5. Distribuciones Conjuntas

Estadística Computacional - San Joaquín - 2022

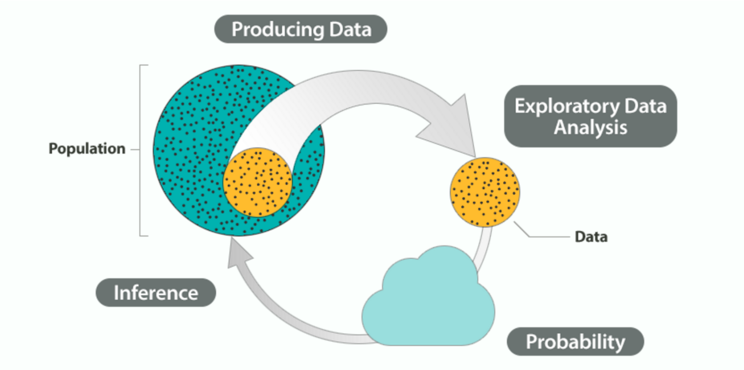
Ricardo Ñanculef, Manuel Goyo
jnancu@inf.utfsm.cl, *manuel.goyo@usm.cl*

Departamento de Informática UTFSM

Departamento de Informática
Universidad Técnica Federico Santa María

¿Cómo estudiaremos la Estadística?

- 3 Partes: Estadística Descriptiva, Teoría de Probabilidades, Inferencia.
- Estudio de **múltiples** fenómenos aleatorios con modelos matemáticos numérico.



Distribuciones Conjuntas

Objetivo General

- Estudiar cómo interactúan dos o más v.a.
- Dadas dos v.a. X, Y, el conocimiento de X ¿Cómo cambia la probabilidad de que Y tome ciertos valores?
- ¿Cuál es el grado de asociación o dependencia entre las variables?

Ejemplo

- Sea X_i una v.a. $\mathcal{N}(\mu, \sigma^2)$ que representa la **temperatura** del i-ésimo día de un cierto año en una cierta ciudad.
- Sea Y_i una v.a. $Poi(\lambda)$ que representa el **número de helados vendidos** el i-ésimo día del mismo año en la misma ciudad.
- ¿Cómo interactúan X_i, Y_i? ¿Depende Y_i de X_i? Si sabemos el valor de X_{i-1}, ¿Podemos predecir Y_i? ¿Podemos acotar el rango de incerteza sobre del valor de Y_i?

Distribuciones para V.A. Multivariada

F.D.P Conjunta - Discreta

Definición. Caso Discreto

Sean X, Y dos variables aleatorias discretas. La función de probabilidad conjunta de X, Y se define como la sgte. función $f : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$:

$$f_{X,Y}(x_i, y_j) := P(X = x_i, Y = y_j)$$

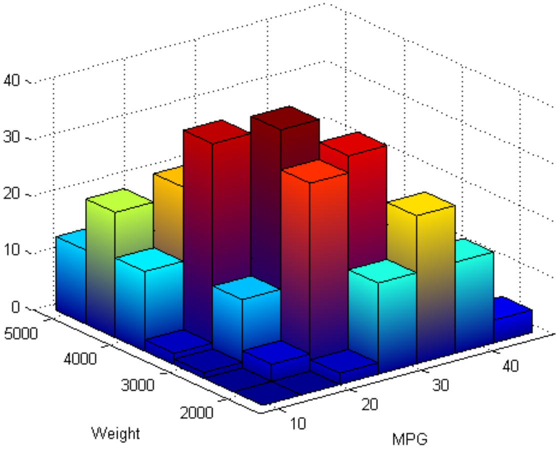
sólo si $X \in D_x$ e $Y \in D_y$, en otro caso $f_{X,Y}(x_i, y_j) = 0 \; \forall x_i \notin D_x$ ó $\forall y_j \notin D_y$.

- Ejemplo: Supongamos que medimos la temperatura usando una v.a. X que puede tomar 3 valores: 0 (baja), 1 (media), 2 (alta). Medimos la cantidad de helados vendidos usando una v.a. Y que puede tomar también 3 valores: 0 (bajas), 1 (medias), 2 (altas). La siguiente tabla resume la fdp conjunta de X e Y.

		Y		
		0	1	2
X	0	.10	.04	.02
	1	.08	.2	.06
	2	.06	.14	.30

F.D.P Conjunta - Discreta

Idea Gráfica - Caso Discreto



F.D.P Conjunta - Discreta

Definición. Caso Discreto

Sean X, Y dos variables aleatorias discretas con recorridos D_x y D_y respectivamente y función de probabilidad conjunta $f_{X,Y}$. Tenemos que $\forall A \subset D_x, B \subset D_y$

$$P(X \in A, Y \in B) = \sum_{x_i \in A} \sum_{y_j \in B} f_{X,Y}(x_i, y_j)$$

- Tenemos además,

$$\sum_{x_i \in D_x} \sum_{y_j \in D_y} f_{X,Y}(x_i, y_j) = P(X \in \mathbb{R}, Y \in \mathbb{R}) = 1$$

F.D.P Conjunta - Discreta

Ejemplo 1 - Discreto

- En el ejemplo anterior ...

		Y		
		0	1	2
X	0	.10	.04	.02
	1	.08	.20	.06
	2	.06	.14	.30

- ¿Es una fdp válida?
- ¿Cuál es la probabilidad de que $X = 0$ e $Y = 2$?
- ¿Cuál es la probabilidad de que $X > 1$ e $Y \geq 2$?

F.D.P Conjunta - Continua

Definición. Caso Continuo.

Sean X, Y dos variables aleatorias continuas. Si existe una función $f_{X,Y} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_0^+$ tal que

$$P(X \in A, Y \in B) = \int_A \int_B f_{X,Y}(x,y)dy\ dx\ ,$$

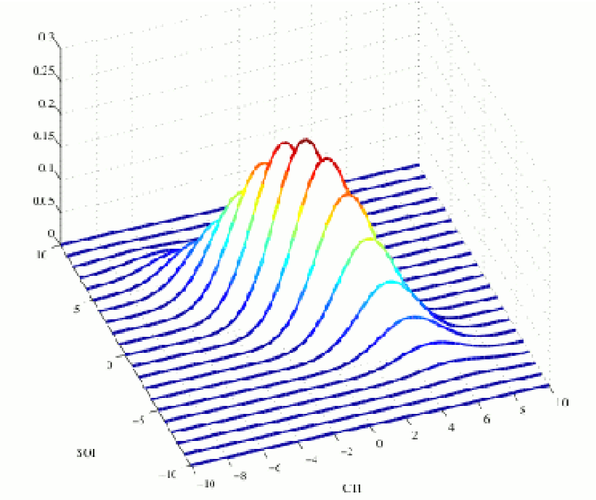
$\forall A, B \in \mathbb{R}$, $f_{X,Y}$ se denomina la función de densidad de probabilidad (fdp) conjunta de X, Y.

- Por lo tanto,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy\ dx = P(X \in \mathbb{R}, Y \in \mathbb{R}) = 1$$

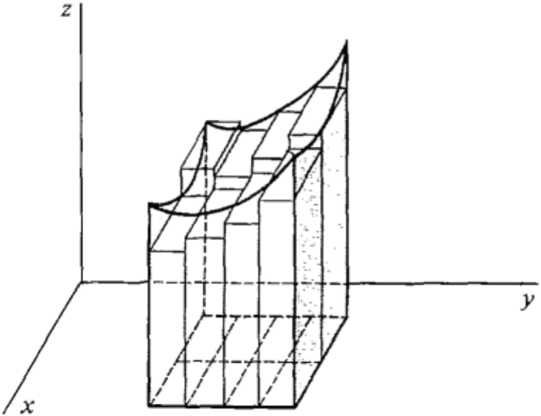
F.D.P Conjunta - Continua

Idea Gráfica - Caso Continuo



F.D.P Conjunta - Continua

Idea Gráfica - Caso Continuo



F.D.P Conjunta - Continua

Ejemplo 3 - Continuo

- Supongamos que tenemos dos ampolletas y medimos el tiempo (en meses) que dura cada una, usando las v.a. X e Y. La fdp conjunta de X e Y podría estar dada por

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{e^{-(2x+5y)/20}}{40} & \text{si } x, y > 0 \\ 0 & \text{etoc} \end{cases} .$$

- ¿Es una fdp válida?
- ¿Cuál es la probabilidad de que la primera dure más de 2 y la segunda más de 4?
- ¿Cuál es la probabilidad de que la primera dure más que la segunda?

F.D.P Conjunta - Continua

Ejemplo - Continuo

- ¿Es una fdp válida? En efecto,

$$
\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx\ dy &= \int_{0}^{\infty}\int_{0}^{\infty} \frac{\exp(-(2x+5y)/20)}{40}\ dx\ dy \\
&= \int_{0}^{\infty} \frac{\exp(-y/4)}{40} \int_{0}^{\infty} \exp(-x/10)\ dx\ dy \\
&= \int_{0}^{\infty} \frac{\exp(-y/4)}{40} \left(-10 \cdot \exp(-x/10)\Big|_{0}^{\infty}\right)\ dy \\
&= 1 \cdot \int_{0}^{\infty} \frac{\exp(-y/4)}{4}\ dy \\
&= 1 \cdot 1 = 1.0
\end{aligned}
$$

F.D.P Conjunta - Continua

Ejemplo - Continuo

- ¿Cuál es la probabilidad de que la primera dure más de $X = 2$ meses y la segunda más de $Y = 4$ meses?

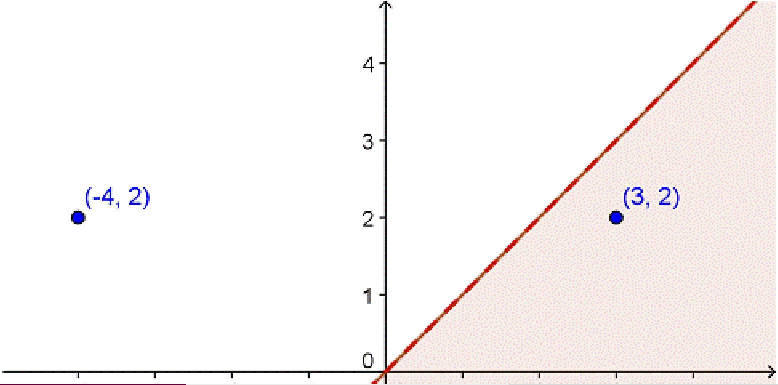
$$
\begin{aligned}
P(X > 2, Y > 4) &= \int_4^\infty \int_2^\infty f_{X,Y}(x,y)\ dx\ dy \\
&= \int_4^\infty \int_2^\infty \frac{\exp(-(2x+5y)/20)}{40}\ dx\ dy \\
&= \int_4^\infty \frac{\exp(-y/4)}{40} \int_2^\infty \exp(-x/10)\ dx\ dy \\
&= \int_4^\infty \frac{\exp(-y/4)}{40} \left(-10 \cdot \exp(-x/10)\big|_{x=2}^{x=\infty}\right)\ dy \\
&= \int_4^\infty \frac{\exp(-y/4)}{4} \exp(-2/10)\ dy \\
&= \exp(-2/10) \int_4^\infty \frac{\exp(-y/4)}{4}\ dy \\
&= \exp(-2/10)\exp(-4/4) = 0.287
\end{aligned}
$$

F.D.P Conjunta - Continua

Ejemplo - Continuo

- ¿Cuál es la probabilidad de que la primera dure más que la segunda?

$$P(X > Y) = \int_0^\infty \boxed{\int_y^\infty} f_{X,Y}(x,y) \boxed{dx} \, dy$$



F.D.P Conjunta - Continua

Ejemplo - Continuo

- ¿Cuál es la probabilidad de que la primera dure más que la segunda?

$$\begin{aligned}
P(X > Y) &= \int_0^\infty \boxed{\int_y^\infty} f_{X,Y}(x,y) \boxed{dx}\ dy \\
&= \int_0^\infty \int_y^\infty \frac{\exp(-(2x+5y)/20)}{40}\ dx\ dy \\
&= \int_0^\infty \frac{\exp(-y/4)}{40} \int_y^\infty \exp(-x/10)\ dx\ dy \\
&= \int_0^\infty \frac{\exp(-y/4)}{40} \left(-10 \cdot \exp(-x/10)\Big|_{x=y}^{x=\infty} \right)\ dy \\
&= \int_0^\infty \frac{\exp(-y/4)}{4} \exp(-y/10)\ dy = \int_0^\infty \frac{\exp(-7y/20)}{4}\ dy \\
&= -\frac{20}{7 \cdot 4} \exp(-7y/20)\Big|_{y=0}^{y=\infty} = \frac{20}{28} = 0.714\ .
\end{aligned}$$

F.D.P Conjunta - Continua

Ejemplo - Continuo

- ¿Cuál es la probabilidad de que la primera dure más que la segunda?

$$P(X > Y) = \int_0^{\infty} \boxed{\int_y^{\infty}} f_{X,Y}(x,y) \boxed{dx} \, dy$$

- **Cálculo alternativo**

$$P(X > Y) = \int_0^{\infty} \boxed{\int_0^{x}} f_{X,Y}(x,y) \boxed{dy} \, dx$$

F.D.P Conjunta - Continua

Ejemplo - Continuo

- **Cálculo alternativo**

$$
\begin{aligned}
P(X > Y) &= \int_0^\infty \boxed{\int_0^x} f_{X,Y}(x,y) \boxed{dy} \; dx = \int_0^\infty \int_0^x \frac{\exp(-(2x+5y)/20)}{40} \; dy \; dx \\
&= \int_0^\infty \frac{\exp(-x/10)}{40} \int_0^x \exp(-y/4) \; dy \; dx \\
&= \int_0^\infty \frac{\exp(-x/10)}{40} \left(-4 \cdot \exp(-y/4)\Big|_{y=0}^{y=x} \right) \; dx \\
&= \int_0^\infty \frac{\exp(-x/10)}{10} (1 - \exp(-x/4)) \; dx \\
&= \int_0^\infty \frac{\exp(-x/10)}{10} \; dx - \int_0^\infty \frac{\exp(-7x/20)}{10} \; dx \\
&= 1 + \frac{20}{7 \cdot 10} \exp(-7x/20) \Big|_{x=0}^{x=\infty} = 1 - \frac{20}{70} = 0.714 \; .
\end{aligned}
$$

Distribuciones Marginales

F.D.P Marginales.

Definición Caso Continuo

Sea $f_{X,Y}$ una f.d.p conjunta asociada a dos v.a. continuas X, Y. La **f.d.p. marginal** correspondiente a X se define como

$$f_X(x) := \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy$$

Análogo para Y: $f_Y(y) := \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx$.

Definición Caso Discreto

Sea $f_{X,Y}$ una f.d.p conjunta asociada a dos v.a. discretas X, Y con dominios D_x y D_y respectivamente. La **f.d.p. marginal** correspondiente a X se define como

$$f_X(x_i) := \sum_{y_j \in D_y} f(x_i, y_j)$$

Análogo para Y: $f_Y(y_i) := \sum_{x_i \in D_x} f(x_i, y_j)$.

F.D.P Marginales

Ejemplo 1 - Discreto

- Supongamos que medimos el tráfico en un cierta gran avenida de Santiago usando una v.a. X que puede tomar 3 valores: 0 (bajo), 1 (medio), 2 (alto). Hacemos lo mismo en otra avenida, usando esta vez una v.a. Y. La siguiente tabla resume la fdp conjunta de X e Y.

		Y		
		0	1	2
X	0	.10	.04	.02
	1	.08	.20	.06
	2	.06	.14	.30

- ¿Cuál es la fdp marginal de X?
- ¿Cuál es la fdp marginal de Y?

F.D.P Marginales

Ejemplo 1 - Discreto

- ¿Cuál es la fdp marginal de X? Veamos, $P(X = 0) = \sum_{y=0,1,2} f(x = 0, y) = 0.16$

		Y		
		0	1	2
X	0	.10	.04	.02
	1	.08	.2	.06
	2	.06	.14	.30

- $P(X = 1) = \sum_{y=0,1,2} f(x = 1, y) = 0.34$

		Y		
		0	1	2
X	0	.10	.04	.02
	1	.08	.2	.06
	2	.06	.14	.30

F.D.P Marginales

Ejemplo 1 - Discreto

- ¿Cuál es la fdp marginal de X? Veamos, $P(X = 2) = \sum_{y=0,1,2} f(x = 2, y) = 0.5$

		Y		
		0	1	2
X	0	.10	.04	.02
	1	.08	.2	.06
	2	.06	.14	.30

- Por lo tanto, la marginal de X es

$$f_X(x) = \begin{cases} 0.16 & \text{si } x = 0 \\ 0.34 & \text{si } x = 1 \\ 0.50 & \text{si } x = 2 \\ 0 & \text{etoc} \end{cases} .$$

- Tarea: calcule la marginal de Y.

F.D.P Marginales

Ejemplo 2 - Continuo

- Supongamos que tenemos dos ampolletas y medimos el tiempo (en meses) que dura cada una, usando las v.a. X e Y. La fdp conjunta de X e Y esta dada por

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{e^{-(2x+5y)/20}}{40} & \text{si } x, y > 0 \\ 0 & \text{etoc} \end{cases} .$$

- ¿Cuál es la fdp marginal de X?
- ¿Cuál es la fdp marginal de Y?

F.D.P Marginales

Ejemplo 2 - Continuo

- ¿Cuál es la fdp marginal de X? Por definición,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy\ .$$

- Si $x \leq 0$, $f_{X,Y}(x,y) = 0 \Rightarrow f_X(x) = 0$. Si $x > 0$,

$$\begin{aligned} f_X(x) &= \int_0^{\infty} \frac{\exp(-(2x+5y)/20)}{40}\ dy \\ &= \exp(-x/10) \int_0^{\infty} \frac{\exp(-y/4)}{40}\ dy \\ &= \exp(-x/10)\ \left(-\frac{\exp(-y/4)}{10}\bigg|_0^{\infty}\right) \\ &= \exp(-x/10)\frac{1}{10} \\ &= \frac{\exp(-x/10)}{10}\ \equiv\ Exp(\lambda = 10)\ . \end{aligned}$$

F.D.P Marginales

Ejemplo 2 - Continuo

- ¿Cuál es la fdp marginal de Y? Por definición,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx\ .$$

- Si $y \leq 0$, $f_{X,Y}(x,y) = 0 \Rightarrow f_Y(y) = 0$. Si $y > 0$,

$$\begin{aligned} f_Y(y) &= \int_0^{\infty} \frac{\exp(-(2x+5y)/20)}{40}\ dx \\ &= \exp(-y/4) \int_0^{\infty} \frac{\exp(-x/10)}{40}\ dx \\ &= \exp(-y/4)\ \left(-\frac{\exp(-y/10)}{4}\Big|_0^{\infty}\right) \\ &= \exp(-y/4)\frac{1}{4} \\ &= \frac{\exp(-y/4)}{4}\ \equiv\ Exp(\lambda = 4)\ . \end{aligned}$$

F.D.P Marginales.

Teorema

Sean X, Y dos v.a. con fdp conjunta $f_{X,Y}(x,y)$ y marginales $f_X(x)$ y $f_Y(y)$:

Probabilidad	v.a. continua	v.a. discreta
$P(X \in A)$	$\int_A f_X(x)\ dx$	$\sum_{x_i \in A} f_X(x_i)$
$P(Y \in B)$	$\int_B f_Y(y)\ dy$	$\sum_{y_j \in B} f_y(y_j)$

- Basta observar que :

$$P(X \in A) = P(X \in A, Y \in \mathbb{R}) = \int_A \underbrace{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy}_{\text{Marginal de } X}\ dx$$

$$= \int_A f_X(x)\ dx\ .$$

F.D.P Marginales

Ejemplo

- Supongamos que tenemos dos ampolletas y medimos el tiempo (en meses) que dura cada una, usando las v.a. X e Y. La fdp conjunta de X e Y esta dada por

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{e^{-(2x+5y)/20}}{40} & \text{si } x, y > 0 \\ 0 & \text{etoc} \end{cases} .$$

- ¿Cuál es la probabilidad de que la primera ampolleta dure más de 2 semanas?
- ¿Cuál es la probabilidad de que la segunda ampolleta dure menos de 1 semana?

F.D.P Marginales

Ejemplo

- En ejercicios anteriores vimos que

$$f_X(x) = \frac{\exp(-x/10)}{10} \equiv Exp(10)$$
$$f_Y(y) = \frac{\exp(-y/4)}{4} \equiv Exp(4) \ .$$

- Entonces,

$$P(X > 14) = \int_{14}^{\infty} \frac{\exp(-x/10)}{10} \, dx = 1 - F_{exp(10)}(14) = 0.246$$
$$P(Y < 7) = \int_{0}^{7} \frac{\exp(-y/4)}{4} \, dy = F_{exp(4)}(7) = 0.826 \ .$$

Observaciones para cálculo de integrales

Observaciones

- Orden de integración.
- Cambios de variable.
- Asociar fdp marginales a distribuciones notables
- Asociar cálculos a integrales conocidas.

Valor Esperado y Momentos

Valor Esperado de una función sobre una VA Multivariada

Caso Continuo

Sea $f_{X,Y}$ una f.d.p conjunta asociada a dos v.a. continuas X, Y. Sea $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ una función de X e Y. Definimos

$$E_{XY}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \cdot f_{X,Y}(x,y) \; dy \; dx \; .$$

Caso Discreto

Sea $f_{X,Y}$ una f.d.p conjunta asociada a dos v.a. discretas X, Y. Sea $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ una función de X e Y. Definimos

$$E_{XY}[g(X,Y)] = \sum_{x_i \in D_x} \sum_{y_j \in D_y} g(x_i, y_j) \cdot f_{X,Y}(x_i, y_j) \; .$$

Momentos Producto

Caso Continuo

Valgan todas las definiciones anteriores. Sea $g(X, Y) = X^k \cdot Y^s$. Entonces,

$$E_{XY}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^k \cdot y^s \; f_{X,Y}(x, y) \; dy \; dx$$

se denomina el k, s-ésimo momento producto.

Caso Discreto

Valgan todas las definiciones anteriores. Sea $g(X, Y) = X^k \cdot Y^s$,

$$E_{XY}[g(X, Y)] = \sum_{x_i \in D_x} \sum_{y_j \in D_y} x^k \cdot y^s \; f_{X,Y}(x_i, y_j)$$

se denomina el k, s-ésimo momento producto.

Momentos Marginales de X

Caso Continuo

Valgan todas las definiciones anteriores. Sea $g(X,Y) = X^k$. Entonces,

$$E_{XY}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k \; f_{X,Y}(x,y)\; dy\; dx = \int_{-\infty}^{\infty} x^k \; \int_{-\infty}^{\infty} f_{X,Y}(x,y)\; dy\; dx$$

$$E_X[X^k] = \int_{-\infty}^{\infty} x^k \; f_X(x)\; dx$$

Caso Discreto

Valgan todas las definiciones anteriores. Sea $g(X,Y) = X^k$. Entonces,

$$E_{XY}[g(x,y)] = \sum_{x_i \in D_x}\sum_{y_j \in D_y} x_i^k \; f_{X,Y}(x_i,y_j)$$

$$E_X[X^k] = \sum_{x_i \in D_x} x_i^k \; f_X(x_i)$$

Análogo para Y.

Valores Esperados Marginales

Caso Continuo

$$E_X[X] = \int_{-\infty}^{\infty} x \ f_X(x) \ dx$$

$$E_Y[Y] = \int_{-\infty}^{\infty} y \ f_Y(y) \ dy$$

Caso Discreto

$$E_X[X] = \sum_{x_i \in D_x} x_i \ f_X(x_i)$$

$$E_Y[Y] = \sum_{y_j \in D_y} y_j \ f_Y(y_j)$$

Independencia y Correlación

Independencia.

Definición Caso Continuo

Dos v.a. continuas X, Y se dicen **independientes** si y solo si

$$f_{X,Y}(x,y) := f_X(x) \cdot f_Y(y) \quad \forall x, y$$

Definición Caso Discreto

Dos v.a. discretas X, Y con dominios D_x y D_y respectivamente se dicen **independientes** si y solo si

$$f_{X,Y}(x_i, y_j) = f_X(x_i) \cdot f_Y(y_i) \quad \forall x_i \in D_x \ \forall y_j \in D_y$$

Independencia.

Ejemplo

- Supongamos que la fdp conjunta de X e Y esta dada por

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{e^{-(2x+5y)/20}}{40} & \text{si } x,y > 0 \\ 0 & \text{etoc} \end{cases} .$$

- ¿Son independientes?

Independencia.

Ejemplo

- En ejercicios anteriores vimos que

$$f_X(x) = \frac{\exp(-x/10)}{10} \text{ si y sólo si } x > 0$$
$$f_Y(y) = \frac{\exp(-y/4)}{4} \text{ si y sólo si } y > 0 \ .$$

- Ya que $f_X(x)$ y $f_Y(y)$ están definidos, podemos verificar en este caso

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{e^{-(2x+5y)/20}}{40} \\ &= \frac{\exp(-x/10)}{10} \cdot \frac{\exp(-y/4)}{4} \\ &= f_X(x) \cdot f_Y(y) \ . \end{aligned}$$

- Por lo tanto, X, Y son independientes. Esto implica que X (tiempo ampolleta 1) no entrega información sobre Y (tiempo ampolleta 2), ni viceversa. Ambos fenómenos ocurren de manera independiente sin relación o dependencia entre sí.

Covarianza

Caso Continuo

Sea $g(X,Y) = (X - E[X])\,(Y - E[Y])$, con $E[X], E[Y]$ los valores esperados de X e Y bajo su respectiva fdp marginal: $f_X(x)$, $f_Y(y)$. Entonces,

$$cov(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - E[X])\,(y - E[Y])\;\; f_{X,Y}(x,y)\; dy\; dx$$

se denomina la covarianza entre X e Y.

Caso Discreto

Sea $g(X,Y) = (X - E[X])\,(Y - E[Y])$, con $E[X], E[Y]$ los valores esperados de X e Y bajo su respectiva fdp marginal: $f_X(x)$, $f_Y(y)$. La covarianza entre X e Y:

$$cov(X,Y) = \sum_{x_i \in D_x}\sum_{y_j \in D_y} (x_i - E[X])\,(y_j - E[Y])\;\; f_{X,Y}(x_i, y_j)$$

Covarianza

- Covarianza: ¿Qué tan fuerte están relacionadas dos v.a. entre sí? ¿Cómo varía (linealmente) una variable cuando la otra varía también?

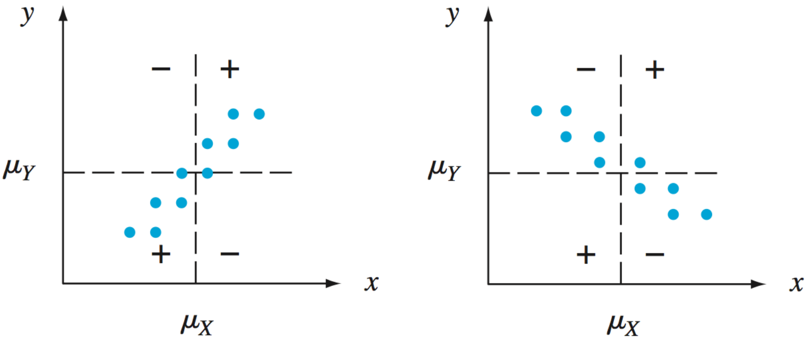
Teorema

Sean X, Y dos variables aleatorias. Tenemos

$$cov(X, Y) = E_{XY}[X \cdot Y] - E_X[X]E_Y[Y]$$

Covarianza

Idea Gráfica



Correlación

Definición

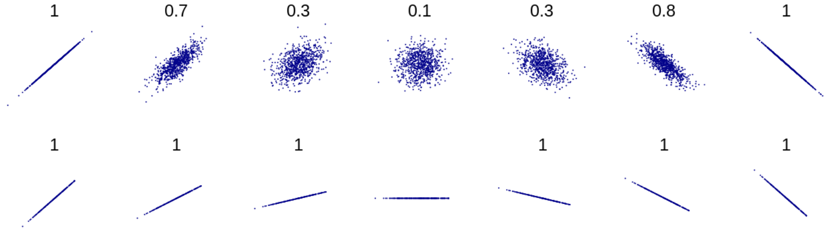
Sean X, Y dos v.a. Entonces, el **Coeficiente de Correlación de Pearson** es una medida de la asociación/dependencia entre X e Y y se define como

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} \ .$$

Claramente, $\rho_{X,Y} \in [-1, 1]$.

Correlación

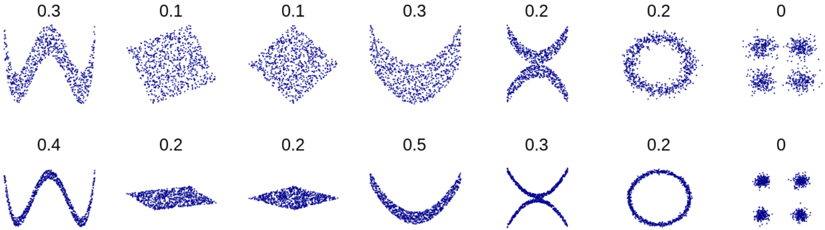
Tipos de Dependencia Detectada usando $|\rho|$



- Si $|\rho| = 1 \leftrightarrow Y = aX + b$, con $a \neq 0$ (dependencia completamente lineal)
- Usualmente:
  - $|\rho| \geq 0.5$ relación fuerte o alta.
  - $|\rho| \geq 0.3$ relación moderadamente alta.

Correlación

Tipos de Dependencia NO Detectada usando $|\rho|$



Propiedades.

Teorema

Sean X, Y dos variables aleatorias independientes,

$$E_{XY}[X \cdot Y] = E_X[X]E_Y[Y]$$

Además, claramente

- X,Y independientes $\implies$ $cov(X, Y) = 0, \rho_{X,Y} = 0$.

WARNING!

$$\begin{aligned} cov(X, Y) = 0 \quad &\not\Longrightarrow \quad \text{X,Y independientes} \\ \rho_{X,Y} = 0 \quad &\not\Longrightarrow \quad \text{X,Y independientes} \end{aligned}$$

Propiedades.

Teorema

Sean X, Y dos v.a. Entonces,

$$Var[X+Y] = Var[X] + Var[Y] + 2cov(X,Y)$$
$$Var[X-Y] = Var[X] + Var[Y] - 2cov(X,Y) \ .$$

Recordemos además que $E_{XY}[X+Y] = E_X[X] + E_Y[Y]$

Corolario

Sean X, Y dos v.a. independientes,

$$Var[X \pm Y] = Var[X] + Var[Y] \ .$$

Ejemplo Covarianza

Ejemplo

- Supongamos que la fdp conjunta de X e Y esta dada por

$$f_{X,Y}(x,y) = \begin{cases} x^2 \cdot e^{-x(1+y)} & \text{si } x, y \geq 0 \\ 0 & \text{etoc} \end{cases} .$$

- Determine $cov(X,Y)$.

Ejemplo Covarianza

Del teorema, tenemos $cov(X, Y) = E[XY] - E[X]E[Y]$. Para calcular $E[X]$ y $E[Y]$ hacemos uso de las marginales, además así verificamos independencia.

Ejemplo

- Marginal de X,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy = \int_0^{\infty} x^2 \cdot e^{-x(1+y)}\ dy$$
$$= x^2 e^{-x} \int_0^{\infty} e^{-xy}\ dy = x^2 e^{-x} \left(-\frac{1}{x} \cdot e^{-xy} \Big|_{y=0}^{y=\infty} \right)$$
$$= xe^{-x}(1-0) = xe^{-x}\ .$$

- Marginal de Y,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx = \int_0^{\infty} x^2 \cdot e^{-x(1+y)}\ dx$$

- Cambiando variable $u = x(1+y)$,

$$f_Y(y) = \int_0^{\infty} \frac{u^2}{(1+y)^2} \cdot e^{-u}\ \frac{du}{(1+y)} = \frac{1}{(1+y)^3}\Gamma(3) = \frac{2}{(1+y)^3}$$

Ejemplo Covarianza

Resulta evidente que X, Y no son independientes.

Ejemplo

- Calculamos $E[X]$ (se trata de una $Gamma(2,1)$),

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)\ dx = \int_{0}^{\infty} x^2 e^{-x}\ dx = \Gamma(3) = 2\ .$$

- Calculamos $E[Y]$,

$$E[Y] = \int_{-\infty}^{\infty} y \cdot f_Y(y)\ dy = \int_{0}^{\infty} \frac{2y}{(1+y)^3}\ dy\ .$$

- Integrando por partes con $u = 2y$ y $dv = (1+y)^{-3}dy$

$$\begin{aligned} E[Y] &= -\frac{y}{(1+y)^2}\Bigg|_{y=0}^{y=\infty} + \int_{0}^{\infty} \frac{1}{(1+y)^2}\ dy \\ &= 0 + -\frac{1}{(1+y)}\Bigg|_{y=0}^{y=\infty} = 1\ . \end{aligned}$$

Ejemplo Covarianza

Ejemplo

- Para calcular $E[XY]$, debemos resolver

$$E[\boxed{XY}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \boxed{xy} \cdot f_{X,Y}(x,y)\ dx\ dy = \int_{0}^{\infty}\int_{0}^{\infty} xy \cdot x^2 \cdot e^{-x(1+y)}\ dx\ dy$$
$$= \int_{0}^{\infty}\int_{0}^{\infty} x^3 y e^{-x(1+y)}\ dx\ dy = \int_{0}^{\infty} y \int_{0}^{\infty} x^3 e^{-x(1+y)}\ dx\ dy$$

- La integral de más adentro la resolvemos cambiando variable $u = x(1+y)$. Tenemos $du = dx(1+y)$, porque (1+y) no depende de x. Entonces,

$$\int_{0}^{\infty} x^3 e^{-x(1+y)}\ dx = \int_{0}^{\infty} \frac{u^3}{(1+y)^3} e^{-u} \frac{du}{(1+y)}$$
$$= \frac{1}{(1+y)^4}\int_{0}^{\infty} u^3 e^{-u} du = \frac{1}{(1+y)^4}\Gamma(4) = \frac{6}{(1+y)^4}$$

- Notemos que $(1+y) > 0$. Por eso mantuvimos los límites de integración.

Ejemplo Covarianza

Ejemplo

- Entonces,

$$E[XY] = \int_0^\infty y \int_0^\infty x^3 e^{-x(1+y)}\ dx\ dy$$
$$= \int_0^\infty \frac{6y}{(1+y)^4}\ dy$$

- Integramos por partes con $u = 6y$, $dv = (1+y)^{-4}dy$, que nos da $du = 6dy$, $v = -3^{-1}(1+y)^{-3}$. Reemplazado en $(\int udv = uv - \int vdu)$,

$$E[XY] = -2y(1+y)^{-3}\Big|_{y=0}^{y=\infty} + 2\int_0^\infty (1+y)^{-3}dy$$
$$= 0 - (1+y)^{-2}\Big|_{y=0}^{y=\infty} = 1\ .$$

- Tenemos entonces que, $cov(X,Y) = E[XY] - E[X]E[Y] = 1 - 2 \cdot 1 = -1$.

Distribuciones Condicionales

F.D.P Condicionales (de X a Y).

Motivación

Sean X, Y dos variables aleatorias. Si sospechamos que al conocer algo de Y podemos inferir algo acerca de X, nos gustaría calcular probabilidades del tipo

$$P(a \leq X \leq b | c \leq Y \leq d)$$
$$P(a \leq X \leq b | Y = c)$$

Si X es independiente de Y

$$P(a \leq X \leq b | c \leq Y \leq d) = P(a \leq X \leq b)$$
$$P(a \leq X \leq b | Y = c) = P(a \leq X \leq b)$$

Pero sino podríamos tener una reducción de la incerteza acerca de los valores que toma X: información!

F.D.P Condicionales (de X a Y).

Motivación

Usando Bayes

$$P(a \leq X \leq b | c \leq Y \leq d) = \frac{P((a \leq X \leq b) \cap (c \leq Y \leq d))}{P(c \leq Y \leq d)}$$

El numerador se puede calcular integrando (o sumando en el caso discreto) la fdp conjunta, y el denominador integrando (o sumando en el caso discreto) la marginal de Y

$$P(a \leq X \leq b | c \leq Y \leq d) = \frac{\int_c^d \int_a^b f_{X,Y}(x,y) dx dy}{\int_c^d f_Y(y) dy}$$

OK, funciona.

F.D.P Condicionales (de X a Y).

Motivación

Sin embargo, en el otro caso (condicional a un valor puntual)

$$P(a \leq X \leq b | Y = c) = \frac{P((a \leq X \leq b) \cap (Y = c))}{P(Y = c)}$$

El cálculo anterior no tiene sentido en el caso continuo porque tanto el numerador como el denominador se anulan.

F.D.P Condicionales (de X a Y).

Definición Caso Discreto

Sea $f_{X,Y}$ una f.d.p conjunta asociada a dos v.a. discretas X, Y con dominios D_x y D_y respectivamente. La f.d.p. condicional de X a Y se define como

$$f_{X|Y=y_j}(x_i) := \frac{f_{X,Y}(x_i, y_j)}{f_Y(y_j)}$$

Definición Caso Continuo

Sea $f_{X,Y}$ una f.d.p conjunta asociada a dos v.a. continuas X, Y. La f.d.p. condicional de X a Y se define como

$$f_{X|Y=y}(x) := \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

F.D.P Condicionales (de X a Y).

Definición

Sean X, Y dos variables aleatorias continuas. Sea $f_{X|Y=y}(x)$ la f.d.p. condicional de X a Y. Entonces

$$P(X \in A|Y = y) := \int_A f_{X|Y=y}(x)\ dx$$

Observación

Sean X, Y dos variables aleatorias independientes,

$$f_{X|Y=y}(x) = f_X(x)$$

IMPORTANTE: Todo esto aplica en sentido recíproco (de Y a X).

Momentos Condicionales (de X a Y).

Definición Caso Discreto

Sea $f_{X,Y}$ una f.d.p conjunta asociada a dos v.a. discretas X, Y con dominios D_x y D_y respectivamente. Sea $g : \mathbb{R} \to \mathbb{R}$ una función de X. Definimos,

$$E[g(X)|Y = y_j] := \sum_{x_i \in D_x} g(x_i) \cdot f_{X|Y=y_j}(x_i)$$

Definición Caso Continuo

Sea $f_{X,Y}$ una f.d.p conjunta asociada a dos v.a. continuas X, Y. Sea $g : \mathbb{R} \to \mathbb{R}$ una función de X. Definimos,

$$E[g(X)|Y = y] := \int_{-\infty}^{\infty} g(x) \cdot f_{X|Y=y}(x)\ dx$$

Momentos Condicionales (de X a Y).

Caso $g(X) = X$ Discreto

$$E[X|Y = y_j] := \sum_{x_i \in D_x} x_i f_{X|Y=y_j}(x_i)$$

Caso $g(X) = X$ Continuo

$$E[X|Y = y] := \int_{-\infty}^{\infty} x \ f_{X|Y=y}(x) \ dx$$

Momentos Condicionales (de X a Y).

Caso $g(X) = (X - E[X])^2$ Discreto

$$Var[X|Y = y_j] := \sum_{x_i \in D_x} (x_i - E[X])^2 \cdot f_{X|Y=y_j}(x_i)$$
$$= E[X^2|Y = y_j] - E[X|Y = y_j]^2$$

Caso $g(X) = (X - E[X])^2$ Continuo

$$Var[X|Y = y_j] := \int_{-\infty}^{\infty} (x - E[X])^2 \cdot f_{X|Y=y}(x)\ dx$$
$$= E[X^2|Y = y] - E[X|Y = y]^2$$

IMPORTANTE: Todo esto aplica en sentido recíproco (de Y a X).

Ejemplo

Ejemplo Condicional

Sea X el número de cámaras digitales Canon vendidas durante una semana particular por una tienda. La función de masa de probabilidad de X es

x	0	1	2	3	4
$f_X(x)$	0.10	0.20	0.30	0.25	0.15

60 % de todos los clientes que compran estas cámaras también compran una garantía extendida. Sea Y el número de compradores durante esta semana que compran una garantía extendida.

- ¿Cuál es la $P(X = 4, Y = 2)$?
- ¿Qué tan probable es que todos los que compran la cámara elijan la garantía?
- Exprese el cálculo para la función de probabilidad marginal de Y.

Generalizaciones

F.D.P Conjuntas. Generalizaciones.

Definición

Sean $X_1, X_2, \ldots, X_n$, n variables aleatorias discretas. La función de densidad probabilidad (fdp) conjunta asociada a $X_1, X_2, \ldots, X_n$ se define como

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) := P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

Definición

Sean $X_1, X_2, \ldots, X_n$, n variables aleatorias continuas. Si existe una función $f_{X_1,X_2,\ldots,X_n} : \mathbb{R}^n \to \mathbb{R}$ tal que

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = \int_{A_1} \ldots \int_{A_n} f_{X_1,X_2,\ldots,X_n}(x_1, \ldots, x_n)\; dx_n \ldots dx_1 \ ,$$

entonces $f_{X_1,X_2,\ldots,X_n}$ se denomina la función de densidad probabilidad (fdp) conjunta asociada a $X_1, X_2, \ldots, X_n$.

Marginales. Generalizaciones.

Definición

Sean $X_1, X_2, \ldots, X_n$, n variables aleatorias discretas. Sea D_i el dominio de X_i. La f.d.p marginal de X_i, se define como

$$f_{X_i}(x_i) = \sum_{x_1 \in D_1} \cdots \sum_{x_{i-1} \in D_{i-1}} \sum_{x_{i+1} \in D_{i+1}} \cdots \sum_{x_n \in D_n} f_{X_1, X_2, \ldots, X_n}(x_1, \ldots, x_n)$$

Definición

Sean $X_1, X_2, \ldots, X_n$, n variables aleatorias continuas. La f.d.p marginal de X_i, se define como

$$f_{X_i}(x_i) = \int_{A_1} \cdots \int_{A_{i-1}} \int_{A_{i+1}} \cdots \int_{A_n} f_{X_1, X_2, \ldots, X_n}(x_1, \ldots, x_n)\; dx_n \ldots dx_{i-1}\; dx_{i+1} \ldots dx_1$$

Independencia. Generalizaciones.

Definición

n variables aleatorias $X_1, X_2, \ldots, X_n$ se dicen independientes si y solo si

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$