# PRELIMINARES

- **Teorema de la probabilidad compuesta.**
  Sean $A_1, \ldots, A_{n-1}$ tales que $P\left(\bigcap_{i=1}^{n-1} A_i\right) > 0$, para todo $A_n \in \mathcal{A}$, entonces se verifica la siguiente identidad:

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2)\ldots P(A_n/A_1 \cap \cdots \cap A_{n-1}).$$

- **Teorema de la probabilidad total.**
  Sea $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ una secuencia de sucesos que define una partición de $\Omega$, siendo $\underbrace{P(A_i) > 0}_{\text{suceso posible}}$, para $i \in \mathbb{N}$, entonces

$$P(B) = \sum_{i=1}^{\infty} P(B/A_i)P(A_i).$$

Recuperamos la probabilidad de B a partir de las prob. condicionadas

- **Teorema de Bayes.**
  Sea $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ una secuencia de sucesos que define una partición de $\Omega$, siendo $P(A_i) > 0$, para $i \in \mathbb{N}$,

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{\underbrace{\sum_{i=1}^{\infty} P(B/A_i)P(A_i)}_{=P(B)}}, \quad \forall i \in \mathbb{N}.$$

## Variable aleatoria

Probar que es v.a. (teoría)

Una **variable aleatoria** sobre un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ es una función medible $X : \Omega \longrightarrow \mathbb{R}$, i.e.,

$$\boxed{X^{-1}(B) \in \mathcal{A}, \quad B \in \mathcal{B}.}$$

- **Teorema de Markov.**

$$\boxed{X \geq 0,\ \exists E[X] \Rightarrow P(X \geq \varepsilon) \leq \frac{E[X]}{\varepsilon}, \quad \varepsilon > 0.}$$

- **Desigualdad de Chebychev.**

$$\boxed{\exists E[X^2] \Rightarrow P(|X - E[X]| \geq k) \leq \frac{\mathrm{Var}[X]}{\underbrace{k^2}_{\text{convergente}}}, \quad k > 0.}$$

## Distribución Degenerada

un sólo valor tiene prob 1 y el resto 0

Una v.a. $X : \Omega \longrightarrow \mathbb{R}$, tal que $P(X = c) = 1$, para un cierto $c \in \mathbb{R}$, se dice que sigue una **Distribución Degenerada** en un punto. Denotamos

$$\boxed{X \sim D(c)}$$

.

- **Función de distribución**: $F_X$ tal que $F_X(c^-) = 0$, y $F_X(c) = F_X(x) = 1$, para cualquier $x > c$.
- **Función generatriz de momentos**: $M_X(t) = e^{tc}$, $t \in \mathbb{R}$. $M_X(t)=E[e^{tx}]=e^{tc} \cdot P_X(c)=e^{tc}$ $t\in\mathbb{R}$ (derivar)
- **Momentos**: $m_k = c^k$ y $\mu_k = 0$, $\forall k \in \mathbb{N}$.
- **Caracterización**: Una variable X es degenerada si y sólo si $\text{Var}(X) = 0$.

$\text{Var}(X) = E[X^2] - (E[X])^2 = c^2 - c^2 = 0$

$E[X] = \frac{d\,e^{tc}}{dt}\Big|_{t=0} = c e^{tc}\Big|_{t=0} = c$

$E[X^2] = \frac{d^2 e^{tc}}{dt^2}\Big|_{t=0} = c^2 e^{tc}\Big|_{t=0} = c^2$

## Distribución Uniforme Discreta

n valores con la misma probabilidad todos

Una v.a. $X : \Omega \longrightarrow \mathbb{R}$, tal que $P(X = x_i) = \frac{1}{n}$, $i = 1, \ldots, n$, se dice que sigue una **Distribución Uniforme** en $n$ puntos. Denotamos

$$\boxed{X \sim \mathcal{U}(x_1, \ldots, x_n),\ x_i \in \mathbb{R}}$$

- **Función de distribución**:

$$F_X(x) = \begin{cases} 0 & x < x_1 \\ \frac{i-1}{n} & x_{i-1} \le x < x_i,\ i = 2, \ldots, n \\ 1 & x \ge x_n \end{cases}$$

- **Función generatriz de momentos**: $M_X(t) = \frac{\sum_{i=1}^n e^{tx_i}}{n}$, $t \in \mathbb{R}$.
- **Momentos**: $m_k = \frac{\sum_{i=1}^n x_i^k}{n}$ y $\mu_k = \frac{\sum_{i=1}^n (x_i - m_1)^k}{n}$, $\forall k \in \mathbb{N}$.
- **Esperanza matemática de una función medible**: $E[g(X)] = \frac{1}{n}\sum_{i=1}^n g(x_i)$.
- **Casos especiales** $E[X] = \overline{X}$, y $\text{Var}(X) = S^2$, denotando, como es usual, por $S^2$ la varianza muestral.

## Distribución de Benoulli

sólo hay éxito/fracaso

Una v.a. $X : \Omega \longrightarrow \mathbb{R}$, se dice que sigue una **Distribución Bernoulli** de parámetro $p$ si y solo toma los valores 0 y 1 con probabilidades $1 - p$ y $p$ respectivamente. **Denotamos**

$$\boxed{X \sim B(1, p), 0 < p < 1}$$

- **Función Masa de Probabilidad**: $P(X = x) = p^x(1 - p)^{1-x}$, $\quad x = 0, 1$. easy
- **Función de Distribución**: $F_X(x) = 0$, para $x < 0$, $F_X(x) = 1 - p$, para $0 \le x < 1$, y $F_X(x) = 1$, para $x \ge 1$.
- **Función Generatriz de Momentos**: $M_X(t) = pe^t + (1 - p)$, $t \in \mathbb{R}$.
- **Momentos**: $m_k = p$, y $\mu_k = (1 - p)^k p + (-p)^k(1 - p)$, para $k \in \mathbb{N}$.
- **Media**: $E[X] = p$.
- **Varianza**: $\text{Var}(X) = p(1 - p)$.

## Distribución Binomial

Una variable aleatoria (discreta), $X : \Omega \longrightarrow \mathbb{R}$, asociada al **número de éxitos de un experimento aleatorio que se repite un número finito de veces con probabilidad de éxito, $p$, constante** se dice que sigue una **Distribución Binomial** de parámetros $n$ y $p$, $0 < p < 1$. **Denotamos**

$$\boxed{X \sim B(n, p), n \in \mathbb{N}}$$

ejemplo: lanzo una moneda no trucada 20 veces, ¿P salgan 7 caras? $n=20$, éxito = salir cara $\to p=1/2$; $X \to B(20, 1/2)$, $P[X=7]$?

.

- **Función Masa de Probabilidad**:

$$P(X = x) = \underbrace{\left[\frac{n!}{x!(n - x)!}\right]}_{\binom{n}{x}} p^x(1 - p)^{n-x};\ x = 0, 1, 2, \ldots, n.$$

- **Función Generatriz de Momentos**: $M_X(t) = (pe^t + (1 - p))^n$, $t \in \mathbb{R}$.
- $E[X] = np$. igual que antes elevado a n
- $\text{Var}(X) = np(1 - p)$.
- Si consideramos que $X$ está asociada **al número de fracasos**, entonces la distribución de probabilidad es, $X \sim B(n, 1 - p)$ $\quad n \in \mathbb{N}$.

## Distribución de Poisson

Una variable aleatoria (discreta), asociada a la **ocurrencia de un número determinado de eventos durante cierto periodo de tiempo** (con probabilidad de ocurrencia pequeña) se dice que sigue una **Distribución de Poisson** de parámetro $\lambda$. **Denotamos**

$$\boxed{X \sim \mathcal{P}(\lambda), (\lambda > 0)}$$

También es conocida como **Ley de los sucesos raros**.

- $\lambda$ expresa la frecuencia media de ocurrencia de un evento, esto es, el número de veces que se espera que ocurra el evento durante un intervalo de tiempo dado.
- **Función Masa de Probabilidad**: $P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$ (que no puede tomar valores negativos!)
- **Aproximación mediante el modelo Binomial**: Cuando el número $n$ de pruebas de Bernoulli tiende a ser infinito para un suceso con escasa probabilidad de ocurrencia, la distribución Binomial aproxima a una distribución de Poisson (ley de los sucesos raros). una Poisson en el límite puede entenderse como una Binomial de sucesos infinitos
- **Función Generatriz de Momentos**: $M_X(t) = \exp(\lambda(e^t - 1))$, $t \in \mathbb{R}$.

  $M_X(t) = E[e^{tX}] = \sum_{x\in\mathbb{N}} e^{tx} p(x) = \sum_{x\in\mathbb{N}} e^{tx} e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\left(\sum_{x\in\mathbb{N}} e^{tx}\frac{\lambda^x}{x!}\right) = e^{-\lambda}\left(\sum_{x\in\mathbb{N}}\frac{(e^t\lambda)^x}{x!}\right) = e^{-\lambda} e^{e^t\lambda} = e^{\lambda(e^t-1)}$

- **Media y Varianza**: $E[X] = \text{Var}(X) = \lambda$. única dist. en la que estos valores coinciden

## Distribución Geométrica

Una variable aleatoria, asociada al **número de fracasos, en sucesivas repeticiones independientes de una prueba de Bernoulli con probabilidad de éxito $p$, antes de que ocurra el primer éxito** se dice que sigue una **Distribución Geométrica** de parámetro $p$, $(0 < p < 1)$.
**Denotamos**

$$\boxed{X \sim \mathcal{G}(p), (0 < p < 1)}$$

- F masa: $P(X=x) = \binom{x}{x}(1-p)^x p^1 = p(1-p)^x$
- **Función de Distribución**: $F_X(x) = 0$, para $x < 0$, y, $F_X(x) = 1 - (1 - p)^{x+1}$ para $x \ge 0$. $P(X=x) = F_X(x) - F_X(x-1)$
- **Función Generatriz de Momentos**: $M_X(t) = \frac{p}{1-(1-p)e^t}$, $\quad t < -\ln(1 - p)$. $\frac{p \cdot \text{exit}}{1 - p\,\text{frac} \cdot e^t}$
- $E[X] = \frac{1-p}{p}$.
- $\text{Var}(X) = \frac{1-p}{p^2}$.
- **Propiedad de falta de memoria**:

$$P(X \le h + k / X \le h) = P(X \le k) \quad \forall h, k \in \mathbb{N} \cup 0.$$

... me da igual el inicio, sólo tengo que calcular los n primeros; si ya sé que han ocurrido h fracasos...

- **Caracterización**: Es la única distribución con valores enteros positivos que verifica la propiedad de falta de memoria.

## Distribución Binomial Negativa

Una variable aleatoria, asociada al **número de fracasos, en sucesivas repeticiones independientes de una prueba de Bernoulli con probabilidad de éxito $p$, antes de que ocurra el $k$-éximo éxito** se dice que sigue una **Distribución Binomial Negativa** de parámetros $n$ y $p$, $(0 < p < 1)$. **Denotamos**

$$\boxed{X \sim \mathcal{BN}(k, p),\ k \in \mathbb{N},\ (0 < p < 1)}$$

.

- La v.a $X + k$ sigue una **Distribución de Pascal o Tiempo de Espera**. Indica el número de pruebas (en total) hasta la $k$-ésimo éxito.
- **Función Masa de Probabilidad**:

$$P(X = x) = \underbrace{\left[\frac{(x + k - 1)!}{x!(k - 1)!}\right]}_{\binom{x+k-1}{x}}(1 - p)^x p^k, \quad x \in \mathbb{N}$$

- **Función Generatriz de Momentos**: $M_X(t) = \left[\frac{p}{1-(1-p)e^t}\right]^k$, $\quad t < -\ln(1 - p)$. igual que antes pero elevado a k
- $E[X] = \frac{k(1-p)}{p}$. igual que antes pero mult · k
- $\text{Var}(X) = \frac{k(1-p)}{p^2}$.

¿pq es interesante dar estos dos momentos? Sirven para comparar una variable consigo misma. E sirve como promedio y Var el nivel de fiabilidad (cómo de dispersos están los datos) (a más varianza, más dispersión).

## Distribución Hipergeométrica

va variando la prob de éxito

Si se considera una población de tamaño $N$ y una subpoblación de tamaño $N_1$, y se extrae una muestra aleatoria de tamaño $n$, sin reemplazamiento o simultáneamente, una v.a. $X$ que describe **el número de individuos de la muestra que pertenecen a la subpoblación** se dice que sigue una **Distribución Hipergeométrica** de parámetros $N$, $N_1$ y $n$. **Denotamos**

$$\boxed{X \sim H(N, N_1, n);\ N, N_1, n \in \mathbb{N},\ N_1, n \le N}$$

EJEMPLO: Población: Peces $N=5$; Subpob.: Peces verdes $N_1=3$

.

- **Función Masa de Probabilidad**: $\frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}}$

$$P(X = x) = \frac{(N_1!/x!(N_1 - x)!)\,((N - N_1)!/(n - x)!(N - N_1 - n + x)!)}{N!/n!(N - n)!}$$

$$x = 0, \ldots, n, \quad x \le N_1,\ n - x \le N - N_1.$$

- $E[X] = n\frac{N_1}{N}$. $\frac{nN_1}{N}$: $N_1 \to$ hijos, $N \to$ madre
- $\text{Var}(X) = n\frac{N_1}{N}\left(1 - \frac{N_1}{N}\right)\frac{N-n}{N-1}$.
- Aproxima a una distribución Binomial $\mathcal{B}(n, p)$ cuando $N, N_1 \to \infty$, con $N_1/N \to p$, y siendo $n$ el parámetro de la Hipergeométrica, que define el número de pruebas de Bernoulli en el modelo Binomial que aproxima.

13, 17,

# TEMA 1

## Definición

La **distribución uniforme continua** sobre un intervalo $(a, b)$, se caracteriza por **tomar valores en dicho intervalo y tener una densidad de probabilidad constante**.
En concreto,

$$X \sim \mathcal{U}(a, b),\ a < b,\ \Leftrightarrow f_X(x) = \frac{1}{b-a},\quad \forall x \in (a, b).$$

RECUERDA: esto no es una probabilidad

### Función de distribución

Para cualquier $x \in (a, b)$, se tiene

$$F_X(x) = \int_{a}^{x} f_X(y)dy = \int_a^x \frac{1}{b-a}dy = \frac{x}{b-a} - \frac{a}{b-a} = \frac{x-a}{b-a}.$$

$-\infty$ (antes de $a$ es 0)

→ longitud intervalo

→ valor cte

Para $x < a$, $F_X(x) = 0$ y para $x \geq b$, $F_X(x) = 1$, según se deriva de la definición de función de distribución de una variable aleatoria continua (ver capítulo de Preliminares).

### Función generatriz de momentos

$$M_X(t) = E[e^{tX}] = \int_a^b \frac{e^{tx}}{b-a}dx = \frac{e^{tb}-e^{ta}}{t(b-a)},\quad t \neq 0;\quad M_X(t) = 1, t = 0.$$

$e^{tX}$ $f_X$

## Distribución Normal $X \sim \mathcal{N}(\mu, \sigma^2); \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}_+$

Se dice que una variable aleatoria (v.a.) continua $X$ se **distribuye según una normal** de parámetros $\mu$ y $\sigma^2$ si su función de densidad $f_X$ viene dada por la siguiente expresión:

media varianza

$$f_X(x) \quad = \quad \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad x \in \mathbb{R}$$

Mediante el procedimiento de **tipificación** se tiene que,

$$X \sim \mathcal{N}(\mu, \sigma^2) \Leftrightarrow \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$$

### Proposición

(i) $X \sim \mathcal{N}(0, 1) \ \Leftrightarrow\ M_X(t) = E[e^{tX}] = e^{\frac{t^2}{2}}$. → lo definimos en todo $\mathbb{R}$

(ii) $X \sim \mathcal{N}(\mu, \sigma^2) \ \Leftrightarrow\ M_X(t) = E[e^{tX}] = e^{t\mu + \frac{t^2\sigma^2}{2}}$.

Poisson y Binomial → corrección por cont.

## Definición

Una v.a $X$ se dice que sigue una **distribución exponencial** si:

se usa para obsolescencia programada, por ejemplo

$$X \sim \exp(\lambda) \Leftrightarrow f_X(x) = \lambda e^{-\lambda x},\ \forall x \geq 0.$$

### Función de distribución

$$F_X(x) = \int_0^x \lambda e^{-\lambda x}dx = 1 - e^{-\lambda x},\quad x \geq 0.$$

### Función generatriz de momentos

$$\begin{aligned} M_X(t) &= E[e^{tX}] = \int_0^\infty \lambda e^{x(t-\lambda)}dx = \left[\frac{\lambda e^{x(t-\lambda)}}{t-\lambda}\right]_{x=0}^{x=\infty} \\ &= -\frac{\lambda}{t-\lambda} = \left(1 - \frac{t}{\lambda}\right)^{-1},\quad t < \lambda. \end{aligned}$$

## Definición

Indica el **tiempo aleatorio transcurrido hasta la ocurrencia del $n$-ésimo suceso** aleatorio, i.e., $X_i \sim \exp(\lambda)$, $i = 1, \ldots, n$, mutuamente independientes,

$$\sum_{i=1}^n X_i = E \sim \mathcal{E}(n, \lambda).$$

- cuando $n=1$, $\mathcal{E} = \exp$
- cuando $n \in \mathbb{R}$, $\mathcal{E} = \Gamma$

Para $n \in \mathbb{N} - \{0\}$, y $\lambda > 0$, la densidad de probabilidad $f_X$ de la v.a. $E$ se define como sigue

$$f_E(x) = \frac{\lambda^n}{\Gamma(n)}x^{n-1}e^{-\lambda x},\quad \Gamma(n) = (n-1)!\quad x \in \mathbb{R}_+.$$

### Función generatriz de momentos

$$M_E(t) = \left[\left(1 - \frac{t}{\lambda}\right)\right]^{-n},\quad t < \lambda.$$

## Función gamma, $\Gamma(u)$

Previamente al estudio de la distribución gamma, se define la **función gamma**, representada por $\Gamma(u)$ como:

lo vimos en AMII

$$\Gamma(u) = \int_0^\infty x^{u-1}e^{-x}dx,\quad u > 0.$$

## Distribución Gamma $X \sim \Gamma(u, \lambda);\ u, \lambda \in \mathbb{R}_+$

Es una **extensión de la distribución de Erlang**, definida cuando el parámetro $n$ no es un número natural sino que es un número real. Es decir, se obtiene ampliando el espacio paramétrico de la distribución de Erlang.

$$X \sim \Gamma(u, \lambda),\ u, \lambda \in \mathbb{R}_+ \Leftrightarrow f_X(x) = \frac{\lambda^u}{\Gamma(u)}x^{u-1}e^{-\lambda x},\quad x > 0.$$

### Función generatriz de momentos

$$M_X(t) = \left[\left(1 - \frac{t}{\lambda}\right)\right]^{-u},\quad t < \lambda.$$

ACTIVIDAD VOLUNTARIA (hacer)

## Función Beta, $\beta(p, q)$

Previamente al estudio de la distribución beta, se define la **función beta**, representada por $\beta(p, q)$ como:

$$\beta(p, q) = \int_0^1 x^{p-1}(1-x)^{q-1}dx.$$

## Distribución Beta $X \sim \beta(p, q);\ p, q > 0$

Una v.a. $X$ sigue una **distribución beta** de parámetros $p$ y $q$ si:

$$X \sim \beta(p, q) \ \Leftrightarrow\ f_X(x) = \frac{1}{\beta(p, q)}x^{p-1}(1-x)^{q-1},\quad x \in (0, 1),\quad p, q > 0.$$

# TEMA 2

**Teorema de medibilidad.**
$X = (X_1, \ldots, X_n)$ es un vector aleatorio sobre $(\Omega, \mathcal{A}, P)$ si y sólo si $X_i$ es una variable aleatoria sobre $(\Omega, \mathcal{A}, P)$, para $i = 1, \ldots, n$.

(vi) Sean $\varepsilon_1, \varepsilon_2 > 0$

$$F(x+\varepsilon_1, y+\varepsilon_2) - F(x, y+\varepsilon_2) - F(x+\varepsilon_1, y) + F(x,y) \overset{?}{\geq} 0$$

Buscamos un punto contraejemplo que no lo verifique. Para ello, sabiendo cómo es F, necesitamos que los tres primeros sumandos sean 1 y el último se anule (pues $a+\varepsilon > a$, $\varepsilon > 0$). Fijamos $\varepsilon_1 = 2$, $\varepsilon_2 = 1$, $\in \mathbb{R}^+$. Tomamos $x = -1$, $y = 0$. Así:

$F(-1+2, 0+1) = 1$ $\quad$ $F(-1, 0+1) = 1$
$F(-1, 0) = 0$ $\quad$ $F(-1+2, 0) = 1$ $\Big\}$ $(-1, 0)$ sirve como contraejemplo

Concluimos que F no es función de distribución ✓

## Caracterización de vector aleatorio discreto

Un vector aleatorio $X = (X_1, \ldots, X_n) : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B}^n)$ **es discreto** si y sólo si, para $i = 1, \ldots, n$, $X_i : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ es **una variable aleatoria discreta unidimensional**.

### Demostración ($\Rightarrow$)

Supongamos que X es un vector aleatorio discreto, i.e., $\exists E_X \subset \mathbb{R}^n$ tal que $P(X \in E_X) = 1$. Definamos el subconjunto

$$\Psi^i_{\mathbb{R}}(E_X) = \{y \in \mathbb{R};\ \exists x \in E_X,\ x = (x_1, \ldots, x_i = y, \ldots, x_n)\}.$$

Dicho conjunto representa la proyección de $E_X$ sobre la $i$-ésima componente. Dado que $E_X$ es numerable, $\Psi^i_{\mathbb{R}}(E_X)$ es también numerable. Adicionalmente,

no numerable $\quad$ numerable

$$\begin{aligned} P(X_i \in \Psi^i_{\mathbb{R}}(E_X)) &= P\left(X_1 \in \mathbb{R}, \ldots, X_i \in \Psi^i_{\mathbb{R}}(E_X), \ldots, X_n \in \mathbb{R}\right) \\ &\geq P(X \in E_X) = 1. \end{aligned}$$

¿pq? Porque $E_X \subseteq \mathbb{R}^n$

como es una probabilidad, no puede ser >1, luego =1 ⇒ $X_i$ es v.a. discreta unidim.

Por tanto, $X_i$ es una variable aleatoria unidimensional discreta, para $i = 1, \ldots, n$.

### Demostración ($\Leftarrow$)

Supongamos ahora que $X_i$ es una variable aleatoria unidimensional discreta, para $i = 1, \ldots, n$, i.e., existe un subconjunto numerable $E_{X_i} \subset \mathbb{R}$, tal que $P(X_i \in E_{X_i}) = 1$, para $i = 1, \ldots, n$. Por tanto el subconjunto de $\mathbb{R}^n$ dado por $E_{X_1} \times \cdots \times E_{X_n}$ es también numerable. Se tiene entonces,

$$\begin{aligned} P\left(X = (X_1, \ldots, X_n) \in E_{X_1} \times \cdots \times E_{X_n}\right) &= P(X_1 \in E_{X_1}, \ldots, X_n \in E_{X_n}) \\ = P\left(\bigcap_{i=1}^n \{X_i \in E_{X_i}\}\right) &\geq 1 - \sum_{i=1}^n P\left(\{X_i \in \overbrace{\mathbb{R} \setminus E_{X_i}}^{E_{X_i}^c}\}\right) \\ &= 1 - \sum_{i=1}^n \overbrace{P(X_i \in E^c_{X_i})}^{=0} = 1 \end{aligned}$$

desigualdad de Boole

## Caracterización de vector aleatorio continuo

Sea $X = (X_1, \ldots, X_n) : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B}^n, P_X)$ es un vector aleatorio continuo. **Entonces,** $X_i : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B}^n, P_{X_i})$ es una variable aleatoria continua, para $i = 1, \ldots, n$.

### Demostración

Para cada $i = 1, \ldots, n$,

todas se mueven en $\mathbb{R}$ menos una que restrinjo

$$\begin{aligned} F_{X_i}(x_i) &= P(X_i \leq x_i) = P(X_1 \in \mathbb{R}, \ldots, X_i \leq x_i, \ldots, X_n \in \mathbb{R}) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{x_i} \cdots \int_{-\infty}^{\infty} f_X(t_1, \ldots, t_n) dt_1 \ldots dt_n \\ &= \int_{-\infty}^{x_i} \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(t_1, \ldots, t_n) dt_1 \ldots dt_{i-1} dt_{i+1} \ldots dt_n\right] dt_i \end{aligned}$$

Definimos entonces la función $f_{X_i} : \mathbb{R} \longrightarrow \mathbb{R}$ dada por

me dejo fuera $t_i$

$$f_{X_i}(t_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(t_1, \ldots, t_i, \ldots, t_n) dt_1 \ldots dt_{i-1} dt_{i+1} \ldots dt_n.$$

Dicha función es no negativa e integrable y su integral sobre $\mathbb{R}$ vale uno. Adicionalmente, se tiene la siguiente igualdad:

$$F_{X_i}(x_i) = \int_{-\infty}^{x_i} f_{X_i}(t_i) dt_i.$$

Por tanto, $X_i$ es de tipo continuo con densidad de probabilidad $f_{X_i}$.

### Contraejemplo

para ver que del revés no es así

Se considera seguidamente un contraejemplo, donde **se observa que un conjunto finito de variables continuas no definen necesariamente un vector aleatorio continuo**.

Sea $(X_1, X_2)$ un vector aleatorio tal que $X_1$ es de tipo continuo y $X_2 = 2X_1 + k$, para un cierto $k > 0$.

Entonces $(X_1, X_2)$ no es un vector aleatorio continuo, ya que si lo fuera

$$P(X_2 = 2X_1 + k) = \int_{\{(x_1, x_2) \in \mathbb{R}^2;\ x_2 = 2x_1 + k\}} f_{(X_1, X_2)}(t_1, t_2) dt_1 dt_2 = 0.$$

conjunto de Borel (es una recta)
recta en $\mathbb{R}^2$ → medida nula → prob = 0

Sin embargo, a partir de la definición de dicho vector se tiene:

$$P(X_2 = 2X_1 + k) = 1.$$

## Función de densidad de $Y = g(X)$, X continuo e Y continuo

Sea $X = (X_1, \ldots, X_n)$ en (5) es una variable aleatoria $n$-dimensional **continua**, con función de densidad de probabilidad $f_X$. Supongamos que $g : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ es una función medible satisfaciendo las siguientes condiciones:

no puedo cambiar la dimensión

(i) $g$ es **derivable** en todos los argumentos

(ii) $g$ es **inyectiva** (que tenga inversa) ($g^{-1} : g(E_X) \longrightarrow E_X$, es una aplicación). Es decir, $\forall y \in g(E_X)$, existe un único $x = (x_1, \ldots, x_n) \in E_X$, tal que $g(x) = g(x_1, \ldots, x_n) = y$. Dado que x está unívocamente determinado por y, utilizaremos la notación

$$g^{-1}(y) = x(y) = (x_1(y), \ldots, x_n(y)).$$

(iii) El **jacobiano** de $g^{-1}$ es **no nulo**, i.e., para cualquier $y \in g(E_X)$,

∃ una derivación inversa

$$J(y) = \det\left[\begin{pmatrix} \frac{\partial x_1(y_1,\ldots,y_n)}{\partial y_1} & \cdots & \frac{\partial x_1(y_1,\ldots,y_n)}{\partial y_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial x_n(y_1,\ldots,y_n)}{\partial y_1} & \cdots & \frac{\partial x_n(y_1,\ldots,y_n)}{\partial y_n} \end{pmatrix}\right] = \begin{vmatrix} \frac{\partial x_1(y_1,\ldots,y_n)}{\partial y_1} & \cdots & \frac{\partial x_1(y_1,\ldots,y_n)}{\partial y_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial x_n(y_1,\ldots,y_n)}{\partial y_1} & \cdots & \frac{\partial x_n(y_1,\ldots,y_n)}{\partial y_n} \end{vmatrix} \neq 0$$

Bajo las condiciones (i)-(iii), $Y = g(X)$ es **una variable aleatoria $n$-dimensional continua**, cuya función de densidad $f_Y$ viene dada por:

Jacobiano del cambio → VALOR ABSOLUTO !!

$$f_Y(y) = f_X(g^{-1}(y))|J(y)|, \quad \forall y \in g(E_X)$$

## Distribuciones el máximo y del mínimo

**Notación**
Sea $X = (X_1, \ldots, X_n) : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B}^n, P_X)$ un vector aleatorio $n$-dimensional. Se definen las variables aleatorias unidimensionales

$$\max(X_1, \ldots, X_n) : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$$
$$\min(X_1, \ldots, X_n) : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$$

cuyas funciones de distribución se definen como sigue: para cualquier $y \in \mathbb{R}$,

$$\begin{aligned} F_{\max(X_1,\ldots,X_n)}(y) &= P(\max(X_1, \ldots, X_n) \leq y) \\ &= P(X_1 \leq y, \ldots, X_n \leq y) = F_X(y, \ldots, y) \\ F_{\min(X_1,\ldots,X_n)}(y) &= P(\min(X_1, \ldots, X_n) \leq y) \\ &= 1 - P(\min(X_1, \ldots, X_n) > y) = 1 - P(X_1 > y, \ldots, X_n > y) \\ &= 1 - P_X((y, \infty) \times \cdots \times (y, \infty)). \end{aligned}$$

luego a la f. dist. del vector aleatorio

## Distribuciones el máximo y del mínimo

La **función de distribución conjunta** de la variable aleatoria bidimensional

$$(Z, T) = (\min(X_1, \ldots, X_n), \max(X_1, \ldots, X_n)),$$

viene dada por

$F_{(Z,T)}(x,y) = P[Z \leq x, T \leq y] = P[\min\{X_1 \ldots X_n\} \leq x, \max\{X_1 \ldots X_n\} \leq y]$ →
→ son indep → $= \underbrace{P[\min\{X_1 \ldots X_n\} \leq x]}_{=1} \cdot P[\max\{X_1 \ldots X_n\} \leq y]$

$$F_{(Z,T)}(x, y) = \begin{cases} F_X(y, \ldots, y), & y \leq x \\ F_X(y, \ldots, y) - P(x < X_1 \leq y, \ldots, x < X_n \leq y), & y > x \end{cases}$$

- En este espacio, la **desigualdad de Cauchy-Schwarz** indica que:

$$[E[XY]]^2 \leq E[X^2]E[Y^2], \quad \forall X, Y \in \mathcal{L}^2(\Omega, \mathcal{A}, P).$$

# TEMA 3

## Teorema de multiplicación de esperanzas

(i) Si $X_1, \ldots, X_n$ son independientes, y existe $E[X_i]$ para $i = 1, \ldots, n$, se tiene que

$$\exists E[X_1 \cdots X_n] = E[X_1] \cdots E[X_n].$$

(ii) Si $X_1, \ldots, X_n$ son independientes y $g_1, \ldots, g_n : \mathbb{R} \longrightarrow \mathbb{R}$ son funciones medibles, las variables aleatorias $g_1(X_1), \ldots, g_n(X_n)$ son independientes. Se tiene entonces

$$E[g_1(X_1) \cdots g_n(X_n)] = E[g_1(X_1)] \cdots E[g_n(X_n)].$$

(iii) Si $X$ e $Y$ son independientes $\Rightarrow$ entonces $\mathrm{Cov}(X, Y) = 0$.

(iv) Si $X_1, \ldots, X_n$ son independientes con momento no centrado de orden dos finito
para poder calcular la Var

$$\boxed{\mathrm{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \mathrm{Var}(X_i),} \quad a_1, \ldots, a_n \in \mathbb{R}.$$

En particular, (i) se deriva de forma directa de la siguiente expresión:

$$\begin{aligned} E[X_1 \cdots X_n] &= \int_{\mathbb{R}^n} x_1 \cdots x_n f_{X_1 \ldots X_n}(x_1, \ldots, x_n) dx_1 \cdots dx_n \\ &= \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 \cdots \int_{-\infty}^{\infty} x_n f_{X_n}(x_n) dx_n = E[X_1] \cdots E[X_n] \end{aligned}$$

(ii) se obtiene considerando que para cualesquiera $B_i$, $i = 1, \ldots, n$, subconjuntos de la $\sigma$-álgebra de Borel $\mathcal{B}$ de $\mathbb{R}^n$,

$$P_{g_1(X_1), \ldots, g_n(X_n)}(B_1, \ldots, B_n) = P(g_1(X_1) \in B_1, \ldots, g_n(X_n) \in B_n) = P(X_1 \in \underbrace{g_1^{-1}(B_1)}_{\text{conjunto de Borel}}, \ldots, X_n \in g_n^{-1}(B_n))$$

$$= P_{X_1, \ldots, X_n}(g_1^{-1}(B_1), \ldots, g_n^{-1}(B_n)) \overset{\text{hipótesis}}{=} \prod_{i=1}^n P_{X_i}(g_i^{-1}(B_i)) = \prod_{i=1}^n P(g_i(X_i) \in B_i) = \prod_{i=1}^n P_{g_i(X_i)}(B_i).$$

(iii) Si $X$ e $Y$ son independientes, entonces $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y] = E[X]E[Y] - E[X]E[Y] = 0$.
(iv) Si $X_1, \ldots, X_n$ son independientes con momento no centrado de orden dos finito

ACT 3 pero falta probar que existen los momentos que se usan (suma finita)

$$\mathrm{Var}\left(\sum_{i=1}^n a_i X_i\right) = E\left(\left[\sum_{i=1}^n a_i X_i\right]^2\right) - \left[\sum_{i=1}^n a_i E[X_i]\right]^2 =$$

$$= \sum_{i,j} a_i a_j E[X_i X_j] - \sum_{i,j} a_i a_j E[X_i] E[X_j] = \sum_{i,j} a_i a_j \left[E[X_i X_j] - E[X_i] E[X_j]\right] \overset{\text{hipótesis independencia}}{=}$$

$$= \sum_{i,j} \delta_{i,j} a_i^2 \left[E[X_i^2] - [E[X_i]]^2\right] = \sum_{i=1}^n a_i^2 \left[E[X_i^2] - [E[X_i]]^2\right] = \sum_{i=1}^n a_i^2 \mathrm{Var}(X_i)$$

donde $\delta_{i,j}$ representa la delta de Kronecker que vale 1 cuando $i = j$ y vale cero cuando $i \neq j$.

## Reproductividad de la Binomial !!

Dadas $X_i \sim B(k_i, p)$, $i = 1, \ldots, n$, v.a. independientes, entonces:
importante

$$\boxed{\sum_{i=1}^n X_i \sim B\left(\sum_{i=1}^n k_i, p\right)}. \quad (1)$$

La **demostración** de esta afirmación se obtiene de forma inmediata a partir de la propiedad 4 de independencia, considerando la expresión de la función generatriz de momentos $M_{X_i}$ de $X_i$, que viene dada por

$$M_{X_i}(t) = (pe^t + (1-p))^{k_i}, \quad \forall t \in \mathbb{R}, \quad i = 1, \ldots, n.$$

Específicamente,

$$\begin{aligned} M_{\sum_{i=1}^n X_i}(t) &\overset{\text{def f.g.m.}}{=} E\left[\exp\left(t \sum_{i=1}^n X_i\right)\right] \overset{e^{a+b}=e^a \cdot e^b}{=} E\left[\prod_{i=1}^n \exp(tX_i)\right] \overset{\text{independencia}}{=} \prod_{i=1}^n E\left[\exp(tX_i)\right] \quad (2) \\ &= \prod_{i=1}^n (pe^t + (1-p))^{k_i} = \underbrace{(pe^t + (1-p))^{\sum_{i=1}^n k_i}}_{\text{f.g.m. caracteriza a la distribución}}. \end{aligned}$$

La expresión (1) se deduce de forma directa de (2), dado que $(pe^t + (1-p))^{\sum_{i=1}^n k_i}$ es la función generatriz de momentos de una Binomial con parámetros $\sum_{i=1}^n k_i$ y $p$.

## Reproductividad de la Poisson !!

Dadas $X_i \sim \mathcal{P}(\lambda_i)$, $i = 1, \ldots, n$, independientes, entonces:

$$\boxed{\sum_{i=1}^n X_i \sim \mathcal{P}\left(\sum_{i=1}^n \lambda_i\right)} \quad (3)$$

Para la **demostración** se aplica la independencia y se utiliza que

$$M_{X_i}(t) = \exp\left(\lambda_i(e^t - 1)\right), \quad t \in \mathbb{R}, \quad i = 1, \ldots, n.$$

dem:

$$M_{\sum X_i}(t) = E[e^{t(\sum X_i)}] = E[\prod_i e^{tX_i}] \overset{\text{indep}}{=} \prod_i E[e^{tX_i}] \overset{\text{Poisson}}{=}$$

$$= \prod_i (e^{\lambda_i(e^t-1)}) = e^{\sum \lambda_i (e^t-1)} = e^{(e^t-1)\sum \lambda_i}$$

## Reproductividad de la Binomial Negativa

Dadas $X_i \sim BN(k_i, p)$, $i = 1, \ldots, n$, independientes, entonces:

$$\boxed{\sum_{i=1}^n X_i \sim BN\left(\sum_{i=1}^n k_i, p\right)} \quad (4)$$

Para la **demostración** se aplica la independencia y se utiliza que

$$M_{X_i}(t) = \left[\frac{p}{1-(1-p)e^t}\right]^{k_i}, \quad t \in \mathbb{R}, \quad i = 1, \ldots, n.$$

dem:

$$E[e^{t(\sum X_i)}] = E[\prod_i e^{tX_i}] \overset{\text{indep}}{=} \prod_i E[e^{tX_i}] \overset{\text{Bin neg}}{=}$$

$$M_{\sum X_i}(t) = \frac{\quad}{(1-p)e^t}\Big]^{k_i} = \left[\frac{p}{1-(1-p)e^t}\right]^{\sum k_i}$$

## Reproductividad de la Geométrica !!

Dadas $X_i \sim G(p)$, $i = 1, \ldots, n$, independientes, entonces:

$$\boxed{\sum_{i=1}^n X_i \sim BN(n, p)} \quad (5)$$

La **demostración** es inmediata a partir de la propiedad anterior, dado que $X_i \sim G(p)$, coincide con $X_i \sim BN(1, p)$, i.e., $k_i = 1$, $i = 1, \ldots, n$.

## Reproductividad de la Normal

Dadas $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$, independientes, entonces:

$$\boxed{\sum_{i=1}^n X_i \sim \mathcal{N}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)} \quad (6)$$

La **demostración** es inmediata a partir de la independencia y de la siguiente expresión:

$$M_{X_i}(t) = \exp\left(t\mu_i + \frac{t^2\sigma_i^2}{2}\right), \quad t \in \mathbb{R}, \quad i = 1, \ldots, n.$$

dem:

$$M_{\sum X_i}(t) = E[e^{t(\sum X_i)}] = E[\prod_i e^{tX_i}] \overset{\text{indep}}{=} \prod_i E[e^{tX_i}] =$$

$$= \prod_i \exp\left(t\mu_i + \frac{t^2\sigma_i^2}{2}\right) = \exp\left(\sum_i t\mu_i + \frac{t^2\sigma_i^2}{2}\right) =$$

$$= \exp\left(t\sum \mu_i + \frac{t^2 \sum \sigma_i^2}{2}\right)$$

## Reproductividad de la Gamma

Dadas $X_i \sim \Gamma(p_i, a)$, $i = 1, \ldots, n$, independientes, entonces:

$$\boxed{\sum_{i=1}^n X_i \sim \Gamma\left(\sum_{i=1}^n p_i, a\right)} \quad (7)$$

La **demostración** es inmediata a partir de la independencia y de la siguiente expresión:

$$M_{X_i}(t) = \left[1 - \frac{t}{a}\right]^{-p_i}, \quad t < a, \quad i = 1, \ldots, n.$$

En **consecuencia**, si $X_i \sim \mathcal{E}(k_i, a)$, $i = 1, \ldots, n$, independientes, entonces:

$$\sum_{i=1}^n X_i \sim \Gamma\left(\sum_{i=1}^n k_i, a\right) \equiv \mathcal{E}\left(\sum_{i=1}^n k_i, a\right). \quad (8)$$

dem.

$$M_{\sum X_i}(t) = E[e^{t(\sum X_i)}] = E[\prod_i e^{tX_i}] \overset{\text{indep}}{=} \prod_i E[e^{tX_i}] \overset{\text{Gamma}}{=}$$

$$= \prod_i [1 - t/a]^{-p_i} = [1 - t/a]^{-\sum p_i}$$

## Reproductividad de la Exponencial

Dadas $X_i \sim \exp(\lambda)$, $i = 1, \ldots, n$, independientes, entonces:

$$\boxed{\sum_{i=1}^n X_i \sim \Gamma(n, \lambda) \equiv \mathcal{E}(n, \lambda)} \quad (9)$$

La **demostración** se obtiene de la independencia y de la siguiente expresión:

$$M_{X_i}(t) = \left[1 - \frac{t}{\lambda}\right]^{-1}, \quad t < \lambda, \quad i = 1, \ldots, n.$$

dem:

$$M_{\sum X_i}(t) = E[e^{t(\sum X_i)}] = E[\prod_i e^{tX_i}] \overset{\text{indep}}{=} \prod_i E[e^{tX_i}] \overset{\text{exp}}{=}$$

$$= \prod_i [1 - t/\lambda]^{-1} = [1 - t/\lambda]^{-\sum 1} = [1 - t/\lambda]^{-n}$$

# TEMA 4

## Caso especial: varianza condicionada

Supongamos que $X$, tal que $\exists E[X^2]$, e $Y$ son dos variables aleatorias sobre el mismo espacio de probabilidad, entonces:

(i). $\exists \text{Var}(X/Y)$, y $\text{Var}(X/Y) \geq 0$.

(ii). $\text{Var}(X/Y) = E[X^2/Y] \underset{\text{negativo}}{-} (E[X/Y])^2$.

(iii). **Descomposición de la varianza**: si $\exists \text{Var}(E[X/Y])$, y $\exists E[\text{Var}(X/Y)]$, entonces:

$$\boxed{\text{Var}(X) = \text{Var}(E[X/Y]) + E[\text{Var}(X/Y)]}.$$

**Demostración:**

$$\begin{aligned}
\text{Var}(X) &\overset{\text{def}}{=} E\left[(X - E[X])^2\right] \overset{E[E[X/Y]]=E[X]}{=} E\left[E\left[(X - E[X])^2/Y\right]\right] \; \textit{(Prop.(v))} \\
&\overset{\text{desarrollo cuadrado}}{=} E\left[E\left[(X^2 + (E[X])^2 - 2XE[X])/Y\right]\right] \quad E[\overset{\text{cte}}{2E[X]}X/Y] = 2E[X]E[X/Y] \\
&\overset{\text{linealidad}}{=} E\left[E[X^2/Y] + (E[X])^2 - 2E[X]E[X/Y]\right] \; \textit{(Linealidad)} \\
&= E\left[E[X^2/Y] \underbrace{- (E[X/Y])^2}_{\text{resto}} \text{ y } \underbrace{+ (E[X/Y])^2}_{\text{sumo}} + (E[X])^2 - 2E[X]E[X/Y]\right] \\
&\overset{\text{linealidad}}{=} E[\text{Var}(X/Y)] + E\left[(E[X/Y] - E[X])^2\right] \quad \text{quito el cuadrado notable desarrollado} \\
&= E[\text{Var}(X/Y)] + \text{Var}(E[X/Y]) \,.\text{(Prop. (v) y def. momento centrado condicionado de orden dos.)}
\end{aligned}$$

La **función $\varphi_{\text{opt}}$, que minimiza el E.C.M.**, en la aproximación de $Y$ mediante una función de $X$ **viene dada** por:

$$\varphi_{\text{opt}}(X) = E[Y/X].$$

$$\begin{aligned}
\text{E.C.M.}(\varphi(X)) &= E\left[(Y - \overset{\varphi(X)}{E[Y/X]})^2\right] \\
&= E\left[E\left[(Y - E[Y/X])^2/X\right]\right] = E[\text{Var}(Y/X)] .
\end{aligned}$$

$$= E[Y^2] - E\left[(E(Y/X))^2\right]$$

(iii). Si $X$ e $Y$ son **independientes**, las curvas de regresión **son rectas paralelas** a los ejes: $y = E[Y]$ y $x = E[X]$.
En este caso **estimamos una variable, sin observar la otra, mediante su esperanza y el E.C.M. es su varianza.**

Se definen **las razones de correlación** de $X$ sobre $Y$, $\eta^2_{X/Y}$ y de $Y$ sobre $X$, $\eta^2_{Y/X}$, como sigue:

razón
bondad
correlación

$$\eta^2_{X/Y} = \frac{\text{Var}(E[X/Y])}{\underset{\to \text{ var. que aproximo}}{\text{Var}(X)}} \overset{\nearrow \text{modelo}}{=} 1 - \frac{E[\text{Var}(X/Y)]}{\text{Var}(X)}$$

$\simeq 1$ buena
$\simeq 0$ mala

- La razón de correlación es una **medida adimensional** que representa la **proporción de varianza de la variable dependiente que queda explicada** por la función de regresión.

Si $\eta^2_{X/Y} = 0$ y $\eta^2_{Y/X} = 0$, sii [redacted] las curvas de regresión de $X/Y$ e $Y/X$ coinciden respectivamente con las rectas $x = E[X]$ e $y = E[Y]$.

(vi). $\eta^2_{X/Y} = 1 \;\Leftrightarrow\; X$ dependen funcionalmente de $Y$. → Ambas ⇒ dep. funcional recip.

## Planteamiento de la regresión lineal de $Y$ sobre $X$ (equivalentemente para $X$ sobre $Y$)

**Intentamos resolver el problema de predecir, por mínimos cuadrados, los valores de una variable $Y$ a partir de una función lineal de otra $X$.** En este sentido se busca una función $\varphi^{L\,(\text{lineal})}_{\text{opt}}(X)$ que minimice el E.C.M.

$$\varphi^L_{\text{opt}}(X) = \min_{a,b} E[(Y \overbrace{- aX - b}^{-\varphi(X),\ \varphi(X)=aX+b})^2].$$

El problema se reduce a obtener $a$ y $b$ que minimicen $L = E[(Y - aX - b)^2]$.

$$L = E[(Y - aX - b)^2] = E[Y^2] + a^2E[X^2] + b^2 + 2abE[X] - 2aE[XY] - 2bE[Y]$$

derivo dos veces, busco mínimo bidim $\}\nabla(a,b)$

Derivando respecto $a$ y $b$ e igualando a cero, se obtiene el siguiente sistema de ecuaciones denominado sistema de **Ecuaciones Normales**:

$$\left.\begin{aligned}
aE[X^2] + bE[X] &= E[XY] \\
\text{(derivada)} \quad b + aE[X] &= E[Y]
\end{aligned}\right\}$$

cuya solución es $a = \dfrac{Cov(X,Y)}{Var(X)}$ y $b = E[Y] - aE[X]$. → MÁS FÁCIL DE TRABAJAR QUE LA CURVA

## Regresión mínimo-cuadrática, cuando $\varphi^L$ es una recta $(R_{Y/X})$

A partir de las soluciones de las ecuaciones normales anteriores se tiene lo siguiente.
**En la regresión lineal mínimo cuadrática de $Y$ sobre $X$,**

$y = \varphi_{opt}(x) = aX + b = \frac{cov(X,Y)}{var X}x + E[Y] - aE[X]$

$$\boxed{Y} \approx \varphi^L_{\text{opt}}(X) = \min_{a,b} E[(Y - aX - b)^2] = \boxed{E[Y] + \frac{\text{Cov}(X,Y)}{\text{Var}(X)}(X - E[X])}.$$

Esta expresión recibe el nombre de **Recta de Regresión de $Y$ sobre $X$**.

$$\boxed{y - E[Y] = \underbrace{\frac{\text{Cov}(X,Y)}{\text{Var}(X)}}_{\text{PENDIENTE}}(x - E[X])}.$$

coeficiente de regresión $\gamma_{Y/X}$

pendiente > 0 ⇒ si una v.a. crece, la otra también (viceversa)
pendiente < 0 ⇒ si una v.a. crece, la otra decrece (y viceversa)
|pendiente| indica la variación de una v.a. según la otra.

El **E.C.M asociado** viene dado por: (Ejercicio voluntario)

$$\begin{aligned}
E.C.M.(\varphi^L_{\text{opt}}(X)) &= \text{Var}(Y) - \frac{[\text{Cov}(X,Y)]^2}{\text{Var}(X)} \\
\text{Por tanto, } \text{Var}(Y) &= \text{Var}(\varphi^L_{\text{opt}}(X)) + E.C.M.(\varphi^L_{\text{opt}}(X)).
\end{aligned}$$

Las rectas de regresión se cortan en el punto $(E[X], E[Y])$.

## Coeficiente de determinación lineal

BONDAD DEL AJUSTE

Una medida de la asociación lineal de las varibles viene determinada por el **coeficiente de determinación lineal** $\rho^2_{X,Y}$, se define como:

$$\boxed{\rho^2_{X,Y} = \frac{[\text{Cov}(X,Y)]^2}{\text{Var}(X)\text{Var}(Y)}} \left( = \frac{\text{Var}(\varphi^L_{\text{opt}}(X))}{\text{Var}(Y)} = \frac{\text{Var}(\varphi^L_{\text{opt}}(Y))}{\text{Var}(X)} \right).$$

Si me piden bondad del ajuste. Indicar en %

**Observación**:

"El 80% de la variabilidad de la v.a. dependiente queda explicada por el modelo"

Tal y como los razones de correlación, el coeficiente de determinación lineal **es una medida de la bondad del ajuste lineal**, o lo que es lo mismo, mide la proporción de varianza de explicada por el modelo de regresión lineal.

(ii). $\rho^2_{X,Y} = \gamma_{X/Y}\gamma_{Y/X}$

(iv). $\rho^2_{X,Y} = 0 \;\Leftrightarrow$ Las rectas de regresión de $Y/X$ y de $X/Y$ son respectivamente $y = E[Y]$, y $x = E[X]$.

(vi). $\rho^2_{X,Y} \leq \eta^2_{Y/X}$ y $\rho^2_{X,Y} \leq \eta^2_{X/Y}$. //significa que la curva siempre ajusta mejor o igual que la recta

Las desigualdades anteriores (prop. (vi).) se convierten en igualdades si y sólo si **las curvas de regresión coinciden con las correspondientes rectas de regresión**.

## Coeficiente de correlación lineal

Independientemente de que el coeficiente de determinación lineal ofrece información acerca de la bondad del ajuste del modelo lineal, la **medida de correlación por excelencia** es el **coeficiente de correlación lineal**, ya que también proporciona información, no sólo de la **fuerza de asociación**, sino también de **su sentido**.

→ tiene signo (el de Cov)

Se define el coeficiente de correlación lineal $\rho_{X,Y}$, como:

$$\boxed{\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}}$$

(iii). $\rho_{X,Y} = 0$ (variables incorreladas linealmente) $\Leftrightarrow$ Las rectas de regresión de $Y/X$ y de $X/Y$ son respectivamente $y = E[Y]$, y $x = E[X]$.

(iv). Si $X$ e $Y$ son independientes, entonces $\rho_{X,Y} = 0$ (el recíproco no es cierto)

(v). $|\rho_{X,Y}| = 1 \;\Leftrightarrow$ Existe dependencia funcional lineal (positiva o negativa, en función del signo) entre $X$ e $Y$.

(vi). $|\rho_{X,Y}| = 1 \;\Leftrightarrow$ Las dos rectas de regresión coinciden entre sí y con la recta de dependencia, además de con las curvas de regresión.

(vii). Si $\rho_{X,Y} = 1$ la dependencia lineal es exacta y positiva (las dos v.a. crecen o decrecen de forma simultánea).

(viii). Si $\rho_{X,Y} = -1$ la dependencia lineal es exacta y negativa (cuando una v.a crece la otra decrece y viceversa).

# TEMA 5

## Distribución multinomial

Consideramos el vector aleatorio $(X_1, X_2, \dots, X_{k+1})$ cuyas componentes $X_i$ **cuentan el número de veces que ocurre el suceso** $A_i$, $\forall i = 1, 2, \dots, k+1$, de modo que $X_{k+1}$ queda completamente determinado por $x_{k+1} = n - \sum_{i=1}^{k} x_i$.

Entonces se dice que $(X_1, X_2, \dots, X_k)$ sigue una **distribución multinomial** $k$-dimensional con parámetros $n$ y $p_1, \dots p_k$, que **denotamos**,

$$(X_1, \dots, X_k) \sim M_k(n; p_1, \dots, p_k),$$

Realmente no es k, sino k+1

si y sólo si su **función masa de probabilidad** $k$-dimensional viene dada por:

la que hemos visto arriba

podemos relacionarla con la de la binomial k=1

$$P(X_1 = x_1, \dots, X_k = x_k) = \frac{n!}{x_1! \cdots x_k! (n - \sum_{i=1}^{k} x_i)!} p_1^{x_1} \cdots p_k^{x_k} \left(1 - \sum_{i=1}^{k} p_i\right)^{n - \sum_{i=1}^{k} x_i},$$

donde $n \in \mathbb{N}$, $0 < p_i < 1$, $i = 1, \dots, k$, $\sum_{i=1}^{k} p_i < 1$, $x_i \in \{0, \dots, n\}$, y $\sum_{i=1}^{k} x_i \leq n$.

## Función generatriz de momentos

$$\begin{aligned} M_{X_1,\dots,X_k}(t_1, \dots, t_k) &= E\left[\exp\left(\sum_{i=1}^{k} t_i X_i\right)\right] \\ &= \left(p_1 e^{t_1} + \cdots + p_k e^{t_k} + \left(1 - \sum_{i=1}^{k} p_i\right)\right)^n, \quad t_1, \dots, t_k \in \mathbb{R}. \end{aligned}$$

para hallar la de $X_1$, pongo el resto a 0. Obtendré la f.g.m. de la binomial.

⇓

Las marginales son binomiales.

En particular, para las **marginales unidimensionales** se cumple:

¡SON BINOMIALES!

$$X_i \sim B(n, p_i), \quad i = 1, \dots, k.$$

En particular, las **condicionadas unidimensionales** cumplen:

$$X_i / X_j = x_j \sim B\left(n - x_j, \frac{p_i}{1 - p_j}\right), \quad i, j = 1, \dots, k; \; i \neq j.$$

## Vector de medias y covarianzas

Sea $X = (X_1, \dots, X_k)$, vector aleatorio $k$-dimensional, con distribución multinomial con parámetros $n$, y $p_1, \dots, p_k$. Se tiene entonces:

i). $E[X] = (np_1, np_2, \dots, np_k)$

ii). $Cov(X_i, X_j) = -np_ip_j$, $\forall i \neq j$

## Regresión y correlación para el caso bidimensional

- La **curva (recta) de regresión** de $X_i / X_j$, viene dada por

  (coinciden)

$$x_i = \frac{np_i}{1 - p_j} - \frac{p_i}{1 - p_j} x_j.$$

- Las **razones de correlación**, que **coinciden con el coeficiente de determinación**, vienen dadas por:

$$\eta^2_{X_i/X_j} = \eta^2_{X_j/X_i} = \rho^2_{X_iX_j} = \frac{p_ip_j}{(1 - p_i)(1 - p_j)}, \quad i \neq j, \; i, j \in \{1, \dots, k\}.$$

- El **coeficiente de correlación lineal** viene dado por

$$\rho_{X_iX_j} = -\sqrt{\frac{p_ip_j}{(1 - p_i)(1 - p_j)}}, \quad i \neq j, \; i, j \in \{1, \dots, k\}.$$

## Distribución Normal Bidimensional

Se dice que $(X_1, X_2)$ se distribuye según una **normal bidimensional** con **vector de medias** $\mu = (\mu_1, \mu_2)$ y con **matriz de varianzas-covarianzas**

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix},$$

si su **función de densidad de probabilidad** viene dada por:

"dividir entre σ²"

$$f_{(X_1,X_2)}(x_1, x_2) = \frac{1}{2\pi [\det(\Sigma)]^{1/2}} e^{\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)}$$

"varianza, σ"

desarrollada

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right)},$$

* ver analogía con la unidimensional

donde $x = (x_1, x_2) \in \mathbb{R}^2$, $\sigma_i^2$ es la varianza de $X_i$, $i = 1, 2$, y $\rho$ es el coeficiente de correlación lineal de $X$ e $Y$.

## Marginales y condicionadas

Ambas distribuciones, **marginales y condicionadas**, son **normales**.

- En el caso de las **marginales**, se tiene

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), \quad X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2).$$

- Para las **condicionadas** se obtienen las siguientes distribuciones normales:

$$X_1 / X_2 = x_2 \sim \mathcal{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1 - \rho^2)\right)$$

$$X_2 / X_1 = x_1 \sim \mathcal{N}\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1 - \rho^2)\right)$$

la que condiciona

EJERCICIO: cae una de las dos en el examen (1 pto) Hay que demostrarla.

$$f_{X_1/X_2=x_2}(x_1) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_2}(x_2)} = \frac{\frac{1}{\sqrt{2\pi}\sigma_2} e^{\left(-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)} \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} e^{\left(-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x_1-\mu_1-\rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)\right]^2\right)}}{\frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}}} =$$

normal 2D; normal 1D

$$= \underbrace{\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}}_{\sigma} \exp\left(-\underbrace{\frac{1}{2\sigma_1^2(1-\rho^2)}}_{\sigma^2}\left[x_1 - \underbrace{\mu_1 - \rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)}_{\mu}\right]^2\right)$$

$$X_1/X_2=x_2 \sim \mathcal{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2), \sigma_1^2(1-\rho^2)\right)$$

## Regresión y correlación

- La **curva de regresión**, que **coincide con la recta de regresión** de $X_i/X_j$, para $i, j = 1, 2$, $i \neq j$, viene dada por la **media de las distribuciones normales condicionadas**, anteriormente especificadas.

  Más concretamente, viene dada por

$$x_i = \mu_i + \rho\frac{\sigma_i}{\sigma_j}(x_j - \mu_j), \quad i, j = 1, 2, \; i \neq j.$$

- Equivalentemente, las **razones de correlación coinciden con el coeficiente de determinación** $\rho^2$, ya que las curvas de regresión coinciden con las rectas de regresión.
- El **E.C.M. asociado a la curva de regresión** de $X_i$ sobre $X_j$ viene dado por $\sigma_i^2(1 - \rho^2)$, $i = 1, 2$, para cada $j = 1, 2$, con $i \neq j$.

## Independencia ⇔ incorrelación ⇔ Σ Diagonal

Observamos que

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix},$$

matriz de covarianzas de una normal bidimensional, es no singular ($\rho^2 \neq 1$) y semidefinida positiva.

Es fácil comprobar que bajo el supuesto de normalidad bivariante:

**Independencia** ⇔ **Incorrelación** ⇔ Σ **Diagonal**

EQUIVALENCIA SÓLO CUANDO ES BINORMAL

De hecho la segunda equivalencia es evidente, y la primera es debida a que

$$\rho = 0 \Leftrightarrow f(x, y) = f_X(x) f_Y(y).$$

**Importante:** esta equivalencia no es válida si cada variable tiene distribución normal unidimensional pero la conjunta no.

## Teorema para la función generatriz de momentos de la normal bivariante

Para $\mathrm{t} = (t_1, t_2) \in \mathbb{R}^2$,

$$
\begin{aligned}
M_{X_1,X_2}(t_1,t_2) &= M_{X_1,X_2}(\mathrm{t}) = E\left[\exp\left(\langle \mathrm{t}, \mathrm{X}\rangle\right)\right] \\
&= \exp\left(\langle \boldsymbol{\mu}, \mathrm{t}\rangle + \frac{\mathrm{t}\Sigma\mathrm{t}^T}{2}\right), \\
&= \exp\left(t_1\mu_1 + t_2\mu_2 + \frac{t_1^2\sigma_1^2 + t_2^2\sigma_2^2 + 2t_1t_2\rho\sigma_1\sigma_2}{2}\right).
\end{aligned}
$$

$$
M_{X_1,X_2}(t_1,t_2) = E\left[\overset{e^{tx}}{e^{(\langle \mathbf{t},\mathbf{X}\rangle)}}\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \overbrace{e^{(t_1x_1+t_2x_2)}}^{\text{esto se puede separar en } x_1,x_2 \to \text{entonces...}} \underbrace{f_{X_1,X_2}(x_1,x_2)}_{\text{quiero separar esto en } x_1,x_2 \text{ (dependencia de una sola variable)}} dx_1\,dx_2
$$

$$
{}^{*} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{(t_1x_1+t_2x_2)} f_{X_1/X_2=x_2}(x_1) f_{X_2}(x_2)\,dx_1\,dx_2
$$

tenemos que integrar primero respecto de $x_1$ por *

$$
= \int_{-\infty}^{\infty} e^{(t_2x_2)} f_{X_2}(x_2)\left[\underbrace{\int_{-\infty}^{\infty} e^{(t_1x_1)} f_{X_1/X_2=x_2}(x_1)\,dx_1}_{\text{es normal (hipótesis)}}\right] dx_2 = \int_{-\infty}^{\infty} e^{(t_2x_2)} \underbrace{M_{X_1/X_2=x_2}(t_1)}_{\text{f.g.m. condicionada}} f_{X_2}(x_2)\,dx_2
$$

$$
= \int_{-\infty}^{\infty} e^{(t_2x_2)} e^{\left(t_1\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)\right)+\frac{t_1^2\sigma_1^2(1-\rho^2)}{2}\right)} f_{X_2}(x_2)\,dx_2
$$

sustituyo $M_{X_1/X_2=x_2}$ por su expresión $\begin{cases} X_1/X_2=x_2 \sim \mathcal{N}\left(\underbrace{\mu_1+\rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)}_{\mu}, \underbrace{\sigma_1^2(1-\rho^2)}_{\sigma^2}\right) \\ M_X(t) = e^{t\mu+\frac{t^2\sigma^2}{2}} \end{cases}$

$$
= \underbrace{e^{\left(t_1\left(\mu_1-\rho\frac{\sigma_1}{\sigma_2}\mu_2\right)+\frac{t_1^2\sigma_1^2(1-\rho^2)}{2}\right)}}_{\text{saco lo que no depende de } x_2} \underbrace{\int_{-\infty}^{\infty} \underbrace{e^{\left(t_2x_2+t_1\rho\frac{\sigma_1}{\sigma_2}x_2\right)}}_{\text{parte de la f.g.m. que sí depende de } x_2} \overbrace{f_{X_2}(x_2)}^{\text{p.h. es f.dist. de una normal}}dx_2}_{\text{f.g.m. marginal rara}}
$$

$$
= e^{\left(t_1\left(\mu_1-\rho\frac{\sigma_1}{\sigma_2}\mu_2\right)+\frac{t_1^2\sigma_1^2(1-\rho^2)}{2}\right)} M_{X_2}\left(t_2+t_1\rho\frac{\sigma_1}{\sigma_2}\right)
$$

uso los parámetros de $X_2$

$$
= e^{\left(t_1\left(\mu_1-\overbrace{\rho\frac{\sigma_1}{\sigma_2}\mu_2}^{\text{se anula *}}\right)+\frac{t_1^2\sigma_1^2(1-\rho^2)}{2}\right)} e^{\left(\left(t_2+\underbrace{t_1\rho\frac{\sigma_1}{\sigma_2}}_{\text{se anula *}}\right)\mu_2+\frac{\left(t_2+t_1\rho\frac{\sigma_1}{\sigma_2}\right)^2\sigma_2^2}{2}\right)}
$$

$$
= e^{\left(t_1\mu_1+t_2\mu_2+\frac{t_1^2\sigma_1^2+t_2^2\sigma_2^2+2t_1t_2\rho\sigma_1\sigma_2}{2}\right)}, \quad t_1, t_2 \in \mathbb{R}.
$$

## Normalidad para combinaciones lineales de las componentes

Sea $\mathrm{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$. Se considera $\mathrm{A}_{2\times q}$ una matriz de rango máximo $q$, con $q = 1, 2$. Entonces:

subcuadrado con determinante $\neq 0$

$$
\mathrm{Y} = \mathrm{X}\mathrm{A}_{2\times q} \sim \mathcal{N}_q\left(\boldsymbol{\mu}\mathrm{A}, \mathrm{A}^T\Sigma\mathrm{A}\right).
$$

**Demostración:**

La demostración se obtiene de forma directa a partir del teorema anterior.

Sea $t \in \mathbb{R}^q$, entonces:

$$
E\left[e^{tY^T}\right] = E\left[e^{(tA^T)X^T}\right] = M_X(tA^T) = e^{(tA^T)\mu^T + \frac{(tA^T)\Sigma(tA^T)^T}{2}} = e^{t(\mu A)^T + \frac{t(A^T\Sigma A)t^T}{2}}
$$

Dado que $A$ es de rango máximo, $|A^T\Sigma A| \neq 0$, se deduce que $Y = XA \sim \mathcal{N}_q(\mu A, A^T\Sigma A)$.

**Observación:**
Se tiene que $Y = XA_{2x2} = (a_{11}X_1 + a_{12}X_2, a_{21}X_1 + a_{22}X_2)$. De modo que **cualquier vector bidimensional cuyas componentes sean combinación lineal** de $X_1$ y $X_2$, componentes de un vector normal bidimensional, de modo que la **matriz** $A$ que define esta combinaciones sea **no singular**, tiene una distribución **normal bidimensional**.

# TEMA 6

- **Convergencia casi segura**:
  casi nunca lo conseguiremos

$$X_n \to^{c.s.} X, \quad n \to \infty \Leftrightarrow P\left[\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right] = 1.$$

- **Convergencia en probabilidad**:

$$X_n \to^{P} X, \quad n \to \infty \Leftrightarrow \forall \varepsilon > 0, \ \lim_{n\to\infty} P\left[\omega : |X_n(\omega) - X(\omega)| \le \varepsilon\right] = 1.$$

- **Convergencia en Ley o en distribución**:

$$X_n \to^{L} X, \quad n \to \infty \Leftrightarrow \lim_{n\to\infty} F_{X_n}(x) = \underbrace{F_X(x)}_{\text{función de distribución}}, \quad \forall x \in C(F_X),$$

siendo $C(F_X)$ el conjunto de puntos de continuidad de $F_X$.

## Leyes de los grandes números

Sean $X_n$, $n \in \mathbb{N}$, variables aleatorias sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$ **independientes**. Adicionalmente, sea $S_n$ la variable aleatoria que define las **sumas parciales** de la sucesión $\{X_n\}_{n\in\mathbb{N}}$, i.e.,

$$S_n = \sum_{i=1}^{n} X_i, \quad n \ge 1.$$

Se consideran dos sucesiones de números reales $\{a_n\}_{n\in\mathbb{N}}$ y $\{b_n\}_{n\in\mathbb{N}}$ tal que $b_n \uparrow \infty$.

- **Ley débil de los grandes números respecto $\{a_n\}_{n\in\mathbb{N}}$ y $\{b_n\}_{n\in\mathbb{N}}$.** Se dice que la sucesión aleatoria $\{X_n\}_{n\in\mathbb{N}}$ satisface la ley débil respecto a $\{a_n\}_{n\in\mathbb{N}}$ y $\{b_n\}_{n\in\mathbb{N}}$ si

$$\frac{S_n - a_n}{b_n} \to^{P} 0.$$

- **Ley fuerte de los grandes números respecto $\{a_n\}_{n\in\mathbb{N}}$ y $\{b_n\}_{n\in\mathbb{N}}$.** Se dice que la sucesión aleatoria $\{X_n\}_{n\in\mathbb{N}}$ satisface la ley fuerte respecto a $\{a_n\}_{n\in\mathbb{N}}$ y $\{b_n\}_{n\in\mathbb{N}}$ si

$$\frac{S_n - a_n}{b_n} \to^{c.s.} 0.$$

## Teorema 1 Ley débil de Bernoulli

Sea $\{X_n\}_{n\in\mathbb{N}}$ una secuencia de variables aleatorias independientes e idénticamente distribuidas según una Bernoulli con parámetro $p$, i.e., $X_n \sim \mathcal{B}(1,p)$, para cualquier $n \in \mathbb{N}$, entonces

$$\frac{S_n - E[S_n]}{n} \to^{P} 0, \text{ y en consecuencia } \frac{S_n}{n} \to^{P} p.$$

## Lema 0 Desigualdad de Chebychev

Sea $X$ una v.a. con esperanaza finita $E[X] < \infty$, entonces $\forall \varepsilon > 0$ se tiene que:

$$P[\{|X| \ge \varepsilon\}] = P[\omega : |X(\omega)| \ge \varepsilon] \le \frac{E[X^2]}{\varepsilon^2}.$$

**Observaciones:**

- Si se considera la v.a. $X - E[X]$ se obtiene la versión clásica de esta desigualdad,

$$P[\{|X - E[X]| \ge \varepsilon\}] = P[\omega : |X(\omega) - E[X]| \ge \varepsilon] \le \frac{E[(X - E[X])^2]}{\varepsilon^2} = \frac{Var(X)}{\varepsilon^2}.$$

- Si se toman complementarios también se puede escribir como sigue:

$$P[\{|X - E[X]| < \varepsilon\}] = P[\omega : |X(\omega) - E[X]| < \varepsilon] \ge 1 - \frac{Var(X)}{\varepsilon^2}.$$

## Teorema 2 Ley débil de Khintchine

*Casi siempre cae en el examen

Sea $\{X_n\}_{n\in\mathbb{N}}$ una secuencia de variables aleatorias independientes e idénticamente distribuidas y existe $E[X_n] = \mu$, para cualquier $n \in \mathbb{N}$, entonces

$$\frac{S_n - E[S_n]}{n} \to^{P} 0, \text{ y en consecuencia } \frac{S_n}{n} \to^{P} \mu.$$

→ Queremos aplicar convergencia en probabilidad: ¿ $\frac{S_n - E(S)}{n} \to^{P} 0 \Rightarrow \lim_{n\to\infty} P[\omega : |\frac{S_n - E(S_n)}{n}| \le \varepsilon] = 1$ ? $\cong$ ¿ $\lim_{n\to\infty} P[\omega : |\frac{S_n - E(S_n)}{n}| > \varepsilon] = 0$ ? para esto, aplico des. Chebyshev

**Demostración:** se derivará para el caso particular en el que existe $E[X_n^2]$, para cualquier $n \in \mathbb{N}$. A partir de la desigualdad de Chebychev, para cualquier $\varepsilon > 0$, teniendo en cuenta que $X_i$, $i = 1, \dots, n$, son independientes y con la misma distribución, y por tanto,

$$\text{Var}(S_n) \overset{\text{no afecta cambio de origen}}{=} \text{Var}(S_n - E[S_n]) = \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) = n\text{Var}(X_1),$$

Desigualdad Chebyshev

$$P\left(\left|\frac{S_n - E[S_n]}{n}\right| \ge \varepsilon\right) \le \frac{\text{Var}(S_n - E[S_n])}{\varepsilon^2 n^2} = \frac{n\text{Var}(X_1)}{n^2\varepsilon^2} \to 0, \quad n \to \infty.$$

En particular, dado que $E[S_n] = n\mu$, se obtiene $S_n/n \to^{P} \mu$, $n \to \infty$.

La ley débil de Bernoulli, se obtiene entonces considerando $\mu = p$.

## Teorema 3 Primer teorema límite (Bernoulli)

Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas según una Bernoulli con parámetro $p$, i.e., $X_n \sim \mathcal{B}(1,p)$, para cualquier $n \in \mathbb{N}$, entonces

$$\frac{S_n - E[S_n]}{n} \to^{L} 0.$$

**Demostracion:** Se obtiene de forma directa del [redacted] Teorema 2, dado que la convergencia [redacted] en [redacted] implica la convergencia en Ley.

Prob

## Teorema 4 Segundo teorema límite (De Moivre y Laplace)

Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas según una Bernoulli con parámetro $p$, i.e., $X_n \sim \mathcal{B}(1,p)$, para cualquier $n \in \mathbb{N}$, entonces

$$\frac{S_n - E[S_n]}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - np}{\sqrt{np(1-p)}} \underbrace{\to^{L}}_{\text{convergencia en Ley}} Z \sim \mathcal{N}(0,1).$$

## Teorema 5 Teorema límite de Lévy

Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas. Se tienen los siguientes límites en ley:

(i) Si $\exists E[X_1] \Rightarrow \exists E[X_n]$, $\forall n$, se tiene $\frac{S_n - E[S_n]}{n} \to^{L} 0.$

(ii) Si $\exists E[X_1^2] \Rightarrow \exists E[X_n^2]$, $\forall n$, se tiene $\frac{S_n - E[S_n]}{\sqrt{\text{Var}(S_n)}} \to^{L} Z \sim \mathcal{N}(0,1).$

necesario para que haya varianzas

### Teorema límite de Levy (Demostración)

La afirmación (i) se obtiene como consecuencia directa del Teorema 2 o bien, del Teorema 3.

La afirmación (ii) se demostrará para el caso particular, donde existe la función generatriz de momentos de las variables aleatorias de la sucesión $\{X_n\}_{n\in\mathbb{N}}$, definidas, todas ellas, sobre un intervalo común.

Se considera la secuencia

$$S_n^* = \frac{S_n - E[S_n]}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - n\mu}{\sqrt{n}\sigma},$$

podemos hacerlo pq son idénticamente distribuidas (i.d.)

i.d. ⇒ Var(S₁)·n

donde $\sigma = \sqrt{\text{Var}(X_n)}$, para cualquier $n \in \mathbb{N}$. Se asume que $\mu = E[X_n] = 0$, para cualquier $n \in \mathbb{N}$. En caso contrario, se considera $\widetilde{X}_n = X_n - \mu$, $n \in \mathbb{N}$.

para generalizar

Dado que $X_n$, $n \in \mathbb{N}$, son independientes e idénticamente distribuidas, con $\mu = 0$, la función generatriz de momentos

$$M_{S_n^*}(t) = \left[M_{X_1}\left(\frac{t}{\sqrt{n}\sigma}\right)\right]^n,$$

donde $M_{X_1} = M_{X_n} = M_X$, para todo $n \in \mathbb{N}$. Se considera ahora el desarrollo de Taylor de segundo orden de $M_X = E[\exp(tX)]$, $t \in (-a,b)$, $a, b \in \mathbb{R}_+$, que relaciona las derivadas en cero de la función generatriz de momentos con los momentos de $X$. Es decir,

importante!

$$\begin{aligned} M_X(t) &= 1 + \frac{dM_X(0)}{dt}t + \frac{1}{2}\frac{d^2 M_X(0)}{dt^2}t^2 + t^2 \underbrace{e_2(t)}_{\text{error del desarrollo en Taylor}} \\ &= 1 + \frac{\sigma^2}{2}t^2 + t^2 e_2(t) = 1 + t^2\left[\frac{\sigma^2}{2} + e_2(t)\right], \quad \lim_{t\to 0} e_2(t) = 0. \end{aligned}$$

Se considera ahora la correspondiente aproximación para $M_{S_n^*}(t)$. Más concretamente,

$$M_{S_n^*}(t) = \left(1 + \left(\frac{t}{\sqrt{n}\sigma}\right)^2\left[\frac{\sigma^2}{2} + e_2\left(\frac{t}{\sqrt{n}\sigma}\right)\right]\right)^n \overset{\text{hago el producto y opero}}{=} \left(1 + \frac{t^2}{n}\underbrace{\left[\frac{1}{2} + \frac{e_2\left(\frac{t}{\sqrt{n}\sigma}\right)}{\sigma^2}\right]}_{\text{lo importante es ver que esto tiene límite finito}}\right)^n$$

Además,

$$t^2\left[\frac{1}{2} + \underbrace{\frac{e_2\left(\frac{t}{\sqrt{n}\sigma}\right)}{\sigma^2}}_{\to 0}\right] \to \frac{t^2}{2}, \quad n \to \infty.$$

Aplicando el Lema 6, fijado un $t$, siendo

$$c_n = t^2\left[\frac{1}{2} + \frac{e_2\left(\frac{t}{\sqrt{n}\sigma}\right)}{\sigma^2}\right], \quad c = \frac{t^2}{2}$$

se obtiene

$$M_{S_n^*}(t) \to \exp\left(\frac{t^2}{2}\right), \quad n \to \infty,$$

$M_X(t) = e^{t\mu + \frac{t^2\sigma^2}{2}}$ ↓ $N(0,1)$ $M_X(t) = e^{t^2/2}$

que corresponde a la función generatriz de momentos de una variable aleatoria normal con media cero y varianza uno. Por el teorema de continuidad para funciones generatrices de momentos se obtiene que

$$S_n^* \to^{L} Z \sim \mathcal{N}(0,1).$$