I: ESPACIOS DE PROBABILIDAD DISCRETOS

Definición: Un espacio de probabilidad es una terna $(\Omega, \mathcal{A}, P)$ donde $\Omega \neq \emptyset$, $\mathcal{A}$ es una σ-álgebra de Ω, y P una probabilidad en $\mathcal{A}$, ie, una medida en el espacio medible $(\Omega, \mathcal{A})$ tal que $P(\Omega) = 1$. Llamaremos a los elementos de $\mathcal{A}$ sucesos. Ω es el espacio de las observaciones.

- En particular notar que $P: \mathcal{A} \longrightarrow [0,1]$.

Proposición (Propiedades de P): Sea P una prob. en $(\Omega, \mathcal{A})$. Se cumple:

1) $P(A) + P(A^c) = 1, \qquad \forall A \in \mathcal{A}$

2) $P(\emptyset) = 0$.

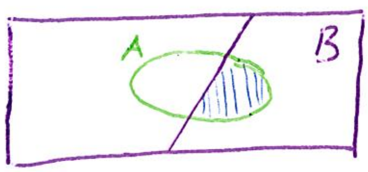
3) $A \subseteq B \implies P(A) \leq P(B)$, y en particular $P(B-A) = P(B) - P(A)$.

4) $P(A \cup B) + P(A \cap B) = P(A) + P(B)$

5) (Desigualdad de Bonferroni, σ-subaditividad): $P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$.

Definición. Sean $A, B \in \mathcal{A}$, $P(B) \neq 0$. Se llama probabilidad condicionada de A sabiendo que ha ocurrido B (ó respecto a B) a

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



- Notar que la aplicación $\begin{array}{l} P_A: \mathcal{A} \longrightarrow [0,1] \\ B \longmapsto P_A(B) := P(B|A) \end{array}$ es una probabilidad, centrada en el suceso A. Se llama probabilidad condicionada.

Definición: Dos sucesos $A, B \in \mathcal{A}$ se dicen **independientes** si $P(A \cap B) = P(A) \cdot P(B)$, es decir, si $P(A|B) = P(A)$. En general, $(A_i)_{i \in I} \subset \mathcal{A}$ se dicen independientes si $\forall J \subseteq I$ finito $P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$.

• ¡OJO!: Para que n sucesos $A_1, .., A_n$ sean indep, no basta con que $P(A_1 \cap \dots \cap A_n) = P(A_1) \cdots P(A_n)$; si no tienen que ser indep. 2 a 2, 3 a 3, ..., $n-1$ a $n-1$ y finalmente los n.

Teorema (de la probabilidad compuesta): Sean $A_1, .., A_n \in \mathcal{A} : P(A_1 \cap .. \cap A_n) \neq 0$. Entonces

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n | A_1 \cap \dots \cap A_{n-1})$$

Teorema (de la probabilidad total): Sean $A_1, .., A_n \in \mathcal{A} : \dot{\bigcup} A_i = \Omega$, $P(A_i) \neq 0$, $B \in \mathcal{A}$.

$$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B|A_i)$$

Teorema (Bayes): Sean $A_1, .., A_n \in \mathcal{A} : \dot{\bigcup} A_i = \Omega$, $P(A_i) \neq 0$, $B \in \mathcal{A}$, $P(B) \neq 0$.

$$P(A_k | B) = \frac{P(A_k) \cdot P(B|A_k)}{\sum_{i=1}^{n} P(A_i) \cdot P(B|A_i)}$$

Definición: Un espacio medible $(\Omega, \mathcal{A})$ se dice **discreto** si Ω es finito ó infinito numerable y $\mathcal{A} = \mathcal{P}(\Omega)$.

• Una probabilidad P en un espacio de prob. discreto queda determinada por una colección $\{p_\omega\}_{\omega \in \Omega}$ tal que $P(\{\omega\}) = p_\omega$. Es claro que $P(A) = \sum_{\omega \in A} p_\omega$ y $P(\Omega) = \sum_{\omega \in \Omega} p_\omega$.

COMBINATORIA: Consideremos el conjunto $\Omega = \{1, .., n\}$.

- Variaciones con repetición: $VR(n, k) = n^k$: Es el número de k-uplas $(a_1, \dots, a_k)$, con $a_i \in \{1, .., n\}$.

- Variaciones sin repetición y sin orden (o ordinarias): $VSR(n, k) = \frac{n!}{(n-k)!}$

Es el número de k-uplas $(a_1, .., a_k)$ con $a_i \in \{1, .., n\}$ pero distintos, i.e. que haya dos a_i, a_j iguales en la misma upla.

* CASO PARTICULAR: Permutaciones: $n = k$. : $|S_n| = n!$. Es el número de n-uplas $(a_1, .., a_n)$, con $a_i \in \{1, .., n\}$ y distintos, i.e., el número de reordenaciones de $\{1, .., n\}$.

- Variaciones sin repetición y orden (o combinaciones): $VSRO(n, k) = \binom{n}{k}$.

Es el número de k-uplas $(a_1, .., a_k)$ con $a_i \in \{1, .., n\}$, pero distintos y con $a_1 < a_2 < \dots < a_k$ (con orden). En otras palabras, es el número de subconjuntos $\{a_1, .., a_k\}$, (con $a_i \in \{1, .., n\}$), posibles, i.e., el número de combinaciones de n elementos tomados de k en k.

ESPACIOS DISCRETOS

- La definición clásica de probabilidad de Laplace es que, para sucesos elementales equiprobables, dentro un número finito,

$$P(A) = \frac{\text{casos favorables}}{\text{casos posibles}}.$$

¿Cuál es la descripción matemática de esto? Se trata de $\Omega = \{1, \dots, n\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ y $P(\{i\}) = \frac{1}{n}$.

__Definición__: En el espacio de probabilidad discreto $(\Omega = \{1, \dots, n\}, \mathcal{A} = \mathcal{P}(\Omega), P(\{i\}) = \frac{1}{n})$, P se dice que es la __distribución de probabilidad uniforme discreta__.

__Definición__. Consideremos el espacio medible $(\Omega = \{0, \dots, n\}, \mathcal{A} = \mathcal{P}(\Omega))$. Se llama __distribución de probabilidad binomial__, dados $n \in \mathbb{N}$ y $p \in (0,1)$, a la probabilidad sobre $(\Omega, \mathcal{A})$

$$b_n(p)(\{k\}) := \binom{n}{k} p^k (1-p)^{n-k}.$$

- La distribución binomial es la que siguen fenómenos aleatorios de __éxito - fracaso__: extracción de bolas __con reemplazamiento__, sacar 6 veces "2" al tirar 10 veces un dado, ...

__Definición__: En el espacio medible $(\Omega = \{1, \dots, N\}, \mathcal{A} = \mathcal{P}(\Omega))$, dados $N \in \mathbb{N}$, $n \leq N$, $0 \leq k \leq n$, y $C < N$, se llama __distribución de probabilidad hipergeométrica__ a

$$H(N, C, n)(\{k\}) := \frac{\binom{C}{k}\binom{N-C}{n-k}}{\binom{N}{n}}$$

• En la distribución hipergeométrica, N representa el tamaño de la población, n el tamaño de la muestra, C es la cantidad de elementos que cumplen la característica deseada y K es la cantidad de éxitos que queremos (la cantidad de individuos que queremos que cumplan la característica).

• Dicha distribución es la que se sigue en extracción de bolas sin reemplazamiento.

Definición: Se llama distribución de probabilidad de Poisson sobre el espacio ($\Omega = \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $\mathcal{A} = \mathcal{P}(\mathbb{N}_0)$), de parámetro $\lambda > 0$, y la denotamos con P_λ, a la que

$$P_\lambda(\{k\}) := e^{-\lambda} \frac{\lambda^k}{k!} .$$

• La distrib. de Poisson se basa en el conteo de las veces que se presenta un fenómeno dentro de un área de oportunidad dada (intervalo de tiempo, e.g.).

• A continuación se describen las generalizaciones de la distrib. hipergeométrica y binomial:

Definición: Sean $N \in \mathbb{N}$, $n \leq N$, $0 \leq k_1 \leq n_1, \dots, 0 \leq k_r \leq n_r$, $r \in \mathbb{N}$. Se llama distribución de probabilidad hipergeométrica múltiple, definida sobre el espacio medible ($\Omega = \{(k_1, \dots, k_r) \in \prod_{i=1}^{r} \{0, \dots, n_i\} : \sum_{i=1}^{r} k_i = n\}$, $\mathcal{A} = \mathcal{P}(\Omega)$) a

$$P(k_1, \dots, k_r) := \frac{\prod_{i=1}^{r} \binom{n_i}{k_i}}{\binom{N}{n}}$$

• Es la prob. de obtener k_1 bolas del primer color que tiene n_1 bolas, ..., k_r bolas del color r-ésimo que tiene n_r bolas; y $N = n_1 + \dots + n_r$, con reemplazamiento.

Definición: Sean $n, N \in \mathbb{N}$, $k_1 + \dots + k_r = n$, $p_i = \frac{n_i}{N}$. Se llama distribución de probabilidad multinomial $M(n; p_1, \dots, p_r)$, definida en $(\Omega = \{(k_1, \dots, k_r) \in \mathbb{N}_0^r : \sum k_i = n\}, \mathcal{A} = \mathcal{P}(\Omega))$, a la probabilidad

$$P(k_1, \dots, k_r) := \binom{n}{k_1, \dots, k_r} p_1^{k_1} \cdots p_r^{k_r},$$

donde $\binom{n}{k_1, \dots, k_r} = \frac{n!}{k_1! \cdots k_r!}$.

• Es la seguida en el ejemplo anterior pero sin reemplazamiento.

II. VARIABLES ALEATORIAS

<u>Definición</u>. Sean $(\Omega, \mathcal{A})$ y $(\Omega', \mathcal{A}')$ dos espacios medibles. En probabilidad, llamaremos <u>variables aleatorias</u> (v.a.) a las funciones medibles $X : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$, i.e., funciones tales que $X^{-1}(\mathcal{A}') \subseteq \mathcal{A}$. Si $(\Omega', \mathcal{A}') = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n) = \mathcal{B}_n)$, diremos que X es una v.a. n-dimensional ó que es una función <u>Borel-medible</u>. En el caso $n=1$, X se dice <u>variable aleatoria real</u> (v.a.r.).

<u>Proposición</u>: Se cumple

1) Toda función cte es medible
2) La composición de medibles es medible
3) Las aplicaciones continuas $X : (\Omega, \mathcal{B}(\Omega)) \longrightarrow (\Omega', \mathcal{B}(\Omega'))$ son medibles.
4) Las funciones crecientes $\mathbb{R} \longrightarrow \mathbb{R}$ son Borel-medibles
5) Las proyecciones de un espacio producto en sus factores (con la σ-alg. producto) son medibles.
6) $(\Omega, \mathcal{A}) \xrightarrow{X = (X_1, .., X_n)} \mathbb{R}^n$ es medible $\iff$ X_i medible $\forall i$.

<u>Definición</u>: Sea $(\Omega, \mathcal{A}, P)$ un esp. de prob, y $X : \Omega \to \mathbb{R}$ Borel-medible. Llamaremos <u>esperanza ó media</u> de X a $E_p(X) = E(X) = EX := \int X \, dP$, cuando existe.

• Delta de Dirac: Sea $(\Omega, \mathcal{A})$ medible, y $\omega \in \Omega$. La función de conjuntos

$$\delta_\omega : \mathcal{A} \longrightarrow [0, \infty]$$
$$A \longmapsto \delta_\omega(A) := \begin{cases} 1 & , \text{si } \omega \in A \\ 0 & \quad \notin \end{cases}$$

es una probabilidad, llamada probabilidad de Dirac ó degenerada. Si $X : \Omega \to \mathbb{R}$ v.a.,

$\int f \, d\delta_\omega = f(\omega)$.

• Medida de contar (ó cardinal): En $(\Omega, \mathcal{A})$ espacio medible, $A \in \mathcal{A}$, la función $\mu(A) := \# A$

es la medida de contar. Si $\Omega = \mathbb{N}$, las funciones $f : \mathbb{N} = \Omega \to \mathbb{R}$ se identifican con las

sucesiones (x_n) de dicho espacio, con la medida de contar, se extraen las funciones p-integrables

$\ell_p \equiv$ sucesiones p-sumables. Para $p = 1$, $f : \mathbb{N} \to \mathbb{R}$ μ-int $\Leftrightarrow \sum_{n=1}^{\infty} |f(n)| < \infty$, y

en este caso $\int_{\mathbb{N}} f \, d\mu = \sum f(n)$.

• Si μ es una medida sobre $(\Omega, \mathcal{A})$, y $c > 0$, $c\mu$ es otra medida; $(c\mu)(A) = c \cdot \mu(A)$

y $\int f \, d(c\mu) = c \int f d\mu$.

• Si (μ_n) es una suc. de medidas sobre $(\Omega, \mathcal{A})$, $\sum_n \mu_n$ es otra medida: $(\sum_n \mu_n)(A) := \sum_n \mu_n(A)$ y

$\int f \, d(\sum_n \mu_n) = \sum_n \int f \, d\mu_n$.

• Si (P_n) suc. de probabilidades sobre $(\Omega, \mathcal{A})$ y (c_n) suc. de números reales tales que $\sum c_n = 1$, entonces

$P := \sum_n c_n P_n$ es una probabilidad, y si X v.a.r, $E_P(X) = \sum_n c_n E_{P_n}(X)$.

• Si μ es una medida en $(\Omega, \mathcal{A})$, y fijamos $A \in \mathcal{A}$, entonces

$$\begin{aligned} \mu_{|A} : \mathcal{A} &\longrightarrow [0, \infty] \\ B &\longmapsto \mu_{|A}(B) := \mu(A \cap B) \end{aligned} \qquad \mu_{|A} = \mu(A \cap \cdot)$$

es una medida en $\mathcal{A}$, y $\int f \, d\mu_{|A} = \int_A f \, d\mu$.

• Notar que la probabilidad condicionada $P_A(B) := P(B|A)$ verifica $P_A = \frac{1}{P(A)} P_{|A}$. Así,

<u>Definición</u>: Sea $(\Omega, \mathcal{A}, P)$ esp. prob., $A \in \mathcal{A}$ no nulo, y X una v.a. Se llama <u>esperanza condicional de X respecto A</u> a

$$E_P(X|A) := E_{P_A}(X) = \frac{1}{P(A)} \int_A X \, dP.$$

<u>IMPORTANTE</u>: $E(X)$ se interpreta como el valor medio, el valor esperado de X en una larga serie de pruebas del experimento; y $E(X|A)$ lo mismo pero solo teniendo en cuenta aquellas en las que ha ocurrido A.

<u>Teorema</u> (de la medida imagen). Sea $(\Omega, \mathcal{A}, \mu)$ un esp. medida y $X : (\Omega, \mathcal{A}, \mu) \to (\Omega', \mathcal{A}')$ una v.a. Podemos definir una medida en $(\Omega', \mathcal{A}')$ "llevándonos μ a $(\Omega', \mathcal{A}')$", definiendo

$$\begin{aligned} \mu^X : \mathcal{A}' &\longrightarrow [0, \infty] \\ A' &\longmapsto \mu^X(A') := \mu\left(X^{-1}(A')\right) \end{aligned}$$

es en efecto una medida en $(\Omega', \mathcal{A}')$, llamada <u>medida imagen de μ por X</u>. Si $f : (\Omega', \mathcal{A}') \to \mathbb{R}$,

$$\int f \, d\mu^X = \int f \circ X \, d\mu$$

Definición: Sea $(\Omega, \mathcal{A}, P)$ un esp. prob, y $X : (\Omega, \mathcal{A}, P) \longrightarrow (\Omega', \mathcal{A}')$. Llamaremos distribución de probabilidad de X a la medida imagen P^X.

• El TMI dice que $E_{P^X}(f) = E_P(f \circ X)$.

Definición: Sea X v.a.r. en $(\Omega, \mathcal{A}, P)$. Se llama función de distribución de X a

$$F_X : \mathbb{R} \longrightarrow [0,1]$$
$$x \longmapsto F_X(x) := P(X \leq x) = P^X((-\infty, x]).$$

• F_X es creciente y cont. a la derecha. En particular, F_X es la func. de distribución de la medida de Lebesgue-Stieltjes P^X.

Definición: Sea X v.a.r. Se llama varianza de X a $\operatorname{Var} X := E[(X - E(X))^2]$.

• $\operatorname{Var} X = E(X^2) - [E(X)]^2$.

Teorema (de la Carga): Sea μ una medida en $(\Omega, \mathcal{A})$, y $h : \Omega \to \mathbb{R}$, $h \geq 0$. Entonces $\lambda(A) := \int_A h \, d\mu$ es una medida sobre $\mathcal{A}$, y $\lambda \ll \mu$ (λ la func. de cjtos sobre $\mathcal{A}$ se dice abs. cont. resp. μ ó que λ está dominada por μ). Dicha h recibe el nombre de derivada RN de λ resp. μ ó densidad de λ resp μ.

Definición: Sea $X:(\Omega,\mathcal{A},P) \longrightarrow (\Omega',\mathcal{A}',\mu')$. Se dice que X tiene densidad h respecto μ' si h es una derivada RN (la densidad) de P^X resp. μ', ie, $h = \frac{dP^X}{d\mu'}$. Si $X:(\Omega,\mathcal{A},P) \longrightarrow \mathbb{R}$ es una v.a.r., se dice que X tiene densidad h si tiene densidad h resp. la medida de Lebesgue m, ie, si $h = \frac{dP^X}{dm}$. Se dice que X es absolutamente continua. Se dice también que h es la función de densidad de X

• CASO V. CONT:

• En el caso en el que X v.a.r. tenga densidad h,

$$E(X) = \int_{-\infty}^{\infty} x\, h(x)\, dx \quad , \quad Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 h(x)\, dx \,,$$

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} h(t)\, dt \,.$$

Definición: Sea $(\Omega,\mathcal{A},P)$ esp. de prob. Se llama función de probabilidad a

$$\begin{aligned} f:\ &\Omega \longrightarrow [0,1] \\ &\omega \longmapsto f(\omega) := P(\omega) \end{aligned}$$

CASO DISCRETO

Si $X:(\Omega,\mathcal{A},P) \longrightarrow \mathbb{N}_0 = \mathbb{N}\cup\{0\}$ es una v.a., su distribución P^X (en $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$) queda determinada por su función de probabilidad

$$\begin{aligned} f:\ &\mathbb{N}_0 \longrightarrow [0,1] \\ &n \longmapsto f(n) = P^X(n) = P(X=n). \end{aligned}$$

Si μ es la medida de contar, $f = \frac{dP^X}{d\mu}$ (f es la densidad de X resp. μ), pq si $A \subseteq \mathbb{N}_0$, $P^X(A) = \sum_{n\in A} f(n) = \int_A f\, d\mu$. Además,

$$E(X) = \sum_{n=0}^{\infty} n f(n) \,, \quad Var\, X = \sum_{n=0}^{\infty} (n - EX)^2 f(n) \,, \quad F_X(x) = \begin{cases} f(0)+\dots+f(n) & x\in[n,n+1) \\ 0\,, & x<0 \end{cases}$$

Teorema (de cambio de variables): Sean U, V abiertos de $\mathbb{R}^n$ y $T: U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^n$ un difeomorfismo.

Sea $J_T = \det\left(\frac{\partial T_i}{\partial x_j}\right)$ el Jacobiano de T. Entonces, para $B \in \mathcal{B}(U)$,

$$m(T(B)) = \int_B |J_T| \, dm \,. \qquad \left(\text{ie, } |J_T| = \frac{dm^{T^{-1}}}{dm}\right)$$

y si $f: V \to \mathbb{R}$ Borel-medible,

$$\int_V f \, dm = \int_U (f \circ T) \, |J_T| \, dm \,.$$

Teorema (de transformación de variables): Sean U, V abiertos de $\mathbb{R}^n$ y $T: U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^n$ ~~un~~ difeomorfismo. Sea $X: (\Omega, \mathcal{A}, P) \to U$ una v.a. (n-dimensional) de densidad f_X (ie, $f_X = \frac{dP^X}{dm}$). Entonces la v.a. $Y := T \circ X$ tiene densidad

$$f_Y(y) = \frac{dP^{T \circ X}}{dm} = (f \circ T^{-1})(y) \cdot |J_{T^{-1}}(y)| \,.$$

y $\forall B \in \mathcal{B}(V)$,

$$m^T(B) = m(T^{-1}(B)) = \int_{T^{-1}(B)} dx = \int_B |J_{T^{-1}}(y)| \, dy \,.$$

ie, $|J_{T^{-1}}| = \frac{dm^T}{dm}$.

Teorema (clásico de la medida producto): Sean $(\Omega_1, \mathcal{A}_1, \mu_1)$, $(\Omega_2, \mathcal{A}_2, \mu_2)$ dos espacios de medida σ-finitos. Entonces existe una única medida μ en $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ tal que $\forall E \in \mathcal{A}$,

$$\mu(E) = \int \mu_2(E_{\omega_1})\, d\mu_1 = \int \mu_1(E_{\omega_2})\, d\mu_2$$

y para los productos de medibles $\mu(A_1 \times A_2) = \mu_1(A_1)\, \mu_2(A_2)$. Además μ es σ-finita, y es una probabilidad si μ_1 y μ_2 lo son.

Teorema (clásico de FUBINI): Si $f: \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ es Borel medible y existe $\int f\, d\mu$,

$$\int f\, d\mu = \int_{\Omega_1} \left(\int f\, d\mu_2 \right) d\mu_1 = \int_{\Omega_2} \left(\int f\, d\mu_1 \right) d\mu_2$$

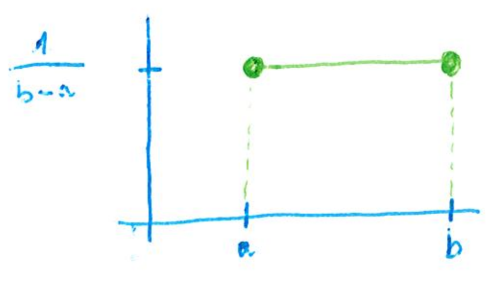
y las integrales iteradas existen c.s.

DISTRIBUCIONES IMPORTANTES

Definición: Sea $[a,b] \subseteq \mathbb{R}$. Se llama distribución uniforme continua sobre $[a,b]$ a la probabilidad en $\mathbb{R}$ que queda definida por la función de densidad

$$f(x) = \frac{1}{b-a} I_{[a,b]}(x)$$

(en general, con $B \in \mathcal{B}_1$, $m(B) > 0$, la distr. uniforme es la prob. def. por la func. de densidad $f(x) = \frac{1}{m(B)} I_B$).



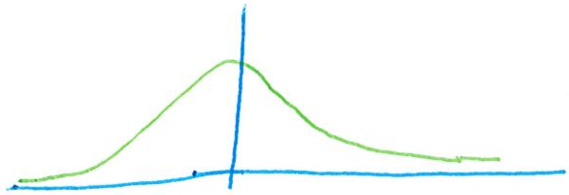
Definición: Se dice que una v.a.r. X tiene distribución de probabilidad normal

$P^X = N(\mu, \sigma^2)$, donde μ es la media y σ^2 la varianza; si su función de

densidad es
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}.$$

- La distribución normal estándar $N(0,1)$ es una probabilidad que queda definida en $(\mathbb{R}, \mathcal{B}_1)$ mediante su densidad resp. la medida de Lebesgue

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



"Campana de Gauss".

Definición: Una v.a.r. X tiene distribución gamma $X \sim G(\alpha, \beta)$ de parámetros $\alpha, \beta > 0$ si tiene como función de densidad

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} e^{-\frac{y}{\beta}} I_{(0,\infty)}(y)$$

- Si $\alpha \in \mathbb{Z}$, $G(\alpha, \beta)$ se llama distribución de Erlang
- $G(1, \beta)$ es la distribución exponencial de parámetro β, $\mathcal{E}(\beta)$, $f(x) = \frac{1}{\beta} e^{-y/\beta} I_{(0,\infty)}(y)$.
- Si $n \in \mathbb{N}$, y $X \sim G(\frac{n}{2}, 2)$, se dice que X tiene distribución chi-cuadrado con n grados de libertad, $X \sim \chi^2(n)$.

MOMENTOS Y DESIGUALDADES

Definición: Sea $X : (\Omega, \mathcal{A}, P) \longrightarrow \mathbb{R}$ v.a.r. y $k > 0$. Llamamos <u>momento de orden k</u> a $E(X^k)$, <u>momento absoluto de orden k</u> a $E(|X|^k)$ y <u>momento central de orden k</u> a $E[(X - EX)^k]$.

- La media es el momento de orden 1 de X, y la varianza el momento central de orden 2 de X.

Proposición:

1) Si X tiene momento de orden $k < \infty$ $\Rightarrow$ tiene momento de orden j, $0 < j \leq k$.

2) Si X tiene momento finito de orden $n-1$, y existe el momento de orden n, entonces el momento central de orden n es

$$E[(X - EX)^n] = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} E(X^k) E(X)^{n-k}.$$

Proposición (desigualdad de Markov): Si X v.a.r., $k, \alpha > 0$,

$$P(|X| \geq \alpha) \leq \alpha^{-k} E(|X|^k)$$

Proposición (Desigualdad de Chebyschev): Sea X v.a.r. con media μ finita y varianza σ^2 finita, y $k > 0$. Entonces

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Es decir, la prob. de encontrar valores de X que disten menos de la media más que k desviaciones típicas es $\leq \frac{1}{k^2}$.

<u>Proposición</u> (Desigualdad de Cauchy-Schwarz): Sean X, Y var 2-integrables en $(\Omega, \mathcal{A}, P)$. Ent.

$$E(|XY|) \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}.$$

<u>Definición</u>: Sean $X, Y \in \mathcal{L}(\Omega, \mathcal{A}, P)$. Se llama <u>covarianza</u> de X e Y a

$$Cov(X,Y) := E[(X-EX)(Y-EY)].$$

<u>Definición</u>: Llamamos <u>coeficiente de correlación (de Pearson)</u> entre X e Y a

$$\rho(X,Y) := \frac{Cov(X,Y)}{\sqrt{Var X}\sqrt{Var Y}}$$

- La desigualdad CS prueba que $\rho(X,Y)$ está bien def. ($\exists$ la integral)
- $Cov(X,Y) = E(XY) - E(X)E(Y)$
- $\rho(X,Y) \in [-1,1]$
- $Cov(aX+b, cY+d) = ac\, Cov(X,Y)$; $\rho(aX+b, cY+d) = \rho(X,Y)$.

<u>Definición</u>: Sea $X = (X_1, .., X_n)$ un v.a. n-dim. en $(\Omega, \mathcal{A}, P)$, $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$.
Se llama <u>matriz de covarianzas</u> de X a

$$Cov\, X := \left(Cov(X_i, X_j) \right)_{i,j=1,..,n} = E\left[(X-EX)(X-EX) \right]$$

Lema: Sea X v.a.r. con densidad f, e $Y := a + bX$, $(b \neq 0)$. Entonces Y tiene densidad $\frac{1}{|b|} f\left(\frac{y-a}{b}\right)$.

Proposición:

1) $X \sim N(\mu, \sigma^2) \Rightarrow a + bX \sim N(a + b\mu, b^2\sigma^2)$ $(a, b \in \mathbb{R}, b \neq 0)$.

2) $Z \sim N(0,1) \Rightarrow \mu + \sigma Z \sim N(\mu, \sigma^2)$, $(\mu \in \mathbb{R}, \sigma > 0)$

3) $X \sim N(\mu, \sigma^2) \Rightarrow \frac{X-\mu}{\sigma} \sim N(0,1)$. A dicho hecho se le conoce como tipificación de $N(\mu, \sigma^2)$.

Definición: Una clase $\mathcal{C} \subset \mathcal{P}(\Omega)$ se dice π-sistema si $A_1, \dots, A_n \in \mathcal{C} \Rightarrow \bigcap_{i=1}^{n} A_i \in \mathcal{C}$.

Definición: Una clase $\mathcal{D}$ se dice d-sistema si

i) $\Omega \in \mathcal{D}$

ii) $D_1, D_2 \in \mathcal{D}$, $D_1 \subset D_2 \Rightarrow D_2 - D_1 \in \mathcal{D}$

iii) $D_1 \subset D_2 \subset D_3 \subset \dots$ suc. creciente en $\mathcal{D}$ $\Rightarrow \bigcup_n D_n \in \mathcal{D}$.

Teorema (Dinkyn): Si $\mathcal{C}$ es un π-sistema en Ω, el d-sistema engendrado por $\mathcal{C}$ (luego que hacer), i.e, (el menor d-sistema que contiene a $\mathcal{C}$) coincide con la σ-álgebra generada por $\mathcal{C}$

$$d(\mathcal{C}) = \sigma(\mathcal{C}).$$

Teorema: Sean $(\Omega_i, \mathcal{A}_i, P_i)$ esp. de prob. y consideremos $(\Omega = \prod \Omega_i, \mathcal{A} = \bigotimes \mathcal{A}_i, P = \prod P_i)$ el esp. de prob. producto. Entonces, las proyecciones $\pi_i : (\Omega, \mathcal{A}, P) \longrightarrow (\Omega_i, \mathcal{A}_i, P_i)$, $\pi_i(\omega) = \omega_i$ son independientes y $P^{\pi_i} = P_i$.

Teorema: Sean $(\Omega_n, \mathcal{A}_n)$ esp. medibles y $X_n : (\Omega, \mathcal{A}, P) \longrightarrow (\Omega_n, \mathcal{A}_n)$ v.a. Para cada $n \in \mathbb{N}$, sea la v.a. $Y_n : (\Omega^{\mathbb{N}} := \prod_{i=1}^{\infty} \Omega, \mathcal{A}^{\mathbb{N}} = \bigotimes_{i=1}^{\infty} \mathcal{A}, P^{\mathbb{N}} = \prod P) \longrightarrow (\Omega_n, \mathcal{A}_n)$, $Y_n(\omega_1, \omega_2, \dots) := X_n(\omega_n)$. Entonces Y_n tiene respecto $P^{\mathbb{N}}$ la misma distribución que X_n resp. P y las Y_n son indep.

Teorema (Corol. de Indep. con densidades): Sean $X_i : (\Omega, \mathcal{A}, P) \longrightarrow (\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1, \dots, n$ v.a., y μ_i una medida σ-finita. Sean $X = (X_1, \dots, X_n)$ y $\mu = \mu_1 \times \dots \times \mu_n$.

1) Si $f = \dfrac{dP^X}{d\mu} \Rightarrow X_i$ tiene resp. μ_i densidad

$$\frac{dP^{X_i}}{d\mu_i} = f_i(\omega_i) = \int_{\prod_{j \neq i} \Omega_j} f(\omega_1, \dots, \omega_n) \, d\Big(\prod_{j \neq i} \mu_j\Big)$$

2) $\left[\cancel{\text{Si } f_i = \dfrac{dP^{X_i}}{d\mu_i}, \text{ entonces}}\right]$ (X_i) son indep. $\Leftrightarrow f = \prod_{i=1}^{n} f_i$.

Definición: Dichas f_i se llaman densidades marginales i-ésimas.

Definición: Dos v.a. X, Y se dicen incorreladas si $Cov(X,Y) = 0$.

Proposición: Si $X_1, \ldots X_n$ son v.a. reales ó complejas integrables independientes, et.

$$E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n)$$

y si tienen varianza finita,

$$Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n).$$

Proposición: X, Y v.a.r. independientes (e integrables) $\Rightarrow$ X, Y incorreladas.

¡OJO!: En general $\not\Leftarrow$.

Proposición: Si $X_1, \ldots, X_n$ v.a. independientes e idénticamente distribuidas, con distribución común $b_1(p)$ de Bernoulli, de parámetro p $\Rightarrow$ La v.a. $S_n := \sum X_i$ sigue una distribución binomial $b_n(p)$.

IMPORTANTE: X, Y iid $\Rightarrow E[X] = E[Y]$ y $Var[X] = Var[Y]$

IV: ESPERANZA CONDICIONAL

Teorema (Radon-Nikodym): Sean λ, μ medidas σ-finitas sobre $(\Omega, \mathcal{A})$.

$$\lambda << \mu \iff \dot{\exists} (\mu\text{-c.s.})\ h : \Omega \to \mathbb{R},\ h \geq 0 : \lambda(A) = \int_A h \, d\mu \quad \forall A \in \mathcal{A}.$$

Teorema (Existencia de la Esperanza Condicional): Sea $X:(\Omega, \mathcal{A}, P) \to (\Omega', \mathcal{A}', P')$ v.a., $Y:(\Omega, \mathcal{A}) \to \mathbb{R}$, si $Y \geq 0$ (ó $E(Y) < \infty$) $\Rightarrow \dot{\exists} (P^X\text{-c.s.})\ g:(\Omega', \mathcal{A}') \to \mathbb{R}$, $g \geq 0$ (ó $E(g) < \infty$) :

$$\int_{X^{-1}(A')} Y \, dP = \int_{A'} g \, dP^X$$

$$\begin{tikzcd}
(\Omega, \mathcal{A}) \arrow[r, "X"] \arrow[dr, "Y"'] & (\Omega', \mathcal{A}') \arrow[d, dashed, "\exists g"] \\
& \mathbb{R}
\end{tikzcd}$$

Definición: Llamaremos <u>esperanza condicional de Y respecto X</u> a dicha g del Th, y la denotaremos como $E(Y|X)$.

- Notar que $E(Y|X)$ es, más bien, una clase de equiv. de funciones que coinciden P^X-c.s.

Definición: Se llama <u>probabilidad condicional de A respecto X</u> a $P(A|X) := E(I_A|X)$

- Recordar que ya sabías que $E(X|A) = \frac{1}{P(A)} \int_A X \, dP$.

Entonces $g(x) = E(Y|X)(x) \overset{NOT}{=} E(Y|X=x)$.

Definición: Si $Y = (Y_1, \ldots, Y_n)$ v.a. n-dim, definimos $E(Y|X) := \big(E(Y_1|X), \ldots, E(Y_n|X)\big)$.

(ó ...

Definición: Sea $\mathcal{B}$ una sub-σ-álgebra de $\mathcal{A}$, $Y: \Omega \to \mathbb{R}$, $Y \geq 0$. Llamaremos **esperanza condicional de Y respecto $\mathcal{B}$**, y la denotaremos $E(Y|\mathcal{B})$, a toda función $\Omega \to \mathbb{R}$ ≥ 0 (ó $E(\ )<\infty$): $\mathcal{B}$-medible

$$\int_B Y \, dP = \int_B E(Y|\mathcal{B}) \, dP \quad \forall B \in \mathcal{B}$$

$$\begin{tikzcd}
(\Omega, \mathcal{A}) \arrow[dr, "Y"] & \\
(\Omega, \mathcal{B}) \arrow[r, "E(Y|\mathcal{B})"'] & \mathbb{R}
\end{tikzcd}$$

Definición: Se llama esperanza condicional de $A \in \mathcal{A}$ respecto $\mathcal{B}$, a $P(A|\mathcal{B}) := E(I_A|\mathcal{B})$.

- Nota: La existencia de $E(Y|\mathcal{B})$ queda asegurada por el Th Exist.

- $\int_B P(A|\mathcal{B}) \, dP = P(A \cap B)$.

- La esperanza condicional recupera la def clásica de prob. condicional $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

- Y v.a.r. $\mathcal{B}$-medible $\Rightarrow$ $E(Y|\mathcal{B}) = Y$

- $f: (\Omega', \mathcal{A}') \to \mathbb{R}$ v.a.r. ≥ 0 ó P^X-int. $\Longrightarrow$ $E(f \circ X | X) = f$.

Proposición: Sea $X: (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ y $\mathcal{B} = X^{-1}(\mathcal{A}')$ una sub-σ-álg. de $\mathcal{A}$, y $h: \Omega \to \mathbb{R}$.

$$h: (\Omega, \mathcal{B}) \to \mathbb{R} \text{ medible} \iff \exists f: (\Omega', \mathcal{A}') \to \overline{\mathbb{R}} \text{ v.a.} : h = f \circ X$$

$$\begin{tikzcd}
(\Omega, \mathcal{A}) \arrow[r, "X"] & (\Omega', \mathcal{A}') \arrow[d, "f"] \\
(\Omega, \mathcal{B}) \arrow[u, phantom, "\cup" description] \arrow[r, "h"'] & \mathbb{R}
\end{tikzcd}$$

Corolario: Sea $X: (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$, $Y: (\Omega, \mathcal{A}) \to \mathbb{R}$ v.a.r ≥ 0 (ó $E(Y) < \infty$), y $\mathcal{B} = X^{-1}(\mathcal{A}')$. Entonces la aplicación $g \in E(Y|X) \longrightarrow g \circ X \in E(Y|\mathcal{B})$ es epiyectiva (g se entiende un representante de su clase, y $g \circ X$ análogo). Por lo cual

$$E(Y|\mathcal{B}) = E(Y|X) \circ X.$$

Proposición (Propiedades): Sean Y, Y_1, Y_2 v.a.r. ≥ 0 (ó con esperanza finita).

1) $E_P(Y) = E_{P^X}[E(Y|X)]$.

2) (Monotonía): $Y_1 \leq Y_2 \Rightarrow E(Y_1|X) \leq E(Y_2|X)$ P^X-c.s.

3) (Linealidad): $E(a_1 Y_1 + a_2 Y_2|X) = a_1 E(Y_1|X) + a_2 E(Y_2|X)$ P^X-c.s, $a_1, a_2 \in \mathbb{R}$.

Teorema (de la Convergencia Monótona para la Esp. Cond.): Sea $(Y_n) \uparrow Y$ una suc. creciente de v.a.r. $\Omega \to \mathbb{R}$, Entonces $(E(Y_n|X))$ es una suc. de v.a.r. crecientes que convergen a $E(Y|X)$.

Corolario:

1) Si $(Y_n) \geq 0$ $\Rightarrow$ $E(\sum Y_n|X) = \sum E(Y_n|X)$, P^X-c.s.

2) $(A_n) \subset \mathcal{A}$ disjuntos $\Rightarrow$ $P(\dot{\bigcup_n} A_n|X) = \sum_n P(A_n|X)$, P^X-c.s.

Teorema (de la Convergencia Dominada para la Esp. Cond.): Sean (Y_n) v.a.r $\Omega \to \mathbb{R}$. Si existe una v.a.r. $Z: \Omega \to \mathbb{R}$ con esperanza finita: $|Y_n| \leq Z$ $\forall n$ (P-c.s), y $Y_n \xrightarrow[P]{P\text{-c.s}} Y$

$\Rightarrow E(Y_n|X) \xrightarrow[P]{P^X\text{-c.s.}} E(Y|X)$.

Teorema: $Y, Z: (\Omega, \mathcal{B}) \to \mathbb{R}$ v.a. Si $Y, YZ \in \mathcal{L}^1$,

1) Sea $\mathcal{B} \subseteq \mathcal{A}$ sub-σ-álgebra, $Y: (\Omega, \mathcal{A}) \to \mathbb{R}$,

$$E(YZ|\mathcal{B}) = Z \cdot E(Y|\mathcal{B}) \qquad P\text{-c.s.}$$

2) $f: (\Omega', \mathcal{A}') \to \mathbb{R}$ v.a., $Y, Y(f \circ X)$ son v.a.r. $\in \mathcal{L}^1$,

$$E(Y(f \circ X)|X) = f \cdot E(Y|X) \qquad P^X\text{-c.s.}$$

__Proposición__: Sea $X : (\Omega, \mathcal{A}, P) \longrightarrow (\Omega', \mathcal{A}' \text{ discreto}, P^X)$, $Y : (\Omega, \mathcal{A}, P) \to \mathbb{R}$ v.a.r. Entonces

$$E(Y \mid X = x) = \frac{1}{P(X = x)} \int_{\{X = x\}} Y \, dP .$$

__Proposición__: Si X, Y son v.a.r. independientes, $E(XY \mid X = t) = E(X \mid X = t) \cdot E(Y \mid X = t)$.

V: DISTRIBUCIÓN CONDICIONAL

Definición: Sea X v.a., Y v.a.r. como en el diag. Se llama función de distribución condicional regular de Y resp. X a toda aplicación $F : \Omega' \times \mathbb{R} \to \mathbb{R}$ que verifique:

$$\begin{tikzcd}(\Omega, \mathcal{A}, P) \arrow[r, "X"] \arrow[d, "Y"'] & (\Omega', \mathcal{A}', P^X) \\ \mathbb{R} & \end{tikzcd}$$

i) $\forall x \in \Omega'$, $F(x, \cdot)$ es una fun. de distribución de $\mathbb{R}$

ii) $\forall y \in \mathbb{R}$, $F(\cdot, y)$ es una versión de $P(Y \leq y \mid X)$.

Teorema (Existencia de la función de distrib. cond. reg.): Para cualquier v.a.r. $Y : (\Omega, \mathcal{A}, P) \to \mathbb{R}$ existe una función de distribución condicional regular de Y resp. X.

Definición: Sean X, Y v.a. como en el diag. Se llama probabilidad condicional regular de Y resp. X a toda aplicación $Q \equiv P^{Y|X} : \Omega' \times \mathcal{A}_1 \to [0,1]$ que verifique:

$$\begin{tikzcd}(\Omega, \mathcal{A}, P) \arrow[r, "X"] \arrow[d, "Y"'] & (\Omega', \mathcal{A}', P^X) \\ (\Omega_1, \mathcal{A}_1, P^{Y|X=x}) & \end{tikzcd}$$

i) $\forall x \in \Omega'$, $Q(x, \cdot) = P^{Y|X=x}$ es una probabilidad en $(\Omega_1, \mathcal{A}_1)$, que recibe el nombre de distribución condicional de Y resp. X

ii) $\forall A_1 \in \mathcal{A}_1$, $Q(\cdot, A_1) = P^{Y|X=\cdot}(A_1)$ es una versión de $P(Y \in A_1 \mid X)$.

Notar que una prob. cond. regular es en particular una medida de transición.

Teorema (Existencia de una probabilidad condicional regular): Para cualquier v.a.r. $Y : (\Omega, \mathcal{A}, P) \to \mathbb{R}$ existe una probabilidad condicional regular de Y resp. X.

PROBLEMA DE REGRESIÓN

<u>Teorema</u>: Sea $1 \le p < \infty$, $Y \in \mathcal{L}_p(\Omega, \underline{\mathcal{A}}, P; \mathbb{R})$ y $\mathcal{B} \subseteq \mathcal{A}$ sub-σ-álgebra.

$\Rightarrow E(Y|\mathcal{B}) \in \mathcal{L}_p(\Omega, \mathcal{A}, P; \mathbb{R})$ y $\|E(Y|\mathcal{B})\|_p \le \|Y\|_p$. Además para $p=2$,

$$\|Y - E(Y|\mathcal{B})\|_2 \le \|Y - Z\|_2 \qquad \forall Z \in \mathcal{L}_2(\Omega, \underline{\mathcal{B}}, P; \mathbb{R})$$

• <u>Problema general de regresión</u>: Dados

$$\begin{tikzcd} & (\Omega', \mathcal{A}') \arrow[d, dashed, "¿\exists f?"] \\ (\Omega, \mathcal{A}, P) \arrow[ur, "X"] \arrow[r, "Y"'] & \mathbb{R} \end{tikzcd}$$

¿$\exists f : Y = f \circ X$? En general <u>no</u>! Entonces, ¿cuál es la mejor $f : (\Omega', \mathcal{A}') \to \mathbb{R}$ tal que $Y \simeq f \circ X$ en el sentido de los mínimos cuadrados, i.e., ¿cuál es la f que hace mínima la expresión $\|Y - f \circ X\|_2$?

Si $\mathcal{B} = X^{-1}(\mathcal{A}')$, $E(Y|\mathcal{B}) = E(Y|X) \circ X$, y el Th anterior dice que

$$\|Y - E(Y|X) \circ X\|_2 \le \|Y - f \circ X\|_2 \qquad \forall f \in \mathcal{L}^2$$

Si X es v.a.r., se llama <u>curva de regresión</u> a con la cual $\underline{f = E(Y|X)}$.

$$E(Y|X) : \mathbb{R} \longrightarrow \mathbb{R}$$
$$x \longmapsto E(Y|X=x).$$

FÓRMULAS IMPORTANTES

I :

$$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B|A_i) \quad , \quad P(A_k|B) = \frac{P(A_k)\cdot P(B|A_k)}{\sum P(A_i)\cdot P(B|A_i)}$$

$$b_{n(p)}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k} \quad , \quad H(N,C,n)(\{k\}) = \frac{\binom{C}{k}\binom{N-C}{n-k}}{\binom{N}{n}}$$

$$P(k_1,\dots,k_r) = \binom{n}{k_1,\dots,k_r} p_1^{k_1}\cdots p_r^{k_r} \quad , \quad P(k_1,\dots,k_r) = \frac{\prod \binom{m_i}{k_i}}{\binom{N}{n}}$$

II :

CASO ABS. CONT.:
$X:(\Omega,\mathcal{A},P) \to \mathbb{R}$ v.a.r.
$h = \dfrac{dP^X}{dm}$

$$E(X) = \int_{\mathbb{R}} x\, h(x)\, dx \quad , \quad Var\, X = \int_{\mathbb{R}} (x - EX)^2 h(x)\, dx ,$$

$$F_X(x) = \int_{-\infty}^{x} h(t)\, dt .$$

CASO DISCRETO
$X:(\Omega,\mathcal{A},P) \to \mathbb{N}_0$ v.a.
f func. de p.d.: $f(n) = P(X=n) = P^X(n)$
$f(\cdot) = \dfrac{dP^X}{d\mu} = P^X(\{\cdot\})$ ← med. de contar

$$E(X) = \sum_{n=0}^{\infty} n\, f(n) \quad , \quad Var\, X = \sum_{n=0}^{\infty} (n - EX)^2 f(n) ,$$

$$F_X(x) = \begin{cases} f(0) + \dots + f(n) & , \ x \in [n, n+1) \\ 0 & \ x < 0 \end{cases} .$$

Th (Transf. Var): Sea $(\Omega,\mathcal{A},P) \xrightarrow{X} U \subseteq \mathbb{R}^n \xrightarrow{T} V \subseteq \mathbb{R}^n$, $Y = T\circ X$. f_X dens. de X

$$f_X = \frac{dP^X}{dm} \Rightarrow \boxed{f_Y = \frac{dP^{T\circ X}}{dm} = (f \circ T^{-1}) \cdot |J_{T^{-1}}|}$$

III :

Th : $X=(X_1,\dots,X_n) : (\Omega,\mathcal{A},P) \longrightarrow (\prod \Omega_i, \bigotimes \mathcal{A}_i, \mu=\mu_1\times\dots\times\mu_n)$

$$f=\frac{dP^{(X_1,\dots,X_n)}}{d\mu} \Rightarrow \boxed{f_i=\frac{dP^{X_i}}{d\mu_i}=\int_{\prod_{j\neq i}\Omega_j} f \; d(\prod_{j\neq i}\mu_j)}$$

$$(X_i) \text{ indep} \iff P^{(X_1,\dots,X_n)}=P^{X_1}\dots P^{X_n}$$
$$\iff f = f_1\cdots f_n .$$
$$\Rightarrow E(X_1\cdots X_n)=E(X_1)\cdots E(X_n) \;;\; Var(\sum X_i)=\sum Var(X_i).$$

IV :

$$\int_{X^{-1}(A')} Y\,dP=\int_{A'} E(Y|X)\,dP^X \quad ; \quad \int_B Y\,dP=\int_B E(Y|\mathcal{B})\,dP \quad \forall B\in\mathcal{B}\subset\mathcal{A}.$$

$$E(a_1 Y_1+a_2 Y_2|X)=a_1E(Y_1|X)+a_2E(Y_2|X).$$

X,Y indp. $\Rightarrow$ $E(XY|X=t)=E(X|X=t)\cdot E(Y|X=t)$.

En genl, $E(f\circ X|X=t)=f(t)$, $E(Y(f\circ X)|X=t)=f(t)\cdot E(Y|X=t)$.

V :

X,Y indep. $\iff \forall A_1\in\mathcal{A}_1,\ P(Y_1\in A_1|X)=\text{cte}=P(Y_1\in A_1)$.

$\Rightarrow$ si Y v.ar. $E(Y|X=x)=E(Y)$.

$$f=\frac{dP^{(X,Y)}}{d(\mu'\times\mu_1)} \Rightarrow \boxed{f^{Y|X=x}(y)=\frac{f(x,y)}{\underbrace{f_X(x)}_{\text{dist. marg. resp. } X}}}$$

$$\boxed{E(Y|X=x)=\int y\, f^{Y|X=x}(y)\,dy}$$

VI: FUNCIONES CARACTERÍSTICAS

Definición: Sea P una probabilidad en el espacio de Borel $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Se llama función característica de P a

$$\begin{aligned} \varphi_P : \mathbb{R} &\longrightarrow \mathbb{C} \\ t &\longmapsto \varphi_P(t) := \int e^{itx}\, dP(x) = E\left[e^{itx}\right] \end{aligned}$$

Definición: Sea X una v.a.r. definida en $(\Omega, \mathcal{A}, P)$. Se llama función característica de X a la función característica de P^X, ie,

$$\begin{aligned} \varphi_X : \mathbb{R} &\longrightarrow \mathbb{C} \\ t &\longmapsto \varphi_X(t) := \varphi_{P^X}(t) = \int e^{itx}\, dP^X(x) = \int e^{itX(x)}\, dP(x) = E\left[e^{itX(x)}\right] \end{aligned}$$

La existencia está garantizada por la integrabilidad de $e^{itX(x)}$.

Proposición:

1) Si X es una v.a.r. discreta que toma valores $x_1, x_2, \dots$ con probabilidades $p_1, p_2, \dots$; entonces

$$\varphi_X(t) = \sum_{k \geq 1} p_k\, e^{itx_k}$$

2) Si X es una v.a.r. abs. continua resp. m con función de densidad f, entonces

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx$$

Proposicion: Sean $X_1, \dots, X_n$ v.a.r. independientes. Entonces $\varphi_{\sum_i X_i}(t) = \prod_i \varphi_{X_i}(t)$.

Proposición: Sea X v.a.r, $a, b \in \mathbb{R}$. Entonces $\varphi_{aX+b}(t) = e^{ibt} \varphi_X(at)$.

Proposición: Si X tiene momento absoluto de orden k finito $\Rightarrow$ φ_X admite derivada k-ésima, y es

$$\varphi_X^{(k)}(t) = \int_{\mathbb{R}} (ix)^k e^{itx} \, dP^X_{(x)}$$

luego, en particular, $E[X^k] = \frac{1}{i^k} \varphi_X^{(k)}(0)$.

Teorema: $\varphi_X = \varphi_Y \iff P^X = P^Y$. ($\iff$ X e Y siguen la misma distr. de prob.)

Definición: Sea $X = (X_1, \dots X_n)$ una v.a. n-dim. Se llama función característica de X a

$$\varphi_X : \mathbb{R}^n \longrightarrow \mathbb{C}$$
$$u \longmapsto \varphi_X(u) := \int e^{i\langle u, X\rangle} dP = E\left[e^{i\sum u_j X_j}\right]$$

Teorema (Caracterización de Independencia): Sea $X_1, \dots X_n$ v.a.r y consideremos $(X_1, \dots X_n)$ vector aleatorio.

$$X_1, \dots, X_n \text{ independientes} \iff \varphi_{(X_1, \dots X_n)} = \varphi_{X_1} \cdots \varphi_{X_n}$$

Proposición: Sea $X \sim N(\mu, \sigma^2)$. Entonces

$$\boxed{\varphi_X(t) = e^{it\mu - \frac{1}{2}t^2\sigma^2}}$$

Proposición. Sea $X \sim G(\alpha, \beta)$. Entonces

$$\boxed{\varphi_X(t) = \frac{1}{(1-\beta i t)^{\alpha}}}$$

Proposición. Sea $X \sim b_n(p)$. Entonces

$$\boxed{\varphi_X(t) = (p e^{it} + 1 - p)^n}$$

Corolario: Si $X_1 \dots X_n$ indep, $X_i \sim b_{n_i}(p) \implies \sum X_i \sim b_{\sum n_i}(p)$.

Proposición: Sea $X \sim P_\lambda$. Entonces

$$\boxed{\varphi_X(t) = e^{\lambda(e^{it}-1)}}$$

Corolario: Si $X_1 \dots X_n$, indep, $X_i \sim P_{\lambda_i} \Rightarrow \sum X_i \sim P_{\sum \lambda_i}$.

Definición: Una v.a. n-dim $X : (\Omega, \mathcal{A}, P) \longrightarrow \mathbb{R}^n$ se dice normal si su función característica es

$$\varphi_X(t) = e^{i u^t b - \frac{1}{2} u^t C u}, \quad u \in \mathbb{R}^n$$

donde $b \in \mathbb{R}^n$ y $C = (C_{ij}) \in \mathcal{M}_n(\mathbb{R})$ es simétrica y semidefinida positiva (i.e, $u^t C u \geq 0 \ \forall u \in \mathbb{R}^n$).

En esta situación se dice que X tiene distribución normal multivariante de media b y matriz de covarianzas C, y pondremos $X \sim N_n(b, C)$

Proposición: Sea $A \in \mathcal{M}_{m\times n}$, $b \in \mathbb{R}^m$ y $X' = (X'_1 \dots X'_n) : (\Omega, \mathcal{A}, P) \longrightarrow \mathbb{R}^n$ de componentes v.a.r. independientes, y cada $X'_j \sim N(0, \lambda_j)$ $(j = 1..n)$. Entonces la v.a. m-dimensional $X := AX' + b \sim N_m(b, C)$, donde $C = ADA^t$, con $D = diag(\lambda_1, \dots \lambda_m)$.

- Es decir, y dado un vector aleatorio n-dim, cuyas componentes tienen distr. normal (1-variante), podemos construir una variable multidim' que siga normal multivariante. ¿Y el recíproco? Sí.

Proposición: Sea $X \sim N_n(b, C)$, C matriz simétrica y semidef positiva. Entonces existe $A \in \mathcal{M}_n$ ortogonal y una v.a. n-dim $X' = (X'_1, \dots, X'_n)$ con $X_j \sim N(0, \lambda_j)$ e independientes tales que $X = AX' + b$, y $\{\lambda_i\}$ son los autovalores de C.

Corolario: $X \sim N_n(b, C) \Longrightarrow E[X] = b$, $Cov[X] = C$.

- Ya sí le hemos puesto bien el nombre antes.
- Proposición: X_1, X_2 normales e indep. $\Longrightarrow (X_1, X_2)$ normal $\quad (X = AX' + b$ como antes$)$

Proposición: Sea $X \sim N_n(b, C)$, $X' = (X'_1 \dots X'_n)$, $X'_j \sim N(0, \lambda_j)$, C simétrica y semidef positiva. Si $\exists C^{-1} \Longrightarrow X$ es abs. continua respecto m_n y la función de densidad de la normal multivariante es

$$\boxed{f(x) = \frac{1}{\sqrt{(2\pi)^n |C|}} e^{-\frac{1}{2}(x-b)^t C^{-1} (x-b)}} \qquad x \in \mathbb{R}^n.$$

Proposición: X tiene distr. normal (multivariante) $\Longleftrightarrow u^t X$ tiene distribución normal (unidim) $\forall u \in \mathbb{R}^n$.

Proposición: Sea X una v.a. n-dim normal, $X = (X_1 \dots X_n)$.

$$X_1 \dots X_n \text{ independientes} \Longleftrightarrow X_1 \dots X_n \text{ incorreladas.}$$

VII : MUESTRAS

Definición: Se llama muestra a una colección de variables aleatorias iid. En el caso finito, se llama tamaño de la muestra al número de v.a. consideradas. Para nosotros serán v.a. reales.

Definición: Sea $X_1 \dots X_n$ una muestra de tamaño n. Se llama función de distribución empírica o muestral a

$$F_n : \mathbb{R} \times \Omega \longrightarrow \mathbb{R}$$
$$(x, \omega) \longmapsto F_n(x,\omega) := \frac{1}{n}\sum_{j=1}^{n} I_{(-\infty,x]}\left(X_j(\omega)\right)$$

• Notar que dicha suma (sin $\frac{1}{n}$) es el número de variables de la muestra tal que $X_j(\omega) \leq x$.

Proposición:

1) Dado $x \in \mathbb{R}$ fijo, $F_n(x,\cdot) : \Omega \to \mathbb{R}$ es una v.a. de distribución $Bn(F(x))$.

2) Dado $\omega \in \Omega$ fijo, $F_n(\cdot,\omega) : \mathbb{R} \to \mathbb{R}$ es la función de distribución de una v.a. discreta.

Definición: Sea $X_1 \dots X_n$ una muestra de tamaño n. Llamaremos

i) media muestral a la v.a. $\overline{X}(\omega) := \frac{1}{n}\sum_{i=1}^{n} X_i(\omega)$.

ii) momento muestral de orden k a la v.a. $a_k(\omega) := \frac{1}{n}\sum_{i=1}^{n} X_i^k(\omega)$

iii) momento muestral absoluto de orden k a la v.a. $A_k(\omega) := \frac{1}{n}\sum_{i=1}^{n} |X_i^k(\omega)|$

iv) momento central muestral de orden k a la v.a. $b_k(\omega) := \frac{1}{n}\sum_{i=1}^{n}\left(X_i(\omega) - \overline{X}(\omega)\right)^k$.

v) momento central muestral absoluto de orden k a la v.a. $B_k(\omega) := \frac{1}{n}\sum_{i=1}^{n}\left|X_i(\omega) - \overline{X}(\omega)\right|^k$

vi) varianza muestral (o cuasi-varianza muestral) a la v.a. $S^2(\omega) := \frac{1}{n-1}\sum\left(X_i(\omega) - \overline{X}(\omega)\right)^2$

Proposición. Sea $X_1 \dots X_n$ una muestra de tamaño n.

1) X_1 tiene media finita $\Rightarrow E[\bar{X}] = E[X_1]$

2) X_1 tiene varianza finita $\Rightarrow Var[\bar{X}] = \frac{1}{n} Var[X_1]$

3) X_1 tiene momento de orden k finito $\Rightarrow E[a_k] = E[X_1^k]$

4) X_1 tiene momento finito de orden 2 $\Rightarrow E[S^2] = Var[X_1]$.

• Por 4) se le llama a S^2 "varianza muestral" y no a b_2.

Definición: Sea $X \sim N(0,1)$ y $Y \sim \chi^2(n) = G(\frac{n}{2}, 2)$ v.a.r. independientes. Llamaremos distribución t de Student de n grados de libertad a la que sigue la v.a.

$$\sqrt{n} \frac{X}{\sqrt{Y}} \sim t(n).$$

Definición: Sea $X \sim \chi^2(m)$, $Y \sim \chi^2(n)$ dos v.a.r. independientes. Se llama distribución F de Snedecor (ó F de Fisher), con (m,n) grados de libertad a la que sigue la v.a.

$$\frac{X/m}{Y/n} \sim F(m,n).$$

Proposición:

1) $X \sim N(\mu, \sigma^2)$, $a,b \in \mathbb{R} \Rightarrow aX+b \sim N(a\mu+b, a^2\sigma^2)$

2) $X_1 \dots X_n$ v.a.r. indep, $X_i \sim N(\mu_i, \sigma_i^2) \Rightarrow \sum X_i \sim N(\sum \mu_i, \sum \sigma_i^2)$.

3) " $X_i \sim G(\alpha_i, \beta) \Rightarrow \sum X_i \sim G(\sum \alpha_i, \beta)$

4) " $X_i \sim \chi^2(n_i) \Rightarrow \sum X_i \sim \chi^2(\sum n_i)$

Teorema: Sean $X_1, \dots, X_n$ una muestra de tamaño n de distribución $N(\mu, \sigma^2)$. iid

1) $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$

2) $\overline{X}$ y $(X_1 - \overline{X}, \dots, X_n - \overline{X})$ son v.a. independientes

3) (Teorema de Fisher): $\overline{X}$ y S^2 son independientes.

4) $(n-1) \frac{1}{\sigma^2} S^2 \sim \chi^2(n-1)$

5) $\sqrt{n} \dfrac{\overline{X} - \mu}{\sqrt{S^2}} \sim t(n-1)$.

- 5) dice cómo se distribuye la normal cuando se usa la varianza muestral y la media muestral. Se probará que $t(n) \to N(0,1)$.

Proposición: Sean $X_1 \dots X_m$; $Y_1 \dots Y_n$ dos muestras: $X_i \sim N(\mu_X, \sigma_X^2)$ y $Y_i \sim N(\mu_Y, \sigma_Y^2)$.
Consideremos las varianzas muestrales de ambas $S_X^2 = \frac{1}{m-1} \sum_{i=1}^{m} (X_i - \overline{X}.)^2$ y $S_Y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2$.

1) $\dfrac{S_X^2 / \sigma_X^2}{S_Y^2 / \sigma_Y^2} \sim F(m-1, n-1)$

2) Si $\sigma_X^2 = \sigma_Y^2$, $\dfrac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{S^2} \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2)$, donde

$$S^2 = \frac{1}{m+n-2} \left[\sum_{i=1}^{m} (X_i - \overline{X})^2 + \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \right]$$ es la varianza muestral combinada

VIII: SUCESIONES DE VVAA

Definición: Sea $(X_n) : (\Omega, \mathcal{A}, P) \longrightarrow \mathbb{R}$ una sucesión de v.a.r. Diremos que

i) (X_n) converge c.s. a X, $(X_n) \xrightarrow{c.s.} X$, si $\exists A \in \mathcal{A}$, $P(A) = 1$: $X_n(\omega) \xrightarrow{en \ \mathbb{R}} X(\omega)$ $\forall \omega \in A$, i.e., si $P\{\omega \in \Omega : X_n(\omega) \not\to 0\} = 0$.

ii) (X_n) converge en probabilidad a X, $(X_n) \xrightarrow{P} X$, si $P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) \to 0$, $\forall \varepsilon > 0$, o bien si $P(|X_n - X| \leq \varepsilon) \to 1$ $\forall \varepsilon$.

iii) $(X_n) \in \mathcal{L}^1$ converge en L^1 ó en media a $X \in \mathcal{L}^1$, $(X_n) \xrightarrow{L^1} X$ si $E[|X_n - X|] \longrightarrow 0$.

iv) $(X_n) \in \mathcal{L}^p$ converge en L^p ó en media p a $X \in \mathcal{L}^p$, $(X_n) \xrightarrow{L^p} X$ si $E[|X_n - X|^p] \longrightarrow 0$.

v) (X_n) converge en distribución a X, $(X_n) \xrightarrow{d} X$ si $F_{X_n}(x) \longrightarrow F_X(x)$ $\forall x$ pto de cont. de F_X.

Proposición: En todas las convergencias salvo en distribución el límite, si existe, es único P-c.s.

- La convergencia c.s es la versión probabilística de la convergencia puntual, y la convergencia en probabilidad de la uniforme.
- Notar que estas nuevas son más débiles, pues permiten que no se cumpla en un conjunto de prob. nula.

- $(X_n) \xrightarrow{L^1} X \Rightarrow (E[X_n]) \longrightarrow E[X]$.

Definición: (X_n) se dice de Cauchy c.s. si $\exists A \in \mathcal{A}, P(A) = 1$; $|X_n(\omega) - X_m(\omega)| \underset{n,m \to \infty}{\longrightarrow} 0$, o bien, si $P\{\omega : |X_n(\omega) - X_m(\omega)| \not\to 0\} = 0$.

Proposición: (X_n) convergente c.s. $\Longleftrightarrow$ (X_n) de Cauchy c.s.

Definición: Sea $(A_n)_{n\in\mathbb{N}}$ una sucesión de conjuntos. Se llama

i) **límite superior** a $\overline{\lim} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \{\omega \in \Omega$ que pertenecen a infinitos $A_n\}$.

ii) **límite inferior** a $\underline{\lim} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m = \{\omega \in \Omega$ que pertenecen a todos los A_n salvo a lo sumo finitos$\}$.

Teorema: Sea (X_n) una suc. de v.a.r. Son equivalentes:

1) $(X_n) \xrightarrow{c.s.} X$

2) $P\left(\bigcup_{m=n}^{\infty} \{|X_m - X| > \varepsilon\}\right) \xrightarrow[n\to\infty]{} 0 \quad \forall \varepsilon$

3) $P\left(\bigcap_{k=1}^{\infty} \underline{\lim} \{|X_n - X| \leq \frac{1}{k}\}\right) = 1$.

Proposición. (X_n) Cauchy c.s. $\iff P\left(\bigcup_{m=1}^{\infty} \{|X_{n+m} - X_n| > \varepsilon\}\right) \xrightarrow[n\to\infty]{} 0$

Corolario: $(X_n) \xrightarrow{c.s.} X \iff \left(\sup_{m\geq n} |X_m - X|\right) \xrightarrow{P} 0$.

Corolario: $\sum_{n=1}^{\infty} P(\{|X_n - X| > \varepsilon\}) < \infty \quad \forall \varepsilon \Rightarrow (X_n) \xrightarrow{c.s.} X$. En este caso se dice que (X_n) **converge completamente** a X.

Teorema (de representación de Skorokhod): Sea $(X_n) \xrightarrow{d} X$ def. en $(\Omega, \mathcal{A}, P)$. Entonces existe $(\Omega', \mathcal{A}', P')$ y (Y_n), Y:

1) $P^{X_n} = P^{Y_n}$, $P^X = P^Y$

2) $(Y_n) \xrightarrow{c.s.} Y$.

· Denotamos $\mathcal{C}_0 := \{ f:\mathbb{R}\to\mathbb{R}$ continuas y acotadas $\}$.

Proposición: $(X_n) \xrightarrow{d} X \iff E[f(X_n)] \longrightarrow E[f(X)] \quad \forall f\in\mathcal{C}_0$.

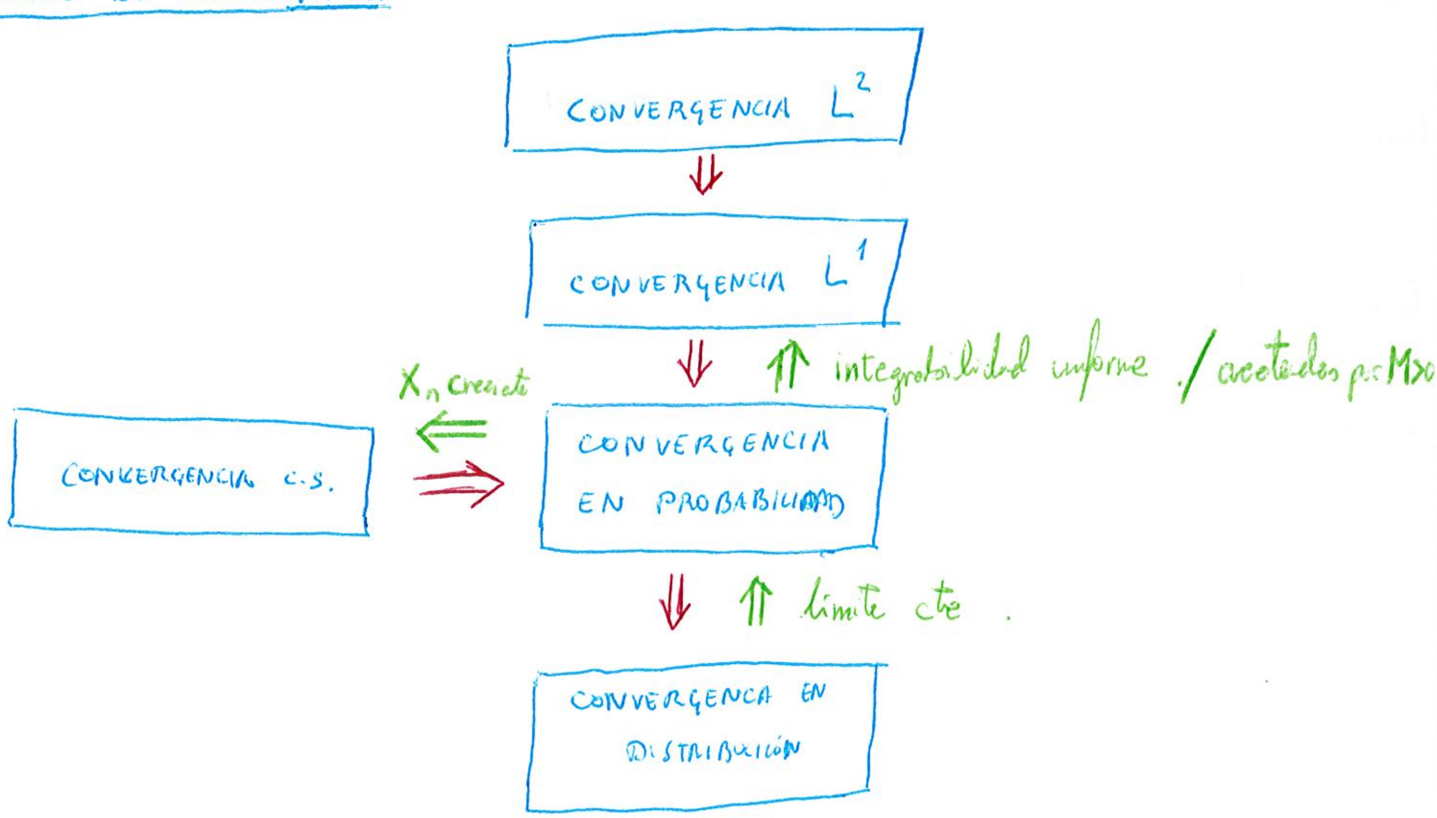
Corolario: $E[f(X_n)] \longrightarrow E[f(X)] \quad \forall f\in C^k, k\in\mathbb{N} \Rightarrow (X_n) \xrightarrow{d} X$.

Teorema (de continuidad de Lévi): Sea (X_n) una suc. de v.a.s. con funciones características φ_{X_n}.

1) $(X_n) \xrightarrow{d} X \iff (\varphi_{X_n}(t)) \longrightarrow \varphi_X(t) \quad \forall t$.

2) Si $\lim_{n\to\infty} \varphi_{X_n}(t) =: \varphi(t)$ existe $\forall t\in\mathbb{R}$ y es cont. en 0, existe X v.a. tal que $(X_n) \xrightarrow{d} X$ y tiene como func. característica φ.

RELACIONES DE CONVERGENCIA



Proposición: $(X_n) \xrightarrow{c.s.} X \Rightarrow (X_n) \xrightarrow{P} X$.

Definición: (X_n) se dice de Cauchy en probabilidad si $\forall \varepsilon > 0 \quad P(\{|X_n - X_m| > \varepsilon\}) \xrightarrow[n,m\to\infty]{} 0$

Proposición. $(X_n) \xrightarrow{P} X \iff (X_n)$ de Cauchy en probabilidad.

Proposición: $(X_n) \xrightarrow{L^1} X \Rightarrow (X_n) \xrightarrow{P} X \qquad ((X_n), X \in \mathcal{L}^1)$.

Teorema (Kolmogorov): $(X_n) \xrightarrow{P} X \Rightarrow (X_n) \xrightarrow{d} X$.

• CONTRAEJEMPLOS:

- $(X_n) \xrightarrow{d} X \not\Rightarrow (X_n) \xrightarrow{P} X$: $X_n = I_{[0, \frac{1}{2}+\frac{1}{n}]}$, $n \geq 2$.
- $(X_n) \xrightarrow{P} X \not\Rightarrow (X_n) \xrightarrow{c.s.} X$: $X_{n,k} := n \cdot I_{[\frac{k-1}{n}, \frac{k}{n})}$, $\begin{matrix} k=1\dots n \\ n=2,3,\dots \end{matrix}$
- $(X_n) \xrightarrow{P} X \not\Rightarrow (X_n) \xrightarrow{L^1} X$: $X_n := n \cdot I_{[0,\frac{1}{n})}$
- $(X_n) \xrightarrow{L^1} X \not\Rightarrow (X_n) \xrightarrow{L^2} X$. : $X_n := \sqrt{n}\, I_{[0,\frac{1}{n})}$.

Lema: $(X_n) \xrightarrow{d} c \in \mathbb{R} \Rightarrow (X_n) \xrightarrow{P} c \in \mathbb{R}$.

Proposición: $(X_n) \uparrow_P X \Rightarrow (X_n) \xrightarrow{c.s.} X$

Proposición: $(X_n) \xrightarrow{P} X$, $|X_n| \leq M \Rightarrow (X_n) \xrightarrow{L^1} X$.

Lema (de absoluta continuidad): Sea $X \in \mathcal{L}^1$.

$$\forall \varepsilon > 0 \ \exists \delta > 0 : A \in \mathcal{A},\ P(A) < \delta \Rightarrow E[|X| I_F] < \varepsilon.$$

Corolario: Sea $X \in \mathcal{L}^1$.

$$\forall \varepsilon > 0 \ \exists K > 0 : E[|X| I_{\{|X| > K\}}] < \varepsilon.$$

Definición: Una familia de v.a.s $\mathcal{C}$ se dice uniformemente integrable si

$$\forall \varepsilon > 0 \ \exists K \geq 0 \ : E\left[|X| \, I_{\{|X| > K\}}\right] < \varepsilon \quad \forall X \in \mathcal{C}.$$

Lema: Toda familia $\mathcal{C}$ u.i. es acotada en $\mathcal{L}^1$ (ie, todas son integrables y dichas integrales están acotadas)

OJO!: $\mathcal{C}$ acotada en $\mathcal{L}^1 \not\Rightarrow \mathcal{C}$ u.i. $\quad : X_n = n \, I_{[0, \frac{1}{n})}$.

Proposición (Condiciones Suficientes de u.i):

1) $\mathcal{C}$ acotada en L^p, $p > 1 \Rightarrow \mathcal{C}$ u.i.

2) $\mathcal{C}$ dominadas por una v.a. integrable $\geq 0 \Rightarrow \mathcal{C}$ u.i.

Teorema: Sea $(X_n), X \in \mathcal{L}^1$.

$$(X_n) \xrightarrow{L^1} X \iff \begin{cases} 1) \ (X_n) \xrightarrow{P} X \\ 2) \ \mathcal{C} = \{X_n\} \text{ es u.i.} \end{cases}$$

CONVERGENCIA BAJO TRANSFORMACIONES

Teorema: Sean $(X_n), (Y_n), X, Y$ v.a. Si $(X_n) \to X, (Y_n) \to Y$ casi seguro (en probabilidad) (en $L^p, p > 1$) $\Rightarrow (X_n + Y_n) \to X + Y$ casi seguro (en probabilidad) (en L^p).

Proposición (Slutsky): $(X_n) \xrightarrow{d} X, (Y_n) \xrightarrow{d} c \in \mathbb{R} \Rightarrow (X_n + Y_n) \xrightarrow{d} X + c$

Teorema: Sean $(X_n), (Y_n), X, Y$ v.a. Si $(X_n) \to X, (Y_n) \to Y$ casi seguro (en probabilidad) (en L^p y en L^q, resp, p.q exp conj) $\Rightarrow (X_n \cdot Y_n) \to X \cdot Y$ casi seguro (en probabilidad) (en L^1).

Proposición (Slutsky) $(X_n) \xrightarrow{d} X, (Y_n) \xrightarrow{d} c \in \mathbb{R} \Rightarrow (X_n \cdot Y_n) \xrightarrow{d} c \cdot X$.

Proposición: Sea $g: \mathbb{R} \to \mathbb{R}$ continua, y (X_n) una suc. de v.a.a. Si $(X_n) \to X$ casi seguro (en probabilidad) (en distribución) $\Rightarrow (g(X_n)) \to g(X)$ casi seguro (en probabilidad) (en distribución).

CONVERGENCIA DE VECTORES ALEATORIOS

- Ahora las sucesiones serán $(X_n) = ((X_n^1, \dots, X_n^k))$, y cambiamos el valor absoluto por la norma euclídea.

Definición: Sea (X_n) una sucesión de vectores aleatorios k-dimensionales, def en $(\Omega, \mathcal{A}, P)$.

i) $(X_n) \xrightarrow{c.s.} X$ si $\exists A \in \mathcal{A}$, $P(A) = 1$: $\forall \omega \in A$ $\quad X_n(\omega) \xrightarrow{\text{en } \mathbb{R}^n} X(\omega)$

ii) $(X_n) \xrightarrow{P} X$ si $P(\{ \|X_n(\omega) - X(\omega)\|_2 > \varepsilon) \xrightarrow[n \to \infty]{} 0 \quad \forall \varepsilon$.

iii) $(X_n) \xrightarrow{L^2} X$ si $E[\|X_n - X\|_2] \to 0$.

iv) $(X_n) \xrightarrow{L^1} X$ si $E[\|X_n - X\|_1] \to 0$.

v) $(X_n) \xrightarrow{d} X$ si $E[f(X_n)] \to E[f(X)] \quad \forall f: \mathbb{R}^k \to \mathbb{R}$ cont. y acotada.

Proposición: Para la convergencia c.s, en probabilidad, en L^2 y L^1 se tiene:

$$(X_n) \to X \Longleftrightarrow (X_n^i) \to X^i \quad , i = 1, \dots, k. \quad ,$$

y para la convergencia en distribución se cumple:

$$(X_n) \xrightarrow{d} X \underset{\not\Longleftarrow}{\Longrightarrow} (X_n^i) \xrightarrow{d} X^i \quad , i = 1 \dots k$$

Corolario (Tchebycheff): Si (X_n) suc de vvaa incorreladas e ident. distr. con $\mu = E[X_i]$, y $\sigma^2 = Var[X_i]$, entonces $\frac{S_n - n\mu}{n} \xrightarrow{P} 0$, o bien

$$\frac{S_n}{n} \xrightarrow{P} \mu \quad \text{(que es una v.a. degenerada)}.$$

es decir, $\frac{1}{n} S_n = \frac{1}{n} \sum X_i = \overline{X} \xrightarrow{P} \mu$.

Corolario: (X_n) suc. de v.a. incorreladas, $X_i \sim b_1(p)$, $\frac{S_n}{n} \equiv$ frecuencia relativa con la que ocurre el suceso las n primeras veces, converge a $p \equiv$ prob. con la que ocurre el suceso.

Corolario: Dado $\varepsilon > 0$ y $0 < \delta \leq 1$, eligiendo $n > \frac{p(1-p)}{\varepsilon^2 \delta}$ (ó $n > \frac{1}{4\varepsilon^2 \delta}$, si p es desconocido) se tiene que $P\left(\left|\frac{S_n}{n} - p\right| \leq \varepsilon\right) \geq 1 - \delta$.

Teorema (Khintchine): Sea (X_n) una suc. de vvaa iid de momento absoluto finito y $\mu = E[X_i]$ $\forall i$. Entonces $\frac{S_n - n\mu}{n} \xrightarrow{P} 0$, o bien

$$\frac{S_n}{n} \xrightarrow{P} \mu$$

Definición: Una sucesión de vvaa (X_n) se dice que obedece la Ley Fuerte de los Grandes Números respecto a la sucesión de ctes normalizadoras (B_n) si existe una sucesión de ctes (A_n) tal que

$$\frac{S_n - A_n}{B_n} \xrightarrow{c.s.} 0.$$

__Teorema__: Sea (X_n) suc de v.a.a. indep con $E[X_i]=0$ y $0<\sigma_i^2 = Var[X_i] < \infty$ $\forall i$.

Si $\sum_{i=1}^{\infty} \sigma_i^2 < \infty \Longrightarrow \sum_{i=1}^{\infty} X_i$ converge c.s.

__Lema__ (Cesàro - Kronecker): Sea $(x_n) \xrightarrow{\text{en } \mathbb{R}} x$, y $(b_n) \xrightarrow{\text{en } \mathbb{R}} \infty$. Entonces $\left(\frac{1}{b_n}\sum_{i=1}^{n}(b_i - b_{i-1})x_i\right) \longrightarrow x$. En particular, con $b_n = n$, $\frac{1}{n}\sum x_i \equiv$ media $\longrightarrow x$. Además, $\left(\frac{1}{b_n}\sum_{i=1}^{n} b_i x_i\right) \longrightarrow 0$.

__Teorema__: Sea (X_n) una suc de v.a.a. indep con $\mu_i = E[X_i]$ y $0 < \sigma_i^2 = Var[X_i] < \infty$ $\forall n$.

Sea $(B_n) \longrightarrow \infty$ y >0. Si $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{B_n^2} < \infty$, entonces

$$\frac{S_n - \sum_{i=1}^{n}\mu_i}{B_n} \xrightarrow{c.s.} 0 .$$

__Corolario__*: (X_n) indep, $\mu_i = E[X_i]$ y $\sigma_i^2 = Var[X_i] < \infty$. Si $\sum \frac{\sigma_n^2}{n^2} < \infty$, entonces

$$\frac{S_n - \sum \mu_i}{n} \xrightarrow{c.s.} 0$$

__Corolario__: (X_n) indep, $\mu_i = E[X_i]$ y $\sigma_i^2 = Var[X_i] \leq M$, $M \geq 0$, entonces

$$\frac{S_n - \sum \mu_i}{n} \xrightarrow{c.s.} 0$$

__Corolario__: (X_n) iid, $\mu = E[X_i]$, $\sigma^2 = Var[X_i] < \infty$ $\forall i$, entonces $\frac{S_n - n\mu}{n} \xrightarrow{c.s.} 0$, o bien,

$$\frac{S_n}{n} \xrightarrow{c.s.} \mu$$

Teorema (Kolmogorov): (X_n) iid, $E[|X_i|] < \infty$ y $\mu = E[X_i]$. Entonces $\frac{S_n - n\mu}{n} \xrightarrow{c.s.} 0$, o bien,

$$\frac{S_n}{n} \xrightarrow{c.s.} \mu.$$

TM CENTRAL DEL LIMITE

Teorema: Sea (X_n) iid de distribución común $b_1(p)$, $0<p<1$, y sea $S_n = \sum_{i=1}^{n} X_i$.

1) (Bernoulli): $\frac{S_n}{n} \xrightarrow{d} p$

2) (De Moivre-Laplace): $\frac{S_n - E[S_n]}{\sqrt{Var[S_n]}} \xrightarrow{d} Z$, $Z \sim N(0,1)$.

Teorema (Poisson): Para cada $n \in \mathbb{N}$, sea $X_{n1}, \dots, X_{nn}$ v.a. iid con distribución $b_1(p_n)$, $0<p_n<1$, y sea $S_n = \sum_{i=1}^{n} X_{ni}$. Si $np_n \to \lambda \in (0,\infty)$, entonces

$$S_n \xrightarrow{d} Y, \quad Y \sim P(\lambda).$$

- Esto dice que $b_n(p_n) \longrightarrow P_\lambda$ cuando $n \to \infty$, pues $S_n \sim b_n(p)$.

Teorema (Lindenberg-Lévi): Sea (X_n) suc. de v.a.a. iid, $\mu = E[X_i]$ y $\sigma^2 = Var[X_i] < \infty$ $\forall i$.
Si $S_n = \sum_{i=1}^{n} X_i$, entonces

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} Z, \quad Z \sim N(0,1).$$

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{S_n - E[S_n]}{\sqrt{Var[S_n]}}$$