

# Estadística y Procesos Estocásticos

## Tema 3: Distribuciones Multivariantes

Grado en Ingeniería en Tecnologías de la Telecomunicación

A detailed illustration of a satellite in space. The satellite is white and cylindrical with two large solar panel arrays extending from its sides. It has two large parabolic dish antennas at the rear. In the background, a large, bright sun is partially obscured by the horizon of a planet, creating a lens flare effect. The sky is dark with some faint stars and a large, glowing orange and red nebula or aurora-like structure.

# 1. Variables Aleatorias bidimensionales. Introducción.



## Ejemplo:



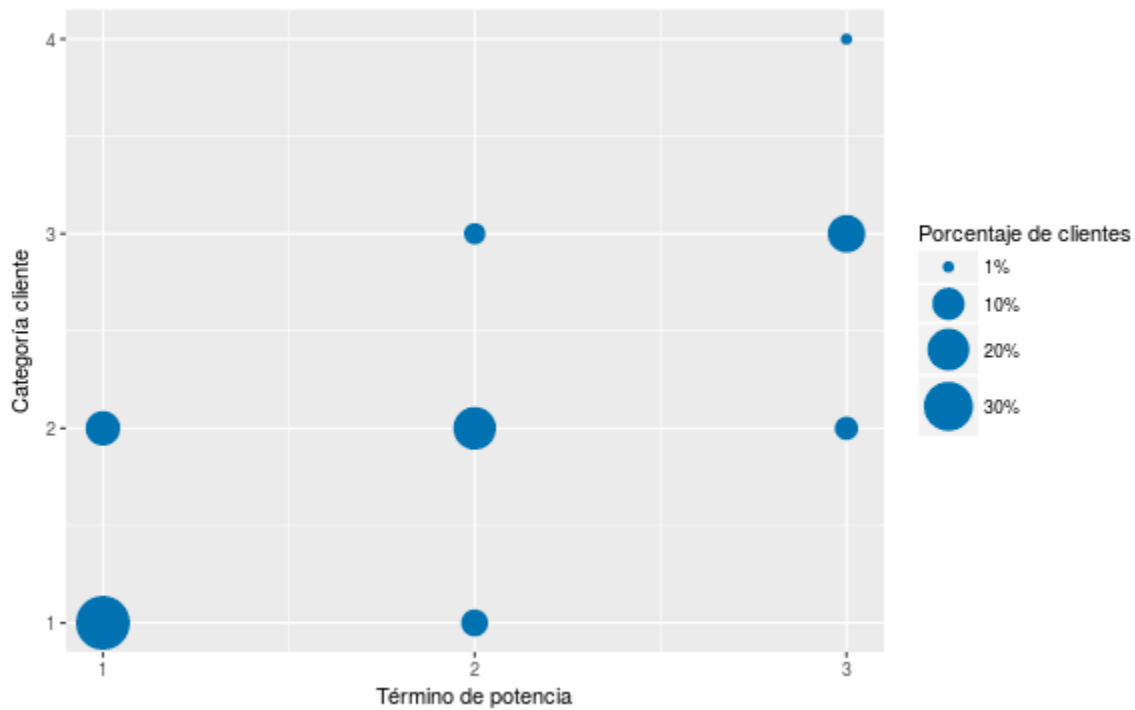
Una compañía eléctrica clasifica a sus clientes en tres categorías: 1 (domésticos), 2 (empresas del sector servicios), y 3 (empresas del sector industrial). Asimismo, el término de potencia contratado por cada cliente puede ser 1 (bajo), 2 (medio), 3 (alto) ó 4 (muy alto). La tabla siguiente muestra como se distribuyen los clientes de la compañía según categoría y término de potencia contratado:

Categoría	Término de potencia			
	1 (bajo)	2 (medio)	3 (alto)	4 (muy alto)
<b>1 (uso doméstico)</b>	38%	12%	0%	0%
<b>2 (sector servicios)</b>	6%	21%	3%	0%
<b>3 (sector industrial)</b>	0%	4%	15%	1%

Varias representaciones gráficas posibles:

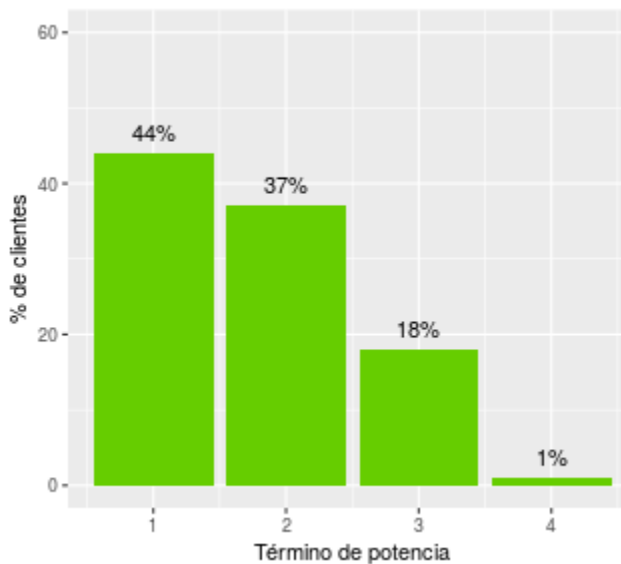
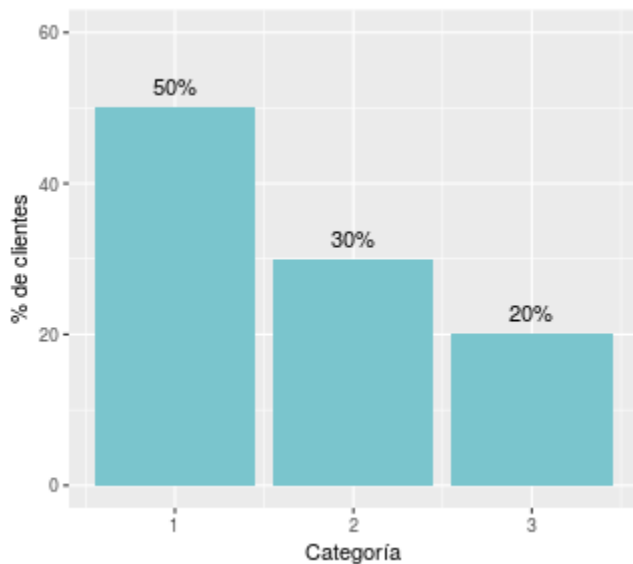
	<b>Término de potencia</b>			
<b>Categoría</b>	<b>1 (bajo)</b>	<b>2 (medio)</b>	<b>3 (alto)</b>	<b>4 (muy alto)</b>
<b>1 (uso doméstico)</b>	38%	12%	0%	0%
<b>2 (sector servicios)</b>	6%	21%	3%	0%
<b>3 (sector industrial)</b>	0%	4%	15%	1%

	Término de potencia			
Categoría	1 (bajo)	2 (medio)	3 (alto)	4 (muy alto)
<b>1 (uso doméstico)</b>	38%	12%	0%	0%
<b>2 (sector servicios)</b>	6%	21%	3%	0%
<b>3 (sector industrial)</b>	0%	4%	15%	1%



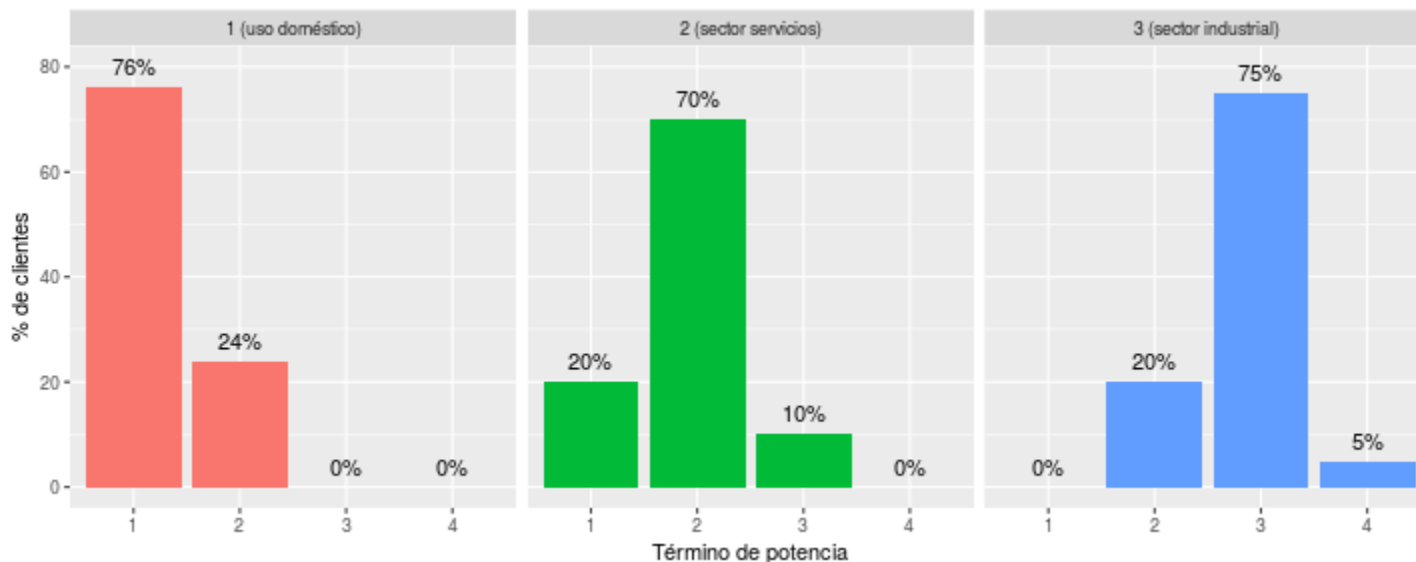
Sumando los totales por filas y por columnas se obtienen los **porcentajes de clientes en cada categoría** (en azul) y los **porcentajes de clientes según potencia contratada** (en verde)

Categoría	Término de potencia				Total
	1 (bajo)	2 (medio)	3 (alto)	4 (muy alto)	
<b>1 (uso doméstico)</b>	38%	12%	0%	0%	50%
<b>2 (sector servicios)</b>	6%	21%	3%	0%	30%
<b>3 (sector industrial)</b>	0%	4%	15%	1%	20%
<b>Total</b>	44%	37%	18%	1%	100%



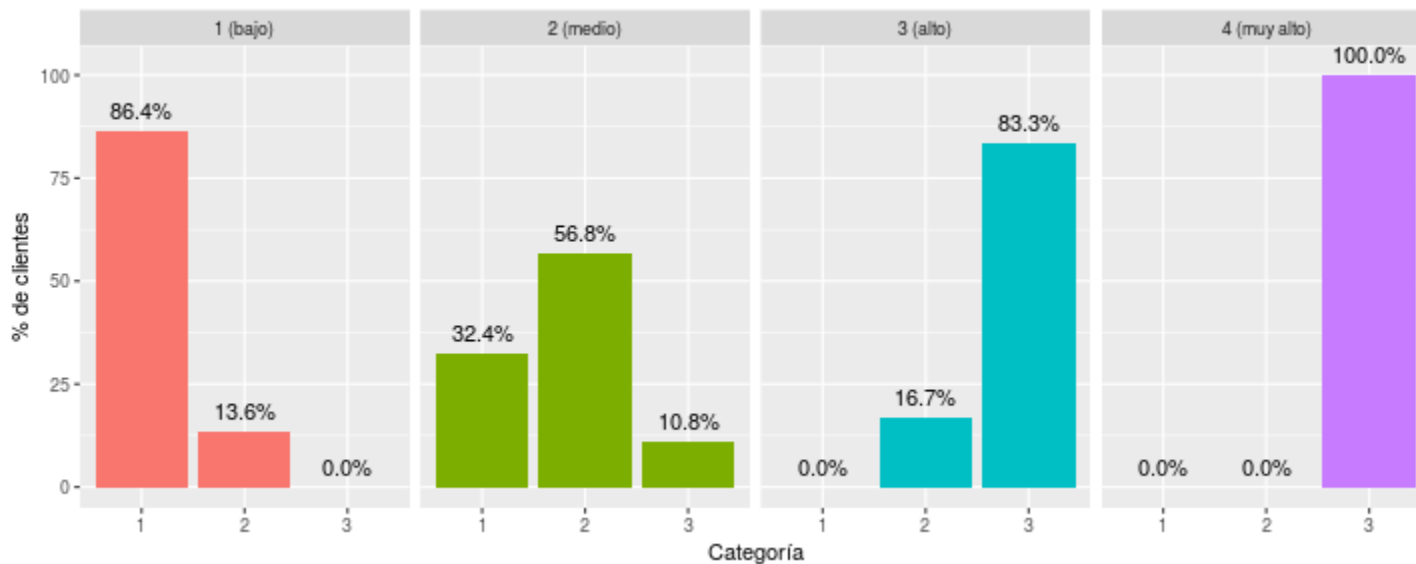
Si dividimos el porcentaje de cada casilla por el total de su fila, obtenemos los porcentajes por fila, que indican **como se reparten los términos de potencia contratados según la categoría del cliente**:

Categoría	Término de potencia				Total
	1 (bajo)	2 (medio)	3 (alto)	4 (muy alto)	
<b>1 (uso doméstico)</b>	76%	24%	0%	0%	100%
<b>2 (sector servicios)</b>	20%	70%	10%	0%	100%
<b>3 (sector industrial)</b>	0%	20%	75%	5%	100%
<b>Total</b>	44%	37%	18%	1%	100%



Si dividimos el porcentaje de cada casilla por el total de su columna, obtenemos los porcentajes por columna, que indican **como se reparten las categorías de clientes según la potencia contratada**:

Categoría	Término de potencia				Total
	1 (bajo)	2 (medio)	3 (alto)	4 (muy alto)	
<b>1 (uso doméstico)</b>	86.4%	32.4%	0%	0%	50%
<b>2 (sector servicios)</b>	13.6%	56.8%	16.7%	0%	30%
<b>3 (sector industrial)</b>	0%	10.8%	83.3%	100%	20%
<b>Total</b>	100%	100%	100%	100%	100%





Supongamos ahora que una empresa de auditoría debe revisar las cuentas de la compañía eléctrica. Para ello debe elegir al azar algunos clientes.

- ¿Cuál es la probabilidad de que elija un cliente doméstico que tenga contratado un término de potencia bajo?
- ¿Cuál es la probabilidad de que elija un cliente doméstico?
- ¿Cuál es la probabilidad de que elija un cliente que tenga contratado un término de potencia bajo?
- Si los auditores han elegido un cliente doméstico, ¿cuál es la probabilidad de que tenga contratado un término de potencia bajo?
- Si los auditories han elegido un cliente que tiene contratado un término de potencia bajo, ¿cuál es la probabilidad de que sea un cliente doméstico?

**Las respuestas a estas preguntas nos conducen a las distintas distribuciones de probabilidad que estudiaremos en este tema**

En el proceso de revisión de cuentas de la compañía eléctrica, cuando la empresa de auditoría elige un cliente al azar, consideraremos las siguientes variables aleatorias:

- $X = \text{"Categoría del cliente (1:doméstico, 2:servicios, 3:industria)"}$
- $Y = \text{"Término de potencia contratado (1:bajo, 2:medio, 3:alto, 4:muy alto)"}$

Si el cliente se elige al azar, la **probabilidad** de que pertenezca a una categoría  $X$  concreta y tenga contratado un término de potencia  $Y$  determinado coincide con la **proporción** de clientes de la compañía eléctrica que cumplen esas condiciones.

Por tanto, expresando la tabla inicial en forma de proporciones obtenemos las probabilidades con que podría ser elegido un cliente según categoría y término de potencia contratada:

	Y (Término de potencia)			
X (Categoría)	1 (bajo)	2 (medio)	3 (alto)	4 (muy alto)
1 (uso doméstico)	0.38	0.12	0.00	0.00
2 (sector servicios)	0.06	0.21	0.03	0.00
3 (sector industrial)	0.00	0.04	0.15	0.01

Entonces:

- ¿Cuál es la probabilidad de que sea elegido un cliente doméstico que tenga contratado un término de potencia bajo?

$$P(X = 1, Y = 1) = 0.38$$

Aquí hemos calculado la **probabilidad conjunta** de  $X$  e  $Y$

La suma de filas y columnas nos da las proporciones de clientes en cada categoría y según término de potencia contratada:

	Y (Término de potencia)				
X (Categoría)	1 (bajo)	2 (medio)	3 (alto)	4 (muy alto)	Total
1 (uso doméstico)	0.38	0.12	0	0	0.5
2 (sector servicios)	0.06	0.21	0.03	0	0.3
3 (sector industrial)	0	0.04	0.15	0.01	0.2
<b>Total</b>	<b>0.44</b>	<b>0.37</b>	<b>0.18</b>	<b>0.01</b>	<b>1</b>

- ¿Cuál es entonces la probabilidad de que sea elegido un cliente doméstico?

$$P(X = 1) = 0.5$$

- ¿Cuál es la probabilidad de que se elija un cliente que tenga contratado un término de potencia bajo?

$$P(Y = 1) = 0.44$$

Estas probabilidades se denominan **probabilidades marginales** de X e Y respectivamente

Dividiendo la proporción en cada casilla por el total de su fila, obtenemos las proporciones por fila, que indican **como es la distribución de probabilidad del término de potencia contratado para cada categoría de cliente:**

	Y (Término de potencia)				
X (Categoría)	1 (bajo)	2 (medio)	3 (alto)	4 (muy alto)	Total
1 (uso doméstico)	0.76	0.24	0	0	1
2 (sector servicios)	0.2	0.7	0.1	0	1
3 (sector industrial)	0	0.2	0.75	0.05	1
<b>Total</b>	0.44	0.37	0.18	0.01	<b>1</b>

Entonces:

- Si se ha elegido un cliente doméstico, ¿cuál es la probabilidad de que tenga contratado un término de potencia bajo?

$$P(Y = 1 | X = 1) = 0.76$$

Hemos calculado aquí la *probabilidad de Y condicionada por X*

Dividiendo la proporción en cada casilla por el total de su columna, obtenemos las proporciones por columna, que indican **como es la distribución de probabilidad de las categorías de clientes para cada término de potencia contratada**:

	Y (Término de potencia)				
X (Categoría)	1 (bajo)	2 (medio)	3 (alto)	4 (muy alto)	Total
1 (uso doméstico)	0.86	0.32	0	0	0.5
2 (sector servicios)	0.14	0.57	0.17	0	0.3
3 (sector industrial)	0	0.11	0.83	1	0.2
<b>Total</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>

Entonces:

- Si se ha elegido un cliente que tiene contratado un término de potencia bajo, ¿cuál es la probabilidad de que sea un cliente doméstico?

$$P(X = 1 | Y = 1) = 0.864$$

Hemos calculado aquí la *probabilidad de X condicionada por Y*

## 2. Distribución conjunta de dos Variables Aleatorias Discretas



# Probabilidad conjunta de variables aleatorias discretas.

- Si  $X$  e  $Y$  son dos variables aleatorias **discretas**, la función:

$$f(s, t) = P(X = s, Y = t)$$

se denomina **función de probabilidad conjunta** del vector  $(X, Y)$

## Ejemplo:

$(X, Y) = (\text{Categoría del cliente, Término de potencia contratado})$

	Y			
X	1	2	3	4
1	0.38	0.12	0.00	0.00
2	0.06	0.21	0.03	0.00
3	0.00	0.04	0.15	0.01



# Distribuciones de probabilidad marginales

Para un vector  $(X, Y)$  de variables aleatorias discretas, se define la **distribución de probabilidad marginal** de  $X$  como la distribución de probabilidad **unidimensional** de dicha variable. Se obtiene como:

$$\Pr(X = s) = \sum_t \Pr(X = s, Y = t)$$

De la misma forma, la distribución marginal de  $Y$  se obtiene mediante:

$$\Pr(Y = t) = \sum_s \Pr(X = s, Y = t)$$

En nuestro ejemplo de la compañía eléctrica hemos obtenido las distribuciones marginales sumando las probabilidades, respectivamente, por filas (distribución marginal de  $X$ ) o por columnas (distribución marginal de  $Y$ ).

# Distribuciones de probabilidad condicionales

Para un vector  $(X, Y)$  de variables aleatorias discretas, la **distribución de  $Y$  condicionada por el valor de  $X$**  se calcula como:

$$\Pr(Y = t \mid X = s) = \frac{\Pr(X = s, Y = t)}{\Pr(X = s)}$$

(recuérdese la definición de probabilidad de un suceso  $A$  condicionado por  $B$ ,  $P(A|B) = P(A \cap B) / P(B)$ )

Análogamente, la distribución de  $X$  condicionada por el valor de  $Y$  se calcula como:

$$\Pr(X = s \mid Y = t) = \frac{\Pr(X = s, Y = t)}{\Pr(Y = t)}$$

# Función de Distribución acumulativa conjunta de dos variables aleatorias.

Dadas dos variables aleatorias  $X$  e  $Y$  definidas sobre un mismo espacio de probabilidad  $(\Omega, \mathcal{F}, P)$ , se define la **función de distribución conjunta** del vector  $(X, Y)$  como:

$$F(s, t) = \Pr(X \leq s, Y \leq t) = \Pr(\{X \leq s\} \cap \{Y \leq t\})$$

Esta definición es válida tanto para variables discretas como continuas.

En el caso de que  $X$  e  $Y$  sean discretas:

$$F(s, t) = \sum_{x \leq s} \sum_{y \leq t} \Pr(X = x, Y = y)$$