Asignatura : INFERENCIA ESTADÍSTICA
Titulación : LICENCIADO EN CCSS. Y T. ESTADÍSTICAS
Profesor : ISABEL MOLINA PERALTA
Capítulo 1 : PRELIMINARES

1. PROBABILIDAD Y VARIABLES ALEATORIAS

1.1 EXPERIMENTOS ALEATORIOS

Experimento aleatorio: Proceso que verifica:

✓ resultado imposible de predecir;

✓ si se repite en idénticas condiciones, se puede obtener un resultado distinto;

✓ se tiene información sobre el conjunto de posibles resultados.

Resultados elementales(ω): posibles resultados indivisibles del experimento aleatorio. **Espacio muestral**(Ω): conjunto de resultados elementales.

Suceso(A): subconjunto de Ω. Una vez realizado el experimento, se puede decir que A "ha ocurrido", si el resultado del experimento está en A.

Ejemplo 1 Experimentos aleatorios:

(a) Lanzamiento de una moneda.
Espacio muestral: $\Omega = \{\text{"}cara\text{"}, \text{"}cruz\text{"}\}$.
Sucesos: $\emptyset$, $\{\text{"}cara\text{"}\}$, $\{\text{"}cruz\text{"}\}$, Ω.

(b) Observación del número de accidentes de tráfico en un minuto de un día en España. Espacio muestral: $\Omega = \mathbb{N} \cup \{0\}$.

(c) Selección aleatoria de una mujer española entre 20 y 40 años y medición de su peso (en kgs).
Espacio muestral: $\Omega = [m, \infty)$, siendo m el peso mínimo.

2. ESTRUCTURAS SOBRE ESPACIOS MUESTRALES

Se necesita una estructura de conjuntos a los que se les va a dar probabilidad.

σ-álgebra ($\mathcal{A}$) sobre Ω: colección de subconjuntos de Ω, con las propiedades:

- $\emptyset \in \mathcal{A}$;
- $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$;
- $A_1, A_2 \ldots \in \mathcal{A} \Rightarrow \cup_{n=1}^{\infty} A_i \in \mathcal{A}$.

Ejemplos de σ-álgebras comunes:

(a) σ-**álgebra de las partes de** Ω $(\mathcal{P}(\Omega))$: colección de todos los subconjuntos de Ω.

(b) Sea el espacio muestral $\Omega = \mathbb{R}$, y sea $I = \{(-\infty, a] : \ a \in \mathbb{R}\}$ el conjunto de intervalos de la forma $(-\infty, a]$.

σ-**álgebra de Borel**: $(\mathcal{B})$ menor σ-álgebra que contiene a I. Intersección de todas las σ-álgebras que contienen a I.

$\mathcal{B}$ contiene a todos los complementarios, intersecciones numerables y uniones numerables de elementos de I.

$\mathcal{B}$ contiene todo tipo de intervalos y puntos aislados de $\mathbb{R}$, pero $\mathcal{B}$ no es $\mathcal{P}(\mathbb{R})$.

- $(a, \infty) \in \mathcal{B}$, ya que $(a, \infty) = (-\infty, a]^c$, y $(-\infty, a] \in \mathbb{R}$.

- $(a, b] \in \mathbb{R}, \forall a < b$, ya que $(a, b] = (-\infty, b] \cap (a, \infty)$, donde $(-\infty, b] \in \mathcal{B}$ y $(a, \infty) \in \mathcal{B}$.

- $\{a\} \in \mathcal{B}, \forall a \in \mathbb{R}$, ya que $\{a\} = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, a\right] \in \mathcal{B}$.

Sea un subconjunto $A \subset \mathbb{R}$.
σ-**álgebra de Borel restringida a** A: $\mathcal{B}_A = \{B \cap A : \; B \in \mathcal{B}\}$.

Espacio medible: $(\Omega, \mathcal{A})$, Ω espacio muestral, $\mathcal{A}$ σ-álgebra sobre Ω.

Ejemplo 2 En el Ejemplo 1

(a) Espacio medible finito: $\Omega = \{\text{``}cara\text{''}, \text{``}cruz\text{''}\}$, $\quad \mathcal{A} = \{\emptyset, \{\text{``}cara\text{''}\}, \{$

(b) Espacio medible discreto: $\Omega = \mathbb{N} \cup \{0\}$, $\quad \mathcal{P}(\Omega)$.

(c) Espacio medible continuo: $\Omega = [m, \infty) \subset \mathbb{R}$, $\quad \mathcal{B}_{[m,\infty)}$.

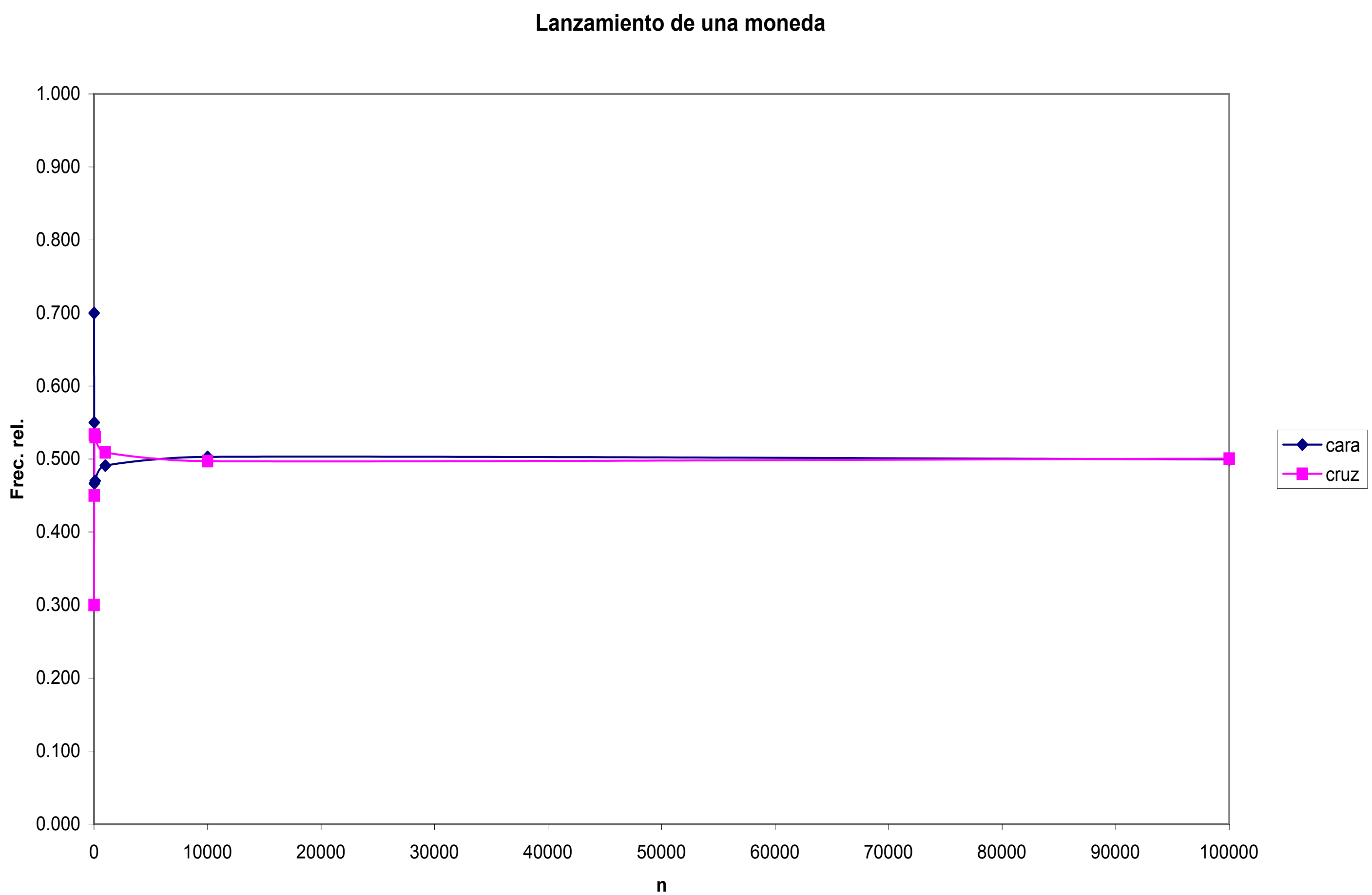
3. DEFINICIONES DE PROBABILIDAD

Definición frecuencial de Probabilidad

Frecuencias relativas de los resultados de los experimentos del Ejemplo 1.1, cuando se repiten n veces.

(a) Lanzamiento de una moneda n veces.

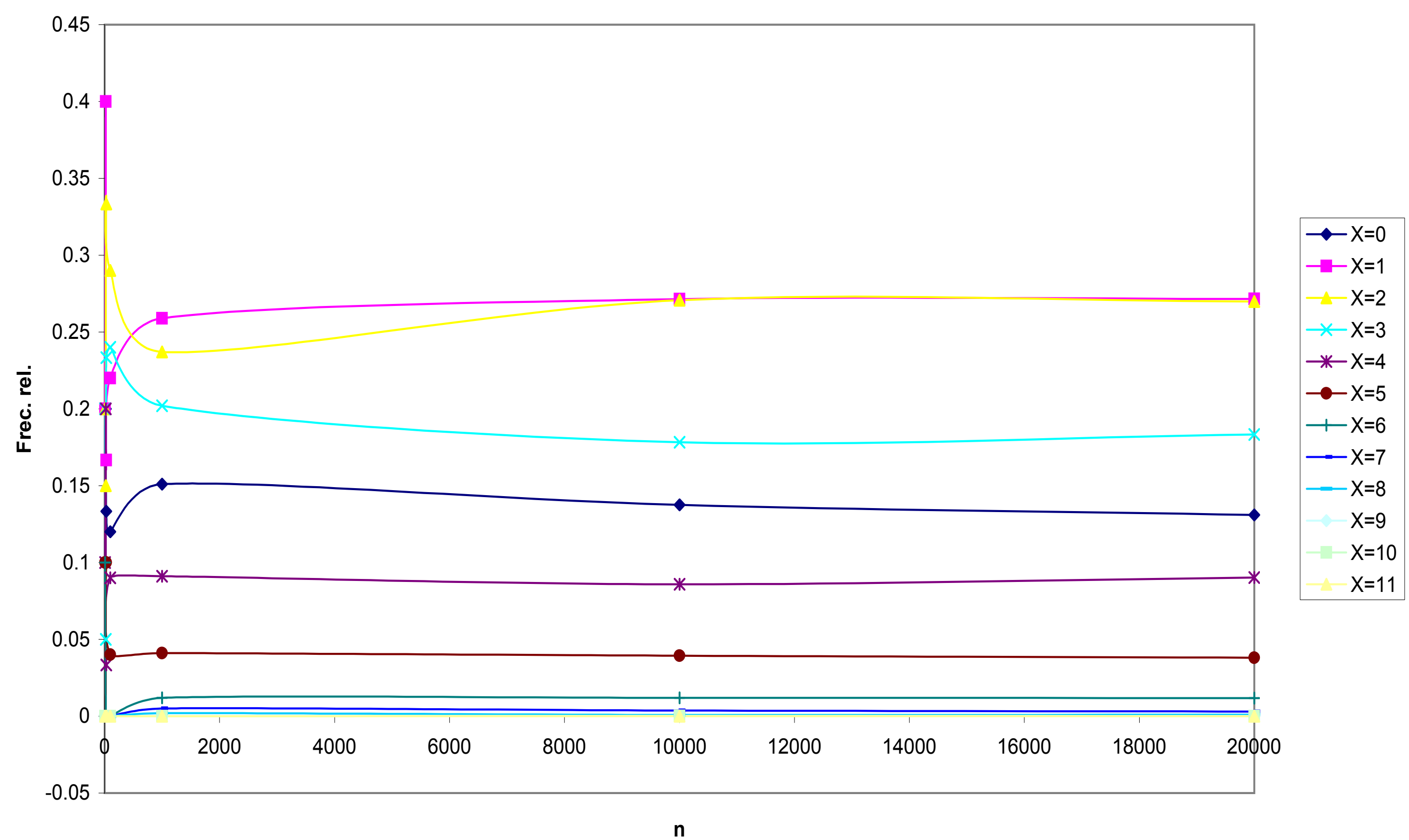
	Resultado	
n	"cara"	"cruz"
10	0.700	0.300
20	0.550	0.450
30	0.467	0.533
100	0.470	0.530
1000	0.491	0.509
10000	0.503	0.497
100000	0.500	0.500

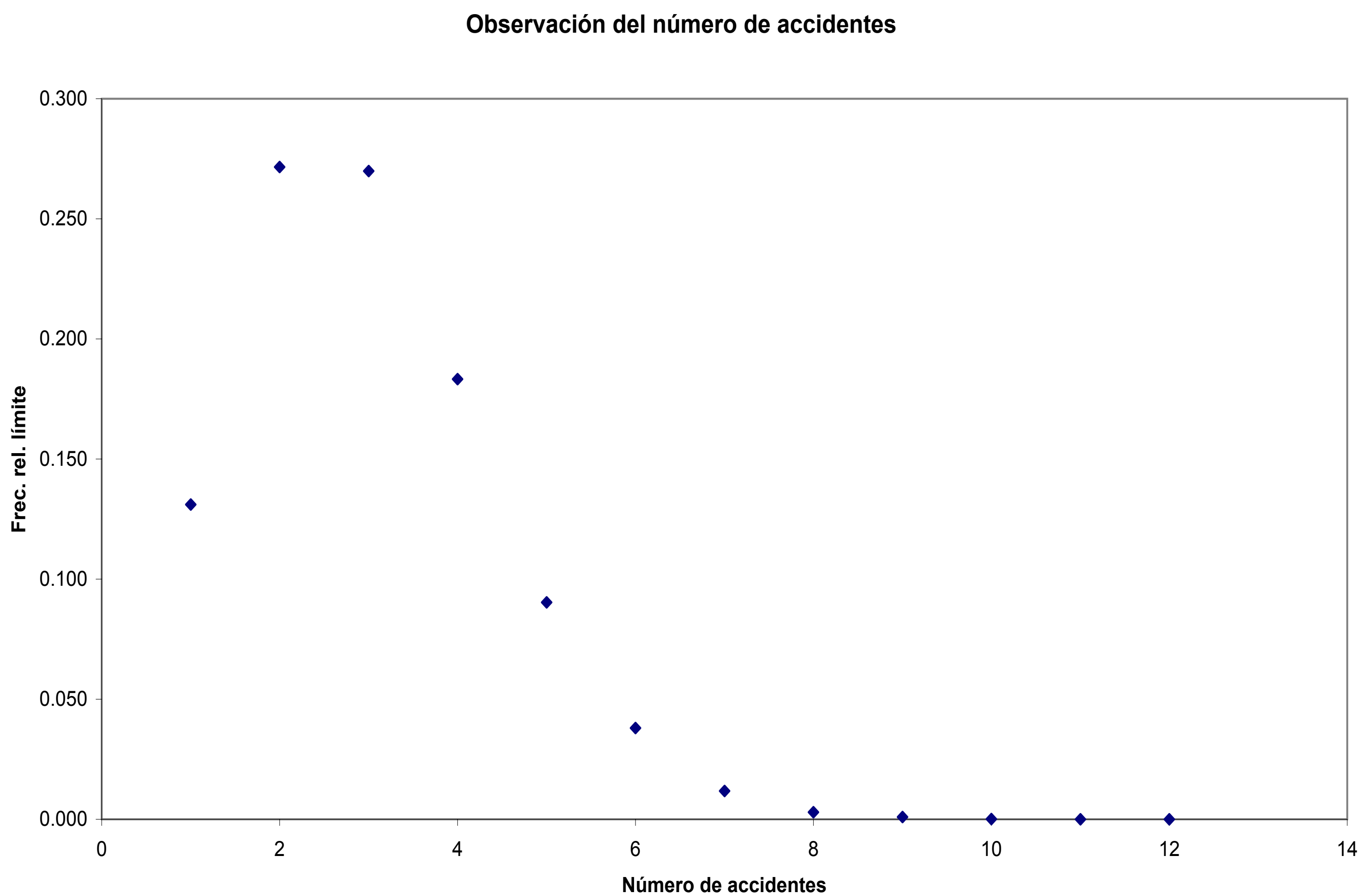

Lanzamiento de una moneda
Frec. rel.
1.000
0.900
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
n
cara
cruz

(b) Observación del número de accidentes de tráfico en n minutos.

	Resultado								
n	0	1	2	3	4	5	6	7	8
10	0.1	0.2	0.2	0.2	0.1	0.1	0.1	0	0
20	0.2	0.4	0.15	0.05	0.2	0	0	0	0
30	0.13	0.17	0.33	0.23	0.03	0	0	0	0
100	0.12	0.22	0.29	0.24	0.09	0	0	0	0
1000	0.151	0.259	0.237	0.202	0.091	0.012	0.002	0	0
10000	0.138	0.271	0.271	0.178	0.086	0.012	0.0008	0.0001	0
20000	0.131	0.272	0.270	0.183	0.090	0.012	0.00095	0.00005	0

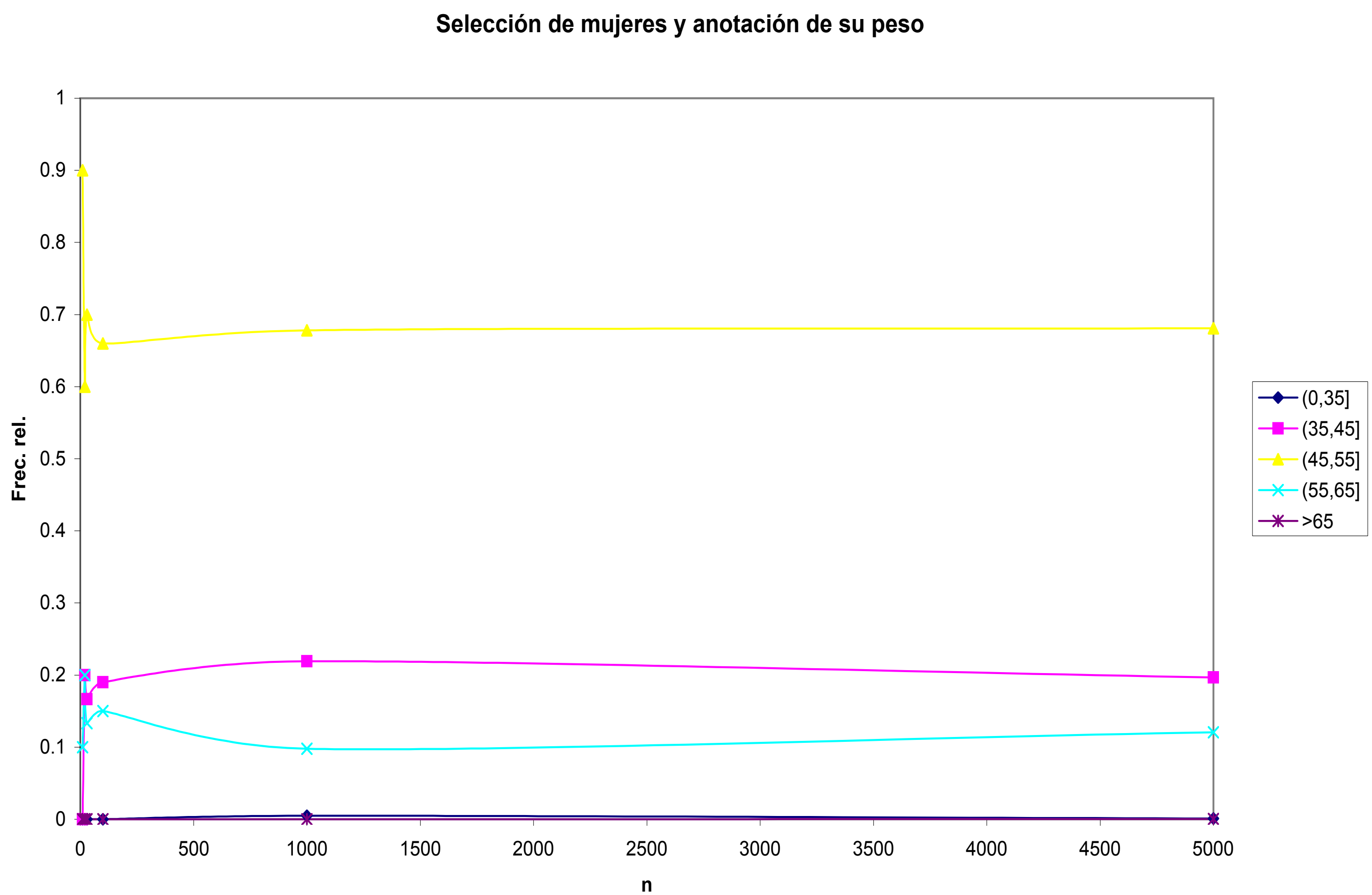
Número de accidentes de tráfico

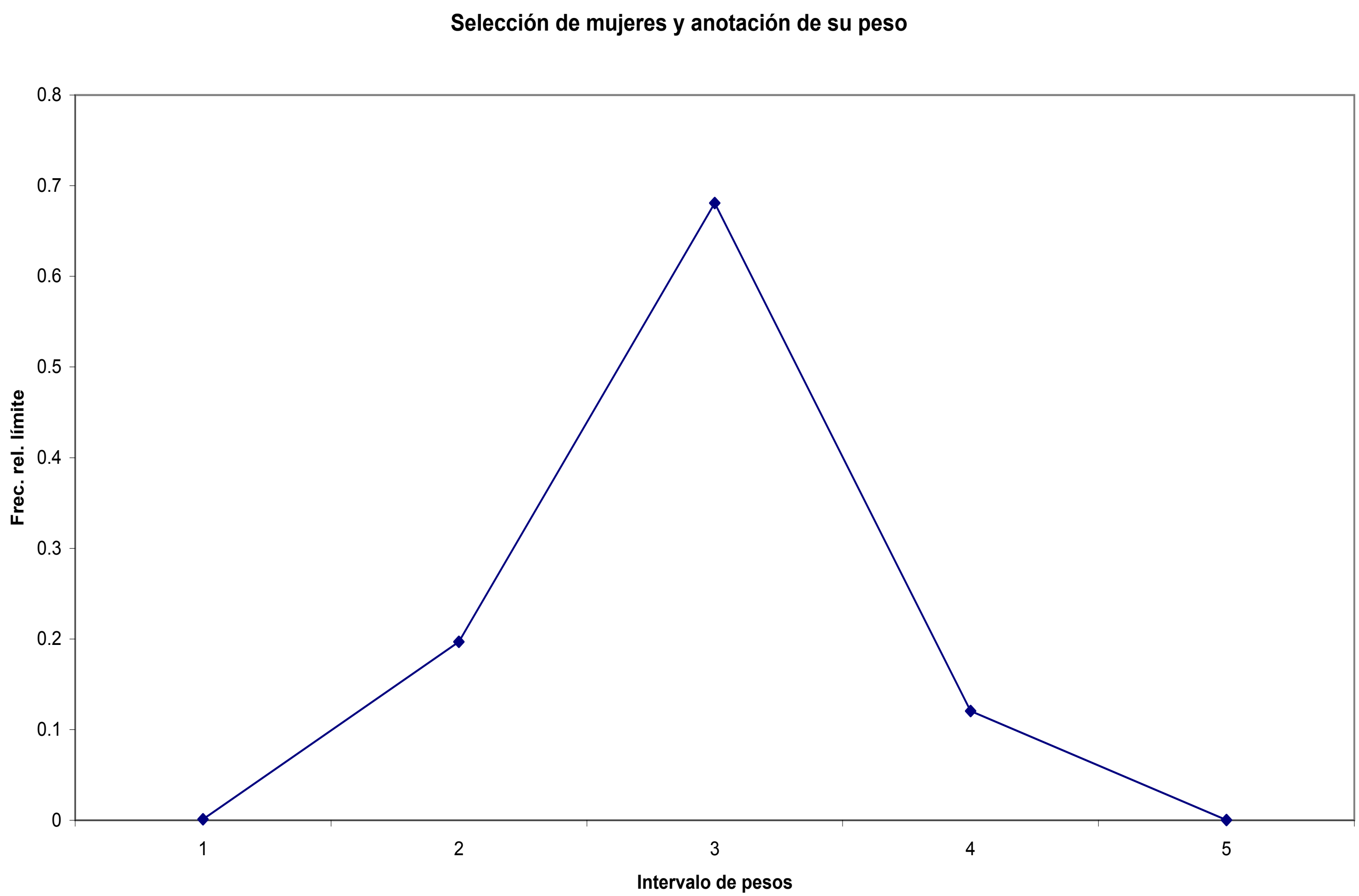



Observación del número de accidentes
Frec. rel. límite
0.300
0.250
0.200
0.150
0.100
0.050
0.000
0
2
4
6
8
10
12
14
Número de accidentes

(c) Selección aleatoria independiente de n mujeres entre 20 y 40 años y observación de su peso (en kgs.).

	Intervalos de Pesos				
n	$(0, 35]$	$(35, 45]$	$(45, 55]$	$(55, 65]$	$(65, \infty)$
10	0	0	0.9	0.1	0
20	0	0.2	0.6	0.2	0
30	0	0.17	0.7	0.13	0
100	0	0.19	0.66	0.15	0
1000	0.005	0.219	0.678	0.098	0
5000	0.0012	0.197	0.681	0.121	0.0004


Selección de mujeres y anotación de su peso
Frec. rel.
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
n
(0,35]
(35,45]
(45,55]
(55,65]
>65


Selección de mujeres y anotación de su peso
Frec. rel. límite
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
1
2
3
4
5
Intervalo de pesos

Probabilidad de un suceso A: Límite de la frecuencia relativa de A, cuando $n \to \infty$.

n núm. de repeticiones del experimento.
n_A número de repeticiones en las que A ocurre.

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

Definición de probabilidad de Laplace

Para experimentos con sucesos elementales igualmente frecuentes y espacio muestral finito.

Probabilidad de un suceso A: Proporción de resultados favorables a A.

k número de resultados posibles del experimento (cardinal de Ω).
$k(A)$ número de resultados contenidos en A.

$$P(A) = k(A)/k.$$

Definición axiomática de probabilidad de Kolmogoroff

$(\Omega, \mathcal{A})$ un espacio medible.
Función de probabilidad (P): Aplicación $P : \mathcal{A} \to I\!R$ que cumple los siguiente axiomas

✓ $\forall A \in \mathcal{A}$, $P(A) \geq 0$;

✓ $P(\Omega) = 1$;

✓ Para cualquier sucesión $A_1, A_2, \ldots$ de sucesos disjuntos de $\mathcal{A}$,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

En la estadística clásica se utilizan funciones de probabilidad que verifican las definiciones frecuencial y axiomática.

Ejemplo 3 Para el Ejemplo 1 (a),
Espacio medible:

$$\Omega = \{\text{``}cara\text{''}, \text{``}cruz\text{''}\}, \quad \mathcal{A} = \{\emptyset, \{\text{``}cara\text{''}\}, \{\text{``}cruz\text{''}\}, \Omega\}$$

Función de probabilidad:

$$P_1(\emptyset) = 0,\ P_1(\{\text{``}cara\text{''}\}) = P_1(\{\text{``}cruz\text{''}\}) = 1/2,\ P_1(\Omega) = 1$$

Esta función verifica las tres definiciones de probabilidad.

Función de probabilidad:

$$P_2(\emptyset) = 0,\ P_2(\{\text{``}cara\text{''}\}) = p < 1/2,\ P_2(\{\text{``}cruz\text{''}\}) = 1-p,\ P_2(\Omega) = 1$$

Si la moneda está trucada, no verifica la definición de Laplace, ya que los resultados no son equiprobables. Sí verifica la definición axiomática de Kolmogoroff.

4. VARIABLES ALEATORIAS

Espacio de probabilidad: $(\Omega, \mathcal{A}, P)$, P función de probabilidad definida sobre el espacio medible $(\Omega, \mathcal{A})$.

La definición de variable aleatoria permite transformar los elementos de Ω en elementos de $I\!R$.

Sea $(\Omega, \mathcal{A})$ un espacio medible y $\mathcal{B}$ la σ-álgebra de Borel sobre $I\!R$.
Variable aleatoria (v.a.): Aplicación

$$\begin{aligned} X \ : \ & \Omega \to I\!R \\ & \omega \rightsquigarrow X(\omega) \end{aligned}$$

que es **medible**, es decir, que verifica

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

donde $X^{-1}(B)$ es la antiimagen de B,

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}.$$

Una v.a. transforma un espacio medible $(\Omega, \mathcal{A})$ en otro espacio medible con mejores propiedades, $(\mathbb{R}, \mathcal{B})$.

La condición de medibilidad permite transferir probabilidades de subconjuntos $A \in \mathcal{A}$ a subconjuntos $B \in \mathcal{B}$.

Ejemplo 4 Para los experimentos del Ejemplo 1, las siguientes son variables aleatorias:

(a) Espacio medible:

$$\Omega = \{\text{``}cara\text{''}, \text{``}cruz\text{''}\}, \quad \mathcal{A} = \{\emptyset, \{\text{``}cara\text{''}\}, \{\text{``}cruz\text{''}\}, \Omega\}$$

Variable aleatoria: "Número de caras al lanzar una moneda"

$$X(\omega) = \begin{cases} 1 & si\ \omega = \text{``}cara\text{''}, \\ 0 & si\ \omega = \text{``}cruz\text{''}; \end{cases}$$

(b) Espacio medible: $(\Omega, \mathcal{P}(\Omega))$, donde $\Omega = I\!N \cup \{0\}$
Variable aleatoria: “Número de accidentes de tráfico en un minuto en España”

$$X_1(\omega) = \omega$$

Variable aleatoria: “Ocurrencia o no de un accidente de tráfico en un minuto en España”

$$X_2(\omega) = \begin{cases} 1 & \text{si } \omega \in I\!N; \\ 0 & \text{si } \omega = 0. \end{cases}$$

(c) Espacio medible: $(\Omega, \mathcal{B}_\Omega)$, donde $\Omega = [m, \infty)$.
Variable aleatoria: $X_1(\omega) = \omega$
Variable aleatoria: “Peso de al menos 65 kilogramos”.

$$X_2(\omega) = \begin{cases} 1 & si\ \omega \geq 65, \\ 0 & si\ \omega < 65; \end{cases}$$

$\mathcal{B}$ σ-álgebra de Borel sobre $\mathbb{R}$.

Probabilidad inducida por la v.a. X:

$$\begin{aligned} P_X \ : \ & \mathcal{B} \to \mathbb{R} \\ & B \rightsquigarrow P_X(B) = P(X^{-1}(B)). \end{aligned}$$

Ejemplo 5 Sea P_1 función de probabilidad definida en el Ejemplo 3, y X la v.a. definida en el Ejemplo 4 (a).

✓ $P_{1X}(\{0\}) = P_1(X = 0) = 1/2$.

✓ $P_{1X}((-\infty, 0]) = P_1(X \leq 0) = 1/2$.

✓ $P_{1X}((0, 1]) = P_1(0 < X \leq 1) = 1/2$.

Función de distribución (f.d.) de una v.a. X: función $F_X : I\!R \rightarrow I\!R$ que verifica

$$F_X(x) = P_X((-\infty, x]), \ \forall x \in I\!R.$$

Propiedades de la función de distribución:

(i) $\forall x \in I\!R, F_X(x) \in [0, 1]$.

(ii) La función de distribución es monótona creciente, es decir,

$$x < y \Rightarrow F_X(x) \leq F_X(y).$$

(iii) Sean a, b dos números reales con $a < b$; entonces

$$P(a < X \leq b) = F_X(b) - F_X(a).$$

(iv) $F_X(-\infty) = \lim_{x \to -\infty} F(x) = 0$ y $F_X(+\infty) = \lim_{x \to +\infty} F(x) = 1$.

(v) F_X es continua por la derecha.

(vi) El conjunto de puntos de discontinuidad de F_X es finito, o a lo sumo, infinito numerable.

Rango de una v.a. X (R_X): Conjunto imagen de X.

Variable aleatoria **discreta**: R_X es finito o infinito numerable.

Para variables discretas se define una función que proporciona las probabilidades de los puntos individuales de su rango R_X.

Función de cuantía de una variable aleatoria discreta X: función $p_X : \mathbb{R} \to \mathbb{R}$ tal que

$$p_X(x) = P_X(\{x\}), \ \forall x \in \mathbb{R}.$$

Notación: $p(X = x) = p_X(x)$.

"Distribución" de una v.a. discreta X: indistintamente P_X, F_X, o p_X.

Variable aleatoria **continua** (o **absolutamente continua**): verifica las condiciones

✓ La función de distribución F_X es siempre continua.

✓ La función de distribución F_X es derivable, con derivada continua, salvo a lo sumo en un conjunto numerable de puntos.

Función de densidad de probabilidad de una v.a. X: función $f_X : \mathbb{R} \to \mathbb{R}$ definida de la forma

$$f_X(x) = \begin{cases} 0 & \text{si } x \in S; \\ F'_X(x) & \text{si } x \notin S. \end{cases}$$

S conjunto de puntos donde F_X no tiene derivada continua.

$f_X(x)$ mide la densidad de probabilidad de un intervalo infinitesimal centrado en x.

Propiedades de la función de densidad

(i) $f_X(x) \geq 0, \ \forall x \in \mathbb{R}$.

(ii) $P(a < X \leq b) = \int_a^b f_X(x)dx$.

(iii) $\int_{\mathbb{R}} f_X(x)dx = 1$.

Una v.a. continua está determinada si se proporciona su función de probabilidad inducida P_X, o su función de distribución F_X, o su función de densidad de probabilidad f_X.

Transformación de variables aleatorias

Si $Y = g(X)$ es una transformación biyectiva de X, con inversa $X = g^{-1}(Y)$, donde g es diferenciable. La densidad de Y es

$$f_Y(y) = f_X(g^{-1}(y)) \left| \left. \frac{dx}{dy} \right|_{x=g^{-1}(y)} \right| .$$

Esperanza:

X discreta, $E[g(X)] = \sum_x g(x) p_X(x)$.

X continua, $E[g(X)] = \int_{\mathbb{R}} g(x) f_X(x) dx$.

Varianza:

$$Var(X) = E\left\{[X - E(X)]^2\right\} = E(X^2) - E^2(X)$$

Covarianza:

$$Cov(X_1, X_2) = E\{[X_1 - E(X_1)][X_2 - E(X_2)]\} = E(X_1 X_2) - E(X_1)E(X$$

Coeficiente de correlación:

$$\rho(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)Var(X_2)}}$$

Covarianza y correlación son medidas de la relación lineal entre dos v.a.

Propiedades:

(i) $-1 \leq \rho(X_1, X_2) \leq 1$.

(ii) Si $X_2 = aX_1 + b$, entonces $\rho(X_1, X_2) = signo(a)$.

(iii) $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$.

(iv) $Var(X_1 - X_2) = Var(X_1) + Var(X_2) - 2Cov(X_1, X_2)$.

(v) Si $X_1, \ldots, X_n$ son independientes:

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

Función generatriz de momentos (f.g.m.) de una v.a. X:

$$M_X(s) = E[e^{sX}],$$

para s dentro de un entorno del cero $(-h, h) \subset I\!R$, $h > 0$, en el que la esperanza exista. Si la esperanza no existe para ningún entorno del cero, entonces se dice que la f.g.m. **no existe**.

Ejemplo 1.10. Una v.a. X con distribución gamma posee función de densidad

$$f_X(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \ 0 < x < \infty, \ \alpha > 0, \ \beta > 0$$

Función gamma:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

Propiedades:

(i) $\Gamma(1) = 1$.

(ii) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$, para cualquier $\alpha \geq 1$.

(iii) $\Gamma(n) = (n - 1)!$.

Vamos a obtener la f.g.m. de X.

$$M_X(s) = E[e^{sX}] = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty e^{sx} x^{\alpha-1} e^{-x/\beta} dx = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-(1/\beta - s)x} dx.$$

Dentro de la integral nos aparece el nucleo de una v.a. gamma, de parámetros

$$\alpha, \quad (1/\beta - s)^{-1} = \beta/(1 - s\beta).$$

Si $s \geq 1/\beta$, entonces la integral es ∞. Sin embargo, si $s < 1/\beta$, entonces la integral es finita, y multiplicando y dividiendo por las constantes necesarias, obtenemos la f.g.m.

$$M_X(s) = \frac{1}{\beta^\alpha}\left(\frac{\beta}{1-s\beta}\right)^\alpha \int_0^\infty \frac{1}{\Gamma(\alpha)}\left(\frac{\beta}{1-s\beta}\right)^\alpha x^{\alpha-1} e^{-\frac{x}{\beta/(1-s\beta)}}dx = \frac{1}{(1-s\beta)^\alpha}, \; s < 1/\beta.$$

Obsérvese que como $\beta > 0$, tomando $h = 1/\beta$, la f.g.m. existe en el intervalo $(-h, h)$.

Ejemplo 1.11. Una v.a. binomial mide el número de éxitos en n repeticiones idénticas e independientes de un experimento con dos resultados posibles, "éxito" o "fracaso", con probabilidad de éxito π. La función de cuantía de dicha v.a. es

$$p(X = x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}, \; x = 0, 1, 2, \ldots$$

Obtenemos su f.g.m.

$$M_X(s) = \sum_{x=0}^{\infty} e^{sx} \binom{n}{x} \pi^x (1-\pi)^{n-x} = \sum_{0}^{\infty} \binom{n}{x} (\pi e^s)^x (1-\pi)^{n-x}.$$

Pero por el binomio de Newton, sabemos que

$$(a+b)^n = \sum_{x=0}^{\infty} \left(\begin{array}{c} n \\ x \end{array} \right) a^x b^{n-x},$$

con lo cual, la función generatriz de momentos es

$$M_X(s) = (\pi e^s + 1 - \pi)^n, \ \forall s \in I\!R.$$

La f.g.m. de una v.a. X genera los momentos respecto al origen de dicha variable aleatoria.

Teorema 1 Sea una v.a. X con f.g.m. $M_X(s)$. Entonces,

$$E[X^n] = M_X^{(n)}(0),$$

donde $M_X^{(n)}(0)$ denota la derivada n-ésima de la función generatriz de momentos repecto de s, evaluada en el punto $s = 0$.

Teorema 2 (Unicidad de la f.g.m.)
Sean X e Y dos v.a. con f.g.m. $M_X(s)$ y $M_Y(s)$ respectivamente. Si

$$M_X(s) = M_Y(s),\ \forall s \in (-h, h),$$

entonces X e Y tienen la misma distribución de probabilidad.

La distribución de probabilidad de una v.a. está determinada si se proporciona su función generatriz de momentos.

Propiedades de la f.g.m.:

(i) Sea X una v.a. con f.g.m. M_X, y definimos la v.a. $Y = aX + b$, donde $a, b \in \mathbb{R}$, $a \neq 0$. La f.g.m. de Y es

$$M_Y(s) = e^{sb} M_X(as).$$

(ii) Sean $X_1, \ldots, X_n$ v.a. independientes con f.g.m. $M_1, \ldots, M_n$ respectivamente. Sea la v.a. $Y = X_1 + \cdots + X_n$. La f.g.m. de Y es

$$M_Y(s) = \prod_{i=1}^{n} M_i(s).$$

Prueba. Comenzamos con (i);

$$M_Y(s) = E[e^{sY}] = E[e^{s(aX+b)}] = E[e^{saX}e^{sb}] = e^{sb}E[e^{saX}] = e^{sb}M_X(as).$$

Demostramos ahora (ii);

$$\begin{aligned} M_Y(s) &= E[e^{sY}] = E\left[e^{s\sum_{i=1}^{n} X_i}\right] = E\left[e^{sX_1} \cdots e^{sX_n}\right] \\ &= E[e^{sX_1}] \cdots E[e^{sX_n}] = \prod_{i=1}^{n} M_i(s). \ \square \end{aligned}$$

Teorema 1.3. (Convergencia de la f.g.m.)
Sea X_n, $n = 1, 2, 3, \ldots$ una sucesión de v.a. con f.g.m. $M_{X_n}(s)$, $n = 1, 2, 3, \ldots$. Supongamos que para todo s en un entorno del cero, se cumple

$$\lim_{n\to\infty} M_{X_n}(s) = M_X(s),$$

donde $M_X(s)$ es una f.g.m. Entonces existe una única $f.d.$ F_X cuyos momentos vienen determinados por $M_X(s)$, y para todo x donde $F_X(x)$ sea continua, se verifica

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

Es decir, la convergencia de las funciones generatrices de momentos implica la convergencia de las funciones de distribución.

2. VECTORES ALEATORIOS BIDIMENSIONALES

2.1 DEFINICIÓN DE VECTOR ALEAT. BIDIMENSIONAL

σ-álgebra de Borel sobre $I\!R^2$: σ-álgebra engendrada por todos los subconjuntos de $I\!R^2$ de la forma

$$(\infty, a] \times (-\infty, b], \quad a, b \in I\!R.$$

Vector aleatorio: Sea $(\Omega, \mathcal{A})$ un espacio medible. Sea $\mathcal{B}_2$ la σ-álgebra de Borel sobre $I\!R^2$. Un *vector aleatorio (v.a.)* bidimensional es una aplicación

$$\begin{aligned} \mathbf{X} \ : \ & \Omega \longrightarrow I\!R^2 \\ & \omega \rightsquigarrow \mathbf{X}(\omega) = (X_1(\omega), X_2(\omega)) \end{aligned}$$

que es *medible*, es decir, que verifica

$$\forall B \in \mathcal{B}_2, \quad \mathbf{X}^{-1}(B) = \{\omega \in \Omega : (X_1(\omega), X_2(\omega)) \in B\} \in \mathcal{A}.$$

Un v.a. $\mathbf{X}$ transforma un espacio medible $(\Omega, \mathcal{A})$ en otro espacio medible $(I\!R^2, \mathcal{B}_2)$. La condición de medibilidad de la v.a. permite obtener la probabilidad de un subconjunto $B \in \mathcal{B}_2$ como la probabilidad de un subconjunto $\mathbf{X}^{-1}(B) \in \mathcal{A}$.

Teorema 3 $\mathbf{X} = (X_1, X_2)$ es un vector aleatorio bidimensional si y solo si X_1 y X_2 son variables aleatorias.

2.2 DISTRIBUCIÓN DE PROBABILIDAD

Probabilidad inducida por $\mathbf{X}$:

$$\begin{aligned} P_{\mathbf{X}} \ &: \ \mathcal{B}_2 \longrightarrow [0;1] \\ & \quad B \rightsquigarrow P_{\mathbf{X}}(B) = P(\mathbf{X}^{-1}(B)). \end{aligned}$$

Función de distribución de X:

$$\begin{aligned} F_{\mathbf{X}} \ &: \quad I\!R^2 \longrightarrow [0;1] \\ & (x_1, x_2) \rightsquigarrow F_{\mathbf{X}}(x_1, x_2) = P_{\mathbf{X}}((-\infty, x_1] \times (-\infty, x_2]). \end{aligned}$$

X_1, X_2 discretas $\Rightarrow \mathbf{X} = (X_1, X_2)$ *discreto*.

X_1, X_2 continuas $\Rightarrow \mathbf{X} = (X_1, X_2)$ *continuo*.

X_1 discreta, X_2 continua $\Rightarrow \mathbf{X} = (X_1, X_2)$ *mixto*.

Función de cuantía o de probabilidad de $\mathbf{X}$:

$$\begin{array}{rcl} p_{\mathbf{X}} & : & I\!\!R^2 \longrightarrow [0, 1] \\ & & (x_1, x_2) \rightsquigarrow p_{\mathbf{X}}(x_1, x_2) = P_{\mathbf{X}}(\{(x_1, x_2)\}). \end{array}$$

También se suele denotar $p(X = x_1, X_2 = x_2) = p_{\mathbf{X}}(x_1, x_2)$.

Se puede especificar la "distribución" de un v. a. discreto bien por su función de probabilidad inducida $P_{\mathbf{X}}$, bien por su función de distribución $F_{\mathbf{X}}$, o bien por su función de cuantía $p_{\mathbf{X}}$.

Función de densidad conjunta de $\mathbf{X}$: Función $f_{\mathbf{X}} : \mathbb{R}^2 \to \mathbb{R}$ definida de la forma

$$f_{\mathbf{X}}(x_1, x_2) = \begin{cases} 0 & \text{si } (x_1, x_2) \in S; \\ \frac{d^2 F_{\mathbf{X}}(x_1, x_2)}{dx_1 dx_2} & \text{si } (x_1, x_2) \notin S, \end{cases}$$

donde S es el conjunto de puntos donde $F_{\mathbf{X}}$ no es derivable.

2.3 DISTRIBUCIONES MARGINALES

Distribuciónes marginales: Sea $\mathbf{X} = (X_1, X_2)$ un v.a. discreto. Las funciones de cuantía marginales de X_1 y de X_2 son

$$p_{X_1}(x_1) = \sum_{x_2} p_{\mathbf{X}}(x_1, x_2), \quad p_{X_2}(x_2) = \sum_{x_1} p_{\mathbf{X}}(x_1, x_2).$$

Si $\mathbf{X} = (X_1, X_2)$ es continua, las densidades marginales de X_1 y X_2 son

$$f_{X_1}(x_1) = \int_{I\!R} f_{\mathbf{X}}(x_1, x_2) dx_2, \qquad f_{X_2}(x_2) = \int_{I\!R} f_{\mathbf{X}}(x_1, x_2) dx_1.$$

Ejemplo 6 Distribución conjunta del número de clientes entre las 11h y las 12h en dos cajas registradoras de un supermercado:

X_1/X_2	0	1	2	3
0	0.15	0.10	0	0
1	0.10	0.20	0.10	0
2	0	0.10	0.15	0.05
3	0	0	0.05	0

2.4 DISTRIBUCIONES CONDICIONADAS

Distribución de X_1 condicionada a $X_2 = x_2$:

✓ Si $\mathbf{X} = (X_1, X_2)$ es *discreto*, y $p_{X_2}(x_2) > 0$,

$$p_{X_1}(x_1 | X_2 = x_2) \;=\; \frac{p_{\mathbf{X}}(x_1, x_2)}{p_{X_2}(x_2)},$$

Se verifica que $\sum_{x_1} p_{X_1}(x_1 | X_2 = x_2) = 1$.

✓ Si $\mathbf{X} = (X_1, X_2)$ es *continuo*, y $f_{X_2}(x_2) > 0$,

$$f_{X_1}(x_1 | X_2 = x_2) \;=\; \frac{f_{\mathbf{X}}(x_1, x_2)}{f_{X_2}(x_2)}.$$

Por tanto $\int_{\mathbb{R}} f_{X_1}(x_1 | X_2 = x_2) dx_1 = 1$.

Ejemplo 7 (continuación) Distribución condicionada de X_2 dado X_1:

X_2/X_1	0	1	2	3
0	0.60	0.40	0	0
1	0.25	0.50	0.25	0
2	0	1/3	0.50	1/6
3	0	0	1	0

Teorema 4 (Teorema de Bayes)

Si $\mathbf{X} = (X_1, X_2)$ es un vector aleatorio discreto,

$$p_{X_1}(x_1|X_2 = x_2) \;=\; \frac{p_{X_2}(x_2|X_1 = x_1)p_{X_1}(x_1)}{p_{X_2}(x_2)}.$$

Para vectores aleatorios continuos, se reemplazan las funciones de cuantía por las de densidad.

2.5 INDEPENDENCIA ENTRE VARIABLES ALEATORIAS

Independencia de v.a: Dos variables aleatorias X_1 y X_2 son *independientes*, si el valor que tome una de ellas no influye en el valor de la otra, es decir:

$$p_{X_1}(x_1|X_2 = x_2) = p_{X_1}(x_1)$$

Por tanto para cualquier valor x_2 con $p_{X_2}(x_2) > 0$,

$$p_{\mathbf{X}}(x_1, x_2) = p_{X_1}(x_1) \cdot p_{X_2}(x_2).$$

Nota 1 X_1 y X_2 independientes $\Rightarrow Cov(X_1, X_2) = \rho(X_1, X_2) = 0$, pero la implicación contraria no es cierta.

Contraejemplo: $X \sim N(0, 1) \Rightarrow Cov(X, X^2) = 0$. Sin embargo, estas variables son fuertemente dependientes (de forma no lineal).

Sin embargo, si X_1 y X_2 tienen distribución conjuntamente normal, covarianza nula implica independencia.

Ejemplo 8 En el Ejemplo 6, ¿son X_1 y X_2 independientes?

Ejemplo 9 Lanzamos un dado azul y otro rojo. Consideremos el valor del dado azul (X), el valor del dado rojo (Y), y suma par ($Z = 0$) o impar ($Z = 1$). ¿Cuales de las v.a. X, Y y Z son independientes?

2.6 ESPERANZAS, COVARIANZAS Y CORRELACIONES

Vector de medias:

$$E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix}.$$

Matriz de varianzas y covarianzas:

$$Cov(\mathbf{X}) = E[(\mathbf{X}-E(\mathbf{X}))(\mathbf{X}-E(\mathbf{X}))'] = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{bmatrix}$$

Esta matriz es semidefinida positiva: $\mathbf{u}'Cov(\mathbf{X})\mathbf{u} \geq 0$, $\forall \mathbf{u} \in I\!R^2$.

Transformación lineal de un v.a: Sea A una matriz con dimensiones $\tilde{d} \times d$, e $\mathbf{Y} = A\mathbf{X}$ un vector aleatorio obtenido a partir del v.a. $\mathbf{X}$. Entonces

(i) $E(\mathbf{Y}) = AE(\mathbf{X})$;

(ii) $Cov(\mathbf{Y}) = ACov(\mathbf{X})A'$.

Ejemplo 10 Obtener el número total de clientes medio, su varianza y la matriz de varianzas y covarianzas de $(X_1, X_2)'$ para los datos del Ejemplo 6.

2.7 ESPERANZAS Y VARIANZAS CONDICIONADAS

Esperanza de X_1 condicionada a $X_2 = x_2$: Esperanza de la distribución de X_1 condicionada a $X_2 = x_2$.

Para variables discretas:

$$E[X_1|X_2 = x_2] = \sum_{x_1} x_1 p_{X_1}(x_1|X_2 = x_2)$$

Para variables continuas:

$$E[X_1|X_2 = x_2] = \int_{\mathbb{R}} x_1 f_{X_1}(x_1|X_2 = x_2) dx_1$$

Si x_2 es un valor conocido, $E[X_1|X_2 = x_2]$ es una constante.

Si x_2 es desconocido, $E[X_1|X_2 = x_2]$ es una función de x_2.

Si X_2 es aleatorio, $E[X_1|X_2]$ es una v.a. Aplicando el operador de esperanza a esta v.a. obtenemos la media de X_1:

$$E[E(X_1|X_2)] = E[X_1]$$

Esto nos permite calcular la esperanza de X_1 en etapas: primero obtener la esperanza condicionada de X_1 como función de X_2, y luego calcular la media de esta esperanza condicionada respecto a la distribución de X_2.

Varianza de X_1 condicionada a $X_2 = x_2$: Matriz de varianzas y covarianzas de la distribución de X_1 condicionada a $X_2 = x_2$.

Descomposición de la varianza:

$$Var(X_1) = E[Var(X_1|X_2)] + Var[E(X_1|X_2)]$$

El primer término promedia las varianzas condicionadas, mientras que el segundo mide la variabilidad de las medias condicionadas.

La varianza total de X_1 será tanto mayor que el promedio de las varianzas condicionadas cuanto mayor sea la variabilidad entre las medias condicionadas.

2.8 LA DISTRIBUCIÓN MULTINOMIAL

Suponemos que observamos con reemplazamiento n individuos al azar y los clasificamos en d grupos. Definimos las d v.a.s

$$X_i = \text{no. de individuos en la clase } i, \quad i = 1, \ldots, d.$$

Sea π_i la probabilidad de que un individuo cualquiera caiga en la clase i. Se verifica que $\sum \pi_i = 1$.

Función de probabilidad conjunta de $\mathbf{X} = (X_1, \ldots, X_d)'$:

$$P(X_1 = x_1, \ldots, X_d = x_d) = \frac{n!}{x_1! \cdots x_d!} \pi_1^{x_1} \cdots \pi_d^{x_d},$$

donde $\sum x_i = n$.

Propiedades:

(i) Las distribuciones marginales de las X_i son binomiales con $E(X_i) = n\pi_i$ y $Var(X_i) = n\pi_i(1-\pi_i)$, $i = 1, \ldots, d$.

(ii) Las covarianzas son

$$Cov(X_i, X_j) = -n\,\pi_i\,\pi_j.$$

(iii) Cualquier distribución condicionada es multinomial.

$$(X_1, \ldots, X_j)|(X_{j+1} = x_{j+1}, \ldots, X_d = x_d) \sim Mult(n'; \pi'_1, \ldots, \pi'_d),$$

donde

$$n' = n - x_{j+1} - \cdots - x_d, \quad \pi'_1 = \pi_1/(\pi_1 + \cdots + \pi_j), \ldots, \pi'_j = \pi_j/(\pi_1 + \cdots + \pi_j)$$

(iv) La matriz de varianzas y covarianzas de una distribución multinomial siempre es singular, puesto que hay una relación exacta entre sus componentes.

2.9 LA DISTRIBUCIÓN NORMAL MULTIVARIANTE

Un v.a. $\mathbf{X}$ sigue una distribución normal d-dimensional si su función de densidad es

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{|\Sigma|^{1/2}(2\pi)^{d/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

para una matriz simétrica semidefinida positiva Σ y un vector $\boldsymbol{\mu}$. Σ es la matriz de varianzas y covarianzas de $\mathbf{x}$, y $\boldsymbol{\mu}$ su vector de medias.

Propiedades:

(i) Cualquier distribución condicionada y marginal de v.a. conjuntamente normales es normal: Sea $\mathbf{X} = (X_1, X_2)$ y

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

con dimensiones correspondientes, entonces

$$\begin{aligned} X_1 &\sim N\left(\boldsymbol{\mu}_1, \Sigma_{11}\right) \\ X_1 | X_2 = x_2 &\sim N\left(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right) \end{aligned}$$

(ii) Si dos variables conjuntamente normales son incorreladas, también son independientes. En general, si Σ es una diagonal, la densidad conjunta se descompone en el producto de las marginales: Entre v.a. conjuntamente normales sólo pueden existir relaciones lineales.

(iii) Cualquier combinación lineal de v.a. conjuntamente normales es normal. Si A es una matriz $\tilde{d} \times d$,

$$A\mathbf{X} \sim N(A\boldsymbol{\mu}, A\Sigma A').$$

(iv) Si Σ es no singular, podemos obtener un v.a. $\mathbf{Z}$ normal d-dimensional estándar, $\mathbf{Z} \sim N(\mathbf{0}, I)$, aplicando

$$\mathbf{Z} = A^{-1}(\mathbf{x} - \boldsymbol{\mu}),$$

donde A es una raíz de Σ, $\Sigma = AA'$.

Ejemplo 11 Según el censo de 1980 de Estados Unidos, las proporciones de adultos clasificados en cinco categorías de edad son:

Edad	Proporción
18-24	0.18
25-34	0.23
35-44	0.16
45-64	0.27
65+	0.16

Si se seleccionan al azar cinco adultos de esta población, encuentre la probabilidad de que la haya una persona entre 18 y 24 años, dos entre 25 y 34 años, y dos entre 45 y 64.

Ejemplo 12 Las demandas semanales X_1 y X_2 de los artículos A y B siguen una distribución conjunta normal bivariante con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas Σ,

$$\boldsymbol{\mu} = \begin{pmatrix} 200 \\ 150 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 10 & 4 \\ 4 & 6 \end{pmatrix}.$$

(a) Un comercio C_1 vende el artículo A a 6 euros la unidad y el artículo B a 10 euros. Otro comercio C_2 vende los artículos A y B a 10 y 6 euros. Hallar la distribución conjunta de los ingresos semanales de los dos comercios por las ventas de los artículos.

(b) Obtener la probabilidad de que los ingresos semanales obtenidos con las ventas de los dos artículos en el comercio C_2 superen en por lo menos 200 euros a los del comercio C_1.