Fundamentos de Probabilidad y Análisis Exploratorio de Datos

Francisco Javier Mercader Martínez

Índice

Tema 1: Estadística Descriptiva Univariante

1.1) Introducción a la Estadística

- Estudia cómo obtener conclusiones de la investigación empírica mediante el uso de modelos matemáticos.
- Disciplina puente entre los modelos matemáticos y los fenómenos reales.
  - Modelo matemáticos: abstracción simplificada de la realidad.
- Metodología para evaluar y juzgar las discrepancias entre Realidad y Teoría.
- Aplicaciones: Economía, Sociología, Ecología, Tecnología, ...

Partes de la Estadística

- **Estadística Descriptiva:**
  - **Objetivo:** Resumir la información contenida en los datos. Para ellos utiliza *métodos gráficos* y *medidas numéricas resumen*.
  - Si no es posible estudiar todos los elementos $\Longrightarrow$ seleccionamos una **Muestra**.
- **Inferencia Estadística:**
  - **Objetivo:** Inferir (generalizar) la información de la muestra a todos los elementos.
  - Contrastar una hipótesis.
  - Determinar y medir relaciones entre variables.
- **Teoría de la Probabilidad:**
  - **Objetivo:** Estudiar fenómenos aleatorios, donde interviene incertidumbre.
  - Construye modelos matemáticos que sirve de puente entre la Teoría y la Realidad.

1.2) Conceptos básicos

- **Población:** Conjunto de individuos objeto de estudio. Teóricamente puede ser ”finita” o ”infinita” (tamaño muy grande).

- **Muestra:** Subconjunto de individuos seleccionados de la población y que serán analizados en el experimento.
- **Tamaño muestral:** Número de individuos seleccionados de la población y que serán analizados en el experimento.
- **Muestreo:** Técnica o proceso empleado para obtener una muestra.
- **Experimento:** Proceso que permite asociar a cada individuo de una población un símbolo (o un grupo de símbolos), numérico o no, de entre un conjunto de símbolos dado a priori.
  Tipos de experimentos
  - Aleatorios.
  - Deterministas.
  - Pseudoaleatorios.
- **Variable estadística:** Cualquier característica que podemos observar un individuo. Ejemplos ¡: "altura", "edad", "nivel de calidad de un producto", "tiempo de vida de una marca de baterías", etc.

Tipos de variables:

- **Variables cualitativas:** El resultado de la observación no es un valor numérico sino un atributo, una cualidad. Ejemplo: "color de pelo", "intención de voto", ...
- **Variables cuantitativas:** Toman valores numéricos. Dos tipos:
  - **Cuantitativas discretas:** Toman un número finito de valores diferentes (o teóricamente infinito numerable). Ejemplo: "el número de hijos de las familias en España".
  - **Cuantitativas continuas:** Toman valores dentro de un intervalo (corresponden a medidas de magnitudes continuas). Ejemplo: "peso", "altura", ...
- En adelante, denotaremos por $\{x_1, x_2, \ldots, x_m\}$ a la muestra o conjunto de datos en estudio.

1.3) Las tablas de frecuencias

- Los datos se agrupan en clases (que denotaremos por A_i), que deben cubrir todo el rango de datos y ser disjuntas dos a dos.
  - Variables cualitativas: cada clase coincide con un atributo (o varios).

- Variables cuantitativas discretas: cada clase coincide con un valor (o varios) de los que toma la variable.
- Variables cuantitativas continuas: cada clase coincide con un intervalo de valores (intervalo de la recta Real).

- Para cada clase, se debe indicar su frecuencia.
  - **Frecuencia absoluta de una clase A_i:** número de veces que se observa dicha clase y se representa por f_i.
  - **Frecuencia relativa de una clase A_i:** cociente entre la frecuencia absoluta de la clase y el tamaño de la muestra, $h_i = \dfrac{f_i}{n}$.
  - **Frecuencia absoluta acumulada de una clase A_i:** es la suma de las frecuencias absolutas de las clases A_j, con $j = 1, 2, \dots, i$. Se representa por $F_i = f_1 + f_2 + \dots + f_i$.
  - **Frecuencia relativa acumulada de una clase A_i:** Se representa por $H_i = h_1 + h_2 + \dots + h_i$ (suma de frecuencias relativas).

¿Cómo determinamos las clases a realizar?

- Cuando la variable es continua (o bien discreta con muchos valores distintos), las clases se corresponden con intervalos de la recta Real. Pero, **¿cuántas clases hacemos y cómo las determinamos?**
- Existen varias reglas para determinar el número de clases (k) a realizar, a partir del tamaño muestral (n):
  - **Regla de la raíz cuadrada:** tomar $k \simeq \sqrt{n}$.
  - **Regla de Sturges:** tomar $k \simeq 1 + \log_2(n) = 1 + \dfrac{\log(n)}{\log(2)}$.
    En ambos casos podemos, por ejemplo, redondear al entero más próximo.
- Determinado k, calculamos el recorrido $R = x_{max} - x_{min}$ y la amplitud de cada clase $d = \dfrac{R}{k}$. Los intervalos que definen cada clase serán

$$[x_{min}, x_{min} + d], (x_{min} + d, x_{min} + 2d], \dots, (x_{max} - d, x_{max}]$$

- Llamaremos **marca de clase** (m_i) al centro del intervalo.

EJEMPLO: Tabla de frecuencias de los pesos de 500 alumnos de una universidad

Clases	f_1	F_i	h_i	H_i
[45,50]	8	8	0.016	0.016
(50, 55]	50	58	0.1	0.116
(55, 60]	82	140	0.164	0.280
(60, 65]	205	345	0.41	0.690
(65, 70]	79	424	0.158	0.848
(70, 75]	56	480	0.112	0.960
(75, 80]	20	500	0.04	1

1.4) Medidas numéricas resumen

- Permiten dar un resumen cuantitativo del fenómeno que estamos estudiando.
- Ponen de manifiesto los rasgos o características principales de la población en estudio.
- Son informativas cuando los datos son homogéneos y pueden ser muy engañosas cuando mezclamos poblaciones distintas.
- Podemos distinguir varios **tipos:**
  - *Medidas de centralización:* informan acerca de la tendencia central de los datos.
  - *Medidas de posición:* informan de la posición cuando consideramos los datos de la muestra ordenados de menor a mayor.
  - *Medidas de dispersión:* informan de lo concentrados o dispersos que están los datos de la muestra.
  - *Medidas de forma:* informan acerca de la forma en la que se distribuyen los datos de la muestra.

1.4.1) Medidas de centralización

- **Media (aritmética):** Representa el centro de gravedad de los datos.

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}; \overline{x} = \frac{\sum_{i=1}^{k} x_i \cdot f_i}{n} = \sum_{i=1}^{k} x_i \cdot h_i \text{ donde } k \text{ es el n}^{\text{o}} \text{ de valores distintos}$$

- **Mediana:** Divide a la muestra ordenada en dos partes, dejando a su izquierda el 50% de las observaciones y a su derecha el otro 50%.

$$Me = \begin{cases} x_{(\frac{n+1}{2})} & \textit{si } n \textit{ es impar} \\ \dfrac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} & \textit{si } n \textit{ es par} \end{cases}$$

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)} \text{ muestra ordenada de menor a mayor.}$$

- **Moda:** Es el valor de la muestra con mayor frecuencia (Mo).

- *Propiedad:* La media es muy sensible a valores extremos, mientras que mediana y moda no.

¿Y si tenemos datos guardados agrupados en clases?

- **Media:** Se toman como observaciones las marcas de clase.

$$\overline{x} = \frac{\sum_{i=1}^{k} m_i \cdot f_i}{n} = \sum_{i=1}^{k} m_i \cdot h_i, \text{ donde} \qquad \begin{aligned} k &= \text{ nº de clases} \\ m_i &= \text{ marca de clase} \end{aligned}$$

- **Mediana:** Una opción es identificar el intervalo donde se encontraría la mediana y proporcionar su marca de clase. Existen expresiones alternativas.

- **Moda:** Consideraremos la moda como la marca de la clase con mayor frecuencia. La moda no tiene por qué ser única, la distribución puede ser *unimodal* y *multimodal*.

1.4.2) Medidas de posición

- **Cuartiles:** Dividen a la muestra ordenada en 4 partes "iguales". Hay 3 cuartiles (Q_1, Q_2 y Q_3), siendo $Me = Q_2$.

- **Deciles:** Dividen a la muestra ordenada en 10 partes "iguales". Hay 9 deciles ($D_1, \ldots, D_9$) siendo $Me = D_5$.

- **Percentiles:** Dividen a la muestra ordenada en 100 partes "iguales". Hay 99 percentiles ($P_1, \ldots, P_99$), siendo $Me = P_50$.

- **Cuantiles:** Fijado $\alpha \in [0, 1]$, se llama *cuantil alpha* al valor C_α que deja a su izquierda el $(100 \cdot \alpha)$% de las observaciones y a su derecha el $(100 \cdot (1 - \alpha))$% restante.

  - Para $\alpha = 0.25, 0.5, 0.75 \rightarrow$ cuartiles.

- Para $\alpha = 0.10, 0.20, \ldots, 0.90 \rightarrow$ deciles.
- Para $\alpha = 0.01, 0.02, \ldots, 0.99 \rightarrow$ percentiles.

- **Rango o recorrido:** $R = x_{max} - x_{min}$ (poco representativa).
- **Varianza:** $s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} = \overline{x^2} - (\overline{x})^2$ (unidades al cuadrado).
- **Cuasivarianza:** $(s^*)^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$ (unidades al cuadrado).
- **Desviación típica:** $s = \sqrt{s^2}$ (mismas unidades que los datos).
- **Cuasidesviación típica:** $s^* = \sqrt{(s^*)^2}$ (mismas unidades).
- **Rango intercuartílico (RIC o IQR):** $RIC = Q_3 - Q_1$ (es el recorrido del 50% datos centrales).
- **Coeficiente de variación:** $CV = \frac{s}{\overline{x}}$ (adimensional). Software: $CV = \frac{s^*}{\overline{x}}$.
- **MEDA:** Mediana de las desviaciones absolutas de los datos respecto a la mediana.
- *Propiedad:* RIC y MEDA no son sensibles a valores extremos, el resto sí.

1.4.3) Medidas de forma

- **Coeficiente de asimetría:** $CA = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^3}{n \cdot s^3}$
Si CA>>0 (asimetría cola derecha), si CA<<0 (asimetría cola izquierda). Si hay simetría CA$\simeq$0.
- Existen varios coeficientes de asimetría (Pearson, Bowley, etc...).
- **Coeficientede apuntamiento o curtosis:** $CA_p = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^4}{n \cdot s^4}$, si $CA_p < 3$ (mesocúrtica).

1.4.4) Momentos muestrales

- **Momentos respecto al origen:**

$$a_r = \frac{\sum_{i=1}^{k} x_i^r}{n} = \frac{\sum_{i=1}^{k} x_i^r \cdot f_i}{n}$$

- **Momentos respecto a la media:**

$$m_r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^r}{n} = \frac{\sum_{i=1}^{k}(x_i - \overline{x})^r \cdot f_i}{n}$$

- Relación entre momentos:

$$m_r = a_r - \binom{r}{1} a_{r-1} \cdot \overline{x} + \binom{r}{2} a_{r-2} \cdot \overline{x}^2 - \ldots \pm \overline{x}^r$$

$$a_r = m_r - \binom{r}{1} m_{r-1} \cdot \overline{x} + \binom{r}{2} m_{r-2} \cdot \overline{x}^2 + \ldots + \overline{x}^r$$

1.5) Representaciones gráficas

- Resumen visual del conjunto de datos.

- Diferentes gráficos según el tipo de variable en estudio.

- **Variables cualitativas:** diagrama de barras (Pareto) y diagrama de sectores (tarta).

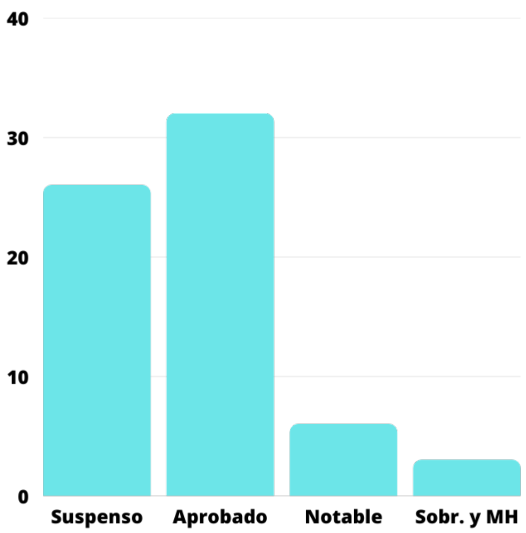


Diagrama de barras

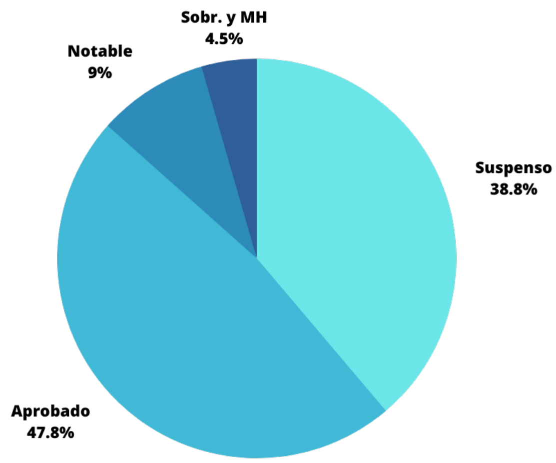


Diagrama de sectores(tarta)

- **Variables discretas:** diagrama de barras y barras acumulativo.

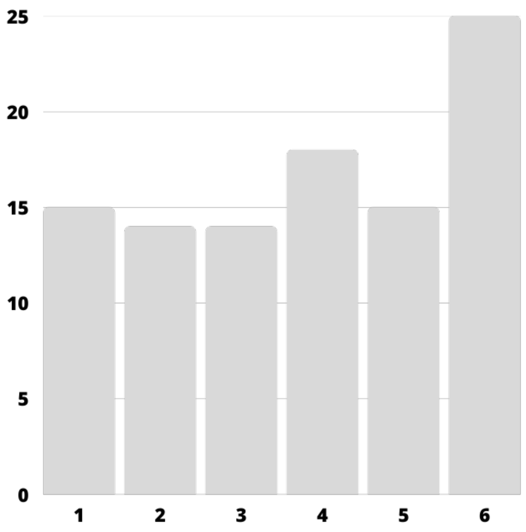


Diagrama de barras

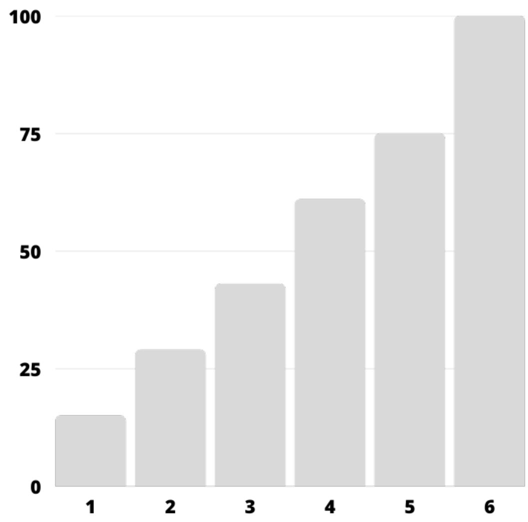
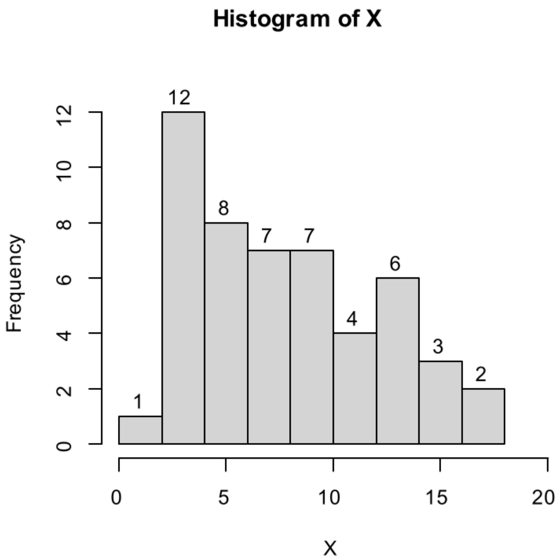


Diagrama de barras acumulativo

- **Variables continuas:** histograma y diagrama de caja-bigotes.



Histograma

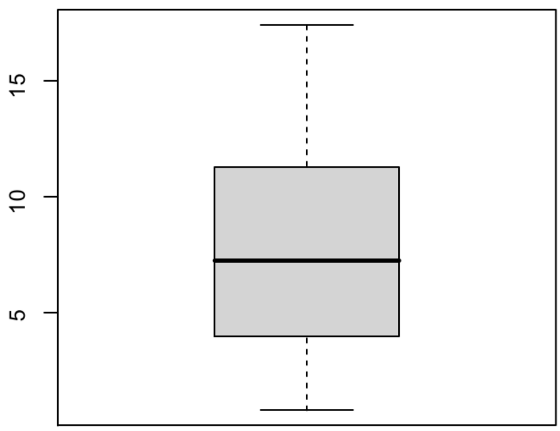


Diagrama de caja-bigotes

Histograma

Es la representación gráfica de la tabla de frecuencias para datos agrupados en intervalos. Permite responder a cuestiones del tipo:

- ¿Una o varias poblaciones? (Unimodal o multimodal)
- ¿Tendencia central de los datos?
- ¿Mucha o poca dispersión?
- ¿Forma simétrica o asimétrica? ¿Apuntamiento?

Diagrama de Caja-Bigotes

Permite responder a cuestiones del tipo:

- ¿Tendencia central de los datos?
- ¿Mucha o poca dispersión?
- ¿Forma simétrica o asimétrica?
- ¿Existen datos atípicos?

Ejemplo 1: Color de pelo

Clase	f_i
Moreno	15
Castaño	6
Rubio	3
Rojo	1
	25

Cualitativa

Ejemplo 2: Número de hermanos

Clase	f_i	F_i	h_i	H_i
0	9	9	$\frac{9}{25} = 36\%$	$\frac{9}{25} = 0.36$
1	8	17	$\frac{8}{25} = 32\%$	$\frac{27}{25} = 0.68$
2	5	22	$\frac{5}{25} = 20\%$	$\frac{22}{25} = 0.88$
≥ 3	3	25	$\frac{3}{25} = 12\%$	$\frac{25}{25} = 1$
	25	73	100%	

Discreta

Ejemplo 3: Peso alumnos

Clases	f_i	F_i	h_i	H_i
$[45, 50]$	8	8	0.016	0.016
$(59, 55]$	50	58	0.1	0.116
$(55, 60]$	82	140	0.164	0.280
$(60, 65]$	205	345	0.41	0.690
$(65, 70]$	79	424	0.158	0.848
$(70, 75]$	56	480	0.112	0.960
$(75, 80]$	20	500	0.04	1

Continua

PROBLEMAS. RELACIÓN 1: Estadística Descriptiva Univariante.

FUNDAMENTOS DE PROBABILIDAD Y ANÁLISIS EXPLORATORIO DE DATOS.

GRADO EN CIENCIA E INGENIERÍA DE DATOS.

1. Los siguientes datos corresponden al nivel de glucosa en sangre de diez niños;

56	62	63	65	65	65	65	68	70	72

   a) Calcula la media, la mediana y la moda. ¿Son parecidas? ¿Por qué?
   b) Dibuja el diagrama de barras y el gráfico acumulativo de frecuencias.
   c) Calcula el rango, la varianza y la desviación típica.
   d) Calcula el coeficiente de variación, el coeficiente de asimetría y el coeficiente de apuntamiento. Interpreta sus resultados.

2. Con el fin de analizar los sueldos de los habitantes de una comarca, se entrevistó a 20 trabajadores en activo que proporcionaron los siguientes valores de sus sueldos brutos anuales (en miles de euros):

7	9	7	158	11	37	37	31	89	60
98	26	69	41	16	22	1	7	29	125

   a) Para los datos SIN agrupar calcula la media aritmética, la mediana, los cuartiles, la desviación típica, la varianza, el coeficiente de variación y el rango intercuartílico. Compara los valores de la media y la mediana, ¿son parecidos o distintos?, ¿a qué se debe?
   b) Dibuja de forma detallada el diagrama de caja y bigotes del conjunto de datos. Comentar las características más relevantes: ¿los datos se distribuyen de forma simétrica o asimétrica?, ¿cuál es el intervalo de valores admisibles, es decir, dónde se encuentran los no atípicos?, ¿existen valores atípicos?
   c) Si queremos agrupar los datos en clases, ¿cuántas clases deberíamos realizar?, ¿de qué amplitud?
   d) Realizar la agrupación de los datos en 5 clases y determinar la tabla de frecuencias absolutas, relativas y sus acumuladas. ¿Qué porcentaje de individuos de la muestra cobran un salario menor o igual a 96000 euros brutos anuales?, ¿y entre 64000 y 128000?
   e) Calcular la media aritmética y la varianza para los datos agrupados en el apartado anterior. Determinar la clase que contiene a la mediana, la clase modal y proporciona sus marcas de clase. Compara la magnitud de las tres medidas de centralización (media, mediana y moda)
   f) Representar el histograma que resulta de los datos agrupados y comentar sus características más relevantes (unimodal o plurimodal, simétrico o asimétrico, mucha o poca dispersión, posibles atípicos).
   g) Atendiendo a la forma del diagrama de caja-bigotes y al histograma, ¿qué medidas de centralización y de dispersión te parecen más representativas para este conjunto de datos?

3. a) Sea $\{x_1, x_2, ..., x_{100}\}$ una muestra de media aritmética 21,5 y sean $x_{101} = 22, \quad x_{102} = 19, \quad x_{103} = 20{,}5$ tres observaciones más. Calcular la media aritmética de la nueva muestra $\{x_1, x_2, ..., x_{100}, x_{101}, x_{102}, x_{103}\}$.

b) Sea $\bar{x}$ la media aritmética de la muestra $\{x_1, x_2, ..., x_n\}$ y sea $\bar{y}$ la media aritmética de la muestra $\{y_1, y_2, ..., y_m\}$. ¿Cuál será la media aritmética de la unión de ambas muestras?

4. Probar que la media aritmética y la mediana son operadores lineales, es decir si tenemos una muestra $\{x_1, x_2, ..., x_n\}$ con media $\bar{x}$ y cada una de las observaciones x_i la multiplicamos por una constante *"a"* y le sumamos otra constante *"b"*, es decir obtenemos una nueva muestra $\{y_1, y_2, ..., y_n\}$ donde $y_i = ax_i + b \quad \forall i = 1, ..., n$, entonces la media de la nueva muestra $\bar{y}$, verifica $\bar{y} = a\bar{x} + b$ (igual con la mediana).

5. Sea $\{x_1, x_2, ..., x_n\}$ una muestra con media $\bar{x}$ y varianza s_X^2. Si cada una de las observaciones x_i la multiplicamos por una constante *"a"* (con $a \neq 0$) y le sumamos otra constante *"b"*, es decir obtenemos una nueva muestra $\{y_1, y_2, ..., y_n\}$ donde $y_i = ax_i + b \quad \forall i = 1, ..., n$:

a) ¿Qué relación existe entre la varianza de la Y $\left(s_Y^2\right)$ y la varianza de la X $\left(s_X^2\right)$?

b) ¿Qué relación existe entre la desviación típica de la Y $\left(s_Y\right)$ y la desviación típica de la X $\left(s_X\right)$?

1) Los siguientes datos corresponden al nivel de glucosa en sangre de diez niños:

56	62	63	65	65	65	65	68	70	72

a) Calcula la media, la mediana y la moda. ¿Son parecidas? ¿Por qué?

$\overline{x}, Me, Mo$?

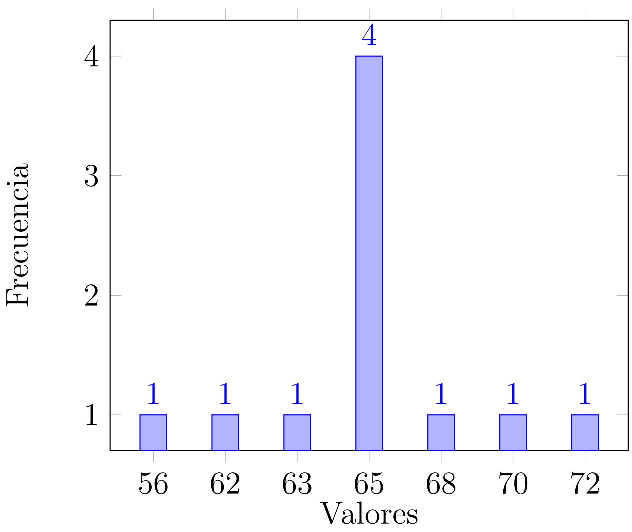
x_i	f_i	F_i	$x_i f_i$	$x_i^2 f_i$	$x_i^3 f_i$	$x_i^4 f_i$
56	1	1	56	3136	175616	9834496
62	1	2	62	3844	238328	14776336
63	1	3	63	3969	250047	15752961
65	4	7	260	16900	1098500	71402500
68	1	8	6	4624	314432	21381376
70	1	9	70	4900	343000	24010000
72	1	10	72	5184	373248	26873856
	10		654	42557	2793171	184031525

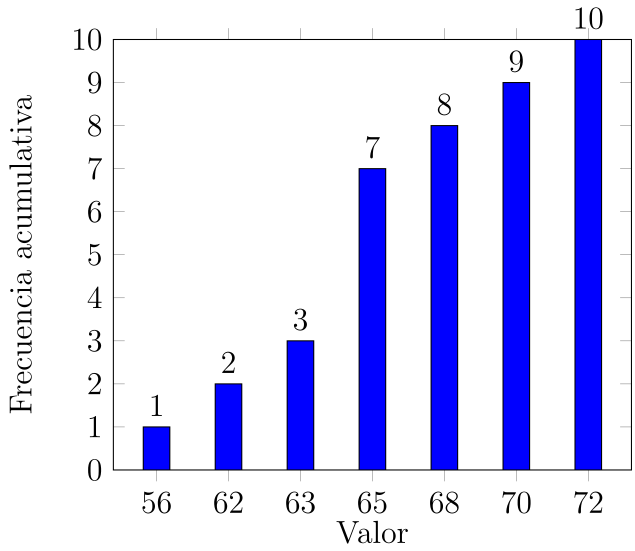
$$\overline{x} = \frac{1}{n}\sum x_i f_i = \frac{651}{10} = \boxed{65.1}$$

Mediana: $\frac{n}{2} = \frac{10}{2} = 5 \longrightarrow$ Miramos el volumen de frecuencias acumuladas (F_i), la más próxima por exceso $\longrightarrow \boxed{Me = 65}$

Moda: Moda es el valor con mayor frecuencia acumulada $\longrightarrow \boxed{\text{Moda} = 65}$

b) Dibuja el diagrama de barras y el gráfico acumulativo de frecuencias.





c) Calcula el rango, la varianza y la desviación típica.

Rango $= x_{\max} - x_{\min} = 72 - 56 = \boxed{16}$

Varianza:

$$s^2 = a_2 - \overline{x}^2 = \frac{1}{n}\sum x_i^2 f_i - \overline{x}^2 = \frac{42557}{10} - (65.1)^2 = 42557 - 4238.01 = \boxed{17.69}$$

Desviación típica:

$$s) + \sqrt{s^2} = +\sqrt{17.69} = \boxed{4.2059}$$

d) Calcula el coeficiente de variación, el cociente de asimetría y el cociente de apuntamiento. Interpreta sus resultados.

$CV = \dfrac{s}{\overline{x}} = \dfrac{4.2059}{65.1} = 0.0646 \longrightarrow \boxed{CV = 6.46\%}$

$CA = \dfrac{m_3}{s^3}$

$m_3 = a_3 - 3a_2\overline{x} + 2\overline{x}^3$

$a_3 = \dfrac{1}{n}\sum x_i^3 f_i = \dfrac{2793171}{10} = 279317.1$

$a_2 = \dfrac{1}{n}\sum x_i^2 f_i = 4255.7$

$m_3 = 279317.1 - 3 \cdot 4255.7 \cdot 65.1 - 2 \cdot (65.1)^3 = -1.103$

$CA = \dfrac{m_3}{s^3} = \dfrac{-1.103}{(4.2059)^3} = \boxed{-0.014} < 0 \longrightarrow$ Asimetría cola izquierda

$CAp = \dfrac{m_4}{s^4} = \dfrac{988.3217}{312.9361} = \boxed{3.1583 > 3} \longrightarrow$ Leptovértice

$m_4 = a_4 - 4a_3\overline{x} + 6a_2\overline{x}^2 - 3\overline{x}^4 = \dfrac{1}{n}\sum (x_i - \overline{x})^4 f_i = 988.3217$

2) Con el fin de analizar los sueldos de los habitantes de una comarca, se entrevistó a 20 trabajadores en activo que proporcionaron los siguientes valores de sus sueldos brutos anuales (en miles de euros):

7	9	7	158	11	37	37	31	89	60
98	26	69	41	16	22	1	7	29	125

a) Para los datos **sin** agrupar calcula la media aritmética, la mediana, los cuartiles, la desviación típica, la varianza, el coeficiente de variación y el rango intercuartílico. Compara los valores de la media y la mediana, ¿son parecido o distintos?, ¿a qué se debe?

$$\overline{x} = \frac{\sum x_i}{n} = \frac{7 + 9 + \cdots + 125}{20} = 44$$

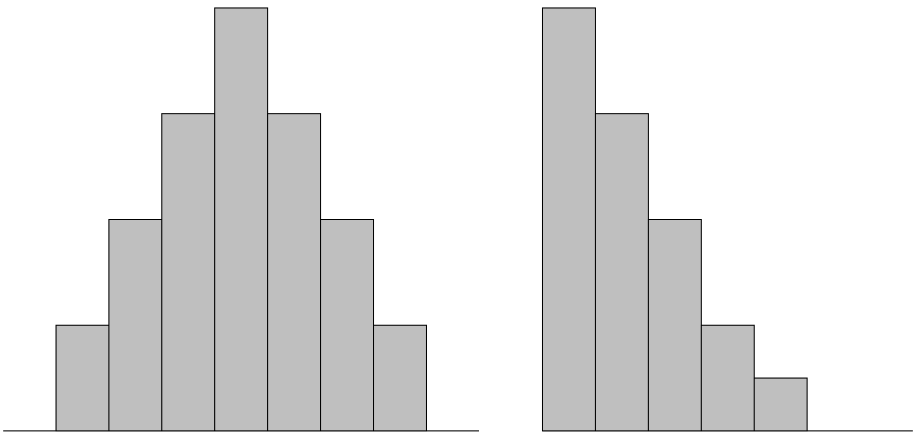
$$Me = \frac{\chi_{10} + \chi_{11}}{2} \text{ porque } n = 20 \text{ (par)}$$

Ordeno datos de menor a mayor:

$$1 \quad 7 \quad 7 \quad 7 \quad \underset{Q_1=\frac{9+11}{2}=10}{\underset{\downarrow}{\boxed{9 \quad 11}}} \quad 16 \quad 22 \quad 26 \quad \underset{Me=\frac{29+31}{2}=30}{\underset{\downarrow}{\boxed{29 \quad 31}}} \quad 373 \quad 37 \quad 41 \quad \underset{Q_3=\frac{60+69}{2}=64.5}{\underset{\downarrow}{\boxed{60 \quad 69}}} \quad 89 \quad 98 \quad 125 \quad 158$$

$Q_1 = 1^{\circ}$ Cuartil$\longrightarrow$ Lo que obtenemos como la mediana de los datos por debajo de la mediana (excluida esta).

- Histograma



Si la distribución es simétrica ¿$\overline{x}$, Me, Moda?

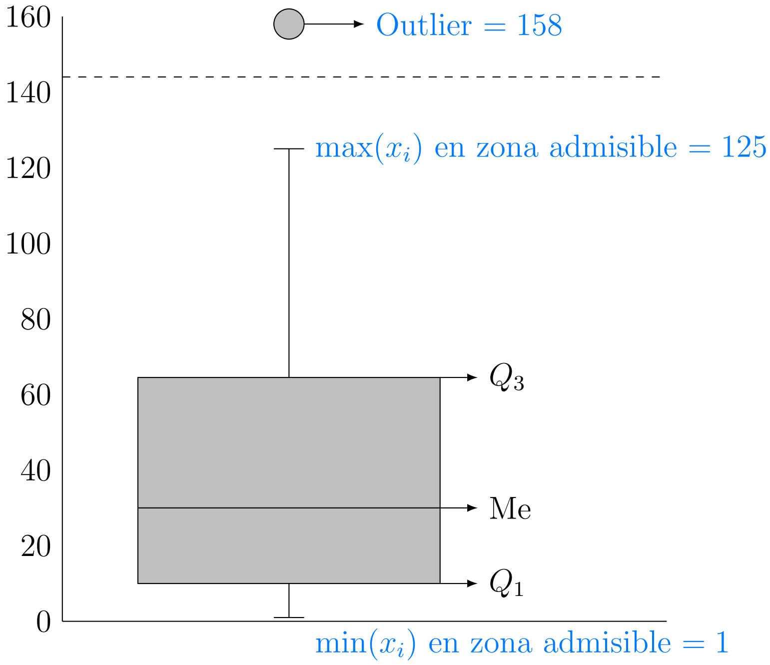
$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x}^2)}{n} = \overline{x^2} - (\overline{x})^2$$

$$\overline{x^2} = \frac{\sum x_i^2}{n} = \frac{7^2 + 9^2 + \cdots + 125^2}{20} = \frac{77491}{20}$$

$$s_x^2 = \frac{77491}{20} - (44)^2 = 1938.55 \longrightarrow s_x = \sqrt{s_x^2} = 44.03$$

b) Dibuja de forma detallada el diagrama de caja y bigotes del conjunto de datos. Comentar las características más relevantes: ¿los datos se distribuyen de forma simétrica o asimétrica?, ¿cuál es el intervalo de valores admisibles, es decir, dónde se encuentran los no atípicos?, ¿existen valores atípicos?

- $LSA = Q_3 + 1.5 \cdot RIC = 64.5 + 1.5 \cdot 54.5 = 145.25$
- $LIA = Q_1 + 1.5 \cdot RIC = 10 - 1.5 \cdot 54.5 = -71.75$



c) Si queremos agrupar los datos en clases, ¿cuántas clase deberíamos realizar?, ¿de qué amplitud?

$$d = \frac{\max(x_i) - \min(x_i)}{k} = \frac{158 - 1}{5} = 31.4$$

d) Realizar la agrupación de los datos en 5 clases y determinar la tabla de frecuencias absolutas, relativas y sus acumuladas. ¿Qué porcentaje de individuos de la muestra cobran un salario menor o igal a 96000 euros brutos anuales?, ¿y entre 64000 y 128000?

Clases	f_i	F_i	h_i	H_i
[0,32]	11	11	$\frac{11}{20}$	$\frac{11}{20}$
(32,64]	4	15	$\frac{4}{20}$	$\frac{15}{20}$
(64,96]	2	17	$\frac{2}{20}$	$\frac{17}{20}$
(96,128]	2	19	$\frac{2}{20}$	$\frac{19}{20}$
(128,160]	1	20	$\frac{1}{20}$	1
	20			

$\frac{17}{20} = 0.85 \longrightarrow 85\%$ cobran ≤ 96000€

$100 \cdot \left(\frac{19}{20} - \frac{15}{20}\right) = 20\%$ cobran entre 64000 y 128000

e) Calcular la media aritmética y la varianza para los datos agrupados en el apartado anterior. Determinar la clase que contiene a la mediana, la clase modal y proporciona sus marcas de clase. Compara la magnitud de las tres medidas de centralización (media, mediana y moda)

$$\overline{x} = \frac{16 \cdot 11 + 48 \cdot 4 + 80 \cdot 2 + 112 \cdot 2 + 144 \cdot 1}{20} = 44.8$$

$$Me = \frac{\chi_{10} + \chi_{11}}{2} \in [0, 32]$$

3) a) Sea $\{x_1, x_2, \ldots, x_{100}\}$ una muestra de media aritmética 21.5 y sean $x_{101} = 22, x_{102} = 19, x_{103} = 20.5$ tres observaciones más. Calcular la media aritmética de la nueva muestra $\{x_1, x_2, \ldots, x_{100}, x_{101}, x_{102}, x_{103}\}$

$\{x_1, x_2, \ldots, x_{100}\}$

$\{x_1, x_2, \ldots, x_{100}, 22, 19, 20.5\} \longrightarrow \overline{x} =¿?$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{103} x_i f_i = \frac{1}{103}\left(\sum_{i=1}^{100} x_i f_i + \sum_{i=1}^{103} x_i f_i\right) = \frac{1}{103}(100 \cdot 21.5 + 22 + 19 + 20.5) = 21.4709$$

b) Sea $\overline{x}$ la media aritmética de la muestra $\{x_1, x_2, \ldots, x_n\}$ y sea $\overline{y}$ la media aritmética de la muestra $\{y_1, y_2, \ldots, y_m\}$. ¿Cuál será la media aritmética de la unión de ambas muestras?

$$\left.\begin{array}{l}\{x_1, x_2, \ldots, x_n\} \longrightarrow \overline{x} \\ \{y_1, y_2, \ldots, y_m\} \longrightarrow \overline{y}\end{array}\right\} \{x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m\} \longrightarrow ¿\overline{z}?$$

$$z = \frac{1}{n+m}\sum_{i=1}^{n+m} z_i f_i = \frac{1}{n+m}\left[\sum_{i=1}^{n} x_i f_i + \sum_{i=1}^{m} y_i f_i\right] = \frac{1}{n+m}[n \cdot \overline{x} + m \cdot \overline{y}]$$

$$\boxed{z = \frac{m \cdot \overline{x} + m \cdot \overline{y}}{n+m}}$$

4) Probar que la media aritmética y la mediana son operadores lineales, es decir, si tenemos una muestra $\{x_1, x_2, \ldots, x_n\}$ con media $\overline{x}$ y cada una de las observaciones x_i la multiplicamos por una constante "a" y le sumamos otra constante "b", es decir, obtenemos una nueva muestra $\{y_1, y_2, \ldots, y_n\}$ donde $y_I = ax_i + b$ $\forall i = 1, \ldots, n$, entonces la media de la nueva muestra $\overline{y}$, verifica $\overline{y} = a\overline{x} + b$ (igual con la mediana).

$$\underset{\substack{\downarrow \\ \text{media} = \overline{x}}}{\{x_1, x_2, \ldots, x_n\}} \longrightarrow \underbrace{\{\underbrace{ax_1 + b}_{y_1}, \underbrace{ax_2 + b}_{y_2}, \ldots, \underbrace{ax_n + b}_{y_n}\}}_{\text{¿media } (y) = \overline{y} = a\overline{x} + b?}$$

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n}(ax_i + b)}{n} = \frac{\sum_{i=1}^{n} ax_i + \sum_{i=1}^{n} b}{n} = \frac{a\sum x_i + b_n}{n} = a \cdot \underbrace{\frac{\sum x_i}{n}}_{\overline{x}} + b = a\overline{x} + b$$

¿Qué pasa con las medianas?

Ordenamos datos de $\{x_1, \ldots, x_n\}$ de menor a mayor.

$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)} \longrightarrow Me_{(x)} = x_{\left(\frac{n+1}{2}\right)}$ (supongo n impar)

- Si $a > 0 \longrightarrow y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)} \longrightarrow Me_{(y)} = y_{\left(\frac{n+1}{2}\right)} = \boxed{ax_{\left(\frac{n+1}{2}\right)} + b}$
- Si $a < 0 \longrightarrow$ Se invierte el orden pero la posición central la ocupa el mismo dato.

5) Sea $\{x_1, x_2, \ldots, x_n\}$ una muestra con media $\overline{x}$ y varianza s_X^2. Si cada una de las observaciones x_i la multiplicamos por una constante "a" (con $a \neq 0$) y le sumamos otra constante "b", es decir, obtenemos una nueva muestra $\{y_1, y_2, \ldots, y_n\}$ donde $y_i = ax_i + b$ $\forall i =, \ldots, n$

$\{x_1, \ldots, x_n\}$ aumenta con moda $\overline{x}$ y varianza s_X^2. Sea $\{y_1 = ax_1 + b, y_2 = ax_2 + b, \ldots, y_n = ax_n + b\}$

a) ¿Qué relación existe entre la varianza de la $Y(S_Y^2)$ y la varianza de la $X(S_X^2)$?

$$S_X^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}$$

$$S_Y^2 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n} = \frac{\sum_{i=1}^{n}\left((ax_i + b) - (a\overline{x} + b)\right)^2}{n} = \frac{\sum_{i=1}^{n}\left(a^2(x_i - \overline{x})\right)^2}{n} = a^2 \cdot \frac{\sum_{i=1}^{n}(x - i - \overline{x})^2}{n} = a^2 \cdot S_X^2$$

b) ¿Qué relación existe entre la desviación típica de la $Y(S_Y)$ y la desviación típica de la $X(S_X)$?

$$S_Y = \sqrt{a^2 \cdot S_X^2} = \boxed{|a| \cdot S_X}$$

Tema 2: Estadística Descriptiva Multivariante

2.1) Introducción

¿Cómo presumir si hay más de una variable?

- Realizar un estudio descriptivo de cada variable por separado (visto en Tema 1).
- Descriptiva Bivariante. Estudiar relaciones entre pares de variables. Usaremos métodos gráficos y medidas numéricas.
- Descriptiva Multivariante: Estudiar relaciones entre 3 o más variables.

2.2) Estadística Descriptiva Bivariante

- En adelante, denotaremos por $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ al conjunto de datos en estudio (muestra de datos bidimensionales), correspondientes a dos variables X e Y.

¿Qué herramientas contemplamos?

- **Tablas de doble entrada:** tablas de frecuencias con dos variables.
- **Medidas numéricas de asociación:** para medir la relación entre dos varaibles. Dependen del tipo de variables en estudio.
- **Métodos gráficos para datos bidimensionales:** para visualizar la relación entre dos variables. Dependen del tipo de variables en estudio.

2.2.1) Tablas de doble entrada

- Sean $A_1, A_2, \ldots, A_k$ las clases (partición) de la variable X y $B_1, B_2, \ldots, B_p$ las clases (partición) de la variable Y. La **tabla de doble entrada** viene dada por:

X \ Y	B_1	B_2	$\cdots$	B_j	$\cdots$	B_p
A_1	f_{11}	f_{12}	$\cdots$	f_{1j}	$\cdots$	f_{1p}
A_2	f_{21}	f_{22}	$\cdots$	f_{2j}	$\cdots$	f_{2p}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
A_i	f_{i1}	f_{i2}	$\cdots$	f_{ij}	$\cdots$	f_{ip}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
A_k	f_{k1}	f_{k2}	$\cdots$	f_{kj}	$\cdots$	f_{kp}

- $f_{ij} \longrightarrow$ **Frecuencia absoluta de la clase bidimensional** $A_i \times B_j$ (n$^{\underline{o}}$ de elementos de la muestra cuya primera componente pertenece a A_i y la segunda a B_j). **Frecuencia relativa:** $h_{ij} = \dfrac{f_{ij}}{n}$.

- **Frecuencias Marginales:** Hace referencia a las frecuencias (absolutas o relativas) de cada variable individual por separado. Se obtienen sumando por filas y por columnas las frecuencias(absolutas o relativas) de la tabla de doble entrada.

X \ Y	B_1	B_2	$\cdots$	B_j	$\cdots$	B_p	
A_1	f_{11}	f_{12}	$\cdots$	f_{1j}	$\cdots$	f_{1p}	$f_{1\bullet}$
A_2	f_{21}	f_{22}	$\cdots$	f_{2j}	$\cdots$	f_{2p}	$f_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	
A_i	f_{i1}	f_{i2}	$\cdots$	f_{ij}	$\cdots$	f_{ip}	$f_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	
A_k	f_{k1}	f_{k2}	$\cdots$	f_{kj}	$\cdots$	f_{kp}	$f_{k\bullet}$
	$f_{\bullet 1}$	$f_{\bullet 2}$	$\cdots$	$f_{\bullet j}$	$\cdots$	$f_{\bullet p}$	n

- También se puede construir la **tabla de doble entrada usando frecuencias relativas**, contemplando a su vez las frecuencias marginales relativas.

Y \ X	B_1	B_2	$\cdots$	B_j	$\cdots$	B_p	
A_1	h_{11}	h_{12}	$\cdots$	h_{1j}	$\cdots$	h_{1p}	$h_{1\bullet}$
A_2	h_{21}	h_{22}	$\cdots$	h_{2j}	$\cdots$	h_{2p}	$h_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	
A_i	h_{i1}	h_{i2}	$\cdots$	h_{ij}	$\cdots$	h_{ip}	$h_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	
A_k	h_{k1}	h_{k2}	$\cdots$	h_{kj}	$\cdots$	h_{kp}	$h_{k\bullet}$
	$h_{\bullet 1}$	$h_{\bullet 2}$	$\cdots$	$h_{\bullet j}$	$\cdots$	$h_{\bullet p}$	1

- Si X e Y son **independientes** (no relacionadas), entonces

$$h_{ij} = h_{i\bullet} \times h_{\bullet j} \longleftrightarrow f_{ij} = \frac{f_{i\bullet} \times f_{\bullet j}}{n} \forall i = 1, \ldots, k, \forall j = 1, \ldots, p.$$

- La tabla de doble entrada es una herramienta para analizar si existe relación entre dos vairables.

- **Idea y si tedel test Chi-cuadrado de independencia:** comparar si hay poca o mucha discrepancia entre las frecuencias observadas de las clases $A_i \times B_j$ y las frecuencias esperadas en el caso de independencia

$$\mathcal{X}_0^2 = \sum_{i=1}^{k} \sum_{i=1}^{k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \text{ donde } \begin{array}{l} O_{ij} = f_{ij} \rightarrow \text{frecuencia observada} \\ E_{ij} = \dfrac{f_{i\bullet} \times f_{\bullet j}}{n} \rightarrow \text{frecuencia esperada} \end{array}$$

$\mathcal{X}_0^2 \longrightarrow$ discrepancia entre observado y esperado bajo independencia.

2.2.2) Medidas numéricas de asociación

Existen diferentes medidas según los tipos de variables.

- Si X e Y son **variables cualitativas:** Coeficiente de contigencia, Phi y V de Cramer.

  Son medidas basadas en el estadístico Chi-cuadrado, que intentan "estandarizar" su valor entre 0 y 1, así como minimizar el efecto del tamaño de la muestra sobre la cuantificación del grado de asociación. Valores próximos a cero indican independencia y próximos a 1 indican fuerte asociación entre X e Y.

$$C = \sqrt{\frac{\mathcal{X}_0^2}{n + \mathcal{X}_0^2}} \qquad \phi = \sqrt{\frac{\mathcal{X}_0^2}{n}} \quad V_{Cramer} = \sqrt{\frac{\mathcal{X}_0^2}{n \cdot \min(k-1, p-1)}}$$

(Coeficiente de contigencia) (Phi) (V de Cramer)

- En realidad, el *coeficiente de contingencia* no alcanza el valor 1.
- En el caso de *Phi*, su uso se suele limitar a tablas 2×2 con el fin de dicho coeficiente no supere 1.
- La *V de Cramer* es una extensión del coeficiente de Phi, que se mueve entre 0 y 1, pero tiende a subestimar el grado de asociación.

- Si X e Y son **variables cualitativas ordinales:** Si las variables cualitativas permiten establecer un orden **(variables ordinales)**, se puede usar como medida de asociación el *coeficiente gamma de Goodman y Kruskall.*

  Esta medida está basada en el número de concordancias y discordancias entre los pares de datos que conforman la muestra. Si los dos valores de un caso en ambas variables son mayores (o menores) que los dos valores de otro caso, se da una concordancia.

  $$\gamma = \frac{Con - Dis}{Con + Dis} \text{ donde } \begin{array}{l} Con = \text{nº de concordancias} \\ \\ Dis = \text{nº de discordancias} \end{array}$$

  $\gamma = -1$(asociación negativa), $\gamma = 0$(independiente) y $\gamma = 1$ (asociación positiva)

- Existen otras medidas de asociación basadas en las concordancias como la *"d de Sommers"* y las *medidas Tau de Kendall* (Tau-a, Tau-b y Tau-c).

- Si X e Y son **variables cuantitativas:** Covarianza y correlación de Pearson. Son medidas del grado de asociación lineal entre X e Y.

  $$s_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n} = \overline{x \cdot y} - \overline{x} \cdot \overline{y} \quad s^*_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n-1} = \overline{x \cdot y} - \overline{x} \cdot \overline{y}$$

  Covarianza Cuasi-covarianza(software)

  $$R_{X,Y} = \frac{s_{X,Y}}{s_X \cdot s_Y} = \frac{s^*_{X,Y}}{s^*_X \cdot s^*_Y}$$

  Correlación de Pearson (adimensional)

- La covarianza de X con X es su varianza, $s_{X,X} = s_x^2$.

- Se cumple que $s_{X,Y}^2 \leq s_X^2 \cdot s_Y^2$, y por tanto, $R_{X,Y} \in [-1, 1]$.

$$R_{X,Y} \simeq -1 \quad \text{(fuerte relación lineal decreciente)}$$

$$R_{X,Y} \simeq 0 \qquad \text{(ausencia de relación lineal)}$$

$$R_{X,Y} \simeq 1 \qquad \text{(fuerte relación lineal creciente)}$$

- Si X e Y son **variables cuantitativas pasadas a ordinales:** Correlación de Spearman.

- El *coeficiente de correlación de Spearman* es un caso particular de la correlación de Pearson pero, en lugar de usar el valor numérico de los datos, se usa el valor de su posición al ordenarse de menor a mayor en cada variable (pasamos a datos ordinales).

  $$RS_{X,Y} = 1 - \frac{6\sum_{i=1}^{d} d_i^2}{n(n^2-1)} \quad \text{Correlación de Spearman}$$

  donde d_i es la diferencia entre el orden que ocupa el sujeto i-ésimo en la ordenación "X" y en la ordenación "Y".

- Mide el grado de asociación monótona (creciente o decreciente) entre las variables X e Y, pues una relación monótonaentre ellas se traduce en un relación lineal entre sus ordenaciones.

2.2.3) Momentos bivariantes

- **Momentos respecto al origen:**

$$a_{rs} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i^r \cdot y_j^s}{n} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{p} x_i^r \cdot y_j^s \cdot f_{ij}}{n}$$

- **Momentos respecto a la media:**

$$m_{rs} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} (x_i - \overline{x})^r (y_j - \overline{y})^s}{n} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{p} (x_i - \overline{x})^r (y_j - \overline{y})^s f_{ij}}{n}$$

- Relaciones

$$\begin{aligned} s_{XY} &= m_{11} = a_{11} - \overline{x} \cdot \overline{y} \\ s_X^2 &= m_{20} = a_{20} - \overline{x}^2 \\ s_Y^2 &= m_{02} = a_{02} - \overline{y}^2 \end{aligned}$$

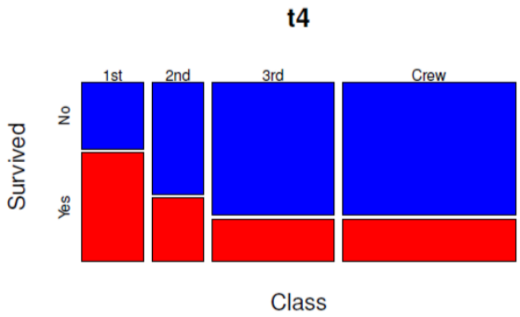
2.2.4) Gráficos para datos bidimensionales

Existen diferentes gráficos según los tipos de variables.

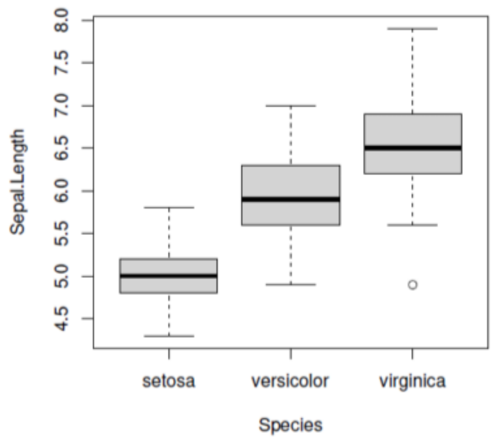
- Si X e Y son **variables cualitativas:** diagramas de áreas y de barras comparativo.

- Si X e Y es **cuantitativa y la otra cualitativa:** caja de bigotes por grupos.

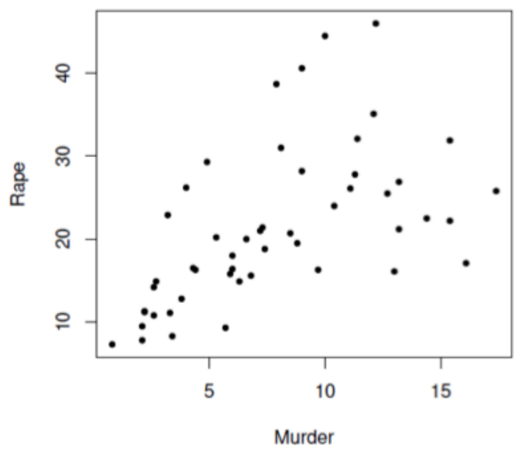
- Si X e Y son **cariables cuantitativas:** diagrama de dispersión o nube de puntos.



Áreas comparativo



Caja-bigotes por grupos



Nube puntos

2.3) Ajuste por mínimos cuadrados (datos bidimensionales)

Sea $\{(x_1, y1), (x_2, y2), \ldots, (x_n, y_n)\}$ un conjunto de datos correspondientes a dos varaibles cuantitativas X e Y.

- **La nube de puntos** informa sobre si existe o no asociación entre X e Y, y el tipo de relación (lineal, cuadrática, exponencial, etc.).
- **Las relaciones lineales** tienen especial importancia porque muchas relaciones no lineales se pueden linealizar mediante una transformación de los datos.
- **Regresión Lineal (RL):** Técnica inferencial para modelizar la relación lineal existente entre dos o más variables.
- **Ajuste por mínimos cuadrados:** Técnica descriptiva para modelizar la relación lineal existente entre dos o más variables.
- En RL se hace uso del ajuste por mínimos cuadrados.
- Nomenclatura: $\begin{array}{ll} Y \longrightarrow & \text{variable respuesta (explicada o dependiente),} \\ X \longrightarrow & \text{regresor (predictor o variable independiente).} \end{array}$
- **Objetivo:** Encontrar la ecuación de la recta $y = a + b \cdot x$ que minimiza la suma de los cuadradados de las distancias verticales entre los puntos de la muestra y dicha recta. Es decir, buscamos $(a, b) \in \mathbb{R}^2$ que minimiza $\displaystyle\sum_{i=1}^{n} [y_i - (a + b \cdot x_i)]^2$
- **La solución viene dada por:**

$$\hat{a} = \overline{y} - \frac{s_{X,Y}}{s_X^2 \cdot \overline{x}} \qquad \hat{b} = \frac{s_{X,Y}}{s_X^2 \cdot \overline{x}}$$

(ordenada en el origen) (pendiente)

- Recordar que la **corrección de Pearson** $(R_{X,Y})$ cuantifica el grado de asociación lineal (bondad del ajuste).
- ¿Y si queremos la recta ajustada de "X" en función de "Y"?

$$x - \overline{x} = \frac{s_{X,Y}}{s_Y^2}(y - \overline{y})$$

- **Valores ajustados y residuos:**

$$\hat{y} = \hat{a} + \hat{b} \cdot x_i \qquad e_i = y_i - \hat{y}_i$$

(valores ajsutados) (residuos)

- **Ejemplo:** nube de puntos, recta ajustada por mínimos cuadradados y gráfico de residuos.

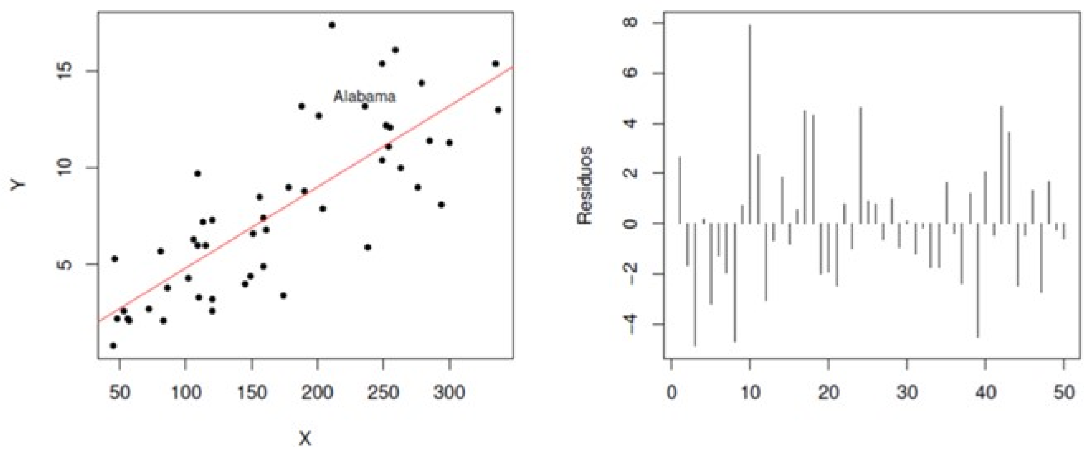


Gráfico de puntos (izquierda) para 'Assault' (eje x) y 'Murder' (eje y) y la recta de regresión (rojo). Errores (residuos) en las predicciones para cada estado (derecha).

2.4) Estadística Descriptiva Multivariante

¿Y si tenemos más de dos variables?

- Podemos realizar un análisis descriptivo bivariante entre cada par de variables. La información se suele poner en forma matricial: matriz de nube de puntos, matriz de covarianzas, matriz de correlaciones, etc.

- Tablas de triple entrada o de dimensión superior.

- Gráficos en 3D y gráficos por grupos usando varios factores.

- Ajuste por mínimos cuadrados para el caso de dos o más regresores (predictores).

PROBLEMAS. RELACIÓN 2: Estadística Descriptiva Bivariante.

FUNDAMENTOS DE PROBABILIDAD Y ANÁLISIS EXPLORATORIO DE DATOS.

GRADO EN CIENCIA E INGENIERÍA DE DATOS.

1.Las calificaciones obtenidas por un grupo de alumnos en Biología y Física son las siguientes:

Bio.	3	4	6	7	5	8	7	3	5	4	8	5	5	8	8	8	5
Fís.	5	5	8	7	7	9	10	4	7	4	10	5	7	9	10	5	7

(Cada pareja de datos corresponde a un alumno)

a) ¿Cuántos alumnos tiene el grupo?
b) Escribir la tabla de doble entrada de frecuencias absolutas.
c) Hallar las distribuciones de frecuencias marginales, así como la media y la varianza de dichas distribuciones.
d) Calcular la recta por mínimos cuadrados que nos ayuda a predecir la nota de Física conocida la de Biología.
e) ¿Cómo de buena es esta predicción?

2. De una variable estadística bidimensional se conoce la recta por mínimos cuadrados de Y sobre X: $y = (1/2)x + 2$, la recta por mínimos cuadrados de X sobre Y: $x = 2y - 4$ y la desviación típica de X: $s_x = 3$.

a) Estudiar si es posible determinar el punto $(\bar{x}, \bar{y})$.

b) Hallar el coeficiente de correlación lineal, así como la covarianza y la desviación típica de Y.

c) Si $\bar{x} = 2$ determinar $\bar{y}, a_{20}, a_{02}$ y a_{11}.

3. Conocemos la siguiente información sobre las medidas de una variable estadística bidimensional (X, Y):

$$(CV)_x = 2(CV)_y, \quad \bar{x} = 4, \quad \bar{y} = 7, \quad R = 1.$$

Determinar la recta por mínimos cuadrados de la variable Y sobre la variable X.

4. Consideremos los siguientes datos de la variable bidimensional (X, Y):

X	Y
6,5	20,3
11,5	14,6
20,1	11,4
25,7	7,2
34,2	6,3

a) Representar gráficamente la nube de puntos.
b) Realizar un ajuste exponencial del tipo $y = a.e^{bx}$ a estos datos y representar gráficamente el resultado.
c) Comprobar la bondad del ajuste realizado.

1) Las calificaciones obtenidas por un grupo de alumnos en Biología y Física son las siguientes:

Bio.	3	4	6	7	5	8	7	3	5	4	8	5	5	8	8	8	5
Fís.	5	5	8	7	7	9	10	4	7	4	10	5	7	9	10	5	7

(Cada pareja de datos corresponde a un alumno)

a) ¿Cuántos alumnos tiene el grupo?

En grupo tiene 17 alumnos

b) Escribir la tabla de doble entrada de frecuencias absolutas

X \ Y	4	5	7	8	9	10	$f_{i\bullet}$
3	1	1	0	0	0	0	2
4	1	1	0	0	0	0	2
5	0	1	4	0	0	0	5
6	0	0	0	1	0	0	1
7	0	0	1	0	0	1	2
8	0	1	0	0	2	2	5
$f_{\bullet j}$	2	4	5	1	2	3	

c) Hallar las distribuciones de frecuencias marginales, así como la media y la varianza de dichas distribuciones.

$$\overline{x} = \frac{3 \cdot 2 + 4 \cdot 2 + 5 \cdot 5 + 6 \cdot 1 + 7 \cdot 2 + 8 \cdot 5}{17} = 5.82$$

$$\overline{y} = \frac{4 \cdot 2 + 5 \cdot 4 + \cdots + 10 \cdot 3}{17} = 7$$

$$s_X^2 = \overline{x^2} - (\overline{x})^2 = \overline{x^2} - (5.82)^2 = 3.08 \longrightarrow s_X = \sqrt{3.08}$$

$$\overline{x^2} = \frac{3^2 \cdot 2 + 4^2 \cdot 2 + 5^2 \cdot 5 + 6^2 \cdot 1 + 7^2 \cdot 2 + 8^2 \cdot 5}{17} = 37$$

d) Calcular la recta por mínimos cuadrados que nos ayuda a predecir la nota de Física conocida de la Biología.

$$y-\overline{y}=\frac{s_{XY}}{s_X^2}(x-\overline{x}) \longrightarrow y-7=\frac{2.7095}{3.08}(x-5.82) \longleftrightarrow y=0.877x+1.888$$

$$s_{XY}=\overline{xy}-\overline{x}\cdot\overline{y}=43.47-5.82\cdot 7=2.7095$$

$$\overline{xy}=\frac{3\cdot 5+4\cdot 5+\cdots+5\cdot 7}{17}=43.47$$

$$\frac{s_{XY}}{s_X^2}=\frac{2.7095}{3.08}=0.877$$

e) ¿Cómo de buena es esta predicción?

Calculamos la correlación de Pearson.

$$R_{xy}=\frac{S_{xy}}{S_x\cdot S_y}=\frac{2.7095}{1.75\cdot 2.029}=0.76 \text{ (Ajuste no muy bueno pero aceptable)}$$

2) De una variable estadística bidimensional se conoce la recta por mínimos cuadrados de Y sobre $X: y=\frac{1}{2}x+2$, la recta por mínimos cuadrados de X sobre $Y: x=2y-4$ y la desviación típica de $X: s_X=3$.

a) Estudiar si es posible determinar el punto $(\overline{x},\overline{y})$.

¿$\overline{x},\overline{y}$? En general, este punto $(\overline{x},\overline{y})$ se obtiene resolviendo el sistema de ecuaciones dado por las dos rectas. En este caso, ese sistema es compatible indeterminado porque ambas rectas coinciden $\longrightarrow$ no pueda calcular $(\overline{x},\overline{y})$.

b) Hallar el coeficiente de correlación lineal, así como la varianza y la desviación típica de Y.

$$y=\frac{1}{x}+2,\ S_x=3\longrightarrow S_x^2=9$$

$$y-\overline{y}=\frac{S_{xy}}{S_x^2}(x-\overline{x})$$

- $\frac{S_{xy}}{S_x^2}=\frac{1}{2}\longleftrightarrow S_{xy}=\frac{1}{2}\cdot S_x^2=\frac{1}{2}\cdot 9=\frac{9}{2}$

  Recta de x sobre $y: x=2y+4$

$$x-\overline{x}=\frac{S_{xy}}{S_y^2}(y-\overline{y})$$

- $\frac{S_{xy}}{S_y^2}=2\longrightarrow\frac{\frac{9}{2}}{S_y^2}=2\longrightarrow S_y^2=\frac{9}{4}\longrightarrow S_y=\frac{3}{2}$

c) Si $\overline{x}=2$ determinar $\overline{y}, a_{20}, a_{02}$ y a_{11}.

$$y=\frac{1}{2}x+2\longrightarrow \overline{y}=\frac{1}{2}\overline{x}+2=\frac{1}{2}\cdot 2+2=3$$
$$a_{20}=\frac{\sum_{i=1}^n x_i^2}{n}=\overline{x^2},\quad a_{02}=\frac{\sum y_i^2}{n}=\overline{y^2}$$
$$S_x^2=\overline{x^2}-(\overline{x})^2\longrightarrow \overline{x^2}=S_x^2+(\overline{x})^2=9+2^2=13$$
$$a_{11}=\frac{\sum x_i\cdot y_i}{n}=\overline{xy}$$
$$S_{xy}=\overline{xy}-\overline{x}\cdot\overline{y}\longrightarrow \frac{9}{2}=\overline{xy}-2\cdot 3\longrightarrow \overline{xy}=\frac{9}{2}+6=10.5$$

3) Conocemos la siguiente información sobre las medidas de una variable estadísticas bidimensional (X,Y):

$$(CV)_x=2(CV)_y,\quad \overline{x}=4,\quad \overline{y}=7,\quad R=1$$

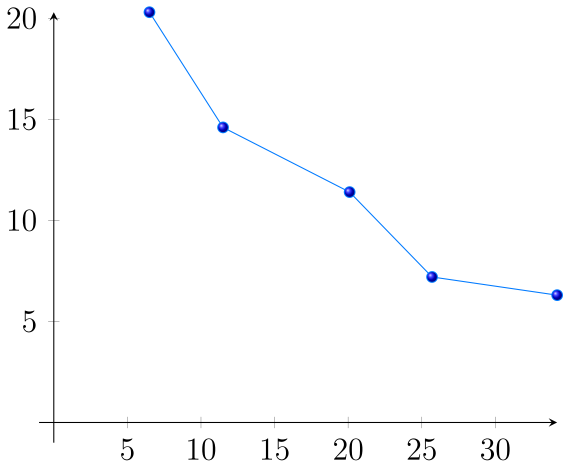
Determinar la recta por mínimos cuadrados de la variable Y sobre la variable X.

$$(y-\overline{y})=\frac{S_{xy}}{S_x^2}(x-\overline{x})$$
$$\frac{S_{xy}}{S_x^2}=\frac{S_{xy}}{S_x^2}\cdot\frac{S_y}{S_y}=\underbrace{\frac{S_{xy}}{S_x\cdot S_y}}_{1}\cdot\frac{S_y}{S_x}=1\cdot\frac{7}{8}=\boxed{\frac{7}{8}}$$
$$R_{xy}\longleftrightarrow\frac{S_{xy}}{S_x\cdot S_y}=1$$
$$\begin{matrix}(CV_x)=\dfrac{S_x}{\overline{x}}\\[2ex](CV_y)=\dfrac{S_y}{\overline{y}}\end{matrix}\quad\begin{cases}\text{Sabemos que}\\ \dfrac{S_x}{\overline{x}}=2\left(\dfrac{S_y}{\overline{y}}\right)\longrightarrow\dfrac{S_y}{S_x}=\dfrac{\overline{y}}{2\overline{x}}=\dfrac{7}{2\cdot 4}=\dfrac{7}{8}\end{cases}$$
$$\boxed{y-7=\frac{7}{8}(x-4)\longrightarrow y=\frac{7}{8}x-\frac{7\cdot 4}{8}+7\longrightarrow y=\frac{7}{8}+\frac{7}{2}}$$

4) Consideremos los siguientes datos de la variable bidimensional (X,Y):

X	Y
6.5	20.3
11.5	14.6
20.1	11.4
25.7	7.2
34.2	6.3

a) Representar gráfica la nube de puntos.



b) Realzar un ajuste exponencial del tipo $y = a \cdot e^{bx}$ a estos datos y representar gráficamente el resultado.

Ajuste exponencial de $y = a \cdot e^{bx}$

Hay que pasarlo a lineal tomando logaritmos y dando sus valores nuevos a y.

$(y_{\text{new}} - \overline{y}_{\text{new}} = \dfrac{S_{x,y_{\text{new}}}}{S_x^2})(x - \overline{x})) \longrightarrow y_{\text{new}} - 2.35 = \dfrac{-4.7}{97.56}(x - 19.6) \longrightarrow y_{\text{new}} = -0.048x + 3.299 \equiv \ln(y) \longrightarrow y = e^{-0.48x} \cdot e^{3.299} = \underbrace{27.086}_{a} \cdot \underbrace{e^{0.048x}}_{b=-0.048}$

$$\overline{y}_{new} = \frac{3.01 + \cdot 1.66}{5} = 2.35$$

$$\overline{x} = \frac{6.5 + \cdots + 34.2}{5} = 19.6$$

$$S_{x,y_{\text{new}}} = \overline{xy}_{new} - \overline{x} \cdot \overline{y}_{new} = \frac{6.5 \cdot 3.01 + \cdots + 34.2 \cdot 1.66}{5} - (19.6) \cdot (2.35) = -4.7$$

$$S_x^2 = \overline{x^2} - (\overline{x})^2 = \frac{6.5^2 + \cdots + 34.2^2}{5} = 97.56$$

c) Comprovar la bondad del ajuste realizado

$$R_{x,y_{\text{new}}} \frac{S_{x,y_{\text{new}}}}{S_x \cdot S_{y_{\text{new}}}} = \frac{-4.7}{9.87 \cdot 0.48} = -0.99$$

$$S_x = \sqrt{97.56} = 9.87; \; S_{y_{\text{new}}} = 0.48$$

Tema 3: Espacios de Probabilidad

3.1) Instroducción y definiciones de probabilidad

Disitnguen varios tipos de experimentos:

- Experimentos deterministas.
- Experimentos aleatorios: proporcionan resultados distintos aunque se realicen bajo las mismas pruebas.
- Experimento pseudoaleatorios.

La Teoría de la Probabilidad estuida los experimentos aleatorios.

- Definición (Laplace)

Si todos los resultados del experimentos son equiprobables, se define

$$P(A) = P(\text{"Ocurre } A\text{"}) = \frac{\text{n}^{\underline{\text{o}}} \text{ de casos favorables } A}{\text{n}^{\underline{\text{o}}} \text{ de casos posibles}}$$

Ejemplo

$$P(\text{"Sacar n}^{\underline{\text{o}}}\text{ par al lanzar un dado"}) \longrightarrow \frac{3}{6} = \frac{1}{2}$$

- Definición (Teoría de Von Misses)

Vale en el caso de experimentos con resultados no equiprobables.

$$P(A) = \lim_{n\to\infty} \frac{\text{N}^{\underline{\text{o}}} \text{ de veces que obtengo } A}{n} = \lim_{n\to\infty} f_r(A) \text{(frecuencia relativa)}$$

donde $n =$"n$^{\underline{\text{o}}}$ de veces que repito el experimento".

- Inconveniente: no siempre puedo calcular el límite o repetir el experimento.

- Definición (Axiomática de Kolmogorov)

Nota: Como $30 << \overline{x} = 44 \longrightarrow$ sospechamos asimetría cola derecha o bien atípico.

Es la que se usa en Teoría de la Probabilidad y permite construir una teoría consistente.

Nota: Para contar casos favorables y posibles usamos reglas de conteo (Matemática Discreta).

- Espacios de probabilidad y σ-álgebras

Consideraremos un experimento aleatorio.

- Definición
  - Llamaremos espacio muestral, Ω, al conjunto de todos los resultados individuales posibles del experimento aleatorio.
  - Llamaremos suceso elemental a cada uno de los resultados individuales.
  - Llamaremos suceso a cualquier subconjunto de Ω (agrupaciones de resultados.)

Denotaremos por $\mathcal{S}$ al conjunto de todos los sucesos sobre los que queremos calcular su probabilidad. Observar que $\mathcal{S} \subseteq P(\Omega)$.

Los sucesos suelen denotar por letras mayúsculas $A, B, C, \ldots$

Ejemplo

Experimento aleatorio: "Observar el resultado al lanzar un dado".

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Sucesos elementales: $\{\{1\}, \{2\}, \ldots, \{6\}\}$
- Algunos sucesos:

  $A =$"Obtener número par"$= \{2, 4, 6\} \subseteq \Omega \longrightarrow$ un suceso.

  $B =$"Obtener valor $\leq 3 = \{1, 2, 3\} \subseteq \Omega \longrightarrow$ un suceso.

Operaciones con sucesos:

- $A \cup B$
- $A \cap B$
- $A^c = \overline{A} = \Omega - A$
- $A + B = A \cap B^c$
- $A \triangle B = (A - B) \cup (B - A)$

- Propiedades
  - $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
  - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
  - $A \cup A^c = \Omega$
  - Leyes de Morgan:
    - $(A \cup B)^c = A^c \cap B^c$
    - $(A \cap B)^c = A^c \cup B^c$

- Definición Axiomática de Kolmogorov

Llamaremos espacio de probabilidad a una terna $(\Omega, \mathcal{S}, P)$ formada por un conjunto Ω, un conjunto $\mathcal{S} \subseteq P(\Omega)$ y una función de posibilidad

$$\begin{aligned} P : & \mathcal{S} \longrightarrow \mathbb{R} \\ & A \longrightarrow P(A) \end{aligned}$$

Verificando los siguientes axiomas:

Ax1) $\Omega \in \mathcal{S}$

Ax2) Si $A \in \mathcal{S}$, entonces $A^c \in \mathcal{S}$

Ax3) Si $A_1, A_2, \dots \in \mathcal{S}$, entonces $A_1 \cup A_2 \cup \dots \in \mathcal{S}$

Ax4) $P(\Omega) = 1$

Ax5) $P(A) \geq 0 \, \forall A \in \mathcal{S}$

Ax6) Si $A_1, A_2, \dots \in \mathcal{S}$ sucesos dispuestos dos a dos $A_i \cap A_j = \varnothing \forall i \neq 0$, entonces $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P_(A_2) + \cdots$

- Definición

Si Ω es un conjunto y $\mathcal{S}$ cumples los axiomas 1, 2 y 3 se dice que $\mathcal{S}$ es una σ-álgebra.
La σ-álgebra generada por los intervalos abierto de $\mathbb{R}$ (de $\mathbb{R}^n$) se llama σ-álgebra de Borel.

Ejemplo

Experimento aleatorio con Ω finito. $\Omega = \{w_1, w_2, \dots, w_n\}$
$\mathcal{S} = P(\Omega)$ es una σ-álgebra (porque $P(\Omega)$ siempre es σ-álgebra).

$$\begin{aligned} P : & \mathcal{S} \longrightarrow \mathbb{R} \\ & A \longrightarrow P(A) = \frac{\text{casos favorables a } A}{\text{casos posibles experimento}} = \frac{\text{cardinal}(A)}{n} = \frac{|A|}{n} \end{aligned}$$

Entonces $(\Omega, \mathcal{S}, P)$ es un espacio de probabilidad.

- Demostración

Ax1,2,3,4 y 5 triviales. Como Ω es finito bastaría probar Ax6 para el caso de dos sucesos. Sean A y B sucesos disjuntos $A \cap B = \varnothing$.

¿$P(A \cup B) = P(A) + P(B)$?

$$P(A \cup B) = \frac{|A \cup B|}{n} = \{\text{porque } A \cap B = \varnothing\} = \frac{|A| + |B|}{n} = \frac{|A|}{n} + \frac{|B|}{n} = P(A) + P(B)$$

3.2) Propiedades de las σ-álgebras y de la función probabilidad

- Propiedades

Sea $\mathcal{S}$ una σ-álgebra de Ω entonces:

1) $\varnothing \in \mathcal{S}$: Por Ax1 sabemos que $\Omega \in \mathcal{S}$ y por Ax2 sabemos que $\Omega^c = \varnothing \in \mathcal{S}$.

2) Si A y $B \in \mathcal{S} \longrightarrow A \cup B \in \mathcal{S}$: Aplico el Ax3 con los sucesos $A_1 = A, A_2 = B, A_3 = A_4 = \cdots = \varnothing$.

3) Si $A, B \in \mathcal{S} \longrightarrow A \cap B \in \mathcal{S}$: Sbamos que $\underbrace{A^c, B^c \in \mathcal{S}}_{Ax2}$ y también $\underbrace{A^c \cup B^c}_{\text{Prop. 2}}$

4) $A, B \in \mathcal{S} \longrightarrow A - B \in \mathcal{S} : A - B = A \cap B^c \in \mathcal{S}$

5) $A, B \in \mathcal{S} \longrightarrow A \triangle B \in \mathcal{S} : A \triangle B = (A - B) \cup (B - A) \in \mathcal{S}$ por ser unión de sucesos.

6) Si $A_1, A_2, \cdots \in \mathcal{S}$, entonces $A_1 \cap A_2 \cap \cdots \in \mathcal{S}$: Similar a (3).

- Propiedades

Nota: Las σ-algebraicas son cerradas para uniones e intersecciones numerables y también para complementarios, diferencias y diferencias simétricas.

Sea $(\Omega, \mathcal{S}, P)$ un espacio de probabilidad. Entonces:

1) $P(\varnothing) = 0$

2) $P(A^c) = 1 - P(A)$

3) Si $A, B \in \mathcal{S}$ tal que $A \cap B = \varnothing \longrightarrow P(A \cap B) = P(A) + P(B)$

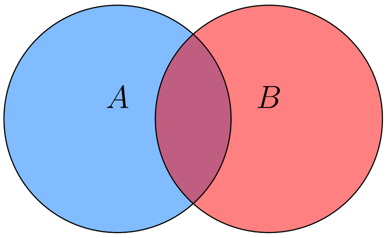
4) Si $A \subseteq B \longrightarrow P(B - A) = P(B) - P(A)$

5) Si $A \subseteq B \longrightarrow P(A) \leq P(B)$

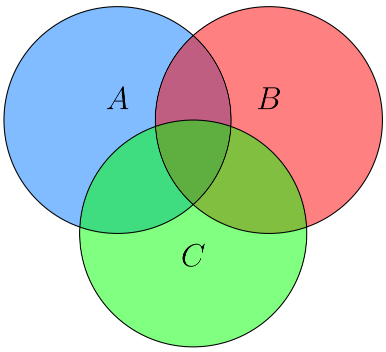
6) Si $A \in \mathcal{S} \longrightarrow P(A) \leq 1$

7) $A, B \in \mathcal{S} \longrightarrow P(A \cup B) \leq P(A) + P(B)$

8) $A, B \in \mathcal{S} \longrightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$



9) $A, B, C \in \mathcal{S} \longrightarrow P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cup B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$



- Propiedades (Función probabilidad)

Proposición 1

$P(\varnothing) = 0$

Ax4 $\quad P(\Omega) = 1$

Ax6 $\quad P(UA) = \sum P(A_i)$

$A - i \cap A_j = \varnothing \; \forall i \neq j$

$A_1 = \Omega, \; A_2 = A_3 = \cdots = \varnothing$

$$\begin{aligned} P(UA_i) &= P(\Omega) \\ &= P(\Omega) + P(\varnothing) + \cdots + P(\varnothing) \\ &\longrightarrow P(\varnothing) = 0 \end{aligned}$$

Proposición 2

$P(A^c) = 1 - P(A)$

– Demostración:

$\Omega = A \cup A^c \longrightarrow \underbrace{P(\Omega)}_{1} = P(A) + P(A^c)$

con A y A^c incompatibles

PROBLEMAS. RELACIÓN 3: Espacio de probabilidad.

FUNDAMENTOS DE PROBABILIDAD Y ANÁLISIS EXPLORATORIO DE DATOS.

GRADO EN CIENCIA E INGENIERÍA DE DATOS.

1. Un experimento aleatorio puede tener cuatro resultados diferentes $\Omega = \{a, b, c, d\}$. Estudiar si las siguientes asignaciones de probabilidad pueden ser correctas:

   a) P(a) = 0.1, P(b) = 0.3, P(c) = 0.4 y P(d) = 0.2
   b) P(a) = 1/5, P(b) = 1/5, P(c) = 1/5 y P(d) = 1/4
   c) P(a) = 0.6, P(b) = −0.2, P(c) = 0.4 y P(d) = 1.2
   d) P(a) = 1/5, P(b) = 1/4, P(c) = 1/3 y P(d) = 1/6

2. Si A, B, C y D son cuatro sucesos mutuamente excluyentes de un espacio de probabilidad $(\Omega, \mathcal{S}, \mathrm{P})$, estudiar si las siguientes asignaciones pueden ser correctas:

   a) P(A) = 0.1, P(B) = 0.3, P(C) = 0.4 y P(D) = 0.2
   b) P(A) = 1/5, P(B) = 1/5, P(C) = 1/5 y P(D) = 1/4
   c) P(A) = 0.6, P(B) = −0.2, P(C) = 0.4 y P(D) = 1.2
   d) P(A) = 1/5, P(B) = 1/4, P(C) = 1/3 y P(D) = 1/2
   e) Responder a las cuatro cuestiones anteriores suponiendo que los sucesos A, B, C y D no son mutuamente excluyentes.

3. Poner un ejemplo de un espacio de probabilidad donde $\mathcal{S} \neq \wp(\Omega)$.

4. Construir el conjunto de sucesos S de un espacio de probabilidad sabiendo que $A, B \in S$. Calcular $P(A)$ y $P(B)$ sabiendo que $P(A \cup B) = \frac{3}{5}$, $P(A \cap B) = \frac{1}{5}$ y $P(A - B) = \frac{1}{5}$.

5. Sean $P(A), P(B), P(A \cap B)$ las probabilidades de los sucesos $A, B, A \cap B$, respectivamente. Determinar en función de ellas:
   a) $P(A \cap \overline{B})$.
   b) $P(\overline{A} \cap B)$.

6. Se lanzan dos monedas al azar.
   a) Calcular la probabilidad de que ambas monedas muestren cara después del lanzamiento.
   b) Calcular la probabilidad de obtener al menos una cara.

Nota. En todos los problemas en que se manejan bolas, urnas, dados, monedas, etc. estos objetos son distintos mientras no se diga lo contrario.

7. De una baraja francesa (52 cartas) se sacan al azar tres cartas. Se pide
   a) La probabilidad de que en las tres cartas extraídas haya exactamente un as
   b) La probabilidad de que en las tres cartas extraídas haya al menos un as

8. Se toman al azar 4 cartas de una baraja española (que se compone de 40 cartas distribuidas en 4 palos: oros, copas, bastos y espadas). Se pide la probabilidad de que
   a) ninguna de las 4 cartas sea de oros
   b) exactamente una de las 4 cartas sea de oros
   c) sean las 4 de oros
   d) sean las 4 de palos distintos

9. Una baraja francesa tiene 52 cartas, 26 rojas (los corazones y los diamantes) y 26 negras (picas y tréboles). Se divide la baraja (al azar) en dos partes iguales (26 cartas en cada parte). Calcular la probabilidad de que en cada parte haya igual número de cartas negras que de rojas.

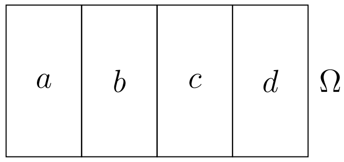
10. En un sobre hay veinte papeletas, ocho llevan dibujado un coche y el resto están en blanco. Hallar la probabilidad de extraer al menos una papeleta con el dibujo de un coche si:
    a) Se saca sólo una papeleta.
    b) Se extraen dos papeletas.
    c) Se extraen tres papeletas.

11. Cuatro matrimonios se reúnen a comer y se sientan al azar en las 8 sillas que hay alrededor de una mesa redonda sin que queden huecos entre dos sillas.
    a) Calcular la probabilidad de que cada caballero quede junto a su esposa.
    b) Calcular la probabilidad de que queden alternados las señoras y los caballeros, o sea, cada señora quede entre dos caballeros.

12. a) ¿Cuál es la probabilidad de que en un aula con 25 alumnos nunca coincidan dos cumpleaños?
    b) ¿Cuál es la probabilidad de que en un aula con 25 alumnos, al menos coincidan dos cumpleaños?

13. Supongamos que lanzamos un dado dos veces:
    a) ¿Cuál es la probabilidad de que la suma de los puntos obtenidos sea igual a 8?
    b) ¿Cuál es la probabilidad de que la suma de los puntos obtenidos sea igual a 9?

1) Un experimento aleatorio puede terner cuatro resultados diferentes $\Omega = \{a, b, c, d\}$. Estudiar si las siguientes asignaciones de probabilidad pueden ser correctas:



a) $P(a) = 0.1, P(b) = 0.3, P(c) = 0.4$ y $P(d) = 0.2$

Es correcto porque suma 1.

b) $P(a) = \frac{1}{5}, P(b) = \frac{1}{5}, P(c) = \frac{1}{5}$ y $P(d) = \frac{1}{4}$

No suma 1 $\longrightarrow$ No es correcto

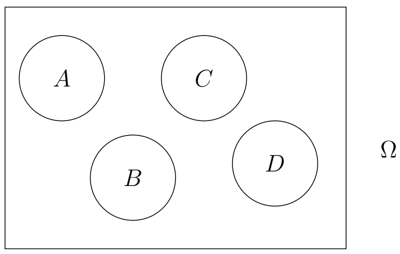
c) $P(a) = 0.6, P(b) = -0.2, P(c) = 0.4$ y $P(d) = 1.2$

$P(b)$ tiene valor negativo $\longrightarrow$ No es correcto

d) $P(a) = \frac{1}{5}, P(b) = \frac{1}{4}, P(c) = \frac{1}{3}$ y $P(d) = \frac{1}{6}$

No suma 1 $\longrightarrow$ No es correcto

2) Si A, B, C y D son cuatro sucesos mutuamente excluyentes de un espacio de probabilidad $(\Omega, \mathcal{S}, P)$, estudiar si las siguientes asignaciones pueden ser correctas:



Debe cumplirse $P(i) \geq 0 \; \forall i = A, B, C, D$ y $P(A) + P(B) + P(C) + P(D) \leq 1$

a) $P(A) = 0.1, P(B) = 0.3, P(C) = 0.4$ y $P(D) = 0.2$

Si es correcta

b) $P(A) = \frac{1}{5}, P(B) = \frac{1}{5}, P(C) = \frac{1}{5}$ y $P(D) = \frac{1}{4}$

$$\sum_{i=A,\dots,D} p(i) = \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{4} = \frac{17}{20} < 1. \text{ Si es correcta}$$

c) $P(A) = 0.6, P(B) = -0.2, P(C) = 0.4$ y $P(D) = 1.2$

No es correcta porque $P(B) = -0.2 < 0$

d) $P(A) = \frac{1}{5}, P(B) = \frac{1}{4}, P(C) = \frac{1}{3}$ y $P(D) = \frac{1}{2}$

$\sum_{i=A,\ldots,D} p(i) = \frac{1}{5} + \frac{1}{4} + \frac{1}{3} + \frac{1}{2} = \frac{77}{60} > 1 \longrightarrow$ No es correcta (pero sí en el apartado (e) sí lo es.)

e) Responder a las cuatro cuestiones anteriores suponiendo que los sucesos A, B, C y D no son mutuamente excluyentes.

3) Poner un ejemplo de un espacio de probabilidad donde $\mathcal{S} \neq \wp(\Omega)$

Exp ="Lanzar un dado".$\Omega = \{1, 2, 3, 4, 5, 6\}$

A ="Obtener número par"

$\mathcal{S} = \{\omega, \varnothing, A, A^c\}$ es una σ-álgebra.

$\mathcal{S} \subseteq P(\Omega)$

4) Construir el conjunto de sucesos S de un espacio de probabilidad sabiendo que $A, B \in S$. Calcular $P(A)$ y $P(B)$ sabiendo que $P(A \cup B) = \frac{3}{5}, P(A \cap B) = \frac{1}{5}$ y $P(A - B) = \frac{1}{5}$.

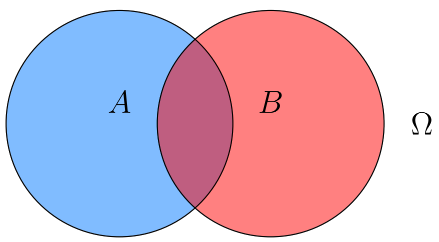
$A, B \in \mathcal{S}$

$\mathcal{S} = \{\Omega, \varnothing, A, B, A^c, B^c, A \cup B, A \cap B, A^c \cap B^c, A^c \cup B^c, A \cap B^c, B \cap A^c, A \cup B^c, B \cup A^c, A \triangle B, (A \triangle B)^c\}$

$P(A \cup B) = \frac{3}{5}, P(A \cap B) = \frac{1}{5}, P(A - B) = \frac{1}{5}$ ¿$P(A)$ y $P(B)$?

- $\underbrace{P(A \cup B)}_{\frac{3}{5}} = P(A) + P(B) - \underbrace{P(A \cap B)}_{\frac{1}{5}} \longrightarrow P(B) = \frac{2}{5}$

- $\underbrace{P(A - B)}_{\frac{1}{5}} = P(A) - \underbrace{P(A \cap B)}_{\frac{1}{5}} \longrightarrow P(A) = \frac{2}{5}$



5) Sean $P(A), P(B), P(A \cap B)$ las probabilidades de los sucesos $A, B, A \cap B$, respectivamente. Determinar en función de ellas:

a) $P(A \cap \overline{B})$

$P(A \cap \overline{B}) = P(A - B) = P(A) - P(A \cap B)$

b) $P(\overline{A} \cap B)$

$P(\overline{A} \cap B) = P(B - A) = P(B) - P(A \cap B)$

6) Se lanzan dos monedas al azar.

a) Calcula la probabilidad de que ambas monedas muestren cara después del lanzamiento.

Experimento aleatorio: "Observar el resultado de lanzar 2 monedas".

C_i ="Ha salido cara en la moneda i", $i = 1, 2$ $+_i$ ="Ha salido cruz en la moneda i", $i = 1, 2$

$\Omega = \{\{C_1, C_2\}, \{C_1, +_2\}, \{+_1, C_2\}, \{+_1, +_2\}\} \longrightarrow$ Espacio muestral finito con 4 elementos (4 sucesos elementales equiprobables).

$P(C_1, C_2) = \dfrac{1}{4}$

b) Calcular la probabilidad de obtener al menos una cara. $P(\Omega - \{x_1, x_2\}) = P\overline{(+_1, +_2)} = 1 - P(+_1, +_2) = 1 - \dfrac{1}{4} = \dfrac{3}{4}$

Nota: En todos los problemas en que se manejan bolas, urnas, dados, monedas, etc. estos objetos son distintos mientras no se diga lo contrario.

7) De una baraja francesa (52 cartas) se sacan al azar tres cartas. Se pide:

a) La probabilidad de que en las tres cartas extraídas al azar haya exactamente un as.

Experimento aleatorio: "Observar el resultado al extraer 3 cartas de la baraja"

$\Omega = \{\{3_T, 2_E, 1_P\}, \{1_E, 1_T, !_p\}, \dots\} \longrightarrow$ Espacio muestra finito. Todas las sucesiones son equiprobables.

¿Cuántas hay? $|\Omega|$ = casos posibles (aplico definición de Laplace)

$|\Omega| = \binom{52}{3} = \dfrac{52!}{3!(52-3)!} = 22100$ n°de casos posibles

P("Exactamente un as")= $\dfrac{\text{n°de casos favorables}}{ncasosposibles} = \dfrac{4512}{22100} \simeq 0.2$

n°casos posibles= $\binom{4}{1} \cdot \binom{48}{2} = 4512$

b) La probabilidad de que en las tres cartas extraídas haya al menos un as.

P("Haya al menos 1 as")=1-P("No haya ningún as")$=1 - 0.78 = 0.22$

P("No haya ningún as")= $\dfrac{\binom{48}{3}}{\binom{52}{3}} \simeq 0.78$

8) Se toman al azar 4 cartas de una baraja española (que se compone de 40 cartas distribuidas en 4 palos: oros, copas, bastos y espadas). Se pide la probabilidad de que:

a) Ninguna de las 4 cartas sean de oros.

Exp. aleatorio="Obtener el resultado de extraer 4 cartas de la baraja"

$\Omega = \{\{3_C, 1_O, 2_E, 4_B\}, \dots\}$ $\longrightarrow$Espacio finito. Todas las sucesiones son equiprobables.

(aplico definición de Laplace)

$\Omega = \binom{40}{4}$ nº de casos posibles

P("Ninguna sea Oros")= $\dfrac{\text{nº de casos favorables}}{ncasosposibles} = \dfrac{\binom{30}{4}}{\binom{40}{4}} \simeq 1.24$

b) Exactamente una de las cartas sea de oros.

P("Exactamente un oro")= $\dfrac{\binom{10}{1}\cdot\binom{30}{3}}{\binom{40}{4}} \simeq 0.44$

c) Sean las 4 de oros.

P("Sean las 4 oros")= $\dfrac{\binom{10}{4}}{\binom{40}{4}} \simeq 0.0023$

d) Sean las 4 de palos distintos.

P("Sean 4 palos distintos")= $\dfrac{\binom{10}{1}\cdot\binom{10}{1}\cdot\binom{10}{1}\cdot\binom{10}{1}}{\binom{40}{4}} \simeq 0.11$

9) Una baraja francesa tiene 52 cartas, 26 rojas (los corazones y los diamantes) y 26 negras (picas y tréboles). Se divide la baraja (al azar) en dos partes iguales (26 cartas en cada parte). Calcular la probabilidad de que en cada parte haya igual número de cartas negras que de rojas.

Exp. aleatorio = ("Observar el resultado de separar 2 montones de 26 cartas").

$\Omega = \{\{\text{Roja} \oplus \text{Negras}\}, \{\text{DiaTreb} \oplus \text{PieCora}, \dots\}\} \longrightarrow$ Espacio muestral finito. Todas las sucesiones son equiprobables. (Laplace)

¿Cuántas hay? $|\Omega|$ = casos posibles

$|\Omega| = \binom{52}{26}$ nº casos posibles.

P("Haya el mismo nº de cartas negras y rojas en cada montón")= $\dfrac{\text{nº de casos favorables}}{\text{nº de casos posibles}} =$

$\dfrac{\binom{26}{13}\cdot\binom{26}{13}}{\binom{52}{26}} \sim 0.212 \equiv$P("13 rojas y 13 negras en cada una de las partes")

10) En un sobre hay veinte papeletas, ochos llevan dibujando un coche y el resto están en blanco. Hallar la probabilidad de extraer al menos una papeleta con el dibujo de un coche si:

$$\text{20 papeletas}\left\langle\begin{array}{l}\text{8 coches}\\ \\ \text{12 blancos}\end{array}\right.$$

a) Se saca sólo una papeleta.

P("No coche")$=\dfrac{\text{n}^{\text{o}} \text{ de casos favorables}}{\text{n}^{\text{o}} \text{ de casos posibles}} = \dfrac{12}{20} = \dfrac{3}{5}$
P("Al menos un coche")=1-P("No salga ningún coche")$=1-\dfrac{3}{5}=\dfrac{2}{5}=0.4$

b) Se extraen dos papeletas.

P("No coche")$=\dfrac{\binom{12}{2}}{\binom{20}{2}} \simeq 0.35$
P("Al menos un coche")$=1-0.35=0.65$

c) Se extraen tres papeletas.

P("No coche")$=\dfrac{\binom{12}{3}}{\binom{20}{3}} \simeq 0.19$
P("Al menos un coche")$=1-0.19=0.81$

11) Cuatro matrimonios se reúnen a comer y se sientan al azar en las 8 sillas que hay alrededor de una mesa redonda sin que queden huevos entre dos sillas.

a) Calcular la probabilidad de que cada caballero quede junto a su esposa.

Experimento aleatorio="Observar en qué silla se situan los 4 caballero y sus esposas al disponerlos aleatoriamente".

4 caballeros y 4 esposas $\left\langle \begin{array}{l} c_1, c_2, c_3, c_4 \\ \\ e_1, e_2, e_3, e_4 \end{array} \right.$

$\Omega = \{(c_1, c_2, c_3, c_4, e_1, e_2, e_3, e_4), \ldots, (e_1, e_2, e_3, e_4, c_1, c_2, c_3, c_4)\} \longrightarrow$ Aplico Laplace
P("Cada caballero con su esposa")$=\dfrac{\text{Casos favorables}}{\text{Casos posibles}} = \dfrac{2 \cdot 4! \cdot 2^4}{8!} = \dfrac{2}{105}$

- <u>Casos favorables:</u> Los matrimonios (4) podrán sentarse juntos en las sillas $\{1,2\},\{3,4\},\{5,6\},\{7$ de 4! formas y además hay 2 opciones de sentar a cada matrimonio en un par de sillas.

- Observarr que tenemos el mismo escenario para los pares de sillas $\{8,1\},\{2,3\},\{4,5\},\{6,7\}$.

b) Calcular la probabilidad de que queden alternados las señoras y los caballeros, o sea, cada señora quede entre dos caballeros.

P("Cada señora entre 2 caballeros")$=\dfrac{2 \cdot 4! \cdot 4!}{8!} = \dfrac{1}{35}$
<u>Casos favorables:</u> Tenemos las siguientes opciones:

1) Todas las señoras a las sillas impares y los caballeros en las pares $\underbrace{\{1,3,5,7\}}_{\text{Señoras}}, \underbrace{\{2,4,6,8\}}_{\text{Caballeros}}$

Ambos se pueden colocar de 4! formas distintas

2) Análogo, pero sillas impares para caballeros.

12) a) ¿Cuál es la probabilidad de que en un aula con 25 alumnos nunca coincidan sus cumpleaños?

b) ¿Cuál es la probabilidad de que en un aula con 25 alumnos, al menos coincidan dos cumpleaños?

13) Supongamos que lanzamos un dado dos veces:

a) ¿Cuál es la probabilidad de que la suma de los puntos obtenidos sea igual a 8?

b) ¿Cuál es la probabilidad de que la suma de los puntos obtenidos sea igual a 9?

Tema 4: Probabilidad condicionada

4.1) Espacio de probabilidad condicionada

- Objetivo

Recalcular la probabilidad de un suceso cuando se dispone de información adicional.

Ejemplo

Experimento aleatorio = "Lanzamiento dado"

$A =$ "Obtener un 6" $\longrightarrow P(A) = \frac{1}{6}$

$B =$ "Obtener número par" $\longrightarrow P(B) = \frac{3}{6} = \frac{1}{2}$

$P(A/B) =$ "Probabilidad de obtener un 6 saliendo que salió n° par" $= \frac{1}{3}$

- Definición

Secuencia $(\Omega, \mathcal{S}, P)$ espacio de probabilidad. Sea $A \in \mathcal{S}$ y $B\mathcal{S}$ tal que $P(B) > 0$. Se define la probabilidad de A condicionada a B como $P(A/B) = \dfrac{P(A \cap B)}{P(B)}$.

Se define: "Probabilidad de que ocurra A sabiendo que ha ocurrido B".

- Proposición

Sea $(\Omega, \mathcal{S}, P)$ espacio de probabilidad y sea B tal que $P(B) > 0$. Entonces, $(\Omega, \mathcal{S}, P(\cdot/B))$ es un espacio de probabilidad.

$$P(\cdot/B) : \mathcal{S} \longrightarrow \mathbb{R}$$
$$A \longrightarrow P(A/B)$$

- Demostración

Ax1, Ax2 y Ax3 $\longrightarrow$ Se cumplen porque tengo la misma σ-álgebra $\mathcal{S}$

Ax4 ¿$P(\Omega/B) = 1$? $P(\Omega/B) = \dfrac{P(\Omega \cap B)}{P(B)} = \dfrac{P(B)}{P(B)} = 1$

Ax5 ¿$P(A/B) \geq 0\, \forall A \in \mathcal{S}$? $P(A/B) = \dfrac{P(A \cap B)}{P(B)} \geq 0\, \forall A \in \mathcal{S}$

Ax6 ¿$P(A_1 \cup A_2 \cup \ldots /B) =^(A_1/B) + P(A_2/B) + \cdots$? Con $A_i \cap A_j = \varnothing\, \forall i \neq j$.

$P(A_1 \cup A_2 \cup \ldots /B) = \dfrac{P\left((A_1 \cup A_2 \cup \ldots) \cap B\right)}{P(B)} = \dfrac{P(A_1 \cap B) \cup P(A_2 \cap B) \cup \cdots}{P(B)}$

$\dfrac{P(A_1 \cap B) + p(A_2 \cap B) + \cdots}{P(B)} = P(A_1/B) + P(A_2/B) + \cdots$

4.2) Propiedades

Además de las propiedades vistas en Tema 3, se cumple las siguientes:

1) $P(A/\Omega) = P(A) \longrightarrow P(A/\Omega) = \dfrac{P(A\cap\Omega)}{P(\Omega)} = \dfrac{P(A)}{P(\Omega)} = P(A)$

2) $P(B/B) = 1 \longrightarrow P(B/B) = \dfrac{P(B\cap B)}{P(B)} = \dfrac{P(B)}{P(B)} = 1$

3) Regla del producto: Si A y B son sucesos tal que $P(A) > 0, P(B) > 0$ entonces,

$$\left.\begin{array}{l} P(A\cap B) = P(B)\cdot P(A/B) \\ P(A\cap B) = P(A)\cdot P(B/A) \end{array}\right\} \longrightarrow \begin{array}{l}\text{Demostración inmediata usando} \\ \text{definición condicionada}\end{array}$$

- Teorema (Probabilidad compuesta)

Sea $(\Omega, \mathcal{S}, P)$ un espacio de probabilidad y sean $A_1, A_2, \ldots, A_n \in \mathcal{S}$ tales que $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$. Entonces

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

Diremos que $\{A_1, A_2, \ldots, A_n\}$ es una partición de Ω si $A_1\cup A_2\cup\cdots\cup A_n = \Omega$ y $A_i\cap A_j = \varnothing$.

- Demostración

Usando la definición de probabilidad condicionada, el término de la derecha quedaría.

$$\cancel{P(A_1)}\cdot\frac{\cancel{P(A_1\cap A_2)}}{\cancel{P(A_1)}}\cdot\frac{P(A_1\cap A_2\cap A_3)}{\cancel{P(A_1\cap A_2)}}\cdot\quad\cdots\quad\cdot\frac{P(A_1\cap A_2\cap\cdots\cap A_n)}{P(A_1\cap A_2\cap\cdots\cap A_{n-1})}$$

Nota: Observar que todos los denominadores son no nulos porque $P(A_1\cap A_2\cap\cdots\cap A_{n-1}) > 0$.

Ejemplo

Se extraen 3 cartas sin reemplazamiento de la baraja española (40 cartas). Calcular la probabilidad de extraer 3 ases.

$A_i =$ "Obtengo un As en la i-ésima extracción", $i = 1, 2, 3$

Me piden $P(A_1\cap A_2\cap A_3) = \left\{\begin{array}{l}\text{Teorema de} \\ \text{probabilidad compuesta}\end{array}\right\} = \left(\dfrac{4}{40}\right)\cdot\left(\dfrac{3}{39}\right)\cdot\left(\dfrac{2}{38}\right) = \dfrac{1}{2470}$

– Otra forma:

$$\frac{\text{casos favorable}}{\text{casos posibles}} = \frac{\binom{4}{3}}{\binom{40}{3}} = \frac{4}{\frac{40!}{3!37!}} = \frac{4}{\frac{40\cdot 38\cdot 39}{3\cdot 2}} = \frac{4\cdot 3\cdot 2}{40\cdot 39\cdot 38} = \frac{1}{2470}$$

- Teorema de la Probabilidad Total

Sea $(\Omega, \mathcal{S}, P)$ espacio de probabilidad y sean $H_1, H_2, \ldots, H_n$ una partición de Ω con $P(H_i) > 0\,\forall i$ y sea $A \in \mathcal{S}$ suceso cualquiera. Entonces:

$$P(A) = P(H_1)\cdot P(A/H_1) + P(H_2)\cdot P(A/H_2) + \cdots + P(H_n)\cdot P(A/H_n).$$

- Demostración

$$A = A\cap\Omega = \left\{\begin{array}{l}\text{por ser } (H_1,\ldots,H_n) \text{ partición}\\[2ex] \text{de } \Omega\end{array}\right\} = A\cap(H_1\cup H_2\cup\cdots\cup H_n) = \left\{\begin{array}{l}\text{propiedad}\\[2ex] \text{distributiva}\end{array}\right\} =$$
$$(A\cap H_1)\cup(A\cap H_2)\cup\cdots\cup(A\cap H_n)$$

Observar que $(A\cap H_i)$ y $(A\cap H_j)$ son incompatibles o distintos $\forall i\neq j$ por ser $H_i\cap H_j = \varnothing\,\forall i\neq j$.

Tomando probabilidades en la igualdad inicial quedaría

$$P(A) = P(A\cap H_1) + P(A\cap H_2) + \cdots + P(A\cap H_n)$$

Regla del Producto $= P(H_1)\cdot P(A/H_1) + P(H_2)\cdot P(A/H_2) + \cdots + P(H_n)\cdot P(A/H_n)$

- Teorema (de Bayes)

En las condiciones del teorema anterior, se tiene que:

$$\underset{i=1,2,\ldots,n}{P(H_i/A)} = \frac{P(H_i)\cdot P(A/H_i)}{P(H_1)\cdot P(A/H_1) + P(H_2)\cdot P(A/H_2) + \cdots + P(H_n)\cdot P(A/H_n)}$$

- Demostración

$$P(H_i/A) = \left\{\begin{array}{l}\text{definición probabilidad}\\[2ex] \text{condicionada}\end{array}\right\} = \frac{P(A\cap H_i)}{P(A)} = \left\{\begin{array}{l}\text{en el numerador aplico la regla del}\\[2ex] \text{producto, en el denominador el}\\[2ex] \text{Teorema de la Probabilidad Total}\end{array}\right\} =$$
$$\frac{P(H_i)\cdot P(A/H_i)}{P(H_1)\cdot P(A/H_1) + P(H_2)\cdot P(A/H_2) + \cdots + P(H_n)\cdot P(A/H_n)}$$

- Nomenclatura

$P(H_i) =$ Probabilidad a priori

$P(A/H_i) =$ Probabilidad a posteriori

Ejemplo

Una urna con 3 bolas blancas y 2 negras. Otra urna con 2 bolas blancas y 3 negras. Se lanza un dado y si sale 1 se extrae una bola de la urna 1. En caso, contrario, se extrae de la urna 2.

a) Probabilidad de sacar bola blanca.

$3B$	$2B$
$2N$	$3N$

$B =$"sacar bola blanca"

$$U_1 = \text{"sacar bola blanca de la urna 1"} \longrightarrow P(U-1) = \frac{1}{6} \qquad \begin{cases} P(B/U_1) = \dfrac{3}{5} \\ P(B/U_2) = \dfrac{2}{5} \end{cases}$$

$$U_2 = \text{"sacar bola blanca de la urna 2"} \longrightarrow P(U_2) = \frac{5}{6}$$

Aplico Teorema de la Probabilidad Total.

$$P(B) = P(U_1) \cdot P(B/U_1) + P(U_2) \cdot P(B/U_2) = \frac{1}{6} \cdot \frac{3}{5} + \frac{5}{6} \cdot \frac{2}{5} = \frac{13}{30}$$

b) Si sale una bola blanca, probabilidad de que fuera en la urna 1.

$$P(U_1/B) = \frac{P(U_1) \cdot P(B/U_1)}{P(B)} = \frac{\frac{1}{6} \cdot \frac{3}{5}}{\frac{13}{30}} = \frac{3}{13}$$

c) Si sale una bola negra, probabilidad de que fuera en la urna 1.

$$P(N) = 1 - P(B) = 1 - \frac{13}{30} = \frac{17}{30} \qquad P(N/U_1) = 1 - P(B/U_1) = \frac{2}{5}$$

$$= (U_1/N) = \frac{P(U_1) \cdot P(N/U_1)}{P(N)} = \frac{\frac{1}{6} \cdot \frac{2}{5}}{\frac{17}{30}} = \frac{2}{17}$$

4.3) Independencia de sucesos

- Definición

Sea $(\Omega, \mathcal{S}, P)$ un espacio de probabilidad y sean $A, B \in \mathcal{S}$ sucesos. Diremos que A y B son independientes si $P(A \cap B) = P(A) \cdot P(B)$

$$\boxed{P(A/B) = \frac{P(A \cap B)}{P(B)} \longrightarrow P(A \cap B) = P(B) \cdot P(A/B) = P(A) \cdot P(B/A)}$$

- Propiedad

1) A y B son independientes.

2) $P(A/B) = P(A)$

3) $P(B/A) = P(B)$

- Demostración trivial usando regla de productos:

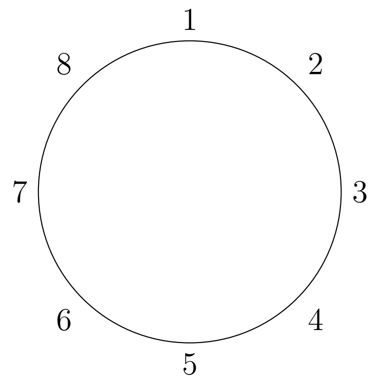
1) Si A y B son independientes

  1.1) A y B^c son independientes

  1.2) A^c y B son independientes

  1.3) A^c y B^c son independientes

- Definición



Diremos que $A_1, \dots, A_n$ son independientes si se cumple que:

$$P(\bigcap_{i\in I} A) = \underset{i\in I}{\pi} \cdot P(A_i) \quad \forall I \leq \{1, 2, \dots, n\}$$

<u>Nota:</u> Observar que si $A_1, \dots, A_n$ son independientes $\longrightarrow$ sin independientes 2 a 2, pero el recíproco **no** es cierto

- Teorema (de Bayes generalizado)

Sea $(\Omega, \mathcal{S}, P)$ un espacio de probabilidad y sea $\{H_1, \dots, H_n\}$ partición Ω. Sean A y $B \in \mathcal{S}$ con $P(A \cap H_i) > 0 \; \forall i = 1, \dots, n$

$$P(B/A) = \frac{\sum_{i=1}^{n} P(H_i) \cdot P(A/H_i) \cdot P(B/A \cap H_i)}{\sum_{i=1}^{n} P(H_i) \cdot P(A/H_i)}$$

- Demostración

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

Aplico el teorema de la Probabilidad Total tanto a numerado como denominador

PROBLEMAS. RELACIÓN 4: Probabilidad condicionada.

FUNDAMENTOS DE PROBABILIDAD Y ANÁLISIS EXPLORATORIO DE DATOS.

GRADO EN CIENCIA E INGENIERÍA DE DATOS.

1. Sabiendo que $P(A\cup B)=0,95, P(A\cap B)=0,35$ y $P(A|B)=0,5$. Hallar $P(\mathrm{A})$, $P(B)$ y $P(\overline{A}\cap\overline{B})$.

2. Sea $(\Omega,\mathscr{S},\mathrm{P})$ un espacio de probabilidad sobre el espacio muestral Ω y sean A, B, C y D sucesos en la σ-álgebra S de forma que A, B y C son disjuntos (incompatibles) y verifican $A\cup B\cup C=\Omega$. Sabiendo que $P(D/A)=0,3; P(D/B)=0,2; P(D)=0,3; P(A)=0,6$ y $P(B)=0,3$. Calcular:
   a) $P(C)$
   b) $P(D\cap A)$
   c) $P(D\cap B)$
   d) $P(D/C)$
   e) $P(D/(A\cup B))$
   f) $P(A/D)$

3. Una urna contiene tres bolas blancas y cinco bolas negras, sacamos dos bolas. ¿Cuál es la probabilidad de que ambas bolas sean blancas
   a) si tras sacar la primera bola, ésta es devuelta a la urna?
   b) si la segunda extracción se realiza sin devolver la primera bola a la urna?

4. Se extraen sucesivamente tres cartas (sin reemplazamiento) de una baraja española (40 cartas con 10 cartas de cada palo, oros, bastos, espadas y copas). ¿Cuál es la probabilidad de que las tres cartas extraídas sean figuras (sota, caballo o rey)?

5. Se lanza una moneda equilibrada dos veces y se consideran los sucesos:

   A={sacar cara la primera tirada} y B={sacar una cara y una cruz}

   a) ¿Son independientes los sucesos A y B?
   b) Supongamos ahora que la moneda no es equilibrada y tiene el doble de probabilidad de obtener cara que cruz. ¿Son independientes los sucesos A y B?

6. Supongamos que lanzamos un dado perfecto dos veces y consideramos los sucesos:
   A={sacar un tres en el primer lanzamiento}
   B={la suma de las puntuaciones obtenidas en los dos lanzamientos es siete}
   C={la suma de las puntuaciones obtenidas en los dos lanzamientos es ocho}
   a) ¿Son los sucesos A y B independientes?
   b) ¿Son los sucesos A y C independientes?

7. a) ¿Cuál es la probabilidad de obtener al menos un seis al lanzar un dado cuatro veces?

   b)¿Cuál es la probabilidad de obtener al menos un seis doble al lanzar dos dados veinticuatro veces?

8. Lanzamos al aire dos monedas equilibradas y consideramos los sucesos:

   A={sacar cruz en la primera moneda lanzada}

   B={sacar cara en la segunda moneda lanzada}

   C={obtener el mismo resultado en las dos monedas lanzadas}

   ¿Son los sucesos A, B y C independientes? ¿Son independientes dos a dos? Razonar la respuesta.

9. Tenemos tres urnas con la siguiente composición. La urna U_1 contiene cinco bolas blancas y cinco bolas negras, la urna U_2 contiene cuatro bolas blancas y cuatro bolas negras y la urna U_3 contiene tres bolas blancas y nueve negras. Se elige una urna al azar y se saca una bola. ¿Cuál es la probabilidad de que la bola extraída sea blanca?

10. La probabilidad de que un jugador de tenis gane un punto con su servicio es "*p*", con independencia de lo que haya sucedido en los demás puntos. En un momento dado, el resultado es DEUCE (40-40). Calcular la probabilidad de que el jugador que sirve gane ese juego.

11. Tenemos tres urnas con la misma composición que en el ejercicio 9.
    a) Se escoge una urna al azar y se extrae una bola que resulta ser blanca. Calcular la probabilidad de que la bola fuese de la urna U_2.
    b) Se escoge una urna al azar y se extrae una bola que resulta ser negra. Calcular la probabilidad de que la bola fuese de la urna U_1.

12. Supongamos que estamos observando el cielo en busca de aviones y que la probabilidad de que haya un avión volando es del 5%. Supongamos también que disponemos de un radar que puede mostrar una señal en la pantalla o no y que la probabilidad de que la muestre es del 99% cuando hay un avión (se le escapa uno de cada cien) y del 10% cuando no lo hay (produce una falsa alarma en estos casos). Si el radar detecta algo, ¿cuál es la probabilidad de que efectivamente haya un aeroplano sobrevolando la zona?

13. Una urna U_1 contiene cinco bolas negras y dos bolas rojas. Otra urna U_2 contiene tres bolas negras y dos bolas rojas. Se traslada una bola de la urna U_1 a la urna U_2 y a continuación se extrae una bola de la urna U_2.

    a) Calcular la probabilidad de que la bola extraída de la urna U_2 sea una bola roja.
    b) Si, efectivamente, la bola extraída de la urna U_2 es una bola roja, calcular la probabilidad de que la bola trasladada fuese una bola negra.

1. Sabiendo que $P(A \cup B) = 0{,}95$, $P(A \cap B) = 0{,}35$ y $P(A|B) = 0{,}5$. Hallar $P(\mathrm{A})$, $P(B)$ y $P(\overline{A} \cap \overline{B})$.

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \longrightarrow P(B) = \frac{P(A \cap B)}{P(A/B)} = \frac{0.35}{0.5} = 0.7$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \longrightarrow P(A) = P(A \cup B) - P(B) + P(A \cap B)$$

$$= 0.95 - 0.7 + 0.35 = 0.6$$

$$\boxed{P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B})} = 1 - P(A \cup B) = 1 - 0.95 = 0.05$$

Leyes de De Morgan

$$\boxed{\begin{array}{l} P(A) = 0.6 \\ P(B) = 0.7 \\ P(\bar{A} \cap \bar{B}) = 0.05 \end{array}}$$

2. Sea $(\Omega, \mathscr{S}, \mathrm{P})$ un espacio de probabilidad sobre el espacio muestral Ω y sean A, B, C y D sucesos en la σ-álgebra S de forma que A, B y C son disjuntos (incompatibles) y verifican $A \cup B \cup C = \Omega$. Sabiendo que $P(D/A) = 0{,}3$; $P(D/B) = 0{,}2$; $P(D) = 0{,}3$; $P(A) = 0{,}6$ y $P(B) = 0{,}3$. Calcular:

a) $P(C)$
b) $P(D \cap A)$
c) $P(D \cap B)$
d) $P(D/C)$
e) $P(D/(A \cup B))$
f) $P(A/D)$

a) $P(C) = 1 - (P(A) + P(B)) = 1 - (0.6 + 0.3) = 0.1$

b) $P(D \cap A) = P(A) \cdot P(D/A) = 0.6 \cdot 0.3 = 0.18$

c) $P(D \cap B) = P(B) \cdot P(D/B) = 0.3 \cdot 0.2 = 0.06$

d) $P(D/C) = \dfrac{P(D \cap C)}{P(C)} = \dfrac{0.06}{0.1} = 0.6$

$P(D) = P(D \cap \Omega) = P(D \cap (A \cup B \cup C)) = P((D \cap A) \cup (D \cap B) \cup (D \cap C)) =$

$P(D) = P(D \cap A) + P(D \cap B) + P(D \cap C) \rightarrow 0.3 = 0.18 + 0.06 + P(D \cap C)$

$P(D \cap C) = 0.06$

e) $P(D/(A \cup B)) = \dfrac{P(D \cap (A \cup B))}{P(A \cup B)} = \dfrac{P((D \cap A) \cup (D \cap B))}{P(A) + P(B)} =$

$\dfrac{P(D \cap A) + P(D \cap B)}{P(A) + P(B)} = \dfrac{0.18 + 0.06}{0.6 + 0.3} = \dfrac{0.24}{0.9} = \dfrac{4}{15}$

f) $P(A/D) = \frac{P(A\cap D)}{P(D)} = \frac{0.18}{0.3} = 0.6$

3. Una urna contiene tres bolas blancas y cinco bolas negras, sacamos dos bolas. ¿Cuál es la probabilidad de que ambas bolas sean blancas
   a) si tras sacar la primera bola, ésta es devuelta a la urna?
   b) si la segunda extracción se realiza sin devolver la primera bola a la urna?

$\boxed{\begin{matrix} 3B \\ 5N \end{matrix}}$ Saco 2 bolas

B_i = "La bola de la extracción i-ésima es blanca", $i=1,2$

Me pide $P(B_1 \cap B_2) = P(B_1) \cdot P(B_2/B_1)$

a) Con reemplazamiento

$P(B_1 \cap B_2) = P(B_1) \cdot P(B_2/B_1) = P(B_1) \cdot P(B_2) = \frac{3}{8} \cdot \frac{3}{8} = \frac{9}{64} \simeq 0.14$

b) Sin reemplazamiento

$P(B_1 \cap B_2) = P(B_1) \cdot P(B_2/B_1) = \frac{3}{8} \cdot \frac{2}{7} = \frac{3}{28} \simeq 0.11$

4. Se extraen sucesivamente tres cartas (sin reemplazamiento) de una baraja española (40 cartas con 10 cartas de cada palo, oros, bastos, espadas y copas). ¿Cuál es la probabilidad de que las tres cartas extraídas sean figuras (sota, caballo o rey)?

F_i: "La carta de la i-ésima extracción es una figura", $i = 1,2,3$

Extracciones sin reemplazamiento.

$P(\text{"3 Figuras"}) = P(F_1) \cdot P(F_2/F_1) \cdot P(F_3/F_1 \cap F_2) =$

$\frac{12}{40} \cdot \frac{11}{39} \cdot \frac{10}{38} = \frac{11}{494} \simeq 0.022$

5. Se lanza una moneda equilibrada dos veces y se consideran los sucesos:
   A={sacar cara la primera tirada} y B={sacar una cara y una cruz}
   a) ¿Son independientes los sucesos A y B?
   b) Supongamos ahora que la moneda no es equilibrada y tiene el doble de probabilidad de obtener cara que cruz. ¿Son independientes los sucesos A y B?

a) Consideremos los sucesos elementales equiprobables, con probabilidad = $1/4$

$\Omega = \{cc, c+, +c, ++\}$

$P(A) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ $\quad$ $P(B) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

$P(A \cap B) = \{c+\} = \frac{1}{4}$

$P(A \cap B) = P(A) \cdot P(B) \rightarrow \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} \rightarrow \frac{1}{4} = \frac{1}{4}$

Si se cumple $\rightarrow$ sí son independientes

b) Ahora P("Obtener cara") = $2/3$

P("Obtener cruz") = $1/3$

Ahora $P(A) = P(\{cc\}) + P(\{c+\}) = 2/3 \cdot 2/3 + 2/3 \cdot 1/3 = 2/3$

Ahora $P(B) = P(\{c+\}) + P(\{+c\}) = 2/3 \cdot 1/3 + 1/3 \cdot 2/3 = 4/9$

Ahora $P(A \cap B) = P(\{c+\}) = 2/3 \cdot 1/3 = 2/9$

Ahora $2/9 \neq 2/3 \cdot 4/9 \longrightarrow$ No son independientes.

6. Supongamos que lanzamos un dado perfecto dos veces y consideramos los sucesos:
   A={sacar un tres en el primer lanzamiento}
   B={la suma de las puntuaciones obtenidas en los dos lanzamientos es siete}
   C={la suma de las puntuaciones obtenidas en los dos lanzamientos es ocho}
   a) ¿Son los sucesos A y B independientes?
   b) ¿Son los sucesos A y C independientes?

a) A y B son sucesos independientes $\longleftrightarrow P(A \cap B) = P(A) \cdot P(B)$

1º lanz	2º lanz
1	6
2	5
$\boxed{3}$	$\boxed{4}$
4	3
5	2
6	1

$p(A) = \frac{1}{6}$

$P(B) = \frac{cf}{cp} = \frac{6}{36} = \frac{1}{6}$

$A \cap B$: P("Sacar A y B")

$P(A \cap B) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$

$\Big\} \rightarrow$ Si son independientes

b) ¿Son A y C independientes?

$P(C) = \frac{5}{36}$

1° lanz	2° lanz
1	
2	6
3	5
4	4
5	3
6	2

$A \cap C = \{$Sacar un 3 en el 1^{er} lanzamiento y un 5 en el 2°$\}$

$$P(A \cap C) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

Luego $P(A \cap C) \neq P(A) \cdot P(C)$

$$\left.\begin{array}{l} P(A \cap C) \neq P(A) \cdot P(C) \\ \frac{1}{36} \neq \frac{1}{6} \quad \frac{5}{36} \end{array}\right\} \rightarrow A \text{ y } C \text{ no son indep.}$$

7. a) ¿Cuál es la probabilidad de obtener al menos un seis al lanzar un dado cuatro veces?

b)¿Cuál es la probabilidad de obtener al menos un seis doble al lanzar dos dados veinticuatro veces?

a) P(Al menos un 6 al lanzar el dado 4 veces)

$S_i = \{$Sacar un 6 en el lanzamiento $i\}$ $i = 1, 2, 3, 4$

$S = \{$Obtener al menos una vez un 6$\}$

$$S = \bigcup_{i=1}^{n} S_i$$

$$P(S) = P\left(\bigcup_{i=1}^{4} S_i\right) = P\left(\left(\cap S_i^c\right)^c\right) = 1 - P\left(\cap S_i^c\right) = 1 - \prod_{i=1}^{4} P(S_i^c) =$$

$$1 - P(S_1^c) \cdot P(S_2^c) \cdot P(S_3^c) \cdot P(S_4^c) = 1 - \left(1 - \frac{1}{6}\right)^4 = 1 - \left(\frac{5}{6}\right)^4 =$$

$$\boxed{0.517746 > \frac{1}{2}}$$

• Más sencillo:

$P($no obtener ningún 6 en los 4 lanzamientos$) = \left(\frac{5}{6}\right)^4$

Por tanto $P($obtener al menos un 6$) = 1 - P($no obtener ninguno$)$

$$= \boxed{1 - \left(\frac{5}{6}\right)^4} = 0.517746$$

• Otra forma:

$P($Obtener al menos un 6 en 4 lanzamientos$) =$

$P($Obtener un 6$) + P($Obtener dos 6es$) + P($Obtener tres 6es$) +$

$$P(\text{Obtener cuatro 6es}) = \binom{4}{1} \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^3 + \binom{4}{2} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^2 + \binom{4}{3} \cdot \left(\frac{1}{6}\right)^3 \cdot \frac{5}{6}$$

$+\binom{4}{4}\cdot\left(\frac{1}{6}\right)^4 = 4\cdot\frac{1}{6}\cdot\left(\frac{5}{6}\right)^3 + 6\cdot\left(\frac{1}{6}\right)^2\cdot\left(\frac{5}{2}\right)^2 + 4\left(\frac{1}{6}\right)^3\frac{5}{6} + \left(\frac{1}{6}\right)^4 =$

$\frac{4\cdot 5^3 + 6\cdot 5^2 + 4\cdot 5 + 1}{6^4} = \frac{500+150+20+1}{6^4} = \frac{671}{6^4} = 0\ 517746$

b) P(Obtener al menos un 6 doble al lanzar 2 dados 24 veces)

$SD_i = \{$Obtener un 6 doble en el lanzamiento $i\}$ $i = 1, \dots, 2$

$SD = \{$Obtener al menos un 6 doble$\}$

$SD = \bigcup_{i=1}^{24} SD_i$

$P(SD) = P\left(\bigcup_{i=1}^{24} SD_i\right) = P\left(\left(\cap SD_i^c\right)^c\right) = 1 - \prod_{i=1}^{24} P(SD_i^c) = 1 - \left(1 - \frac{1}{36}\right)^{24} =$

$1 - \left(\frac{35}{36}\right)^{24} \approx \boxed{0.4914 < \frac{1}{2}}$

- **Más sencillo:**

P(No obtener ningún 6 doble) $= \left(\frac{35}{36}\right)^{24}$

Por tanto: P(Obtener al menos un 6 doble) = 1 - P(no obtener ninguno) $= 1 - \left(\frac{35}{36}\right)^{24}$

$\downarrow$

P(Obtener un 6 doble al lanzar 2 dados) $= \frac{1}{6}\cdot\frac{1}{6}$

P(No obtener un 6 doble al lanzar 2 dados) $= \frac{1}{6}\cdot\frac{1}{6}$

P(No obtener un 6 doble al lanzar 2 dados 24 veces) $= \left(1 - \frac{1}{6}\cdot\frac{1}{6}\right)^{24}$

$= \left(1 - \frac{1}{36}\right)^{24} = \left(\frac{35}{36}\right)^{24}$

8. Lanzamos al aire dos monedas equilibradas y consideramos los sucesos:

A={sacar cruz en la primera moneda lanzada}

B={sacar cara en la segunda moneda lanzada}

C={obtener el mismo resultado en las dos monedas lanzadas}

¿Son los sucesos A, B y C independientes? ¿Son independientes dos a dos? Razonar la respuesta.

$\Omega = \{cc, c+, +c, ++\}$

$P(A) = \{+c, ++\} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

$P(B) = \{cc, +c\} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

$P(A \cap B) = \{+c\} = \frac{1}{4}$

$P(A \cap C) = \{++\} = \frac{1}{4}$

$P(C) = \{cc, ++\} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ $\qquad$ $P(B \cap C) = \{cc\} = \frac{1}{4}$

$P(A \cap B) = P(A) \cdot P(B)$ ✓

$P(A \cap C) = P(A) \cdot P(C)$ ✓

$P(B \cap C) = P(B) \cdot P(C)$ ✓

- ¿A y C independientes? Si
- ¿B y C independientes? Si

- ¿A, B y C independientes?

$A \cap B \cap C = \emptyset$

$P(A \cap B \cap C) = 0 \neq P(A) \cdot P(B) \cdot P(C)$ No son indepen.

9. Tenemos tres urnas con la siguiente composición. La urna U_1 contiene cinco bolas blancas y cinco bolas negras, la urna U_2 contiene cuatro bolas blancas y cuatro bolas negras y la urna U_3 contiene tres bolas blancas y nueve negras. Se elige una urna al azar y se saca una bola. ¿Cuál es la probabilidad de que la bola extraída sea blanca?

u_1	u_2	u_3
5B	4B	3B
5N	4N	9N

$\boxed{P(B/u_1) = \frac{5}{10} = \frac{1}{2} \quad P(B/u_2) = \frac{4}{8} = \frac{1}{2} \quad P(B/u_3) = \frac{3}{12} = \frac{1}{4}}$

$P(B) = P(u_1) \cdot P(B/u_1) + P(u_2) \cdot P(B/u_2) + P(u_3) \cdot P(B/u_3) =$

$\frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{4} = \frac{5}{12}$

10. La probabilidad de que un jugador de tenis gane un punto con su servicio es "*p*", con independencia de lo que haya sucedido en los demás puntos. En un momento dado, el resultado es DEUCE (40-40). Calcular la probabilidad de que el jugador que sirve gane ese juego.

J1 : Jugador que saca

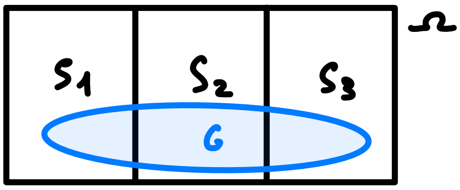
J2 : El otro jugador

Podemos encontramos ante 3 escenarios :

S_1 : { J1 gane los 2 puntos siguientes } $\longrightarrow$ J1 gana

S_2 : { J2 gane los 2 puntos siguientes } $\longrightarrow$ J1 pierde

S_3 : { J1 gana 1 punto y J2 el otro } $\longrightarrow$ DEUCE



$P(S_1) = p \cdot p = p^2$

$P(S_2) = (1-p)^2$

$P(S_3) = 2p(1-p)$

Sea $G = \{J1$ gane el juego$\}$. Queremos calcular $P(G)$, llamémosle x, es decir $P(G) = x$

Por el Teorema de la Probabilidad Total, tenemos:

$$P(G) = P(G|S_1)\cdot P(S_1) + P(G|S_2)\cdot P(S_2) + P(G|S_3)\cdot P(S_3)$$

$$x = 1\cdot p^2 + \cancel{0\cdot(1-p)^2} + x\cdot 2p(1-p) = x(1-2p(1-p)) = p^2 \longrightarrow$$

$$\boxed{x = \frac{p^2}{1-2p+2p^2}}$$

11. Tenemos tres urnas con la misma composición que en el ejercicio 9.
    a) Se escoge una urna al azar y se extrae una bola que resulta ser blanca. Calcular la probabilidad de que la bola fuese de la urna U_2.
    b) Se escoge una urna al azar y se extrae una bola que resulta ser negra. Calcular la probabilidad de que la bola fuese de la urna U_1.

a) $P(U_2/B) = \dfrac{P(B\cap U_2)}{P(B)} = \dfrac{P(U_2)\cdot P(B/U_2)}{P(B)} = \dfrac{1/3 \cdot 1/2}{5/12} = \dfrac{2}{5}$

b) $P(N) = 1 - P(B) = 1 - \dfrac{5}{12} = \dfrac{7}{12}$

$P(U_1/N) = \dfrac{P(U_1)\cdot P(N/U_1)}{P(N)} = \dfrac{1/3\cdot(1-1/2)}{7/12} = \dfrac{2}{7}$

12. Supongamos que estamos observando el cielo en busca de aviones y que la probabilidad de que haya un avión volando es del 5%. Supongamos también que disponemos de un radar que puede mostrar una señal en la pantalla o no y que la probabilidad de que la muestre es del 99% cuando hay un avión (se le escapa uno de cada cien) y del 10% cuando no lo hay (produce una falsa alarma en estos casos). Si el radar detecta algo, ¿cuál es la probabilidad de que efectivamente haya un aeroplano sobrevolando la zona?

$A = \{$hay un avión sobrevolando la zona$\}$

$R = \{$el radar muestra una señal$\}$

Sabemos que $P(A) = 0.05$; $P(R|A) = 0.99$

$P(R|A^c) = 0.01$

y nos piden $P(A|R)$. Usando el Teorema de Bayes.

$$P(A|R) = \frac{P(R|A)\cdot P(A)}{P(R|A)\cdot P(A) + P(R|A^c)\cdot P(A^c)} = \frac{0.99\cdot 0.5}{0.99\cdot 0.05 + 0.1\cdot 0.95} =$$

$$\frac{0.0495}{0.1445} = \boxed{0.34}$$

Esta probabilidad sale baja lo cual puede sorprendernos pese a que estamos usando un radar bastante bueno. Esto sucede porque el suceso A es mucho menos probable que su contrario, lo que hace que las falsas alarmas cobren protagonismo porque su coeficiente de ponderación es alto (0.95).

13. Una urna U_1 contiene cinco bolas negras y dos bolas rojas. Otra urna U_2 contiene tres bolas negras y dos bolas rojas. Se traslada una bola de la urna U_1 a la urna U_2 y a continuación se extrae una bola de la urna U_2.

a) Calcular la probabilidad de que la bola extraída de la urna U_2 sea una bola roja.

b) Si, efectivamente, la bola extraída de la urna U_2 es una bola roja, calcular la probabilidad de que la bola trasladada fuese una bola negra.

u_1	u_2
5 N	3 N
2 R	2 R

$P(BTR)$ = "Bola traslada roja" $= \frac{2}{7}$

$P(BTN)$ = "Bola traslada negra" $= \frac{5}{7}$

$P(R/BTR) = 3/6 = 1/2 \qquad P(R/BTN) = 2/6 = 1/3$

a) $P(R) = P(BTR)\cdot P(R/BTR) + P(BTN)\cdot P(R/BTN) =$

$$\frac{2}{7}\cdot\frac{1}{2} + \frac{5}{7}\cdot\frac{1}{3} = \frac{8}{21}$$

b) Me piden $P(BTN/R)$:

$$P(BTN/R) = \frac{P(BTN)\cdot P(R/BTN)}{P(R)} = \frac{5/7\cdot 1/3}{8/21} = \frac{5}{8}$$

Tema 5: Variables Aleatorias Unidimensionales

5.1) Concepto de Variable Aleatoria

- Definición

Sea $(\Omega, \mathcal{S}, P)$ espacio de probabilidad. Llamaremos variable aleatoria a una aplicación $x : \Omega \longrightarrow \mathbb{R}$ que a cada resultado del espacio muestral le asocia un número de la recta real y tal que

$$x^{-1}((-\infty, x]) = \{w \in \Omega : x(w) \leq x\} \in \mathcal{S}$$

Ejemplo

Nota: Si $\mathcal{S} = P(\Omega)$, entonces $x : \Omega \longrightarrow \mathbb{R}$ siempre es una variable aleatoria.

$X =$"Nº de unidades defectuosas en un lote de 100 unidades" es una variable aleatoria discreta.

$X =$"Concentración de CO_2 en el interior de la clase" es una variable aleatoria continua.
Las variables aleatorias se denotan con letras mayúsculas, las últimas del abecedario: X, Y, Z

5.2) Función de distribución

- Definición

Sea $X : \Omega \longrightarrow \mathbb{R}$ una variable aleatoria llamaremos función de distribución de X a $F_X : R \longrightarrow \mathbb{R}$.

$$x \longrightarrow F_X(x) = P(X \leq x) = P(X \in (-\infty, x])$$

- Propiedades

1) F_X es monótona creciente, linealmente dependiente, si $\underset{x,y\in\mathbb{R}}{x < y} \Longrightarrow F_X(x) \leq F_X(y)$

2) F_X es continua por la derecha, linealmente independiente, $\lim_{t \to x^+} F_X(t) = F_X(x)$

3) $\lim_{x \to -\infty} F_X(x) = 0, \ \lim_{x \to +\infty} F_X(x) = 1$

Nota: Si F_X cumple las propiedades anteriores, entonces existe una variable aleatoria X con distribución F_X.

- Propiedades

Sea $F : \mathbb{R} \longrightarrow \mathbb{R}$ función de distribución. Entonces:

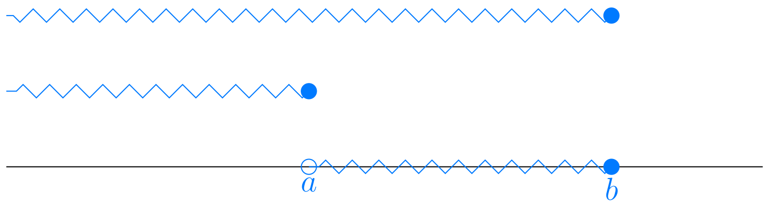
1) $F(x) = P(X \leq x)$

2) $P(X < x) = P(X \in (-\infty, x)) = F(X) = \lim\limits_{t \to x} F(x)$

3) $0 \leq F(x) \leq 1 \; \forall x \in \mathbb{R}$

4) $P(A < X \leq b) = F(b) - F(a)$

5) $P(a < X < b) = F(b^-) - F(a)$



6) $P(X = a) = F(a) - F(a^-)$

- Definición

A la función $\overline{F}(x) = 1 - F(x)$ se le llama función de fiabilidad. Tipos de variables aleatorias:

- Discretas
- Continuas

5.3) Variables Aleatorias Discretas

- Definición

Diremos que $X : \Omega \longrightarrow \mathbb{R}$ es una variable aleatoria discreta si toma un número finito o infinito numerable de valores con probabilidad no nula.

A los valores que toma X se les llama rango de X o soporte de X.

$$\text{Sop}(X) = \{x_1, x_2, \ldots, x_n, \ldots\} \text{ tales que } P(X = x_i) \neq 0 \; \forall 1, 2 \ldots, n, \ldots$$

- Definición

Sea X una variable aleatoria. Se llama función puntual de probabilidad (f.p.p) a la función $p(x) = P(X = x) \; \forall x \in \mathbb{R}$

- $p(x_i) = P(X = x_i) \; \forall x_i \in \text{Sop}(X) \longrightarrow$ Puntos del Soporte(X)

- $p(x) = P(X = x)\ \forall x \in \text{Sop}(X) \longrightarrow$ Puntos fuera del Soporte(X)

- Propiedades

Sea $p(x) = P(X = x)$ función puntual de probabilidad de X. Entonces:

1) $p(x) \geq 0\ \forall x \in \mathbb{R}$

2) $\displaystyle\sum_{x_i \in \text{Sop}(X)} p(x_i) = 1$

> Nota: Si $p(x)$ cumple las propiedades anteriores, entonces existe una variable aleatoria discreta que tiene a $p(x)$ como función puntual de probabilidad .

- Propiedades

$p(x)$ función puntual de probabilidad de la variable aleatoria discreta X y sea $F(x)$ su función de distribución. Entonces:

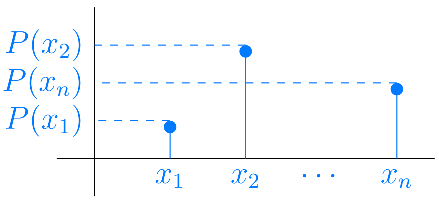
1) $F(x) = \displaystyle\sum_{\substack{x_i \in \text{Sop}(X) \\ x_i \leq x}} p(x_i)$

2) $P(a < X \leq b) = \displaystyle\sum_{\substack{x_i \in (a,b] \\ x_i \in \text{Sop}(X)}} p(x_i)$

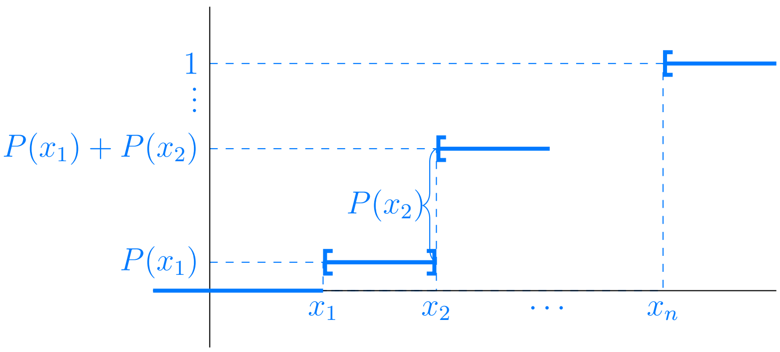
3) $P(X \in A) = \displaystyle\sum_{\substack{x_i \in \text{Sop}(X) \\ x_i \in A}} p(x_i)$

Ejemplo

Representa la función puntual de probabilidad .



$F(x)$: Función de distribución



> Nota: Las variables aleatorias discretas tienen función de distribución escalonada, constante a otros en los puntos de soporte (o rango) y amplitud del salto igual a la probabilidad de dicho punto.

5.4) Variables Aleatorias Continuas

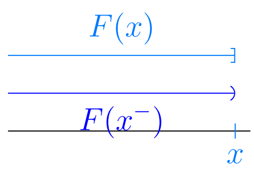
- Definición

Llamaremos variable aleatoria de tipo continuo a las que tienen función de distribución continua.

  - Idea: Las variables aleatorias continuas son las que toman valores en un intervalo de $\mathbb{R}$.

  - Observación: Si X es una variable continua, entonces $P(X = x) = 0 \, \forall x \in \mathbb{R}$.

Vemos que $P(X = x) = F(x) - F(x^-) = 0$



En realidad nos concentraremos en variables aleatorias con función de distribución $F(x)$ derivable en "casi" todo punto, para así disponer de la función de densidad.

- Definición

Diremos que $x : \Omega \longrightarrow \mathbb{R}$ es una variable aleatoria de tipo absolutamente continuo si existe $f : \mathbb{R} \longrightarrow \mathbb{R}^+$ no negativa e integrable tal que

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t) \, \mathrm{d}t$$

A la función $f(t)$ se le llama función de densidad y el soporte X se define como $\mathrm{Sop}(X) = \{x \in \mathbb{R} \text{ tal que } f(x) > 0\}$

- Propiedades

Sea $f : \mathbb{R} \longrightarrow \mathbb{R}$ función de densidad de la variable aleatoria x. Entonces:

1) $f(t) \geq 0 \, \forall t \in \mathbb{R} \longrightarrow$ Demostración: Por definición
2) $\displaystyle\int_{-\infty}^{+\infty} f(t) = 1 \longrightarrow$ Demostración: $\displaystyle\int_{-\infty}^{+\infty} f(t) = \lim_{x \to \infty} F(x) = 1$

Nota: Hay variables aleatorias que son de tipo continuo pero no tienen función de densidad (abuso de notación). En adelante llamaremos variable aleatoria de tipo continuo a las que tengan función de densidad.

- Propiedades

Sea $x : \Omega \longrightarrow \mathbb{R}$ variable aleatoria continua con función de densidad $f_X(x)$ y distribución $F(x)$. Entonces:

1) $P(A < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f_X(t)\,\mathrm{d}t.$

2) $P(X \in A) = \int_A f_X(t)\,\mathrm{d}t$

3) $F'(x) = f(x)$

4) $F(x) = \int_{-\infty}^x f_x(t)\,\mathrm{d}t$

5) $P(X = x) = 0\, \forall x \in \mathbb{R}$

Ejemplo

Si $F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } x \in [0,1) \\ 1 & \text{si } x \geq 1 \end{cases}$ es función de distribución de la variable aleatoria X, calcular su función de densidad $f_X(x)$.

$$f(x) = F'(x) = \begin{cases} 1 & \text{si } x \in (0,1) \\ 0 & \text{en caso contrario} \end{cases}$$

$$P(x > 0.8) = 1 - F(0.8) = 1 - 0.8 = 0.2$$

5.5) Cambios de variable

- Definición

Sea $X : \Omega \longrightarrow \mathbb{R}$ una variable aleatoria y sea $g : \mathbb{R} \longrightarrow \mathbb{R}$ una función tal que la composición $Y = g(X) : \underset{w \longrightarrow g(X(w))}{\Omega \longrightarrow \mathbb{R}}$ es una variable aleatoria . A la variable aleatoria Y se le llama cambio de variable sobre X.

Nota: Si $g : \mathbb{R} \longrightarrow \mathbb{R}$ es continua, entonces $Y = g(X)$ es una variable aleatoria

- Objetivos

1) Obtener la función de distribución de Y a partir de la función de distribución de X.

¿Cómo lo haremos?

Realizando operaciones sobre la función de distribución.

2) Obtener la función puntual de probabilidad de Y en función de la función puntual de probabilidad de X

¿Cómo lo haremos?

Operando directamente.

3) Obtener la función de densidad de Y a partir de la densidad de Y a partir de la densidad de X.

¿Cómo lo haremos?

Dos formas:

a) Obtener la función de distribución y luego derivar.

b) Teorema: Sea $X : \Omega \longrightarrow \mathbb{R}$ variable aleatoria continua con densidad $f_X(t)$ y soporte $\mathrm{Sop}(X) \leq I \subseteq \mathbb{R}$, sea $g : I \longrightarrow J \subseteq \mathbb{R}$ una biyección e $Y = g(X)$ cambio de variable. Entonces:

$$\underbrace{f_Y(y)}_{\substack{\text{función densidad}\\ \text{de } Y}} = f_X\left(g^{-1}(y)\right) \cdot \left|\left(g^{-1}(y)\right)'\right|$$

Ejemplo

Sea $X : \Omega \longrightarrow \mathbb{R}$ variable aleatoria y $a, b \in \mathbb{R}$. Consideramos $Y = aX + b$. Observar que Y es una variable aleatoria porque $g(x) = aX + b$ es función continua. Obtener la función de distribución de Y a partir de la de $X(F(X))$.

$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(aX \leq b - y)$

- Caso 1: Si $a > 0, F_Y(y) = P(aX \leq b - y) = P\left(X \leq \dfrac{b-y}{a}\right) = F_X\left(\dfrac{y-b}{a}\right)$
- Caso 2: Si

  $a < 0, F_Y(y) = P(aX \leq b - y) = P\left(X \geq \dfrac{b-y}{a}\right) = \left\{\begin{array}{c}\text{caso}\\ \\ \text{continuo}\end{array}\right\} = 1 - F_X\left(\dfrac{y-b^-}{a}\right)$
- Caso 3: Si $a = 0, F_Y(y) = P(0 \leq y - b) = \begin{cases} 1 & \text{si } y \geq b \\ 0 & \text{si } y < b \end{cases}$

Ejemplo

Como en el ejemplo anterior, pero ahora suponemos X variable aleatoria discreta con función puntual de probabilidad $P_X(x)$. Considero $Y = aX + b$. Obtener la función puntual de probabilidad de Y a partir de la de X.

$P_X(y) = P(Y = y) = P(aX + b = y)$

- Caso 1: $a \neq 0, P_Y(y) = P(aX + b = y) = P\left(X = \dfrac{y-b}{a}\right) = P_X\left(\dfrac{y-b}{a}\right)$
- Caso 2: $a = 0, P_Y(y) = P(b = y) = \begin{cases} 1 & \text{si } y = b \\ 0 & \text{si } y \neq b \end{cases}$

5.6) Esperanza, varianza y momentos de una variable aleatoria

- Definición

Sea $(\Omega, \mathcal{S}, P)$ espacio probabilidad y sea $X : \Omega \longrightarrow \mathbb{R}$ una variable aleatoria .

- Si X es discreta con $\mathrm{Sop}(X) = \{x_1, x_2, \ldots, x_n, \ldots\}$ y función puntual de probabilidad $p(x)$, se define la esperanza o media de X como

$$\mu_X = E(x) = \sum_{x_i \in \mathrm{Sop}(X)} x_i \cdot p(x_i) = \begin{cases} 0 \cdot \dfrac{1}{2} + 1 \cdot \dfrac{1}{2} = \dfrac{1}{2} \\ 0 \cdot \dfrac{1}{3} + 1 \cdot \dfrac{2}{3} = \dfrac{2}{3} \end{cases}$$

Nota: Para algunas variable aleatoria no existe su media

- Si X es continua con densidad $f(t)$, se define:

$$\mu_X = \int_{-\infty}^{+\infty} t \cdot f(t)\, \mathrm{d}t$$

- Definición

Sea X variable aleatoria y $g : \mathbb{R} \longrightarrow \mathbb{R}$ función. Entonces:

- Si X es discreta con $\mathrm{Sop}(X) = \{x_1, x_2, \ldots, x_n, \ldots\}$ y función puntual de probabilidad $p(x)$, se define la esperanza o media de X como

$$E(g(X)) = \sum_{x_i \in \mathrm{Sop}(X)} g(x_i) \cdot p(x_i)$$

– Si X es continua con densidad $f(t)$, se define:

$$E\left(g(X)\right) = \int_{-\infty}^{+\infty} g(t) \cdot f(t)\, \mathrm{d}t$$

- Definición

Sea $X : \Omega \longrightarrow \mathbb{R}$ una variable aleatoria

– Si X es discreta con $\mathrm{Sop}(X) = \{x_1, x_2, \ldots, x_n, \ldots\}$, se define la varianza de X como

$$\mathrm{Var}(X) = \sigma_X^2 = \sum_{x_i \in \mathrm{Sop}(X)} (x_i - \mu_X)^2 \cdot p(x_i), \text{ con } \mu_X = E(x)$$

– Si X es continua como antes, se define

$$\mathrm{Var}(X) = \sigma_X^2 = \int_{-\infty}^{+\infty} (t - \mu_X)^2 \cdot f(t)\, \mathrm{d}t$$

- Definición

$\sigma_X = \sqrt{\mathrm{Var}(X)}$ desviación típica de X.

- Propiedades

Sean $X : \Omega \longrightarrow \mathbb{R}$ e $Y : \Omega \longrightarrow \mathbb{R}$ variables aleatorias y sean $a, b \in \mathbb{R}$. Entonces:

1) $E(a) = a$

2) $E(aX + b) = a\mu_X + b = a \cdot E(X) + b$

3) $E(X + Y) = E(X) + E(Y)$

4) Si $g(x) \leq h(x)\ \forall x \longrightarrow E\left(f(x)\right) \leq E\left(h(y)\right)$

5) $\mathrm{Var}(a) = 0$

6) $\mathrm{Var}(aX + b) = a^2\, \mathrm{Var}(X)$

7) $\boxed{\mathrm{Var}(X) = E(x^2) - (E(x))^2}$!!

8) Si X tiene media μ_X y varianza σ_X^2, entonces:

$$\frac{X - \mu_x}{\sigma} \text{ tiene media 0 y varianza 1}$$

- Definición

Sea $X : \Omega \longrightarrow \mathbb{R}$ variable aleatoria llamaremos:

a) Momento ordinario o respecto al origen de orden n a $\alpha_n = E(X^n)$

b) Momento centrado o respecto a la media de orden n a

$$\mu_n = E\left((X-\mu_X)^n\right)$$

$$\alpha_1 = E(X) = \mu_X \text{ (media } X)$$

$$\mu_2 = \text{Var}(X) \text{ (varianza } X)$$

5.7) Desigualdades

- Teorema (Desigualdades de Markov)

Sea Z variable aleatoria no negativa con media finita $E(Z)$

$$z : \Omega \longrightarrow \mathbb{R}$$

y sea $a \in \mathbb{R}, a > 0$. Entonces:

$$P(Z \geq a) \leq \frac{E(z)}{a}$$

- Demostración

Supongamos Z variable aleatoria discreta, con $\text{Sop}(Z) = \{z_1, z_2, \ldots, z_n, \ldots\}$ y función puntual de probabilidad $p(x)$.

a) $P(z \geq a) = a \cdot \sum\limits_{\substack{x_i \in \text{Sop}(Z) \\ x_i \geq a}} p(x_i) = \sum\limits_{\substack{x_i \in \text{Sop}(Z) \\ x_i \geq a}} a \cdot p(x_i) \underset{\substack{\text{por ser} \\ a \leq x_i}}{\leq} \sum\limits_{\substack{x_i \in \text{Sop}(Z) \\ x_i \geq a}} x_i \cdot p(x_i) \underset{\substack{\text{Hacemos términos} \\ \text{no negativos}}}{\leq} \sum\limits_{x_i \in \text{Sop}(Z)} x_i \cdot p(x_i) = E(Z)$

Nota: Se puede demostrar de forma similar para caso continuo

- Teorema (Desigualdad de Techebycher generalizada)

Sea X variable aleatoria con media finita y momento central de orden n finito (μ_n finito). Sea $\varepsilon > 0$, entonces:

$$\boxed{P\left(|X-\mu \geq \varepsilon\right) \leq \frac{E\left(|X-\mu|^n\right)}{\varepsilon^n}}$$

- Demostración

Tomamos $z = |X-\mu|^n$ (variable aleatoria no negativa) y $a = \varepsilon^n$ en el teorema anterior. Por tanto, aplicando Desigualdad de Markov se tiene:

$$\boxed{P\left(|X-\mu \geq \varepsilon\right) = P\left(|X-\mu|^n \geq \varepsilon^n\right) \leq \frac{E\left(|X-\mu|^n\right)}{\varepsilon^n}}$$

- Teorema (Desigualdad de Techebycher)

Sea $X:\Omega \longrightarrow \mathbb{R}$ variable aleatoria (discreta o continua) con $\mu_X = E(X)$ y $\sigma_X^2 = \text{Var}(X)$ finitas, y sea $k>0$. Entonces:

a) $P(|X-\mu_X| \geq k\cdot\sigma_x) \leq \dfrac{1}{k^2}$

b) $P(|X-\mu_x| < k\cdot\sigma_x) \geq 1-\dfrac{1}{k^2}$

- Demostración

(a) Tomamos $\varepsilon = k\cdot\sigma_X$ en la desigualdad de Techebycher generalizada, y tomamos $\varepsilon>0$ por ser $k>0$ y $\sigma_X>0$ para variable aleatoria no constantes.

$n=2$, se tiene:

$$P(|X-\mu_X| \geq k\cdot\sigma_X) < \frac{E(|X-\mu_X|^2)}{(k\cdot\sigma_X)^2} = \frac{\sigma_X^2}{(k\cdot\sigma_X)^2} = \frac{1}{k^2}$$

(b) Inmediato a partir de (a) porque usa el suceso complementario.

5.8) Otras características de una variable aleatoria

Igual que una variable aleatoria tiene media y varianza, se puede definir su moda, cuantiles, RIC, coeficiente de asimetría y curtosis.

- Definición

Sea $X:\Omega \longrightarrow \mathbb{R}$ variable aleatoria

- Si X es discreta, entonces se define la moda como el punto M_0 donde se alcanza el máximo de la función puntual de probabilidad ,

$$p(M_0) \geq p(x)\ \forall x \in \mathbb{R} \quad p(x) \text{ función puntual de probabilidad}$$

- Si X es continua, la moda es el punto $M_0 \in \mathbb{R}$ donde se alcanza el máximo de la función densidad

$$f(M_0) \geq f(x) \, \forall x \in \mathbb{R}$$

- Definición

Se define la mediana como el punto $M_e \in \mathbb{R}$ tal que

$$\boxed{P(X \leq M_e) \geq 0.5 \text{ y } P(X > M_e) \geq 0.5}$$

- Definición

Se define el cuantil α con $\alpha \in (0, 1)$ al punto $x_\alpha \in \mathbb{R}$ tal que $P(X \leq X_\alpha \geq \alpha)$ y $P(X \geq x) \geq 1 - \alpha$

$$\boxed{Q_1 = \text{cuantil } 0.25 \text{ y } Q_3 = \text{cuantil } 0.75}$$

- Definición

Medida de dispersión:

1) $RIC = Q_3 - Q_1$

2) $CV = \dfrac{\sigma_X}{N_X}$

- Definición

Medidas de forma:

1) Configuración asimétrica $\gamma_1(x) = \dfrac{E(|x - \mu_1|^3)}{\sigma_x^3} = \dfrac{\mu_1^3}{\sigma_x^3}$

2) Configuración urtosis $\gamma_2(x) = \dfrac{E(|x - \mu_2|^4)}{\sigma_x^4} = \dfrac{\mu_2^4}{\sigma_x^4}$

- Anexos

1) Método transformada inversa

Generación de valores aleatorios de una variable aleatoria continua usando el método de la transformada inversa.

Consideremos U la variable aleatoria que consiste en seleccionar un número al azar en el intervalo $(0, 1)$ (veremos que en la distribución Uniforme en $(0, 1)$). U es una variable aleatoria continua con densidad:

$$f_U(t) = \begin{cases} 0 & \text{si } t \leq 0 \\ 1 & \text{si } t \in (0,1) \\ 0 & \text{si } t \geq 1 \end{cases}$$

y función de distribución:

$$F_U(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } x \in [0,1] \\ 1 & \text{si } x \geq 1 \end{cases}$$

Sea $G : (a,b) \longrightarrow \mathbb{R}$ función verificando las propiedades de la función de distribución, en este caso además estrictamente creciente y continua. Entonces $X = G^{-1}(U)$ es una variable aleatoria como vimos en el tema.

- ¿Cómo podemos generar valores aleatorios de la variable aleatoria X?

Observar que $F_X(x) = P(X \leq x) = P(G^{-1}(U) \leq x) = P(U \leq G(x)) = G(x)$, es decir, $G(x)$ es la función de distribución de X. Para generar valores aleatorios de X, bastará con generar valores aleatorios en $(0,1)$ (software) y luego aplicarlas en G^{-1}.

$$\underset{\text{valores aleatorios en } (0,1)}{\{u_1, u_2, \dots\} \in (0,1)} \longrightarrow \underset{\text{valores aleatorios de } X \text{ con distribución } G(x)}{\{G^{-1}(u_1), G^{-1}(u_2), \dots\}}$$

1) Genero valores en $(0,1) \longrightarrow$ valores $\{u_1, u_2, \dots, u_m, \dots\}$

2) Calculo $F^{-1}(u_i)$ donde F es función distribución de una variable aleatoria X. Estos son valores aleatorios de X.

2) Funciones Gamma y Beta de Euler

2.1) Función Gamma de Euler

La función gamma (Γ) se define en el intervalo $(0, +\infty)$ por:

$$\boxed{\Gamma(p) = \int_0^{+\infty} e^{-t} \cdot t^{p-1} \, \mathrm{dt}} \text{ para } p > 0$$

- Propiedades

1) $\Gamma(p+1) = p\,(\neg(p)) \;\; \forall p > 0$

2) En particular si $n \in \mathbb{Z}^+ \longrightarrow \Gamma(n+1) = n!$

3) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

2.2) Función Beta de Euler

La función beta (β) se define $\boxed{\forall p > 0, q > 0}$ por:

$$\boxed{\beta(p,q) = \int_0^1 t^{p-1} \cdot (1-t)^{q-1}\,\mathrm{d}t}$$

Si $p > 0$ y $q > 0 \longrightarrow \beta(p,q)$ existe y tiene valor finito:

- Propiedades

1) La función beta es simétrica en sus variables, es decir:

$$\beta(p,q) = \beta(q,p)$$

2) La función beta se puede escribir de la forma:

$$\beta(p,q) = \int_0^{+\infty} \frac{x^{p-1}}{(x+1)^{p+q}}\,\mathrm{d}x \quad (p > 0 \text{ y } q > 0)$$

3) Relación entre beta y gamma

$$\forall p, q > 0 \quad \boxed{\beta(p,q) = \frac{\Gamma(p)\cdot\Gamma(q)}{\Gamma(p+q)}}$$

PROBLEMAS. RELACIÓN 5: Variable aleatoria de una dimensión.

FUNDAMENTOS DE PROBABILIDAD Y ANÁLISIS EXPLORATORIO DE DATOS.

GRADO EN CIENCIA E INGENIERÍA DE DATOS.

1. Dada la variable aleatoria *X*, cuya distribución de probabilidad es:

$X = x_i$	1	2	3	4	5
$P(X = x_i)$	2/8	1/8	2/8	2/8	1/8

Determinar:

a) La representación gráfica de la función puntual de probabilidad de la variable aleatoria *X*.

b) La función de distribución de la variable aleatoria *X*.

c) La probabilidad $P(1 < X \leq 2{,}7)$.

d) La probabilidad $P(1 \leq X < 3{,}5)$.

2. Dada la variable aleatoria *X*, cuya función de distribución *F(x)* viene definida por:

$$F(x) = \begin{cases} 0 & si \quad x < 0 \\ 1/4 & si \quad 0 \leq x < 1 \\ 2/4 & si \quad 1 \leq x < 2 \\ 3/4 & si \quad 2 \leq x < 3 \\ 1 & si \quad 3 \leq x \end{cases}$$

Determinar: a) La representación gráfica de *F(x)*.

b) La función puntual de probabilidad que engendra *F(x)*.

c) Las probabilidades: $P(X = 1{,}7)$, $P(X = 2)$ *y* $P(1{,}2 < X < 3)$.

3. Sea X una v.a. cuya función de probabilidad viene dada por

$$P(X = k) = \frac{k}{6} \qquad \text{para} \quad k = 1, 2, 3.$$

a) Calcular la esperanza y la varianza de la v.a. *X*.
b) Calcular la moda y la mediana de la v.a. X.

4. Dada la variable aleatoria X tal que $P(X=r)=\begin{cases} k\dfrac{r-1}{n} & si \quad r=2,3,\dots,n \\ 0 & en\ otro\ caso \end{cases}$ determinar k para que $P(X=r)$ sea la función de masa de probabilidad que define la distribución de probabilidad de la variable aleatoria *X*.

5. Dada la variable aleatoria *X*, cuya distribución de probabilidad viene definida por la función de densidad:

$$f(x)=\begin{cases} 0 & si \quad x\le 0 \\ \dfrac{1}{9}(x+1) & si \quad 0<x\le 1 \\ \dfrac{4}{9}\left(x-\dfrac{1}{2}\right) & si \quad 1<x\le \dfrac{3}{2} \\ \dfrac{4}{9}\left(\dfrac{5}{2}-x\right) & si \quad \dfrac{3}{2}<x\le 2 \\ \dfrac{1}{9}(4-x) & si \quad 2<x\le 3 \\ \dfrac{1}{9} & si \quad 3<x\le 6 \\ 0 & si \quad 6<x \end{cases}$$

a) Hacer la representación gráfica de la función de densidad *f(x)*.

b) Calcular la función de distribución de la variable aleatoria *X*.

c) Calcular la probabilidad $P(1{,}3<X<2{,}4)$.

6. Dada la variable aleatoria *X*, cuya distribución de probabilidad viene definida por la función de densidad:

$$f(x)=\begin{cases} 2x & para \quad 0<x<1 \\ 0 & para \quad cualquier \quad otro \quad valor \quad de \quad x \end{cases}$$

a) Determinar la función de distribución correspondiente a la variable aleatoria *X*.

b) Calcular las probabilidades: $P(X=0{,}75)$, $P(-1<X\le 0{,}5)$, $P(0{,}3\le X\le 0{,}8)$.

c) Calcular la moda y la mediana de la v.a. *X*.

7. Dada una variable aleatoria *X* cuya distribución de probabilidad viene definida por la función de densidad:

$$f(x)=\begin{cases} kx^2 & para \quad 0<x<1 \\ 0 & para \quad cualquier \quad otro \quad valor \quad de \quad x \end{cases}$$

a) Calcular el valor de la constante *k*, para que *f*(*x*) sea la función de densidad que define la distribución de probabilidad de la variable aleatoria *X*.
b) Determinar la función de distribución de la variable aleatoria *X*.

c) Calcular la esperanza y la varianza de la v.a. *X*.

8. Dada la variable aleatoria *X*, cuya distribución de probabilidad viene definida por la función de distribución:

$$F(x)=\begin{cases}0 & para \quad x<0\\ x^2 & para \quad 0\le x<1.\\ 1 & para \quad 1\le x\end{cases}$$

Determinar la esperanza y la varianza de *X*.

9. La distribución del número de veces que es solicitado el servicio de una hormigonera durante la cimentación tiene como función de densidad

$$f(x)=k\cdot e^{-ax} \quad \text{para} \quad x\ge 0.$$

a) ¿Cuál es el valor de *k* suponiendo que *a* es un parámetro conocido en la fábrica?
b) ¿Cuánto vale su media y su varianza?

10. El porcentaje real de zumo de naranja que poseen los refrescos con gas fabricados por una determinada compañía es una variable aleatoria con la siguiente función de densidad:

$$f(x)=\begin{cases}\dfrac{3}{4} & \text{si} \quad a\le x<7.5\\ 1-(8-x)^2 & \text{si} \quad 7.5\le x<8.5\\ 0 & \text{en el resto}\end{cases} \quad \text{siendo } a \text{ una cantidad constante.}$$

a) Obtener el porcentaje mínimo de zumo de naranja que poseen estos refrescos.

b) Determinar la probabilidad de que un bote de estos refrescos, elegido al azar, tenga al menos un 8% de zumo de naranja.

c) Determinar el porcentaje de refrescos cuyo contenido en zumo de naranja oscila entre un 7% y un 8%.

11. Se sabe que el contenido depositado por una máquina expendedora de café (capuchino) es una variable aleatoria de media 150 ml y desviación típica 5 ml.

a) Se desea determinar la probabilidad de que un vaso llenado con dicha máquina contenga entre 140 ml y 160 ml. Con la información proporcionada, ¿podemos obtener dicha probabilidad? En caso negativo, indicar si se puede obtener una cota de la probabilidad solicitada y calcularla.
b) ¿Y qué podemos decir de la probabilidad de que un vaso contenga entre 135 ml y 160 ml?
c) Obtener un intervalo (a, b) centrado en la media (150 ml), tal que al menos el 90% de los vasos expendidos tengan su contenido entre dichos valores (a, b).
d) Si la máquina deposita 200 ml o más, el vaso rebosa. ¿Podemos garantizar que el porcentaje de vasos que rebosan no superará el 80%?.

12. Sea *X* una variable aleatoria discreta que toma los valores $\text{i} = -2, -1, 0, 1, 2$ con $p_i = 1/5$. Calcular la función de probabilidad y la función de distribución de la variable aleatoria $Y = X^2$. Calcular $E(Y)$ y $Var(Y)$.

13. Sea *X* una variable aleatoria continua con función de distribución

$$F(x)=\begin{cases}0 & \text{para} \quad x<0\\ x & \text{para} \quad 0\le x<1\\ 1 & \text{para} \quad 1\le x\end{cases}$$

Calcular la función de densidad y la función de distribución de la variable aleatoria $Y = X^2$. Calcular $E(Y)$ y $Var(Y)$.

14. Sea X una variable aleatoria continua cuya función de densidad viene dada por

$$f(x) = \begin{cases} 2x & \text{para} \quad 0 < x < 1 \\ 0 & \text{para el resto} \end{cases}$$

Calcular la función de densidad y la función de distribución de las variables aleatorias:

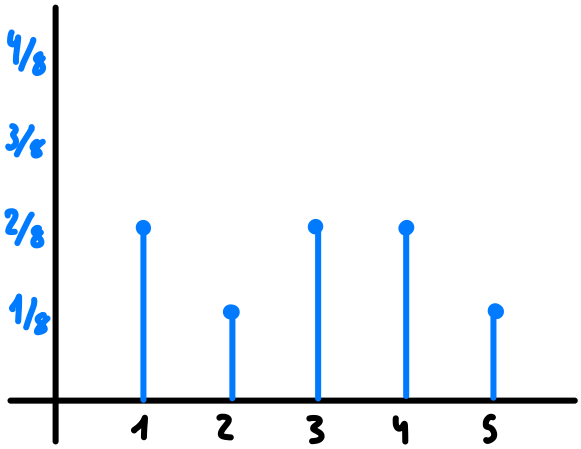
a) $Y = 3X + 5$.
b) $Y = X^2$.
c) $Y = 2X^2 + 5$.

1. Dada la variable aleatoria X, cuya distribución de probabilidad es:

$X = x_i$	1	2	3	4	5
$P(X = x_i)$	2/8	1/8	2/8	2/8	1/8

Determinar:

a) La representación gráfica de la función puntual de probabilidad de la variable aleatoria X.



b) La función de distribución de la variable aleatoria X.

$$F(x) = \begin{cases} 0 & \text{si } x < 1 \\ 2/8 = 1/4 & \text{si } x \in [1,2) \\ 3/8 & \text{si } x \in [2,3) \\ 5/8 & \text{si } x \in [3,4) \\ 7/8 & \text{si } x \in [4,5) \\ 1 & \text{si } x \geq 5 \end{cases}$$

c) La probabilidad $P(1 < X \leq 2{,}7)$.

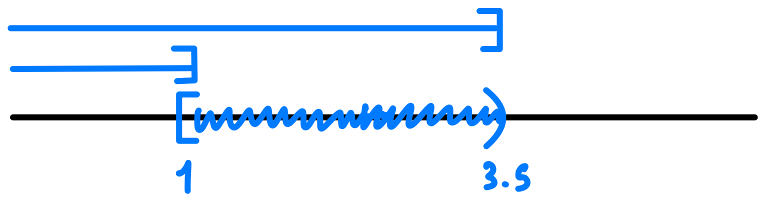
$$P(1 < X \leq 2.7) = \sum_{\substack{x_i \in Sop(X) \\ x_i \in (1,\, 2.7]}} p(x_i) = P(2) = \frac{1}{8}$$

$$\longrightarrow F(2.7) - F(1) = \frac{3}{8} - \frac{2}{8} = \frac{1}{8}$$

d) La probabilidad $P(1 \leq X < 3{,}5)$.

$P(1 \leq x < 3.5) = p(1) + p(2) + p(3) = \frac{2}{8} + \frac{1}{8} + \frac{2}{8} = \frac{5}{8}$

$\hookrightarrow F(3.5^-) - F(1^-) = \frac{5}{8} - 0 = \frac{5}{8}$



2. Dada la variable aleatoria X, cuya función de distribución $F(x)$ viene definida por:

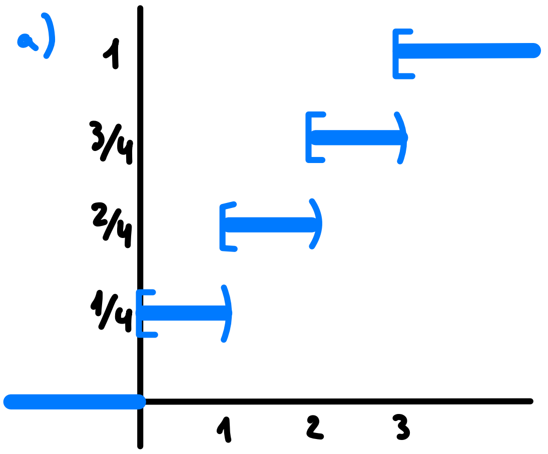
$$F(x) = \begin{cases} 0 & si \quad x < 0 \\ 1/4 & si \quad 0 \leq x < 1 \\ 2/4 & si \quad 1 \leq x < 2 \\ 3/4 & si \quad 2 \leq x < 3 \\ 1 & si \quad 3 \leq x \end{cases}$$

Determinar: a) La representación gráfica de $F(x)$.

b) La función puntual de probabilidad que engendra $F(x)$.

c) Las probabilidades: $P(X = 1{,}7)$, $P(X = 2)$ y $P(1{,}2 < X < 3)$.

a)



b) Sabemos que es v.a discreta

¿Sop(X)? Sop$(X) = \{0, 1, 2, 3\}$ saltos de $F(x)$

• f.p.p de X, $p(0) = p(X=0) = 1/4$

$p(1) = p(X=1) = 2/4 - 1/4 = 1/4$

$p(2) = p(X=2) = 3/4 - 2/4 = 1/4$

$p(3) = p(X=3) = 1 - 3/4 = 1/4$

$p(x) = 0 \quad \forall x \notin \text{Sop}(X)$

3. Sea X una v.a. cuya función de probabilidad viene dada por

$$P(X = k) = \frac{k}{6} \qquad \text{para} \quad k = 1, 2, 3.$$

a) Calcular la esperanza y la varianza de la v.a. X.

$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{2}{6} + 3 \cdot \frac{3}{6} = \frac{14}{6} = \boxed{\frac{7}{3}}$

$Var(X) = E(X^2) - (E(X))^2 = 6 - \left(\frac{7}{3}\right)^2 = \frac{54-49}{9} = \boxed{\frac{5}{9}}$

$E(X^2) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{2}{6} + 3^2 \cdot \frac{3}{6} = 6$

b) Calcular la moda y la mediana de la v.a. X.

$P(X=1) = \frac{1}{6}$

$P(X=2) = \frac{1}{3}$

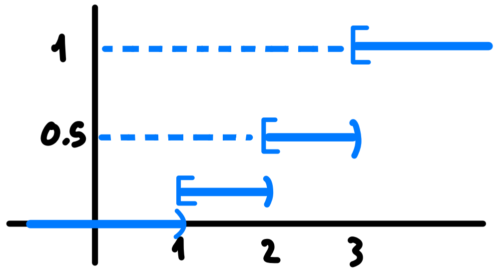
$P(X=3) = \frac{1}{2}$

Moda $(X) = 3$ porque es el punto con mayor probabilidad

¿Mediana $(X) = Me$?

$P(X \leq Me) \geq 0.5$

$P(X \geq Me) \geq 0.5$

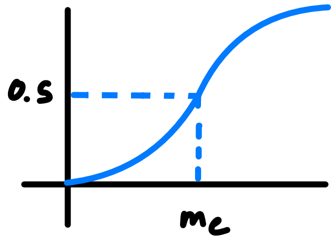


0 si $x < 1$

$1/6$ si $x \in [1, 2)$

$1/6 + 1/3 = 1/2$ si $x \in [2, 3)$

$\frac{1}{6} + \frac{1}{3} + \frac{1}{2} = 1$ si $x \in [3, +\infty)$



Observar que $P(X \leq 2) = p(1) + p(2) = \frac{1}{2}$

$P(X \geq 2) = p(2) + p(3) = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}$

¿Es $X = 2$ mediana de X? Si

¿Es $X = 3$ mediana de X? Si

$P(X \leq 3) = 1$

$P(X \geq 3) = \frac{1}{2}$

$\boxed{[2, 3]}$ intervalo mediana

$P(X \leq 2.7) = \frac{1}{2}$

$P(X \geq 2.7) = P(3) = \frac{1}{2}$

4. Dada la variable aleatoria X tal que $P(X = r) = \begin{cases} k\dfrac{r-1}{n} & si \quad r = 2,3,\ldots,n \\ 0 & en\ otro\ caso \end{cases}$ determinar k para que $P(X = r)$ sea la función de masa de probabilidad que define la distribución de probabilidad de la variable aleatoria X.

X es una v.a discreta puesto que la probabilidad se halla concentrada en un nº finito de putos: $2, 3, \ldots, n$

Para que $P(X=r)$ sea una función de probabilidad se tiene que cumplir:

a) $P(X=r) \geq 0 \quad \forall r$

b) $\sum_r P(X=r) = 1$

Por lo tanto

$$\sum_r P(X=r) = \sum_{r=2}^{n} k\,\frac{r-1}{n} = \frac{k}{n}\sum_{r=2}^{n}(r-1) = \frac{k}{n}\underbrace{(1+2+\cdots+(n-1))}_{} =$$

$$\frac{k}{n}\cdot\frac{(\cancel{1}+(n-\cancel{1}))(n-1)}{2} = \frac{k}{\cancel{n}}\cdot\frac{\cancel{n}(n-1)}{2}$$ s= suma de un pogresión aritmética

$$= \frac{k(n-1)}{2} = 1 \longrightarrow \boxed{k = \frac{2}{n-1}}$$

$$s = \frac{(1^{\circ} + \text{último})\cdot(\text{n}^{\circ}\text{ de término})}{2}$$

Además como $k = \dfrac{2}{n-1} > 0$, la $P(X=r) \geq 0 \quad \forall r$.

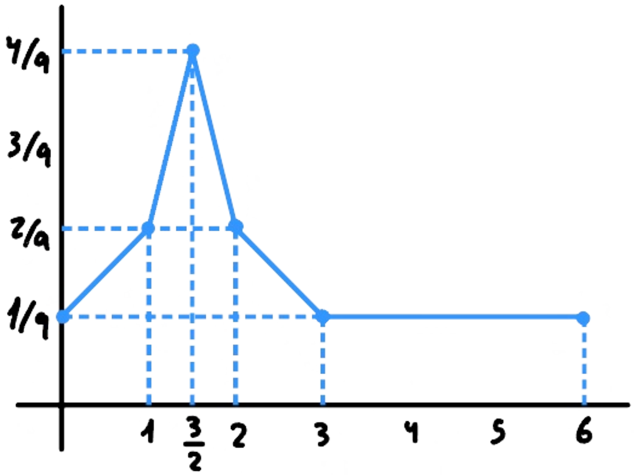
5. Dada la variable aleatoria X, cuya distribución de probabilidad viene definida por la función de densidad:

$$f(x) = \begin{cases} 0 & si \quad x \leq 0 \\ \frac{1}{9}(x+1) & si \quad 0 < x \leq 1 \\ \frac{4}{9}\left(x - \frac{1}{2}\right) & si \quad 1 < x \leq \frac{3}{2} \\ \frac{4}{9}\left(\frac{5}{2} - x\right) & si \quad \frac{3}{2} < x \leq 2 \\ \frac{1}{9}(4-x) & si \quad 2 < x \leq 3 \\ \frac{1}{9} & si \quad 3 < x \leq 6 \\ 0 & si \quad 6 < x \end{cases}$$

a) Hacer la representación gráfica de la función de densidad $f(x)$.

b) Calcular la función de distribución de la variable aleatoria X.

c) Calcular la probabilidad $P(1{,}3 < X < 2{,}4)$.

a) 

c) $P(1.3 < X < 2.4) =$

$F(2.4) - F(1.3) =$

$$\int_{1.3}^{2.4} f(t)\,dt = \int_{1.3}^{3/2} \frac{4}{9}\left(t - \frac{1}{2}\right) dt + \int_{3/2}^{2} \frac{4}{9}\left(\frac{5}{2} - t\right) dt + \int_{2}^{2.4} \frac{1}{9}(4-t)\,dt$$

b) Calcular $F(x) = P(X = x)$

- Si $x \leq 0 \rightarrow F(x) = \int_{-\infty}^{x} 0\,dt = 0$
- Si $x \in (0, 1] \rightarrow F(x) = \int_{0}^{x} \frac{1}{9}(t+1)\,dt = \frac{1}{9}\left(\frac{x^2}{2} + x\right)$
- Si $x \in (1, 3/2] \longrightarrow F(x) = \int_{0}^{1} \frac{1}{9}(t+1)\,dt + \int_{1}^{x} \frac{4}{9}\left(t - \frac{1}{2}\right) =$

  $\frac{1}{3} + \frac{2}{9}(x-1)^2$
- Si $x \in (3/2, 2] \longrightarrow F(x) = \int_{0}^{1} \frac{1}{9}(t+1)\,dt + \int_{1}^{3/2} \frac{4}{9}\left(t - \frac{1}{2}\right) + \int_{3/2}^{x} \frac{4}{9}\left(\frac{5}{2} - t\right) dt$

  $= \frac{5}{9} + \frac{4}{9}\left(x - \frac{3}{2}\right) - \frac{2}{9}\left(x - \frac{3}{2}\right)^2$
- Si $x \in (2, 3] \rightarrow F(x) = \int_{0}^{1} \frac{1}{9}(t+1)\,dt + \int_{1}^{3/2} \frac{4}{9}\left(t - \frac{1}{2}\right) + \int_{3/2}^{2} \frac{4}{9}\left(\frac{5}{2} - t\right) dt +$

  $\int_{2}^{x} \frac{1}{9}(4-t)\,dt = \frac{14}{27} + \frac{1}{18}(x-2)^2$

- Si $x \in (3, 6] \longrightarrow F(x) = \int_{0}^{1} \frac{1}{9}(t+1)\,dt + \int_{1}^{3/2} \frac{4}{9}\left(t - \frac{1}{2}\right) + \int_{3/2}^{2} \frac{4}{9}\left(\frac{5}{2} - t\right) dt +$

  $\int_{2}^{3} \frac{1}{9}(4-t)\,dt + \int_{3}^{x} \frac{1}{9} = \frac{5}{6} + \frac{1}{9}(x-3)$

Juntando todas las piezas obtenidas en cada intervalo, se obtiene la función de de distribución acumulada para la variable aleatoria x:

$$F(x)=\begin{cases} 0 & \text{si } x\leq 0 \\ \frac{1}{9}\left(\frac{x^2}{2}+x\right) & \text{si } 0<x\leq 1 \\ \frac{1}{3}+\frac{2}{9}(x-1)^2 & \text{si } 1<x\leq \frac{3}{2} \\ \frac{5}{9}+\frac{4}{9}\left(x-\frac{3}{2}\right)-\frac{2}{9}\left(x-\frac{3}{2}\right)^2 & \text{si } \frac{3}{2}<x\leq 2 \\ \frac{14}{27}+\frac{1}{18}(x-2)^2 & \text{si } 2<x\leq 3 \\ \frac{5}{6}+\frac{1}{9}(x-3) & \text{si } 3<x\leq 6 \\ 1 & \text{si } x>6 \end{cases}$$

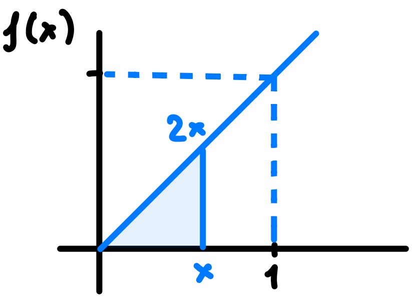
6. Dada la variable aleatoria X, cuya distribución de probabilidad viene definida por la función de densidad:

$$f(x)=\begin{cases} 2x & para \quad 0<x<1 \\ 0 & para \quad cualquier \quad otro \quad valor \quad de \quad x \end{cases}$$

a) Determinar la función de distribución correspondiente a la variable aleatoria X.

b) Calcular las probabilidades: $P(X=0{,}75)$, $P(-1<X\leq 0{,}5)$, $P(0{,}3\leq X\leq 0{,}8)$.

c) Calcular la moda y la mediana de la v.a. X.



a) ¿F_X? $F_X(x)=\int_{-\infty}^{x} f(t)\,dt$

Si $x\leq 0 \longrightarrow F_X(x)=0$

Si $x\in(0,1) \longrightarrow F_X(x)=\int_0^x 2t\,dt=[t^2]_0^x=x^2$

Si $x\geq 1 \longrightarrow F_X(x)=\int_0^1 2t\,dt=[t^2]_0^1=1$

$$F(x)=\begin{cases} 0 & \text{si } x\leq 0 \\ x^2 & \text{si } x\in(0,1) \\ 1 & \text{si } x\geq 1 \end{cases}$$

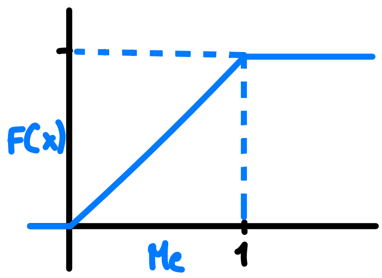
b) $P(X=0.75)=0$

$P(-1\leq X\leq 0.5)=F(0.5)-F(-1)=(0.5)^2-0=0.25$

$P(0.3\leq X\leq 0.8)=F(0.8)-F(0.3)=(0.8)^2-(0.3)^2=0.55$

c) Moda(X). Como el máximo $f(x)$ se alcanza en $x=1$, entonces Moda (X) = $M_o = 1$

¿Mediana (X)? $F(Me) = 0.5 \longrightarrow$ Es el punto "x" único que cumple $x^2 = 0.5 = \frac{1}{2} \longrightarrow \boxed{x = \frac{1}{\sqrt{2}}}$



7. Dada una variable aleatoria X cuya distribución de probabilidad viene definida por la función de densidad:

$$f(x) = \begin{cases} kx^2 & para \quad 0 < x < 1 \\ 0 & para \quad cualquier \quad otro \quad valor \quad de \quad x \end{cases}$$

X v.a continua

$$f(x) = \begin{cases} kx^2 & 0 < x < 1 \\ 0 & \text{resto} \end{cases}$$

a) Calcular el valor de la constante k, para que $f(x)$ sea la función de densidad que define la distribución de probabilidad de la variable aleatoria X.

¿$f(x) \geq 0 \quad \forall x \in \mathbb{R}$?

¿ $\int_{-\infty}^{+\infty} f(t)\,dt = 1$?

Sí, siempre que $k>0$

$$\int_{-\infty}^{+\infty} f(t)\,dt = \int_0^1 k \cdot t^2 \cdot dt = k \cdot \left[\frac{t^3}{3}\right]_0^1 = k \cdot \frac{1}{3}$$

$k \cdot \frac{1}{3} = 1 \longrightarrow \boxed{k=3}$

b) Determinar la función de distribución de la variable aleatoria X.

¿F_X? $F_X(x) = P(X \leq x)$

$$\begin{cases} \bullet \text{ Si } x \leq 0 \longrightarrow F(x) = 0 \\ \bullet \text{ Si } x \in [0,1] \longrightarrow F(x) = \int_0^x 3t^2\,dt = \left[3\frac{t^3}{3}\right]_0^x = x^3 \\ \bullet \text{ Si } x \geq 1 \longrightarrow F(x) = 1 \end{cases}$$

$$F_X(x) = \begin{cases} 0 & si \quad x \leq 0 \\ x^3 & si \quad x \in [0,1] \\ 1 & si \quad x \geq 1 \end{cases}$$

c) Calcular la esperanza y la varianza de la v.a. X.

$$E(X) = \int_{-\infty}^{+\infty} t \cdot f(t)\,dt = \int_0^1 t \cdot 3t^2 dt = 3\left[\frac{t^4}{4}\right]_0^1 = \frac{3}{4}$$

$$Var(X) = E(X^2) - (E(X))^2 = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = \boxed{\frac{3}{80}}$$

$$E(X^2) = \int_{-\infty}^{+\infty} t^2 \, f(t)\,dt = \int_0^1 t^2 \cdot 3t^2 \, dt = 3\left[\frac{t^5}{5}\right]_0^1 = \frac{3}{5}$$

8. Dada la variable aleatoria X, cuya distribución de probabilidad viene definida por la función de distribución:

$$F(x) = \begin{cases} 0 & para \quad x < 0 \\ x^2 & para \quad 0 \leq x < 1 . \\ 1 & para \quad 1 \leq x \end{cases}$$

Determinar la esperanza y la varianza de X.

$$F'(x) = f(x) = \begin{cases} 2x & si \quad x \in (0,1) \\ 0 & resto \end{cases}$$

$$E(X) = \int_{-\infty}^{+\infty} t \cdot f(t)\,dt = \int_0^1 t \cdot 2t\,dt = 2\left[\frac{t^3}{3}\right]_0^1 = \frac{2}{3}$$

$$Var(X) = E(X^2) - (E(X))^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

$$E(X^2) = \int_{-\infty}^{+\infty} t^2 \cdot f(t)\,dt = \int_0^1 t^2 \cdot 2t\,dt = 2\left[\frac{t^4}{4}\right]_0^1 = \frac{1}{2}$$

9. La distribución del número de veces que es solicitado el servicio de una hormigonera durante la cimentación tiene como función de densidad

$$f(x) = k \cdot e^{-ax} \quad \text{para} \quad x \geq 0 .$$

a) ¿Cuál es el valor de k suponiendo que a es un parámetro conocido en la fábrica?
b) ¿Cuánto vale su media y su varianza?

X = "Nº veces que es solicitado el servicio de una hormigonera durante la cimentación.

$$X \sim f(x) = \begin{cases} k \cdot e^{-ax} & \text{si } x \geq 0 \\ 0 & \text{en el resto} \end{cases}$$

a) Si a = parámetro conocido k

$$\int_0^{+\infty} f(x)\,dx = 1$$

$$\int_0^{+\infty} k \cdot e^{-ax}\,dx = k \int_0^{+\infty} e^{-ax}\,dx = \left\{ \begin{array}{l} ax = t \\ a\,dx = dt \\ x=0 \to t=0 \\ x=+\infty \to t=+\infty \end{array} \right\} = k \int_0^{+\infty} e^{-t}\,\frac{dt}{a} =$$

$$k\,\frac{1}{a}\int_0^{+\infty} e^{-t}\,dt = k \cdot \frac{1}{a}\left[\frac{e^{-t}}{-1}\right]_0^{+\infty} = k \cdot \frac{1}{a}\left[\cancel{-e^{-\infty}}^{\,0} + \cancel{e^{-0}}^{\,1}\right] = \frac{k}{a} = 1$$

$$\longrightarrow \boxed{k = a}$$

b) $E(X)$ y $Var(X)$?

- $E(X) = \int_0^{+\infty} x\, f(x)\,dx = \int_0^{+\infty} x\, a e^{-ax}\,dx = a\int_0^{+\infty} x \cdot e^{-ax}\,dx =$

$$\left\{ \begin{array}{ll} u = x & du = dx \\ dv = x e^{-ax} & v = -\frac{1}{a} \cdot e^{-ax} \end{array} \right\} = a \cdot \left(\left[x \cdot \left(-\frac{1}{a} \cdot e^{-ax}\right)\right]_0^{+\infty} + \int_0^{+\infty} \frac{1}{a}\, e^{-ax}\,dx \right) =$$

$$a \cdot \left(+\infty \left(-\frac{1}{a}\, e^{-\infty}\right) - 0 \cdot \left(-\frac{1}{a}\, e^{-0}\right) + \frac{1}{a}\int_0^{+\infty} e^{-ax}\,dx \right) =$$

$$\lim_{t \to +\infty} a\left(t\left(-\frac{1}{a}\right) e^{-at}\right) + \frac{a}{a^2} = \lim_{t\to+\infty} \frac{-\cancel{a}\,t}{\cancel{a} \cdot e^{at}} + \frac{1}{a} = \left(\frac{\infty}{\infty}\right) \xrightarrow{\text{L'Hôpital}}$$

$$\lim_{t\to+\infty} \cancel{\frac{-1}{a\, e^{at}}}^{\,0} + \frac{1}{a} = \frac{1}{a}$$

Esta integral se puede hacer de otra forma, si conocemos la función gamma.

$$\Gamma(p) = \int_0^{+\infty} e^{-t} \cdot t^{p-1}\,dt, \quad p>0 \longrightarrow \Gamma(n) = (n-1)!$$

En nuestro caso:

$$E(X) = \int_0^{+\infty} x \cdot a \cdot e^{-ax}\,dx = a\int_0^{+\infty} x \cdot e^{-ax}\,dx = \left\{ \begin{array}{l} ax = t \\ a\,dx = dt \\ x=0 \to t=0 \\ x=+\infty \to t=+\infty \end{array} \right\} =$$

$$a\cdot\int_0^{+\infty}\frac{t}{a}\cdot e^{-t}\cdot\frac{dt}{a}=\frac{\cancel{a}}{a^{\cancel{2}}}\cdot\int_0^{+\infty}t\cdot e^{-t}\,dt=\frac{1}{a}\cdot\underbrace{\int_0^{+\infty}t\cdot e^{-t}\,dt}_{\Gamma(2)=1!=1}=$$

$p-1=1$

$\boxed{p=2}$

$$\frac{1}{a}\cdot 1=\boxed{\frac{1}{a}}$$

- $Var(X)=E(X^2)-(E(X))^2=(*)$

$$E(X^2)=\int_0^{+\infty}x^2\cdot a\cdot e^{-ax}\,dx=a\int_0^{+\infty}x^2\cdot e^{-ax}\,dx=\left\{\begin{array}{l}ax=t\\ a\,dx=dt\\ x=0\to t=0\\ x=+\infty\to t=+\infty\end{array}\right\}=$$

$$\cancel{a}\int_0^{+\infty}\left(\frac{t}{a}\right)^2e^{-t}\frac{dt}{\cancel{a}}=\frac{1}{a^2}\underbrace{\int_0^{+\infty}t^2e^{-t}\,dt}_{\Gamma(3)=2!=4}=$$

$p-1=2$

$p=3$

$$\frac{1}{a^2}\cdot 4=\frac{2}{a^2}$$

$$(*)=\frac{4}{a^2}-\left(\frac{1}{a}\right)^2=\frac{4}{a^2}-\frac{1}{a^2}=\boxed{\frac{3}{a^2}}$$

10. El porcentaje real de zumo de naranja que poseen los refrescos con gas fabricados por una determinada compañía es una variable aleatoria con la siguiente función de densidad:

$$f(x)=\begin{cases}\frac{3}{4} & \text{si} \quad a\le x<7.5\\ 1-(8-x)^2 & \text{si} \quad 7.5\le x<8.5\\ 0 & \text{en el resto}\end{cases}\quad \text{siendo } a \text{ una cantidad constante.}$$

a) Obtener el porcentaje mínimo de zumo de naranja que poseen estos refrescos.

b) Determinar la probabilidad de que un bote de estos refrescos, elegido al azar, tenga al menos un 8% de zumo de naranja.

a) Me piden calcular "a"

1) ¿$f(x)\ge 0\ \forall x\in\mathbb{R}$? Si

2) $\int_{-\infty}^{+\infty}f(x)\,dx=1$

$$\int_{-\infty}^{+\infty}f(t)\,dt=\int_{-\infty}^{a}0\,dt+\int_a^{7.5}\frac{3}{4}\,dt+\int_{7.5}^{8.5}(1-(8-t)^2)\,dt+\int_{8.5}^{+\infty}0\,dt$$

$$=\frac{3}{4}\left[t\right]_a^{7.5}+\left[0+\frac{(8-t)^3}{3}\right]_{7.5}^{8.5}=\frac{3}{4}(7.5-a)+8.5+\frac{(-0.5)^3}{3}-\left(7.5+\frac{(0.5)^3}{3}\right)=$$

$$\frac{157-18a}{24}=1\longrightarrow a=\frac{133}{18}=7.39\,\%$$

b) $P(X \geq 8) = 1 - F(8)$

$$P(X \geq 8) = \int_{8}^{+\infty} f(t)\,dt = \int_{8}^{8.5} 1-(8-t)^2\,dt = \frac{11}{24} = 0.4583$$

c) Me piden $P(7 \leq X \leq 8) = F(8) - F(7) = 1 - F(X \geq 8) - 0 =$

$$1 - \frac{11}{24} = \frac{13}{24} = 0.5417$$

$$P(X \leq 7) = \int_{-\infty}^{7} 0\,dt = 0$$

$$P(7 \leq X \leq 8) = \int_{7}^{7.39} 0\,dt + \int_{7.39}^{7.5} \frac{3}{4}\,dt + \int_{7.5}^{8} 1-(8-t)^2\,dt$$

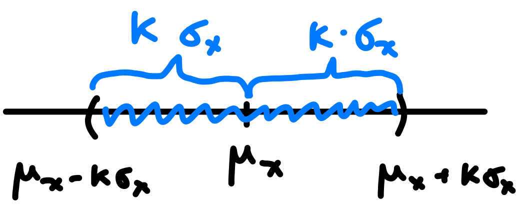
11. Se sabe que el contenido depositado por una máquina expendedora de café (capuchino) es una variable aleatoria de media 150 ml y desviación típica 5 ml.

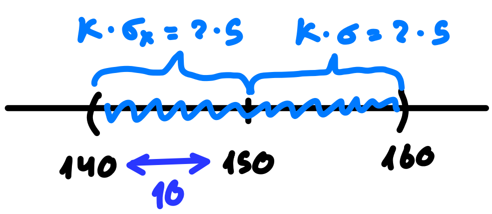
a) Se desea determinar la probabilidad de que un vaso llenado con dicha máquina contenga entre 140 ml y 160 ml. Con la información proporcionada, ¿podemos obtener dicha probabilidad? En caso negativo, indicar si se puede obtener una cota de la probabilidad solicitada y calcularla.

X: "Contenido que provee la máquina de cafe" v.a

$E(X) = 150$ ml $\quad \sigma_X = 5$ ml

Me piden $P(140 \leq x \leq 160) \rightarrow$ No puedo obtenerla porque no controlo la probabilidad de X. Si puedo obtener una cota





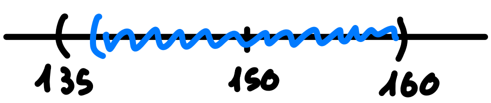
Me piden $P(140 < X < 160) = P(X \in (140, 160)) =$

$$P(|X - 150| < \overset{k\ \sigma_X}{2 \cdot 5}) \geq 1 - \frac{1}{2^2} = 0.75$$

b) ¿Y qué podemos decir de la probabilidad de que un vaso contenga entre 135 ml y 160 ml?

Aplico Tchebycher con $K = 2$

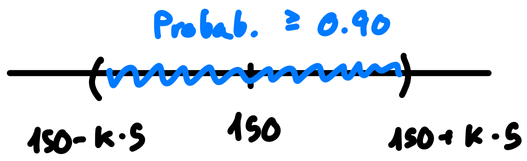
¿ $P(135 < X < 160)$?



$$P(135 < X < 160) \geq P(140 < X < 160) \geq 0.75$$

c) Obtener un intervalo (a, b) centrado en la media (150 ml), tal que al menos el 90% de los vasos expendidos tengan su contenido entre dichos valores (a, b).

Determinar (a, b) tal que $P(a < X < b) \geq 0.90$



Aplico Tchebychev (b) con $1 - \frac{1}{k^2} = 0.90$

$\longrightarrow \frac{1}{k^2} = 0.10 \longrightarrow k = \sqrt{10}$

El intervalo: $(150 - \sqrt{10} \cdot S,\ 150 + \sqrt{10} \cdot S)$

d) Si la máquina deposita 200 ml o más, el vaso rebosa. ¿Podemos garantizar que el porcentaje de vasos que rebosan no superará el 80%?.

¿$P(X > 200) \leq 0.8$? Sí es cierto.

<u>Desigualdad de Markov:</u> $P(Z \geq a) \leq \frac{E(Z)}{a}$ $\quad a > 0$, Z v.a no negativa

$$P(X \geq 200) \leq \frac{E(X)}{200} = \frac{150}{200} = \frac{3}{4} = 0.75$$

12. Sea X una variable aleatoria discreta que toma los valores $i = -2, -1, 0, 1, 2$ con $p_i = 1/5$. Calcular la función de probabilidad y la función de distribución de la variable aleatoria $Y = X^2$. Calcular $E(Y)$ y $Var(Y)$.

Calcular f.p.p de Y, p_Y

$Sop(Y) = \{0, 1, 4\}$

$p_Y(0) = P(Y=0) = P(X^2 = 0) = P(X = 0) = \frac{1}{5}$

$p_Y(1) = P(Y=1) = P(X^2 = 1) = P(X=1) + (P X = -1) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$

$p_Y(4) = P(Y=4) = P(X^2 = 4) = P(X=2) + (P X = -2) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$

$p_Y(y) = 0 \quad \forall y \notin \{0, 1, 4\}$

$$F_Y(y) = \begin{cases} 0 & \text{si } y < 0 \\ 1/5 & \text{si } y \in [0, 1) \\ 3/5 & \text{si } y \in [1, 4) \\ 1 & \text{si } y \geq 4 \end{cases}$$

$E(Y) = 0 \cdot \frac{1}{5} + 1 \cdot \frac{2}{5} + 4 \cdot \frac{2}{5} = 2$

$Var(Y) = E(Y^2) - (E(Y))^2 = \frac{34}{5} - 2^2 = \frac{14}{5}$

$E(Y^2) = 0^2 \frac{1}{5} + 1^2 \cdot \frac{2}{5} + 4^2 \cdot \frac{2}{5} = \frac{34}{5}$

13. Sea X una variable aleatoria continua con función de distribución

$$F_X(x) = \begin{cases} 0 & \text{para} \quad x < 0 \\ x & \text{para} \quad 0 \le x < 1 \\ 1 & \text{para} \quad 1 \le x \end{cases}$$

Calcular la función de densidad y la función de distribución de la variable aleatoria $Y = X^2$. Calcular $E(Y)$ y $Var(Y)$.

Me pide f_Y, F_Y, $E(Y)$ y $Var(Y)$

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

Si $y \le 0 \longrightarrow F_X(\sqrt{y}) - F_X(-\sqrt{y}) = 0 - 0 = 0$

Si $y \in [0,1] \longrightarrow F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \sqrt{y}$

Si $y \ge 1 \longrightarrow F_X(\sqrt{y}) - F_X(-\sqrt{y}) = 1 - 0 = 1$

$$f_Y(y) = F_Y'(y) = \begin{cases} \dfrac{1}{2\sqrt{y}} & \text{si } y \in [0,1] \\ 0 & \text{en el resto} \end{cases}$$

Otro método

1°) Calculo densidad de X

$$f_X(x) = F_X'(x) = \begin{cases} 1 & \text{si } y \in [0,1] \\ 0 & \text{en el resto} \end{cases}$$

2°) Aplico el Teorema que relaciona f_Y con f_X (biyección al ser $\boxed{g(x) = x^2}$ en $[0,1]$), $\boxed{g^{-1}(y) = \sqrt{y}}$

$$f_Y(y) = f_X(g^{-1}(y)) \cdot |(g^{-1}(y))'| = \begin{cases} \dfrac{1}{2\sqrt{y}} & \text{si } y \in [0,1] \\ 0 & \text{en el resto} \end{cases}$$

$$E(Y) = E(X^2) = \int_{-\infty}^{+\infty} t^2\,dt = \int_0^1 t^2\,dt = \left[\frac{t^3}{3}\right]_0^1 = \frac{1}{3}$$

$$Var(Y) = E(Y^2) - (E(Y))^2 = \frac{1}{5} - \left(\frac{1}{3}\right)^2 = \frac{4}{45}$$

$$E(Y^2) = E(X^4) = \int_{-\infty}^{+\infty} t^4\,dt = \int_0^1 t^4\,dt = \left[\frac{t^5}{5}\right]_0^1 = \frac{1}{5}$$

14. Sea X una variable aleatoria continua cuya función de densidad viene dada por

$$f(x)=\begin{cases}2x & \text{para} \quad 0<x<1\\ 0 & \text{para el resto}\end{cases}$$

Calcular la función de densidad y la función de distribución de las variables aleatorias:

a) $Y=3X+5$.

a) $Y = 3X+5$

Calculo la distribución de X integrando su densidad

- Si $x<0 \longrightarrow F_X(x) = \int_{-\infty}^{x} 0 \, dt = 0$

- Si $x \in [0,1] \longrightarrow F_X(x) = \int_{-\infty}^{0} 0 \, dt + \int_{0}^{x} 2t \, dt = 0 + \left[t^2\right]_0^x = x^2$

- Si $x \geq 1 \longrightarrow F_X(x) = 1$

$$F(x) = \begin{cases} 0 & \text{si} \quad x<0 \\ x^2 & \text{si} \quad x\in[0,1] \\ 1 & \text{si} \quad x \geq 1\end{cases}$$

Me piden:

$$F_Y(y) = P(Y\leq y) = P(3X+5 \leq y) = P\left(X \leq \frac{y-5}{3}\right) = F_X\left(\frac{y-5}{3}\right)$$

- Si $\frac{y-5}{3} < 0 \longleftrightarrow y \leq 5 \longrightarrow F_Y(y) = 0$

- Si $\frac{y-5}{3} \in [0,1] \longleftrightarrow 0 \leq \frac{y-5}{3} \leq 1 \longleftrightarrow 5 \leq y \leq 8 \longrightarrow F_Y(y) = F_X\left(\frac{y-5}{3}\right) = \left(\frac{y-5}{3}\right)^2$

- Si $\frac{y-5}{3} \geq 1 \longleftrightarrow y \geq 8 \longrightarrow F_Y(y) = 1$

¿Cómo calcular la densidad de Y?

- Forma 1: $f_Y(y) = F_Y'(y) = \begin{cases} \frac{2}{9}(y-5) \\ 0 \text{ en el resto}\end{cases}$

• **Forma 2:** Aplico el Teorema

$g: \mathbb{R} \longrightarrow \mathbb{R}$ biyección

$y(x) = 3x+5 \longrightarrow$ ¿$g^{-1}(y)$?

$y = 3x+5 \longrightarrow x = \frac{y-5}{3} = g^{-1}(y)$

$$\boxed{f_Y(y) = f_X(g^{-1}(y)) \left|(g^{-1}(y))'\right|}$$

$(g^{-1}(y))' = \left(\frac{y-5}{3}\right)' = \frac{1}{3}$

$f_X(g^{-1}(y)) = f_X\left(\frac{y-5}{3}\right) = \begin{cases} 0 & \text{si } \frac{y-5}{3} \notin [0,1] \longleftrightarrow y \notin [5,8] \\ 2\cdot\left(\frac{y-5}{3}\right) & \text{si } \frac{y-5}{3} \in [0,1] \longleftrightarrow y \in [5,8] \end{cases}$

$f_Y(y) = \begin{cases} 0 & \text{si } y \notin [5,8] \\ \frac{1}{3}\cdot 2\cdot\left(\frac{y-5}{3}\right) = \frac{2}{9}(y-5) & \text{si } y \in [5,8] \end{cases}$

b) $Y = X^2$.

Calcula F_Y y f_Y

$F_X(x) = \begin{cases} 0 & \text{si } x \le 0 \\ x^2 & \text{si } x \in [0,1] \\ 1 & \text{si } x \ge 1 \end{cases}$

$F_Y(y) = P(Y^2 \le y) = P(X^2 \le y) =$

$P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$

$\begin{cases} \bullet \text{ Si } y \le 0 \longrightarrow F_Y(y) = 0 \\ \bullet \text{ Si } \sqrt{y} \in [0,1] \longrightarrow F_Y(y) = (\sqrt{y})^2 - 0 = y \\ \bullet \text{ Si } y \ge 1 \longrightarrow F_Y(y) = 1 \end{cases}$

$F_Y(y) = \begin{cases} 0 & \text{si } y \le 0 \\ y & \text{si } y \in [0,1] \\ 1 & \text{si } y \ge 1 \end{cases}$

$f_Y(y) = F_Y'(y)$

$f_Y(y) = \begin{cases} 1 & \text{si } y \in (0,1) \\ 0 & \text{resto} \end{cases}$

Otra forma: Si $g: \mathbb{R} \longrightarrow \mathbb{R}$ biyección, entonces

$f_Y(y) = f_X(g^{-1}(y)) \cdot \left|(g^{-1}(y))'\right|$

$$g: \mathbb{R}^+ \longrightarrow \mathbb{R}^+$$
$$x \longrightarrow g(x) = x^2$$

$$y = x^2 \longrightarrow x = \sqrt{y} = g^{-1}(y)$$

$$\left(g^{-1}(y)\right)' = \frac{1}{2\sqrt{y}}$$

$$f_Y(y) = f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \begin{cases} 2\sqrt{y}\,\dfrac{1}{2\sqrt{y}} = 1 & \text{en } y \in (0,1) \\ 0 & \text{en el resto} \end{cases}$$

Tema 6: Modelos de Distribución

6.1) Modelos discretos

A) Uniforme discreta: $X \sim UD(n)$

- Definición

Diremos que X sigue una uniforme discreta de parámetro n. Si toma exactamente n valores, todos con la misma probabilidad.

– Rango o Soporte de X

$\text{Sop}(X) = (x_1, x_2, \ldots, x_n)$

– Función puntual de probabilidad

$$p(x) = P(X = x)$$

$$p(x_i) = \frac{1}{n} \; \forall x_i \in \text{Sop}(X)$$

– Esperanza o media de X

$$E(X) = x_1 \cdot \frac{1}{n} + x_2 \cdot \frac{1}{n} + \cdots + x_n \cdot \frac{1}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$$

– Varianza de X

$$\text{Var}(X) = \underbrace{E(X^2)}_{(*)} - (E(X))^2 = \overline{x^2} - (\overline{x})^2 = \sigma_x^2$$

$$(*) \; E(x^2) = x_1^2 \cdot \frac{1}{n} + x_2^2 \cdot \frac{1}{n} + \cdots + x_n^2 \cdot \frac{1}{n} = \frac{\sum x_i^2}{n} = \overline{x^2}$$

Ejemplo: $X =$"Resultado de lanzar un dado"$\sim UD(6)$

B) Modelo Bernoulli de parámetro p, $X \sim B(p)$

Llamaremos experimento Bernoulli de parámetro p a un experimento aleatorio con sólo 2 resultados posibles que llamaremos éxito (E) y fracaso (F).

$$\Omega = \{E, F\}, \text{ con } p = P(E) = P(\text{"éxito"})$$

Ejemplo

Chequearemos una pieza al azar de una producción y miro si es defectuosa.

- Definición

A la variable aleatoria $X = \begin{cases} 1 & \text{si el experimento Bernoulli resultó éxito} \\ 0 & \text{en caso contrario} \end{cases}$

$x : \Omega \longrightarrow \mathbb{R}$ se dice que sigue Bernoulli de parámetro p, $X \sim B(p)$.

  - Media o Esperanza de X

    $E(X) = 0 \cdot (1-p) + 1 \cdot p = p$

  - Varianza X

    $E(X^2) = 0^2 \cdot (1-p) + 1^2 \cdot p = p$

    $\text{Var}(X) = E(X^2) - (E(X))^2 = p - p^2)p(1-p) = p \cdot q$ con $q = 1-p$

Ejemplo

$X \equiv$ "Nº de piezas defectuosas al extraer al azar una de su producción".

$X \sim B(p)$ donde $p = P($"Pieza defectuosa"$)$=Tabla defectuosa en m_i producción

C) Modelo binomial de parámetro n y $p, X \sim B(n,p)$

- Definición

Consideremos un experimento Bernoulli de parámetro p, y supongamos que se repite el experimento n veces de formas independiente.

La variable aleatoria X ="Nº de éxitos obtenidos en las n repeticiones del experimento" sigue un modelo Binomial, $X \sim B(n,p)$

- Soporte o rango X

  $\text{Sop}(X) = \{0, 1, \ldots, n\}$

- Función puntual de probabilidad

$p(x) = P(X = x)$

$$p(0) = P(X = 0) = P(F_1 \cap F_2 \cap \cdots \cap F_n) = \left\{ \begin{array}{l} \text{linealmente} \\ \\ \text{independiente} \end{array} \right\} = \prod_{i=1}^{n} P(F_1)$$

$$p(1) = P(X = 1) = P(E_1 \cap F_2 \cap \cdots \cap F_n) \cup \cdots \cup P(F_1 \cap F_2 \cap \cdots \cap E_n) = np(1-p)^{n-1} = npq^{n-1}$$

$$p(2) = P(X = 2) = \binom{n}{2} \cdot p^2 \cdot (1-p)^{n-2}$$

$$p(x) = P(X = x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \quad \forall x \in \{0, 1, 2, \dots, n\}$$

- Observación

El modelo $B(n, p)$ se obtiene como suma de n modelos Bernoulli de parámetro p independientes.

$X_i \sim B(p),\ i = 1, \dots, n$ independientes

$X = X_1 + X_2 + \cdots + X_n \sim B(n, p)$

$E(X) = E(X_1) + E(X_2) + \cdots + E(X_n) = n \cdot p$

– Varianza de $X \sim B(n, p)$

$$\mathrm{Var}(X) = \mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) = n \cdot p \cdot (1-p) = npq$$

Nota: Si X e Y son variables aleatorias independientes:

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

- Propiedad: la binomial es reproductora respecto al parámetro n. Es decir:
  - Si $X \sim B(n_1, p), Y \sim B(n_2, P)$ independientes, entonces $X + Y \sim B(n_1 + n_2, p)$

Ejemplo 1

Consideremos una urna con 20 bolas, de las cuales 8 son blancas y 12 son negras. Extremos 5 bolas una a una con reemplazamiento.

La variable aleatoria $X \equiv$"Nº de bolas blancas en la muestra de tamaño 5"$\sim B\left(n = 5, p = \dfrac{8}{20} = \dfrac{2}{5}\right)$

Ejemplo 2

Sabemos que la tasa de defectuosas de una linea productiva es del 7%. Suponemos producción muy elevada y empaqueto las piezas en cajas de 15 unidades.

$X \equiv$"Nº de piezas defectuosas en una caja de 15 unidades"$\sim B(n = 15, p = 0.07)$

Nota: Disponemos de tablas para calcular las probabilidades puntuales

D) Modelo hipergeométrico, $X \sim M(N, a, n)$

Es como la Binomial pero los experimentos Bernoullis no se realizan de forma independiente.

Por ejemplo, si se realizan extracciones sin reemplazamiento.

$N \longrightarrow$ nº de elementos de población

$a \longrightarrow$ nº de éxitos de la población

$n \longrightarrow$ tamaño de la muestra extraída

$n = N - a \longrightarrow$ nº de fracasos posibles

La variable aleatoria $X \equiv$"Nº de éxitos en la muestra de tamaño n"$\sim +(N, a, n)$

$$\text{Sop}(X) = \{\underbrace{\max(0, N - b)}, \ldots, \min(n, a)\}$$

$$P(X = x) = P(\underbrace{E \cap E \cap \cdots \cap E}_{x} \cap \underbrace{F \cap F \cap \cdots \cap F}_{n-x}) = \frac{\binom{a}{x} \cdot \binom{N-a}{n-x}}{\binom{N}{n}} = \frac{\binom{a}{x} \cdot \binom{b}{n-x}}{\binom{a+b}{n}}$$

$$E(X) = n \cdot \frac{a}{N}$$

$$\text{Var}(X) = n \cdot \frac{a}{N} \cdot \left(\frac{N-a}{N}\right) \cdot \left(\frac{N-n}{N-1}\right)$$

Nota: Si $\frac{n}{N} \leq 0.10$, entonces podemos aproximar la $H(N, a, n)$ por una

$$B\left(n, p = \frac{a}{N}\right)$$

En este caso, el muestreo sin reemplazamiento sería equivalente al muestreo sobre poblaciones infinitas.

Ejemplo

Una urna con 20 bolas de las cuales son 8 blancas y 12 negras. Se extrae una muestra de tamaño 5 sin reemplazamiento.

$$X \sim\text{"Nº de bolas blancas en la muestra"} \sim H(N = 20, a = 8, n = 5)$$

E) Geométrica, $X \sim G(p)$

Consideramos experimento Bernoulli(p) que se realiza de forma independiente hasta obtener el 1^er^ éxito.

La variable aleatoria $X \equiv$"Nº de fracasos hasta obtener el 1^er^ éxito"$\sim G(p)$.

$\text{Sop}(X) = \{0, 1, 2, \dots\} = \mathbb{N}$

$P(X = x) = P(\underbrace{F_1 \cap F_2 \cap \cdots \cap F_x}_{x \text{ fracasos}} \cap \underbrace{E_1}_{\text{éxito}}) = (1-p)^x \cdot p = p \cdot q^x$

$$E(X) = \sum_{x=0}^{\infty} x \cdot p \cdot q^x = \sum_{x=1}^{\infty} xpq^x = pq \sum_{x=1}^{\infty} xq^{x-1} = pq \sum_{x=1}^{\infty} \frac{\partial}{\partial q}(q^x) = pq \frac{\partial}{\partial q}\underbrace{\left(\sum_{x=1}^{\infty} q^x\right)}_{\frac{q}{1-q}} =$$

$$pq \frac{\partial}{\partial q}\left(\frac{q}{1-q}\right) = pq \frac{1-q+q}{(1-q)^2} = \frac{\cancel{p}q}{p^{\cancel{2}}} = \boxed{\frac{q}{p} = E(X)}$$

Tenemos la suma infinita de una progresión geométrica de razón $q_\infty \in (0,1)$, por tanto converge.

<u>Nota:</u> $\displaystyle\sum_{i=1}^{\infty} a_i = \frac{a_1}{1-r}$, $a_n = a_1 \cdot r^{n-1}$ si $|r| < 1$

$$\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{2q^2}{p^2} + \frac{p}{q} - \left(\frac{q}{p}\right)^2 = \frac{2q^2 + pq - q^2}{p^2} = \frac{q\overbrace{(q+p)}^{1}}{p^2} = \boxed{\frac{q}{p^2}}$$

Vamos a calcular $E(X(X-1)) = E(X^2) - E(X) \longrightarrow E(X^2) = E(X(X-1)) + E(X) = \dfrac{2q^2}{p^2} + \dfrac{q}{p}$

$$E(X(X-1)) = \sum_{x=0}^{\infty} x(x-1) \cdot p \cdot q^x = \sum_{x=2}^{\infty} x(x-1)pq^x = pq^2 \sum_{x=2}^{\infty} x(x-1)q^{x-2} = pq^2 \sum_{x=2}^{\infty} \frac{\partial^2}{\partial q^2}(q^x) =$$

$$pq^2 \frac{\partial^2}{\partial q^2} \underbrace{\left(\sum_{x=2}^{\infty} q^x\right)}_{\frac{q^2}{1-q}} = pq^2 \frac{\partial^2}{\partial q^2}\left(\frac{q^2}{1-q}\right) = (*)$$

$$\frac{\partial}{\partial q}\left(\frac{q^2}{1-q}\right) = \frac{2q(1-q)+q^2}{(1-q)^2} = \frac{2q-q^2}{(1-q)^2} \longrightarrow \frac{\partial^2}{\partial q^2}\left(\frac{q^2}{1-q}\right) = \frac{(2-2q)(1-q)^2 + 2(2q-q^2)(1-q)}{(1-q)^4} =$$

$$\frac{2\cancel{(1-q)}\overbrace{\left((1-q)^2 + (2q-q^2)\right)}^{1-\cancel{2q}+\cancel{q^2}+\cancel{2q}-\cancel{q^2}}}{(1-q)^{\cancel{4}}} = \frac{2}{(1-q)^3} = \frac{2}{p^3}$$

$$(*) = \cancel{p}q^2 \frac{2}{p^{\cancel{3}}} = \frac{2q^2}{p^2}$$

F) Binomial negativa, $X \sim BN(n, p)$

Consideramos experimento Bernoulli(p) y realizamos el experimento de forma independiente hasta obtener el n-ésimo éxito.

La variable aleatoria $X \equiv$"Nº de fracasos hasta obtener el n-ésimo éxito" $X \sim BN(n,p)$

- $\text{Sop}(X) = \{0,1,2,\dots\} = \mathbb{N}$
- $P(X=x) = P(\underbrace{E \cap P \cap \cdots \cap \boxed{E}}_{n+x \text{ posiciones}}) = \boxed{\binom{n+x-1}{x} \cdot (1-p)^x \cdot p^n}$

Sea $y_1 =$ "Nº de fracasos hasta 1$^{\text{er}}$ éxito" $\sim G(p) y_2$ = "Nº de fracasos hasta 1$^{\text{er}}$ y 2

$\vdots$

$y_n =$ "Nº de fracasos entre $(n-1)$-ésimo y n-ésimo éxito" $\sim G(p)$

$x = y_1 + y_2 + \cdots + y_n$ con y_j independiente $\forall j = 1,2,\dots,n$

$E(X) = E(y_1) + \cdots + E(y_n) = \frac{q}{p} + \frac{q}{p} + \cdots + \frac{q}{p} = \boxed{\frac{nq}{p}}$

$\text{Var}(X) = \text{Var}(y_1) + \cdots + \text{Var}(y_n) = \boxed{\frac{nq}{p^2}}$

Nota 1:

$X =$"Nº de fracasos hasta obtener 1$^{\text{er}}$ éxito"$\sim G(p)$

$Y =$"Nº de experimentos necesarios para obtener 1$^{\text{er}}$ éxito"

$Y = X + 1 \cdot E(Y) = E(X) + 1 = \frac{q}{p} + 1 = \boxed{\frac{1}{p}}$

$$\text{Var}(Y) = \text{Var}(X) = \boxed{\frac{q}{p^2}}$$

$P(Y=3) = P(X=2)$

Nota 1:

$X =$"Nº de fracasos hasta obtener n-ésimo éxito"$\sim G(p)$

$Y =$"Nº de experimentos necesarios para obtener el n-ésimo éxito"

$Y = X + n \cdot E(Y) = E(X) + n = \frac{q}{p} + n = n \cdot \left(\frac{q}{p} + 1\right) = \boxed{\frac{n}{p}}$

$$\text{Var}(Y) = \text{Var}(X) = \boxed{\frac{nq}{p^2}}$$

G) Modelo de Poisson, $X \sim P(\lambda)$

- Definición

Diremos que la variable aleatoria X sigue un modelo de Poisson, $X \sim P(\lambda)$, de parámetro λ si su función puntual de probabilidad es:

$$p(x) = P(X = x) = e^{-1} \cdot \frac{\lambda^x}{x!},\ x = 1, 2, \ldots, \mathbb{N}$$

Probamos que $E(X) = \lambda$ y $\mathrm{Var}(X) = \lambda$.

¿Qué fenómenos modeliza la Poisson?

Escenario 1: Cuando tengamos una Binomial, $B(n, p)$, con n grande y p pequeña.

- Regla: $n \geq 20, p \leq 0.1, n \cdot p \leq 5 \longrightarrow$ aproximación $B(n, p)$ por $P(\lambda = np)$.

Escenario 2: Contar el número de veces que ocurre un suceso sobre un soporte continuo (intervalo de tiempo/superficie).

$E(X) = \lambda$

$$\boxed{\text{Nota: } e^a = \sum_{n=0}^{\infty} \frac{a^n}{n!}}$$

$$E(X) = \sum_{x=0}^{\infty} = x \cdot e^{-\lambda} \cdot \frac{\lambda^x}{x!} = \sum_{x=1}^{\lambda} xe^{-1} \cdot \frac{\lambda^x}{x!} = e^{-\lambda} \cdot \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = e^{-\lambda} \cdot \lambda \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{-\lambda} \cdot \lambda \cdot e^{\lambda} = \boxed{\lambda}$$

$\mathrm{Var}(X) = E(X^2) - (E(X))^2$

$E(X^2) = E(X(X-1)) + E(X)$

$X(X-1) = X^2 - X$

$x^2 = X(X-1) + X$

$$E(X(X-1)) \sum_{x=0}^{\infty} x(x-1) \cdot e^{-\lambda} \cdot \frac{\lambda^x}{x!} = e^{-\lambda} \cdot \lambda^2 \sum_{x=0}^{\infty} x(x-1) \cdot \frac{\lambda^{x-2}}{(x-2)!} = e^{-\lambda} \cdot \lambda^2 \cdot \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = \underbrace{e^{-\lambda} \cdot e^{\lambda}}_{1} \cdot \lambda^2 = \boxed{\lambda^2}$$

$$\boxed{\mathrm{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda}$$

- Propiedades

Si $X_1 \sim P(\lambda_1)$ y $X_2 \sim P(\lambda_2)$ independientes $\longrightarrow X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$. La Poisson es reproductiva.

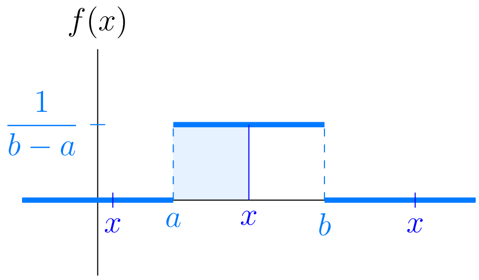
6.2) Modelos continuos

A) Uniforme, $\sim U(a, b)$

- Definición

Diremos que X sigue una distribución uniforme en el intervalo (a,b) si tiene función de densidad constante

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } x \in (a,b) \\ 0 & \text{si } x \notin (a,b) \end{cases}$$



La función de distribución vale:

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

$$F(x) = \begin{cases} 0 & \text{si } x \leq a \\ \dfrac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x \geq b \end{cases}$$

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = (*)$$

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x = \int_a^b x \cdot \frac{1}{b-a}\,\mathrm{d}x = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{\cancel{(b-a)}(b+a)}{2\cancel{(b-a)}} = \boxed{\frac{b+a}{2}}$$

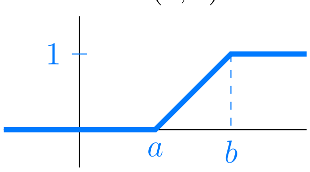
$$E(X^2) = \int_a^b x^2 \cdot \frac{1}{b-a}\,\mathrm{d}x = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{\cancel{(b-a)}(b^2+ab+a^2)}{3\cancel{(b-a)}} = \boxed{\frac{b^2+ab+a^2}{3}}$$

$$(*) = \frac{b+a}{2} - \frac{b^2+ab+a^2}{3} = \frac{3(b+a) - 2(b^2-2ab-2a^2)}{6} = \boxed{\frac{-2b^2-2a^2+3b-2ab+3a}{6}}$$

Ejemplo

$X =$"Seleccionar al azar un número en el intervalo $(0,1)$"

$F(x) \rightarrow X \sim U(0,1)$

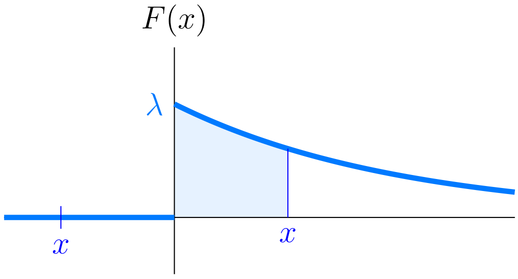


B) Exponencial, $X \sim \mathrm{Exp}(\lambda)$

- Definición

Diremos que X sigue $\mathrm{Exp}(\lambda)$ si su densidad es:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \text{ con } \lambda > 0 \\ 0 & \text{en el resto} \end{cases}$$



La función de distribución:

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

- Si $x \geq 0 \longrightarrow F(x) = \int_0^x \lambda e^{-\lambda t}\,\mathrm{d}t = \left[-e^{-\lambda t}\right]_0^x = -e^{-\lambda x} + 1 = \boxed{1 - e^{-\lambda x}}$

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-\lambda x} & \text{si } x \geq 0 \end{cases}$$

- $E(X) = \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x = \int_0^{+\infty} x\lambda e^{-\lambda x}\mathrm{d}x = \lambda \int_0^{+\infty} x e^{-\lambda x}\mathrm{d}x = \left\{ \begin{array}{ll} x = u & \mathrm{d}x = \mathrm{d}u \\ e^{-\lambda x}\,\mathrm{d}x = \mathrm{d}v & v = -\frac{1}{\lambda} e^{-\lambda x} \end{array} \right\}$

  $\lambda \cdot \left(\left[x \cdot \frac{1}{\lambda} e^{-\lambda x} \right]_0^{+\infty} - \int_0^{+\infty} -\frac{1}{\lambda} e^{-\lambda x}\,\mathrm{d}x \right) = \lambda \cdot \left(-\frac{1}{\lambda}\right) \cdot \left(0 - \left[\frac{1}{\lambda} e^{-\lambda x} \right]_0^{+\infty} \right) = \boxed{\frac{1}{\lambda}}$

  $\lim_{x \to +\infty} x e^{-\lambda x} = \lim_{x \to +\infty} \frac{x}{e^{\lambda x}} = \frac{\infty}{\infty} = \{\text{L'Hôpital}\} = \lim_{x \to +\infty} \frac{1}{\lambda e^{\lambda x}} = \frac{1}{\infty} = 0$

  $\begin{cases} \Gamma(p) = \int_0^{+\infty} t^{p-1} \cdot e^{-t}\,\mathrm{d}t \\ \Gamma(n) = (n-1)! \end{cases}$

- $\mathrm{Var}(X) = E(X^2) - (E(X))^2 = (*)$

  $E(X^2) = \int_0^{+\infty} x^2 \cdot \lambda e^{-\lambda x}\,\mathrm{d}x = \left\{ \begin{array}{l} t = \lambda x \\ x = \frac{t}{\lambda} \longrightarrow \mathrm{d}x = \frac{1}{\lambda}\,\mathrm{d}t \end{array} \right\} = \lambda \cdot \int_0^{+\infty} \frac{t^2}{\lambda^2} e^{-t} \cdot \frac{1}{\lambda}\,\mathrm{d}t =$

  $\frac{1}{\lambda^2} \int_0^{+\infty} t^2 \cdot e^{-t}\,\mathrm{d}t = \boxed{\frac{2}{\lambda^2}}$

  $(*) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \boxed{\frac{1}{\lambda^2}}$

- ¿Moda? El máxumo de $f(x)$ se alcanza en $x = 0 =$ Moda.

- Mediana: Me =punto que cumple $F(Me) = \frac{1}{2} = 1 - e^{-\lambda x} \longrightarrow e^{-\lambda x} = \frac{1}{2} \longrightarrow \frac{1}{e^{\lambda x}} = \frac{1}{2} \longrightarrow$

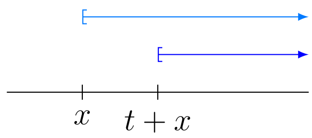
  $e^{\lambda x} = 2 \longrightarrow \lambda x = \ln(2) \longrightarrow \boxed{x = \frac{1}{\lambda} \cdot \ln(2)}$

- Propiedad (falta de memoria)

Si $X \sim \text{Exp}(\lambda)$, entonces $P(X > t + x/X > t) = P(X > x)$

- Demostración

$P(X > t + x/X > t) = \dfrac{P((X > t) \cap (X > t + x))}{P(X > t)} = \dfrac{P(X > t + x)}{P(X > t)} = \dfrac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x} =$
$P(X > x)$



Ejemplo

"Tiempo de vida de un dispositivo que no envejece"

C) Weibull, $X \sim W(a, b)$

Si $Y \sim \text{Exp}(\lambda) \longrightarrow x = Y^c$ es una Weibull $c > 0, c = \dfrac{1}{a} \quad \lambda = b^{-a}$

- Si $c = 1$ sale la exponencial
- Sirve para modelizar tiempo de vida con tasas de fallo decreciente, constante o creciente.

$X =$"Tiempo de vida de un dispositivo"

D) Gamma, $X \sim \text{Gamma}(a, b)$

- Es una generalización Exponencial.
- Para $a = 1$ se obtiene $\text{Exp}(\lambda = b)$
- Si $a \leq n$ entero positivo, se denomina distribución Erlang ($\text{Gamma}(\text{int}(a), b)$), que es suma de exponenciales independientes.

$X =$"Tiempo trascurrido hasta el n-ésimo suceso o evento"

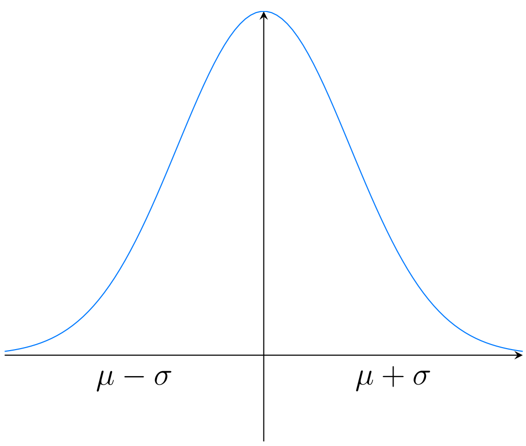
E) Modelo Normal, $X \sim N(\mu, \sigma)$

- Definición

Diremos que X sigue una $N(\mu, \sigma)$ si su densidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \; \forall x \in \mathbb{R}$$

a) ¿Cómo es la densidad?



- Es simétrica con eje de simétrica $x = \mu$.
- En $x = \mu$ tiene un máximo se puede ver que $f'(x = \mu) = 0$ y $f''(x = \mu) < 0$.
- Tiene puntos de inflexión en $x = \mu - \sigma$ y $x = \mu + \sigma$. Comprobar que $f'''(x) = 0 \longleftrightarrow x = \mu \pm \sigma$
- Tiene asíntota horizontal $y = 0$

$$\lim_{x\to -\infty} f(x) = \lim_{x\to +\infty} f(x) = 0$$

Se puede demostrar:

1) Que $f(x)$ es densidad
2) Que $E(x) = \mu$
3) Que $\sigma(X) = \sigma$, $\mathrm{Var}(X) = \sigma^2$

b) ¿Cuándo se usa la Normal?

- Para modelizar errores de medición
- Para aproximar la Binomial y Poisson
- Para Inferencia (Teorema Central del Límite)

c) Función distribución

- No existe la primitiva de $f(x)$
- Se obtiene mediante integración numérica

Tenemos tabulada la $N(\mu = 0, \sigma = 1)$

d) Tipificar: Si $X \sim N(\mu, \sigma) \longrightarrow z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$

Entonces podremos calcular la probabilidad de la $N(\mu, \sigma)$ a partir de la tabla $N(0, 1)$ si tipificamos.

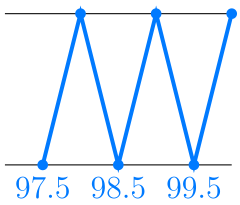
e) Propiedades de la Normal

1) Si $X \sim N(\mu_X, \sigma_X) \longrightarrow Y = \underset{a,b\in\mathbb{R}}{aX + b} \sim N(\mu_Y = a\mu_X + b, \sigma_Y = |a| \cdot \sigma_X)$
2) Si $X_1, X_2, \ldots, X_n$ sea $N(\mu_i, \sigma_i)$ $i = 1, 2, \ldots, n$, independientes entocnes $Y = X_1 + X_2 + \cdots + X_n \sim N(\mu_Y = \mu_1 + \cdots + \sigma_n, \sigma_Y = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2})$

3) - Si $X \sim B(n,p)$ con n grande ($n \geq 30$) y $np(1-p) > 5$, entonces X se puede aproximar por $Y \sim N(\mu = np, \sigma = \sqrt{np(1-p)})$

- Si $X \sim P(\lambda)$ como $\lambda \geq 5$, entonces X se puede aproximar por $Y \sim N(\mu = \lambda, \sigma = \sqrt{\lambda})$. La aproximación se puede mejorar realizando corrección por continuidad

$$P(X = a) \simeq P(a - 0.5 \leq Y \leq a + 0.5)$$



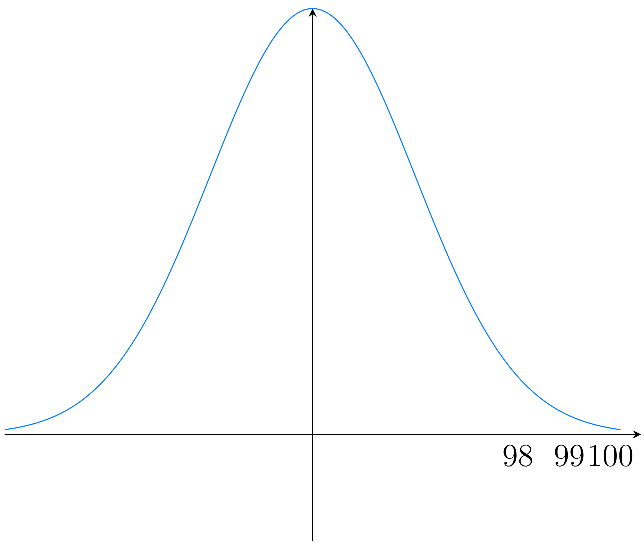
$B(n = 100, p = 0.1)$

$B(n,p)$

$P(17 < X \leq 35) \simeq P(17.5 \leq X \leq 35.5)$

$N(\mu = np = 10, \sigma = 3)$

$P(X > 20) \simeq P(X \geq 20.5)$



PROBLEMAS. RELACIÓN 6: Modelos básicos de distribuciones de probabilidad.

FUNDAMENTOS DE PROBABILIDAD Y ANÁLISIS EXPLORATORIO DE DATOS.

GRADO EN CIENCIA E INGENIERÍA DE DATOS.

1. El 80% de las bolas contenidas en una urna son de color blanco, siendo el 20% restante de color rojo. Determinar la probabilidad de que al efectuar 3 extracciones -sucesivas con reemplazamiento- 2 de las bolas extraídas sean de color blanco y 1 de color rojo.

2. Una empresa, dedicada a la venta de un determinado tipo de artículo, que ofrece a sus habituales clientes dos formas de pago: "pago al contado", o "pago aplazado", sabe que el 20% de las unidades adquiridas de dicho artículo lo son bajo la forma de "pago al contado". Si en un período de tiempo determinado, se han adquirido 5 unidades, determinar la probabilidad de que:
   a) 2 unidades o más, lo hayan sido bajo la forma de "pago al contado".
   b) 2 unidades o menos, lo hayan sido bajo la forma "pago aplazado".

3. El número medio de automóviles que llega a una estación de suministro de gasolina es de 10 por hora. Determinar:
   a) La probabilidad de que en una hora lleguen menos de 3 automóviles.
   b) La probabilidad de que en una hora concreta lleguen menos de 5 automóviles si sabemos que ya llegó uno a repostar.
   c) La probabilidad de que en el periodo de tres horas consecutivas lleguen más de 15 automóviles.

4. Por parte de una compañía de seguros se sabe que el 0,003% de los individuos de una población fallece cada año de un determinado tipo de accidente. Determinar:
   a) La probabilidad de que la compañía tenga que pagar a más de 3 de los 10000 asegurados por fallecer en tal tipo de accidente en un año determinado.
   b) El número de fallecimientos esperados para el conjunto de asegurados y debidos a ese tipo de accidente.

5. En una determinada zona geográfica se pretende introducir un nuevo producto del que es razonable esperar sea demandado por el 0,4% de los habitantes de dicha zona. Determinar la probabilidad de que, consultados 1000 de éstos, dicho producto sea demandado:
   a) Por 3 o más.
   b) Por 5 o menos.

6. De 50 edificios en un parque industrial, 12 no cumplen el código eléctrico. Si se seleccionan aleatoriamente 10 edificios para inspeccionarlos, ¿cuál es la probabilidades que exactamente 3 de los 10 no cumplan el código?

7. Una prueba de resistencia de soldadura consiste en poner carga en uniones soldadas hasta que se dé una ruptura. Para cierto tipo de soladura el 80% de las rupturas ocurre en la propia soldadura mientras que el otro 20% se da en las vigas. Se realizan varias pruebas de resistencia.
   a) ¿Cuál es la probabilidad de que la ruptura de la viga se dé en la tercera prueba?
   b) ¿Cuál es la probabilidad de que tenga que hacer 10 pruebas antes de romper 3 vigas?

8. Acerca de la cantidad aleatoria demandada durante un cierto período de tiempo por parte de una empresa textil, se sabe que sigue una distribución uniforme que no supera la tonelada. Determinar, para dicho período de tiempo:
   a) La probabilidad de que la cantidad demandada no supere los 900 kilos.
   b) La probabilidad de que la cantidad demandada esté comprendida entre 800 y 900 kilos.
   c) La demanda esperada.

9. La duración aleatoria de un determinado tipo de artículo, en horas, viene regulada por la ley de probabilidad $N(\mu = 180, \sigma = 5)$. Determinar la probabilidad de que la duración de tal artículo:
   a) Sea superior a 170 horas.
   b) Sea inferior a 150 horas.

10. Una empresa sabe que el comportamiento en probabilidad de la demanda aleatoria de un artículo que produce, viene explicada por la ley $N(\mu = 10000, \sigma = 100)$. La empresa decide seguir produciendo el artículo en el futuro, sólo si la demanda está comprendida entre 9930 y 10170 unidades, determinar la probabilidad de que no siga produciendo tal artículo.

11. Sabiendo que la demanda aleatoria de gasolina, durante un cierto período de tiempo, se comporta con arreglo a la ley normal de media 150000 litros, con desviación típica igual a 10000 litros, determinar la cantidad que hay que tener dispuesta a la venta en dicho período, para poder satisfacer la demanda con una probabilidad de 0,95.

12. Un establecimiento comercial dispone a la venta diariamente en una de sus secciones, sólo dos artículos a precios p_1 y p_2, de forma que:
    - el 70% de las unidades ofrecidas, lo son del artículo de precio p_1.
    - el 30% de las unidades ofrecidas, lo son del artículo de precio p_2.

    Si en un día determinado, se venden en dicha sección 2000 unidades, determinar la probabilidad de que más de 800 unidades correspondan al artículo de precio p_2.

13. Un agricultor desea vender 4000 kg de naranjas y 1000 kg de limones. El precio del kilo de naranjas sigue una distribución $N(\mu = 0{,}31 \text{ euros}, \sigma = 0{,}06 \text{ euros})$ mientras que el precio del kilo de limones sigue una distribución $N(\mu = 0{,}43 \text{ euros}, \sigma = 0{,}16 \text{ euros})$ independiente de la anterior. ¿Cuál es la probabilidad de que el importe de la venta sea inferior a 1500 euros?

14. El tiempo de vida de un circuito integrado particular tiene una distribución exponencial con media de 2 años.
    a) Hallar la probabilidad de que el circuito dure más de 3 años.
    b) Suponiendo que el circuito tiene actualmente 4 años y aún funciona, calcular la probabilidad de que funcione 3 años más.

1. El 80% de las bolas contenidas en una urna son de color blanco, siendo el 20% restante de color rojo. Determinar la probabilidad de que al efectuar 3 extracciones -sucesivas con reemplazamiento- 2 de las bolas extraídas sean de color blanco y 1 de color rojo.

B: "La bola extraida es blanca" $P(B) = 0.8$

R: "La bola extraida es roja" $P(R) = 0.2$

Exp. Bernoulli: "Extraer una bola y ver si es blanca"

E: "éxito": B Repetimos el exp. 3 veces de forma indep.

F: "fracaso": R (porque es con reemplazo)

X: "Nº bolas blancas en la muestra de tamaño 3."

$X \sim B(n=3, p=0.8)$

• Me piden $P(X=2) = \binom{3}{2} \cdot (0.8)^2 \cdot (0.2) = \frac{3!}{2!1!} \cdot 0.8^2 \cdot 0.2 = 0.384$

Y = "N bolas rojas en la muestra de tamaño 3" $Y \sim B(n=3, p=0.2)$

• Me piden $P(Y=1) = \binom{3}{1} \cdot (0.2) \cdot (0.8)^2 = \frac{3!}{1!2!} \cdot 0.2 \cdot 0.8^2 = 0.384$

2. Una empresa, dedicada a la venta de un determinado tipo de artículo, que ofrece a sus habituales clientes dos formas de pago: "pago al contado", o "pago aplazado", sabe que el 20% de las unidades adquiridas de dicho artículo lo son bajo la forma de "pago al contado". Si en un período de tiempo determinado, se han adquirido 5 unidades, determinar la probabilidad de que:
   a) 2 unidades o más, lo hayan sido bajo la forma de "pago al contado".
   b) 2 unidades o menos, lo hayan sido bajo la forma "pago aplazado".

Éxito: "Pago al contado" $\longrightarrow P(E) = 0.2$

$n = 5 \longrightarrow$ Repito exp. Bernoulli de forma independiente

X = "Nº de unidades pagadas al contado de entre los 5 vendidos"

$X \sim B(n=5, p=0.2)$

a) Me piden $P(X \geq 2) = P(X=2) + P(X=3) + P(X=4) + P(X=5) =$

$0.2048 + 0.0512 + 0.0064 + 0.0003 = \boxed{0.2627}$

$P(X \geq 2) = 1 - (P(X=0) + P(X=1))$

b) Me piden $P(X \geq 3) = P(X=3) + P(X=4) + P(X=5) = \boxed{0.0579}$

3. El número medio de automóviles que llega a una estación de suministro de gasolina es de 10 por hora. Determinar:
   a) La probabilidad de que en una hora lleguen menos de 3 automóviles.
   b) La probabilidad de que en una hora concreta lleguen menos de 5 automóviles si sabemos que ya llegó uno a repostar.
   c) La probabilidad de que en el periodo de tres horas consecutivas lleguen más de 15 automóviles.

X: "Nº coches que llegan en una hora" $\sim P(\lambda = 10)$

a) Me piden $P(X < 3)$

$P(X<3) = P(X=0) + P(X=1) + P(X=2) = e^{-10}\left(\frac{10^0}{0!} + \frac{10^1}{1!} + \frac{10^2}{2!}\right) \approx$

$\boxed{0.028}$

b) Me piden $P(X < 5 / X \geq 1)$

$P(X<5/X \geq 1) = \frac{P(1 \leq X < 5)}{P(X \geq 1)} = \frac{P(X=1) + P(X=2) + P(X=3) + P(X=4)}{1 - P(X=0)}$

$\approx \boxed{0.0292}$

c) X_1: "Nº de personas que llegan en la 1ª hora" $\sim P(\lambda_1 = 10)$

X_2: "Nº de personas que llegan en la 2ª hora" $\sim P(\lambda_2 = 10)$

X_3: "Nº de personas que llegan en la 3ª hora" $\sim P(\lambda_3 = 10)$

$Y = X_1 + X_2 + X_3 \sim P(\lambda = 30)$ = "Nº de coches que llegan en las 3 horas"

Me pide $P(Y > 15)$ (se aproxima con una Normal)

Sea $W \sim N(\mu = 30, \sigma = \sqrt{30})$

Entonces $Y \approx W$

$P(Y>15) \simeq P(W > 15.5) = P\left(\frac{W-30}{\sqrt{30}} \geq \frac{15.5-30}{\sqrt{30}}\right) = P(Z \geq -2.65) =$

$P(Z \leq 2.65) = \boxed{0.9959}$

4. Por parte de una compañía de seguros se sabe que el 0,003% de los individuos de una población fallece cada año de un determinado tipo de accidente. Determinar:
   a) La probabilidad de que la compañía tenga que pagar a más de 3 de los 10000 asegurados por fallecer en tal tipo de accidente en un año determinado.
   b) El número de fallecimientos esperados para el conjunto de asegurados y debidos a ese tipo de accidente.

0.003% Fallece
Resto No fallece

a) P("Pague a más de 3 de los 1000 asegurados")?

$$X_i \begin{cases} 1 & \text{si Fallece} \\ 0 & \text{si No fallece} \end{cases} \quad X_i \sim b(p)$$

$p = P(X_i = 1) = 0.00003$

$q = 1 - p = 0.99997$

$X = \sum_{i=1}^{10000} x_i \quad i = 1, 2, \ldots, 10000 \longrightarrow X \sim B(n = 10000, p = 0.00003)$

Nos piden $P(X > 3)$ ya que X = nº de asegurados (de los 10000) que F.

$P(X = r) = \binom{n}{r} \cdot p^r \cdot q^{n-r} \longrightarrow$ Muy pesado

Vamos a aproximar $B(n, p)$ por una $P(\lambda = np)$

Condiciones: $\left.\begin{array}{l} n \geq 20 \\ p < 0.1 \\ np < 5 \end{array}\right\}$ Se cumplen ya que $n = 10000$, $p = 0.00003$ y $np = 0.3 < 5$. Por tanto:

$B(n, p) \sim P(\lambda = np < 5)$ $\boxed{\lambda = 0.3}$

$P(X > 3) = \sum_{r=4}^{10000} P(X = r) = 1 - \sum_{r=0}^{3} P(X = r)$: $P(X = 0) = \frac{(0.3)^0}{0!} e^{-0.3} = 0.7408$; $P(X = 1) = 0.2222$; $P(X = 2) = 0.0333$; $P(X = 3) = 0.0033$

$P(X > 3) = 1 - 0.9996 = 0.0004$

b) nº accidentes esperados $\longrightarrow E(X)$?

$E(X) = np = 0.3 \longrightarrow \boxed{\lambda = 0.3}$

5. En una determinada zona geográfica se pretende introducir un nuevo producto del que es razonable esperar sea demandado por el 0,4% de los habitantes de dicha zona. Determinar la probabilidad de que, consultados 1000 de éstos, dicho producto sea demandado:
   a) Por 3 o más.
   b) Por 5 o menos.

Éxito $\longrightarrow$ Producto demandado, $P(E) = 0.004$.

X = "Nº de personas entre las 100 consultadas, que demandan el producto" $\sim B(n = 1000,\ p = 0.004)$

Como $n \geq 20$, $p \leq 0.1$, $n \cdot p = 1000 \cdot 0.004 = 4 < 5$, aproximo $B(n, p)$ por $P(\lambda = n \cdot p = 4)$

a) Me piden $P(X \geq 3) = 1 - (P(X = 0) + P(X = 1) + P(X = 2)) =$

$1 - (0.0183 + 0.0733 + 0.1465) = 0.7619$

b) $P(X \leq 5) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) =$

$0.0183 + 0.0733 + 0.1465 + 0.1954 + 0.1954 + 0.1563 = 0.7852$

6. De 50 edificios en un parque industrial, 12 no cumplen el código eléctrico. Si se seleccionan aleatoriamente 10 edificios para inspeccionarlos, ¿cuál es la probabilidades que exactamente 3 de los 10 no cumplan el código?

$N = 50 \longrightarrow$ tamaño población

$a = 12 \longrightarrow$ nº éxitos en la población

$b = 50 - 12 = 38 \longrightarrow$ nº fracaso en la población

$n = 10 \longrightarrow$ tamaño muestral

X = "Nº de edificios que no cumplen el código eléctrico en una muestra de 10 unidades" $\sim H(N = 50,\ a = 12,\ n = 10)$

Me piden $P(X = 3) = \dfrac{\binom{12}{3} \cdot \binom{38}{7}}{\binom{50}{10}} = 0.2703$

7. Una prueba de resistencia de soldadura consiste en poner carga en uniones soldadas hasta que se dé una ruptura. Para cierto tipo de soladura el 80% de las rupturas ocurre en la propia soldadura mientras que el otro 20% se da en las vigas. Se realizan varias pruebas de resistencia.
   a) ¿Cuál es la probabilidad de que la ruptura de la viga se dé en la tercera prueba?
   b) ¿Cuál es la probabilidad de que tenga que hacer 10 pruebas antes de romper 3 vigas?

Exp. Bernoulli $\begin{cases} \text{Se rompe por soldadura (Fracaso)} \longrightarrow P(F) = 0.8 \\ \text{Se rompe la viga (Éxito)} \longrightarrow P(E) = 0.2 \end{cases}$

a) Y = "Nº de pruebas que necesito hasta romper 1ª viga"

Me piden $P(Y=3)$

X = "Nº de soldaduras rotas hasta romper 1ª viga" $\sim G(p=0.2)$

$Y = X+1$, $P(Y=3) = P(X=2) = 0.2 \cdot (0.8)^2 = 0.128$

b) Y = " Nº de pruebas hasta romper la tercera viga "

X : "Nº de soldaduras rotas hasta romper 3ª viga "

$X \sim BN(n=3, p=0.2)$

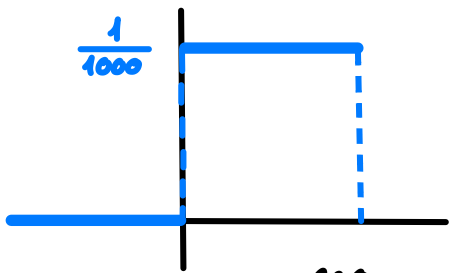
Me piden $P(Y=10) = P(X=7) = (0.2)^3 \cdot (0.8)^7 \cdot \binom{9}{7} = 0.06$

8. Acerca de la cantidad aleatoria demandada durante un cierto período de tiempo por parte de una empresa textil, se sabe que sigue una distribución uniforme que no supera la tonelada. Determinar, para dicho período de tiempo:
   a) La probabilidad de que la cantidad demandada no supere los 900 kilos.
   b) La probabilidad de que la cantidad demandada esté comprendida entre 800 y 900 kilos.
   c) La demanda esperada.

X : " Demanda textil de la empresa en una semana de tiempo "

$X \sim \text{Uniforme}(0, 1000)$

$$f(x) \begin{cases} \frac{1}{b-a} = \frac{1}{1000} & \text{si } x \in (0, 1000) \\ 0 & \text{en el resto} \end{cases}$$



a) $P(X \leq 900) = P(0 \leq X \leq 900) = \int_0^{900} \frac{1}{1000} dx = \frac{1}{1000} [x]_0^{900} = \frac{900}{1000} = \boxed{0.9}$

b) $P(800 \leq X \leq 900) = \frac{900-800}{1000} = \boxed{0.1}$

c) $E(X) = \frac{a+b}{2} = \boxed{500 \text{ kg}}$

$$E(X) = \int_{\mathbb{R}} x f(x) dx = \int_0^{1000} x \cdot \frac{1}{1000} dx = \frac{1}{1000} \left[\frac{x^2}{2}\right]_0^{1000} = \frac{1000^2}{1000 \cdot 2} = \frac{1000}{2} = \boxed{500}$$

9. La duración aleatoria de un determinado tipo de artículo, en horas, viene regulada por la ley de probabilidad $N(\mu = 180, \sigma = 5)$. Determinar la probabilidad de que la duración de tal artículo:

a) Sea superior a 170 horas.
b) Sea inferior a 150 horas.

X : "Duración en horas de un artículo"

$X \sim N(\mu = 180, \sigma = 5)$

a) $P(X > 170) \underset{\uparrow \text{ tipifico}}{=} P\left(\frac{X-180}{5} > \frac{170-180}{5}\right) = P(Z > -2) = P(Z < 2) = \boxed{0.9772}$

b) $P(X < 150) = P\left(\frac{X-180}{5} < \frac{150-180}{5}\right) = P(Z < -6) = P(Z > 6) = 1 - \overbrace{P(Z < 6)}^{\simeq 1}$

$\boxed{\simeq 0}$

10. Una empresa sabe que el comportamiento en probabilidad de la demanda aleatoria de un artículo que produce, viene explicada por la ley $N(\mu = 10000, \sigma = 100)$. La empresa decide seguir produciendo el artículo en el futuro, sólo si la demanda está comprendida entre 9930 y 10170 unidades, determinar la probabilidad de que no siga produciendo tal artículo.

$P(\text{"No producir el artículo"}) = P(X \notin (9930, 10170)) = 1 - P(9930 < X < 10170)$

$P(9930 < X < 10170) = P\left(\frac{9930-10000}{100} < \frac{X-10000}{100} < \frac{10170-10000}{1000}\right) =$

$P(-0.7 < Z < 1.7) = P(Z < 1.7) - P(Z > -0.7) = P(Z < 1.7) - (1 - P(Z < 0.7))$

$= 0.9554 - (1 - 0.7580) = 0.7134$

$(*) = 1 - 0.7134 = \boxed{0.2866}$

11. Sabiendo que la demanda aleatoria de gasolina, durante un cierto período de tiempo, se comporta con arreglo a la ley normal de media 150000 litros, con desviación típica igual a 10000 litros, determinar la cantidad que hay que tener dispuesta a la venta en dicho período, para poder satisfacer la demanda con una probabilidad de 0,95.

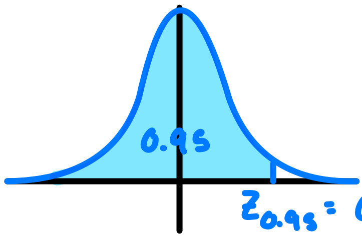
X : "Demanda de gasolina en litros" $\sim N(\mu = 150000, \sigma = 10000)$

Me piden el punto a tal que $P(X \leq a) = 0.95$

$P(X \leq a) = 0.95 \longrightarrow P\left(\underbrace{\boxed{\frac{X-150000}{10000}}}_{Z \sim N(0,1)} \leq \frac{a-150000}{10000}\right) = 0.95 \longleftrightarrow$

$P\left(Z \leq \frac{a-150000}{10\,000}\right) = 0.95$

$z_{0.95} = 1.645$



Luego $\dfrac{a - 150000}{10000} = 1.645 \longrightarrow$ $z_{0.95}$ = Cuartil 0.95 de la N(0,1)

$a = 1.645 \cdot 10000 + 150000 = \boxed{166450 \text{ litros}}$

12. Un establecimiento comercial dispone a la venta diariamente en una de sus secciones, sólo dos artículos a precios p_1 y p_2, de forma que:
    - el 70% de las unidades ofrecidas, lo son del artículo de precio p_1.
    - el 30% de las unidades ofrecidas, lo son del artículo de precio p_2.

    Si en un día determinado, se venden en dicha sección 2000 unidades, determinar la probabilidad de que más de 800 unidades correspondan al artículo de precio p_2.

X : "Nº de unidades vendidas a precio p_2 de entre las 2000 a la venta".

$X \sim B(n = 2000, p = 0.3)$ $Sop(X) = \{0, 1, 2, \dots, 2000\}$

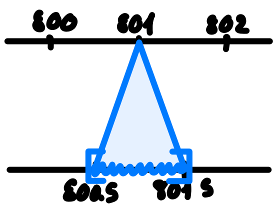
P("Vender una unidad a precio p_2") = 0.30 (éxito)

Me piden $P(X > 800) \rightarrow$ como es muy largo aproximo a una Normal

$W \sim N(\mu = n \cdot p = 600, \ \sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{420})$

Observar que $X \approx W$ porque se cumple
aproximadamente

$P(X > 800) \approx P(W > 800.5)$
corrección por continuidad



B(n, p)

$= P(z > 9.78) =$

N(μ, σ) $1 - \underbrace{P(z < 9.78)}_{\simeq 1} \boxed{\simeq 0}$

13. Un agricultor desea vender 4000 kg de naranjas y 1000 kg de limones. El precio del kilo de naranjas sigue una distribución $N(\mu = 0{,}31 \text{ euros}, \sigma = 0{,}06 \text{ euros})$ mientras que el precio del kilo de limones sigue una distribución $N(\mu = 0{,}43 \text{ euros}, \sigma = 0{,}16 \text{ euros})$ independiente de la anterior. ¿Cuál es la probabilidad de que el importe de la venta sea inferior a 1500 euros?

$$\left.\begin{array}{l} X_1 = \text{"Precio del kg de naranjas"} \sim N(\mu = 0.31, \sigma = 0.06) \\ X_2 = \text{"Precio del kg de limones"} \sim N(\mu = 0.43, \sigma = 0.16) \end{array}\right\} \text{Independiente}$$

Y = " Precio venta de 4000 kg naranjas y 1000 kg limones " $= 4000X_1 + 1000X_2$

Por reproductividad de la Normal sabemos que

$Y \sim N(\mu_Y = 4000 \cdot 0.31 + 1000 \cdot 0.43,\ \sigma_Y = \sqrt{83200} = 288.44) \longrightarrow \boxed{N(\mu_Y = 1670, \sigma = 288.44)}$

$Var(Y) = Var(4000X_1 + 1000X_2) = Var(4000X_1) + Var(1000X_2) = 4000^2 Var(X_1) +$

$1000^2 Var(X_2) = 4000 \cdot 0.06^2 + 1000^2 \cdot 0.16^2 = 83200$

Me piden $P(Y < 1500) = P\left(\frac{Y - 1670}{288.44} < \frac{1500 - 1670}{288.44}\right) = P(Z < -0.59) =$

$1 - P(Z < 0.59) = 1 - 0.7224 = \boxed{0.277}$

14. El tiempo de vida de un circuito integrado particular tiene una distribución exponencial con media de 2 años.
    a) Hallar la probabilidad de que el circuito dure más de 3 años.
    b) Suponiendo que el circuito tiene actualmente 4 años y aún funciona, calcular la probabilidad de que funcione 3 años más.

X = " Tiempo de vida en años del circuito " $\sim Exp(\lambda = 1/2)$

$E(X) = 2$ años

$E(X) = 1/\lambda$ para Exponencial

a) $P(X > 3) = 1 - P(X \leq 3) = 1 - F(3) = 1 - (1 - e^{-\frac{1}{2} \cdot 3}) = e^{-3/2}$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{en el resto} \end{cases} \longrightarrow F(x) = 1 - e^{-\lambda x} \text{ si } x > 0$$

$$P(X > 3) = \int_3^{+\infty} \frac{1}{2} e^{-\frac{1}{2}x} dx = \left[-e^{-1/2 x}\right]_3^{+\infty} = 0 - (-e^{-1/2 \cdot 3}) = e^{-3/2}$$

b) $P(X > 7 / X > 4) = P(X > 3) = e^{-3/2}$

Tema 7: Vectores aleatorios

7.1) Introducción

Una variable aleatoria multidimensional o vector aleatorio es una aplicación $\vec{x} : \Omega \longrightarrow \mathbb{R}$ tal que cada componente

$$w \longrightarrow \vec{x}(w) = (X_1(w), \ldots, X_n(w))$$

x_i es una variable aleatoria unidimensional.
Para $n = 2$, usaremos (X, Y) variable aleatoria bidimensional.

- Definición

Llamaremos distribución conjunta a la distribución de (X, Y) y llamaremos marginales a las distribuciones individuales de X e Y.

- Función de distribución conjunta de (X, Y)

$$F_{(X,Y)}(x, y) = P(X \leq x, Y \leq y) = P\left((X \leq x) \cap (Y \leq y)\right) \; \forall x, y \in \mathbb{R}^2$$

- Función de distribución marginal de X:

$$F_X(x) = P(X \leq x) = F_{(X,Y)}(x, +\infty)$$

- Función de distribución marginal de Y:

$$F_Y(y) = P(Y \leq y) = F_{(X,Y)}(+\infty, y)$$

7.2) Distribución conjunta y marginales

A) Caso discreto

- Definición

Sean X e Y variables aleatorias discreta, se define la función puntual de probabilidad (función puntual de probabilidad) conjunta.

$$P_{(X,Y)}(x, y) = P(X = x, Y = y) = P\left((X = x) \cap (Y = y)\right) \; \forall x, y \in \mathbb{R}^2$$

- Propiedades

1) $P_{(X,Y)}(x,y) \geq 0 \; \forall x,y \in \mathbb{R}^2$

2) $\displaystyle\sum_{(x,y)\in \text{Sop}(X,Y)} P_{(X,Y)}(x,y) = 1$

  - Función puntual de probabilidad conjunta: $P_{(X,Y)}(x,y) = P(X=x, Y=y) \forall (x,y) \in \mathbb{R}^2$
  - Función puntual de probabilidad marginal de X : $p_X(x) = P(X=x) = \displaystyle\sum_{y\in \text{Sop}(Y)} P_{(X,Y)}(x,y)$
  - Función puntual de probabilidad marginal de Y : $p_Y(y) = P(Y=y) = \displaystyle\sum_{x\in \text{Sop}(X)} P_{(X,Y)}(x,y)$

  En general, para calcular la probabilidad en un recinto de $\mathbb{R}^2$ haremos: $P((X,Y)\in A) = \displaystyle\sum_{(x,y)\in A} P_{(X,Y)}(x,y),\ A \subseteq \mathbb{R}^2$

- Definición

Sea $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ tal que $g(X,Y)$ es variable aleatoria unidimensional. Entonces $E(g(X,Y)) = \displaystyle\sum_{(x,y)\in \text{Sop}(X,Y)} g(x,y) \cdot P_{(X,Y)}(x,y)$

Ejemplo

$E(X \cdot Y) = \displaystyle\sum_{(x,y)\in \text{Sop}(X,Y)} x \cdot y \cdot P_{(X,Y)}(x,y)$

B) Caso continuo

(X,Y) variable aleatoria bidireccional con X e Y continuas.

- Definición

Diremos que $f_{(X,Y)} : \mathbb{R}^2 \longrightarrow \mathbb{R}$ es la función de densidad conjunta de (X,Y) si permite obtener la distribución integrando

$$F_{(X,Y)}(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{y} \left(\int_{-\infty}^{x} f_{(X,Y)}(u,v)\, \mathrm{d}u \right) \mathrm{d}v$$

La densidad conjunta permite obtener probabilidades en regiones de $\mathbb{R}^2$

$$P((X,Y)\in A) = \iint_A f_{(X,Y)}(x,y)\, \mathrm{d}x\, \mathrm{d}y$$

- Densidad marginal de X:

$$f_X(x) = \int_{y=-\infty}^{y=+\infty} f_{(X,Y)}(x,y)\,\mathrm{d}y = \int_{y\in \mathrm{Sop}(Y)} f_{(X,Y)}(x,y)\,\mathrm{d}y = \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y)\,\mathrm{d}y$$

- Densidad marginal de Y:

$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x,y)\,\mathrm{d}x$$

- Función de distribución marginal de X : $F_X(x) = P(X \leq x) = \displaystyle\int_{u=-\infty}^{u=x} \int_{y\in\mathbb{R}} f_{(X,Y)}(u,y)\,\mathrm{d}u\,\mathrm{d}y$

- Función de distribución marginal de Y : $F_Y(y) = P(Y \leq y) = \displaystyle\int_{v=-\infty}^{v=y} \int_{x\in\mathbb{R}} f_{(X,Y)}(x,v)\,\mathrm{d}x\,\mathrm{d}v$

- Definición

Sea $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ tal que $g(X,Y)$ es variable aleatoria unidimensional.

$$E(g(X,Y)) = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x,y) \cdot f_{(X,Y)}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

7.3) Distribuciones condicionadas e independencia

A veces, conocer el valor que toma una de las variables aporta información adicional sobre la otra variable.

$$(X,Y) \longrightarrow \text{Dependientes}$$

Otras veces no aporta nada $\longrightarrow$Independencias.

A) Caso discreto: (X,Y) con X,Y discretas

- Definición

Llamaremos función puntual de probabilidad de Y condicionada a $X = x$ a la función

$$P_{Y|X=x}(y) = \frac{P_{(X,Y)}(x,y)}{P_X(x)} \text{ con } p_X(x) > 0 \text{ y } \begin{cases} y \in \mathbb{R} \\ y \in \mathrm{Sop}(Y) \end{cases}$$

- Propiedades:

$$P_{Y|X=x}(y) \text{ es función puntual de probabilidad}$$

- Demostración

1) $P_{Y|X=x}(y) = \dfrac{P_{(X,Y)}(x,y)}{P_X(x)} \geq 0 \; \forall y$

2) $\displaystyle\sum_{y\in\mathbb{R}} P_{Y|X=x}(y) = \frac{\displaystyle\sum_y p(x,y)}{p_X(x)} = \frac{P_X(x)}{P_X(x)} = 1$

$$P_{Y|X=x}(y) = \frac{P(Y=y|X=x) = P\left((Y=y)\cap(X=x)\right)}{P(X=x)} = \frac{P_{(X,Y)}(x,y)}{P_X(x)}$$

Sirve para obtener probabilidades del tipo:

$$P(Y\in A|X=x) = \sum_{\substack{y\in A\\ y\in \mathrm{Sop}(Y)}} P_{Y|X=x}(y)$$

- Definición

Si $g : \mathbb{R} \longrightarrow \mathbb{R}$ aplicación tal que $g(Y)$ es variable aleatoria, se define la esperanza condicionada por $X = x$ como:

$$E\left(g(y)|X=x\right) = \sum_{y\in \mathrm{Sop}(Y)} g(y)\cdot P_{Y|X=x}(y)$$

B) Caso continuo: (X,Y) con X e Y continuas

- Definición

Se define la función de densidad de Y condicionada a $X = x$ como:

$$f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x,y)}{f_X(x)} \text{ si } f_X(x) > 0,\ \forall y\in\mathbb{R}$$

- Propiedades

$f_{Y|X=x}(y)$ es densidad.

- Demostración

1) $f_{Y|X=x}(y) \geq 0 \; \forall y$ por serlo numerador y denominador.

2) $\displaystyle\int_0^{+\infty} f_{Y|X=x}(y)\,\mathrm{d}y = \frac{1}{f_X(x)}\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y = \frac{f_X(x)}{f_X(x)} = 1$

Sirve para obtener probabilidades del tipo:

$$P(Y \in A|X = x) = \int_{y \in A} f_{Y|X=x}(y)\,\mathrm{d}y$$

- Definición

Si $g : \mathbb{R} \longrightarrow \mathbb{R}$ aplicación tal que $g(Y)$ es variable aleatoria, se define la esperanza condicionada por $X = x$ como:

$$E\left(g(Y)|X = x\right) = \int_{-\infty}^{+\infty} g(y) \cdot f_{Y|X=x}(y)\,\mathrm{d}y$$

C) Independencia

- Definición

Sea (X, Y) variable aleatoria bidimensional discreta. Diremos que X e Y son independientes si se cumple alguna de las siguientes condiciones equivalentes:

1) $P_{(X,Y)}(x, y) = p_X(x) \cdot p_Y(y)\ \forall (x, y) \in \mathbb{R}^2$

2) $F_{(X,Y)}(x, y) = F_X(x) \cdot F_Y(y)\ \forall (x, y) \in \mathbb{R}^2$

3) $P_{(Y|X=x)}(y) = P_Y(y)\ \forall x \in \mathrm{Sop}(X),\ \forall y \in \mathrm{Sop}(Y)$

4) $P_{(X|Y=y)}(x) = P_X(x)\ \forall x \in \mathrm{Sop}(X),\ \forall y \in \mathrm{Sop}(Y)$

- Definición

Sea (X, Y) variable aleatoria bidimensional continua. Diremos que X e Y son independientes si se cumple alguna de las siguientes condiciones equivalentes:

1) $f_{(X,Y)}(x, y) = f_X(x) \cdot p_Y(y)\ \forall (x, y) \in \mathbb{R}^2$

2) $F_{(X,Y)}(x, y) = F_X(x) \cdot F_Y(y)\ \forall (x, y) \in \mathbb{R}^2$

3) $f_{(Y|X=x)}(y) = f_Y(y)\ \forall x \in \mathrm{Sop}(X),\ \forall y \in \mathrm{Sop}(Y)$

4) $f_{(X|Y=y)}(x) = f_X(x)\ \forall x \in \mathrm{Sop}(X),\ \forall y \in \mathrm{Sop}(Y)$

- Propiedades

Si X e Y son variables aleatorias independientes, entonces:

1) $P(X \in A, Y \in B) = P\left((X \in A) \cap (Y \in B)\right) = P(X \in A) \cdot P(Y \in B)$

2) $g(X)$ y $h(Y)$ son independientes $\forall g : \mathbb{R} \longrightarrow \mathbb{R}$ y $\forall h : \mathbb{R} \longrightarrow \mathbb{R}$ tales que $g(X)$ y $h(Y)$ son variables aleatorias .

3) $E\left(g(X) \cdot h(Y)\right) = E(g(X)) \cdot E(h(Y))$

7.4) Covarianza

Son medidas de asociación lineal entre X e Y.

- Definición

$$\sigma_{xy} = \mathrm{Cov}(X,Y) = E\left((X-\mu_X)\cdot(Y-\mu_Y)\right) - E(X\cdot Y)\cdot E(X)\cdot E(Y)$$

$$\rho_{xy} = \mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\cdot\sqrt{\mathrm{Var}(Y)}} = \frac{\sigma_{xy}}{\sigma_x\cdot\sigma_x} \in [-1,1]$$

- Si $\rho_{xy} \simeq 0 \longrightarrow$ No hay relación lineal
- Si $\rho_x y \simeq 1 \longrightarrow$ Relación lineal creciente
- Si $\rho_{xy} \simeq -1 \longrightarrow$ Relación lineal decreciente

- Propiedades

1) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$

2) $\mathrm{Var}(aX+bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X,Y)$

3) Si X e Y son independientes, entonces:

$$\sigma_{xy} = \mathrm{Cov}(X,Y) = 0 \text{ y } \rho_{xy} = \mathrm{Corr}(X,Y) = 0$$

El recíproco de (3) no es cierto

4) $\mathrm{Cov}(aX+bY) = ab\,\mathrm{Cov}(X,Y)$

5) $\mathrm{Cov}(X_1+X_2,Y) = \mathrm{Cov}(X_1,Y) + \mathrm{Cov}(X_2,Y)$

7.5) Modelos multivariantes

A) Modelo multinomial, $\vec{X} \sim M(n, p_1, p_2, \ldots, p_k)$

Consideremos un experimento aleatorio con k resultados posibles, $\Omega = \{A_1, A_2, \ldots, A_k\}$. Repito el experimento n veces de forma independiente. Denotemos por $p_1 = P(A_i)$, $i = 1, \ldots, k$.

Llamaremos $X_i =$"Nº de veces que ocurre A_i en las n repeticiones"$\sim B(n, p_i)$

$\boxed{\text{Cov}(X_i, X_j) = -n \cdot p_i \cdot p_j \text{ no son independientes}}$

Entonces, el vector aleatorio

$$\vec{X} = (X_1, X_2, \ldots, X_k) \sim M(n, p_1, p_2, \ldots, p_k)$$

- Función puntual de probabilidad conjunta:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1! \cdot x_2! \cdot \quad \cdots \quad x_k!} \cdot p_1^{x_1} \cdot p_2^{x_2} \cdot \quad \cdots \quad \cdot p_k^{x_k}$$

$$p_1 + p_2 + \cdots + p_k = 1$$

$$P(X_1) + P(X_2) + \cdots + P(X_k) = n$$

B) Normal multivariante $\vec{X} \sim \vec{N}(\vec{\mu}, V)$

Su función densidad es:

$$f(x_1, \ldots, x_n) = \frac{1}{\sqrt{|V(2\pi)^n}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu}), V^{-1}(\vec{c} - \vec{y})\right\}$$

donde $\vec{x} = (x_1, \ldots, x_n)$, V es una matriz simétrica y definida positiva, V^{-1} su inversa y$|V|$ determinante de V.

$$E(X_1, \ldots, X_n) = (\mu_1, \ldots, \mu_n) = \vec{\mu}$$

- Propiedades

1) Si $\vec{X} = (X_1, \ldots, X_n) \sim N(\vec{\mu}, v) \longrightarrow x_j \sim N(\mu_j, \sqrt{v_{jj}})$

2) Si $\vec{X} \sim N(\vec{\mu}, V) \longrightarrow A\vec{X} + B \sim N(A\vec{\mu} + B, AVA') \, \forall A \in M_{m\times n}, \, B \in M_{m\times n}$

3) Si $\vec{X} \sim N(\vec{\mu}, V)$ y $\text{Cov}(X_i, X_j) = 0 \longrightarrow X_i$ y X_j son independientes.

4) Si $\vec{X} \sim N(\vec{\mu}, V) \longrightarrow \exists A$ tal que $Z = A^{-1}(\vec{X} - \vec{\mu}) \sim N(0, In)$.

7.6) Distribuciones multidimensionales $\overrightarrow{X} = (X_1, \ldots, X_n)$

En esta sección extendemos los resultados anteriores al caso de variables n-dimensionales $\vec{X} = (X_1, \ldots, X_n)$

- Función de distribución conjunta:

$$F_{\vec{X}}(x_1, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$$

- Función puntual de probabilidad conjunta:

$$P_{\vec{X}}(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

- Densidad conjunta (permite obtener la función de distribución integrando).

$$F_{\vec{X}}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\vec{X}}(t_1, \ldots, t_n) \, \mathrm{d}t_1, \ldots, \, \mathrm{d}t_u$$

- Función distribución marginal de $X_{k_1}, \ldots, X_{k_s}$:

$$F_{X_{k_1}, \ldots, X_{k_s}}(t_1, \ldots, t_s) = F_{\vec{X}}(-\infty, \ldots, t_i, \ldots, +\infty)$$

- Función puntual marginal de $X_{k_1}, \ldots, X_{k_s}$:

$$p_{k_1, \ldots, k_s}(t_1, \ldots, t_s) = \sum_{x_1, \ldots, x_n} p_{\vec{X}}(x_1, \ldots, t_i, \ldots, x_n)$$

- Densidad marginal de $X_{k_1}, \ldots, X_{k_s}$:

$$f_{k_1, \ldots, k_s}(t_1, \ldots, t_s) = \int_{\mathbb{R}^{n-s}} f_{\vec{X}})(x_1, \ldots, t_i, \ldots, x_n) \, \mathrm{d}x_1, \ldots, \, \mathrm{d}x_s$$

- Función puntual de probabilidad de $(X_{l_1}, \ldots, X_{l_s})$ condicionada a $(X_{k_1} = Z_1, \ldots, X_{l_s} = Z_s)$

$$P_{\left(X_{l_1}, \ldots, X_{l_r} | X_{k_1} = Z_1, \ldots, X_{k_s} = Z_s\right)}(X_{l_1}, \ldots, X_{l_s}) = \frac{P_{\vec{X}}(x_1, \ldots, x_i, \ldots, x_n)}{P_{k_1, \ldots, k_s}}(Z_1, \ldots, Z_s)$$

- Densidad de $(X_{l_1}, \ldots, X_{l_r})$ condicionada a $(X_{k_1} = Z_2, \ldots, X_{k_s} = Z_s)$

$$f_{\left(X_{l_1}, \ldots, X_{l_r} | X_{k_1} = Z_1, \ldots, X_{k_s} = Z_s\right)}(X_{l_1}, \ldots, X_{l_s}) = \frac{f_{\vec{X}}(x_1, \ldots, x_i, \ldots, x_n)}{f_{k_1, \ldots, k_s}(Z_1, \ldots, Z_s)}$$

- $(X_1,\dots,X_n)$ independientes:

$$\left.\begin{array}{l}\text{caso}\\ \\ \text{continuo}\end{array}\right\} f_{\vec{X}}(x_1,\dots,x_n)=f_{X_1}(x_1),\dots,f_{X_n}(x_n) \| \overbrace{P_{\vec{X}}(x_1,\dots,x_n)}^{\text{Caso discreto}}=p_{X_1}(x_1)\cdot \quad \cdots \quad \cdot p_{X_n}(x_n)$$

Consecuencia: Si $(X_1,\dots,X_m)$ independiente $\longrightarrow$ cualquier subconjunto es independiente.

$$\text{MCov}(X_1,\dots,X_n)=\begin{bmatrix} \overbrace{\text{Cov}(X_1,X_1)}^{\sigma^2_{x_1}} & \text{Cov}(X_1,X_2) & \cdots & \text{Cov}(X_1,X_n) \\ \vdots & \vdots & & \vdots \\ \text{Cov}(X_n,X_1) & \cdots & \cdots & \underbrace{\text{Cov}(X_n,X_n)}_{\sigma^2_{x_n}} \end{bmatrix}$$

$$\text{MCov}(X_1,\dots,X_n)=\begin{bmatrix} \rho_{X_1X_1} & \rho_{X_1X_2} & \cdots & \rho_{X_1X_n} \\ \vdots & \vdots & & \vdots \\ \rho_{X_nX_1} & \cdots & \cdots & \rho_{X_nX_n} \end{bmatrix}=\begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ & \ddots & & \\ & & \ddots & 1 \\ \rho_{n1} & \cdots & \cdots & \cdots \end{bmatrix}$$

7.7) Cambios de variables

$\vec{X}=(X_1,\dots,X_n)$ variable aleatoria y sea $g:\mathbb{R}^n\longrightarrow\mathbb{R}^m$ aplicación. ¿Cuándo será $g(\vec{X})$ una variable aleatoria m-dimensional?

Si g es continua $\longrightarrow \vec{Y}=g\left(\vec{X}\right)$ es una variable aleatoria.

- Caso discreto: $P_{\vec{Y}}P(Y_1=y_1,\dots,Y_n=y_n)=P(\vec{Y}=\vec{y})=P\left(g(\vec{X})=\vec{y}\right)=P\left(\vec{X}\in g^{-1}(\vec{y})\right)=\sum_{\vec{x}\in g^{-1}(\vec{y})} P_{\vec{X}}(\vec{x})$
- Caso continuo: Si $g:\mathbb{R}^n\longrightarrow\mathbb{R}^n \quad (n=m)$ biyectiva.

$$\underset{\text{Función densidad conjunta de } Y}{P_{\vec{Y}}(\vec{y})}=\underset{\text{función densidad conjunta de } X}{f_{\vec{X}}\left(g^{-1}(\vec{y})\right)}\cdot\underbrace{\begin{vmatrix} \dfrac{\partial g_1^{-1}(\vec{y})}{\partial y_1} & \cdots & \dfrac{\partial g_1^{-1}(\vec{y})}{\partial y_n} \\ \vdots & & \vdots \\ \dfrac{\partial g_n^{-1}(\vec{y})}{\partial y_1} & \cdots & \dfrac{\partial g_n^{-1}(\vec{y})}{\partial y_n} \end{vmatrix}}_{\text{determinante Jacobina}}$$

- Distribuciones X^2 de Pearson, t-Student y F de Snedecor

Veremos la construcción de estas nuevas distribuciones a partir de la Normal.

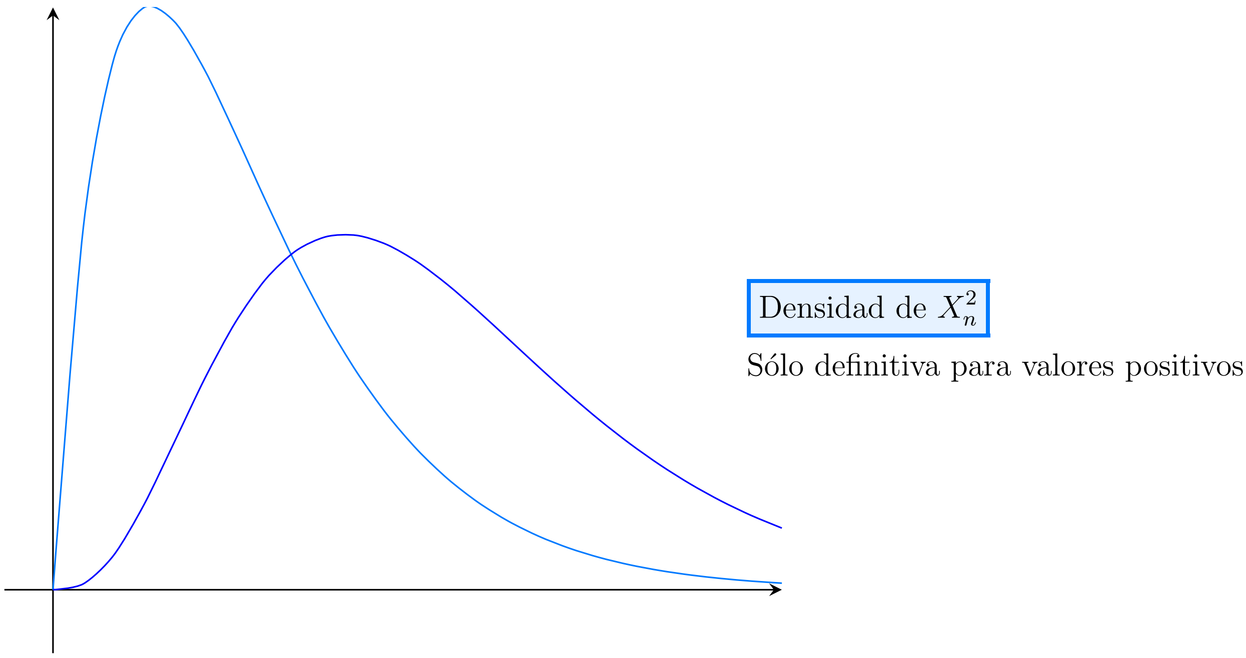
- Definición

Sean $Z_1, Z_2, \ldots, Z_n$ variable aleatoria $N(0,1)$ independientes. Entonces, se dice que la variable aleatoria:

$$\sum_{i=1}^{4} Z_i^2 \sim X_n^2 \text{ sigue una Chi-cuadrado con } n \text{ grafos libertad.}$$

– $E(X_n^2) = n, \ \mathrm{Var}(X_n^2) = 2n$

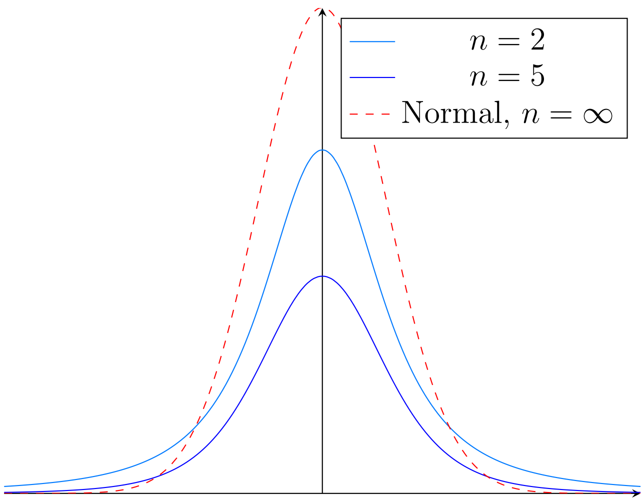
- Propiedad: Si $X \sim X_{n_1}^2$, $Y \sim Y \sim X_{n_2}^2$ independientes $\longrightarrow X + 4 \sim X_{n_1+n_2}^2$



- Definición

Sea $X \sim N(0,1)$ y sea $Y \sim X_n^2$ independientes. Entonces $\dfrac{X}{\sqrt{\frac{Y}{n}}} \sim tn$ t-Student con n grados de libertad.

– $E(t_n) = 0, \mathrm{Var}(t_n) = \dfrac{n}{n-2} \quad (\text{si } n \geq 3)$



Densidad de t_n

- Propiedad: La t_n es simétrica respecto del origen, está definida en todo $\mathbb{R}$ y si $n \geq 30$ se puede aproximar por $N(0,1)$. Para $n = 100$, t_{100} es casi idéntica a $N(0,1)$.

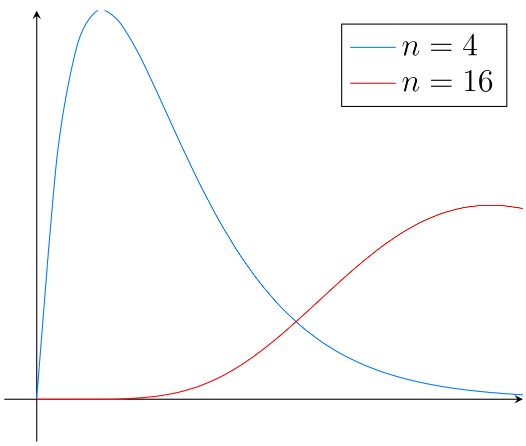
- Definición

Sea $X \sim X_n^2$ e $Y \sim X_m^2$ independiente. Entonces:

$$\frac{\frac{X}{n}}{\frac{Y}{m}} \sim F_{n,m} \qquad \text{F-Snedercor con } n \text{ y en grados de libertad.}$$

- $E(F_{n,m}) = \dfrac{m}{m-2}$ si $m > 2$.
- $\mathrm{Var}(F_{n,m}) = \dfrac{2m^2(m+n-2)}{m(m-2)^2(m-4)}$ si $m > 4$.

Está definida sólo para sólo para valores positivos



- Propiedad: Por definición $F_{n,m} = \dfrac{1}{F_{m,n}}$. Además, $t_n^2 = F_{1,n}$

Nota: Para el cálculo de probabilidades de estas distribuciones usaremos tablas o la calculadora estadística.

PROBLEMAS. RELACIÓN 7: Vectores aleatorios, independencia y modelos multivariantes
FUNDAMENTOS DE PROBABILIDAD Y ANÁLISIS EXPLORATORIO DE DATOS
GRADO EN CIENCIA E INGENIERÍA DE DATOS.

1. Determinar el valor de c de tal manera que las siguientes funciones representen funciones puntuales de probabilidad conjunta de X e Y.

$$a)p(x,y) = c \cdot x \cdot y \text{ para } x = 1,2,3 \text{ e } y = 1,2,3$$
$$b)p(x,y) = c \mid x - y \mid \text{ para } x = -2,0,2 \text{ e } y = -2,3$$

2. Dos variables aleatorias tienen la siguiente función de densidad conjunta:

$$f(x,y) = \begin{cases} k(x^2 + y^2) \ , \ 0 \leq x \leq 2 \ ; 1 \leq y \leq 4 \\ 0 \text{ en caso contrario.} \end{cases}$$

Calcular el valor de la constante k, su distribución, las densidades marginales y $P(1 < X < 2, 2 < Y < 3)$, $P(1 < X < 2)$.

3. En una red de supermercados, en un mes dado, se recogen los datos siguientes : para cada supermercado, la proporción de clientes que compra sólo una vez en el mes, y por otra parte, el beneficio del supermercado al cabo del mes. Podemos así definir dos variables aleatorias , X : proporción de clientes que compra sólo una vez en el mes; Y : beneficio del supermercado (en millones de euros), y obtenemos la siguiente función de densidad conjunta para (X, Y) :

$$f(x,y) = \begin{cases} k(x + y)e^{-y} \ , \text{ si } 0 < x < 1 \ ; \ y > 0 \\ 0 \text{ en caso contrario.} \end{cases}$$

   a) Describir el experimento aleatorio junto con el espacio muestral.
   b) Determinar el valor de k para que $f_{(X,Y)}$ sea efectivamente una función de densidad conjunta.
   c) ¿Son las v.a X y Y independientes?
   d) En un supermercado dado, se sabe que el porcentaje de clientes que compra sólo una vez en el mes es menor que 25 %, ¿cuál es la probabilidad de que el beneficio supere el millón de euros?

4. El índice anual de subida de precios X, y el índice anual de subida de salarios Y son un par de v.a. que se distribuyen conjuntamente según la función de densidad

$$f(x,y) = \begin{cases} k \cdot y^2 \cdot x \ , \text{ si } 0 < x < 7 \ ; \ 0 < y < 6 \\ 0 \text{ en caso contrario.} \end{cases}$$

   a) Determinar el valor de la constante k para que $f_{(X,Y)}$ sea efectivamente una función de densidad.
   b) Si en un año la subida de los precios ha sido del 5 %, calcular cuál habrá sido la función de densidad del índice de subida de salarios. Comparar el resultado obtenido con el que se obtendría con una subida de precios de 4 %. ¿Se puede afirmar que la subida de salarios está influida o condicionada por la subida de precios?

5. Probar que no pueden existir dos variables aleatorias X e Y tales que: $E(X) = 3$, $E(Y) = 2$, $E(X^2) = 10$, $E(Y^2) = 29$ y $E(XY) = 0$.

6. Supongamos que X e Y son dos variables aleatorias con $Var(X) = 9$, $Var(Y) = 4$ y $\rho(X,Y) = -1/6$. Calcular $Var(X + Y)$ y $Var(X - 3Y + 4)$.

7. Consideremos una variable aleatoria bidimensional (X, Y) con función de densidad conjunta:

$$f_{(X,Y)}(x,y) = \begin{cases} k \cdot y \cdot (1-x^2) & \text{si } -1 \leq x \leq 1 \text{ y si } 0 < y < 1 \\ 0 & \text{en otro caso.} \end{cases}$$

(a) Determinar el valor de la constante k para que $f_{(X,Y)}(x,y)$ sea una función de densidad.

(b) Calcular la función de densidad marginal de Y. ¿Puede decirse que X y Y son independientes?

(c) Calcular la siguiente probabilidad condicionada $\Pr(0 < X < 1 | Y \geq 0.5)$.

8. Sea (X, Y) una variable aleatoria bidimensional discreta con función puntual de probabilidad conjunta:

	Y=-1	Y=0	Y=1
X=-1	1/8	1/8	1/8
X=0	1/8	0	1/8
X=1	1/8	1/8	1/8

a) Calcular las f.p.p. marginales de X e Y.

b) Calcular el coeficiente de correlación de Pearson entre X e Y.¿Existe relación lineal entre X e Y? ¿Puede afirmarse que son independientes?

c) Calcular $\Pr(Y > -1 \mid X > -1)$

9. Cierta enfermedad genética se caracteriza por la carencia de una enzima. Algunas personas son portadoras de la enfermedad, lo que significa que pueden transmitirla potencialmente a sus hijos. De acuerdo con las leyes de la herencia genética, un hijo cuyos padres son portadores de dicha enfermedad tiene probabilidad 0.25 de no tener la enfermedad, 0.5 de ser portador y 0.25 de padecer la enfermedad. En una muestra de diez hijos de portadores de dicha enfermedad ¿cuál es la probabilidad de que tres no la tengan, cinco sean portadores y dos la padezcan?

10. De los clientes que compran cierto tipo de ordenadores, el 20 % piden una tarjeta gráfica actualizada, el 30 % piden memoria extra, el 15 % piden tanto tarjeta gráfica actualizada como memoria extra y el resto no pide nada. Se eligen de forma aleatoria 15 clientes. Calcular:

a) Probabilidad de que, entre los 15 clientes seleccionados, 3 pidan tarjeta gráfica, 4 pidan memoria extra, 2 pidan tanto tarjeta gráfica como memoria y el resto no pida nada.

b) Probabilidad de que, entre los 15 clientes seleccionados, 3 pidan tarjeta gráfica.

1. Determinar el valor de c de tal manera que las siguientes funciones representen funciones puntuales de probabilidad conjunta de X e Y.

$$a) p(x,y) = c \cdot x \cdot y \text{ para } x = 1,2,3 \text{ e } y = 1,2,3$$
$$b) p(x,y) = c \mid x - y \mid \text{ para } x = -2,0,2 \text{ e } y = -2,3$$

$$\boxed{\text{f.p.p conjunta} \longleftrightarrow \begin{array}{l} 1)\ P_{(X,Y)}(x,y) \geq 0 \quad \forall (x,y) \in \mathbb{R}^2 \\ 2)\ \sum_{(x,y)\in Sop(X,Y)} P_{(X,Y)}(x,y) = 1 \end{array}}$$

a) $p(x,y) = c \cdot x \cdot y \qquad x = 1,2,3 \ ; \ y = 1,2,3$

1. Se cumple $p(x,y) \geq 0 \quad \forall (x,y) \longleftrightarrow c \geq 0$

2. $\sum p(x,y) = 1 \longrightarrow c(1 \cdot 1 + 1\cdot 2 + 1\cdot 3 + 2\cdot 1 + 2\cdot 2 + 2\cdot 3 + 3\cdot 1 + 3\cdot 2 + 3\cdot 3) = 1 \longrightarrow$

$c \cdot 36 = 1 \longrightarrow \boxed{c = \frac{1}{36}}$

Calcular $P(X \leq 2, Y = 3)$

$P(X \leq 2, Y = 3) = P_{(X,Y)}(1,3) + P_{(X,Y)}(2,3) = \frac{1}{36} \cdot 1 \cdot 3 + \frac{1}{36} \cdot 2 \cdot 3 = \frac{9}{36} = \boxed{\frac{1}{4}}$

Calcular la f.p.p marginal de X:

$P_X(x) = P(X = x) = \sum_{y \in Sop(Y)} P_{(X,Y)}(x,y)$

$Sop(X) = \{1,2,3\}$

$P_X(1) = P(X=1) = p(1,1) + p(1,2) + p(1,3) = \frac{1}{36}(1\cdot 1 + 1\cdot 2 + 1\cdot 3) = \frac{1}{6}$

$P_X(2) = P(X=2) = p(2,1) + p(2,2) + p(2,3) = \frac{1}{36}(2\cdot 1 + 2\cdot 2 + 2\cdot 3) = \frac{1}{3}$

$P_X(3) = P(X=3) = p(3,1) + p(3,2) + p(3,3) = \frac{1}{36}(3\cdot 1 + 3\cdot 2 + 3\cdot 3) = \frac{1}{2}$

2. Dos variables aleatorias tienen la siguiente función de densidad conjunta:

$$f(x,y) = \begin{cases} k(x^2+y^2)\ ,\ 0 \leq x \leq 2\ ; 1 \leq y \leq 4 \\ 0 \text{ en caso contrario.} \end{cases}$$

Calcular el valor de la constante k, su distribución, las densidades marginales y $P(1 < X < 2, 2 < Y < 3)$, $P(1 < X < 2)$.

Nota: $f(x,y)$ es densidad conjunta si:

(1) $f(x,y) \geq 0 \quad \forall (x,y) \in \mathbb{R}^2$

(2) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = 1$

(i) Calcular "K"

1) Se cumple que $f(x,y) \geq 0 \ \forall (x,y) \longleftrightarrow K \geq 0$

2) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = \int_{y=1}^{y=4}\int_{x=0}^{x=2} K(x^2+y^2)\,dx\,dy = K\int_1^4 \left[\frac{x^3}{3}+y^2x\right]_0^2 =$

$K\int_1^4 \frac{8}{3} + 2y^2\,dy = K\left[\frac{8}{3}y + \frac{2y^3}{3}\right]_1^4 = 50K = 1 \longrightarrow \boxed{K = \frac{1}{50}}$

función densidad

- Función distribución $F(x,y) = P(X \leq x, Y \leq y)$
- Si $x \leq 0$, o bien $y \leq 1 \rightarrow F(x,y) = 0$
- Si $x \in [0,2]$, $y \in [1,4]$

$F(x,y) = \int_{v=1}^{v=y}\left(\int_{u=0}^{u=x} \frac{1}{50}(u^2+v^2)\,du\right)dv = \frac{1}{50}\int_1^y \left[\frac{u^3}{3}+uv^2\right]_0^x dv =$

$\frac{1}{50}\int_1^y \frac{x^3}{3} + xv^2\,dv = \frac{1}{50}\left[\frac{x^3}{3}v + x\frac{v^3}{3}\right]_1^y = \frac{1}{50}\left(\frac{yx^3}{3} + \frac{x^3}{3} - \frac{xy^3}{3} - \frac{x}{3}\right)$

- Si: $x \geq 2$, $y \geq 4 \longrightarrow F(x,y) = 1$

— Si $x \geq 2$, $y \in [1,4] \longrightarrow F(x,y) = F(2,y) = \frac{1}{50}\left(\frac{8y}{3} + \frac{8}{3} - \frac{2y^3}{3} - \frac{2}{3}\right)$

— Si $x \in [0,2]$, $y \geq 4 \longrightarrow F(x,y) = F(x,4) = \frac{1}{50}\left(\frac{4x^3}{3} + \frac{x^3}{3} - \frac{64x}{3} - \frac{x}{3}\right)$

- Densidad marginal de x

$f_X(x) = \int_{-\infty}^{+\infty} f(x,y) = \int_1^4 \frac{1}{50}(x^2+y^2)dy = \frac{1}{50}\left[x^2y + \frac{y^3}{3}\right]_1^4 = \frac{1}{50}\left(4x^2 + \frac{64}{3} - x^2 - \frac{1}{3}\right)$

$$f_X(x) = \begin{cases} 1/50\,(3x^2+21) & \text{si } x \in [0,2] \\ 0 & \text{en el resto} \end{cases}$$

- Densidad marginal Y:

$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx$

- Si $y \notin [1,4] \longrightarrow f_Y(y) = 0$
- Si $y \in [1,4] \longrightarrow f_Y(y) = \int_0^2 \frac{1}{50}(x^2+y^2)\,dx = \frac{1}{50}\left[\frac{x^3}{3} + y^2 x\right]_0^2 = \frac{1}{50}\left(\frac{8}{3} + 2y^2\right)$

3. En una red de supermercados, en un mes dado, se recogen los datos siguientes : para cada supermercado, la proporción de clientes que compra sólo una vez en el mes, y por otra parte, el beneficio del supermercado al cabo del mes. Podemos así definir dos variables aleatorias , X : proporción de clientes que compra sólo una vez en el mes; Y : beneficio del supermercado (en millones de euros), y obtenemos la siguiente función de densidad conjunta para (X,Y) :

$$f(x,y) = \begin{cases} k(x+y)e^{-y} \text{ , si } 0 < x < 1 \text{ ; } y > 0 \\ 0 \text{ en caso contrario.} \end{cases}$$

   a) Describir el experimento aleatorio junto con el espacio muestral.
   b) Determinar el valor de k para que $f_{(X,Y)}$ sea efectivamente una función de densidad conjunta.
   c) ¿Son las v.a X y Y independientes?
   d) En un supermercado dado, se sabe que el porcentaje de clientes que compra sólo una vez en el mes es menor que 25 %, ¿cuál es la probabilidad de que el beneficio supere el millón de euros?

Experimento aleatorio : " Seleccionar al alzar un supermercado de la red".

X = " proporción de clientes que compran una vez al mes "

Y = " beneficio supermercado (millones) "

$$f(x,y) = \begin{cases} k(x+y)e^{-y} & 0 < x < 1, \ y > 0 \\ 0 & \text{caso contrario} \end{cases}$$

a) ¿k?

i) $f(x,y) \geq 0 \quad \forall (x,y) \in \mathbb{R}^2$, cierto si $k \geq 0$

ii) $\iint_R f(x,y) = 1$

b) $\iint_{\mathbb{R}^2} f(x,y)\,dx\,dy = k\int_0^1 \underbrace{\int_0^{+\infty} (x+y)e^{-y}\,dy}_{I_1}\,dx = k\int_0^1 (x+1)\,dx = k\left[\frac{x^2}{2} + x\right]_0^1 =$

$\frac{3}{2}k = 1 \longleftrightarrow \boxed{k = \frac{2}{3}}$

$I_1 = \int_0^{+\infty} x e^{-y}\,dy + \int_0^{+\infty} y e^{-y}\,dy = x\left[-e^{-y}\right]_0^{+\infty} + \Gamma(2) = x + 1$

c) ¿ X e Y independientes?

- Marginal de X: $f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy$
  - Si $x \notin (0,1)$ — $f_X(x) = 0$
  - Si $x \in (0,1) \rightarrow f_X(x) = \int_0^{+\infty} \frac{2}{3}(x+y)e^{-y} = \frac{2}{3}\Big(\overbrace{\int_0^{+\infty} x e^{-y}}^{x} + \overbrace{\int_0^{+\infty} y e^{-y}}^{1}\Big) = \boxed{\frac{2}{3}(x+1)}$

- Marginal de Y: $f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx$
  - Si $y \leq 0 \rightarrow f_Y(y) = 0$
  - Si $y > 0 \rightarrow f_Y(y) = \int_0^1 \frac{2}{3}(x+y)e^{-y}dx = \frac{2}{3}\left(\int_0^1 x e^{-y}dx + \int_0^1 y e^{-y}dx\right) =$

$$\frac{2}{3}\left(\left[\frac{x^2}{2}e^{-y}\right]_0^1 + \left[xye^{-y}\right]_0^1\right) = \frac{2}{3}\left(\frac{1}{2}e^{-y} + ye^{-y}\right)$$

$$= \boxed{\frac{2}{3}e^{-y}\left(y + \frac{1}{2}\right)}$$

Vemos que $f(0.25, 2) \neq f_X(0.25) \cdot f_Y(2) \longrightarrow X$ e Y no son independientes

d) Me pide $P(Y > 1 \mid X < 0.25) = \dfrac{\overbrace{P((X<0.25) \cap (Y>1))}^{\text{Numerador}}}{\underbrace{P(X<0.25)}_{\text{Denominador}}} = (*)$

Numerador $= \int_0^{0.25}\int_1^{+\infty} \frac{2}{3}(x+y)e^{-y}\,dy\,dx = 0.13029$

Denominador: $\int_0^{0.25} \frac{2}{3}(x+1)\,dx = \frac{2}{3}\left[\frac{x^2}{2} + x\right]_0^{0.25} = 0.1875$

$(*)$: $\dfrac{0.13029}{0.1875} = \boxed{0.69488}$

4. El índice anual de subida de precios X, y el índice anual de subida de salarios Y son un par de v.a. que se distribuyen conjuntamente según la función de densidad

$$f(x,y) = \begin{cases} k \cdot y^2 \cdot x \text{ , si } 0 < x < 7 \text{ ; } 0 < y < 6 \\ 0 \text{ en caso contrario.} \end{cases}$$

a) Determinar el valor de la constante k para que $f_{(X,Y)}$ sea efectivamente una función de densidad.

b) Si en un año la subida de los precios ha sido del 5 %, calcular cuál habrá sido la función de densidad del índice de subida de salarios. Comparar el resultado obtenido con el que se obtendría con una subida de precios de 4 %. ¿Se puede afirmar que la subida de salarios está influida o condicionada por la subida de precios?

a) Para que $f_{(X,Y)}$ sea una función de densidad tiene que cumplir dos condiciones, de la primera deducimos que k debe de ser positiva y de la segunda :

$$\begin{aligned}
1 &= \int_R \int_R f_{(X,Y)}(x,y)dydx = \int_{x=0}^{7} \int_{y=0}^{+6} ky^2 xdydx \\
&= k \int_{x=0}^{7} x \left(\int_{y=0}^{6} y^2 dy \right) dx = k \int_{x=0}^{7} x \left[\frac{y^3}{3} \right]_0^6 dx \\
&= k \frac{6^3}{3} \left[\frac{x^2}{2} \right]_0^7 = k \cdot 1764
\end{aligned}$$

Por lo tanto, tenemos que $\boxed{k = \frac{1}{1764}}$

b) Nos piden la densidad condicional de Y dado que $(X = 5)$, y compararla con la densidad condicional de Y dado que $(X = 4)$. Para ello tenemos que calcular primero la función de densidad marginal de X. Si $x \notin (0,7)$, entonces $f_X(x) = 0$. En cambio, si $0 < x < 7$,

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)dy = \frac{1}{1764} \int_0^{+6} y^2 \cdot xdy \\
&= \frac{1}{1764} x \left[\frac{y^3}{3} \right]_0^6 = \frac{2x}{49}
\end{aligned}$$

o sea

$$f_X(x) = \begin{cases} \frac{2x}{49} \text{ si } 0 < x < 7 \\ 0 \text{ en caso contrario} \end{cases}$$

Nos será útil calcular la densidad marginal de Y. Procedemos de manera similar al cálculo anterior, si $y \notin (0,6)$ entonces $f_Y(y) = 0$. En cambio, si $y \in (0,6)$,

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)dx = \frac{1}{1764} \int_0^{+7} y^2 \cdot xdx \\
&= \frac{1}{1764} y^2 \left[\frac{x^2}{2} \right]_0^7 = \frac{1}{72} y^2
\end{aligned}$$

o sea

$$f_Y(y) = \begin{cases} \frac{1}{72} y^2 \text{ si } 0 < y < 6 \\ 0 \text{ en caso contrario} \end{cases}$$

Por definición, para x tal que $f_X(x) > 0$, se define la densidad condicionada $f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x,y)}{f_X(x)}, \forall y$. Por lo tanto, si $0 < x < 7$, se tiene que:

$$f_{Y|X=x}(y) = \begin{cases} \frac{y^2 x}{1764} \cdot \frac{49}{2x} = \frac{1}{72} y^2 \text{ si } 0 < y < 6 \\ 0 \text{ en caso contrario} \end{cases}$$

Observamos que la densidad condicionada por $(X = x)$ no depende del valor x, es decir, que $f_{Y|X=5}(y) = f_{Y|X=4}(y) \forall y$. Además comprobamos que la densidad de Y condicionada por $(X = x)$ coincide con la marginal de Y, es decir, $f_{Y|X=x}(y) = f_Y(y) \forall y$

Por lo tanto, podemos afirmar que la subida de salarios no está condicionada por la subida de los precios, o lo que es lo mismo, que las v.a. X e Y son independientes. También se puede comprobar que $f(x,y) = f_X(x) f_Y(y)$.

5. Probar que no pueden existir dos variables aleatorias X e Y tales que: $E(X) = 3$, $E(Y) = 2$, $E(X^2) = 10$, $E(Y^2) = 29$ y $E(XY) = 0$.

- $Cov(X,Y) = E(XY) - E(X)\cdot E(Y) = 0 - 3\cdot 2 = \boxed{-6}$
- $\rho(X,Y) = \dfrac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y} = \dfrac{-6}{1\cdot 5} = -\dfrac{6}{5} \notin [-1,1]$ Imposible

$Var(X) = E(X^2) - (E(X))^2 = 10 - 3^2 = 1 \longrightarrow \sigma_X = \sqrt{1} = 1$

$Var(Y) = E(Y^2) - (E(Y))^2 = 29 - 2^2 = 25 \longrightarrow \sigma_Y = \sqrt{25} = 5$

$$S_{xy} = \overline{xy} - \bar{x}\cdot\bar{y}$$

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

$$S_x^2 = \overline{x^2} - (\bar{x})^2$$

6. Supongamos que X e Y son dos variables aleatorias con $Var(X) = 9$, $Var(Y) = 4$ y $\rho(X,Y) = -1/6$. Calcular $Var(X+Y)$ y $Var(X-3Y+4)$.

- Calcular $Var(X+Y)$

$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y) = 9 + 4 - 2 = \boxed{11}$

$Cov(X,Y) = \rho(x,y)\cdot\sigma_x\cdot\sigma_y = -\frac{1}{6}\cdot 3\cdot 2 = -1$

- $Var(X-3Y+4) = Var(X-3Y) = Var(X) + Var(-3Y) + 2Cov(X,-3Y)$

$= Var(X) + 3^2 Var(Y) + 2\cdot(-3)\cdot Cov(X,Y) = 9 + 9\cdot 4 - 6\cdot(-1) = \boxed{51}$

7. Consideremos una variable aleatoria bidimensional (X,Y) con función de densidad conjunta:

$$f_{(X,Y)}(x,y) = \begin{cases} k\cdot y\cdot(1-x^2) & \text{si } -1 \leq x \leq 1 \text{ y si } 0 < y < 1 \\ 0 & \text{en otro caso.} \end{cases}$$

(a) Determinar el valor de la constante k para que $f_{(X,Y)}(x,y)$ sea una función de densidad.

(b) Calcular la función de densidad marginal de Y. ¿Puede decirse que X y Y son independientes?

(c) Calcular la siguiente probabilidad condicionada $\Pr(0 < X < 1 | Y \geq 0.5)$.

$$f_{(X,Y)}(x,y) = \begin{cases} ky(1-x^2) & \text{si } -1 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

a) ¿k?

i) $f_{(X,Y)}(x,y) \geq 0 \quad \forall (x,y) \in \mathbb{R}^2 \longleftrightarrow k \geq 0$

ii) $\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = 1 \longleftrightarrow \int_0^1\left(\int_{-1}^1 ky(1-x^2)\,dx\right)dy =$

$\displaystyle\int_0^1 ky\left[x - \frac{x^3}{3}\right]_{-1}^1 dy = \int_0^1 ky\left(1 - \frac{1}{3} - \left(-1 + \frac{1}{3}\right)\right)dy = \frac{4}{3}k\int_0^1 y\,dy =$

$\displaystyle\frac{4}{3}k\left[\frac{y^2}{2}\right]_0^1 = \frac{2}{3}k = 1 \longleftrightarrow \boxed{k = \frac{3}{2}}$

b) Marginal de X:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy$$

Si $x \in [-1, 1] \rightarrow f_X(x) = \frac{3}{2}\int_0^1 y(1-x^2)\,dy = \frac{3}{2}\left[\frac{y^2}{2}(1-x^2)\right]_0^1 = \frac{3}{4}(1-x^2)$

- Si $x \notin [-1,1] \rightarrow f_X(x) = 0$

$$f_X(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{si } x \in [-1,1] \\ 0 & \text{en el resto} \end{cases}$$

Marginal de Y:

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx$$

- Si $y \in [0, 1] \longrightarrow f_Y(y) = \int_0^1 f(x,y)\,dx = \frac{3}{2}\int_{-1}^1 y(1-x^2)\,dx = \frac{3}{2}\,y\overbrace{\left[x - \frac{x^3}{3}\right]_{-1}^1}^{\frac{4}{3}}$

$$\frac{3}{2}\cdot\frac{4}{3}\cdot y = 2y$$

- Si $y \notin [0,1] \longrightarrow f_Y(y) = 0$

$$f_Y(y) = \begin{cases} 2y & \text{si } y \in [0,1] \\ 0 & \text{en el resto} \end{cases}$$

¿Son independientes?

Sí, porque $f_{(X,Y)}(x,y) = f_X(x)\cdot f_Y(y) \quad \forall (x,y) \in \mathbb{R}^2$

c) $P(0 < X < 1 / Y \geq 0.5) = \dfrac{P(0 < X < 1) \cap \cancel{P(Y \geq 0.5)}}{\cancel{P(Y \geq 0.5)}} = P(0 < X < 1)$

$$P(0 < X < 1) = \int_0^1 \frac{3}{4}(1-x^2)\,dx = \frac{3}{4}\left[x - \frac{x^3}{3}\right]_0^1 = \frac{3}{4}\cdot\frac{2}{3} = \boxed{\frac{1}{2}}$$

8. Sea (X,Y) una variable aleatoria bidimensional discreta con función puntual de probabilidad conjunta:

	Y=-1	Y=0	Y=1
X=-1	1/8	1/8	1/8
X=0	1/8	0	1/8
X=1	1/8	1/8	1/8

a) Calcular las f.p.p. marginales de X e Y.

b) Calcular el coeficiente de correlación de Pearson entre X e Y.¿Existe relación lineal entre X e Y? ¿Puede afirmarse que son independientes?

c) Calcular $\Pr(Y > -1 \mid X > -1)$

	$y=-1$	$y=0$	$y=1$	
$x=-1$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
$x=0$	$\frac{1}{8}$	0	$\frac{1}{8}$	$\frac{2}{8}=\frac{1}{4}$
$x=-1$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
	$\frac{3}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	

a) $P_X(x) = P(X=x) = \sum_{y \in Sop(Y)} p(x,y)$

$Sop(X) = Sop(Y) = \{-1, 0, 1\}$

- $p_X(-1) = P(X=-1) = p(-1,1) + p(-1,0) + p(-1,1) = \frac{3}{8} = p_Y(-1)$
- $p_X(0) = P(X=0) = \frac{2}{8} = p_Y(0)$
- $p_X(1) = P(X=1) = \frac{3}{8} = p_Y(1)$

b) ¿Son independientes?

¿$p(x,y) = p_X(x) \cdot p_Y(y) \quad \forall (x,y) \in Sop(X,Y)$?

¿$p(0,0) = p_X(0) \cdot p_Y(0)$? $\longleftrightarrow$ ¿$0 = \frac{2}{8} \cdot \frac{2}{8}$? No, por tanto no son independientes.

¿Correlación de Pearson de X e Y?

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y} = \boxed{0}$$ No hay relación lineal

$$Cov(X,Y) = E(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y) = 0 - 0 \cdot 0 = \boxed{0}$$

$$E(X) = -1 \cdot \frac{3}{8} + 0 \cdot \frac{2}{8} + 1 \cdot \frac{3}{8} = 0$$

$$E(Y) = -1 \cdot \frac{3}{8} + 0 \cdot \frac{2}{8} + 1 \cdot \frac{3}{8} = 0$$

$$E(X \cdot Y) = \sum_{(x,y) \in Sop(X,Y)} x \cdot y \cdot P_{(X,Y)}(x,y) = (-1)\cdot(-1)\cdot\frac{1}{8} + \cancel{(-1)\cdot 0 \cdot \frac{1}{8}} + (-1)\cdot 1 \cdot \frac{1}{8} + \cancel{0 \cdot (-1) \cdot \frac{1}{8}} + \cancel{0 \cdot 0 \cdot 0} + \cancel{0 \cdot 1 \cdot \frac{1}{8}} + 1 \cdot (-1) \cdot \frac{1}{8} + \cancel{1 \cdot 0 \cdot \frac{1}{8}} + 1 \cdot 1 \cdot \frac{1}{8} = \frac{1}{8} - \frac{1}{8} - \frac{1}{8} + \frac{1}{8} = \boxed{0}$$

c) $$P(Y > -1 \mid P(X > -1)) = \frac{P((X > -1) \cap (Y > -1))}{P(X > -1)} = \frac{p(0,0) + p(0,1) + p(1,0) + p(1,1)}{p_X(0) + p_X(1)}$$

$$= \frac{0 + 1/8 + 1/8 + 1/8}{2/8 + 3/8} = \frac{3/8}{5/8} = \boxed{\frac{3}{5}}$$

9. Cierta enfermedad genética se caracteriza por la carencia de una enzima. Algunas personas son portadoras de la enfermedad, lo que significa que pueden transmitirla potencialmente a sus hijos. De acuerdo con las leyes de la herencia genética, un hijo cuyos padres son portadores de dicha enfermedad tiene probabilidad 0.25 de no tener la enfermedad, 0.5 de ser portador y 0.25 de padecer la enfermedad. En una muestra de diez hijos de portadores de dicha enfermedad ¿cuál es la probabilidad de que tres no la tengan, cinco sean portadores y dos la padezcan?

X_1 = " Nº hijos de la muestra de tamaño 10 que no tienen la enfermedad "

X_2 = " Nº de hijos de la muestra son portadores "

X_3 = " Nº hijos de la muestra de tamaño 10 que no padece la enfermedad "

$\vec{X} = (X_1, X_2, X_3) \sim M(n=10, p_1 = 0.25, p_2 = 0.5, p_3 = 0.25)$

Me pide $P(X_1 = 3, X_2 = 5, X_3 = 2) = \dfrac{10!}{3!\,5!\,2!} \cdot 0.25^3 \cdot 0.5^5 \cdot 0.25^2 =$

10. De los clientes que compran cierto tipo de ordenadores, el 20 % piden una tarjeta gráfica actualizada, el 30 % piden memoria extra, el 15 % piden tanto tarjeta gráfica actualizada como memoria extra y el resto no pide nada. Se eligen de forma aleatoria 15 clientes. Calcular:

a) Probabilidad de que, entre los 15 clientes seleccionados, 3 pidan tarjeta gráfica, 4 pidan memoria extra, 2 pidan tanto tarjeta gráfica como memoria y el resto no pida nada.

b) Probabilidad de que, entre los 15 clientes seleccionados, 3 pidan tarjeta gráfica.

X_1 = " Nº de clientes en la muestra de 15 que piden tarjeta "

X_2 = " Nº de clientes en la muestra de 15 que piden memoria "

X_3 = " Nº de clientes en la muestra de 15 que piden ambas "

X_4 = " Nº de clientes en la muestra de 15 que no pide nada "

$\vec{X} = (X_1, X_2, X_3, X_4) \sim M(n=15, p_1 = 0.2, p_2 = 0.3, p_3 = 0.15, p_4 = 0.35)$

a) Me piden $P(X_1 = 3, X_2 = 4, X_3 = 2, X_4 = 6) =$

$\dfrac{15!}{3!\,4!\,2!\,6!} \cdot 0.2^3 \cdot 0.3^4 \cdot 0.15^2 \cdot 0.35^6 = \boxed{0.7512}$

b) Me piden $P(X_1 = 3)$

Como $X_1 \sim B(n = 15, p_1 = 0.2)$

$$P(X_1 = 3) = \binom{15}{3} \cdot 0\ 2^3 \cdot 0.8^{12} = \frac{15!}{(15-3)!} \cdot 0.2^3 \cdot 0.8^{12} = \boxed{0.2501}$$

Problemas de Repaso

PROBLEMAS DE REPASO TEMAS 1-2-3-4
FUNDAMENTOS DE PROBABILIDAD Y ANÁLISIS EXPLORATORIO DE DATOS.
GRADO EN CIENCIA E INGENIERÍA DE DATOS.

1. El responsable en control industrial de una empresa somete a un test de fiabilidad 50 dispositivos electrónicos y anota su duración en horas hasta el fallo. La recogida de los datos lleva a las siguientes frecuencias absolutas:

Duración (h)	Núm. dispositivos	Duración (h)	Núm. dispositivos
$0 \leq X < 200$	17	$800 \leq X < 1000$	6
$200 \leq X < 400$	9	$1000 \leq X < 1200$	2
$400 \leq X < 600$	7	$1200 \leq X < 1400$	1
$600 \leq X < 800$	7	$1400 \leq X < 1600$	1

   a) Determinar la tabla de frecuencias completa que contenga frecuencias absolutas, relativas y sus acumuladas.

   b) ¿Qué porcentaje de dispositivos de la muestra tienen una duración inferior a 800h?, ¿y en el intervalo $200 \leq X < 800$?

   c) Calcular la media, desviación típica y coeficiente de variación de estos datos.

   d) Justificar en qué intervalo se encuentra la mediana y cuál es el intervalo modal.

   e) Comparar la magnitud de las tres medidas de centralización (media, mediana y moda) y explicar a qué conclusión podemos llegar.

   f) Representar el histograma que resulta de la tabla del enunciado y comentar sus características más relevantes.

2. Los siguientes datos se corresponden con la dureza de unas piezas antes y después de sometarlas a un proceso térmico:

Pieza	1	2	3	4	5	6	7	8	9	10	11
Dureza antes	182	232	191	200	148	249	276	213	241	480	262
Dureza después	198	210	194	220	138	220	219	161	210	313	226

   a) Calcular la media, desviación típica, mediana y cuartiles para la dureza antes y después del proceso.

   b) Realizar un diagrama de caja y bigotes comparativo de ambas variables (dureza antes y dureza después). Describir sus características más relevantes y qué conclusiones podemos sacar para esta muestra de piezas.

   c) Se quiere buscar un modelo para explicar la dureza después del proceso en función de la dureza previa.

   (i) Representar la nube de puntos correspondiente. ¿Sería adecuado un modelo lineal?

   (ii) Obtener la ecuación de la recta ajustada por mínimos cuadrados.

   (iii) Obtener una medida de la bondad del ajuste y estimar cuál sería la dureza posterior para una pieza con dureza previa igual a 195.

3. Los siguientes datos se refieren al crecimiento de una colonia de bacterias en un medio de cultivo:

x	3	6	9	12	15	18
y	115000	147000	239000	356000	579000	864000

siendo x los días desde la inoculación e y el número de bacterias.

a) Representar y en función de x para verificar que es razonable ajustar a una curva exponencial.

b) Ajustar estos datos a una curva exponencial y dar una medida de la bondad del ajuste.

c) Estimar, usando el modelo ajustado en el apartado anterior, el número de bacterias a los 14 días.

4. Sean A y B dos sucesos cualesquiera. Se sabe que $\Pr(A) = \dfrac{1}{3}$, $\Pr(B) = \dfrac{2}{5}$ y que $\Pr(A \cup B) = \dfrac{3}{5}$. Se pide:

a) Calcular: $\Pr(A \cap B)$; $\Pr(A^c)$; $\Pr(A^c \cap B^c)$; $\Pr(A \mid B)$; $\Pr(B^c \mid A)$.

b) ¿Son A y B incompatibles?

c) ¿Son A y B independientes?

5. Una urna contiene cinco bolas numeradas del 1 al 5. Determinar:

a) La probabilidad de que al sacar tres bolas con reemplazamiento, la suma de los puntos sea impar.

b) La probabilidad de que al sacar dos bolas sin reemplazamiento, la suma de los puntos sea impar.

c) Ahora se lanza una moneda trucada, donde la probabilidad de obtener cara es de $\dfrac{2}{3}$. Si sale cara, se sacan tres bolas de la urna con reemplazamiento. Si sale cruz, se sacan dos bolas de la urna sin reemplazamiento.

(i) Determinar la probabilidad de que la suma de los puntos de las bolas extraidas sea impar.

(ii) Si la suma de las puntuaciones de las bolas extraidas resulta ser par, determinar la probabilidad de que la moneda mostrara cruz en su lanzamiento.

6. Una pieza producida en una empresa puede tener dos tipos de defectos. El 8 % de la producción presenta el defecto de tipo A, el 5 % de la producción presenta el defecto de tipo B, y se supone que no hay piezas que tengan los dos tipos de defectos. Después de ser producida cada pieza es sometida de manera automática a un test de ruptura, con las siguientes posibilidades: Si la pieza tiene el defecto de tipo A, tiene una probabilidad de 0.9 de romperse. Si la pieza tiene el defecto de tipo B, tiene una probabilidad de 0.95 de romperse. Finalmente, si la pieza no tiene ningún tipo de defecto, tiene una probabilidad de 0.01 de romperse.

a) Si el experimento aleatorio consiste en escoger al azar un pieza de la producción, traducir los datos del enunciado, después de haber introducido los sucesos convenientes.

b) ¿Cuál es la probabilidad de que una pieza escogida al azar en la producción se vaya a romper durante el test?

c) Si una pieza escogida al azar se ha roto durante el test, ¿cuál es la probabilidad de que no fuese defectuosa, es decir, que no tuviera el defecto A ni el B?

1. El responsable en control industrial de una empresa somete a un test de fiabilidad 50 dispositivos electrónicos y anota su duración en horas hasta el fallo. La recogida de los datos lleva a las siguientes frecuencias absolutas:

Duración (h)	Núm. dispositivos	Duración (h)	Núm. dispositivos
$0 \leq X < 200$	17	$800 \leq X < 1000$	6
$200 \leq X < 400$	9	$1000 \leq X < 1200$	2
$400 \leq X < 600$	7	$1200 \leq X < 1400$	1
$600 \leq X < 800$	7	$1400 \leq X < 1600$	1

a) Determinar la tabla de frecuencias completa que contenga frecuencias absolutas, relativas y sus acumuladas.

Duración (h)	Num. Dispositivos	F_i	h_i	H_i
$0 \leq X < 200$	17	17	$\frac{17}{50} = 0.34$	0.34
$200 \leq X < 400$	9	26	$\frac{9}{50} = 0.18$	0.52
$400 \leq X < 600$	7	33	$\frac{7}{50} = 0\ 14$	0.64
$600 \leq X < 800$	7	40	$\frac{7}{50} = 0.14$	0.80
$800 \leq X < 1000$	6	46	$\frac{6}{50} = 0.12$	0.92
$1000 \leq X < 1200$	2	48	$\frac{2}{50} = 0.04$	0.96
$1200 \leq X < 1400$	1	49	$\frac{1}{50} = 0.02$	0.98
$1400 \leq X < 1600$	1	50	$\frac{1}{50} = 0.02$	1
	50		1	

b) ¿Qué porcentaje de dispositivos de la muestra tienen una duración inferior a 800h?, ¿y en el intervalo $200 \leq X < 800$?

$$P(X < 800) = \frac{40}{50} \cdot 100 = 80\%$$

$$P(200 \leq X < 800) = P(200 \leq X < 400) + P(400 \leq X < 600) + P(600 \leq X < 800) =$$

$$\frac{9}{50} + \frac{7}{50} + \frac{7}{50} = \frac{23}{50} = 0.46 \longrightarrow 46\%$$

c) Calcular la media, desviación típica y coeficiente de variación de estos datos.

$$\bullet\ \bar{x} = \frac{\sum m_i \cdot f_i}{n} = \frac{100 \cdot 17 + 300 \cdot 9 + 500 \cdot 7 + 700 \cdot 7 + 900 \cdot 6 + 1100 \cdot 2 + 1300 + 1500}{50}$$

$$= \boxed{464}$$

- $s^2 = \overline{x^2} - (\bar{x})^2$, $s = \sqrt{s^2}$ desv. típica

$$\overline{x^2} = \frac{100^2 \cdot 17 + 300^2 \cdot 9 + 500^2 \cdot 7 + 700^2 \cdot 7 + 900^2 \cdot 6 + 1100^2 \cdot 2 + 1300^2 + 1500^2}{50} = 347600$$

$$s^2 = 347600 - (464)^2 = 132304$$

$$s = \sqrt{s^2} = \sqrt{132304} = \boxed{363.763}$$

- $CV = \frac{s}{\bar{x}} = \frac{363.763}{464} = \boxed{0.7839}$ Dispersión relativa del 78.39%.

d) Justificar en qué intervalo se encuentra la mediana y cuál es el intervalo modal.

- Como $n = 50 \longrightarrow Me = \frac{x_{25} + x_{26}}{2}$

La mediana será la media aritmética de los datos que ocupan las posiciones 25 y 26, si se ordenan los datos de menor a mayor.

Mirando la tabla de frecuencias de casualidad, vemos que $x_{(25)} \in [200, 400)$ y $x_{(26)} \in [200, 400) \longrightarrow$

$$Me = \frac{x_{(25)} + x_{(26)}}{2} \in [200, 400)$$

- Intervalo modal $= [0, 200)$ porque tiene mayor frecuencia.

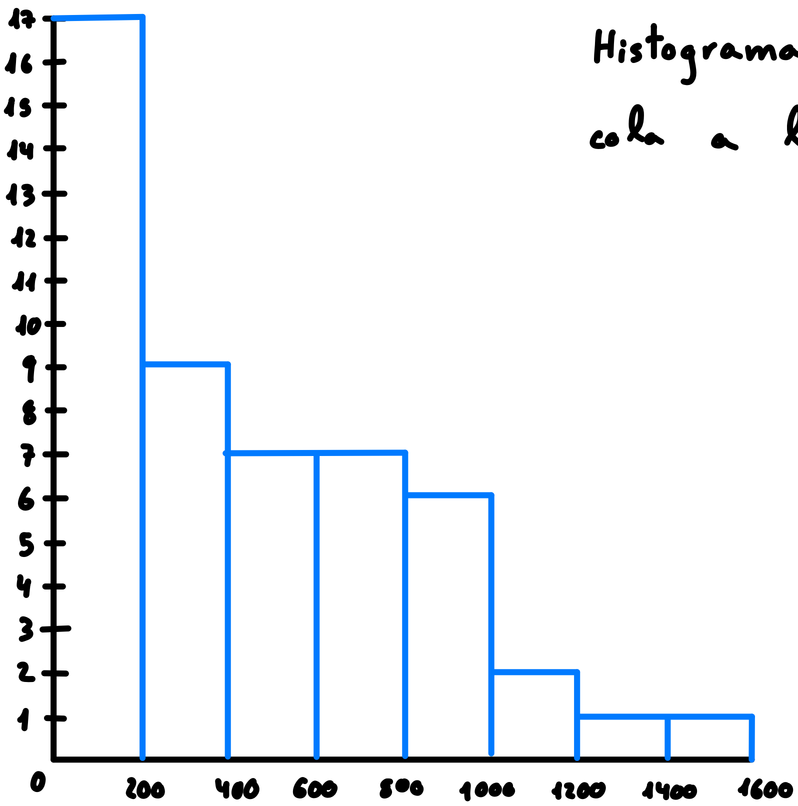
e) Comparar la magnitud de las tres medidas de centralización (media, mediana y moda) y explicar a qué conclusión podemos llegar.

Observamos que Moda < Mediana < Media

$[0, 200)$ $\quad \in [200, 400)$ $\quad 464$

lo que nos hace pensar en una distribución muy asimétrica con cola a la derecha.

f) Representar el histograma que resulta de la tabla del enunciado y comentar sus características más relevantes.



Histograma unimodal, muy asimétrico con cola a la derecha.

2. Los siguientes datos se corresponden con la dureza de unas piezas antes y después de sometarlas a un proceso térmico:

Pieza	1	2	3	4	5	6	7	8	9	10	11
Dureza antes	182	232	191	200	148	249	276	213	241	480	262
Dureza después	198	210	194	220	138	220	219	161	210	313	226

$x =$ " Dureza antes "

$y =$ " Dureza después "

a) Calcular la media, desviación típica, mediana y cuartiles para la dureza antes y después del proceso.

$$\bar{x} = \frac{182 + 232 + 191 + 200 + 148 + 249 + 276 + 213 + 241 + 480 + 262}{11} = 243.09$$

$$\bar{y} = \frac{198 + 210 + 194 + 220 + 138 + 220 + 219 + 161 + 210 + 313 + 226}{11} = 209.9$$

$$s_x^2 = \overline{x^2} + (\bar{x})^2 = 66000.36 + (243.09)^2 = 6907.17$$

$$\overline{x^2} = \frac{182^2 + 232^2 + 191^2 + 200^2 + 148^2 + 249^2 + 276^2 + 213^2 + 241^2 + 480^2 + 262^2}{11} = 66000.36$$

$$s_y^2 = \overline{y^2} + (\bar{y})^2 = 45801 + (209.9)^2 = 1739.17$$

$$\overline{y^2} = \frac{198^2 + 210^2 + 194^2 + 220^2 + 138^2 + 220^2 + 219^2 + 161^2 + 210^2 + 313^2 + 226^2}{11} = 45801$$

Para calcular los cuartiles, ordenamos de menor a mayor cada variable:

$X_{ord.}$: 148 182 191 200 213 232 241 249 262 276 480

$Q_1 = X_3$ $\quad$ $Me = X_6$ $\quad$ $Q_3 = X_9$

$Y_{ord.}$: 138 161 194 198 210 210 219 220 220 226 313

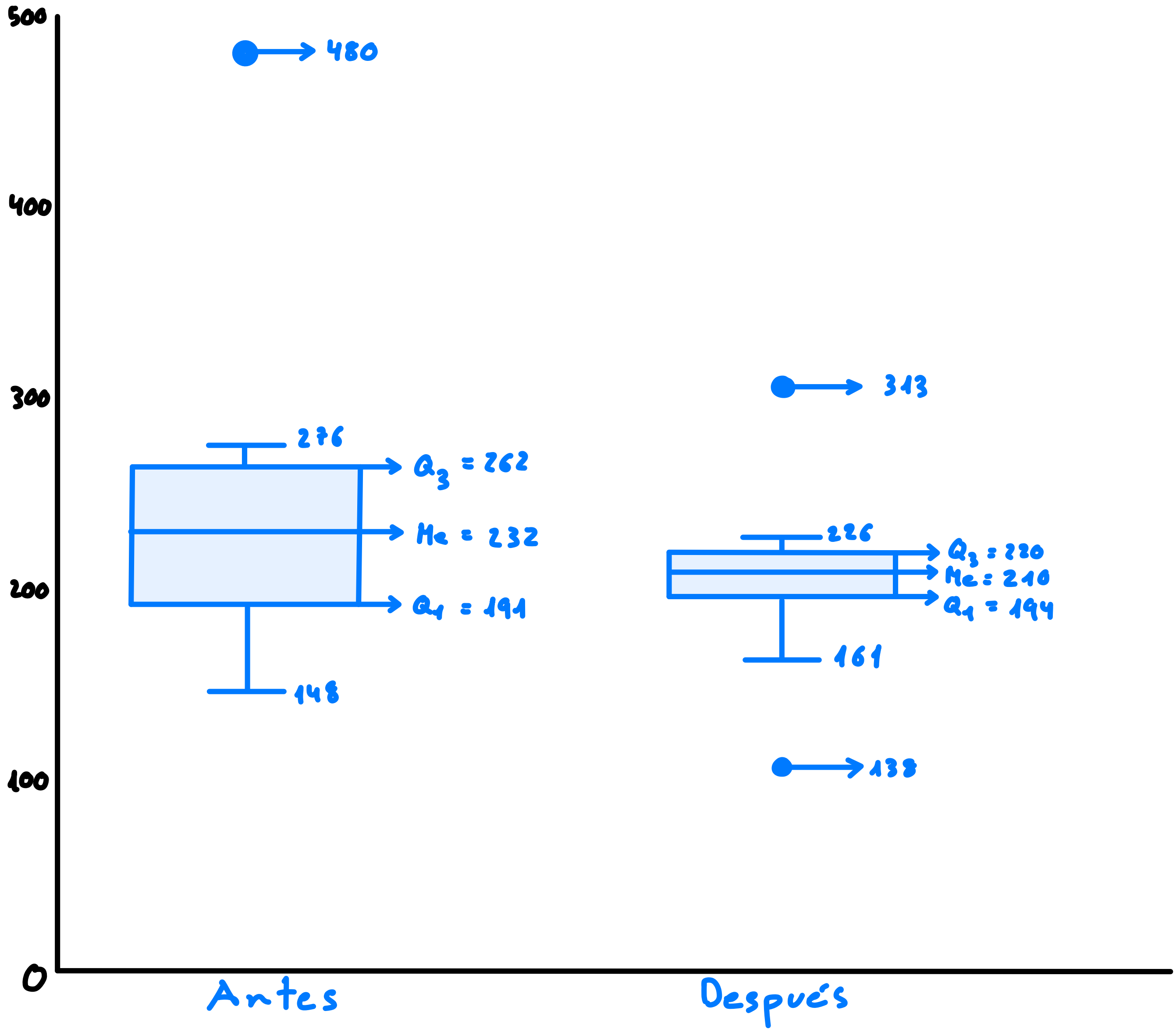
$Q_1 = Y_3$ $\quad$ $Me = Y_6$ $\quad$ $Q_3 = Y_9$

b) Realizar un diagrama de caja y bigotes comparativo de ambas variables (dureza antes y dureza después). Describir sus características más relevantes y qué conclusiones podemos sacar para esta muestra de piezas.

Calculemos los límites de admisibilidad de x e y:

para "x" $\begin{cases} LIA_x = Q_1 - 1.5\cdot(Q_3 - Q_1) = 191 - 1.5(262-191) = 84.5 \\ LSA_x = Q_3 + 1.5\cdot(Q_3 - Q_1) = 262 + 1.5(262-191) = 368.5 \end{cases}$

para "y" $\begin{cases} LIA_y = Q_1 - 1.5\cdot(Q_3 - Q_1) = 194 - 1.5(220-194) = 155 \\ LSA_y = Q_3 + 1.5\cdot(Q_3 - Q_1) = 220 + 1.5(220-194) = 259 \end{cases}$



Observamos que la dureza después del proceso térmico disminuye en cuanto a tendencia central (mediana menor) y también disminuye en cuanto a dispersión (RIC menor). Destacamos la presencia de un outlier "antes" y 2 outliers "después".

c) Se quiere buscar un modelo para explicar la dureza después del proceso en función de la dureza previa.

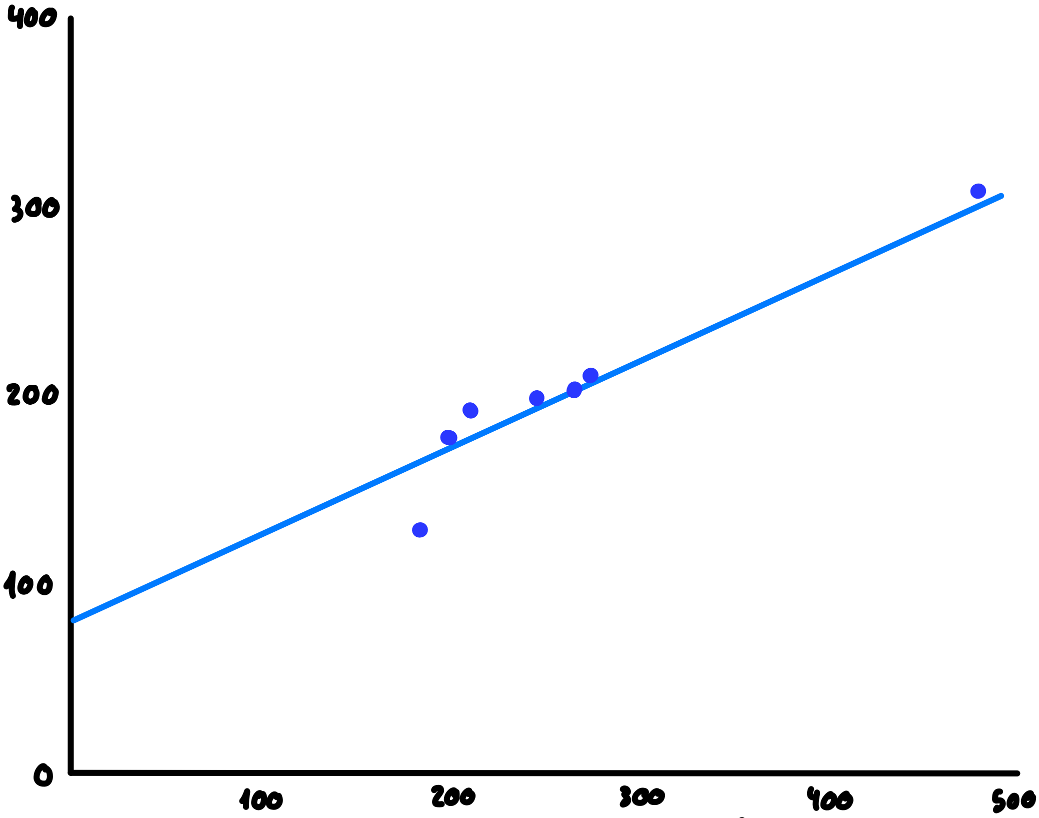
(i) Representar la nube de puntos correspondiente. ¿Sería adecuado un modelo lineal?

(ii) Obtener la ecuación de la recta ajustada por mínimos cuadrados.

(iii) Obtener una medida de la bondad del ajuste y estimar cuál sería la dureza posterior para una pieza con dureza previa igual a 195.

i) y = "despues"

x = "antes"



Parece existir cierta relación lineal creciente entre ambas variables pero especialmente influenciada por el dato atípico (antes = 480, después = 313)

Nota: Aunque no lo haremos en esta ocasión, debería analizarse la relación entre estas dos variables omitiendo el dato (480, 313)

ii) Recta ajustada: $y - \bar{y} = \dfrac{S_{xy}}{S_x^2}(x - \bar{x}) = (*)$

$S_{xy} = \overline{xy} - \bar{x} \cdot \bar{y} = 54164.82 - (243.09)\cdot(209.90) = 3137.83$

$\overline{xy} = \dfrac{182 \cdot 198 + 232 \cdot 210 + \cdots + 262 \cdot 226}{11} = 54164.82$

$(*) = y - 209.90 = \dfrac{3137.83}{6907.17}(x - 243.09) \longrightarrow \boxed{y = 0.4543x + 99.4642}$

y = "Dureza después"

x = "Dureza antes"

iii) $R^2 = R^2_{xy} \longrightarrow$ coeficiente de determinación (es el cuadrado de la correlación de Pearson)

$R_{xy} = \dfrac{S_{xy}}{S_x S_y} \longrightarrow R^2_{xy} = \dfrac{S^2_{xy}}{S_x^2 \cdot S_y^2} = \dfrac{(3137.83)^2}{(6907.17)\cdot(1739.17)} \longrightarrow \boxed{R^2_{xy} = 0.82}$

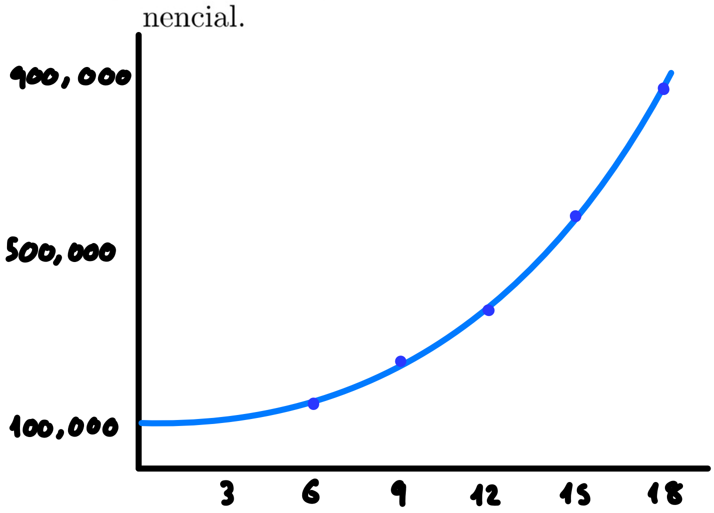
Buen ajuste, pues se interpreta como que el 82% de la variabilidad de y = "dureza después" viene explicada a partir de x = "dureza antes". El 18% restante, no vendría explicado por "x" sino por otros factores que no se hacen contemplando (errores o residuos).

3. Los siguientes datos se refieren al crecimiento de una colonia de bacterias en un medio de cultivo:

x	3	6	9	12	15	18
y	115000	147000	239000	356000	579000	864000

siendo x los días desde la inoculación e y el número de bacterias.

a) Representar y en función de x para verificar que es razonable ajustar a una curva exponencial.



x = "días desde inoculación"

y = "número de bacterias"

La nube de puntos indica que sí tiene sentido ajustar una curva exponencial. $y \simeq a e^{bx}$

b) Ajustar estos datos a una curva exponencial y dar una medida de la bondad del ajuste.

Como solo sabemos hacer ajustes lineales, transformaremos el modelo anterior para que sea lineal. Para ello, tomaremos logaritmos neperianos sobre la variable y: "nº de bacterias".

$y_{new} = \ln(y)$. Ajustamos una recta para (x, y_{new})

x	y	y_{new}
3	115000	$\ln(115000) = 11.65$
6	147000	$\ln(147000) = 11.90$
9	239000	$\ln(239000) = 12.38$
12	356000	$\ln(356000) = 12.78$
15	579000	$\ln(579000) = 13.27$
18	864000	$\ln(864000) = 13.67$

$$y_{new} - \bar{y}_{new} = \frac{S_{xy_{new}}}{S_x^2}(x - \bar{x})$$

Realizando los cálculos de todas las magnitudes de la fórmula, llegamos a:

$\boxed{y_{new} = 0.139x + 11.15}$; $y_{new} = \ln(y)$

Hacemos el cambio

$$\boxed{y = e^{0.139x + 11.15} = 69563.38 \cdot e^{0.139x}}$$

$$R^2_{xy_{new}} = \frac{S^2_{x,y_{new}}}{S_x^2 \cdot S^2_{y_{new}}} = 0.9942 \longrightarrow \text{Muy buen ajuste}$$

c) Estimar, usando el modelo ajustado en el apartado anterior, el número de bacterias a los 14 días.

Si $x = 14$ días $\longrightarrow \hat{y}_{new} = 0.139 \cdot 14 + 11.15 = 13.09584 \longrightarrow$

$\hat{y} = e^{13.09584} = \boxed{486911\ 2}$

Predicción del número de bacterias en 14 días.

4. Sean A y B dos sucesos cualesquiera. Se sabe que $\Pr(A) = \frac{1}{3}$, $\Pr(B) = \frac{2}{5}$ y que $\Pr(A \cup B) = \frac{3}{5}$. Se pide:

   a) Calcular: $\Pr(A \cap B)$; $\Pr(A^c)$; $\Pr(A^c \cap B^c)$; $\Pr(A \mid B)$; $\Pr(B^c \mid A)$.

$P(A \cup B) = P(A) + P(B) - P(A \cap B) \longrightarrow P(A \cap B) = P(A) + P(B) - P(A \cup B) =$

$\frac{1}{3} + \frac{2}{5} - \frac{3}{5} = \boxed{\frac{2}{15}}$

$P(A^c) = 1 - P(A) = 1 - \frac{1}{3} = \boxed{\frac{2}{3}}$

$P(A^c \cap B^c) = P((A \cup B)^c) = 1 - P(A \cup B) = 1 - \frac{3}{5} = \boxed{\frac{2}{5}}$

$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{2/15}{2/5} = \frac{5}{15} = \boxed{\frac{1}{3}}$

$P(B^c|A) = 1 - P(B|A) = 1 - \frac{2}{5} = \boxed{\frac{3}{5}}$

$P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{2/15}{1/3} = \frac{6}{15} = \frac{2}{5}$

b) ¿Son A y B incompatibles?

Son incompatibles si $A \cap B = \emptyset$

Como $P(A\cap B) = \frac{2}{15} \neq 0 \longrightarrow A \cap B \neq \emptyset \longrightarrow$ No son incompatibles

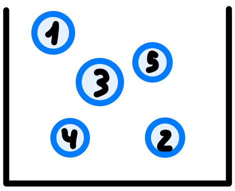
c) ¿Son A y B independientes?

A y B son independientes si $P(A\cap B) = P(A) \cdot P(B)$

$\underbrace{P(A\cap B)}_{\frac{2}{15}} = \underbrace{P(A)}_{\frac{1}{3}} \cdot \underbrace{P(B)}_{\frac{2}{5}} \longrightarrow$ Si se cumple $\longrightarrow$ A y B independientes

5. Una urna contiene cinco bolas numeradas del 1 al 5. Determinar:

   a) La probabilidad de que al sacar tres bolas con reemplazamiento, la suma de los puntos sea impar.



Urna

S = "La suma de los puntos es impar" $\longrightarrow$ suceso de interés.

Nos piden $P(S)$

Definimos los sucesos.

I_j = "El número de extracción j es impar", $j = 1, 2, 3$

Observar que $S = (I_1 \cap I_2 \cap I_3) \cup (I_1 \cap \bar{I}_2 \cap \bar{I}_3) \cup (\bar{I}_1 \cap I_2 \cap \bar{I}_3) \cup (\bar{I}_1 \cap \bar{I}_2 \cap I_3)$

siendo las anteriores uniones de sucesos disjuntos, luego sumaremos las probabilidades de cada suceso entre paréntesis. Calculemos cada uno por separado:

- $P(I_1 \cap I_2 \cap I_3) = \frac{3}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} = \left(\frac{3}{5}\right)^3$

  independientes al ser con reemplazamiento.

- $P(I_1 \cap \bar{I}_2 \cap \bar{I}_3) = \frac{3}{5} \cdot \frac{2}{5} \quad \frac{2}{5} = \frac{3}{5} \cdot \left(\frac{2}{5}\right)^2$

Observar que el resto de probabilidades coinciden con esta última:

$$P(S) = \left(\frac{3}{5}\right)^3 + 3 \cdot \frac{3}{5} \cdot \left(\frac{2}{5}\right)^2 = \frac{3^3 + 3^2 \cdot 2^2}{5^3} = \boxed{\frac{63}{125}}$$

b) La probabilidad de que al sacar dos bolas sin reemplazamiento, la suma de los puntos sea impar.

$S = (I_1 \cap \bar{I}_2) \cup (\bar{I}_1 \cap I_2)$

- $P(I_1 \cap \bar{I}_2) = P(I_1) \cdot P(\bar{I}_2 | I_1) = \frac{3}{5} \cdot \frac{2}{4} = \boxed{\frac{3}{10}}$

- $P(\bar{I}_1 \cap I_2) = P(\bar{I}_1) \cdot P(I_2 | \bar{I}_1) = \frac{2}{5} \quad \frac{3}{4} = \boxed{\frac{3}{10}}$

$$P(S) = \frac{3}{10} + \frac{3}{10} = \frac{6}{10} = \boxed{\frac{3}{5}}$$

c) Ahora se lanza una moneda trucada, donde la probabilidad de obtener cara es de $\frac{2}{3}$. Si sale cara, se sacan tres bolas de la urna con reemplazamiento. Si sale cruz, se sacan dos bolas de la urna sin reemplazamiento.

(i) Determinar la probabilidad de que la suma de los puntos de las bolas extraidas sea impar.

(ii) Si la suma de las puntuaciones de las bolas extraidas resulta ser par, determinar la probabilidad de que la moneda mostrara cruz en su lanzamiento.

Sea C: "La moneda sale cara".

$P(C) = \frac{2}{3}$, $P(\bar{C}) = \frac{1}{3}$

Observar que $\{C, \bar{C}\}$ es una partición de Ω.

- Si sale cara, estamos en el escenario del apartado (a).
- Si sale cruz, estamos en el escenario del apartado (b).

i) Observar que nos piden $P(\bar{S})$ y que tenemos:

- $P(S/G) = \frac{63}{125}$ Apartado (a)
- $P(S/\bar{G}) = \frac{3}{5}$ Apartado (b)

Usando el Teorema de Probabilidad Total tenemos:

- $P(S) = P(G) \cdot P(S/G) + P(\bar{G}) \cdot P(S|\bar{G}) = \frac{2}{3} \cdot \frac{63}{125} + \frac{1}{3} \cdot \frac{3}{5} = \boxed{\frac{67}{125}}$

ii) Nos piden $P(\bar{G}/\bar{S})$

$$P(\bar{G}|\bar{S}) = \frac{P(\bar{G}) \cdot P(\bar{S}/\bar{G})}{P(\bar{S})} = (*)$$

Teorema de Bayes

- $P(\bar{S}/\bar{G}) = 1 - P(S/\bar{G}) = 1 - \frac{3}{5} = \frac{2}{5}$
- $P(\bar{S}) = 1 - P(S) = 1 - \frac{67}{125} = \frac{58}{125}$

$$(*) = \frac{1/3 \cdot 2/5}{58/125} = \frac{2/15}{58/125} = \frac{2 \cdot 125}{58 \cdot 15} = \frac{25}{3 \cdot 29} = \boxed{\frac{25}{87}}$$

6. Una pieza producida en una empresa puede tener dos tipos de defectos. El 8 % de la producción presenta el defecto de tipo A, el 5 % de la producción presenta el defecto de tipo B, y se supone que no hay piezas que tengan los dos tipos de defectos. Después de ser producida cada pieza es sometida de manera automática a un test de ruptura, con las siguientes posibilidades: Si la pieza tiene el defecto de tipo A, tiene una probabilidad de 0.9 de romperse. Si la pieza tiene el defecto de tipo B, tiene una probabilidad de 0.95 de romperse. Finalmente, si la pieza no tiene ningún tipo de defecto, tiene una probabilidad de 0.01 de romperse.

   a) Si el experimento aleatorio consiste en escoger al azar un pieza de la producción, traducir los datos del enunciado, después de haber introducido los sucesos convenientes.

A = "La pieza presenta defecto tipo A".

B = "La pieza presenta defecto tipo B".

C = "La pieza no representa ningún defecto".

$$\left.\begin{array}{l} P(A) = 0.08 \\ P(B) = 0.05 \\ P(C) = 0.87 \end{array}\right\}$$

Observar que $\{A, B, C\}$ forman una partición de Ω.

En realidad $C = \bar{A} \cap \bar{B} = \overline{(A \cup B)}$

R = "La pieza seleccionada se rompe en el test"

Por el resultado sabemos que:

- $P(R/A) = 0.9$
- $P(R/B) = 0.95$
- $P(R/C) = 0.01$

b) ¿Cuál es la probabilidad de que una pieza escogida al azar en la producción se vaya a romper durante el test?

Nos piden $P(R)$. Usamos el Teorema de Probabilidad Total

$$P(R) = P(A) \cdot P(R/A) + P(B) \cdot P(R/B) + P(C) \cdot P(R/C) =$$

$$0.08 \cdot 0.9 + 0.05 \cdot 0.97 + 0.87 \cdot 0.01 = \boxed{0.1282}$$

c) Si una pieza escogida al azar se ha roto durante el test, ¿cuál es la probabilidad de que no fuese defectuosa, es decir, que no tuviera el defecto A ni el B?

Nos piden $P(C/R)$

$$P(C/R) = \frac{P(C) \cdot P(R/C)}{P(R)} = \frac{0.87 \cdot 0.01}{0.1282} = \boxed{0.06786}$$

Teorema de Bayes

PROBLEMAS DE REPASO TEMAS 5-6
FUNDAMENTOS DE PROBABILIDAD Y ANÁLISIS EXPLORATORIO DE DATOS.
GRADO EN CIENCIA E INGENIERÍA DE DATOS.

1. El número de interrupciones que se producen en la línea de envasado a lo largo de un día es una variable aleatoria X con función de distribución:

$$F(x) = \begin{cases} \alpha & \text{si } x < 0 \\ \frac{1}{2} & \text{si } 0 \leq x < 1 \\ \frac{3}{4} & \text{si } 1 \leq x < 2 \\ \frac{7}{8} & \text{si } 2 \leq x < 3 \\ \beta & \text{si } x \geq 3 \end{cases}$$

   a) Determinar el valor de α y β para que sea verdaderamente función de distribución. Hallar la función puntual de probabilidad de X.
   b) Calcular la probabilidad de que en un día se produzcan entre 1 y 3 interrupciones, ambas inclusive.
   c) Son las 10 de la mañana y ya se ha producido una interrupción en la línea de envasado. Calcular la probabilidad de que en dicho día se produzcan al menos 2 interrupciones.
   d) Calcular el número medio de interrupciones que se producen diariamente en la línea de envasado, así como su desviación típica.
   e) Determinar la mediana y la moda de X.
   f) Determinar la función puntual de probabilidad, función de distribución, media y varianza de la variable aleatoria $Y = 2X^3 - 5$.

   NOTA El problema anterior también se podría haber enunciado proporcionando la función puntual de probabilidad en lugar de la función de distribución.

2. La demanda semanal (miles de litros) de gasóleo A en una estación de servicio es una v.a. X con la siguiente función de densidad:

$$f(x) = \begin{cases} k(1 - \frac{1}{x^2}) \text{ , si } 1 \leq x \leq 2 \\ 0 \text{ , en el resto} \end{cases}$$

   a) Determinar el valor de k para que $f(x)$ sea verdaderamente una función de densidad. Determinar la función de distribución de X.
   b) Calcular la demanda semanal media y su desviación típica.
   c) Determinar la probabilidad de que en una semana se suministren entre 1300 y 1800 litros de gasóleo A.

d) Determinar la mediana y la moda de X. ¿Son únicas?

e) La demanda semanal de ese mismo combustible en otra estación de servicio viene dada por la v.a. $Y = X^2 - 1$. Determinar la función de densidad, función de distribución, media y desviación típica de Y.

NOTA El problema anterior también se podría haber enunciado proporcionando la función de distribución en lugar de la función de densidad.

3. Una envasadora se utiliza para el llenado de botellines de agua. Los botellines poseen el etiquetado de "contenido: 33 cl", considerándose aceptables sólo los botellines con contenidos entre 31 cl y 35 cl. Asumiendo que el "volumen de llenado" sigue una distribución Normal de media 33 cl y desviación típica 1 cl y que la producción de botellines es muy grande, se pide:

   a) Determinar el porcentaje de botellines defectuosos, es decir, la probabilidad de que el contenido de un botellín seleccionado al azar esté fuera del intervalo (31, 35) cl.

   b) Si se desea disminuir el porcentaje de botellines defectuosos al 1 %, determinar a cuánto deberíamos ajustar la desviación típica del volumen de envasado (suponiendo que nosotros pudiéramos ajustarla).

   c) El departamento de calidad inspecciona al azar botellines envasados. Determinar la probabilidad de que necesiten inspeccionar más de 5 botellines hasta encontrar el primero defectuoso. Calcular también la probabilidad de que seleccionen exactamente 10 botellines aceptables antes de extraer el tercer botellín defectuoso.

   d) Los botellines de agua se comercializan en paquetes de 6 unidades. Determinar la probabilidad de que en un paquete haya al menos un botellín defectuoso.

   e) Los pequeños comercios suelen adquirir los botellines en palés de 300 unidades. Determinar la probabilidad de que en un palé haya más de 15 botellines defectuosos.

   f) El kiosko del barrio tiene 20 botellines de agua para vender, 5 de los cuales son defectuosos. Si un cliente compra 8 botellines (seleccionados al azar por el vendedor), determinar la probabilidad de que haya entre 2 y 7 botellines aceptables (ambos inclusive).

4. Una empresa se dedica a la elaboración de envases (latas) para conservas. La materia prima para la elaboración de dichos envases se recibe por medio de camiones que llegan a la empresa a lo largo de todo el día. El número de camiones que llegan diariamente con material sigue una variable Poisson de media 8. Supongamos por otro lado que el tiempo, en horas, entre llegadas de camiones consecutivos sigue una distribución Exponencial de media 3 horas. Se pide:

   a) Determinar la probabilidad de que un día lleguen menos de 2 camiones de reparto.

   b) Determinar la probabilidad de que un día lleguen más de 2 y menos de 6 camiones de reparto.

   c) Determinar la probabilidad de que en 5 días consecutivos lleguen en total al menos 35 camiones de reparto.

   d) Calcular la probabilidad de que transcurran más de 5 horas entre dos camiones consecutivos.

   e) Han pasado 2 horas desde que llegó el último camión de reparto, determinar la probabilidad de que debamos esperar en total más de 5 horas hasta que llegue el siguiente camión.

1) El número de interrupciones que se producen en la línea de envasado a lo largo de un día es una variable aleatoria X con función de distribución:

$$F(x) = \begin{cases} \alpha \text{ si } x < 0 \\ \dfrac{1}{2} \text{ si } 0 \leq x < 1 \\ \dfrac{3}{4} \text{ si } 1 \leq x < 2 \\ \dfrac{7}{8} \text{ si } 2 \leq x < 3 \\ \beta \text{ si } x \geq 3 \end{cases}$$

$X =$ "Nº interrupciones en un día"

a) Determinar el valor de α y β para que sea verdaderamente función de distribución. Hallar la función puntual de probabilidad de X.

Para que $F(x)$ sea función de distribución, debe cumplir, entre otras, las siguientes propiedades:

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

Por tanto, $\alpha = 0$ y $\beta = 1$

Calcular la f.p.p:

$p(x) = P(X = x) \quad \forall x \in \mathbb{R}$

Bastará con determinar los puntos del soporte (aquellos donde se produce salto de la función de distribución) y la probabilidad de cada punto (que coincide con la amplitud del salto).

$Sop(X) = \{0, 1, 2, 3\}; \; p(x) = 0 \; \forall x \notin Sop(X)$

- $p(0) = P(X = 0) = F(0) - F(0^-) = \dfrac{1}{2} - 0 = \dfrac{1}{2}$
- $p(1) = P(X = 1) = F(1) - F(1^-) = \dfrac{3}{4} - \dfrac{1}{2} = \dfrac{1}{4}$

- $p(2) = P(X = 2) = F(2) - F(2^-) = \frac{7}{8} - \frac{3}{4} = \frac{1}{8}$
- $p(3) = P(X = 3) = F(3) - F(3^-) = 1 - \frac{7}{8} = \frac{1}{8}$

Recordar $F(a^-) = \lim_{\substack{x \to a \\ x < a}} F(x)$ (límite por la izquierda)

b) Calcular la probabilidad de que en un día se produzcan entre 1 y 3 interrupciones, ambas inclusive.

Me piden $P(1 \leq x \leq 3)$

$$P(1 \leq x \leq 3) = F(3) - F(1^-) = 1 - \frac{1}{2} = \boxed{\frac{1}{2}}$$

c) Son las 0 de la mañana y ya se ha producido una interrupción en la línea de envasado. Calcular la probabilidad de que en dicho día se produzcan al menos 2 interrupciones.

Me piden $P(X \geq 2 | X \geq 1)$

$$P(X \geq 2 | X \geq 1) = \frac{P((X \geq 1) \cap (X \geq 2))}{P(X \geq 1)} = \frac{P(X \geq 2)}{P(X \geq 1)} = \frac{p(2) + p(3)}{p(1) + p(2) + p(3)} =$$

$$\frac{\frac{1}{8} + \frac{1}{8}}{\frac{1}{4} + \frac{1}{8} + \frac{1}{8}} = \frac{\frac{1}{4}}{\frac{1}{2}} = \boxed{\frac{1}{2}}$$

d) Calcular el número medio de interrupciones que se producen diariamente en la línea de envasado, así como su desviación típica.

Me piden $E(X)$ y $\sigma(X)$

$$E(X) = \sum_{x_i \in Sop(X)} x_i \cdot p(x_i)$$

$$E(X) = 0 \cdot p(0) + 1 \cdot p(1) + 2 \cdot p(2) + 3 \cdot p(3) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} = \boxed{\frac{7}{8}}$$

$$Var(X) = E(X^2) - (E(X))^2 = (\times)$$

$$E(X^2) = 0^2 \cdot p(0) + 1^2 \cdot p(1) + 2^2 \cdot p(2) + 3^2 \cdot p(3) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} + 4 \cdot \frac{1}{8} + 9 \cdot \frac{1}{8} = \frac{15}{8}$$

$(\times) = \frac{15}{8} - \left(\frac{7}{8}\right)^2 = \frac{120-49}{64} = \frac{71}{64}$

$\sigma(X) = \sqrt{Var(X)} = \sqrt{\frac{71}{64}} = \boxed{1.053}$

e) Determinar la mediana y la moda de X.

Me piden $Me(X)$ y $Moda(X)$.

- La Moda es el punto (o puntos) del soporte con mayor probabilidad, en este caso

  $Moda(X) = 0$ porque $\max\limits_{x} p(x) = p(0) = \frac{1}{2}$.

- La Mediana es el punto (o puntos) verificando:

$$P(X \geq Me) \geq 0.5 \text{ y } P(X \leq Me) \geq 0.5$$

Mirando la función de distribución del enunciado, observamos que:

- $P(X \leq 0) = F(0) = \frac{1}{2}$, $P(X \geq 0) = 1 \longrightarrow$ El punto "$x = 0$" es una mediana de X.

- $P(X \leq 1) = F(1) = \frac{3}{4}$, $P(X \geq 1) = \frac{1}{2} \longrightarrow$ El punto "$x = 1$" es una mediana de X.

- Todos los puntos del intervalo $[0,\ 1]$ son mediana de X.

Observamos que en este caso la Moda es única ($Moda(X) = 0$), pero la mediana no es única ($Me(X) = [0,\ 1]$).

f) Determinar la función puntual de probabilidad, función de distribución, media y varían a de la variable aleatoria $Y = 2X^3 - 5$.

- <u>Calcular la f.p.p de Y:</u>

Recordemos que $Sop(X) = \{0,\ 1,\ 2,\ 3\}$. Entonces, $Sop(Y) = \{-5,\ -3,\ 11,\ 49\}$

Si $x = 0 \longrightarrow y = 2 \cdot 0^3 - 5 = -5$

$x = 1 \longrightarrow y = 2 \cdot 1^3 - 5 = -3$

$x = 2 \longrightarrow y = 2 \cdot 2^3 - 5 = 11$

$x = 3 \longrightarrow y = 2 \cdot 3^3 - 5 = 49$

- $P_Y(-5) = P(Y = -5) = P(X = 0) = p_X(0) = \frac{1}{2}$

- $P_Y(-3) = P(Y = -3) = P(X = 1) = p_X(1) = \dfrac{1}{4}$

- $P_Y(11) = P(Y = 11) = P(X = 2) = p_X(2) = \dfrac{1}{8}$

- $P_Y(49) = P(Y = 49) = P(X = 3) = p_X(3) = \dfrac{1}{8}$

- Calcular la función de distribución de Y

$$F_y(y) = \begin{cases} 0 \text{ si } y < -5 \\ \dfrac{1}{2} \text{ si } -5 \leq y < -3 \\ \dfrac{3}{4} \text{ si } -3 \leq y < 11 \\ \dfrac{7}{8} \text{ si } 11 \leq y < 49 \\ 1 \text{ si } y \geq 49 \end{cases}$$

Calcular $E(Y)$; $E(Y) = \displaystyle\sum_{y_i \in Sop(Y)} y_i \cdot P_Y(y_i)$

$$E(Y) = (-5)\cdot\frac{1}{2} + (-3)\cdot\frac{1}{4} + 11\cdot\frac{1}{8} + 49\cdot\frac{1}{8} = \frac{(-10)+(-6)+11+49}{8} = \frac{34}{8} = \frac{17}{4}$$

Otra forma:

$$E(Y) = E(2X^3 - 5) = 2E(X^3) - 5 = (\times)$$

$$E(X^3) = 0^3\cdot\frac{1}{2} + 1^3\cdot\frac{1}{4} + 2\cdot\frac{1}{8} + 3^3\cdot\frac{1}{8} = \frac{2+8+27}{8} = \frac{37}{8}$$

$$(\times) = 2\cdot\frac{37}{8} - 5 = \frac{74-40}{8} = \frac{34}{8} = \boxed{\frac{17}{4}}$$

Calcular $Var(Y)$:

$$Var(Y) = E(Y^2) - (E(Y))^2 = (\times)$$

$$E(Y^2) = (-5)^2\cdot\frac{1}{2} + (-3)\cdot\frac{1}{4} + +11^2\cdot\frac{1}{8} + 49^2\cdot\frac{1}{8} = \frac{100+18+121+2401}{8} = \frac{2640}{8} = 330$$

$$(\times) = 330 - \left(\frac{17}{4}\right) = \frac{5280-289}{16} = \frac{4991}{16} = \boxed{311.9375}$$

Otra forma:

$$E(Y^2) = E((2X^3 - 5)^2) = E(4X^6 - 20X^3 + 25) = 4E(X^6) - 20E(X^3) + 25$$

Se puede calcular $E(X^6)$ y $E(X^3)$ a partir de la f.p.p de X.

NOTA: El problema anterior también se podría haber enunciado proporcionando la función puntual de probabilidad en lugar de la función de distribución.

2) La demanda semanal (miles de litros) de gasóleo A en una estación de servicio es una v.a. X con la siguiente función de densidad:

$$f(x) = \begin{cases} k\left(1 - \dfrac{1}{x^2}\right), & \text{si } 1 \leq x \leq 2 \\ 0, & \text{en el resto} \end{cases}$$

$X =$ "Demanda semanal de gasóleo A en miles de litros"

a) Determinar el de valor de k para que $f(x)$ sea verdaderamente una función de densidad. Determinar la función de distribución de X.

Para que $f(x)$ sea densidad, debe cumplir:

(i) $f(x) \geq 0 \ \forall x \in \mathbb{R}$

(ii) $\displaystyle\int_{\mathbb{R}} f(x)dx = 1$

Observar que (i) se cumple siempre que $k \geq 0$, ya que $1 - \dfrac{1}{x^2} \geq 0 \ \forall x \in [1, 2]$.

Veamos el valor de "k" para que cumple (ii):

$$\int_{\mathbb{R}} f(x)dx = \int_1^2 \left(1 - \frac{1}{x^2}\right) dx = k\left([x]_1^2 + \left[\frac{1}{x}\right]_1^2\right) = k\left(1 + \frac{1}{2} - 1\right) = k \cdot \frac{1}{2}$$

Luego $k \cdot \dfrac{1}{2} = 1 \longleftrightarrow \boxed{k = 2}$

- Calcular función distribución $F(X) = P(X \leq x)$

$$F(X) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

- Si $x \leq 1 \longrightarrow F(x) = 0$

- Si $x \in [1,\, 2] \longrightarrow F(x) = \int_1^x 2\left(1 - \frac{1}{2}\right) dt = 2\left([t]_1^x + \left[\frac{1}{t}\right]_1^x\right) =$

$$2\left((x-1) + \left(\frac{1}{x} - 1\right)\right) = 2\left(x + \frac{1}{x} - 2\right)$$

- Si $x \geq 2 \longrightarrow F(x) = 1$

Luego:

$$F(x) = \begin{cases} 0 & \text{si } x \leq 1 \\ 2\left(x + \dfrac{1}{x} - 2\right) & \text{si } x \in [1,\, 2] \\ 1 & \text{si } x \geq 2 \end{cases}$$

b) Calcular la demanda semanal media y su desviación típica.

Nos piden $E(X)$ y $\sigma(X)$

$$E(X) = \int_{\mathbb{R}} x f(x) dx = \int_1^2 x \cdot 2 \cdot \left(1 - \frac{1}{x^2}\right) dx = 2 \cdot \int_1^2 \left(x - \frac{1}{x}\right) dx =$$

$$2\left(\int_1^2 x dx - \int_1^2 \frac{1}{x} dx\right) = 2 \cdot \left(\left[\frac{x^2}{2}\right]_1^2 - [\ln(x)]_1^2\right) = 2 \cdot \left(\left(2 - \frac{1}{2}\right) - \left(\ln(2) - \ln(1)\right)\right)$$

$$= 2 \cdot \left(\frac{3}{2} - \ln(2)\right) = 3 - 2\ln(2) = \boxed{1.6137}$$

$$Var(X) = E(X^2) - (E(X))^2 = (\times)$$

$$E(X^2) = \int_{\mathbb{R}} x^2 f(x) dx = \int_1^2 x^2 \cdot 2 \cdot \left(1 - \frac{1}{x^2}\right) dx = 2\left(\int_1^2 x^2 dx - \int_1^2 1 dx\right) =$$

$$2\left(\left[\frac{x^3}{3}\right]_1^2 - [x]_1^2\right) = 2\left(\left(\frac{2^3}{3}-\frac{1}{3}\right)-(2-1)\right) = 2\left(\frac{7}{3}-1\right) = 2\cdot\frac{4}{3} = \frac{8}{3}$$

$$(\times) = \frac{8}{3} - (1{,}6137)^2 = 0{,}06262$$

$$\sigma(X) = \sqrt{0{,}06262} = \boxed{0{,}25}$$

c) Determinar la probabilidad de que en una semana se suministren entre 1300 y 1800 litros de gasóleo A.

Me piden $P(1{,}3 \le X \le 1{,}8)$

$$P(1{,}3 \le X \le 1{,}8) = F(1{,}8) - F(1{,}3) = 2\left(1{,}8 + \frac{1}{1{,}8} - 2\right) - 2\left(1{,}3 + \frac{1}{1{,}3} - 2\right) = \boxed{0{,}5726}$$

Otra forma:

$$P(1{,}3 \le X \le 1{,}8) = \int_{1{,}3}^{1{,}8} f(x)dx = \int_{1{,}3}^{1{,}8} 2\cdot\left(1-\frac{1}{x^2}\right)dx$$

d) Determinar la mediana y la moda de X. ¿Son únicas?

Me piden $Me(X)$ y $Moda(X)$.

La Mediana es el punto (o puntos) que cumple $P(X \le Me) \ge 0{,}5$ y $P(Me \ge Me) \ge 0{,}5$

Luego en v.a continuas resolveremos $F(Me) = 0{,}5$ para obtener "Me" (la mediana).

Mirando la función de distribución calculada en (a) tenemos:

$$F(X) = 2\left(x + \frac{1}{x} - 2\right) = 0{,}5 \iff 2x + \frac{2}{x} = 4{,}5 \iff 2x^2 - 4{,}5x + 2 = 0 \iff$$

$$x = \frac{4{,}5 \pm \sqrt{4{,}5^2 - 4\cdot 2\cdot 2}}{2\cdot 2} = \begin{cases} \nearrow \ \boxed{x = 1{,}64 = Me} \\ \searrow \ x = 0{,}6096 \longrightarrow \text{Solución descartada} \notin [1,\ 2] \end{cases}$$

Por definición, la moda es el punto (o puntos) donde la densidad alcanza el máximo.

$$\max_{x\in[1,\ 2]}\left\{2\left(1-\frac{1}{x^2}\right)\right\} = f(2) = 2\cdot\left(1-\frac{1}{2^2}\right) = \frac{3}{2} \Longrightarrow \boxed{Moda(X) = 2}$$

En este caso, mediana y moda son únicas.

e) La demanda semanal de ese mismo combustible en otra estación de servicio viene dada por la v.a $Y = X^2 - 1$. Determinar la función de densidad, función de distribución, media y desviación típica de Y.

$Y = x^2 - 1$

- Calcular densidad de Y, $f_Y(y)$

Observar que $g(x) = x^2 - 1$ es biyección para $x \in [1, 2]$

$y = x^2 - 1 \longrightarrow x^2 = y + 1 \longrightarrow x = \sqrt{y+1} = g^{-1}(y).$

Usamos el teorema que relaciona $f_Y(y)$ con $f_X(x)$:

$$F_Y(y) = f_X(g^{-1}(y)) \cdot \left| \left(g^{-1}(y)\right)' \right|$$

$$\left(g^{-1}(y)\right)' = \frac{\partial}{\partial y}\left(g^{-1}(y)\right) = \frac{\partial}{\partial y}\left(\sqrt{y+1}\right) = \frac{1}{2\sqrt{y+1}}$$

- Si $x \in [1, 2] \longrightarrow y = x^2 - 1 \in [0, 3]$

$$f_Y(y) = \begin{cases} 0 \text{ si } y \notin [0, 3] \\ 2 \cdot \left(1 - \dfrac{1}{\left(\sqrt{y+1}\right)^2}\right) \cdot \dfrac{1}{2\sqrt{y+1}} = \dfrac{1}{\sqrt{y+1}}\left(1 - \dfrac{1}{y+1}\right) \text{ si } y \in [0, 3] \end{cases}$$

- Calcular función de distribución de Y, $F_Y(y)$

$$F_Y(y) = P(Y \le y) = P(X^2 - 1 \le y) = P(X^2 \le y + 1) = P\left(-\sqrt{y+1} \le X \le \sqrt{y+1}\right) =$$

$$F_X\left(\sqrt{y+1}\right) - \underbrace{F_X\left(-\sqrt{y+1}\right)}_{=0} = 2 \cdot \left(\sqrt{y+1} + \frac{1}{\sqrt{y+1}} - 2\right) \text{ si } y \in [0, 3]$$

$$F_Y(y) = \begin{cases} 0 \text{ si } y \leq 0 \\ 2 \cdot \left(\sqrt{y+1} + \dfrac{1}{\sqrt{y+1}} - 2 \right) \text{ si } y \in [0,\ 3] \\ 1 \text{ si } y \geq 3 \end{cases}$$

- Calcular $E(Y)$

$$E(Y) = \int_{\mathbb{R}} y \cdot f_Y(y) dy \longrightarrow \text{forma 1}$$

Otra forma:

$$E(Y) = E(X^2 - 1) = E(X^2) - 1 = \frac{8}{3} - 1 = \boxed{\frac{5}{3}}$$

$E(X^2) = \dfrac{8}{3}$ (calculado en apartado (b))

- Calcular $\sigma(Y)$

$$Var(Y) = E(Y^2) - \left(E(Y)\right)^2$$

$$E(Y^2) = E\left(\left(X^2 - 1\right)^2\right) = E\left(X^4 - 2X^2 + 1\right) = E(X^4) - 2E(X^2) + 1$$

Faltaría calcular $E(X^4)$.

Nota: El problema anterior también se podría haber enunciado proporcionando la función de distribución en lugar de la función de densidad.

3) Una envasadora se utiliza para el llenado de botellines de agua. Los botellines poseen el etiquetado de "contenido: 33cl", considerándose aceptables sólo los botellines no contenidos entre 31 cl y 35 cl. Asumiendo que el "volumen de llenado" sigue una distribución Normal de media 33 cl y desviación típica 1cl y que la producción de botellines es muy grande, se pide:

$X =$ "Volumen de llenado en cl"

$X \sim N(\mu = 33,\ \sigma = 1)$

Botellín aceptable si $X \in (31,\ 35)$

a) Determinar el porcentaje de botellines defectuosos, es decir, la probabilidad de que el contenido de un botellín seleccionado al azar esté fuera del intervalo (31, 35) cl.

$P(X \notin (31,\ 35) = 1 - P(X \in (31,\ 35)$

$$P(31 < X < 35) = P\left(\frac{31-33}{1} < \boxed{\frac{X-33}{1}} < \frac{35-33}{1}\right) = P(-2 < Z < 2) =$$

Tipificar

$Z \sim N(0,1)$

$P(Z < 2) - P(Z < -2) = P(Z < 2) - (1 - P(Z < 2)) = Ø(2) - (1 - Ø(2)) =$

$0{,}9772 - (1 - 0{,}9772) = 0{,}9772 - 0{,}0228 = \boxed{0{,}9544}$

$P(\text{"Botellín aceptable"}) = \boxed{0{,}9544}$

$P(\text{"Botellín defectuoso"}) = 1 - 0{,}9544 = \boxed{0{,}0456}$

b) Si se desea disminuir el porcentaje de botellines defectuosos al 1%, determinar a cuánto deberíamos ajustar la desviación típica del volumen de envasado (suponiendo que nosotros supiéramos ajustarla).

Consideramos ahora "σ" nuestra incógnita, para que se cumpla $P(31 < X < 35) = 0{,}99$

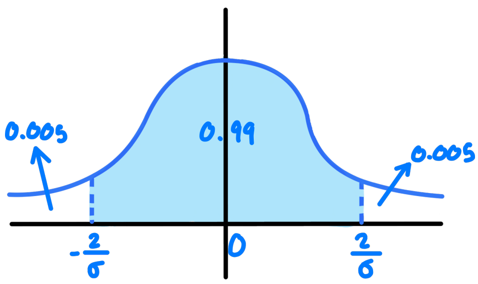
$Z \sim N(0,1)$

$$P(31 < X < 35) = 0{,}99 \longleftrightarrow P\left(\frac{31-33}{\sigma} < \boxed{\frac{X-33}{\sigma}} < \frac{35-33}{\sigma}\right) = 0{,}99 \Longleftrightarrow$$

$$P\left(\frac{-2}{\sigma} < Z < \frac{2}{\sigma}\right) = 0{,}99 \Longleftrightarrow \frac{2}{\sigma} = z_{0{,}995} = 2{,}575$$

Cuadril 0.995 de la N(0, 1)

$$\longrightarrow \sigma = \frac{2}{2{,}575} = \boxed{0{,}7767}$$



c) El departamento de calidad inspecciona al azar botellines envasados. Determinar la probabilidad de que necesiten inspeccionar más de 5 botellines hasta encontrar el primero defectuoso. Calcular también la probabilidad de que seleccionen exactamente 10 botellines aceptables antes de extraer el tercer botellín defectuoso.

Sea $Y =$ "Nº de botellines inspeccionados hasta encontrar el 1er defectuoso".

Sea $V=$ "Nº botellines aceptables hasta encontrar el 1er defectuoso".

$V \sim G(p=0{,}0456)$

- Me piden $P(Y>5)$

$P(Y>5)=P(V+1>5)=P(V>4)=1-\big(P(V=0)+P(V=1)+$

$P(V=2)+P(V=3)+P(V=4)\big)=1-\big(0{,}0456+0{,}9544\cdot 0{,}0456+0{,}9544^2\cdot$

$0{,}0456+0{,}9544^3\cdot 0{,}0456+0{,}9544^4\cdot 0{,}0456)\big)=\boxed{0{,}8012}$

Sea $W=$ "Nº botellines aceptables antes de obtener el 3er defectuoso"

$W \sim BN(n=3,\ p=0{,}0456)$

- Me piden $P(W=10)=\binom{3+10-1}{10}\cdot 0{,}0456^3+0{,}9544^{10}=$

$$\frac{12\cdot 11}{2}\cdot 0{,}0456^3\cdot 0{,}9544^{10}=\boxed{0{,}0039}$$

d) Los botellines de agua se comercializan en paquetes de 6 unidades. Determinar la probabilidad de que en un paquete haya al menos un botellín defectuoso.

Sea ahora $Y=$ "Nº botellines defectuosas en un paquete de 6"

$Y \sim B(n=6,\ p=0{,}0456)$

Me piden $P(Y\geq 1)=1-P(Y\leq 1)=1-P(Y=0)=1-0{,}9544^6=\boxed{0{,}2442}$

e) Los pequeños comercios suelen adquirir los botellines en palés de 300 unidades. Determinar la probabilidad de que en un palé hay más de 15 botellines defectuosos.

Sea ahora $Y=$ "Nº de botellines defectuosos en el palé de 300 unidades"

$Y \sim B(n=300{,}p=0{,}456)$

Me piden $P(Y>15)$

Como $n=300$ grande y $n\cdot p\cdot(1-p)=300\cdot 0{,}0456\cdot 0{,}9544=13{,}0{,}56>5\longrightarrow$

podemos aproximar la Binomial por una Normal.

$$P(Y>15)\backsimeq P(W\geq 15{,}5)=\left(\frac{W-13{,}68}{3{,}6133}\geq\frac{15{,}5-13{,}68}{3{,}6133}\right)=P(Z\geq 0{,}5036)=$$

Aproximó por Normal y hago corrección por continuidad

$1 - P(Z \leq 0{,}5036) = 1 - 0{,}6915 = \boxed{0{,}3085}$

f) El kiosko del barrio tiene 20 botellines de agua para vender, 5 de los cuales son defectuosos. Si un cliente compra 8 botellines (seleccionados al azar por el vendedor), determinar la probabilidad de que haya entre 2 y 7 botellines aceptables (ambos inclusive).

$$\underbrace{\text{20 botellines}}_{N} \begin{cases} \nearrow \text{ 5 defectuosos} \longrightarrow a = 5 \\ \searrow \text{ 15 aceptables} \longrightarrow b = 15 \end{cases}$$

Seleccionamos $\underbrace{8}_{n}$ botellines de las 20 disponibles.

$Y =$ "Nº de botellines defectuosos entre los 8 seleccionados".

$Y \sim H(N = 20,\ a = 5,\ n = 8);\ \ Sop(Y) = \{0,\ 1,\ 2,\ 3,\ 4,\ 5\}$

Me piden $P(2 \leq Y \leq 7)$

$$P(2 \leq Y \leq 7) = 1 - \big(P(Y=0) + P(Y=1)\big) = 1 - \left(\frac{\binom{15}{8}}{\binom{20}{8}} + \frac{\binom{5}{1} \cdot \binom{15}{7}}{\binom{20}{8}} \right) =$$

$1 - 6 \cdot 0{,}05108 = \boxed{0{,}6935}$

$$\frac{\binom{15}{8}}{\binom{20}{8}} = \frac{\dfrac{15!}{8! \cdot 7!}}{\dfrac{20!}{8! \cdot 12!}} = \frac{15! \cdot 12!}{7! \cdot 20!} = \frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8}{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16} = 0{,}05108$$

$$\frac{\binom{15}{7}}{\binom{20}{8}} = \frac{\binom{15}{8}}{\binom{20}{8}} = 0{,}05108$$

4) Una empresa se dedica a la elaboración de envases (latas) para conservas. La materia prima para la elaboración de dichos envases se recibe por medio de camiones que llegan a la empresa a lo largo de todo el día. El número de camiones que llegan diariamente con material sigue una variable Poisson de media 8. Supongamos por otro lado que el tiempo, en horas, entre entradas de camiones consecutivos sigue una distribución Exponencial de media 3 horas. Se pide:

$X =$ "Nº de camiones que llegan al día"

$X \sim P(\lambda_X = 8)$ (recordar que la media de la Poisson es el parámetro λ)

$E(X) = 8$ camiones

$Y =$ "Tiempo en horas entre dos camiones consecutivos"

$Y \sim Exp\left(\lambda_Y = \frac{1}{3}\right)$ $\left(\text{recordar que la media de la Exponencial es } \frac{1}{\lambda}\right)$

a) Determinar la probabilidad de que un día llegue menos de 2 camiones de reparto.

$$P(X<2) = P(X=0) + P(X=1) = e^{-8} \cdot \frac{8^0}{0!} + e^{-8} \cdot \frac{8^1}{1!} = 0{,}0003 + 0{,}027 = \boxed{0{,}003}$$

$$P(X=x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

b) Determinar la probabilidad de que un día lleguen más de 2 y menos de 6 camiones de reparto.

$$P(2<X<6) = P(X=3) + P(X=4) + P(X=5) = e^{-8}\left(\frac{8^3}{3!} + \frac{8^4}{4!} + \frac{8^5}{5!}\right) =$$

$$0{,}0286 + 0{,}0573 + 0{,}0916 = \boxed{0{,}1775}$$

c) Determinar la probabilidad de que en 5 días consecutivos lleguen en total al menos 35 camiones de reparto.

Sea $X_i =$ "Nº de camiones que llegan el día i", $i = 1, 2, \dots, 5$

$V = X_1 + X_2 + X_3 + X_4 + X_5 =$ "Nº de camiones que llegan en los 5 días consecutivos"

Como $X_i \sim P(\lambda = 8)$ $\forall_i = 1, \dots, 5$ e independientes $\longrightarrow V = \sum_{i=1}^{5} X_i \sim P(\lambda_V = 40)$

Me piden $P(V \geq 35)$. Obtener esta probabilidad a mano de forma exacta es muy largo, así que usaremos la aproximación por la Normal al ser $\lambda_V = 40 > 5$.

Sea $W \sim N(\mu = 40,\ \sigma = \sqrt{40})$, entonces $\underbrace{V}_{\text{Poisson}} \approx \underbrace{W}_{\text{Normal}}$

$$P(V \geq 35) \simeq P(W \geq 34{,}5) \underset{\text{Tipificar}}{=} P\left(\underbrace{\boxed{\frac{W-40}{\sqrt{40}}}}_{Z \sim N(0,1)} \geq \frac{34{,}5-40}{\sqrt{40}}\right) = P(Z \geq -0{,}8696)$$

$$P(Z \leq -0{,}8696) = \boxed{0{,}8078}$$

d) Calcular la probabilidad de que transcurran más de 5 horas entre dos camiones consecutivos.

Me piden $P(Y > 5)$

$$P(Y > 5) = \int_5^{+\infty} \lambda e^{-\lambda x} dx = \int_5^{+\infty} \frac{1}{3} \cdot e^{-\frac{1}{3}x} dx = \left[-e^{-\frac{1}{3}x}\right]_5^{+\infty} =$$

$$0 - (-e^{-\frac{5}{3}}) = e^{-\frac{5}{3}} = 0{,}1889$$

e) Han pasado 2 horas desde que llegó el último camión de reparto, determinar la probabilidad de que debamos esperar en total más de 5 horas hasta que llegue el siguiente camión.

Me piden $P(Y > 5 \mid Y > 2)$

$$P(Y > 5 \mid Y > 2) \underset{\text{Propiedad Falta Memoria}}{=} P(Y > 3)$$

$$P(Y > 3) = \int_3^{+\infty} \lambda e^{-\lambda x} dx = \int_3^{+\infty} \frac{1}{3} \cdot e^{-\frac{1}{3}x} dx = \left[-e^{-\frac{1}{3}x}\right]_3^{+\infty} =$$

$$0 - (-e^{-1}) = e^{-1} = \boxed{0{,}36788}$$