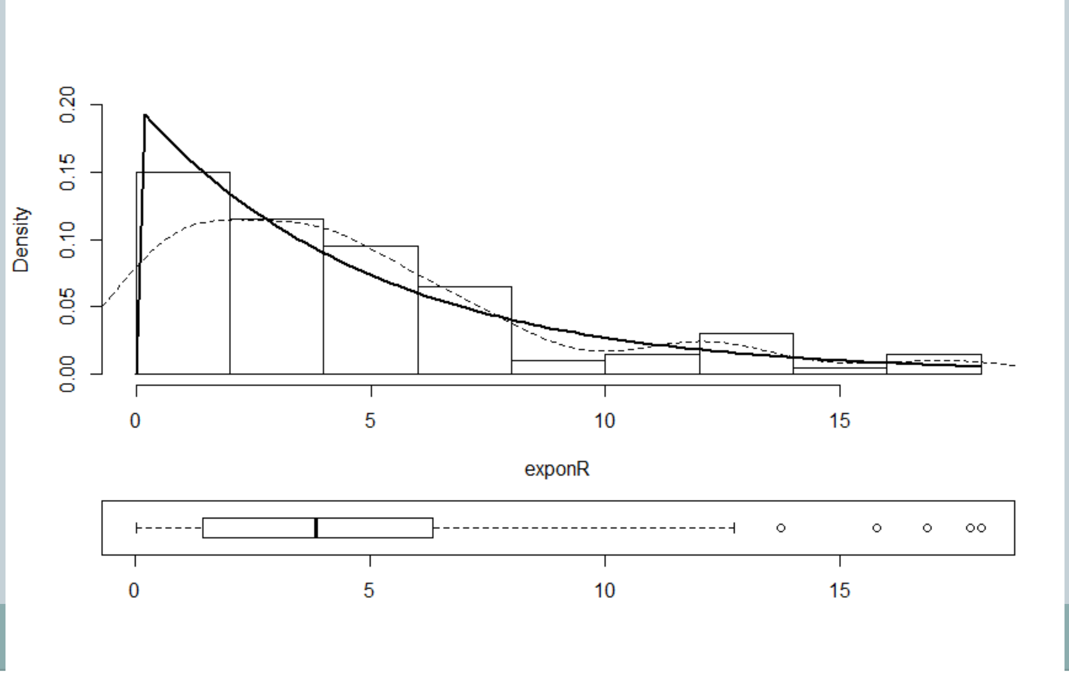
>**d**exp(x, rate=λ) # *pdf*

>**p**exp(x, rate=λ) # *cdf*

>**r**exp(k, rate=λ) # *Muestra aleatoria (simulación)*

Distribución Exponencial


```
>exponR<-rexp(100,rate=1/5);exponR
>par(fig=c(0,1,0,.4) ) #figura con 40% del diagrama
>boxplot(exponR,horizontal=TRUE)
>par(fig=c(0,1,.25,1), new=TRUE)
>hist(exponR,prob=TRUE,main="",ylim=c(0,0.2))
>curve(dexp(x,rate=1/5), lwd=2, add=TRUE); lines(density(exponR),lty=2)
```



Teorema Central del Límite

Sean X_1, X_2,…, X_n, n variables aleatorias independientes y $S_n = X_1+X_2+\ldots+X_n$ con $E(S_n) = \mu$ y $Var(S_n) = \sigma^2$, entonces se verifica que:

$$\frac{S_n - \mu}{\sigma} \approx N(0,1)$$

- En condiciones muy generales, la distribución de la suma de *n* variables aleatorias tiende (cuando *n* es grande) a una Distribución Normal.

$\{X_1, \ldots, X_n\}$ RANDOM SAMPLE

X_i $(i = 1,\ldots,n)$ INDEPENDENT AND IDENTICALLY DISTRIBUTED R.V.

$E(X_i) = \mu$, $VAR(X_i) = \sigma^2$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\xRightarrow{n\to\infty} \quad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad \frac{\bar{X} - \mu}{\sqrt{(\sigma^2/n)}} \sim N(0,1)$$

Teorema Central del Límite

- Ejemplo: X~ **Poiss(λ) Discreta**.
- Tomando *X1, X2,…,Xn* de *k* valores cada uno (todos con la misma media y varianza) para valores grandes de *n*:
  X1+X2+…+Xn=S~N(μ_S , σ_S *)* **Continua.**

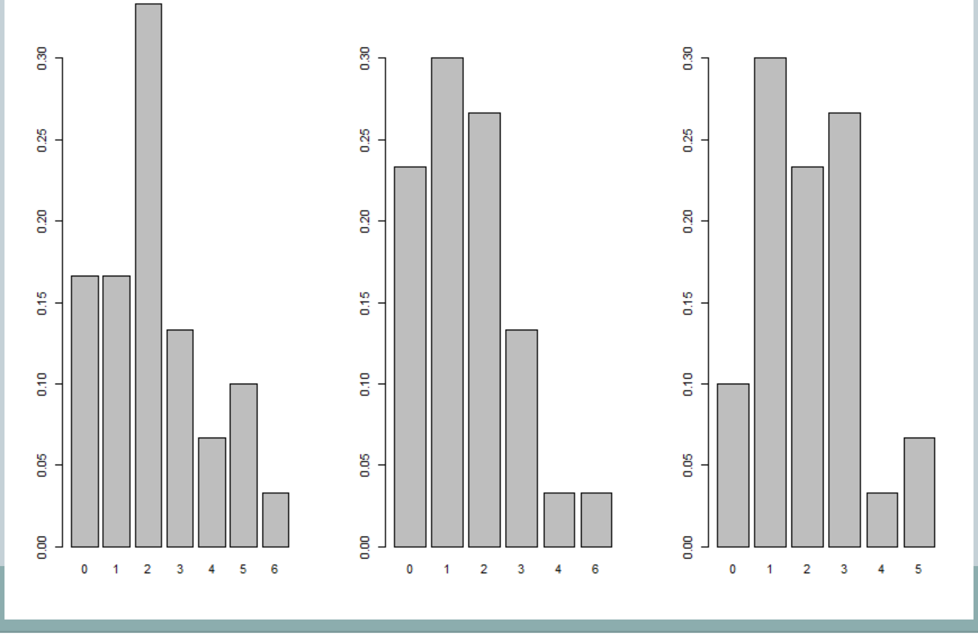
$$\mu_S = n\lambda = \sum_{i=1}^{\lambda} \lambda$$

$$\sigma_S = \sqrt{n\lambda} = \sqrt{\sum_{i=1}^{\lambda} \lambda}$$

$$\frac{S - n\lambda}{\sqrt{n\lambda}} \approx N(0,1)$$

Teorema Central del Límite

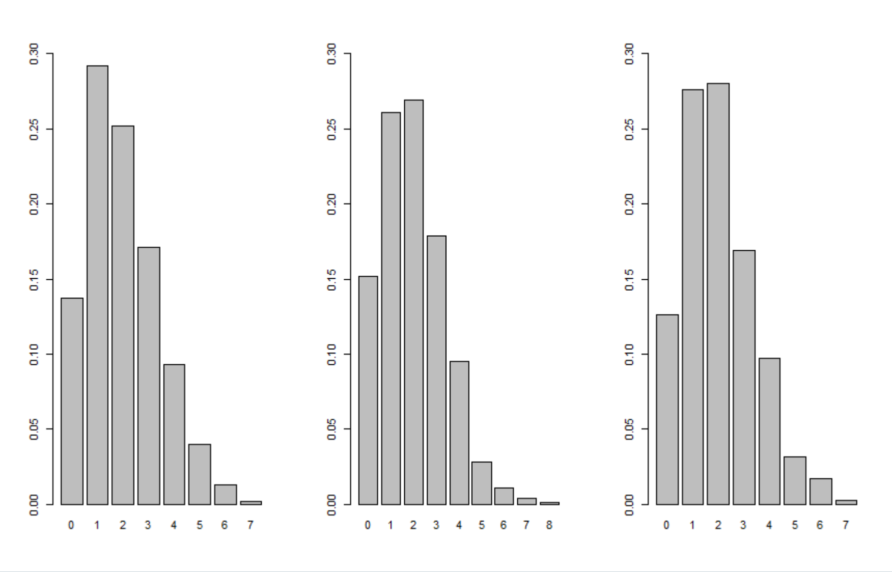

```
>x1<-rpois(30,2);x2<-rpois(30,2);x3<-rpois(30,2);
>par(mfrow=c(1,3))
>barplot(table(x1)/length(x1),ylim=c(0,0.3));
>barplot(table(x2)/length(x2),ylim=c(0,0.3));
>barplot(table(x3)/length(x3),ylim=c(0,0.3))
>S3a<-((x1+x2+x3)-3*2)/sqrt(3*2)
```

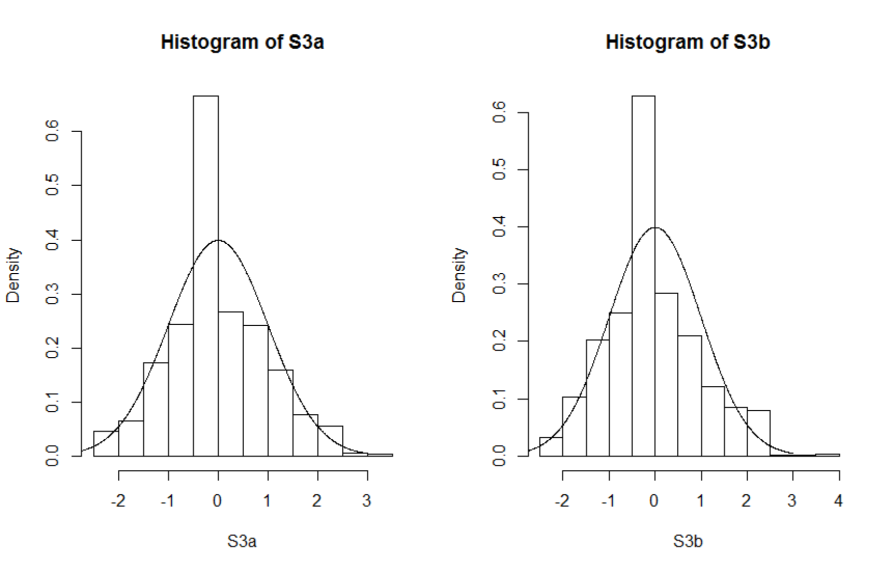


Teorema Central del Límite


```
x1<-rpois(1000,2);x2<-rpois(1000,2);x3<-rpois(1000,2);
barplot(table(x1)/length(x1),ylim=c(0,0.3))
barplot(table(x2)/length(x2),ylim=c(0,0.3))
barplot(table(x3)/length(x3),ylim=c(0,0.3))
S3b<-((x1+x2+x3)-3*2)/sqrt(3*2)
par(mfrow=c(1,2))
points <- seq(-3,3,length=length(S3b))
points.normal <- dnorm(points, mean = 0,sd = 1)
hist(S3a,probability=TRUE);lines(points, points.normal)
hist(S3b,probability=TRUE);lines(points, points.normal)
```

Teorema Central del Límite





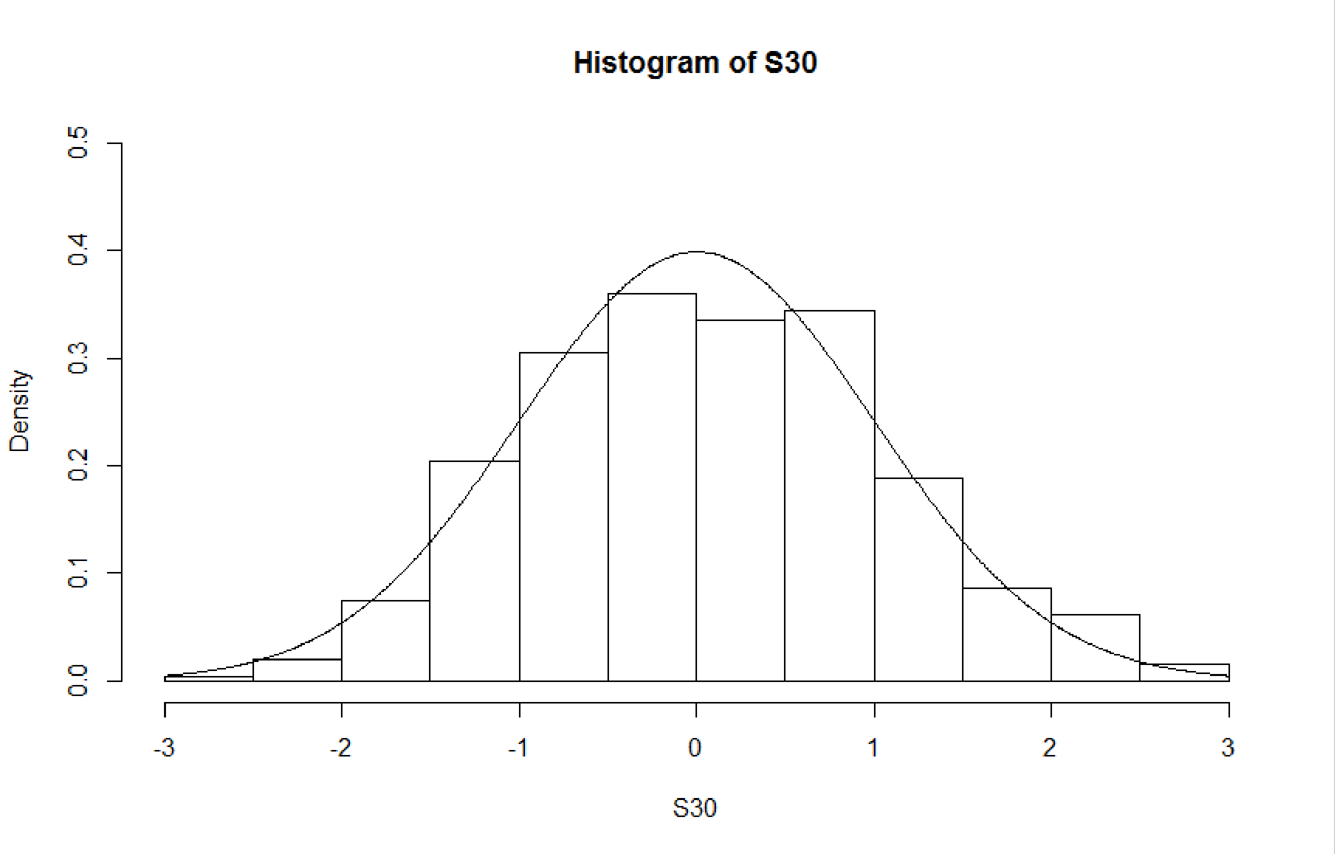
Teorema Central del Límite


```
x4<-rpois(1000,2);x5<-rpois(1000,2);x6<-rpois(1000,2);
...

S30<-((x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+
      x11+x12+x13+x14+x15+x16+x17+x18+x19+x20+
      x21+x22+x23+x24+x25+x26+x27+x28+x29+x30)-30*2)/sqrt(30*2)

par(mfrow=c(1,1))
points <- seq(-3,3,length=length(S30))
points.normal <- dnorm(points, mean = 0,sd = 1)
hist(S30,probability=TRUE, ylim=c(0,0.5))
lines(points, points.normal)
```

Teorema Central del Límite



Problemas resolver con R

Ejercicio 4.2.7 Consideramos que la v. a. T “tiempo de fallo en años” de un cierto componente de un sistema se modela bien por una distribución exponencial con tiempo medio para el fallo de 5 años.

a)Calcular la probabilidad de que un componente dado funcione después de ocho años.
b)Si se instalan cinco de estos componentes en diferentes sistemas, ¿cuál es la probabilidad de que dos funcionen al final de ocho años?

Ejercicio 4.2.14 Una normativa europea obliga a que en los envases de yogur no debe haber menos de 120 gramos. La máquina dosificadora de una empresa láctea hace los envases de yogur según una ley normal de desviación típica de 2 gramos y media 122 gramos. Contestar, razonadamente, las cuestiones:

a) ¿Qué tanto por ciento de los envases de yogur de esta empresa cumplirá la normativa?
b) ¿Cuál deberá de ser la media (μ) de la ley normal con la que la máquina dosificadora hace envases para que el 98% de la producción de yogures de esta empresa cumpla la normativa? (manteniéndose la desviación típica igual a 2gr.)