Repaso

Basado en el curso de Ciencia de Datos con R Fundamentos Estadísticos

por

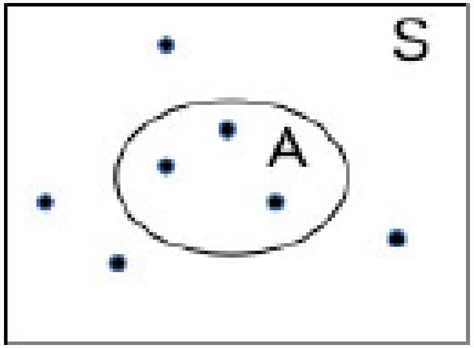
M. Sued, A. Bianco y M. Valdora

Teoría de Probabilidades

1. $\mathcal{S}$ espacio muestral: conjunto con los posibles resultados del experimento.

2. A, B, C: eventos a los cuales vamos a asignarles probabilidad.

3. $\mathbb{P}$: función de probabilidad.



$m = 7$
$m_A = 3$

$$\boxed{\frac{m_A}{m} \to \mathbb{P}(A)}$$

$\mathbb{P}(A)$ *representa el porcentaje de veces que esperamos que* A *ocurra en* **infinitas** *repeticiones*

Función de Probabilidad

$0 \leq \mathbb{P}(A) \leq 1$

La probabilidad del espacio muestral es uno: $\mathbb{P}(\mathcal{S}) = 1$.

Si A_1 y A_2 son eventos disjuntos($A_1 \cap A_2 = \emptyset$), entonces

$$\mathbb{P}(A_1 \cup A_2) \ = \ \mathbb{P}(A_1) + \mathbb{P}(A_2) \ .$$

$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

Dados dos eventos A y B, tenemos que

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Concepto de Independencia

Dado un evento B con $\mathbb{P}(B) > 0$, definimos la probabilidad del evento A dado que B aconteció mediante la formula:

Probabilidad Condicional

$$\boxed{\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}}$$

A y B se dicen independientes si

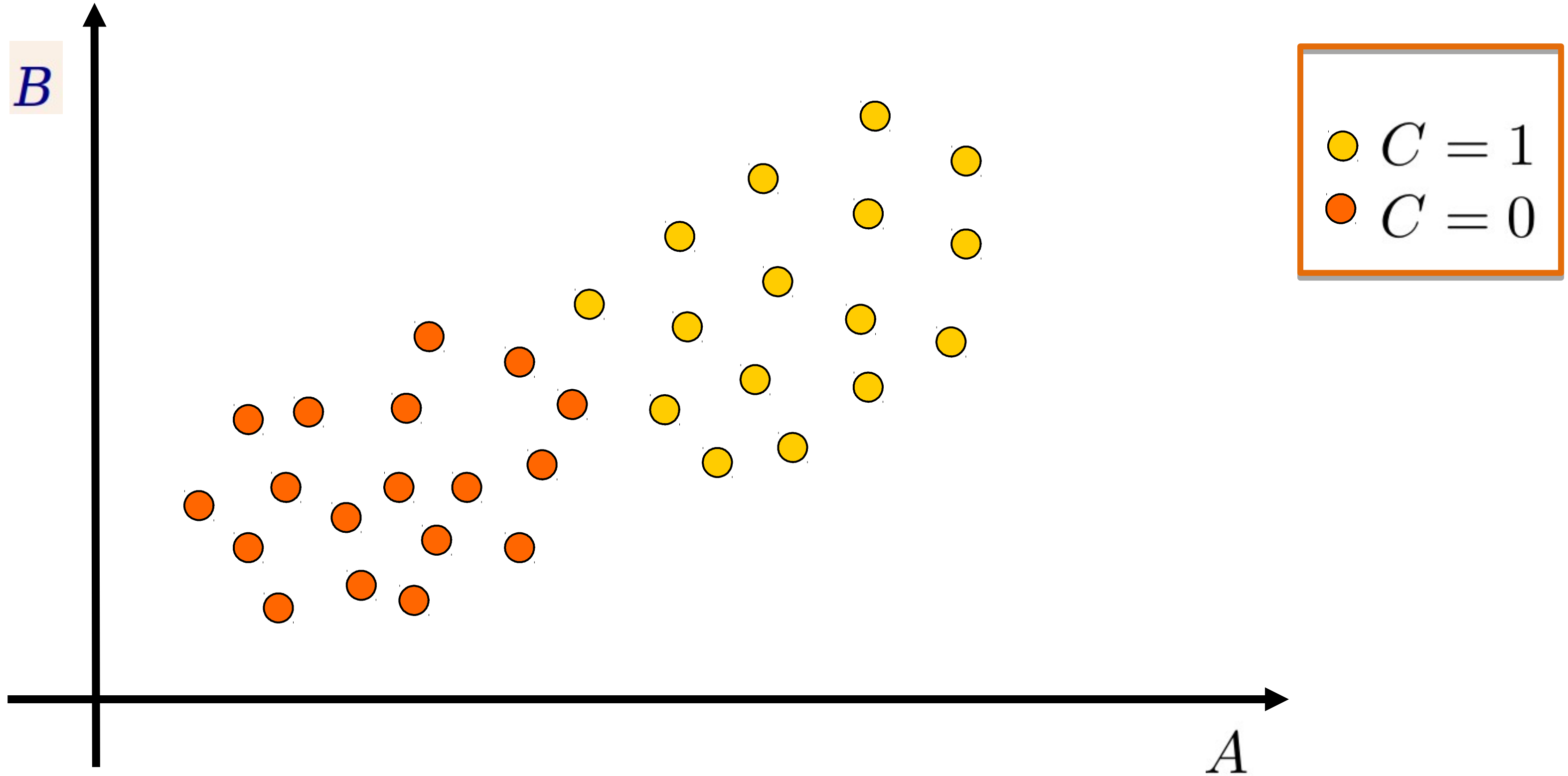
$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)}$$

Lema: Si A y B son independientes y $\mathbb{P}(B) > 0$, entonces

$$\mathbb{P}(A \mid B) = \mathbb{P}(A)$$

Concepto de Independencia Condicional

$$\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\mathbb{P}(B \mid C)$$



Teorema de Bayes

Inversión del Condicional

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B|A) * P(A) + P(B|\bar{A}) * P(\bar{A})}$$

Variables Aleatoeias

Una Variable Aleatoria X es una función definida sobre el espacio muestral que toma valores en los reales:

$$X : \mathcal{S} \to \mathbb{R}$$

Conjunto de valores de la Variable Aleatoria

Evento de Interes

Elementos del Espacio Muestral

$$X \in A = X^{-1}(A) = \{s \in \mathcal{S} : X(s) \in A\}$$

$$\mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}\left(\{\omega \in \mathcal{S} : X(\omega) \in A\}\right)$$

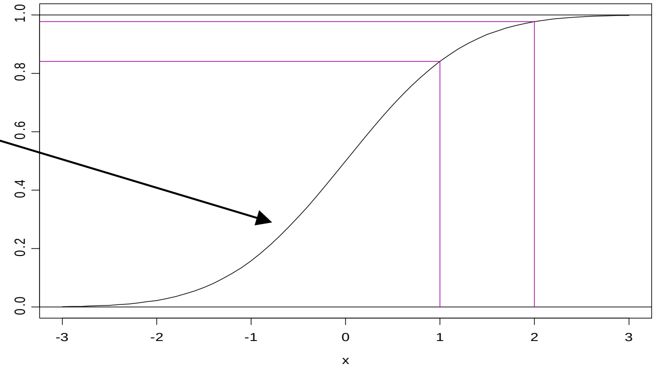
$$\mathbb{P}(X = t) = \mathbb{P}(X^{-1}(t)) = \mathbb{P}\left(\{\omega \in \mathcal{S} : X(\omega) = t\}\right)$$

Valor puntual

La Función de Distribución

La función de distribución acumulada F_X de la variable (aleatoria) X está definida por:

$F_X : \mathbb{R} \to [0, 1] \qquad F_X(t) = \mathbb{P}(X \leq t) = F(t)$



1. $0 \leq F_X(t) \leq 1$ para todo $t \in \mathbb{R}$.
2. F_X creciente: si $s \leq t$, entonces $F_X(s) \leq F_X(t)$.
3. $\lim_{t \to -\infty} F_X(t) = 0$ y $\lim_{t \to +\infty} F_X(t) = 1$.
4. F_X es continua a derecha con límite a izquierda:
5. $\mathbb{P}(X = a)$ es el salto en a de F_X

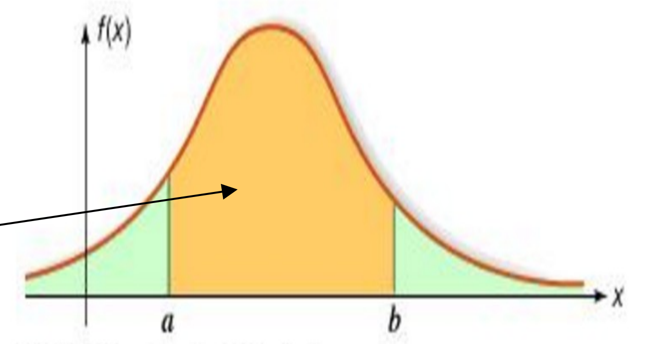
La Función de Densidad

Una variable aleatoria X se dice ($^{(*)}$ absolutamente) continua sii existe una densidad

$$f_X : \mathbb{R} \to \mathbb{R}_{\geq 0}$$

tal que

$$\mathbb{P}(X \in A) = \int_A f_X(u)\, du.$$



En particular,

$$F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^{t} f_X(u)\, du\ .$$

En tal caso, diremos que f_X es la función de densidad de la variable aleatoria X.

$f : \mathbb{R} \to \mathbb{R}$ se dice densidad si

- $f(u) \geq 0$ para todo $u \in \mathbb{R}$
- $\int_{-\infty}^{+\infty} f(u)\, du = 1$

La Normal Univariada

- Z normal estandart si

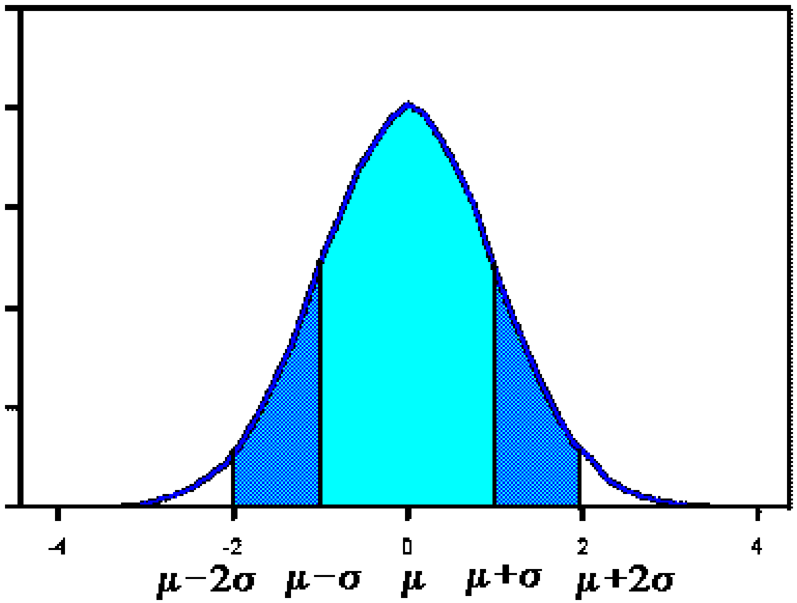
$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

- f_Z simétrica en el origen: $f_Z(z) = f_Z(-z)$
- Siendo f_Z simétrica, tenemos que $F_Z(-u) = 1 - F_Z(u)$
- $F_Z(z) = \int_{-\infty}^{z} f_Z(u)du$ no se puede calcular.
- Hay tabla con valores de $F_Z(u)$ para $u > 0$.
- $\phi(z) = F_Z(z)$ se llama función phi.

- Z normal estandar, Sea $X := \sigma Z + \mu$

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

- $F_X(x) = \phi\left((x-\mu)/\sigma\right)$
- $X \sim \mathcal{N}(\mu, \sigma^2)$



Esperanza y Varianza

Dada una variable aleatoria discreta X con $\mathbb{R}g(X) = \{x_1, x_2, \cdots\}$ y función de probabilidad puntual $p_X(x_i)$, definimos la esperanza de X mediante la fórmula,

$$\mathbb{E}[X] \ = \sum_{i=i}^{\infty} x_i \, p_X(x_i) \ ,$$

Dada una variable aleatoria continua X con función de densidad f_X, definimos la esperanza de X como

$$\mathbb{E}[X] \ = \int_{-\infty}^{\infty} u \, f_X(u) \, du \ .$$

X v.a.discreta con $\mathbb{E}[X] = \mu$. La varianza de X, está definida mediante la fórmula

$$\mathbb{V}(X) = \mathbb{E}\Big[(X - \mu)^2\Big] \ .$$

Covarianza y Correlación

- Medidas de asociación lineal entre variables (x e y)

Covarianza Empírica

$$cov\left(x,y\right)=\sum_{i=1}^{n}\frac{1}{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)$$

Correlación

Medias

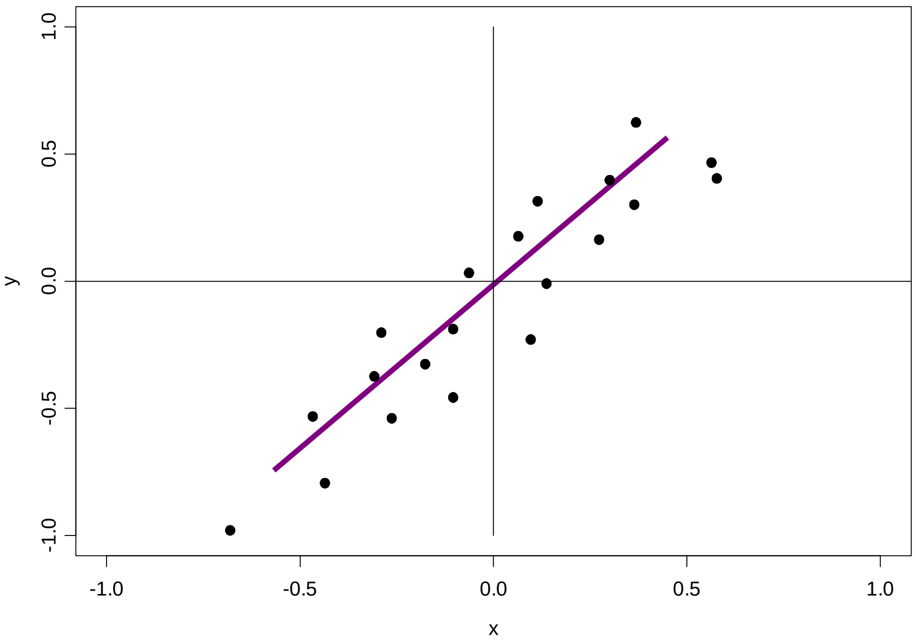
$$\rho\left(x,y\right)=cor\left(x,y\right)=\frac{cov(x,y)}{\sigma(x)\sigma(y)}$$

$$-1\leq\rho(x,y)\leq 1$$

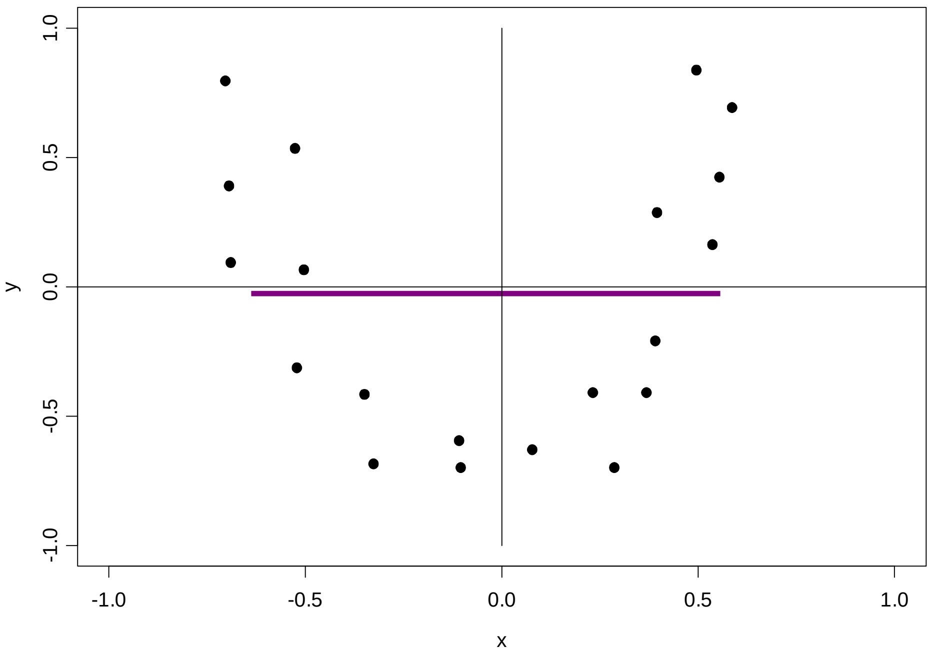
Desvios

Asociación Vs. Correlación

Cor(x,y) = 0.9



Cor(x,y) = 0.05



Tchevichev

Sea W una v.a. con media μ_W y $V[W] = \sigma_W^2$.Luego, $\epsilon > 0$

$$\boxed{\mathbb{P}\left(|W - \mu_W| \geq \varepsilon\right) \leq \frac{\sigma_W^2}{\varepsilon^2}}$$

Ley de los Grandes Números

Sean $(X_i)_{i\geq 1}$ i.i.d., con $E[X_i] = \mu$ y $V[X_i] = \sigma^2$, para todo i.
Entonces, el promedio converge a μ en probabilidad:

$$\lim_{n\to\infty} \mathbb{P}((|\bar{X}_n - \mu| > \varepsilon) = 0$$

Promedio

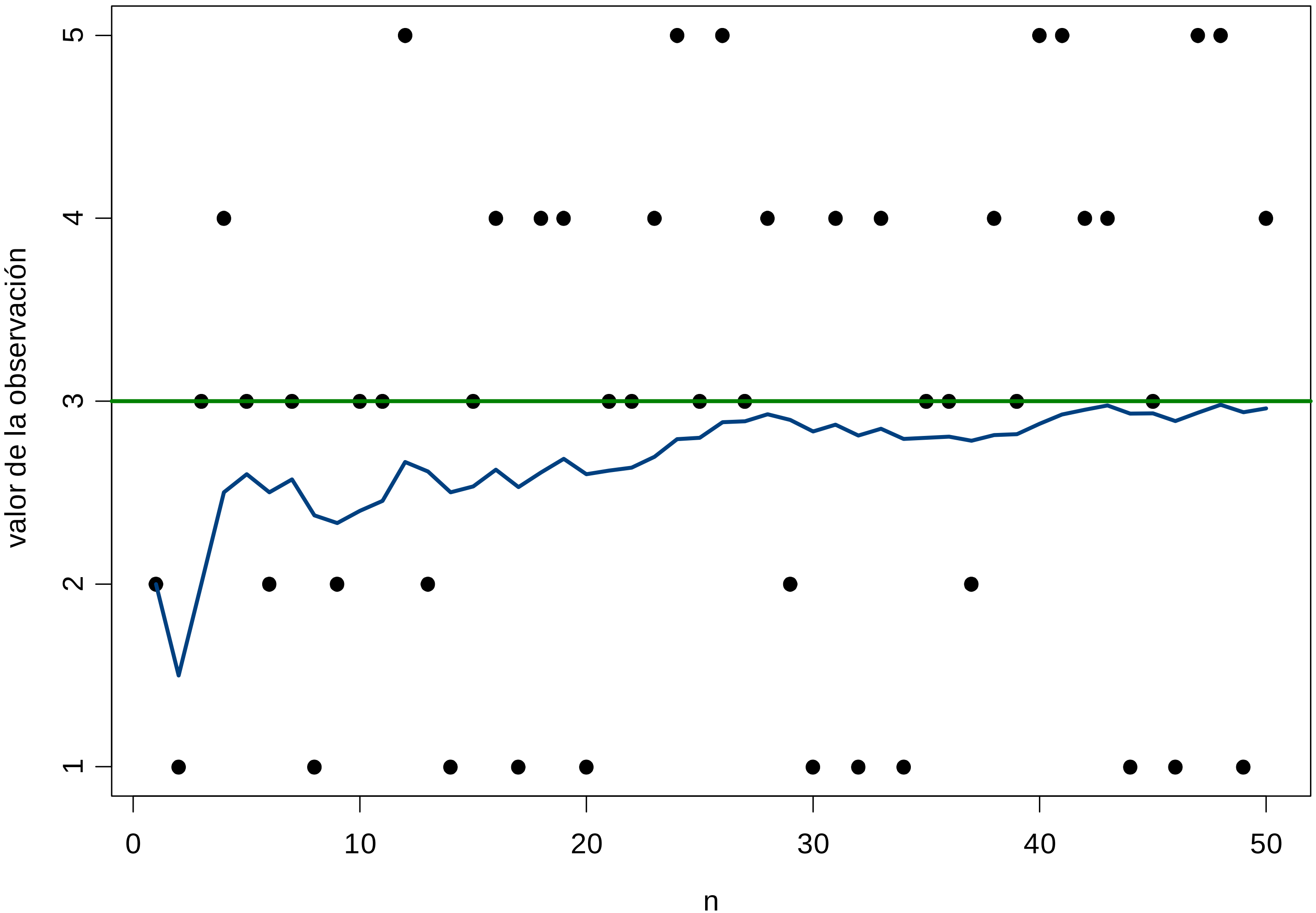
Es decir, para todo $\varepsilon > 0$, vale que

$$\boxed{\bar{X}_n \to \mu \text{ en probabilidad}}$$

Ejemplo de Ley de los Grandes Números

Valores de la Variable

Val	Prob
1	0.2
2	0.2
3	0.2
4	0.2
5	0.2



Teorema Central del Límite

Variables Independiantes e Identicamente Distribuidas

Sean $(X_i)_{i\geq 1}$ v.a. i.i.d. con $E[X] = \mu$ y $V[X] = \sigma^2$, entonces tenemos que

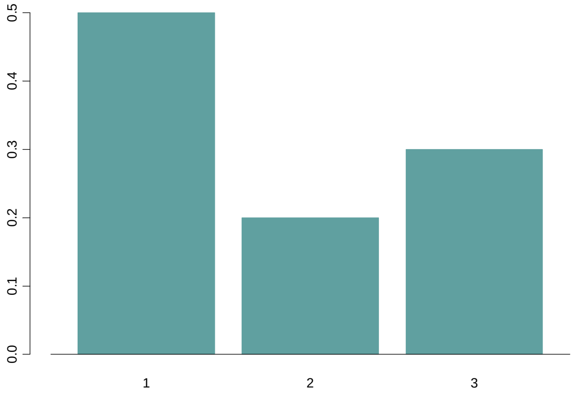
$$\mathbb{P}\left(\frac{\bar{X}_n - E(X_1)}{\sqrt{Var(\bar{X}_n)}} \leq x\right) \underset{n\to\infty}{\to} \phi(x)\ \forall x \in \mathbb{R}\ ,$$

Promedio de Variables Aleatorias

Función de Distribución Normal

Ejemplo del Teorema Central

X_i	$P(x_i)$
1	0.5
2	0.2
3	0.3



$(X_1+x_2)/2$	$P((X_1+x_2)/2)$
2/2	0.5*0.5
3/2	0.5*0.2 + 0.2*0.5
4/2	0.5*0.3 + 0.2*0.2 + 0.3*0.5
5/2	0.2*0.3 + 0.3*0.2
6/2	0.3*0.3

Distribución de $(x_1+x_2+\ldots+x_{20})/20$