Fundamentos De Matemáticas

UNIVERSIDAD DEL ATLÁNTICO
FACULTAD DE CIENCIAS BÁSICAS
DEPARTAMENTO DE MATEMÁTICAS
Autores

Índice general

I Estadística Descriptiva

Estadística Descriptiva

CAPÍTULO

1 Conceptos previos

Contenido Del Capítulo

1.1 Introducción

En este capítulo hablaremos de la estadística en forma general de sus conceptos básicos y de su utilidad, pero al final pretendemos introducir al estudiante en el manejo de los datos obtenidos, como son: distinguir y clasificar los datos, organizarlos, tabularlos y representarlos gráficamente, para así obtener de ellos una mejor información y realizar así una buena inferencia.

No obstante al realizar todos los procedimientos que enseñaremos en este capítulo debemos ser prudentes al analizar los gráficos obtenidos ya que los mismos datos se pueden representar de diferentes formas y no todas son las más pertinentes, por lo que nuestro objetivo en este capítulo además de construir las bases conceptuales de la estadística desde un punto de vista elemental es enseñarles a realizar los gráficos que mejor representen las variables de interés.

1.2 ¿Qué es la estadística?

A pesar de que la estadística es hoy por hoy una ciencia axiomatizada y por ende una rama de las matemáticas, en sus orígenes era una ciencia puramente experimental, como lo indica la historia de la humanidad, ya que se tiene testimonio de que los egipcios 3050 ac llevaban datos acerca de los esclavos, población, agrimensura y riquezas; los chinos llevaban estadística de la población; la Biblia en el libro de los números se habla de los censos realizados por los levitas, lo mismo que los romanos lo cual lo hacían cada 5 años; en América los Aztecas y los Incas también tabulaban datos.

Pero fue en la edad media cuando Jhonn Graunt realizó un estudio estadístico serio sobre las muertes ocurridas en Londres entre 1603 y 1624 y publicado en 1662, luego Edmond Halley en 1693 realizó la primera tabla de seguros de vida.

Hasta ese momento la estadística aún era una ciencia puramente experimental y fueron los matemáticos Jacobo Bernoulli, Fermat, Pascal, Laplace, Leibniz, Huyghens, Bayes, Tchebitchef y Borel los que la axiomatizaron.

La palabra estadística fue usada por primera vez por Godofredo Achenwall convencido de que los datos estudiados por esta nueva ciencia sería de gran utilidad para los gobernantes.

Como lo indica este pequeño resumen histórico la estadística generalmente se toma como una ciencia que se encarga de presentar informes de encuestas, como nos lo hace pensar también los medios de comunicación masivos, pero la estadística es algo más que eso.

Ella se encarga de establecer los métodos y procedimientos para recolectar, clasificar y resumir datos para luego analizarlos siempre que exista variabilidad e incertidumbre causada intrínsecamente por ellos y después realizar inferencias con el fin de hacer predicciones y tomar decisiones.

Aunque no es fácil definir la estadística y por eso no vamos a tratar de hacerlo aquí por que lo más importante es comprender `que es y como se utiliza` y en el párrafo anterior expresamos lo que es la estadística.

En el siguiente esquema representamos en forma general la estructura de esta ciencia, aunque faltan eslabones estos lo representaremos en el segundo curso.

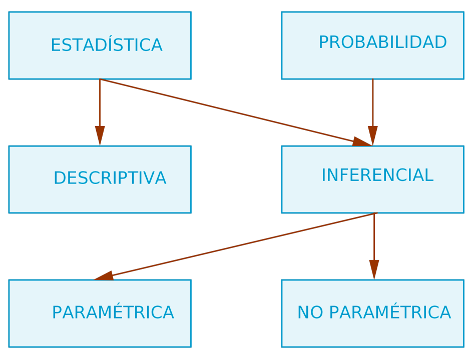


Figura 1.1: Diagrama

Ahora clasificaremos la estadística.

Estadística descriptiva: Describe y analiza los datos utilizando métodos numéricos elementales y gráficos para resumir y presentar la información contenida en ellos.

Estadística inferencial: apoyándose del cálculo de probabilidades y utilizando los datos de una muestra, efectúa estimaciones, decisiones y generaliza sobre un conjunto mayor de datos llamado población.

1.3 Conceptos previos

Definición 1.1 Individuos o elementos:

Individuos o elementos:Son personas u objetos que contienen cierta información de interés.

Definición 1.2 Población:

Es el conjunto de individuos o elementos que poseen la misma información de interés, es decir poseen ciertas propiedades comunes, y lo denotaremos Ω.o S

Definición 1.3 Muestra:

Es un subconjunto representativo de una población.

Definición 1.4 Caracteres o atributos:

Son rasgos o cualidades de los elementos de la población, los cuales pueden ser cualitativos o cuantitativos.

Definición 1.5 Modalidades:

Son diferentes situaciones posibles de un carácter, los cuales deben ser mutuamente excluyentes y exhaustivos, es decir cada elemento posee una y solo una de las modalidades posibles y un carácter debe tener todas las modalidades posibles.

Definición 1.6 Clases:

Conjunto de una o más modalidades en el que se verifica que cada modalidad pertenece a una y solo una de las clases

Definición 1.7 Variabilidad:

Al estudiar unos individuos escogemos uno varios caracteres, pero la medida de le los caracteres presentan desviaciones con respecto a una modalidad cuando cada carácter es analizado con las mismas condiciones, pero con la imposibilidad de controlar esas desviaciones. A cada una de esas desviaciones se le llama variabilidad.

1.3.1. Variables estadísticas

Cuando hablamos de variables estamos en realidad hablando de una función definida:

$$f: \begin{array}{ccc} A & \longrightarrow & B \\ x_i & \longmapsto & X(x_i) \end{array}$$

Donde $A \subseteq \Omega$ y B es un conjunto determinado de modalidades.

Al conjunto B se le denomina dominio de la variable o rango, de acuerdo con el tipo de dominio las variables se clasifican en:

Definición 1.8 Variables cualitativas:

Son aquellas que las modalidades posibles son de tipo nominal o categórico.

Ejemplo: 1.3.1

{si, no}, {hombre, mujer}, {blanco, negro, amarillo,...,etc.}

Definición 1.9 Variables casicuantitativas:

Son variables de tipo nominal, pero se puede establecer un orden entre ellas.

Ejemplo: 1.3.2

{1^o,2^o,3^o,...}, {alto, medio, bajo}

Definición 1.10 Variables cuantitativas:

Son las que tienen por modalidades un conjunto de elementos dotados de operaciones y estas a su vez se clasifican en discretas y continuas

Definición 1.11 Discretas:

Cuando las variables toman valores puntuales

Ejemplo: 1.3.3

Cuando el rango es subconjunto de los números $I\!N$,

Definición 1.12 Continuas:

si X toma todos los valores en un intervalo

Ejemplo: 1.3.4

si $X \in I\!R$

1.4 Organización de datos

Generalmente los datos se organizan usando tablas y determinando algunas cantidades o valores que definiremos a continuación:

Consideremos una población finita de n elementos o individuos descrita según un carácter o variable X cuyas modalidades han sido agrupadas en un número k de clases, que denotaremos $c_1, c_2, c_3, ..., c_k$ y para cada clase $c_i, i = 1, 2, 3, ..., k$ definimos lo siguiente:

Definición 1.13 Distribución de frecuencias:

Es una función que a cada clase c_i de C (el conjunto de todas las clases) le asigna un valor $n_i \in I\!N_0$, es decir, es una función de la forma

$$\begin{array}{rccc} \mathfrak{F}: & C & \rightarrow & I\!N_0 \\ & c_i & \longmapsto & n_i \end{array}$$

tal que $\mathfrak{F}(c_i) = n_i$

La función de distribución no es única existen cuatro, las cuales definiremos ahora.

Definición 1.14 Frecuencia absoluta:

La frecuencia absoluta de una clase c_i es el número de veces n_i que se observa una modalidad perteneciente a esa clase.

Definición 1.15 Frecuencia relativa:

La frecuencia relativa de una clase c_i es el cociente f_i entre la frecuencia absoluta de la clase c_i y el número total de observaciones de todas las modalidades pertenecientes a todas las clases.

Generalmente esta frecuencia se multiplica por 100 y se da en porcentaje

Definición 1.16 Frecuencia absoluta acumulada:

La frecuencia absoluta acumulada N_i es el número de elementos de la población cuya modalidad es inferior o equivalente a la modalidad c_i es decir,

$$N_i = \sum_{j=1}^{k} n_j$$

Definición 1.17 Frecuencia relativa acumulada:

Se denota F_i y es el tanto por uno de los elementos de la población que están en alguna de las clases y que presenta una modalidad inferior o igual a la c_i, es decir

$$F_i = \frac{N_i}{n} = \sum_{j=1}^{k} f_j$$

De éstas definiciones se pueden deducir algunas propiedades evidentes ya que las modalidades son exhaustivas y mutuamente excluyentes

[0]La población también puede ser infinita, pero ese caso lo estudiaremos más adelante en el capítulo 3

- $$n = \sum_{j=1}^{k} n_j,$$

- $$\sum_{j=1}^{k} f_j = 1$$

Definición 1.18 Tabla de frecuencia:

Es una representación de $\mathfrak{F}$ y generalmente una tabla como se muestra en la tabla 1.1
Donde:

- M: representa modalidades,
- F.A.: frecuencia absoluta,
- F.R.: frecuencia relativa,
- F.A.A: frecuencia absoluta acumulada y
- F.R.A.: frecuencia relativa acumulada.

M	F.	F.R.	F.A.A.	F.R.A.
C	n_i	f_i	N_i	F_i
c_1	n_1	$f_1=\frac{n_1}{n}$	$N_1=n_1$	$F_1=\frac{N_1}{n}=f_1$
Cc_2	n_2	$f_2=\frac{n_2}{n}$	$N_2=n_1+n_2$	$F_2=f_1+f_2$
c_3	n_3	$f_3=\frac{n_3}{n}$	$N_3=n_1+n_2+n_3$	$F_3=f_1+f_2+f_3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
c_j	n_j	$f_j=\frac{n_j}{n}$	$N_j=\sum_{i=1}^{j} n_i$	$F_j=\sum_{i=1}^{j} f_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
c_k	n_k	$f_k=\frac{n_k}{n}$	$N_k=n$	$F_k=1$
[a]	n	1		

Tabla 1.1: Tabla de Frecuencias

[a]Aunque hemos definido la tabla sólo para $\mathfrak{F}$ en realidad en una tablas de frecuencia se tabulan las otras frecuencias definidas en este apartado

Ejemplo: 1.4.5

Se lanzan cinco monedas 1000 veces. El número de lanzamientos en los que han salido 0,1,2,3,4,5 caras se indican en la Tabla 1.2:

Nº de caras	n_i	f_i	N_i	F_i
0	38			
1	144			
2	342			
3	287			
4	164			
5	25			
Total	1000			

Tabla 1.2: Lanzamiento de monedas

a. Completar la tabla

b. Determinar para que clase F_i es mayor que el 60 %: $F_{---} =$ ---

1.4.0.1. Elección de clases

Las clases se pueden seleccionar de diferentes formas, pero siempre hay que seguir los criterios que se ajustan al tipo de variables que estudiamos.

- Cuando se trata de una variable nominal las clases c_i serán de tipo nominal
- Cuando la variable es cuantitativa discreta las clases serán valores numéricos del tipo $x_1, x_2, x_3, \cdots, x_k$.
- Si las variables son cuantitativas continuas las clases se definen mediante intervalos abiertos o semiabiertos, es decir de la forma:

$$(x_{i-1}, x_i), (x_i, x_{i+1}), [x_{i-1}, x_i), [x_i, x_{i+1}), (x_{i-1}, x_i], (x_i, x_{i+1}].$$

En estos casos las modalidades que contienen una clase son todos los valores numéricos posibles contenidos en el intervalo.

Por convención nosotros de aquí en adelante tomaremos siempre los $(k-1)$ primeros intervalos de la forma $(x_{i-1}, x_i]$ y el último $[x_{k-1}, x_k]$. A cada intervalo lo representaremos $x_{i-1} - x_i = I_i$.

Definición 1.19 Amplitud:

La amplitud de un intervalo se define $a_i = x_i - x_{i-1}$.

Definición 1.20 Marca de clase:

Es un valor $m_i \in I_i$ que representa a la clase y se define

$$m_i = \frac{x_i + x_{i-1}}{2}$$

La marca de clase es una forma abreviada de representar la clase.

m_i se determina de esta forma si las clases son conjuntos acotados.

1.4.0.2. Elección de clases para variables continuas

Cuando tenemos una muestra es importante escoger en una forma adecuada las clases y el número de clases k y para ello indicaremos los siguientes pasos:

- Lo primero es determinar k, entre mayor sea su valor mejor[1], pero de todas formas hay que acotarlo por que la idea es reducir el número de datos en la muestra.

$$k = \left\{ \begin{array}{cc} \sqrt{n} & \text{si } n \text{ es muy grande} \\ 1 + 3{,}22 \log n & \text{en otro caso} \end{array} \right\}$$

n se considera grande si $n \geq 40$

- Como segundo paso se determina el rango $R = x_k - x_0$
- Determinado el rango de la muestra podemos hallar la amplitud de cada intervalo que generalmente la tomamos constante:

$$a = \frac{R}{k}, \quad a_i = a \quad \forall i = 1, 2, 3, ..., k$$

- Ahora determinaremos los intervalos[2]

$$\begin{aligned} x_0 &= x_{\text{mín}} \\ x_1 &= x_0 + a \\ x_2 &= x_1 + 2a \\ x_k &= x_{\text{máx}} + ka \end{aligned}$$

Como se puede observar es posible que el valor de a no sea un número fácil de manejar entonces en estos casos como

$$x_k \geq x_{\text{máx}} > x_{\text{mín}} \geq x_0$$

entonces se varían los extremos simétricamente y a se aproxima al mayor entero es decir $a\prime = [[a]] + 1$

Ejemplo: 1.4.6

Queremos observar el peso de las personas en una población y se toma una muestra de 21 individuos, los cuales están tabulados en la siguiente tabla

Peso en Kg

58	42	51	54	40	40	49
56	58	57	59	63	58	66
70	73	71	69	70	68	64

Construir la tabla de frecuencias:

Solución

Lo primero que hay que identificar es la variable y en este caso vemos que la variable es de tipo cuantitativa continua por lo que ahora hay que determinar los intervalos y su longitud para que la perdida de información no sea significativa entonces sea X la variable peso, utilizaremos la fórmula $k = 1 + 3{,}22 * \log_{10} 21 = 5.2575 \approxeq 6$ aunque podríamos escoger también $k = \sqrt{21} \approxeq 5$.

[1] Se aconseja que se escojan entre 5 y 20 clases dependiendo del número de datos

[2] Se aconseja que las marcas de clases coincidan con un gran número de datos y que los datos no sean extremos de las clases, para que los cálculos posteriores queden mejor aproximados.

Ahora hallemos $R = 73 - 40 = 33 \Longrightarrow a_i = a = \frac{33}{6} = 5.5$
$x_0 = x_{\text{mín}} = 40$
$x_5 = x_{\text{máx}} = 73$

i	Intervalos	m_i	n_i	f_i	N_i	F_i
1	40-45,5					
2	45,5-51					
3						
4						
5	62-67.5					≈ 1
					21	≈ 1
			21	≈ 1		

Existe otra posibilidad de construir la tabla y es escogiendo por ejemplo $a\prime = 7 \Longrightarrow R\prime = a\prime \cdot 5 = 35$
$d = R\prime - R = 2 \Longrightarrow x_0 = x_{\text{mín}} - \frac{d}{2} = 39, \quad x_5 = x_{\text{máx}} + \frac{d}{2} = 74$

i	Intervalos	m_i	n_i	f_i	N_i	F_i
1	39-46					
2	46-53					
3						
4						
5	67-74					
6					21	≈ 1
			21	≈ 1		

Utilizando un software estadístico ***Statgraphics*** ®, obtenemos la siguiente tabla

```
Tablas de Frecuencia para el Peso

--------------------------------------------------------------------------------
       Límite   Límite   Marca      Absoluta   Relativa   Acumulada  Acumulada.Rel.
Clases Inferior Superior de Clase   Frecuencia Frecuencia Frecuencia Frecuencia
--------------------------------------------------------------------------------
   Antes de        38,0                0        0,0000       0
   1      38,0   44,6667  41,3333      3        0,1429       3      0,1429
   2   44,6667   51,3333     48,0      2        0,0952       5      0,2381
   3   51,3333      58,0  54,6667      6        0,2857      11      0,5238
   4      58,0   64,6667  61,3333      3        0,1429      14      0,6667
   5   64,6667   71,3333     68,0      6        0,2857      20      0,9524
   6   71,3333      78,0  74,6667      1        0,0476      21      1,0000
Después de 78,0                        0        0,0000      21      1,000
--------------------------------------------------------------------------------
```

Podemos usar R para resolver el ejercicio, en este caso cargamos el paquete *Agricolae* como se muestra a continuación

```
round(table.freq(h1), 2)

##   Lower Upper Main Frequency Percentage CF   CPF
## 1  40.0  46.6 43.3         3       14.3  3  14.3
## 2  46.6  53.2 49.9         2        9.5  5  23.8
```

```
## 3   53.2   59.8 56.5          7       33.3 12   57.1
## 4   59.8   66.4 63.1          3       14.3 15   71.4
## 5   66.4   73.0 69.7          6       28.6 21 100.0
```

```
stem(peso)
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   4 | 0029
##   5 | 14678889
##   6 | 34689
##   7 | 0013
```

1.5 Representaciones gráficas y diagramas

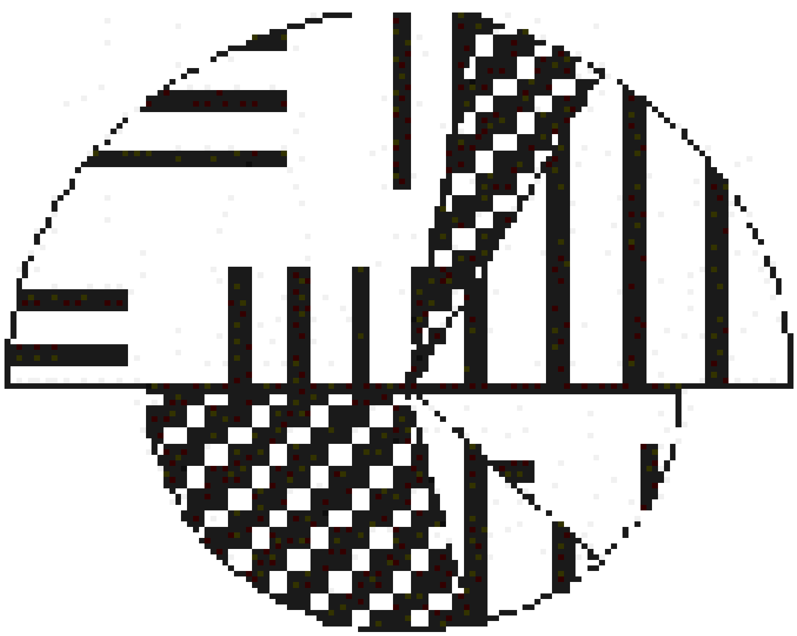
En la sección anterior hemos visto que al organizar los datos en una tabla se reduce la información, pero con ello podemos analizarlos de manera más sistemática y de esa manera podemos concentrarnos en los más importante, pero aún así a veces no es fácil observar todo lo que queremos, sobre todo si la persona interesada en los resultados no le interesa la estadística y sabe muy poco de ella, por lo que una representación gráfica simplifica aún más los datos.

1.5.1. Gráficos para variables cualitativas

i. Diagramas de barras

Se establece una especie de plano cartesiano representando las modalidades en el eje de ordenadas y las frecuencias absolutas o relativas en el eje de las abscisas, con este gráfico si se comparan varias poblaciones entre sí es conveniente utilizar las frecuencias relativas

Diagrama de barra para una variable cualitativa



ii. Diagramas circulares o sectores:

En estos diagramas se toma un círculo o cilindro y se dividen en tantos sectores como clases existan de modo que cada sector es proporcional a su frecuencia absoluta o acumulada, como se indica en los siguientes diagramas:

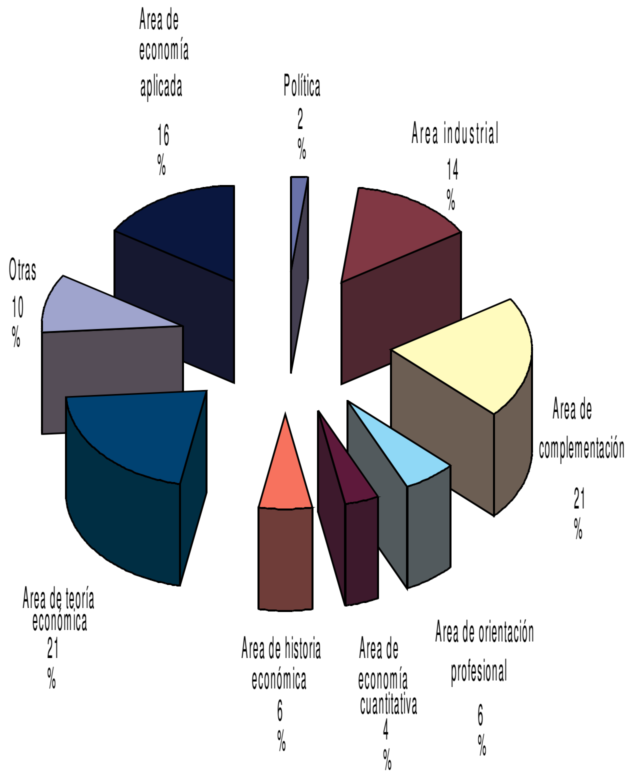
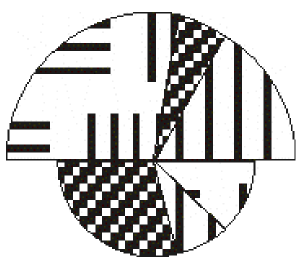


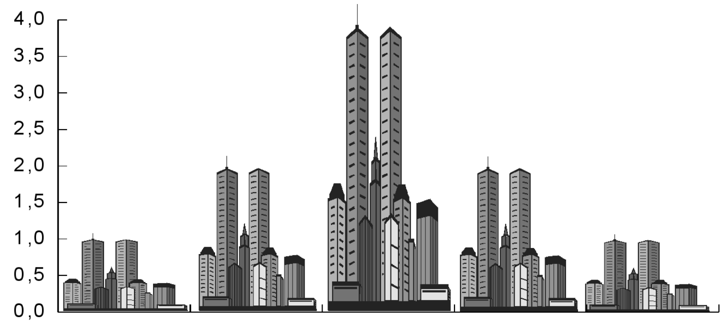
Diagrama circular

Con estos diagramas también se pueden comparar dos poblaciones utilizando dos semicírculos de radios r_1 y r_2 tal que $r_2 > r_1$ y cumplan la relación: $r_2 = r_1\sqrt{\frac{n_2}{n_1}}$ donde las n representan el tamaño de las poblaciones.



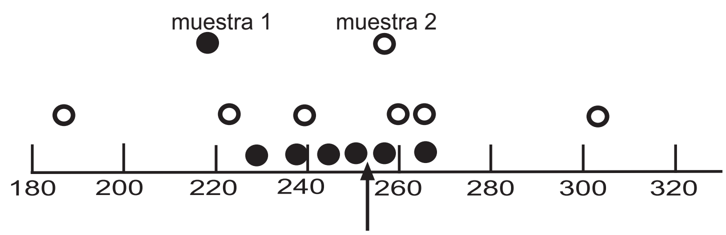
DOS MUESTRAS

iii. Pictogramas: Cuando se expresan dibujos alusivos al tema en estudio estos dibujos se hacen de tal forma que se utilizan diferentes escalas para representar la frecuencia absoluta o relativa.



1.5.2. Gráficos para variables cuantitativas

i. Diagrama de puntos: En este diagrama se coloca la frecuencia absoluta o relativa de una modalidad en una recta numérica y nos sirve para analizar dos o más modalidades cuando el número de datos es pequeño, con este gráfico analizamos fácilmente la tendencia y la variabilidad de la muestra. lo mismo que características poco usuales.



ii. Diagrama de tallo y hoja: Cuando el conjunto de datos

$$x_1, x_2, x_3, \cdots, x_n.x_1, x_2, x_3, \cdots, x_n.$$

es grande cada x_i, $\quad i = 1, 2, 3, ..., n$ tiene más de dos dígitos, entonces se dividen los x_i en dos partes . Un tallo formado por los primeros dígitos. Una hoja formada por el resto de dígitos

Ejemplo: 1.5.7

En un examen de clasificación para seleccionar alumnos que pueden ver directamente cálculo en el primer semestre en la facultad de ingeniería se obtuvieron los siguientes resultados:

$$\begin{bmatrix} 95 & 95 & 100 & 100 & 100 & 100 & 100 & 105 & 105 & 105 \\ 110 & 110 & 110 & 110 & 110 & 110 & 110 & 110 & 110 & 115 \\ 115 & 115 & 115 & 115 & 115 & 115 & 115 & 115 & 115 & 115 \\ 120 & 120 & 120 & 120 & 120 & 120 & 120 & 120 & 125 & 125 \\ 125 & 125 & 130 & 130 & 130 & 130 & 135 & 135 & 140 & 140 \end{bmatrix}$$

09	5	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	0	0	0	0	0	5	5	5	-	-	-	-	-	-	-	-	-	-	-	-
11	0	0	0	0	0	0	0	0	0	5	5	5	5	5	5	5	5	5	5	5
12	0	0	0	0	0	0	0	0	5	5	5	5	-	-	-	-	-	-	-	-
13	0	0	0	0	5	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

iii. Diagramas diferenciales: Son aquellos en los que se representan gráficamente frecuencias absolutas y relativas

iv. Diagramas integrales: Son los diagramas en los que se representan gráficamente el número de elementos que presentan una modalidad inferior o igual a una modalidad dada y se generan a tráves de las frecuencias acumuladas

Debido a que existen dos tipos de variables cuantitativas entonces debemos clasificar estos dos tipos de diagramas de acuerdo con el tipo de variable cuantitativa en estudio.

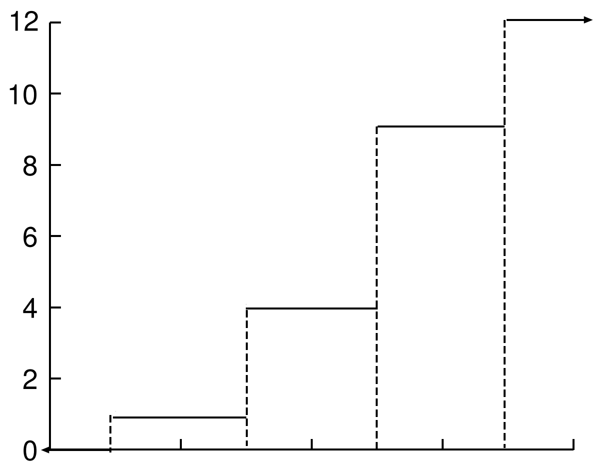
Gráficos para variables discreta

Al representar gráficamente la frecuencia absoluta o relativa de una variable discreta usamos los diagrama de barras, pero a diferencia de los diagrama de barra de las variables cualitativas las barras aquí se presentan con líneas delgadas, para indicar así la naturaleza de la variable. En el caso de los diagramas integrales tienen la forma del gráfico de una función escalonada

Ejemplo: 1.5.8

La siguiente tabla representa el número de hijos que tenían 12 familias encuestadas de un caserío cerca a Baranoa:

x_i	n_i	f_i	N_i
1	1		
2	3		
3	5		
4	3		
	12		



FRECUENCIAS ABSOLUTAS ACUMULADAS

Para las variables continuas también se pueden representar con diagramas circulares

Ejemplo: 1.5.9

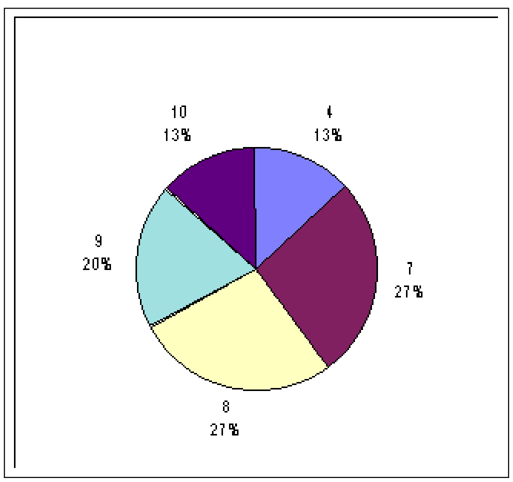
La tabla representa las notas obtenidas por los alumnos de 11en un examen de Matemáticas en 1990

Notas	**Frecuencia**	**frecuencia Relativa**
10	2	$2/15 = ,133$
9	3	$3/15 = ,200$
8	4	$4/15 = ,267$
7	4	$4/15 = ,267$
4	2	$2/15 = ,133$

(Tabla 1)

notas	$\mathbf{n}_i$	$\mathbf{f}_i$	$f_i\,\%$	$\mathbf{N}_i$	$\mathbf{F}_i$
10	2	$2/15 = ,133$	13,3	15	1,000
9	3	$3/15 = ,200$	20,0	13	,867
8	4	$4/15 = ,267$	26,7	10	,667
7	4	$4/15 = ,267$	26,7	6	,400
4	2	$2/15 = ,133$	13,3	2	,133

(Tabla 2)



Gráficos para variables continuas

Para las variables continuas existen dos tipos de gráficos:

- Los histogramas: Los cuales se construyen representando sobre cada intervalo, un rectángulo que tiene la longitud del segmento como base y la altura debe ser un valor proporcional a el valor de la frecuencia para ese intervalo.
- Polígono de frecuencias: Este se elabora uniendo los puntos que corresponden a las imágenes de las marcas de clase.

En el caso de los diagramas integrales a estos polígonos se les llama ojiva y se obtienen uniendo las abscisas a partir de los extremos de los intervalos en los que se han organizado los datos

Ejemplo: 1.5.10

Representar gráficamente la información que aparece en la siguiente tabla

Intervalos	m_i	n_i	N_i
0-2	1	1	
2-4	3	4	
4-6	5	10	
6-8	7	3	
8-10	9	1	

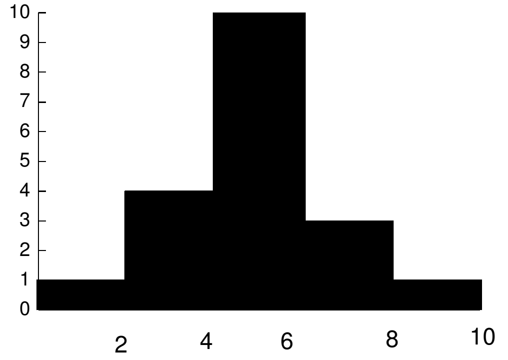
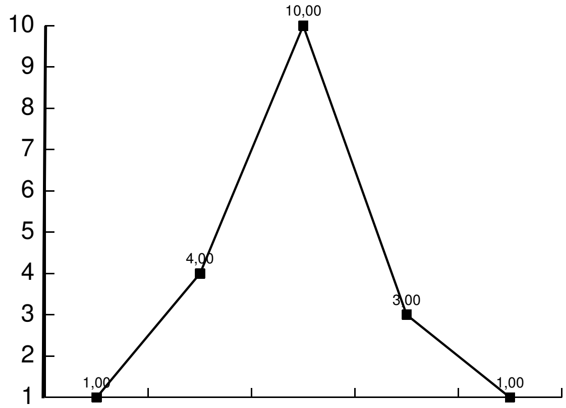


Diagrama Diferencial



Polígono de frecuencias

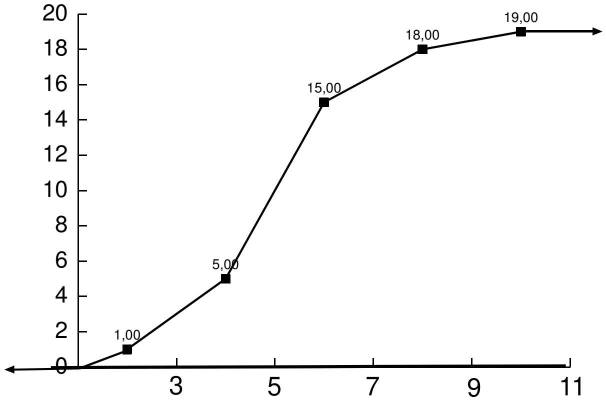


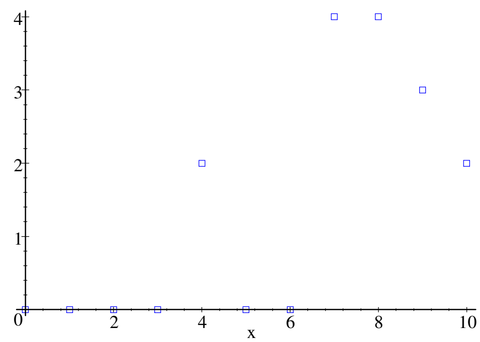
Diagrama acumulado

Ejemplo: 1.5.11

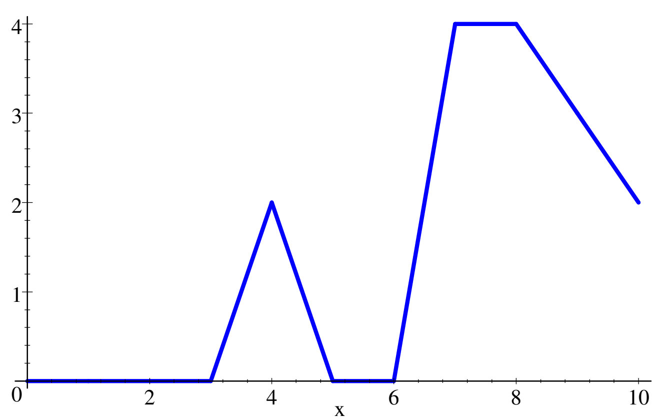
Representar gráficamente los datos de la tabla1

Notas	Frecuencia
10	2
9	3
8	4
7	4
6	0
5	0
4	2
3	0
2	0
1	0
0	0

(Tabla 1)



Test Scores



Test Scores

* Realizar una ojiva que represente los datos de la tabla2

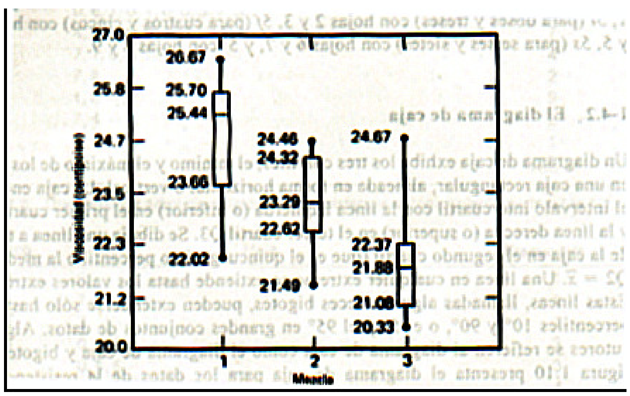
- Diagrama de caja: El diagrama de caja es una representación visual que describe varias características importantes tales coma son: la dispersión, la variabilidad, la simetría e identifica los datos que se alejan de manera poco usual del resto de los datos.

En un diagrama de caja se concentran el 50 % de los datos en un rectángulo, ubicando los valores para los cuales la frecuencia relativa acumulada son el 25 % y el 75 % respectivamente en líneas horizontales o verticales formando los lados paralelos del rectángulo, también se utiliza una línea que pasa por el centro del rectángulo para representar el dato que le corresponde a la frecuencia relativa acumulada del 50 %. De los dos lados del rectángulo antes mencionados se extiende una línea perpendicular hasta los extremos más alejados, los valores que se encuentran en esta región se llaman valores atípicos o extremos.

Ejemplo: 1.5.12

La siguiente tabla muestra los datos sobre la viscosidad de tres mezclas

mezcla 1	mezcla 2	mezcla 2
22.02	21.49	20.33
23.83	22.67	21.67
26.67	24.62	24.67
25.38	24.18	22.45
25.48	22.78	22.28
23.50	22.56	21.95
25.90	24.46	20.49
24.98	23.79	21.81

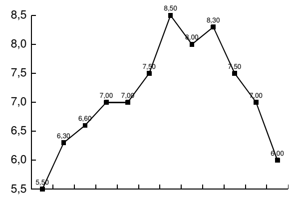


- Gráficas de serie de tiempo: Estas gráficas de series de tiempo se utilizan cuando nos interesa averiguar si el momento en que se tomaron afecta su variabilidad.

Ejemplo: 1.5.13

El número de piezas (en miles) existentes en el almacén de una determinada fábrica el último día de cada mes del año 1981, viene dado por la tabla

Meses	Ene.	Feb.	Mar.	Abr.	Ma..	Jun.	Jul.	Ag.	Sep.	Oct.	Nov.	Dic.
Piezas	5,5	6,3	6,6	7	7	7,5	8,5	8	8,3	7,5	7	6



1.5.2.1. Diagrama de Tallo y Hoja

Salida del programa estadístico ***Statgraphics*** ®, para el peso de las 27 personas

Diagrama de Tallo y Hoja para Peso: unidad = 1,0   1|2 representa 12,0

3	4|002
5	4|99
7	5|14
(10)	5|6778888899
10	6|3334
6	6|689
3	7|001

Este diagrama muestra la tabulación de frecuencias para Peso. El rango de los datos se ha dividido en 7 intervalos (llamados tallos), cada uno representado por un renglón en la tabla. Los tallos se etiquetan utilizando uno ó más dígitos indicadores para los valores que caen dentro de ese intervalo. En cada renglón, los valores individuales se representan por un dígito (llamado hoja) a la derecha de la línea vertical. Esto resulta en un histograma para los datos del cual uno puede recuperar, al menos, dos dígitos significativos de cada valor. Si hay algunos puntos muy alejados del resto (llamados puntos lejanos), se colocan en tallos alto y bajo separados. En este caso, no hay puntos alejados

Salida obetnida con R para el mismo ejemplo

```
stem(peso)

##
##   The decimal point is 1 digit(s) to the right of the |
##
##   4 | 0029
##   5 | 14678889
##   6 | 34689
##   7 | 0013
```

1.6 Problemas

1. Decir si los datos que representan los siguientes item son continuos o,discretos o cualitativos
   a) Color de los ojos
   b) Letras usadas en un párrafo
   c) Velocidad de un coche
   d) Edad de un trabajador
   e) Número de billetes de 20000 que circulan en una ciudad
   f) Número de estudiantes matriculados en la Universidad
   g) Caudal del río,medido en diferentes días del año

2. El conjunto de datos adjunto está formado de observaciones del gasto de agua en regaderas (L/min) para una muestra de $n = 128$ casas de un sector exclusivo de Barranquilla

4,6	12,3	7,1	4,0	9,2	6,7	6,9	11,5	5,1	3,8
11,2	10,5	14,3	8,0	8,8	6,8	5,1	5,6	9,6	7,5
7,5	6,2	5,8	2,3	3,4	10,4	9,8	6,6	3,7	6,4
6,0	8,3	6,5	7,6	9,3	9,2	7,3	5,0	6,3	13,8
6,2	5,4	4,8	7,5	6,0	6,9	10,8	7,5	6,6	5,5
3,3	7,6	3,9	11,9	2,2	15,0	7,2	6,1	15,3	18,9
7,2	5,4	5,5	4,3	9,0	12,7	11,5	7,4	5,0	3,5
8,2	8,4	7,3	10,3	11,9	6,0	5,6	9,5	9,3	10,4
9,7	5,1	6,7	10,2	6,2	8,4	7,0	4,8	5,6	10,5
14,6	10,8	15,5	7,5	6,4	3,4	5,5	6,6	5,9	15,0
9,6	7,8	7,0	6,9	4,1	3,5	11,9	3,7	5,7	6,8
11,3	9,3	9,6	10,4	9,3	6,9	9,8	9,1	10,6	4,5
6,2	8,3	3,2	4,9	5,0	6,0	8,2	6,3		

   a) Dibuje un histograma de frecuencias absolutas en el eje vertical

   b) Dibuje un diagrama acumulado

   c) Construya una representación de tallo y hoja

   d) Dibuje una gráfica de series de tiempo

   e) Tome tres opciones diferentes de intervalos de clase para el conjunto de datos

      1) Intervalos largos e iguales

      2) Intervalos cortos e iguales

      3) Intervalos de diferente rango

   f) Realice un diagrama circular

   g) Compare las gráficas y determine cual de ellas representa mejor los datos

   h) ¿La gráfica está centrada o sesgada?

   i) ¿Existen Puntos inusuales?, es decir muy alejados de la mayoría

3. Un artículo sobre semilla de maní (Sept de 1990) reportó los siguientes resultados

Cremoso	56	44	62	36	39	53	50	65	56
	68	41	30	40	50	56	30	22	40
Crujiente	62	53	75	42	47	40	34	62	52
	50	34	42	36	75	80	47	56	62

   a) Construya un diagrama de tallo y hojas para las muestras

   b) Construya un histograma para los muestras

   c) Realice un diagrama de puntos en ambos casos

   d) Realice sendos diagramas de cajas

   e) Compare las dos muestras

   f) Obtuvo los mismos resultados al comparar

   g) ¿Que decisión tomaría usted? Explique porqué.

4. Construya un diagrama de barras para los datos representados en la siguiente tabla.

X	n
480	75
498	56
516	42
534	30
552	21
570	15
588	11
606	6
624	2
720	1

   Donde X representa el sueldo en miles de pesos que reciben los empleados de una empresa industrial de acuerdo con la labor que desempeñan

5. Considere la muestra de calificaciones de un grupo de primaria

a	c	d	b	c	c	c	d	f	f
d	f	a	d	c	b	c	d	d	b

   Construya un

   a) Diagrama de barras

   b) Diagrama circular

6. En un párrafo se utilizaron las siguientes letras

a	b	c	e	h	t	r	o	i	u	w	x
20	15	12	15	3	10	4	20	12	10	1	1

   a) Diagrama de barras

   b) Diagrama circula

7. ¿Qué clase de gráficas son apropiadas para los datos

   a) Nominales

   b) Ordinales

   c) Cuantitativos discretos

   d) Cuantitativos continuos

8. Use los datos de la gráfica 1.3 para realizar una tabla de frecuencias

9. De acuerdo con los datos de la gráfica 1.4 realice una tabla de frecuencias

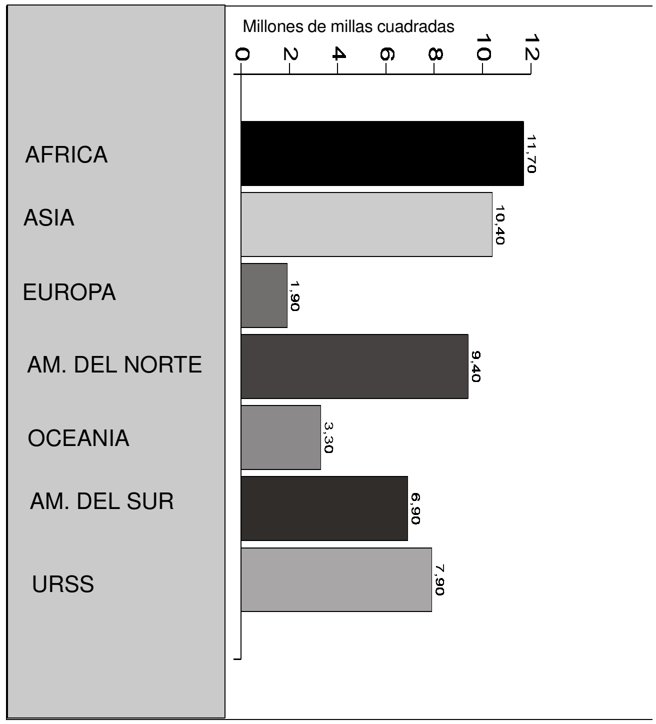


Figura 1.2: AREAS DE LOS CONTINENTES

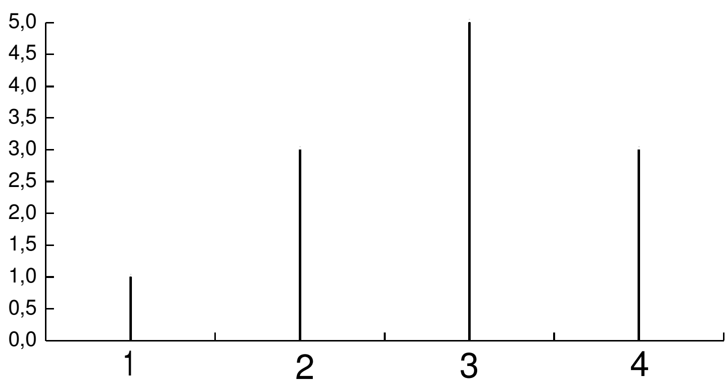
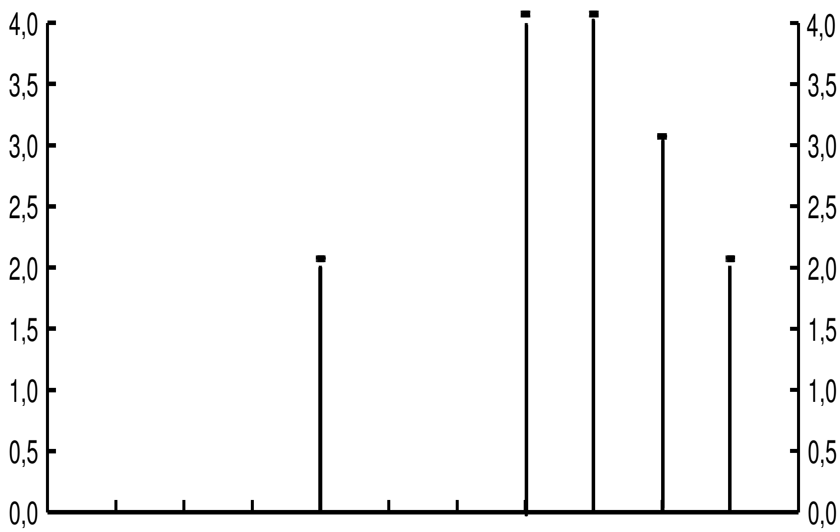


Figura 1.3: FRECUENCIAS ABSOLUTAS


4,0
3,5
3,0
2,5
2,0
1,5
1,0
0,5
0,0

CAPÍTULO

2 Medidas descriptivas

Contenido Del Capítulo

2.1 Introducción

En el capítulo anterior analizamos los los datos de una muestra utilizando diagramas o gráficos con lo que nos dimos cuenta que al tratar de realizar inferencias a partir de ellos, los gráficos no estaban bien definidos ya que para un mismo grupo de datos podemos realizar gráficos diferentes, por los que debemos definir expresiones bien definidas para realizar mejores inferencias.

Estas expresiones
son cantidades matemáticas con las que podemos obtener conclusiones validando las inferencias, estas expresiones matemáticas que representan propiedades importantes de la muestra se llaman medidas descriptivas.

2.2 Medidas de tendencia central

2.2.1. Medias

2.2.1.1. Media aritmética

La media aritmética de una muestra es la suma de cada uno de los valores posibles multiplicado por su frecuencia, es decir.

Si la siguiente tabla representa la tabla de frecuencia de la muestra

M	n_i	f_i
x_1	n_1	f_1
$\vdots$	$\vdots$	$\vdots$
x_k	n_k	f_k

La media es el valor :

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i n_i}{n} \tag{2.1}$$

y si los datos no están ordenados entonces

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i}{n} \tag{2.2}$$

En la definición de media se consideró que la variable de interés X es discreta, pero si la variable X no es discreta sino continua. En la fórmula se reemplaza cada valor x_i por la marca de clase correspondiente es decir

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i n_i}{n} \tag{2.3}$$

Este proceso hace que la media aritmética difiera de la media obtenida según (2.1), es decir habrá una perdida de precisión que será mayor en cuanto mayor sea la diferencia entre las marcas de clase y los valores reales, o sea entre mayor sea la longitud a_i de los intervalos

Desviaciones

$$\sum_{i=1}^{k} (x_i - \overline{x})\, n_i \approx 0 \qquad \textit{Dispersión total} \tag{2.4}$$

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 \gtrapprox 0 \quad \textit{Error cuadrático} \tag{2.5}$$

$$\sum_{i=1}^{n} |x_i - \bar{x}| \gtrapprox 0 \quad \textit{Error absoluto} \tag{2.6}$$

Estas fórmulas determinan la variabilidad entre los datos.

Ejemplo: 2.2.1

Obtener la media y las desviaciones con respecto a la media

$N^o de\ caras$	n_i	$x_i n_i$	$x_i - \bar{x}$
0	38	0	
1	144	144	
2	342	684	
3	287	861	
4	164	656	
5	25	125	
	1000		

Ejemplo: 2.2.2

Determinar la media aritmética del ejemplo 1.10

Intervalos	m_i	n_i	N_i	$m_i n_i$
$0-2$	1	2		
$2-4$	3	1		
$4-6$	5	4		
$6-8$	7	3		
$8-10$	9	2		

Proposición 1

Dados r grupos con $n_1, n_2, n_3, \cdots, n_r$ observaciones, si $\bar{x}_1$, $\bar{x}_2, \bar{x}_3, \cdots, \bar{x}_r$ son las respectivas medias, entonces la media de las $n = n_1 + n_2 + n_3 + \cdots + n_r$ observaciones es:

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i \bar{x}_i}{n} \tag{2.7}$$

Desventajas de la media La media es una medida muy usada en estadística, pero a pesar de eso posee ciertas desventajas

- La media aritmética es muy sensible a los valores extremos, es decir si una medida se aleja mucho de las otras hará que la media se aproxime mucho a ella
- No se recomienda usar cuando los datos se desplazan hacia los extremos
- En el caso de variables continuas depende de los intervalos de clase
- en le caso de variables discretas el valor puede no ser un valor de la muestra.

2.2.1.2. Otras medias

Media geométrica [1] $\bar{x}_g$ es la media de los logaritmos de las medidas

$$\lg \bar{x}_g = \frac{\sum_{i=1}^{k} \lg \bar{x}_i}{n}$$

es decir

$$\bar{x}_g = \sqrt[n]{x_1.x_2.x_3 \cdots x_3}$$

y en el caso de datos agrupados tenemos

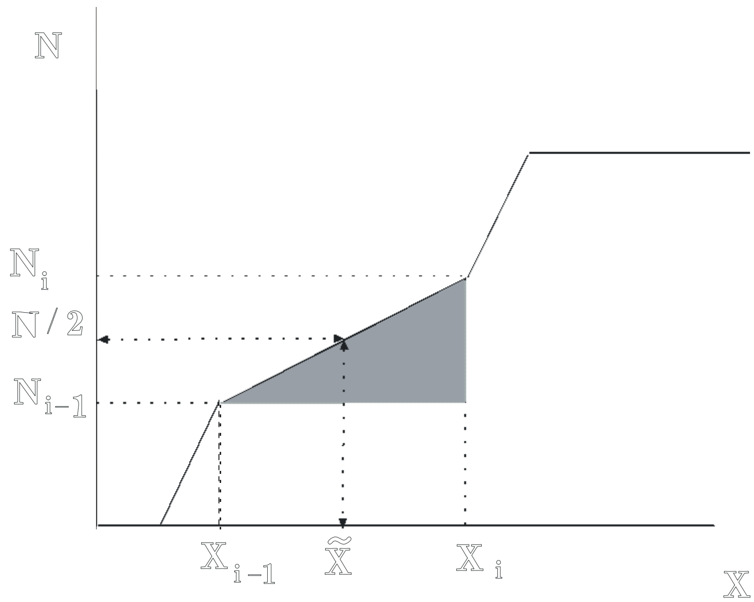
$$\bar{x}_g = \sqrt[n]{x_1^{n_1}.x_2^{n_2}.x_3^{n_3} \cdots x_3^{n_k}}$$

2.2.1.3. Media armónica

$\bar{x}_a$ se define como el recíproco de la media aritmética de los recíprocos :

$$\frac{1}{\bar{x}_a} = \frac{\sum_{i=1}^{n} \frac{1}{x_i}}{n}$$

[1] Se aconseja usar cuando la diferencia entre los extremos es muy grande comparada con la diferencia entre los otros datos o cuando la diferencia entre un dato y otro siempre aumenta.



2.2.1.4. Media cuadrática

[2]$\bar{x}_c$ es la raíz cuadrada de la media aritmética de los cuadrados

$$\bar{x}_c = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

2.2.2. La mediana

Sea X una variable discreta cuyas observaciones han sido ordenadas de mayor de mayor a menor, entonces se le llama mediana $\widetilde{x}$ al primer valor de la variable que deja por debajo de si el 50 % de las observaciones es decir si n es el número de observaciones, la mediana será la observación $\left[|\frac{n}{2}|\right]+1$

Definición 2.1 Mediana

Sea $x_{(1)}, x_{(2)}, x_{(3)}, \cdots, x_{(n)}$ las observaciones de una muestra para una variable X donde $x_{(1)}$ representa la observación más pequeña, $x_{(2)}$ la observación que le sigue en valor y así sucesivamente $x_{(n)}$ denota la observación de mayor valor, entonces la mediana se define

$$\widetilde{x} = \begin{cases} x_{([n+1]/2)} & \text{si } n \text{ impar} \\ \frac{x_{(n/2)}+x_{([n/2]+1)}}{2} & \text{si } n \text{ par} \end{cases}$$

En el caso de variables continuas, las clases vienen dadas por intervalos como se indicó en el capítulo anterior por tal razón para determinar la mediana se escoge el intervalo donde se encuentra el valor para el cual están debajo de él la mitad de los datos. Entonces a partir de ese intervalo se observan las frecuencias absolutas acumuladas y se aplica la siguiente fórmula

$$\widetilde{x} = x_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i} a_i$$

de aquí se puede deducir que $\widetilde{x}$ el "punto" que divide al histograma en dos partes de áreas iguales

Gráficamente la moda para variables continuas se determina de acuerdo con la siguiente gráfica

2.2.2.1. Propiedades y desventajas de la mediana

1. Tiene la ventaja de no ser afectada por los valores extremos y por eso se aconseja para distribuciones para las cuales los datos no se concentran en el centro
2. Es fácil de calcular
3. En el caso de variables discretas el valor de la mediana es un valor de la variable

[2]Es usada muy a menudo cuando se estudian experimentos científicos

4. El mayor defecto es que las propiedades matemáticas son muy complicadas y esto hace que muy poco se use para realizar inferencias

5. Es función de los intervalos escogidos en el caso de variables continuas

Ejemplo: 2.2.3

Sea X una variable discreta que ha presentado las siguientes observaciones

$X \rightsquigarrow 2, 5, 7, 9, 12 \Rightarrow \widetilde{x} = 7$ y $\overline{x} = 7$

Ahora si cambiamos la ultima observación, no se afectará la mediana pero si la media

$X \rightsquigarrow 2, 5, 7, 9, 125 \Longrightarrow \widetilde{x} = 7$ y $\overline{x} = 29, 6$

Además la media no es valor de la variable

Ejemplo: 2.2.4

Media

Obtener la media y la mediana de la tabla frecuencias siguientes

$x_{i-1} - x_i$	m_i	n_i	a_i	$x_i n_i$	N_i
0-10		80			
10-20		60			
20-30		30			
30-100		5			
100-500		5			

$\overline{x} = \frac{\quad}{\quad} \qquad \widetilde{x} = \qquad + \frac{\qquad}{\qquad} \qquad =$

Realizar el histograma de frecuencia y comparar la media y la mediana

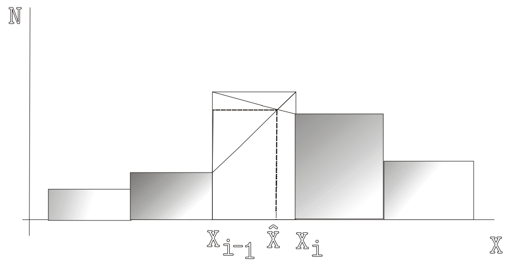


Figura 2.2: La moda

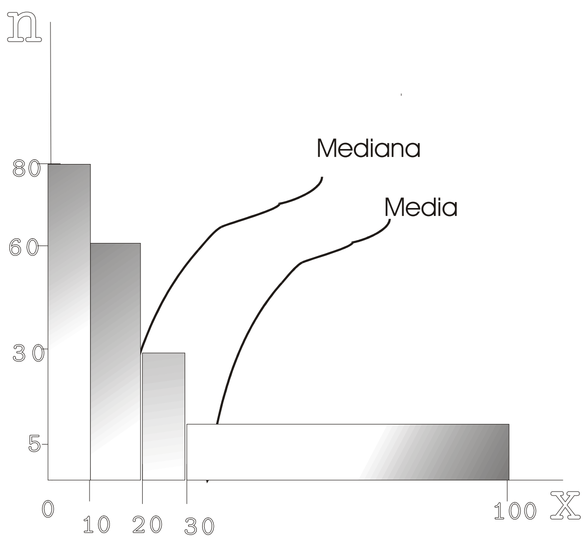


Figura 2.1: Comparación de la media y la mediana

2.2.2.2. La moda

Llamaremos moda $\widehat{x}$ a cualquier máximo relativo de la distribución de frecuencias, es decir, cualquier valor de la variable que posea una frecuencia mayor que su anterior y posterior valor.

En el caso de variables continuas es más correcto hablar de intervalos modales. Luego de determinar el intervalo de clase o intervalo modal,que es aquel para el cual la distribución de frecuencia posee un máximo relativo, se determina la moda utilizando la siguiente fórmula

$$\widehat{x} = x_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} a_i$$

La cual se obtiene geométricamente como se indica en la fig. 2.3

Propiedades de la moda

La moda posee la siguientes propiedades

- Es muy fácil de calcular

- Puede no ser única
- Es función de los intervalos de su amplitud,número y límites

2.2.2.3. Relación entre la media, la moda y la mediana

En el caso de distribuciones unimodales, la mediana está con frecuencia comprendida entre la media y la moda (incluso más cerca de la media).

En distribuciones que presentan cierta inclinación, es más aconsejable el uso de la mediana. Sin embargo en estudios relacionados con propósitos estadísticos y de inferencia suele ser más apta la media.

Veamos un ejemplo de cálculo de estas tres magnitudes.

Ejemplo: 2.2.5

Consideramos una tabla estadística relativa a una variable continua, de la que nos dan los intervalos, las marcas de clase m_i, y las frecuencias absolutas, n_i, para determinar las tres medidas de tendencia central básicas

Intervalos	m_i	n_i	N_i	$m_i n_i$
0-2	1	2		
2-4	3	1		
4-6	5	4		
6-8	7	3		
8-10	9	2		

La mediana es el valor que deja debajo de sí el 50 % de los datos, en la tabla en la columna de las frecuencias absolutas acumuladas observamos que sexto dato se encuentra en la tercera clase, es decir

$i = 3$

$(x_{i-1}, x_i] = (4, 6]$ intervalo donde se encuentra la mediana

$\widetilde{x} = \widetilde{x} = x_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i} a_i = 4 + \frac{\frac{12}{2} - 3}{4} \cdot 2 = 5,5$

para calcular la moda,debemos encontrar el o los intervalos modales, buscando los máximos relativos en las columnas de las frecuencias absolutas n_i.

Encontramos dos intervalos con un máximo relativo,

para $i = 1$ cuya frecuencia absoluta es 2

para $i = 3$ cuya frecuencia absoluta es 4.

Por tanto la moda en cada caso se calcula

cuando $i = 1 \Rightarrow$

$\widehat{x} = x_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} a_i = 0 + \frac{2-0}{(2-0)+(2-1)} \cdot 2 = 1,\widehat{3}$

cuando $i = 3 \Rightarrow$

$\widehat{x} = x_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} a_i = 4 + \frac{4-1}{(4-1)+(4-3)} \cdot 5 = 5,5$

Ahora lo resolveremos geométricamente

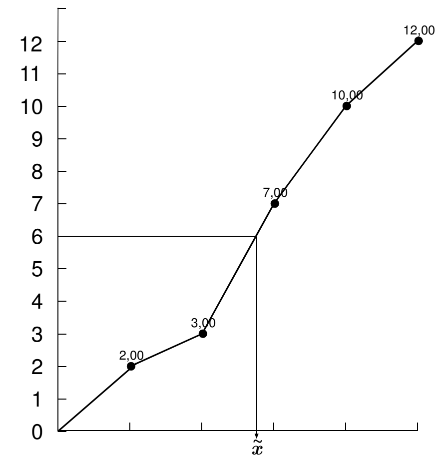


Figura 2.3: La media

$$\tilde{x} = 5,5$$

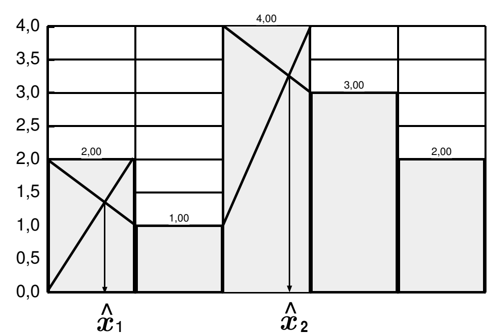


Figura 2.4: La moda

$$\hat{x}_1 = 1,33\ ;\ \ \hat{x}_2 = 5,5$$

La media es $\bar{x} = 5.\hat{3}$

2.3 Medidas de posición

A veces es importante obtener los valores de la variable que dividen la población en cuatro,diez o cien partes iguales,usualmente llamados cuartiles deciles y percentiles respectivamente.

El procedimiento es similar al utilizado para determinar la mediana, como lo indicaremos ahora.

2.3.0.1. Percentil

Para una variable discreta, se define el percentil de orden k, como la observación que deja por debajo de si el k % de la población es decir

$$N_k = n\frac{k}{100},\ \text{si } n \text{ es impar}$$

$$\text{es decir } p_k = x_{[|(n+1)\frac{k}{100}|]}$$

donde en sub índice $[|(n+1)\frac{k}{100}|]$ indica que es la posición k a la que le corresponde ese valor de la frecuencia absoluta acumulada. Si n es par

$$p_k = \frac{x_{[|n\frac{k}{100}|]} + x_{[|n\frac{k}{100}|]+1}}{2}$$

En el caso de variables continuas se busca el intervalo donde se encuentra p_k, es decir se busca el valor que deja por debajo de si el $k\,\%$ de las observaciones y se determina el intervalo $(x_{i-1}, x_i]$ donde se encuentra y se utiliza la relación

$$p_k = x_{i-1} + \frac{n\frac{k}{100} - N_{i-1}}{n_i} \cdot a_i$$

para determinarlo geométricamente observe la figura

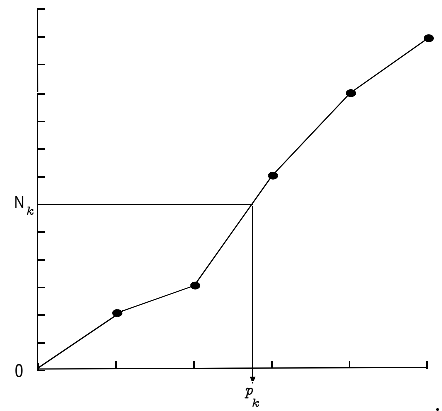


Figura 2.5: Percentil

2.3.0.2. Quartiles

Los cuartiles son tres y se definen:

- $Q_1 = p_{25}$
- $Q_2 = p_{50}$
- $Q_3 = p_{75}$

2.3.1. Deciles

De manera análoga se definen los deciles

los deciles son los valores que dividen las observaciones en 10 grupos de igual tamaño es decir son el conjunto

$D_1, D_2, D_3, \cdots, D_{10}$ y se definen

$D_i = p_{10\cdot i} \qquad i = 1, 2, 3, \cdots, 10$

Ejemplo: 2.3.6

1. Dada la siguiente tabla de distribución de frecuencia Determine los tres cuartiles

x_i	n_i	N_i
0	14	
1	10	
2		
3	26	
4	20	
5	15	100
	100	

Solución
Primer cuartil

$$\frac{n}{4} = 25 \Rightarrow N_i > \frac{n}{4} = 39 \Rightarrow Q_1 = 2$$

Segundo cuartil

$$\frac{2n}{4} = 50 \Rightarrow N_i > \frac{2n}{4} = 65 \Rightarrow Q_2 = 3$$

Tercer cuartil

$$\frac{3n}{4} = 75 \Rightarrow N_i > \frac{3n}{4} = 85 \Rightarrow Q_3 = 4$$

Ejemplo: 2.3.7

Calcular los cuartiles en las siguiente distribución de una variable continua

x_{i-1}	n_i	N_i
0-1	10	
1-2	12	
2-3		
3-4	10	
4-5	7	51
	51	

primer cuartil

$$\frac{n}{4} = 12,75 \Rightarrow 22 = N_i > \frac{n}{4} \Rightarrow Q_1 \in (1, 2]$$
$$Q_1 = x_{i-1} + \frac{n/4 - N_{i-1}}{n_i} \cdot a_i = 1 + \frac{12,75 - 10}{12} \cdot 1 = 1,23$$

segundo cuartil

$$\frac{2n}{4} = 25,5 \Rightarrow 34 = N_i > \frac{2n}{4} \Rightarrow Q_2 \in (2, 3]$$
$$Q_2 = x_{i-1} + \frac{2n/4 - N_{i-1}}{n_i} \cdot a_i = 2 + \frac{25,5 - 22}{12} \cdot 1 = 2,29$$

Tercer cuartil

$$\frac{3n}{4} = 38,5 \Rightarrow 44 = N_i > \frac{3n}{4} \Rightarrow Q_3 \in (3, 4]$$
$$Q_3 = x_{i-1} + \frac{3n/4 - N_{i-1}}{n_i} \cdot a_i = 3 + \frac{38,25 - 34}{10} \cdot 1 = 3,45$$

Ejemplo: 2.3.8

La distribución de una variable tiene por polígono acumulativo de frecuencias relativas.Si el número total de observaciones es de 50

Intervalo	m_i	F_i	a_i
0-5	2,5	0,2	5
5-7	6	0,7	2
7-12	9,5	0,8	5
12-15	13,5	1	3

1. Elaborar una tabla estadística con los siguientes elementos : intervalos de clase, marcas de clase, frecuencia absoluta, frecuencia absoluta acumulada, frecuencia relativa y frecuencia relativa acumulada.
2. Cuantas observaciones tuvieron un valor inferior a 10, cuantas inferior a 8 y cuantas fueron superior a 11.
3. Calcule las modas.
4. Determine los cuartiles

1. En la siguiente tabla se proporciona la información pedida y algunos cálculos auxiliares que nos permitirán responder otras preguntas.

Intervalo	m_i	n_i	N_i	f_i	F_i	a_i
0-5	2,5	10			0,2	5
5-7	6	25			0,7	2
7-12	9,5	5			0,8	5
12-15	13,5	10			1	3

2. Calcularemos ahora el número de observaciones que se han pedido

- Hasta el segundo intervalo van 35 datos ahora analizaremos cuantos datos tienen un valor menor que diez en el tercer intervalo

$$\begin{array}{lcc} 7\,a\,12 & \longrightarrow & 5 \\ 7\,a\,10 & \longrightarrow & x \end{array} \Longleftrightarrow \begin{array}{ccc} 5 & \longrightarrow & 5 \\ 3 & \longrightarrow & x \end{array} \Longrightarrow x = 3$$

  Por tanto el número de observaciones que tomaron un valor inferior a 10 es 38

- De igual forma que en el paso anterior hasta el segundo intervalo van 35 observaciones y en el tercer intervalo tenemos que

$$\begin{array}{lcc} 7\,a\,12 & \longrightarrow & 5 \\ 7\,a\,8 & \longrightarrow & x \end{array} \Longleftrightarrow \begin{array}{ccc} 5 & \longrightarrow & 5 \\ 1 & \longrightarrow & x \end{array} \Longrightarrow x = 1$$

  Por tanto el número de observaciones que tomaron un valor menor que 8 es 36

- En el cuarto intervalo hay 10 observaciones mayores que 11, por tanto hay que determinar cuantas mayores que 11 hay en el tercer intervalo

$$\begin{array}{lcc} 7\,a\,12 & \longrightarrow & 5 \\ 7\,a\,11 & \longrightarrow & x \end{array} \Longleftrightarrow \begin{array}{ccc} 5 & \longrightarrow & 5 \\ 4 & \longrightarrow & x \end{array} \Longrightarrow x = 4$$

  Entonces hay cuatro observaciones entre 7 y 11 por tanto mayor que 11 hay 1 es decir hay 11 observaciones mayores que 11

3. Tenemos dos modas

- Cuando $i = 2$

$$\widehat{x_1} = x_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} a_i$$
$$\widehat{x_1} = 5 + \frac{25 - 10}{(25 - 10) + (25 - 5)} \cdot 2 = 5,86$$

- Cuando $i = 4$

$$\widehat{x_2} = x_{i-1} + \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} a_i$$
$$\widehat{x_2} = 12 + \frac{10 - 5}{(10 - 5) + (10 - 0)} \cdot 7 = 14,\widehat{3}$$

4 Los Cuartiles son

$$Q_1 = x_{i-1} + \frac{n/4 - N_{i-1}}{n_i} \cdot a_i = 5 + \frac{12,5 - 10}{25} \cdot 2 = 5,2$$
$$Q_2 = x_{i-1} + \frac{2n/4 - N_{i-1}}{n_i} \cdot a_i = 5 + \frac{25 - 10}{25} \cdot 2 = 6,2$$
$$Q_3 = x_{i-1} + \frac{3n/4 - N_{i-1}}{n_i} \cdot a_i = 7 + \frac{37,5 - 35}{5} \cdot 5 = 9,5$$

2.4 Medidas de variabilidad o dispersión

Las medias de posición o medidas de tendencia central nos indican donde se sitúa un grupo de medidas. Los de variabilidad o dispersión nos indican si esas medidas o valores están próximas entre sí o si por el contrario están dispersas.

Una medida razonable da la variabilidad podría ser la amplitud o rango, el cual se obtiene restando el valor o medida mínima de la medida máxima de un conjunto de observaciones

Esta forma de determinar la variabilidad es fácil de calcular y tiene las mismas unidades de las medidas, pero tiene algunos inconvenientes

◦ No utiliza todas las observaciones si no dos de ellas

◦ Es afectada por las observaciones extremas

◦ El rango aumenta con el número de medidas, o bien queda igual. En ningún caso disminuye

Por eso a continuación determinaremos otras formas de medir la dispersión de una variable

2.4.1. Desviación media

Se define la desviación media D_m como la media de las diferencias en valor absoluto de los valores de la variable a la media, es decir, si tenemos un conjuntos de n observaciones $x_1, x_2, x_3, \cdots, x_n$ entonces

$$D_m = \frac{1}{n} \sum_{i=1}^{n} |x_i - \overline{x}|$$

Si los datos están agrupados en una tabla estadística es más sencillo usar la relación

$$D_m = \frac{1}{n} \sum_{i=1}^{k} |x_i - \overline{x}| \cdot n_i$$

Como se observa, la desviación media guarda las mismas dimensiones que las observaciones. La suma de los valores absolutos es relativamente fácil de calcular, pero esta simplicidad tiene un inconveniente: Desde el punto de vista geométrico, la distancia que induce la desviación media en el espacio de observaciones no es natural (no permite definir ángulos entre dos conjuntos de observaciones). Esto hace que sea muy engorroso trabajar con ella a la hora de hacer inferencias.

2.4.2. Varianza y desviación típica

Como forma de medir la dispersión de los datos hemos descartado:

- $\sum_{i=1}^{n} (x_i - \overline{x})$, pues sabemos que esa suma vale 0, ya que las desviaciones con respecto a la media se compensan al haber términos en esa suma que son de signos distintos.
- Para tener el mismo signo al sumar las desviaciones con respecto a la media podemos realizar la suma con valores absolutos. Esto nos lleva a la D_m, pero como hemos

mencionado, tiene poco interés por las dificultades que presenta.

- Si las desviaciones con respecto a la media las consideramos al cuadrado, $(x_i - \overline{x})^2$, de nuevo obtenemos que todos los sumandos tienen el mismo signo (positivo). Esta es además la forma de medir la dispersión de los datos de forma que sus propiedades matemáticas son más fáciles de utilizar. Vamos a definir entonces dos medidas que serán fundamentales en el resto del curso:

La varianza y la desviación típica.

La varianza, S_n^2, se define como la media de las diferencias cuadráticas de n puntuaciones con respecto a su media aritmética, es decir

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

Para datos agrupados en tablas, usando las notaciones establecidas en el capítulo anterior, la varianza se puede escribir como

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x})^2 \cdot n_i$$

Una fórmula equivalente para el cálculo de la varianza está basada en lo siguiente:

$$\begin{aligned} S_n^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 n_i - \overline{x}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - 2x_i\overline{x} + \overline{x}^2\right) \\ \frac{1}{n}\sum_{i=1}^{n} x_i^2 &- 2\overline{x}\frac{\sum_{i=1}^{n} x_i}{n} + \frac{1}{n}n\overline{x}^2 \end{aligned}$$

Con lo cual se tiene

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

Si los datos están agrupados en tablas, es evidente que

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 n_i - \overline{x}^2$$

La varianza no tiene la misma magnitud que las observaciones (ej. si las observaciones se miden en metros, la varianza lo hace en $metros^2$). Si queremos que la medida de dispersión sea de la misma dimensionalidad que las observaciones bastará con tomar su raíz cuadrada. Por ello se define la desviación típica, S_n, como

$$S_n = \sqrt{S_n^2}$$

Ejemplo: 2.4.9

Calcular la varianza y desviación típica de las siguientes cantidades medidas en metros:

3,3,4,4,5

Solución

Para calcular dichas medidas de dispersión es necesario calcular previamente el valor con respecto al cual vamos a medir las diferencias. Éste es la media:

$$\overline{x} = 3,8\, metros$$

La varianza es:

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 n_i - \overline{x}^2$$
$$= \frac{1}{5}(9+9+16+16+25) - 3,8^2 = 0,56m^2$$

siendo la desviación típica su raíz cuadrada:

$$S_n = \sqrt{0,56m^2} = 0,748m$$

Las siguientes propiedades de la varianza (respectivamente, desviación típica) son importantes a la hora de hacer un cambio de origen y escala a una variable. En primer lugar, la varianza (resp. Desviación típica) no se ve afectada si al conjunto de valores de la variable se le añade una constante. Si además cada observación es multiplicada por otra constante, en este caso la varianza cambia en relación al cuadrado de la constante (resp. La desviación típica cambia en relación al valor absoluto de la constante). Esto queda precisado en la siguiente proposición:

Proposición 2

Si $Y = aX + b$ entonces $S_{n,Y}^2 = a^2 S_{n,X}^2$

Prueba

Para cada observación x_i de X,$i = 1, 2, \cdots, n$, tenemos una observación de Y que es por definición $y_i = ax_i + b$, es fácil demostrar que $\overline{y} = a\overline{x} + b$. Por tanto, la varianza de Y es

$$S_{n,Y}^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - \overline{y})^2$$
$$= \frac{1}{n}\sum_{i=1}^{n} [(ax_i + b) - (a\overline{x} + b)]^2$$
$$= \frac{1}{n}\sum_{i=n}^{n} a^2 \left(x_i - \overline{x}^2\right)$$
$$= a^2 S_{n,X}^2$$

■(q.e.d)

Las consecuencias del anterior resultado eran de esperar: Si los resultados de una medida son trasladados una cantidad b, la dispersión de los mismos no aumenta. Si estos mismos datos se multiplican por una cantidad $a < 1$, el resultado tenderá a concentrarse alrededor de su media (menor varianza). Si por el contrario $a > 1$ habrá mayor dispersión.

Otra propiedad fundamental de la varianza es la siguiente:

Proposición 3

Dados r grupos, cada uno de ellos formado por n_i observaciones de media $\overline{x_i}$ y de varianza S_i^2. Entonces la varianza, S^2, del conjunto de todas las $n = n_1, n_2, n_3, \cdots, n_r$ observaciones vale

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{r} n_i S_{n_i}^2 + \frac{1}{n}\sum_{i=1}^{r} n_i \left(\overline{x_i} - \overline{x}\right)^2$$

Proposición 4

Además de las propiedades que hemos demostrado sobre la varianza (y por tanto sobre la desviación típica), será conveniente tener siempre en mente otras que enunciamos a continuación:

- Ambas son sensibles a la variación de cada una de las puntuaciones, es decir, si una puntuación cambia, cambia con ella la varianza. La razón es que si miramos su definición, la varianza es función de cada una de las puntuaciones.
- Si se calculan a través de los datos agrupados en una tabla, dependen de los intervalos elegidos. Es decir, cometemos cierto error en el cálculo de la varianza cuando los datos han sido resumidos en una tabla estadística mediante intervalos, en lugar de haber sido calculados directamente como datos no agrupados. Este error no será importante si la elección del número de intervalos, amplitud y límites de los mismos ha sido adecuada.
- La desviación típica tiene la propiedad de que en el intervalo

$$(\overline{x} - 2S_n, \overline{x} + 2S_n) = \overline{x} \pm 2S_n$$

se encuentra, al menos, el 75 % de las observaciones (véase más adelante el teorema de Thebycheff). Incluso si

tenemos muchos datos y estos provienen de una distribución normal (se definirá este concepto más adelante), podremos llegar al 95 %

- No es recomendable el uso de ellas, cuando tampoco lo sea el de la media como medida de tendencia central.

2.4.2.1. Grados de libertad

Los grados de libertad de un medida calculado sobre n datos se refieren al número de cantidades independientes que se necesitan en su cálculo, menos el número de restricciones que ligan a las observaciones y la medida.

Ilustrándolo con un ejemplo. Consideraremos una serie de valores de una variable,

$X \rightsquigarrow 2, 5, 7, 9, 12$

que han sido tomados de forma independiente.

Su media es $\overline{x} = 7$ y se ha calculado a partir de las $n = 5$ observaciones independientes x_i, que están ligadas a la media por la relación:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i$$

Luego el número de grados de libertad de la media es $n - 1 = 4$.

$$\frac{(x_i - \overline{x})^2}{n}$$

Sin embargo esas cantidades no son totalmente independientes, pues están ligadas por una restricción:

$$\sum_{i=1}^{n}\left(x_i - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)\right) = 0$$

El número de grados de libertad de la medida es el número de observaciones de la variable menos el número de restricciones que se verifican, así que en este caso, los grados de libertad de la varianza sobre los $n = 5$ datos son también $n - 1 = 4$.

Un principio general de la teoría matemática nos dice que si pretendemos calcular de modo aproximado la varianza de una población a partir de la varianza de una muestra suya, se tiene que el error cometido es generalmente más pequeño, si en vez de considerar como estimación de la varianza de la población, a la varianza muestral

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

consideramos lo que se denomina cuasivarianza muestral, S^2 que se calcula como la anterior, pero cambiando el denominador por el número de grados de libertad, $n - 1$:

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

Sobre este punto insistiremos más adelante, ya que es fundamental en estadística inferencial.

2.4.2.2. Tipificación

Se conoce por tipificación al proceso de restar la media y dividir por su desviación típica a una variable X. De este modo se obtiene una nueva variable

$$Z = \frac{X - \overline{x}}{S}$$

de media $\overline{z} = 0$ y desviación típica, $S_Z = 1$. que denominamos variable tipificada.

Esta nueva variable carece de unidades y permite hacer comparables dos medidas que en un principio no lo son, por aludir a conceptos diferentes. Así por ejemplo nos podemos preguntar si un elefante es más grueso que una hormiga determinada, cada uno en relación a su población. También es aplicable al caso en que se quieran comparar individuos semejantes de poblaciones diferentes. Por ejemplo si deseamos comparar el nivel académico de dos estudiantes de diferentes Universidades para la concesión de una beca de estudios, en principio sería injusto concederla directamente al que posea una nota media más elevada, ya que la dificultad para conseguir una buena calificación puede ser mucho mayor en un centro que en el otro, lo que limita las posibilidades de uno de los estudiante y favorece al otro. En este caso, lo más correcto es comparar las calificaciones de ambos estudiantes, pero tipificadas cada una de ellas por las medias y desviaciones típicas respectivas de las notas de los alumnos de cada Universidad.

2.4.2.3. Coeficiente de variación

Hemos visto que las medidas de centralización y dispersión nos dan información sobre una muestra. Nos podemos preguntar si tiene sentido usar estas magnitudes para comparar dos poblaciones. Por ejemplo, si nos piden comparar la dispersión de los pesos de las poblaciones de elefantes de dos circos diferentes, S nos dará información útil.

¿Pero qué ocurre si lo que comparamos es la altura de unos elefantes con respecto a su peso? Tanto la media $\overline{x}$, como la desviación típica, S, se expresan en las mismas unidades que la variable. Por ejemplo, en la variable altura podemos usar como unidad de longitud el *metro* y en la variable peso, el kilogramo. Comparar una desviación (con respecto a la media) medida en *metros* con otra en *kilogramos* no tiene ningún sentido.

El problema no deriva sólo de que una de las medidas sea de longitud y la otra sea de masa. El mismo problema se plantea si medimos cierta cantidad, por ejemplo la masa, de dos poblaciones, pero con distintas unidades. Este es el caso en que comparamos el peso en *toneladas* de una población de 100 elefantes con el correspondiente en *miligramos* de una población de 50 hormigas.

El problema no se resuelve tomando las mismas escalas para ambas poblaciones. Por ejemplo, se nos puede ocurrir medir a las hormigas con las mismas unidades que los elefantes (toneladas). Si la ingeniería genética no nos sorprende con alguna barbaridad, lo lógico es que la dispersión de la variable peso de las hormigas sea prácticamente nula (¡Aunque haya algunas que sean 1.000 veces mayores que otras!)

En los dos primeros casos mencionados anteriormente, el problema viene de la dimensionalidad de las variables, y en el tercero de la diferencia enorme entre las medias de ambas poblaciones. El coeficiente de variación es lo que nos permite evitar estos problemas, pues elimina la dimensionalidad de las variables y tiene en cuenta la proporción existente entre medias y desviación típica. Se define del siguiente modo:

$$Cv = \frac{S_X}{\overline{x}}$$

Basta dar una rápida mirada a la definición del coeficiente de variación, para ver que las siguientes consideraciones deben ser tenidas en cuenta:

∘ Sólo se debe calcular para variables con todos los valores positivos. Todo índice de variabilidad es esencialmente no negativo. Las observaciones pueden ser positivas o nulas, pero su variabilidad debe ser siempre positiva. De ahí que sólo debemos trabajar con variables positivas, para la que tenemos con seguridad que $\overline{x} > 0$.

∘ No es invariante ante cambios de origen. Es decir, si a los resultados de una medida le sumamos una cantidad positiva, $b > 0$, para tener $Y = X + b$, entonces, $C_{vY} < C_{vX}$ ya que la desviación típica no es sensible ante cambios de origen, pero si la media. Lo contrario ocurre si restamos ($b < 0$).

$$C_{vY} = \frac{S_Y}{\overline{y}} = \frac{S_X}{\overline{x} + b} < \frac{S_X}{\overline{x}} = C_{vX}$$

∘ Es invariante a cambios de escala. Si multiplicamos X por una constante a, para obtener,$Y = aX$, entonces

Es importante destacar que los coeficientes de variación sirven para comparar las variabilidades de dos conjuntos de valores (muestras o poblaciones), mientras que si deseamos comparar a dos individuos de cada uno de esos conjuntos, es necesario usar los valores tipificados.

Dada la distribución de edades (medidas en años) en un colectivo de 100 personas, obtener:

1.

La variable tipificada Z.

2.

Valores de la media y varianza de Z.

3.

Coeficiente de variación de Z.

Horas trabajadas

Horas trabajadas	*Nº de empleados*
$0-4$	47
$4-10$	32
$10-20$	17
$20-40$	4
	100

Para calcular la variable tipificada

$$Z = \frac{X - \overline{x}}{S_X}$$

partimos de los datos del enunciado. Será necesario calcular en primer lugar la media y desviación típica de la variable original ($X = años$).

Intervalos	m_i	n_i	$m_i n_i$	$m_i^2 n_i$
$0-4$	2	47	94	188
$4-10$	7	32	224	1568
$10-20$	15	17	255	3825
$20-40$	30	4	120	
Total			693	9181

$$\overline{x} = \frac{693}{100} = 6,93\,años$$
$$S_X^2 = \frac{9181}{100} - 6,93^2 = 43,78\,(años)^2$$
$$S_X = \sqrt{43,78} = 6,6\,años$$

A partir de estos valores podremos calcular los valores tipificados para las marcas de clase de cada intervalo y construir su distribución de frecuencias:

$$z_1 = \frac{2 - 6,93}{6,6} = -0,745$$
$$z_2 = \frac{7 - 6,93}{6,6} = 0,011$$
$$z_2 = \frac{15 - 6,93}{6,6} = 1,22$$
$$z_2 = \frac{30 - 6,93}{6,6} = 3,486$$

z_i	n_i	$z_i n_i$	$z_i^2 n_i$
$-0,745$	47	$-35,015$	26,086
0,011	32	0,352	0,004
1,220	17	20,720	25,303
3,486	4	13,944	48,608
	$n = 100$	0,021	100,002

$$\overline{z} = \frac{0,021}{100} \approx 0$$
$$S_Z^2 = \frac{100,02}{100} - 0^2 \approx 1$$
$$S_X = \sqrt{1} = 1$$

A pesar de que no se debe calcular el coeficiente de variación sobre variables que presenten valores negativos (y Z los presenta), lo calculamos con objeto de ilustrar el porqué:

$$C_v = \frac{S_Z}{\overline{z}} = \frac{1}{0} = \infty$$

Es decir, el coeficiente de variación no debe usarse nunca con variables tipificadas.

2.4.3. Asimetría y apuntamiento

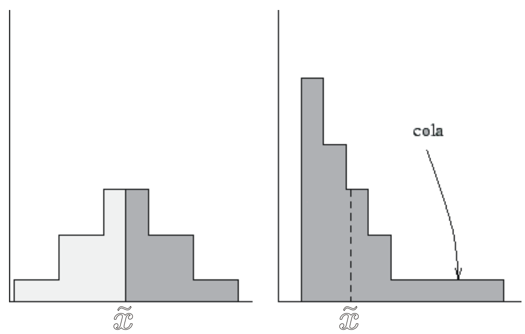
Sabemos cómo calcular valores alrededor de los cuales se distribuyen las observaciones de una variable sobre una muestra y sabemos cómo calcular la dispersión que ofrecen los mismos con respecto al valor de central. Nos proponemos dar un paso más allá en el análisis de la variable. En primer lugar, nos vamos a plantear el saber si los datos se distribuyen de forma simétrica con respecto a un valor central, o si bien la gráfica que representa la distribución de frecuencias es de una forma diferente del lado derecho que del lado izquierdo.

Si la simetría ha sido determinada, podemos preguntarnos si la curva es más o menos apuntada (larga y estrecha). Este apuntamiento habrá que medirlo comparado a cierta distribución de frecuencias que consideramos normal (no por casualidad es éste el nombre que recibe la distribución de referencia).

Estas ideas son las que vamos a desarrollar en lo que resta del capítulo.

2.4.3.1. Índices de asimetría o sesgo

Para saber si una distribución de frecuencias es simétrica, hay que precisar con respecto a qué. Un buen candidato es la mediana, ya que para variables continuas, divide al histograma de frecuencias en dos partes de igual área. Podemos basarnos en ella para, de forma natural, decir que una distribución de frecuencias es simétrica si el lado derecho de la gráfica (a partir de la mediana) es la imagen por un espejo del lado izquierdo(figura).



Cuando la variable es discreta, decimos que es simétrica, si lo es con respecto a la media.

- Se podría pensar que definir la simetría usando la mediana para variables continuas y usando la media para variables discretas es una elección arbitraria. En realidad esto no es así, pues si una variable es continua, coinciden ambos criterios de simetría (con respecto a la media y a la mediana). Es más, se tiene que media y mediana coinciden para distribuciones continuas simétricas. Por otro lado,
- En el caso de variables discretas, la distribución es simétrica si el lado derecho del diagrama se obtiene por imagen especular desde la media. En este caso coincide la media con la mediana si el número de observaciones es impar.
- Si la variable es continua simétrica y unimodal, coinciden la media, la mediana y la moda.

Dentro de los tipos de asimetría posible, vamos a destacar los dos fundamentales (2.6):

2.4.3.2. Asimetría positiva:

Si las frecuencias más altas se encuentran en el lado izquierdo de la media, mientras que en derecho hay frecuencias más pequeñas (cola).

2.4.3.3. Asimetría negativa:

Cuando la cola está en el lado izquierdo.

Cuando realizamos un estudio descriptivo es altamente improbable que la distribución de frecuencias sea totalmente simétrica. En la práctica diremos que la distribución de frecuencias es simétrica si lo es de un modo aproximado. Por otro lado, aún observando cuidadosamente la gráfica, podemos no ver claro de qué lado están las frecuencias más altas. Conviene definir entonces unos estadísticos que ayuden a interpretar la asimetría, a los que llamaremos índices de asimetría, y que denotaremos mediante A_s. Vamos a definir a continuación algunos de los índices de asimetría más usuales como son el índice basado en los tres cuartiles, el momento de tercer orden y la distancia entre la moda y la media o la media y la mediana.

2.4.3.4. Índice basado en los tres cuartiles (Yule-Bowley)

Si una distribución es simétrica, es claro que deben haber tantas observaciones entre la que deja por debajo de sí las tres cuartas partes de la distribución y la mediana, como entre la mediana y la que deja por debajo de sí un cuarto de todas las observaciones. De forma abreviada esto es,

$$Q_3 - Q_2 = Q_2 - Q_1$$

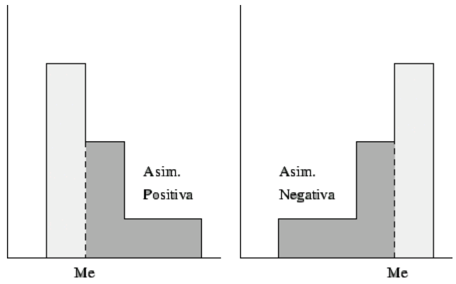


Figura 2.6: Asimetría

Una pista para saber si una distribución de frecuencias tiene una asimetría positiva, es observando la figura):

$$Q_3 - Q_2 > Q_2 - Q_1$$

Índice de asimetríaPor analogía, si es asimétrica negativa, se tendrá

$$Q_3 - Q_2 < Q_2 - Q_1$$

Para quitar dimensionalidad al problema, utilizamos como índice de asimetría la cantidad, llamada índice intercuartílico

$$A_s = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} \tag{2.8}$$

Es claro que

$$1 \geq \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1) + (Q_2 - Q_1)} \geq -1$$

El número obtenido, A_s, es invariante ante cambios de origen de referencia y de escala.

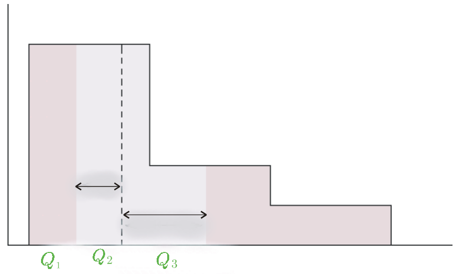


Figura 2.7: Índice de asimetría

2.4.4. Índice basado en el momento central de tercer orden

Sea X una variable cuantitativa y Llamamos momento de orden p a:

$$\mu_p = \frac{1}{n} \sum_{i=1}^{n} x_i^p$$

Se denomina momento central de orden p a la cantidad

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

Si los datos están agrupados en una tabla, m_p admite otra expresión equivalente:

$$m_p = \frac{1}{n}\sum_{i=1}^{n} n_i(x_i - \overline{x})^p$$

si $p = 1$ se tiene que $m_1 = 0$.

El momento de orden 2 es la varianza muestral:

$$m_2 = S^2$$

Es sencillo comprobar que los momentos de orden p impar, son siempre nulos en el caso de variables simétricas, ya que para cada i que esté a un lado de la media, con,$(x_i - \overline{x}) < 0$, le corresponde una observación j del otro lado de la media tal que. $(x_i - \overline{x}) = (x_j - \overline{x})$ Elevando cada una de esas cantidades a p impar, y sumando se tiene que $m_p = 0$ si la distribución es simétrica

Si la distribución fuese asimétrica positiva, las cantidades,$(x_i - \overline{x})^p$ con, $p \geq 3$ impar positivas estarían muy aumentadas al elevarse a p. Esta propiedad nos indica que un índice de asimetría posible consiste en tomar $p = 3$y definir

$$A_s = a_3 = \frac{m_3}{m_2\sqrt{m_2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{S^3} \tag{2.9}$$

que para datos organizados en una tabla sería

$$a_3 = \frac{\frac{1}{n}\sum_{i=1}^{n} n_i (x_i - \overline{x})^3}{S^3} \tag{2.10}$$

Apoyándonos en este índice, diremos que hay asimetría positiva si $a_3 > 0$, y que la asimetría es negativa si $a_3 < 0$.

Hemos dividido m_3 por el cubo de para S que a_3 sea un número abstracto sin dimensiones, independiente de la variabilidad de la variable. Por otro lado, la cantidad A_s definida por la relación (2.8) no es la misma que la definida en (2.9). Simplemente las notamos A_s para simbolizar que es un índice de asimetría.

2.9.2.6 Otros índices de asimetría

Basándonos en que si una distribución de frecuencias es simétrica y unimodal, entonces la media, la mediana y la moda coinciden, podemos definir otras medidas de asimetría, como son:

$$A_s = \frac{\overline{x} - \widehat{x}}{S} \tag{2.11}$$

o bien,

$$A_s = \frac{3\left(\overline{x} - \widetilde{x}\right)}{S} \tag{2.12}$$

Diremos que hay asimetría positiva si $A_s < 0$ y negativa si $A_s > 0$ (véase la figura).

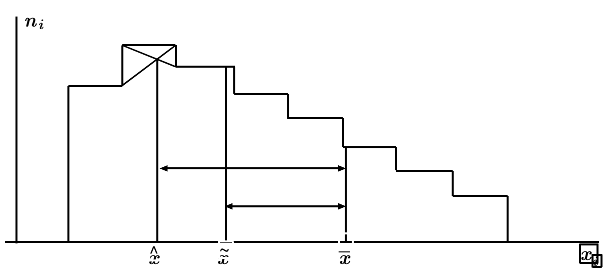


Figura 2.8: Otros índices de asimetría

Ejemplo: 2.4.10

Las edades de un grupo de personas se reflejan en la tabla siguiente:

Intervalos	n$_i$
7-9	4
9-11	18
11-12	14
12-13	27
13-14	42
14-15	31
15-17	20
17-19	1

Determinar la variabilidad de la edad mediante los estadísticos varianza, desviación típica, coeficiente de variación y rango intercuartílico. Estudie la simetría de la variable.

Solución

En primer lugar realizamos los cálculos necesarios a partir de la tabla de frecuencias:

$Intervalos$	n_i	m_i	N_i	$m_i x_i$	$m_i^2 n_i$
$7-9$					
$9-11$					
$11-12$					
$12-13$					
$13-14$					
$14-15$					
$15-17$					
$17-19$					

La media es $\overline{x} = 13,15$ años. La varianza la calculamos a partir de la columna de $m_i^2 n_i$ como sigue:

$$S^2 = 27742,25/157 - 13,15^2 = 3,78\, años^2$$

$$S = \sqrt{3,78} = 1,94\, años$$

El coeficiente de variación no posee unidades y es:

$$C_v = \frac{1,94}{13,15} = 0,15$$

$$15\,\% \;\; de\, variabilidad$$

En lo que concierne a la simetría podemos utilizar el coeficiente de asimetría de Yule-Bowley, para el cual es preciso el cálculo de los cuartiles:

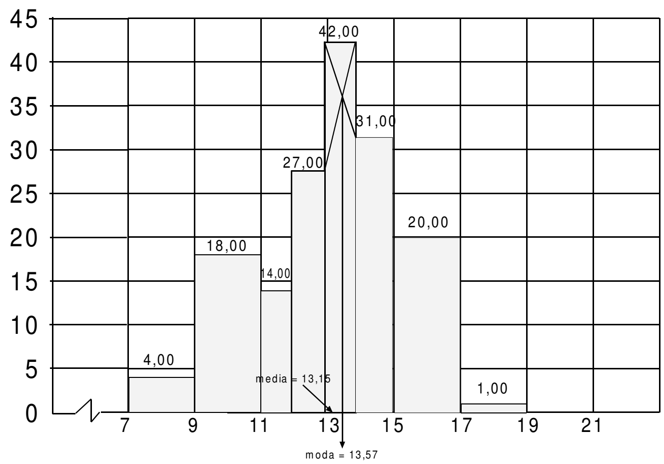


Figura 2.9: Ejemplo de asimetría

$$Q_1 = 12 + \frac{39,25 - 36}{27} \times 1 = 12,12$$
$$Q_2 = 13 + \frac{78,5 - 63}{42} \times 1 = 13,37$$
$$Q_3 = 14 + \frac{117,75 - 105}{31} \times 1 = 14,41$$

Lo que nos dice que aproximadamente en un rango de $Q_3 - Q_2 = 2,29$ años se encuentra el 50 % central del total de observaciones Además:

$$A_s = \frac{(14,41 - 13,37) - (13,37 - 12,12)}{14,41 - 12,12} = -0,09$$

Este resultado nos indica que existe una ligera asimetría a la izquierda (negativa). Un resultado similar se obtiene si observamos que la distribución de frecuencias es unimodal, siendo la moda:

$$\widehat{x} = 13 + \frac{42 - 27}{(42 - 27) - (42 - 31)} \times 1 = 13,57$$

en cuyo caso podemos usar como medida del sesgo:

$$A_s = \frac{\overline{x} - \widehat{x}}{S} = -0,21$$

2.4.4.1. Medidas de apuntamiento o curtosis

Se define el coeficiente de aplastamiento de Fisher como:

$$\gamma = \frac{m_4}{\sigma^4} - 3 \tag{2.13}$$

donde m_4 es el momento empírico de cuarto orden. Es éste un coeficiente adimensional, invariante ante cambios de escala y de origen. Sirve para medir si una distribución de frecuencias es muy apuntada o no. Para decir si la distribución es larga y estrecha, hay que tener un patrón de referencia. El patrón de referencia es la distribución normal o gaussiana para la que se tiene

$$\gamma = 0$$

De este modo, atendiendo a, γ se clasifican las distribuciones de frecuencias en

- Leptocúrtica:Cuando,

  $\gamma > 0$ o sea, si la distribución de frecuencias es más apuntada que la normal;

- Mesocúrtica:Cuando,

Apuntamiento$\gamma = 0$ es decir, cuando la distribución de frecuencias es tan apuntada como la normal;

- Platicúrtica:Cuando,

  $\gamma < 0$ o sea, si la distribución de frecuencias es menos apuntada que la normal;

  Estadísticos de tendencia central o posición nos indican donde se sitúa un grupo de puntuaciones. Los de variabilidad o dispersión nos indican si esas puntuaciones o valores están próximas entre sí o si por el contrario están o muy dispersas.

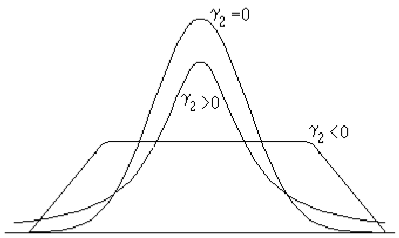


Figura 2.10: apuntamiento

2.5 Problemas

1. Calcule en los ejercicios de la tarea del capítulo 1 las siguientes medidas
   a) Tendencia central
      1) Posición
      2) Variabilidad
      3) Asimetría
      4) Apuntamiento
   b) Determine cuales medias representan mejor la muestra
   c) Interprete desde un punto de vista estadístico los resultados.
2. En un experimento de aprendizaje se registró el tiempo que los alumnos emplearon al resolver un problema de matemáticas x_i : en el primer intento, y_i : en el segundo intento.

	i	1	2	3	4	5	6
Individuo	x_i	23,7	36,9	25,5	30,2	28,0	34,8
	y_i	15,0	25,4	21,0	22,3	25,2	28,8

   a) Calcule la diferencia $d = x_i - y_i$ en cada caso y luego determine las medidas de
      1) Tendencia central
      2) Posición
      3) Variabilidad
      4) Asimetría
      5) Apuntamiento
   b) Determine lo mismo para x_i y y_i
   c) Compare
      1) $\bar{x} - \bar{y}$ con $\bar{d}$
      2) $S_x^2 - S_y^2$ con S_d^2
      3) Que puede concluir de la comparación
   d) Determine $\sum_{i=1}^{6} \frac{(x_i - c)^2}{N}$ donde c representa una medida de tendencia central diferente de la media aritmética
   e) Comparando el resultado del inciso b. con los resultados del inciso d. para que valor de c se obtuvo un valor mínimo
   f) Minimice la función

$$f(c) = \sum_{i=1}^{n} \frac{(x_i - c)^2}{N}$$

      ¿Está de acuerdo este resultado con el del inciso anterior.?
   g) Si

$$z = \frac{x_i - \bar{x}}{S}$$

      es. Calcule $\bar{z}$ y S_z^2 ¿Qué puede concluir?
   h) Realice los incisos del d. al g. para las variables y, d y z
   i) Realice histogramas para las variables: x, y, d y z. Compare los resultados con los histogramas

3. calcule la media y la varianza a partir de los datos tabulados en la siguiente tabla

Clases	n_i
$10 \leq x < 20$	121
$20 \leq x < 30$	165
$30 \leq x < 40$	184
$40 \leq x < 50$	173
$50 \leq x < 60$	142
$60 \leq x < 70$	120
$70 \leq x < 80$	118
$80 \leq x < 90$	110
$90 \leq x < 100$	90

4. Calcule las medidas de asimetría y de apuntamiento para los datos de la siguiente muestra

Clases	n_i
$600 \leq x < 650$	41
$650 \leq x < 700$	46
$700 \leq x < 750$	50
$750 \leq x < 800$	52
$800 \leq x < 850$	60
$850 \leq x < 900$	64
$900 \leq x < 950$	65
$950 \leq x < 1000$	66
$1000 \leq x < 1050$	70

5. Si el ingreso medio de 20 trabajadores es de 40000 dólares, ¿Cuál es su ingreso total?

6. Si la estatura media de una muestra de 25 jugadores de basketball es 6,9 pies, ¿cuál es la suma de la estatura de los 23 jugadores?

7. En un esfuerzo de reducir su consumo de café, un trabajador de oficina registra los números siguientes de tazas de café consumidas durante un periodo de 20 días

4	5	3	6	7	1	2	3	0	5
6	5	8	4	0	2	3	7	5	6

¿Qué medida de tendencia central le servirá mejor a su propósito? ¿Explique por qué? ¿ Cuál es el valor numérico?

8. En un investigación realizada por la secretaria de un médico para averiguar los tiempos de espera en minutos de los pacientes que acuden con el doctor, una muestra de pacientes de un día arrojó los resultados:

35	25	35	50	25	55	30	50	35	35
5	5	35	60	30	30	25	55	30	20
60	25	25	40	80	20	20	5	5	10

a) Describa un tiempo típico de espera usando la media.

b) Describa un tiempo típico de espera usando la mediana.

c) ¿Cuál medida, media, moda o mediana, se considera usted que es más representativa del conjunto de datos? Explique.

d) Determine los tres cuartiles

e) Determine los deciles

f) Realice un diagrama de caja

g) Determine los índices de asimetría y de apuntamiento

h) Realice una inferencia sustentada en los resultados.

9. Se escogió una muestra de 705 conductores de autobús y se registró en la tabla siguiente el número de accidentes de tránsito que tuvieron durante cuatro años.

Número de accidentes	Frecuencia
0	144
1	157
2	158
3	115
4	78
5	44
6	21
7	7
8	6
10	3
11	1

a) ¿Cuál es la moda?

b) Calcule la media

c) Determine la mediana, analítica y geométricamente

d) Determine el rango medio

e) Calcule el sesgo

f) ?Cuál medida de tendencia central usaría para determinar el valor central? Explique

g) Calcule los cuartiles

h) calcule los deciles

i) Realice un diagrama de caja

10. Si 20 puntajes tienen una media de 15 y 30 puntajes , una media de 30, ¿cuál es la media del grupo total de 50 puntajes?

11. La media armónica que a menudo se utiliza para promediar velocidades desarrolladas en distancias iguales. A continuación se presentan se darán las velocidades (en millas por hora) de un automóvil que viaja a velocidad constante cada 20 millas

30 60 40 60 30 20 25

¿Cuál es la rapidez promedio en el viaje de 140 millas?

12. Con los datos del ejercicio anterior determine:

a) La media aritmética

b) La media geométrica

c) La media cuadrática

d) Escoja una y explique por qué

13. Si la desviación estándar de un conjunto de datos es 0, ¿qué puede afirmarse de dicho conjunto?

14. ¿Será posible que el rango y la varianza sean iguales ? explique

15. ¿Puede ser la varianza negativa?

16. Suponga que una muestra tiene una media de 25 y una desviación estándar de 3.3 Aproxime un intervalo que contenga el 90 % de los datos

17. ¿Qué efecto tiene sobre la desviación el tamaño de la muestra?

II Probabilidad

CAPÍTULO

3 Teoría de probabilidades

Contenido Del Capítulo

3.1 Introducción

Si el único propósito del investigador es describir los resultados de un experimento concreto, los métodos descriptivos analizados en los capítulos anteriores pueden considerarse suficientes. No obstante, si lo que se pretende es utilizar la información obtenida para extraer conclusiones generales sobre todos aquellos objetos del tipo de los que han sido estudiados, entonces estos métodos constituyen sólo el principio del análisis, y debe recurrirse a métodos de inferencia estadística, los cuales implican el uso de una de las ramas de la matemática, llamada teoría de la probabilidad.

Analizaremos en este párrafo la noción de probabilidad y la terminología subyacente a esta área de las matemáticas, ya que la probabilidad constituye por sí misma un concepto básico que refleja su relación con la faceta del mundo exterior que pretende estudiar: los fenómenos aleatorios, los cuales obedecen unas ciertas reglas de comportamiento.

El concepto de probabilidad es importante cuando se estudian procesos físicos, biológicos, químicos, sociales que generan observaciones que no es fácil o factible predecir con exactitud, pero se puede determinar la frecuencia relativa con que ocurren con cierta precisión, si realizamos un gran número de observaciones.

Los eventos que poseen esta propiedad se denominan eventos aleatorios estocásticos y la frecuencia relativa con que ocurren es una interpretación intuitiva de la probabilidad, pero no nos proporciona una definición exacta de ella.

En su definición y en sus aplicaciones nos dedicaremos en este capítulo.

3.2 Experimentos y sucesos aleatorios

Definición 3.1

Diremos que un experimento es aleatorio si se verifican las siguientes condiciones,\

1. Se puede repetir indefinidamente, siempre con las mismas condiciones.
2. Antes de realizarlo no se puede predecir el resultado.
3. El resultado "e" que se obtiene pertenece a un conjunto de resultados posibles conocido previamente el cual llamaremos espacio muestral y lo denotaremos con la letra Ω *o* S. Los elementos del espacio muestral se denominan sucesos elementales.

Es decir si $e_1, e_2 \in S \Longrightarrow e_1, e_2$ son sucesos elementales o puntos muestrales. En otras palabras: Un suceso es elemental si su ocurrencia o no ocurrencia no está relacionada con ningún otro suceso.

Cualquier subconjunto A de S se denomina suceso o evento aleatorio.

El espacio muestral puede ser de dos tipos:

- Discreto si está formado por un conjunto finito o numerable de resultados.
- Continuo si está compuesto por un conjunto no numerable de elementos.

Definición 3.2

[Suceso determinista]Se denomina experimento determinista a el experimento que al realizarlo varias veces con las mismas condiciones iniciales obtenemos siempre el mismo resultado

Ejemplo: 3.2.1

Si dejamos caer dos cuerpos de diferentes masas y a la misma altura en el vacío, los dos cuerpos caerán al mismo tiempo.

Cuando en un experimento no se puede predecir el resultado final, decimos que el experimento es aleatorio

Ejemplo: 3.2.2

En una ruleta legal no se sabe que número va a salir

3.2.1. Tipos de eventos

1. Evento seguro es aquel que siempre se verifica después del experimento aleatorio, es decir S.
2. Evento imposible es aquel que nunca se verifica como resultado del experimento aleatorio y como un evento debe ser un subconjunto de S el suceso imposible es el conjunto vacío ϕ.
3. Evento complementario. Se denomina suceso complementario de un suceso A al que se verifica si no se verifica A es decir El complemento de A con respecto a S es el conjunto de todos los elementos de S que no están en A. Denotaremos el complemento del conjunto A como $A^{'}$

Ejemplo: 3.2.3

Si realizamos el experimento aleatorio de lanzar un dado al aire. Determinar el conjunto de eventos posibles

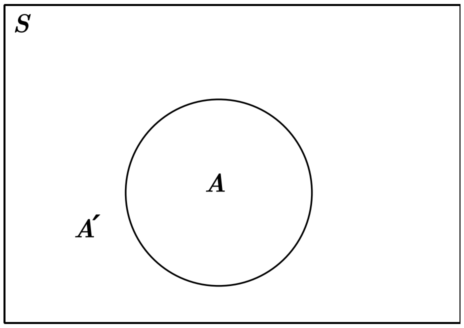


Figura 3.1: Evento complementario

Solución

En este caso tenemos el espacio muestral es
$S = \{1, 2, 3, 4, 5, 6\}$
y los sucesos elementales son
$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$

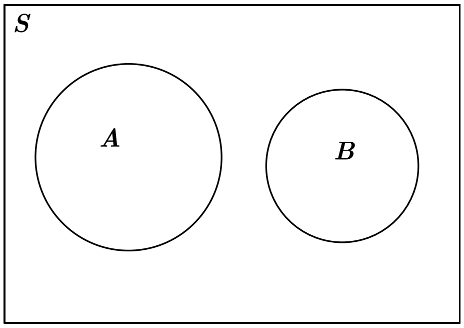
El conjunto de eventos posibles es $\mathcal{A} = \begin{cases} \phi \text{ suceso imposible} \\ S \text{ suceso seguro} \\ \{1, 2, 3\} \\ \{4, 5\} \\ \{2, 4, 6\} = \{1, 2, 3\}' \\ \vdots \end{cases}$

3.3 Operaciones básicas con eventos aleatorios

Como los eventos son en realidad conjuntos podemos aplicarles el álgebra de conjuntos tratada en su curso de álgebra elemental visto en I semestre.

1. Intersección. La intersección de dos eventos A y $B \subset S$ que se representa $A \cap B$, es el evento que contiene a todos los sucesos elementales comunes de A y B, es decir
$$A \cap B = \{e \in S : e \in A \land e \in B\}$$

2. Dos eventos son mutuamente excluyentes o disjuntos si
$A \cap B = \phi$, esto es si A y $B \subset S$ no tienen elementos comunes.



Eventos disjuntos

3. La unión. Dados dos sucesos aleatorios A y B, se denomina suceso unión de A y $B \subset S$ al evento formado por todos los sucesos elementales que pertenecen a A o bien a B, es decir
$$A \cup B = \{e \in S : e \in A \lor e \in B\}$$

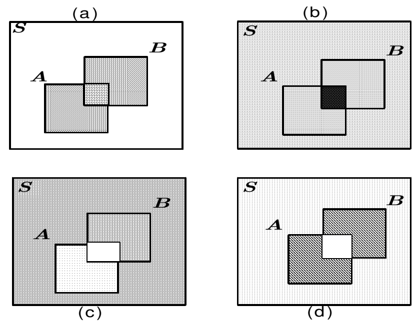
4. Dados dos sucesos aleatorios $A, B \subset S$, se llama diferencia de A y B, y se representa $A - B$ o $A\backslash B$, al suceso aleatorio formado por todos los sucesos elementales que pertenecen a A, pero no a B, es decir

$$A\backslash B = \{e \in S : e \in A \wedge e \notin B\} = A \cap B'$$

Donde a B' se le llama complemento de B y se define $S\backslash B$

5. Diferencia simétrica. Si $A, B \subset S$ se denomina evento diferencia simétrica de A y B y se denota por $A \triangle B$ al evento formado por todos los sucesos elementales que pertenecen a A y no a B, y los que están en B y no en A es decir,

$$A \triangle B = \{e \in S : e \in (A \cup B) \wedge e \notin (A \cap B)\}$$



Dados dos eventos $A, B \subset S$ se presenta en: (a) $A \cup B$; en (b) $A \cap B$; en (c) $A - B$; en (d) $A \triangle B$.

Hay algunas propiedades importantes que valen la pena recordar

- Leyes de Morgan

i $(A \cup B)' = A' \cap B'$

ii $(A \cap B)' = A' \cup B'$

- $A \triangle B = (A\backslash B) \cup (B\backslash A) = (A \cup B) \setminus (A \cap B)$
- $A \cap A' = \phi$
- $A \cap B' = A\backslash B$
- $A' = S\backslash A$
- $A \cup A' = S$

Definición 3.3

Sea $A_1, A_2, A_3, \cdots A_n$ una serie de eventos se tiene,
∘ El evento formado por todos los sucesos comunes y no comunes de la serie es,

$$A_1 \cup A_2 \cup A_3 \cdots \cup A_n = \cup_{i=1}^{n} A_i$$

∘ El evento formado por los sucesos comunes entre los eventos de la serie es el evento,

$$A_1 \cap A_2 \cap A_3 \cdots \cap A_n = \cap_{i=1}^{n} A_i$$

Definición 3.4

Sea A un conjunto finito o enumerable, Se le denomina $n(A)$, al cardinal de A, es decir el número de elementos que posee el conjunto A.

3.4 Probabilidad

3.4.1. Probabilidad estocástica

En los experimentos aleatorios observamos que si conservamos las mismas condiciones iniciales al aumentar el números de ensayos la frecuencia relativa con la que ocurre el suceso A es,

$$f_n(A) = \frac{número\, de\, veces\, que\, ocurre\, A}{n}$$

y tiende a converger hacia un valor, el cual se denomina probabilidad de A

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

Esta noción de probabilidad no se puede aplicar en la práctica ya que:

- Se requiere realizar un número infinito de veces el experimento para calcular la probabilidad
- Los experimentos aleatorios a veces no pueden ser realizados en la práctica por ejemplo. Si se quiere saber cuantas personas mueren al fumigar con un tóxico

3.4.2. Probabilidad de laplace

Si un experimento cualquiera se puede repetir obteniendo un número finito de resultados posibles y no existe ninguna razón para pensar que un resultado tiene privilegios sobre otro, se calcula la probabilidad del suceso A según la regla de Laplace

$$P(A) = \frac{número\, de\, casos\, favorables\, para\, A}{número\, de\, casos\, posibles} = \frac{n(A)}{n(S)}$$

Ejemplo: 3.4.4

Calcular la probabilidad de que al lanzar un dado se obtenga un número par

Solución

El espacio muestral es $S = \{1, 2, 3, 4, 5, 6\}$ llamaremos A al suceso que da como resultado un número par el lanzamiento del dado, $A = \{2, 4, 6\}$, si suponemos que el dado es legal, es decir ninguna cara tiene privilegio sobre otra para salir, entonces

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

3.4.3. Definición axiomática de la probabilidad

Como en toda rama de las matemáticas debemos establecer una serie de axiomas y definiciones básicas

Definición 3.5

Sea $\mathcal{A}$ una clase no vacía formada por subconjuntos de S, diremos que la clase $\mathcal{A}$ es un $\sigma - álgebra$ de eventos si los eventos complementarios de cada elemento de $\mathcal{A}$ está también en $\mathcal{A}$ y además las uniones numerables de eventos de $\mathcal{A}$ es elemento de $\mathcal{A}$ es decir.
Se dice que una clase $\mathcal{A}$ no vacía de subconjuntos de S es una $\sigma - álgebra$ si y sólo si cumple las siguientes propiedades

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

$$\forall A \in \mathcal{A} \Longrightarrow A' \in \mathcal{A}$$

$$\forall A_1, A_2, A_3, \dots \in \mathcal{A} \Longrightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$$

Definición 3.6

Definición axiomática de probabilidad
Dado un espacio muestral S, y una $\sigma - \acute{a}lgebra$ de sucesos $\mathcal{A}$ sobre él, diremos que P es una probabilidad sobre $\mathcal{A}$ si cumple las siguientes propiedades

A$_1$ La probabilidad es una función definida sobre $\mathcal{A}$, que toma solo valores positivos comprendidos entre 0 y 1, es decir

$$\begin{array}{rcl} P : \mathcal{A} & \rightarrow & [0,1] \subset IR \\ A \in \mathcal{A} & \mapsto & 1 \geq P(A) \geq 0 \end{array}$$

A$_2$ La probabilidad del suceso seguro es 1

$$P(S) = 1$$

A$_3$ Para cualquier sucesión infinita $A_1, A_2, A_3, \cdots$ de sucesos disjuntos de $\mathcal{A}$ se tiene que la probabilidad de el evento $\bigcup_{i=1}^{\infty} A_i$ es la serie infinita de las probabilidades, es decir

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Ejemplo: 3.4.5

Cuando el conjunto S formado por todos los posibles resultados de un experimento es finito o infinito enumerable podemos considerar que la clase $\mathcal{A}$ es el conjunto denominado **partes de** S

$$\mathbb{P}(S) = \{A | A \subseteq S\}$$

el cual es un $\sigma - \acute{a}lgebra$.

Ejemplo: 3.4.6

Consideremos el experimento de lanzamiento de un dado, entonces

$$S = \{1, 2, 3, 4, 5, 6\}$$
$$\mathbb{P}(S) = \{\phi, S, \{1\}, \{2\}, \cdots, \{6\}, \{1, 2\}, \{1, 3\}, ...\}$$

Es decir $\mathbb{P}(S)$ tiene $2^6 = 64$ elementos

Ejemplo: 3.4.7

Cuando S es infinito no enumerable por ejemplo queremos realizar el experimento de esperar el tiempo necesario para que el polietileno se desintegre naturalmente, observamos que $S = I\,R+$ y consideramos $\mathcal{A}$ como el conjunto de intervalos abiertos o cerrados, y sus uniones finitas, es decir

$$\mathcal{A} = \{\phi, I\,R^+, [2, 2], (2, 3), (4, 5] \cup [8, +\infty), ...\}$$

De acuerdo con el nivel de estas notas no trabajaremos con $\sigma - \acute{a}lgebras$ diferentes a las indicadas en los ejemplos 3.6 y 3.7[1], por lo que de aquí en adelante suponemos que estas son las clases $\mathcal{A}$

[1] El lector interesado en $\sigma - \acute{a}\lg ebras$ puede remitirse a Real Analysis probability de Robert B. Ash

Teorema 3.1 Probabilidad del vacío

$P(\Phi) = 0$

Prueba

consideremos la sucesión infinita de eventos $A_1, A_2, A_3, \cdots \in \mathcal{A}$ los cuales son todos vacíos y además disjuntos ya que $\phi \cap \phi = \phi$, de acuerdo con esto tenemos que $\bigcup_{i=1}^{\infty} A_i = \phi$ al aplicar A_3 tenemos

$$P(\phi) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

y el único real que cumple esta propiedad es el cero por lo que $P(\phi) = 0$ ■(q.e.d)

Teorema 3.2

ara cualquier sucesión finita de n eventos disjuntos $A_1, A_2, A_3, \cdots A_n \in \mathcal{A}$

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

Prueba

Consideremos la sucesión infinita de eventos $A_1, A_2, A_3, \cdots \in \mathcal{A}$ en los cuales hay $A_1, A_2, A_3, \cdots A_n$ eventos disjuntos y los $A_{n+1}, A_{n+2}, A_{n+3}, \cdots$ son vacíos. Entonces $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{n} A_i$. Por A_3 tenemos

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) \\ &= \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(A_i) \\ & \sum_{i=1}^{n} P(A_i) + 0 \\ & \sum_{i=1}^{n} P(A_i) \end{aligned}$$

■(q.e.d)

Teorema 3.3

ara cualquier suceso $A \in \mathcal{A}$ se tiene que $P(A') = 1 - P(A)$

Prueba

Como $A \in \mathcal{A} \Longrightarrow A' \in \mathcal{A}$ y además $A \cup A' = S$ y $P(S) = 1$ tenemos por el teorema 4 que

$$P\left(A \cup A'\right) = P(A) + P\left(A'\right) = P(S)$$

pero de acuerdo con A_2 $P(S) = 1$, entonces

$$\begin{aligned} P(A) + P\left(A'\right) &= 1 \\ P\left(A'\right) &= 1 - P(A) \end{aligned}$$

■(q.e.d)

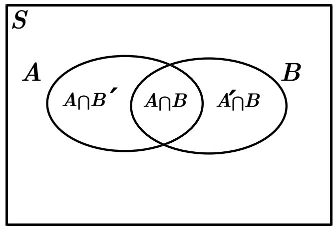


Figura 3.2: aaaa

Teorema 3.4

i $A, B \in \mathcal{A}$ y $A \subset B$, entonces $P(B) \geq P(A)$

Prueba

Como se ilustra en la fig 3.1 El evento B lo podemos escribir como $B = A \cup (B \cap A')$ entonces

$$\begin{aligned} P(B) &= P\left(A \cup \left(B \cap A'\right)\right) \\ &= P(A) + P\left(B \cap A'\right) \end{aligned}$$

y del A_1 $P(B \cap A') \geq 0$ de lo que se deduce $P(B) \geq P(A)$ ■(q.e.d)

Para dos sucesos cualesquiera $A, B \subset \mathcal{A}$ cualesquiera

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

Como se ilustra en la fig 3.2

$$A \cup B = \left(A \cap B'\right) \cup (A \cap B) \cup \left(A' \cap B\right)$$

como los eventos $(A' \cap B), (A \cap B')$ y $(A \cap B)$ son disjuntos podemos aplicar el teorema 3.1

$$\begin{aligned} &P(A \cup B) = \\ &\qquad P\left(A \cap B'\right) + P(A \cap B) + P\left(A' \cap B\right) \end{aligned}$$

pero en la gráfica se observa que

$$\begin{aligned} A &= (A \cap B') \cup (A \cap B) \\ &y \\ B &= (A \cap B) \cup (A' \cap B) \end{aligned}$$

al aplicar el teorema 5 queda

$$\begin{aligned} P(A) &= P\left(A \cap B'\right) + P(A \cap B) \\ P(B) &= P(A \cap B) + P\left(A' \cap B\right) \end{aligned}$$

al despejar y sustituir obtenemos lo que queremos

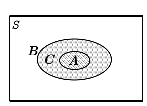


Figura 3.1 $B = A \cup (B \cap A')$ donde $C = (B \cap A')$

$B = A \cup (B \cap A')$ donde $C =$

Teorema 3.5

ara dos sucesos $A, B \subset \mathcal{A}$ cualesquiera

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

Como se ilustra en la fig 3.2

$$A \cup B = \left(A \cap B'\right) \cup (A \cap B) \cup \left(A' \cap B\right)$$

como los eventos $(A' \cap B), (A \cap B')$ y $(A \cap B)$ son disjuntos podemos aplicar el teorema 3.1

$$P(A \cup B) =$$
$$P\left(A \cap B'\right) + P(A \cap B) + P\left(A' \cap B\right)$$

pero en la gráfica se observa que

$$A = (A \cap B') \cup (A \cap B)$$
$$y$$
$$B = (A \cap B) \cup (A' \cap B)$$

al aplicar el teorema 5 queda

$$P(A) = P\left(A \cap B'\right) + P(A \cap B)$$
$$P(B) = P(A \cap B) + P\left(A' \cap B\right)$$

al despejar y sustituir obtenemos lo que queremos

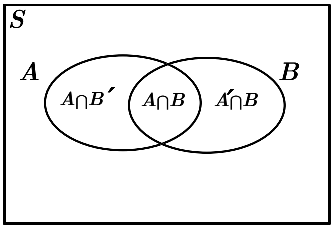


Figura 3.2 Partición de $A \cup B$

Teorema 3.6

ean $A_1, A_2, A_3, \cdots A_n \in \mathcal{A}$ una sucesión finita de eventos se tiene

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j=2}^{n} P(A_i \cap A_j) +$$
$$+ \sum_{i<j<r=3}^{n} P(A_i \cap A_j \cap A_r) +$$
$$+ \cdots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

Ejemplo: 3.4.8

Un estudiante de clase puede ser hombre o mujer. Si la probabilidad de que un hombre sea seleccionado es 0.3 ¿Cuál es la probabilidad de que sea seleccionada una mujer?

Solución

Sea A el evento de que se seleccione un hombre de la clase y sea B el evento de seleccionar una mujer, entonces es obvio que $S = A \cup B$, y además $A \cap B = \phi$, aplicando el teorema 7

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

y de los teorema 4 y A_2 obtenemos

$$\begin{aligned} 1 &= P(A) + P(B) \\ 1 &= 0,3 + P(B) \\ P(B) &= 0,7 \end{aligned}$$

Ejemplo: 3.4.9

Se selecciona una bola de una urna que contiene bolas rojas, blancas, azules, amarillas y verdes. Si la probabilidad de seleccionar una bola roja es $\frac{1}{5}$ y la de seleccionar una blanca es $\frac{2}{5}$ ¿Cuál es la probabilidad de seleccionar una bola azul, amarilla o blanca?

Solución

Sean $A_1, A_2, A_3, A_4, A_5, A_6$ los eventos de seleccionar de la urna una bola roja, blanca, azul, amarilla y verde respectivamente. Nos dicen que $P(A_1) = \frac{1}{5}$ y $P(A_2) = \frac{2}{5}$, entonces

$$P(A_3 \cup A_4 \cup A_5 \cup A_6) = P(B) = ?$$

Como los eventos son disjuntos

$$S = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6$$

y por A_2 y el th.3

$$\begin{aligned} 1 &= P(A_1) + P(A_2) + P(B) \\ P(B) &= 1 - \frac{1}{5} - \frac{2}{5} = \frac{2}{5} \end{aligned}$$

Ejemplo: 3.4.10

Si la probabilidad de que un estudiante A pierda un examen de estadística es de 0,5, la probabilidad de que un estudiante B pierda el mismo examen es de 0,2 y la probabilidad de que ambos pierdan el examen es de 0,1

1. *¿Cuál es la probabilidad de que al menos uno de estos estudiantes gane el examen?*
2. *¿Cuál es la probabilidad de de que ninguno de los dos pierda el examen?*

Solución

Sea A el evento que el estudiante A pierda el examen y sea B el evento de que el estudiante B pierda el examen de estadística

Nos *piden la probabilidad de que uno de los dos gane el examen pero no los dos, es decir si C es este evento*

$$P(C) = P\left(A'\right) + P\left(B'\right) - P\left(A' \cap B'\right)$$
$$P\left(A'\right) = 1 - P(A) = 1 - 0,5 = 0,5$$
$$P\left(B'\right) = 1 - P(B) = 1 - 0,2 = 0,8$$
$$A' \cap B' = (A \cup B)'$$
$$P\left(A' \cap B'\right) = 1 - P(A \cup B)$$

Pero

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,6$$
$$P\left(A' \cap B'\right) = 0,4$$

entonces

$$P(C) = 0,5 + 0,8 - 0,4 = 0,9$$

$P(A' \cap B') = 0,4$

3.5 Técnica para la enumeración de puntos muestrales

Cuando S o cualquiera de sus subconjuntos tiene muchos eventos elementales describirlo por extensión para determinar los casos favorables y casos posibles se hace engorroso por lo que en esta sección utilizaremos el análisis combinatorio para determinar los casos favorables y posibles de una manera más simple.

3.5.1. Diagrama de árbol

En experimentos simples es muy útil utilizar un método llamado diagrama de árbol el cual explicaré con un ejemplo.

Ejemplo: 3.5.11

Si se lanza una moneda legal tres veces ¿Cuál es la probabilidad de que en la primera y la ultima lanzada salga cara?.

Solución

En el diagrama observamos que para el primer lanzamiento hay dos posibilidades sello (S) y cara (C) en el segundo lanzamiento para cada posibilidad anterior hay dos nuevas posibilidades y en el tercer se da lo mismo por lo que al final hay 8 casos posibles y dos casos favorables de acuerdo con Laplace, entonces si A es el evento que sale cara en el primer y último lanzamiento

$$P(A) = \frac{2}{8} = \frac{1}{4}$$

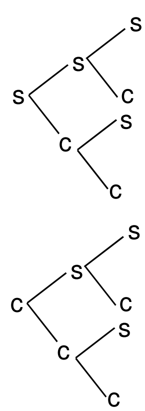


Diagrama de árbol

Teorema 3.7

onsidérese un experimento que tiene las dos características siguientes

- *El experimento se realiza en dos partes*
- *La primera parte del experimento tiene m resultados posibles $x_1, x_2, x_3, \cdots, x_m$ independientemente del resultado x_i obtenido la segunda parte del experimento tiene n resultados posibles $y_1, y_2, y_3, \cdots, y_n$.*
  Cada resultado del espacio muestral S del experimento será por tanto, un par de la forma (x_i, y_j) es decir

$$S = \{(x_i, y_j) \, | i = 1, 2, 3, \cdots m; j = 1, 2, 3 \cdots n\}$$
$$n(S) = mn$$

Este teorema puede generalizarse de la siguiente forma

Si un experimento puede realizarse con las siguientes características

- El experimento se realiza en k partes
- la primera parte puede realizarse con n_1 resultados posibles $x_{11}, x_{12}, x_{13}, \cdots, x_{1n_1}$ e independientemente del resultado x_{1i} se pueden realizar la segunda parte con n_2 resultados posibles $x_{21}, x_{22}, x_{23}, \cdots, x_{2n_2}$ y así sucesivamente cada parte del experimento puede realizarse de n_l formas, $l = 1, 2, 3, \cdots k$ obteniéndose un espacio muestral de la forma

$$S = \{(x_{1i}, x_{2j}, x_{3r}, \cdots, x_{km})\}$$
$$n(S) = n_1 \cdot n_2 \cdot n_3 \cdot \cdots \cdot n_k$$

Ahora enunciamos este teorema de una manera diferente.

Teorema 3.8 Fundamental

Si una acción puede efectuarse de una de p maneras diferentes, y si después de que esta acción ha sido efectuada de una de esas maneras, una segunda acción puede efectuarse de una de q maneras diferentes, entonces el número total de maneras diferentes en que las acciones pueden efectuarse siguiendo el orden mencionado es pq.

Corolario 1 S

una acción puede efectuarse de p maneras diferentes, y una segunda acción puede efectuarse de q maneras diferentes, y una tercera acción puede efectuarse de r maneras diferentes, y así sucesivamente, entonces el número total de maneras diferentes en que pueden realizarse todas estas acciones en el orden mencionado es pqr...

Corolario 2 S

r acciones pueden efectuarse sucesivamente de p maneras diferentes cada una, entonces el número total de maneras diferentes en que pueden efectuarse las r acciones sucesivamente es p^r.

3.6 Permutaciones

Definición 3.7

Una permutación es un arreglo de objetos distintos de tal manera que una permutación difiere de otra si el orden del arreglo o su contenido difieren

Conviene observar que el orden es una característica de especial importancia en una permutación. Cuando cambiamos el orden de los elementos de este arreglo, se dice que permutamos dichos elementos

3.6.1. Muestreo sin reemplazo

Consideremos un experimento en el cual se selecciona un objeto de n objetos distintos, y luego se selecciona un segundo objeto de los $n-1$ objetos restantes, y así sucesivamente hasta seleccionar el último objeto Este proceso se llama muestreo sin reemplazo de acuerdo con el teorema anterior los n objetos se pueden seleccionar de $n\,(n-1)\,(n-2)\cdots 3\cdot 2\cdot 1 = n!$ formas diferentes Ahora si no se escogen todos los objetos, si no k objetos, los k objetos se pueden seleccionar

$$\begin{aligned} p_{n,k} &= n\,(n-1)\,(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!} \quad r \leq n \\ 0! &= 1 \\ 1! &= 1 \\ p_{n,n} &= n\,(n-1)\,(n-2)\cdots 1 = n! \end{aligned}$$

Ejemplo: 3.6.12

Supongamos que se necesitan dos representantes del grupo 02 de estadística I-ad para asistir a un congreso

Solución

Como el grupo 02 de estadística tiene 40 alumnos y se necesitan dos, eso quiere decir que que escogen 2 de 40

$$p_{40,2} = \frac{40!}{(40-2)!} = 1560$$

Ejemplo: 3.6.13

Se necesita colocar 7 libros en un estante ¿De cuantas formas posibles se pueden colocar?

Solución

Como se escogerán 7 de 7 entonces

$$p_{7,7} = 7! = 5040$$

3.6.2. Muestreo con reemplazo

Si suponemos que tenemos una urna con n objetos numerados del 1 al n y se selecciona un objeto de la urna y se anota su número, y luego se coloca nuevamente en la urna, luego se selecciona otro objeto el cual pude ser el primero y así sucesivamente se pueden seleccionar tantos objetos como se quieran. Este proceso se denomina muestreo con reemplazo.

Si queremos realizar un total de k selecciones distintas, se nos presentan dos posibilidades

- Si $k > n$, es imposible ya que hay n números posibles y no están repetidos.
- Si $n \geq k$ Entonces existen n^k posibles formas de escoger los k objetos

Ejemplo: 3.6.14

Si se colocan al azar 12 bolas en 20 urnas, ¿Cuál es la probabilidad de que ninguna urna contenga más de una bola

Solución

Como hay 20 urnas y 12 bolas entonces podemos escoger 12 de 20 posibilidades de colocar las bolas, pero hay 20^{12}, posibles formas de escoger una urna con una bola

$$p = \frac{20!}{8! * 20^{12}} = \frac{235\,702\,467}{16\,000\,000\,000}$$

Coeficiente multinomial

El número de formas en las que se pueden asignar n objetos distintos de k grupos diferentes que contienen $n_1, n_2, n_3, \cdots, n_k$ objetos respectivamente es

$$N = \frac{n!}{n_1!n_2!n_3!\cdots n_k!} = \left(\begin{array}{c} n \\ n_1 n_2 n_3 \cdots n_k \end{array} \right)$$

$$\text{donde } n = \sum_{i=1}^{k} n_i$$

3.6.3. Combinación

Una combinación es un arreglo de objetos distintos donde una combinación difiere de otra si difiere el contenido del arreglo. Si nos interesa determinar el número de combinaciones cuando en n objetos distintos deben seleccionarse r a la vez entonces

$$C_{n,r} = \frac{p_{n,r}}{r!} = \left(\begin{array}{c} n \\ r \end{array} \right) \quad r \leq n$$

ya que el numerador es el número de permutaciones al escoger r objetos de n posibles, pero hay que descontar los casos en que el orden determina para la combinación el mismo elemento, que es exactamente $r!$.

3.6.3.1. Propiedades

- $\left(\begin{array}{c} n \\ 0 \end{array} \right) = \left(\begin{array}{c} n \\ n \end{array} \right) = 1$
- $\left(\begin{array}{c} n \\ k \end{array} \right) = \left(\begin{array}{c} n \\ n-k \end{array} \right)$
- $\left(\begin{array}{c} n \\ k \end{array} \right) + \left(\begin{array}{c} n \\ k-1 \end{array} \right) = \left(\begin{array}{c} n+1 \\ k \end{array} \right)$
- $C_{n,r} = C_{n-1,r-1} + C_{n-1,r}$
- El número de maneras en que mn objetos diferentes pueden dividirse en m grupos de n objetos cada uno, en donde el orden de los objetos en cada grupo no se toma en consideración, es
  $\frac{(mn)!}{(n!)^m}$, considerando el orden en que se forman los grupos,
  $\frac{(mn)!}{(n!)^m m!}$, sin considerar el orden en que se forman los grupos.

Ejemplo: 3.6.15

Calcular el número de maneras distintas en que 15 libros diferentes pueden dividirse en tres grupos de 9, 4 y 2 libros respectivamente.

Solución

En este caso

$$N_m = \frac{15!}{9!4!2!} = 75075.$$

Ejemplo: 3.6.16

Se tiene una baraja[a] de 52 cartas diferentes.

[a]Muchos problemas de probabilidad están relacionados con cartas, y aunque algunos estudiantes están relacionados con este juego, lo describiremos brevemente para que los problemas sean entendidos con claridad.

Una baraja ordinaria tiene 52 cartas divididas en cuatro grupos o palos con trece cartas cada uno. Los nombres de los palos y sus colores son los siguientes:

Bastos(negros), diamantes(rojos), corazones(rojos) y espadas(negras.

Cada palo consiste de 9 cartas numeradas del 2 al 10 inclusive y 4 cartas más llamadas as, rey, reina y sota (ordenadas en valor descendente). La expresión de que una carta se escoge o saca al azar, significa que la carta se toma de una baraja bien mezclada del modo que todas las cartas tengan igual oportunidad de ser escogida.

Encontrar

1. El número de maneras en que pueden repartirse las cuatro manos de 13 cartas a cuatro jugadores de bridge
2. El número de maneras en que las 52 cartas pueden dividirse en cuatro grupos de 13 cartas cada uno.

1. En el juego de bridge cada distribución diferente de las manos entre los jugadores constituye una división diferente. Por tanto, en este caso, los grupos aparecen permutados, y por lo tanto, el número de maneras diferentes es

$$\frac{52!}{(13!)^4} =$$
$$53\,644\,737\,765\,488\,792\,839\,237\,440\,000$$

2. En este caso, no importa el orden de los grupos, así el número de maneras es

$$\frac{52!}{(13!)^4\,4!} =$$
$$2235\,197\,406\,895\,366\,368\,301\,560\,000$$

Ejemplo: 3.6.17

Una moneda se tira 10 veces. Calcular la probabilidad de que aparezcan exactamente 7 caras

Solución

Ya que la moneda tiene dos formas diferentes de aparecer en cada tiro, en 10 sería 2^{10} formas. Y de las 10 caras vamos a seleccionar 7 caras, por lo que serían $C_{10,7}$ formas diferentes, por lo que la probabilidad buscada es

$$P = \frac{\binom{10}{7}}{2^{10}} = \frac{15}{128}$$

Ejemplo: 3.6.18

Si se sacan 3 cartas al azar de una baraja de 52 cartas, calcular la probabilidad de que sean as, rey y reina.

Solución

Se pueden seleccionar 3 cartas al azar entre 52, por lo que hay $C_{52,3}$ formas diferentes. Y como hay 4palos y en cada palo hay un as, un rey y una reina, entonces resulta que estas barajas pueden obtenerse de $4 \times 4 \times 4$ formas diferentes .Por lo que la probabilidad buscada es

$$\frac{4^3}{\binom{52}{3}} = \frac{16}{5525} = 2.8959 \times 10^{-3}$$

Ejemplo: 3.6.19

De una bolsa que que contiene 4 bolas blancas, 2 negras y 3 rojas, se sacan 5 al azar. Calcular la probabilidad de que 2 sean blancas 1 negra y 2 rojas.

Solución

Del total de $4 + 2 + 3 = 9$ bolas se pueden seleccionar 5 bolas en $C_{9,5}$ formas diferentes. Ahora entre las 4 bolas blancas 2 de ellas pueden seleccionarse $C_{4,2}$, entre las 2 blancas $C_{2,1}$ y entre las 3 rojas $C_{3,2}$ formas por lo que el total de casos favorables es $C_{4,2}C_{2,1}C_{3,2}$, así

$$P = \frac{\binom{4}{2}\binom{2}{1}\binom{3}{2}}{\binom{9}{5}} = \frac{2}{7}$$

3.7 Probabilidad condicionada e independencia de eventos

Definición 3.8

Sean $A, B \in \mathcal{A}$ y sea B un evento de probabilidad no nula para el evento A, llamamos probabilidad condicionada de A a B a la cantidad que representamos $P(A|B)$ y que definimos

$$P(A|B) = \frac{P(AB)}{P(B)},$$

la cantidad $P(A|B)$ se lee la probabilidad de A dada la ocurrencia de B

Ejemplo: 3.7.20

Se lanza al aire un dado ¿Cuál es la probabilidad de que salga el número 4? si sabemos que el resultado ha sido par

Solución

Sea $A = \{4\}$, entonces $P(A) = \frac{1}{6}$,
$B = \{2, 4, 6\}$, entonces $P(B) = \frac{3}{6} = \frac{1}{2}$
por tanto

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Definición 3.9

Sean $A, B \in \mathcal{A}$ dos eventos de probabilidad no nula se dice que son independientes si y solo si

$$P(AB) = P(A)P(B)$$

Ejemplo: 3.7.21

Para cierta población de empleados, los porcentajes de quienes aprueban un examen de aptitud para un trabajo, especificados según el sexo, se muestran en la tabla. Es decir todas las personas que presentan el examen el 24 % cae en la categoría de hombre aprobado, el 16 % en la categoría de de hombre reprobado, y así sucesivamente. Se selecciona al azar un empleado de esta población. Sea A el evento de que el empleado aprueba el examen y H el evento de que se selecciona un hombre ¿Son independientes los eventos A y H ?

	sexo		
Resultado	Mujer(M)	Hombre(H)	Total
Aprueba (A)	24	36	60
Reprueba(A')	16	24	40
Total	40	60	100

Solución

Determinemos

$$P(A|H) =$$
$$= \text{------}$$
$$= \text{------}$$
$$= \text{-------}$$

3.7.1. Reglas multiplicativas

Teorema 3.9

ea $A_1, A_2, A_3, \cdots A_n \in \mathcal{A}$ una sucesión de eventos aleatorios, entonces

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)\,P(A_2|A_1)\,P(A_3|A_1A_2)\cdots P\left(A_n|\bigcap_{i=1}^{n-1} A_i\right)$$

Definición 3.10

Se dice que una colección de eventos independientes $A_1, A_2, A_3, \cdots A_n \in \mathcal{A}$ es un evento exhaustivo y excluyente de sucesos, si se verifican las siguientes condiciones

$$\bigcup_{i=1}^{n} A_i = S$$
$$A_i \cap A_j = \phi \qquad \forall i \neq j$$

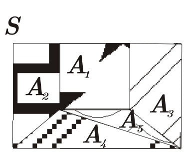


Figura 3.3: A_1, A_2, A_3, A_4 y A_5 forman un sistema exhaustivo y excluyente de sucesos

Teorema 3.10

ea $A_1, A_2, A_3, \cdots A_n \in \mathcal{A}$ un sistema exhaustivo y excluyente de sucesos. Entonces

$$\forall B \subset \mathcal{A}, \Longrightarrow P(B) = \sum_{i=1}^{n} P(B|A_i) P(A_i)$$

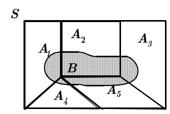


Figura 3.4: Si A_1, A_2, A_3, A_4 forman un sistema exhaustivo y excluyente de sucesos , podemos calcular la probabilidad de B a partir de las cantidades $P(B \cap A_i)$

Prueba

Como $B \subset S$ podemos decir

$$\begin{aligned} P(B) &= P\left(B \cap \left(\bigcup_{i=1}^{n} A_i\right)\right) \\ &= P\left(\bigcup_{i=1}^{n} (B \cap A_i)\right) \\ &= \sum_{i=1}^{n} P(B \cap A_i) \\ &= \sum_{i=1}^{n} P(B|A_i) P(A_i) \end{aligned}$$

Ejemplo: 3.7.22

Se tienen dos urnas, y cada una de ellas contiene un número diferente de bolas blancas y rojas:

■(q.e.d)

- *primera urna U_1: 3 bolas blancas y 2 rojas*
- *Segunda urna U_2 4 bolas blancas y 2 rojas*

Se realiza el siguiente experimento :

Se tira una moneda al aire y si sale cara se elige una bola de la primera urna y si sale sello se escoge una bola de la segunda urna.

¿Cuál es la probabilidad de que salga una bola blanca?

Solución

La situación que tenemos la podemos esquematizar de la siguiente manera

Si B: es el evento de sacar una bola blanca y R : el evento de sacar una bola roja, entonces

$$\begin{aligned} P(U_1) &= \frac{1}{2} \\ P(B|U_1) &= \frac{3}{5} \\ P(U_2) &= \frac{1}{2} \\ P(B|U_2) &= \frac{4}{6} \end{aligned}$$

Como U_1 y U_2 forman un sistema exhaustivo y excluyente de sucesos aplicando el teorema de la probabilidad total podemos afirmar que,

$$\begin{aligned} P(B) &= P(B|U_1)\,P(U_1) + P(B|U_2)\,P(U_2) \\ &= \frac{3}{5}\cdot\frac{1}{2} + \frac{4}{6}\cdot\frac{1}{2} = \frac{19}{30} \end{aligned}$$

Teorema 3.11

Bayes) Sea $A_1, A_2, A_3, \cdots A_n \in \mathcal{A}$ un sistema exhaustivo y excluyente de eventos. Sea $B \subset \mathcal{A}$ un suceso del que conocemos todas las cantidades

$$P(B|A_i)\,, i = 1, 2, 3, \cdots, n$$

a las que denominamos verosimilitudes, entones se verifica

$$\forall j = 1, 2, 3, \cdots, n, \qquad P(A_j|B) = \frac{P(B|A_j)\,P(A_j)}{\sum_{i=1}^{n} P(B|A_i)\,P(A_i)}$$

Prueba

Es consecuencia de la definición de probabilidad condicionada en términos de la intersección, y del teorema de la probabilidad total

$$\begin{aligned} P(A_j|B) &= \frac{P(A_j \cap B)}{P(B)} \\ &= \frac{P(B|A_j)\,P(A_j)}{\sum_{i=1}^{n} P(B|A_i)\,P(A_i)} \end{aligned}$$

■(q.e.d)

Ejemplo: 3.7.23

Se tienen tres urnas. Cada una de ellas contiene un número diferente de bolas blancas y rojas. La primera urna tiene 3 bolas blancas y 2 rojas, la segunda 4 bolas blancas y 2 rojas y la tercera 3 bolas rojas.
Se realiza el siguiente experimento:
Alguién elije al azar y con la misma probabilidad una de las tres urnas, saca una bola.
Si el resultado del experimento ha sido sacar una bola blanca ¿Cuál es la probabilidad de que provenga de la primera urna? Calcular lo mismo para las otras dos urnas.

Solución

Si B: es el evento de sacar una bola blanca y R: el evento de sacar una bola roja, entonces como
U_1 : 3 bolas blancas y 2 rojas
U_2 : 4 bolas blancas y 2 rojas
U_3 : 3 bolas rojas

tenemos

$$
\begin{aligned}
P\left(U_1\right) &= \frac{1}{3}\\
P\left(U_2\right) &= \frac{1}{3}\\
P\left(U_3\right) &= \frac{1}{3}\\
P\left(B|U_1\right) &= \frac{3}{5}\\
P\left(B|U_2\right) &= \frac{4}{6}\\
P\left(B|U_3\right) &= 0
\end{aligned}
$$

En este caso U_1, U_2, U_3 forman un sistema incompatible y excluyente de eventos, por lo que es posible aplicar el teorema de Bayes

$$
\begin{aligned}
P\left(U_1|B\right) &= \frac{P\left(B|U_1\right)P\left(U_1\right)}{P\left(B|U_1\right)P\left(U_1\right)+P\left(B|U_2\right)P\left(U_2\right)+P\left(B|U_3\right)P\left(U_3\right)}\\
&= \frac{\left(\frac{3}{5}\right)\left(\frac{1}{3}\right)}{\left(\frac{3}{5}\right)\left(\frac{1}{3}\right)+\left(\frac{4}{6}\right)\left(\frac{1}{3}\right)+0\left(\frac{1}{3}\right)}\\
&= \frac{9}{19}
\end{aligned}
$$

Para los otros dos casos resulta de manera equivalente

$$
\begin{aligned}
P\left(U_2|B\right) &= \frac{P\left(B|U_2\right)P\left(U_2\right)}{P\left(B|U_1\right)P\left(U_1\right)+P\left(B|U_2\right)P\left(U_2\right)+P\left(B|U_3\right)P\left(U_3\right)}\\
&= \frac{10}{19}
\end{aligned}
$$

$$
\begin{aligned}
P\left(U_3|B\right) &= \frac{P\left(B|U_3\right)P\left(U_3\right)}{P\left(B|U_1\right)P\left(U_1\right)+P\left(B|U_2\right)P\left(U_2\right)+P\left(B|U_3\right)P\left(U_3\right)}\\
&= 0
\end{aligned}
$$

Ejemplo: 3.7.24

En la segunda guerra mundial, uno de los primeros intentos de investigación de operaciones en la Gran Bretaña se orientaba en establecer patrones de búsqueda de submarinos desde vuelos de escuadrones o mediante un sólo avión. Por algún tiempo, la tendencia fue concentrar los vuelos en la costa, pues se pensaba que le mayor número de avistamientos que ocurrían ahí. El grupo de investigación registró 1000 vuelos de un solo avión con los siguientes resultados(Los datos no son reales)

	En la playa	Fuera de la costa	Total
Observación	80	20	100
No observación	820	80	900
Total de salidas	900	100	1000

sea

S_1 Hubo un avistamiento
S_2 No hubo avistamiento
B_1 Salida a la costa
B_2 Salida en altamar

Vemos de inmediato que

$$
\begin{aligned}
P\left(S_1|B_1\right) &= \frac{80}{900} = \frac{4}{45} = 8.8889\times 10^{-2}\\
P\left(S_2|B_2\right) &= \frac{20}{100} = 0.2
\end{aligned}
$$

lo cual indica una estrategia de búsqueda es contraria a la primera práctica.

Ejemplo: 3.7.25

Supóngase que se va a seleccionar una muestra aleatoria de tamaño 2 de un lotes de tamaño 100, y que se sabe que 98 de los 100 artículos se encuentran en buen estado. La muestra se toma de manera tal que el primer artículo se observa y se regresa antes de seleccionar el segundo artículo

Solución

. Si aceptamos que
A : El primer artículo observado está en buen estado
B : El segundo artículo observado está en buen estado
Y si deseamos determinar la probabilidad de que ambos artículos estén en buen estado, entonces

$$P(A\cap B) = P(A)\,P(B) = \left(\frac{98}{100}\right)\left(\frac{98}{100}\right) = 0.960\,4$$

Si se selecciona el artículo sin reemplazo de modo que que el primer artículo no se regresa antes de seleccionar el segundo, entonces

$$P(A\cap B) = P(A)\,P(B|A) = \frac{98}{100}\cdot\frac{97}{99} = 0{,}9602$$

Los resultados son muy parecidos por lo que generalmente suponemos que los eventos son independientes cuando la fracción de muestreo (tamaño de la muestra/tamaño de la población) es menor que 0.1

Ejemplo: 3.7.26

Tres industrias suministran microprocesadores a un fabricante de equipos de telemetría. Todos se elaboran supuestamente con las mismas especificaciones. No obstante, el fabricante ha probado durante varios años los microprocesadores, y los registros indican la siguiente información

Industria	Fracción de defectos	Fracción suministrada por
1	0.02	0.15
2	0.01	0.80
3	0.03	0.05

EL fabricante ha interrumpido las pruebas por causa de los costos involucrados, y puede ser razonable suponer que la proporción defectuosa y la mezcla de inventarios son las mismas que durante el periodo en el que se efectuaron los registros. El director de manufactura selecciona un microprocesador al azar, lo lleva al departamento de pruebas y descubre que está defectuoso. Determinar la probabilidad que el artículo proviene de la industria 3
Sea A el evento de que el artículo es defectuoso y B_i $i = 1, 2, 3$ es el evento que el artículo proviene de la empresa i

$$\begin{aligned} P(B_3|A) &= \frac{P(A|B_2)\,P(B_2)}{P(A|B_1)\,P(B_1) + P(A|B_2)\,P(B_2) + P(A|B_3)\,P(B_3)} \\ &= \frac{(0{,}05)\,(0{,}03)}{(0{,}15)\,(0{,}02) + (0{,}8)\,(0{,}01) + (0{,}05)\,(0{,}03)} = 0.\,12 \end{aligned}$$

Ejemplo: 3.7.27

Una caja contiene 8 bolas rojas, 3 bolas blancas y nueve azules. Si se sacan 3 bolas al azar, determinar la probabilidad de que

a. *Las 3 sea rojas*

b. *Las tres sean blancas*

c. *2 sean rojas*

d. *Al menos una blanca*

e. *Sea una de cada color*

f. *Salgan en el orden roja,blanca y azul*

Solución

Denotaremos por R_i $i = 1, 2, 3$ *los eventos que la primera, la segunda y la tercera bola sean rojas respectivamente. Denotaremos por* B_i $i = 1, 2, 3$ los eventos que la primera, la segunda y la tercera bola sean blancas respectivamente, y *denotaremos por* A_i $i = 1, 2, 3$ *los eventos que la primera, la segunda y la tercera bola sean rojas respectivamente, entonces*

a.

$$P\left(\bigcap_{i=1}^{3} R_i\right) = \frac{\binom{8}{3}}{\binom{20}{3}} = \frac{14}{285}$$

b.

$$P\left(\bigcap_{i=1}^{3} B_i\right) = \frac{\binom{3}{3}}{\binom{20}{3}} = \frac{1}{1140}$$

c.

$$P\left(\bigcap_{i=1}^{2} R_i \cap B\right) = \frac{\binom{8}{2}}{\binom{20}{3}} = \frac{}{}$$

d.

$$P\,(\text{ninguna es blanca}) = \frac{\binom{17}{3}}{\binom{20}{3}} = \frac{}{}$$

e.

$$P\,(\text{sacar una de cada color}) = \frac{\binom{8}{1}\binom{3}{1}\binom{9}{1}}{\binom{20}{3}} = \frac{}{}$$

f.

$$\begin{aligned} &P\,(\text{bolas en el orden R, B,A}) \\ &= \frac{1}{3!}P\,(\text{una de cada colore}) = \frac{3}{95} \end{aligned}$$

Ejemplo: 3.7.28

De una baraja de 52 naipes bien mezclada se sacan 5 naipes. Hallar la probabilidad de que 3 sean de un palo y 2 de otro

Solución

$$\begin{aligned} P\,(3 \text{ de cualquier figura y } 2 \text{ de otra}) &= \\ &= \frac{4\binom{13}{3} \cdot 3\binom{13}{2}}{\binom{52}{5}} \\ &= 0.103 \end{aligned}$$

Ejemplo: 3.7.29

Supóngase que se lanzan 12 dados. Se determina la probabilidad p de que cada uno de los seis números distintos aparezca dos veces

Solución

tenemos que S es el conjunto formado por la sucesión de 12-tupla donde la i-ésimo valor de la sucesión es el resultado i-ésimo lanzamiento, de lo que se deduce que $n(S) = 6^{12}$ donde todos los 6^{12} resultados posibles son igualmente probables, ahora podemos considerar los seis valores posibles que se puede dar en cada dado como seis casillas las cuales se le pueden asignar solo dos posibilidades. Por lo que al utilizar el coeficiente multinomial para determinar los casos favorables tenemos

$$\begin{aligned} n &= 12 \\ k &= 6 \\ n_1 &= n_2 = \cdots = n_6 = 2 \\ N &= \frac{12!}{(2!)(2!)\cdots(2!)} = \frac{12!}{(2!)^6} \end{aligned}$$

Por lo que la probabilidad buscada es
$P = \frac{12!}{2^6 6^{12}} = \frac{1925}{559\,872} = 3.\,438\,3 \times 10^{-3}$

Ejemplo: 3.7.30

Una baraja de 52 cartas contiene 13 corazones. Supóngase que se barajan las cartas y se distribuyen entre cuatro jugadores A,B,C y D de tal manera que a cada jugador le correspondan 13 cartas, Se determinará la probabilidad P de que cada jugador reciba 6, 4, 2 y 1 corazones respectivamente

Solución

El número de combinaciones distintas posibles de las 13 posiciones ocupadas en la baraja por los corazones es $\binom{52}{13}$. Si el jugador A recibe 6 corazones, hay $\binom{13}{6}$ combinaciones posibles de las 6 posiciones que ocupan estos corazones entre las 13 cartas que recibe el jugador A. De la misma manera podemos establecer las combinaciones posibles de cada jugador entre las 13 cartas que recibe cada uno como $\binom{13}{4}$, $\binom{13}{2}$ y $\binom{13}{1}$ respectivamente para B,C y D. Por tanto como los eventos A,B,C y D son independientes

$$P(ABCD) = \frac{\binom{13}{6}\binom{13}{4}\binom{13}{2}\binom{13}{1}}{\binom{52}{13}} = 1.\,9592 \times 10^{-3}$$

Ejemplo: 3.7.31

Supóngase que una moneda equilibrada va a ser lanzada diez veces y se desea determinar

a. *la probabilidad P de obtener exactamente 3 caras*

b. *La probabilidad P de obtener a lo sumo tres caras*

Solución

a. El número total de posibles combinaciones distintas de 10 caras y sellos es 2^{10} y se puede suponer que todas estas combinaciones son igualmente probables y el número de casos favorables es $\binom{10}{3}$ por lo que

$$p = \frac{\binom{10}{3}}{2^{10}} = 0.117\,19$$

Solución

b. Puesto que lo que me piden es la union de los eventos A_i : obtener 0,1,2 y 3 caras $i = 0, 1, 2, 3,$ y como estos eventos son disjuntos tenemos que

$$P\left(\bigcup_{i=0}^{3} A_i\right) = \frac{\binom{10}{0} + \binom{10}{1} + \binom{10}{2} + \binom{10}{3}}{2^{10}} = 0.171\,88$$

Ejemplo: 3.7.32

Supóngase que en una clase hay 15 hombres y 30 mujeres, y que se van a seleccionar al azar 10 estudiantes para una tarea especial. se determinará la probabilidad P de seleccionar exactamente 3 hombres

Solución

el número de combinaciones distintas que se pueden formar con 45 estudiantes para obtener una muestra de 10 es $\binom{45}{10}$, y como el muestreo es al azar y sin reemplazo entonces todas estas combinaciones son igualmente posibles por lo que hay que determinar el números de combinaciones distintas que se pueden seleccionar con 3 hombres y 7 mujeres, entonces con los hombres se pueden realizar $\binom{15}{3}$ y con las mujeres $\binom{30}{7}$ por lo que

$$P = \frac{\binom{15}{3}\binom{30}{7}}{\binom{45}{10}} = \frac{3958\,500}{13\,633\,279} = 0.290\,36$$

Ejemplo: 3.7.33

Supóngase que se baraja un naipe de 52 cartas que contiene 4 ases y que las cartas se reparten entre cuatro jugadores, de forma que cada una reciba 13 cartas. se determinará la probabilidad de que cada jugador reciba un as.

Solución

El número de combinaciones diferentes es $\binom{52}{4}$. Y podemos suponer que todas estas combinaciones son igualmente probables. si cada jugador recibe un as entonces debe haber un as entre las 13 cartas que recibe por ejemplo el primer jugador y un as recibirá cada uno de los jugadores de las 13 cartas que recibirán, es decir hay 13 posiciones posibles para el as que recibe cada jugador por lo que los casos favorables será 13^4, entonces

$$P = \frac{13^4}{\binom{52}{4}} = \frac{2197}{20\,825} = 0.105\,5$$

Ejemplo: 3.7.34

Supóngase que una máquina produce un artículo defectuoso con probabilidad $p = 0{,}4$ y produce un artículo no defectuoso con una probabilidad $q = 0, 6$. Supóngase además que se seleccionan aleatoriamente para su control seis de los artículos producidos por la máquina y que los resultados de control son independientes para estos seis artículos. Se determinará la probabilidad de que exactamente dos de los seis artículos sean defectuosos

Solución

Si consideramos los eventos D_i el evento del que el $i-ésimo$ artículo sea defectuoso y N_j el evento de que el $i-ésimo$ artículo sea no defectuoso para, $i = 1, 2, 3, 4, 5, 6$, como los eventos son independientes entonces por ejemplo

$$P(N_1 N_3 D_2 D_4 D_5 D_6) = qqpppp$$

como esta es una de las posibles combinaciones hay que determinar las otras que no es más que $\binom{6}{2}$ combinaciones distintas y por el th 3.13 tenemos

$$P = \binom{6}{2}(0{,}4)^2(0{,}6)^4 = 0.311\,04$$

Ejemplo: 3.7.35

considérese una máquina que produce un artículo defectuoso con una probabilidad $p = 0{,}4$ y uno no defectuoso con probabilidad $q = 0{,}6$. Supóngase que el control se realiza seleccionando artículos al azar y de uno en uno hasta obtener exactamente cinco artículos defectuosos. Se determinará la probabilidad P de que deban ser seleccionados 20 artículos para obtener 5 defectuosos

Solución

El quinto artículo defectuoso será el $n-ésimo$ controlado si, y sólo si, hay exactamente cuatro defectuosos entre los primeros $n-1$ artículos y el $n-ésimo$ es defectuoso entonces por el th 3.13 tenemos que

$$P = \binom{n-1}{4}p^5 q^{n-5} = \binom{19}{4}(0{,}4)^5(0{,}6)^{15}$$
$$= 1.866\,2 \times 10^{-2}$$

Ejemplo: 3.7.36

Supóngase que se van a extraer dos bolas aleatoriamente y sin reemplazo, de una urna que contiene 8 bolas rojas y 10 bolas azules. Se determinará la probabilidad P de obtener la primera bola roja y la segunda azul

Solución

Sea A el evento de que la primera bola sea roja y B de que la segunda bola sea azul, entonces $P(A) = \frac{8}{18} = 0.444\,44$, además si el suceso A ha ocurrido, entonces se ha obtenido una bola roja de la urna en la primera extracción por lo que la probabilidad de obtener una bola azul en la segunda extracción es

$$P(B|A) = \frac{10}{17} = 0.588\,24$$

entonces

$$P(AB) = (0.444\,44)(0.588\,24) = 0.261\,44$$

Ejemplo: 3.7.37

Para la fabricación de un gran lote de artículos similares se utilizaron tres máquinas M_1, M_2 y M_3. Supóngase que el 20 % de los artículos fueron fabricados por la máquina M_1, el 30 % por la máquina M_2 y el 50 % por la máquina M_3 Suponemos además que el 1 % de los artículos fabricados por la máquina M_1 son defectuosos y así respectivamente 2 % y 3 % de los artículos fabricados por la máquinas M_2 y M_3. Se quiere seleccionar al azar uno de los artículos del lote que resultan defectuosos. Determine la probabilidad de que el artículo haya sido fabricado por la máquina M_2.

Solución

Sea A_i el evento de que el artículo haya sido fabricado por la máquina $M_i, i = 1, 2, 3$, y sea B el suceso de que el artículo seleccionado sea defectuoso, de acuerdo con esto hay que determinar $P(A_2|B)$, entonces

$$\begin{aligned}
P(A_1) &= 0{,}2 \\
P(A_2) &= 0{,}3 \\
P(A_3) &= 0{,}5 \\
P(B|A_1) &= 0{,}01 \\
P(B|A_2) &= 0{,}02 \\
P(B|A_3) &= 0{,}03
\end{aligned}$$

aplicando el teorema de Bayes

$$P(A_2|B) = \frac{P(A_2)\,P(B|A_2)}{\sum_{i=1}^{3} P(A_i)\,P(B|A_i)} = 0{,}26$$

3.8 Problemas

1. Tres clases diferentes tienen 20, 18 y 25 estudiantes, respectivamente, y cada estudiante pertenece a una sola clase. Si se forma un equipo con un estudiante de cada una de estas tres clases, ¿de cuántas maneras distintas se pueden seleccionar los miembros del equipo?

2. ¿De cuántas maneras distintas se pueden ordenar cinco letras a, b, c, d y e?

3. Si un hombre tiene seis camisas distintas y cuatro pares distintos de pantalones, ¿de cuántas formas distintas se puede vestir combinando esas prendas?

4. Si se lanzan cuatro dados, ¿cuál es la probabilidad de que los cuatro números que aparecen sean distintos?

5. Si se lanzan seis dados, ¿cuál es la probabilidad de que cada uno de los seis números posibles aparezcan exactamente una vez?

6. Si se colocan al azar 12 bolas en 20 urnas, ¿cuál es la probabilidad de que ninguna urna contenga más de una bola?

7. El ascensor de un edificio empieza a subir con cinco personas y para en siete pisos. Si la probabilidad de que cualquier pasajero salga del ascensor en un piso concreto es igual para todos los pisos y los pasajeros salen independientemente unos de otros, ¿cuál es la probabilidad de que no haya dos pasajeros que salgan en el mismo piso?

8. Si k personas se sientan aleatoriamente en una fila de n asientos ($n > k$), ¿cuál es la probabilidad de que ocupen k asientos contiguos en la fila?

9. Sik personas se sientan aleatoriamente en n sillas dispuestas en círculo ($n > k$), ¿cuál es la probabilidad de que ocupen k sillas contiguas del círculo?

10. Si n personas se sientan aleatoriamente en una fila de $2n$ asientos, ¿cuál es la probabilidad de que no haya dos personas sentadas en asientos contiguos?

11. Una caja contiene 24 bombillas, de las cuales 2 son defectuosas. Sí una persona selecciona 10 bombillas al azar, sin reemplazo, ¿cuál es la probabilidad de seleccionar las 2 bombillas defectuosas?

12. Supóngase que se ha de seleccionar un comité de 12 personas aleatoriamente escogidas entre un grupo de 100. Determínese la probabilidad de que dos personas concretas, A y B, sean seleccionadas.

13. Supóngase que 35 personas se dividen aleatoriamente en dos equipos de manera que uno de los equipos consta de 10 personas y el otro de 25. ¿Cuál es la probabilidad de que dos personas concretas, A y B, estén en el mismo equipo?

14. Una caja contiene 24 bombillas, de las cuales 4 son defectuosas. Si una persona selecciona aleatoriamente 10 bombillas de la caja, y una segunda persona toma entonces las 14 bombillas restantes, ¿cuál es la probabilidad de que la misma persona seleccione las 4 bombillas defectuosas?

15. Demuéstrese que, para cualquier entero positivo n y k ($n > k$),
$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$$

16. Demuéstrese que
$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = 2^n$$

17. El Senado de Estados Unidos está constituido por dos senadores de cada uno de los 50 estados,

    a) Si se selecciona aleatoriamente un comité de 8 senadores, ¿cuál es la probabilidad de que contenga al menos uno de los dos senadores de un estado concreto?

    b) ¿Cuál es la probabilidad de que un grupo de 50 senadores seleccionados aleatoriamente contenga un senador de cada estado?

18. Supóngase que$A$ es un suceso tal que $P(A) = 0$ y que B es cualquier otro suceso. Demuéstrese que A yB son sucesos independientes.

19. Supóngase que una persona lanza tres veces dos dados equilibrados. Determínese la probabilidad de que en cada uno de los tres lanzamientos la suma de los dos números que aparecen sea 7.

20. Supóngase que la probabilidad de que el sistema de control utilizado en una nave espacial no funcione en un vuelo concreto es 0.001. Supóngase además que la nave también tiene instalado un segundo sistema de control idéntico, pero completamente independiente del primero, que toma el control cuando el primero falla. Determínese la probabilidad de que en un vuelo concreto la nave espacial esté bajo control, ya sea del sistema original o del sistema duplicado.

21. Supóngase que una lotería consta de 10 000 boletos y que otra lotería consta de 5000. Si una persona compra 100 boletos de cada lotería, ¿cuál es la probabilidad de que gane al menos un primer premio?

22. Dos estudiantes A y B están inscritos en un curso. Si el estudiante A asiste a las clases el 80 % de las veces y el estudiante B el 60 %, y si las ausencias de los dos estudiantes son independientes, ¿cuál es la probabilidad de que al menos uno de los dos estudiantes esté en clase un día concreto?

23. Si se lanzan tres dados equilibrados, ¿cuál es la probabilidad de que los tres números que aparecen sean iguales?

24. Considérese un experimento en el cual se lanza una moneda equilibrada hasta que aparece una cara por primera vez. Si este experimento se repite tres veces, ¿cuál es la probabilidad de que se necesite exactamente el mismo número de lanzamientos para cada una de las tres repeticiones?

25. Supóngase que A, B y C son tres sucesos independientes tales que $P(A) = \frac{1}{4}$, $P(B) = \frac{1}{3}$, y $P(C) = \frac{1}{2}$.

    a) Determínese la probabilidad de que ninguno de estos tres sucesos ocurra

    b) Determínese la probabilidad de que ocurra exactamente uno de estos tres sucesos.

26. Supóngase que la probabilidad de que una partícula emitida por un material radiactivo penetre en cierto campo es 0.01. Si se emiten diez partículas,

    a) ¿cuál es la probabilidad de que exactamente una de ellas penetre en el campo?

    b) Si se emiten diez partículas, ¿cuál es la probabilidad de que al menos una de ellas penetre en el campo?

    c) ¿Cuántas partículas tienen que ser emitidas para que la probabilidad de que al menos una partícula penetre en el campo sea al menos 0.8?

27. En la Serie Mundial de Béisbol, dos equipos A y B juegan una serie de partidos uno contra otro y el primer equipo que gana un total de cuatro partidos es el ganador de la Serie Mundial. Si la probabilidad de que el equipo A gane un partido contra el equipo B es $\frac{1}{3}$, ¿cuál es la probabilidad de que el equipo A gane la Serie Mundial?

28. 15. El Senado de Estados Unidos está constituido por dos senadores de cada uno de los 50 estados, (a) Si se selecciona aleatoriamente un comité de 8 senadores, ¿cuál es la probabilidad de que contenga al menos uno de los dos senadores de un estado concreto? (b) ¿Cuál es la probabilidad de que un grupo de 50 senadores seleccionados aleatoriamente contenga un senador de cada estado?

29. La probabilidad de que una persona nade es de 0.45 y la probabilidad de que una persona cace es de 0.58. Si la probabilidad de que una persona cace sabiendo que también caza es 0.21, encuentre la probabilidad de que:

    a) Cace y nade

    b) Cace si también nada

    c) Cace y no nade

    d) Cace o nade

30. La tabla adjunta muestra las frecuencias relativas para el daltonismo en hombres y mujeres, donde H representa hombres, M mujeres, D daltónico y ND no daltónico

	H	M
D	0.042	0.007
ND	0.485	0.466

    Si se escoge a una persona al azar, use la tabla para determinar las probabilidades siguientes:

    a) $P(H)$

    b) $P(H \cap D)$

    c) $P(D)$

    d) $P(H \cap ND)$

31. Si $P(E) = 0{,}2$ y $P(F) = 0{,}3$. Responda si puede ser cierto en cada una de las preguntas dadas y si es posible plantee un ejemplo.

    a) ¿$P(E \cup F) = 0{,}5$?

    b) ¿$P(E \cup F) = 0{,}7$?

    c) ¿$P(E \cup F) = 0{,}4$?

    d) ¿$P(E \cap F) = 0{,}2$?

    e) ¿$P(E \cap F) = 0{,}3$?

    f) ¿$P(E \cap F) = 0{,}1$?

    g) ¿$P(E \cap F) = 0{,}4$?

32. Si cuatro hombres y cuatro mujeres se colocan en fila, ¿Cuál es la probabilidad de que un arreglo aleatorio de los ocho individuos tenga

    a) hombres y mujeres alternados?

    b) A los hombres todos juntos?

33. De un conteo de tarjetas numeradas dei 1 al 10000. ¿Cuál es la probabilidad de que el número que te toque sea divisible exactamente por 5?

34. Un especialista en alergias alega que el 50 % de sus pacientes sufre de alergia ¿Cuál es la probabilidad de que

    a) de que 3 de sus siguientes cuatro pacientes sufran de alergia?

    b) Ninguno de los cuatro pacientes sufran de alergias?

35. De una caja que contiene 6 pelotas negras y 4 verdes, se sacan 3 en sucesión, reemplazándose cada una en la caja antes de extraer la siguiente. ¿Cuál es la probabilidad de que

    a) las 3 sen del mismo color

    b) sea al menos una de cada color

36. Un embarque de 12 televisores contiene 3 defectuosos. ¿En cuantas formas puede un hotel comprar 5 y recibir al menos 2 de los defectuosos?

37. En una cierta ciudad, 40 % de los votantes son Liberales y el 60 % son conservadores; 70 % de los republicanos y el 80 % de los demócratas están a favor de una de una consulta popular. al seleccionar al azar un votante de la ciudad, ¿Cuál es la probabilidad de qué está a favor de la consulta ?

38. ¿Cuántas manos de bridge que contengan 4 espadas, 6 diamantes, 1 de bastos y 2 de corazones son posibles?

39. Una empresa industrial grande utiliza 3 hoteles locales para proporcionar alojamiento a sus clientes durante la noche. De pasadas experiencias se sabe que al 20 % de ellos se les asigna habitación en el hotel de Santa Marta, el 50 % en Cartagena y al 30 % en Tolú. Si existe una falla en el servicio de plomería en el 5 % de la habitaciones en Santa Marta , en el 4 % en Cartagena y del 8 % en Tolú. ¿Cuál es la probabilidad de que

    a) a un cliente se le asigne un cuarto con problemas en plomería ?

    b) a una persona con un problema en plomería se asigne el hotel en Cartagena?

40. Un espacio muestral de 200 adultos se clasifica de acuerdo con su sexo y nivel de educación

Educación	Hombre	Mujer
Primaria	38	45
Secundaria	28	50
Bachillerato	22	17

    Si se selecciona aleatoriamente a una persona de ese grupo, encuentre la probabilidad de que

    a) no tenga grado de profesional dado de que sea mujer

    b) sea hombre dado que tiene educación superior

41. Se lanzan un par de dados, si se sabe que uno de ellos resulta en un 4, ¿cuál es la probabilidad de que

    a) el otro caiga en 6

    b) el total de ambos sea 9

42. Supongamos que se selecciona al azar un individuo de la población de todos los adultos hombres que en los Estados Unidos. Sea a el evento en que el individuo seleccionado tenga una estatura de más de 6 pies, y B el evento de que el individuo seleccionado sea un jugador profesional de baloncesto. ¿Cuál considera que sea mayor $P(B|A)$ o $P(A|B)$?. ¿Por qué?

43. Un circuito flexible se selecciona al azar de una corrida de producción de 1000 circuitos. Los defectos de manufactura se clasifican en tres diferentes tipos, denominados A. B y C. Los defectos de tipo A ocurren el 2 por ciento de las veces, los del tipo B, el 1 por ciento, y los de tipo C, el 1.5 por ciento. Además, se sabe que el 0.5 por ciento tienen los defectos de tipo A y B, el 0.6 por ciento, los defectos B y C, y el 0.4 por ciento presenta los defectos B y C, en tanto que el 0.2 por ciento tiene los tres defectos. ¿Cuál es la probabilidad de que el circuito flexible seleccionado tenga al menos uno de los tres tipos de defectos?

44. En un laboratorio de factores humanos, se miden tiempos de reacción, por ejemplo, el tiempo que transcurre desde el instante en que se despliega un número de posición en un tablero digital hasta que el sujeto presiona un botón localizado en la posición indicada. Participan dos sujetos, midiéndose el tiempo en segundos para cada individuo (t_1, t_2). ¿Cuál es el espacio muestral para este experimento? Presente los siguientes eventos como subconjuntos y márquelos sobre un diagrama:

$$\left(\frac{t_1+t_2}{2}\right) \leq 0{,}15, \ \text{máx}\,(t_1,t_2) \leq 0{,}15, \ |t_1-t_2| < 0{,}6.$$

45. Durante un periodo de 24 horas se entrará a un procesamiento computarizado. En un tiempo X y se saldrá en tiempo Y $\geq$ X. Considérense X y Y medido en horas en la linea del tiempo con el inicio del periodo de 24 horas como el origen. El experimento consiste en observar X y Y, (X, Y).

    a) Describa el espacio muestral S.

    b) Dibuje los siguientes eventos en el plano X, Y.

        1) El tiempo de utilización es una hora o menos

        2) El acceso es antes de t_1 y la salida después de t_2, donde $0 \leq t_1 \leq t_2 \leq 24$

        3) El tiempo de utilización es menor que el 20 por ciento del periodo.

46. Se prueban diodos de un lote uno a la vez y se marcan ya sea como defectuosos o como no defectuosos. Esto continúa hasta encontrar dos artículos defectuosos o cuando se han probado cinco artículos. Describa el espacio muestral para este experimento.

47. Un conjunto tiene cuatro elementos $A = \{a, b, c, d\}$. Describa el conjunto partes de A .

48. Describa el espacio muestral para cada uno de los siguientes experimentos:

    a) Un lote de 120 tapas de baterías para celdas de marcapasos contiene varias defectuosas debido a un problema con el material de barrera que se aplica en el sistema de alimentación. Se seleccionan tres tapas al azar (sin reemplazo) y se inspeccionan con cuidado siguiendo una reducción.

    b) Una paleta de 10 piezas fundidas contiene una unidad defectuosa y nueve en buen estado. Se seleccionan cuatro piezas al azar (sin reemplazo) y se inspeccionan.

49. El gerente de producción de cierta compañía está interesado en probar un producto terminado, que se encuentra disponible en lotes de tamaño 50. A él le gustaría volver a elaborar un lote si tiene la completa seguridad de que el 10 por ciento de los artículos son defectuosos. Decide seleccionar una muestra al azar de 10 artículos sin reemplazo y volver a producir el lote si éste contiene uno o más artículos defectuosos. ¿Este procedimiento parece razonable?

50. Una firma de transporte tienen un contrato para enviar una carga de mercancías de la ciudad W a la ciudad Z. No hay rutas directas que enlacen W con Z, pero hay seis carreteras de W a X y cinco de X a Z. ¿Cuántas rutas en total deben considerarse?

51. Un estado tienen un millón de vehículos registrados y está considerando emplear placas de licencia con seis símbolos en los que los primeros tres sean letras y los últimos tres, dígitos. ¿Es éste esquema factible?

52. El gerente de una pequeña planta desea determinar el número de maneras en que puede asignar trabajadores al primer turno. Cuenta con 15 hombres que pueden servir como operadores del equipo de producción, 8 que pueden desempeñarse como personal de mantenimiento y 4 que

pueden ser supervisores. Si el turno requiere 6 operadores, 2 trabajadores de mantenimiento, y 1 supervisor, ¿de cuántas maneras puede integrarse el primer turno?

53. Un lote de producción tiene 100 unidades de las cuales 20 se sabe que están defectuosas. Una muestra aleatoria de 4 unidades se selecciona sin reemplazo. ¿Cuál es la probabilidad de que la muestra no contenga más de 2 unidades defectuosas?

54. En la inspección de lotes de mercancías que están por recibirse, se emplea la siguiente regla de inspección en lotes que contienen 300 unidades; se selecciona una muestra al azar de 10 artículos. Si no hay más que un artículo defectuoso en la muestra, se acepta el lote. De otro modo se regresa al vendedor. Si la fracción defectuosa en el lote original es p', determinar la probabilidad de aceptar el lote como una función de p'

55. En una planta de plásticos, 12 tubos vacían diferentes químicos en un tanque de mezcla. Cada tubo tienen una válvula de cinco posiciones que mide el flujo dentro del tanque. Un día, mientras se experimenta con diferentes mezclas, se obtiene una solución que emite un gas venenoso, no habiéndose registrado los valores en las válvulas. ¿Cuál es la probabilidad de obtener esta misma solución cuando se experimenta de nuevo de manera aleatoria?

56. Ocho hombres y ocho mujeres con las mismas habilidades solicitan dos empleos. Debido a que los dos nuevos empleados deben trabajar estrechamente, sus personalidades deben ser compatibles. Para lograr esto, el administrador de personal ha aplicado una prueba y debe comparar las calificaciones para cada posibilidad. ¿Cuántas comparaciones debe efectuar el administrador?

57. En forma casual, un químico combinó dos sustancias de laboratorios que produjeron un producto conveniente. Desafortunadamente, su asistente no registró los nombres de los ingredientes. Hay cuarenta sustancias disponibles en el laboratorio. Si las dos en cuestión deben encontrarse mediante experimentos sucesivos de ensayo y error, ¿cuál es el número máximo de pruebas que pueden realizarse?

58. Un prisionero político será enviado a Siberia o a los Urales. Las probabilidades de que lo envíen a estos dos lugares son 0.6 y 0.4, respectivamente. Se sabe además que si un residente de Siberia se elige al azar hay una probabilidad de 0.5 de que lleve un abrigo de piel, en tanto que la probabilidad para lo mismos es de 0.7 en el caso de un residente de los Urales. Al llegar al exilio, la primera persona que ve el prisionero no lleva un abrigo de piel. ¿Cuál es la probabilidad de que esté en Siberia?

59. Se diseña un dispositivo de frenado para evitar que un automóvil patine en el que incluye un.sistema electrónico e hidráulico. El sistema completo puede descomponerse en tres subsistemas en serie que operan de manera independiente: un sistema electrónico, un sistema hidráulico y un accionador mecánico. En un frenado particular, las confiabilidades de estas unidades son aproximadamente 0.995, 0.993 y 0.994, respectivamente. Estime la confiabilidad del sistema.

60. Dos bolas se extraen de una urna que contiene m bolas numeradas del 1 a m. Se conserva la primera bola si tiene el número 1, y se regresa en caso contrario. ¿Cuál es la probabilidad de que la segunda bola extraída tenga el número 2?

61. Se eligen dos dígitos al azar de los dígitos del 1 al 9 y la selección es sin reemplazo (el mismo dígito no puede escogerse en ambas selecciones). Si la suma de los dígitos es par, encuentre la probabilidad de que ambos dígitos sean impares.

62. En cierta universidad 20 por ciento de los hombre y 1 por ciento de las mujeres miden más de dos metros de altura. Asimismo, 40 por ciento de los estudiantes son mujeres. Si se selecciona un estudiante al azar y se observa que mide de más de dos metros de altura, ¿Cuál es la probabilidad de que sea mujer?

63. En un centro de maquinaria hay cuatro máquinas automáticas que producen tornillos. Un análisis de los registros de inspección anteriores produce los siguientes datos:

Máquina	Porcentaje de producción	Porcentaje de defectos producidos
1	15	4
2	30	3
3	20	5
4	35	2

Las máquinas 2 y 4 son más nuevas y se les ha asignado más producción que a las máquinas 1 y 3. Suponga que la combinación de inventarios refleja los porcentajes de producción indicados.

a) Si se elige un tornillo al azar del inventario, ¿cuál es la probabilidad de que esté defectuoso?

b) Si se elige un tornillo y se encuentra que está defectuoso, ¿cuál es la probabilidad de que se haya producido en la máquina 3?

64. Supóngase que $A \subset B$. Demuéstrese que $B' \subset A'$.

65. Para tres sucesos cualesquiera A, B y C, demuéstrese que

$$A(B \cup C) = (A \cap B) \cup (A \cap C).$$

66. Para dos sucesos cualesquiera A y B, demuéstrese que

$$(A \cup B)' = A' \cap B' \text{ y } (A \cap B)' = A' \cup B'.$$

67. 4. Para cualquier conjunto de sucesos $A_i (i \in I)$, demuéstrese que

a) $\left(\bigcup_{i\in I} A_i\right)' = \bigcap_{i\in I} A_i'$

b) $\left(\bigcap_{i\in I} A_i\right)' = \bigcup_{i\in I} A_i'$

68. 5. Supóngase que se selecciona una carta de una baraja de veinte cartas que contiene diez cartas rojas numeradas del 1 al 10 y diez cartas azules numeradas del 1 al 10. Sea A el suceso de seleccionar una carta con un número par, sea B el suceso de seleccionar una carta azul y sea C el suceso de seleccionar una carta con un número menor que 5. Descríbanse el espacio muestral S y cada uno de los siguientes sucesos en palabras y en subconjuntos de S:

a) $A \cap B \cap C$.

b) $B \cap C'$.

c) $A \cup B \cup C$.

d) $A \cap (B \cup C)$.

e) $A' \cap B' \cap C'$.

69. Un estudiante seleccionado de una clase puede ser chico o chica. Si la probabilidad de que un chico sea seleccionado es 0.3, ¿cuál es la probabilidad de que sea seleccionada una chica?

70. Se selecciona una bola de una urna que contiene bolas rojas, blancas, azules, amarillas y verdes. Si la probabilidad de seleccionar una bola roja es $\frac{1}{5}$ y la de seleccionar una blanca es $\frac{2}{5}$ ¿cuál es la probabilidad de seleccionar una bola azul, amarilla o verde?

71. Si la probabilidad de que un estudiante A suspenda un cierto examen de estadística es 0.5, la probabilidad de que un estudiante B suspenda el examen es 0.2 y la probabilidad de que ambos estudiantes A y B suspendan el examen es 0.1,

    a) ¿cuál es la probabilidad de que al menos uno de estos dos estudiantes suspenda el examen?

    b) ¿cuál es la probabilidad de que ni el estudiante A ni el B suspendan el examen?

    c) ¿cuál es la probabilidad de que exactamente uno de los dos estudiantes suspenda el examen?

72. Considere dos sucesos Ay B tales que $P(A) = \frac{1}{3}$ y $P(B) = \frac{1}{2}$. Determínese el valor de $P(B \cap A')$ para cada una de las siguientes condiciones:

    a) A y B son disjuntos

    b) $A \subset B$

    c) $P(A \cap B) = \frac{1}{8}$

73. Si el 50 % de las familias de cierta ciudad están suscritas al periódico matinal, el 65 % de las familias al periódico vespertino y el 85 % al menos a uno de los dos periódicos, ¿ cuál es la proporción de familias que están suscritas a los dos periódicos?

74. Considérense dos sucesos A y B con $P(A) = 0{,}4$ y $P(B) = 0{,}7$. Determínense los posibles valores máximo y mínimo de $P(A \cap B)$ y las condiciones en las cuales se consigue cada uno de estos valores.

75. Demuéstrese que para dos sucesos A y B cualesquiera, la probabilidad de que exactamente uno de los dos sucesos ocurra está dada por la expresión

$$P(A) + P(B) - 2P(A \cap B).$$

76. Se selecciona un punto (x, y) del cuadrado S que contiene todos los puntos (x, y) tales que $0 \le x \le 1$ y $0 \le y \le 1$. Supóngase que la probabilidad de que el punto seleccionado pertenezca a cualquier subconjunto específico de S es igual al área de ese subconjunto. Determínese la probabilidad de cada uno de los siguientes subconjuntos:

    a) el subconjunto de puntos tales que $(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 > \frac{1}{4}$

    b) el subconjunto de puntos tales que $\frac{1}{2} < x + y < \frac{3}{2}$

    c) el subconjunto de puntos tales que $y < 1 - x^2$

    d) el subconjunto de puntos tales que $x = y$.

77. Sea $A_1, A_2, A_3, \cdots$ una serie numerable de eventos y sea $B_1, B_2, B_3, \cdots$ otra serie numerable de eventos tal que

$$B_1 = A_1,\ B_2 = A_1' \cap A_2,\ B_3 = A_1' \cap A_2' \cap A_3 ...$$

    Demuéstrese que

    a) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(B_i)$

    b) $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(B_i)$ para $n = 1, 2, 3, ...$

78. Sea $A_1, A_2, A_3, \cdots A_n$ una serie cualquiera de eventos, demostrar que

$$P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)$$

79. Supóngase que una urna contiene una carta azul y cuatro cartas rojas. A, B, C y D. Supóngase también que dos de estas cinco cartas se extraen al azar sin reemplazo

    a) Si se sabe que se ha extraído la carta A, ¿cuál es la probabilidad de que ambas cartas sean rojas?

    b) Si se sabe que se ha extraído una carta roja, ¿cuál es la probabilidad de que ambas cartas sean rojas?

80. La probabilidad de que cualquier niño de una familia determinada tenga ojos azules es 1/4, y esta característica es heredada por cada niño de la familia independientemente de los demás. Si hay cinco niños en la familia y se sabe que al menos uno de estos niños tiene ojos azules,

    a) ¿cuál es la probabilidad de que al menos tres de los niños tengan ojos azules?

    b) Considérese la familia de cinco nios

        1) Si se sabe que el niño más pequeño de la familia tiene los ojos azules, ¿cuál es la probabilidad de que al menos tres de los niños tengan ojos azules?

        2) Explíquese por qué la respuesta es diferente de la respuesta de la parte (a)

81. Considérese la siguiente versión del juego de dados: El jugador lanza dos dados. SÍ la suma en el primer lanzamiento es 7 u 11, el jugador gana el juego. Si la suma en el primer lanzamiento es 2, 3 ó 12, el jugador pierde. Sin embargo, si la suma en el primer lanzamiento es 4, 5, 6, 8, 9 ó 10, entonces se lanzan los dos dados una y otra vez hasta que la suma sea 7, lío el valor original. Si el valor original se obtiene por segunda vez antes de obtener 7 u 11, entonces el jugador gana. Si se obtiene un total de 7 o de 11 antes de obtener el valor original por segunda vez, entonces el jugador pierde. Determínese la probabilidad de que el jugador gane este juego.

82. En una ciudad determinada, el 30 % de las personas son conservadores, el 50 % son liberales y el 20 % son independientes. Los registros muestran que en unas elecciones concretas, votaron el 65 % de los conservadores, el 82 % de los liberales y el 50 % de los independientes. Si se selecciona al azar una persona de la ciudad y se sabe que no votó en las elecciones pasadas, ¿ cuál es la probabilidad de que sea un liberal?

83. Supóngase que cuando una máquina está correctamente ajustada, el 50 % de los artículos que produce son de alta calidad y el otro 50 % son de calidad media. Supóngase, sin embargo, que la máquina está mal ajustada durante el 10 % del tiempo y que, en estas condiciones, el 25 % de los artículos producidos por ella son de alta calidad y el 75 % de los artículos son de calidad media.

    a) Supóngase que cinco artículos producidos por la máquina en un tiempo determinado son seleccionados al azar e inspeccionados. Si cuatro de estos artículos son de alta calidad y uno es de calidad media, ¿cuál es la probabilidad de que la máquina estuviera correctamente ajustada durante ese tiempo?

b) Supóngase que se selecciona un artículo adicional, que fue producido por la máquina al mismo tiempo que los otros cinco, y resulta ser de calidad media. ¿Cuál es la nueva probabilidad final de que la máquina estuviera correctamente ajustada?

84. Supóngase que una caja contiene cinco monedas y que la probabilidad de obtener cara en un lanzamiento es distinta para cada moneda. Sea p_i la probabilidad de obtener cara al lanzar la $i-ésima$ moneda $(i = 1, ..., 5)$ y Supóngase que $p_1 = 0$, $p_1 = \frac{1}{4}$, $p_3 = 1/2$, $p_4 = 3/4$ y $p_5 = 1$. Si se selecciona una moneda de la caja al azar y que al lanzarla una vez se obtiene una cara. ¿Cuál es la probabilidad final de que se haya seleccionado la $i-ésima$ moneda $(i = 1, ..., 5)$?

CAPÍTULO

4 Variables aleatorias

Contenido Del Capítulo

4.1 Introducción

Cuando definimos la probabilidad en el axioma 1 del capítulo 3 se estableció de la siguiente manera

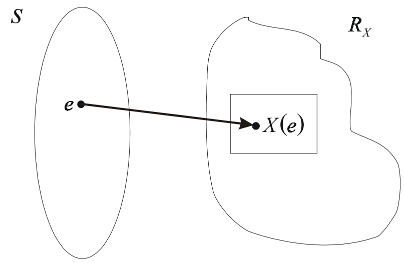
$$\begin{array}{rcl} P:\mathcal{A} & \rightarrow & [0,1]\subset \mathbb{R} \\ A\in\mathcal{A} & \mapsto & 1\geq P(A)\geq 0 \end{array}$$

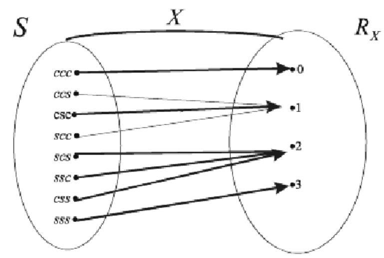
Tomamos un elemento de $\mathcal{A}$ y le asignamos un número, pero esta forma de trabajar a veces dificulta la interpretación del concepto de probabilidad, porque los elementos de $\mathcal{A}$ no son números si no conjuntos, pero que tal si $\mathcal{A}$ estuviera formado no por subconjuntos de S, sino por números, entonces consideremos el ejemplo de lanzar tres monedas al aire para mostrar como podríamos hacer esto.

En este caso el espacio muestral es

$$S=\{ccc,ccs,csc,css,scc,scs,ssc,sss\}$$

ahora si definimos una función, por decir X tal que a cada evento fundamental le asignamos un número de la siguiente forma





De este modo aparece el concepto de variable unidimensional

$$\begin{array}{ccc} X:S & \longrightarrow & \mathbb{R} \\ e & \longmapsto & X(e)=x \end{array}$$

de acuerdo con el ejemplo anterior se define la variable

$$X \equiv número \quad de \quad caras$$

de la siguiente forma

$$\begin{gathered} X:\longrightarrow \mathbb{R} \\ X(CCC)=3 \\ X(CCS)=X(CSC)=X(SCC)=2 \\ X(CSS)=X(SSC)=X(SCS)=1 \\ X(SSS)=0 \end{gathered}$$

la variable X no recibe el calificativo de aleatoria por el echo de asignarle a un elemento $e \in S$ un valor numérico, (por que de echo este valor esta definido de forma precisa), sino por el echo de que al realizar el experimento no sabemos que elemento de S puede ocurrir

En función del espacio del rango R_X esta función puede ser clasificada como

- Variable aleatoria discreta . Si toma un número finito o numerable de valores, por ejemplo

$$X:\longrightarrow \mathbb{N}$$

- Variable aleatoria continua. Si toma un número de valores no numerables, por ejemplo

$$X:\longrightarrow \mathbb{R}$$

Definición 4.1

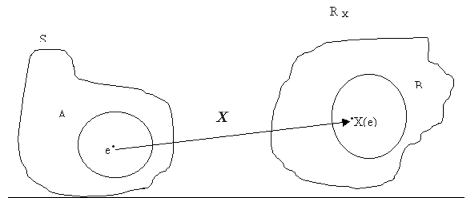
Si E es un experimento que tiene como espacio muestral a S, y X es una función que le asigna un número real $X(e)$ a todo resultado $e \in S$, entonces X se llama variable aleatoria

Definición 4.2

Si S es el espacio muestral de un experimento E y una variable aleatoria X con rango el espacio R_X se define en S, y además si el $A \subset S$ y $B \subset R_X$, entonces A y B son eventos equivalentes si

$$\begin{gathered} P_X(B)=P(A) \\ donde \quad A=\{e \in S : X(e) \in B\} \end{gathered}$$

como indica la 4.1.



Definición 4.3

Si $A \subset S$ y $B \subset R_X$ definimos la probabilidad de B como

$$P_X = P\left(X^{-1}(B)\right) = P(A)$$

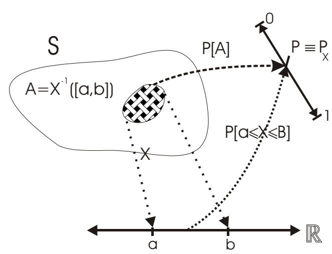
donde

$$A = \{e \in S : X(e) \in B\}$$
$$X^{-1}(B) = \{e \in S : X(e) \in B\}$$

si sobre los elementos de S existe una distribución de probabilidad, esta se transmite a los valores que toma la variable X. Es decir toda variable aleatoria conserva la estructura probabilística del experimento aleatorio que describe, es decir si P es la función de probabilidad definida sobre $S \in \mathcal{A}$, esta induce otra función P_X definida sobre $I\!R$, de forma que conserva los valores de las probabilidades

$$P_X(X = x) = P(\{e \in S : X(e) = x\})$$
$$P_X(X \in (a,b)) = P(\{e \in S : X(e) \in (a,b)\})$$

como indica la 4.1



Espacio probabilístico

De ahora en adelante denotaremos P_X como P pero no se debe confundir con la probabilidad P que definimos en el capitulo 3

4.2 Variables aleatorias discretas

Definición 4.4

Si X es una variable aleatoria discreta, asociamos un número

$$f\left(x_i\right)=P\left(X=x_i\right)$$

como cada resultado x_i en R_X para $i=1,2,3\cdots,n,\cdots$, donde los números $f\left(x_i\right)$ satisfacen

1. $f\left(x_i\right)\geq 0 \quad$ para toda i
2. $\sum_{i=1}^{x} f\left(x_i\right)=1$La función $f\left(x_i\right)$ se llama función de probabilidad o ley de probabilidad de la variable aleatoria, y la colección de pares $\left(x_i, f\left(x_i\right)\right)$ se llama distribución de probabilidad de X
3. Dada una variable aleatoria discreta $X:S\longrightarrow \mathbb{N}$, su función de probabilidad f se define de modo que $f\left(x_i\right)$ es la probabilidad de que X tome ese valor

$$\begin{array}{ccc} f:\mathbb{N} & \longrightarrow & [0,1] \\ x_i & \longmapsto & f\left(x_i\right)=P\left(X=x_i\right) \end{array}$$

si x_i no es uno de los valores que puede tomar X, entonces $f\left(x_i\right)=0$

Definición 4.5

[Función de distribución]De una variable aleatoria discreta, F que se define de modo que si $x_i\in\mathbb{R}, F\left(x_i\right)$ es igual a la probabilidad de que X tome un valor inferior o igual a x_i, es decir,

$$\begin{array}{ccc} F:\mathbb{N} & \longrightarrow & [0,1] \\ x_i & \longmapsto & F\left(x_i\right)=P\left(x_i\geq X\right) \end{array}$$

La función de distribución F, es una función no decreciente, es decir. Si

$$x_{1<}x_2 \Longrightarrow F\left(x_2\right)\geq F\left(x_1\right)$$

Además, es continua a la derecha

$$\lim_{x\longrightarrow a^+} F(x)=F(a)$$

y

$$F(-\infty)=\lim_{x\longrightarrow -\infty} F\left(x_i\right)=0$$

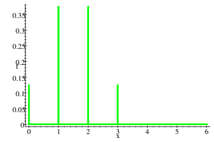
$$F(+\infty)=\lim_{x\longrightarrow +\infty} F\left(x_i\right)=1$$

Ejemplo: 4.2.1

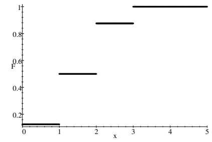
[] Analicemos el experimento del lanzamiento de la moneda tratado en la introducción y determinemos $f(x), F(x)$

Presentación tabular

x	$f(x)$
0	1/8
1	3/8
2	3/8
3	1/8



Función de probabilidad



Función de distribución

Ejemplo: 4.2.2

[] Supóngase que tenemos una variable aleatoria X con una distribución de probabilidad dada por la relación

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, ..., n \\ 0 & \text{de otro modo} \end{cases}$$

donde n es un entero positivo y $0 \leq p \leq 1$. El ejemplo anterior es un caso particular de ésta distribución de probabilidad llamada Binomial

Ejemplo: 4.2.3

[] Supóngase que hay 100 artículos de los cuales hay 5 defectuosos. Se toma una muestra de de 4 artículos sin reemplazo Determine la distribución de probabilidad para los artículos defectuosos

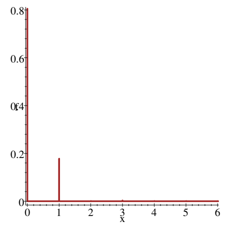
Solución

Si X representa el número de artículos defectuosos entonces la probabilidad de que se escojan x artículos defectuosos es

$$f(x) = P(X = x) = \frac{\binom{5}{x}\binom{100-5}{4-x}}{\binom{100}{4}} \text{ si } x = 0, 1, 2, 3, 4$$

Ya que los casos posibles serían $\binom{100}{4}$ porque de 100 se escogen 4 y los casos favorables sería $\binom{5}{x}\binom{100-5}{4-x}$, debido a que es una muestra sin reemplazo el número de combinaciones distintas que se pueden formar con x artículos defectuosos para formar una muestra de 5 es $\binom{5}{x}$ y la afirmación de que se van a escoger 5 al azar significa que todas las $\binom{100-5}{4-x}$ son igualmente posibles

x	$f(x)$
0	0,805
1	0,178
2	0,014
3	0,003
4	0,00



Función de probabilidad ej 4.3

Ejemplo: 4.2.4

[] Supóngase que una variable aleatoria X tiene una distribución discreta con la siguiente función de probabilidad

$$f(x) = \begin{cases} cx & \text{para} \quad x = 1,2,3,4,5 \\ 0 & \text{en otro caso} \end{cases}$$

Determine el valor de la constante c

Solución

Debido a que f(x) es una función de probabilidad entonces

$$\sum_{i=1}^{5} f(x_i) = 1 \Longrightarrow$$

$$15c = 1 \Longrightarrow$$

$$c = \frac{1}{15}$$

Si la variable X puede tomar un número enumerable de valores $x_1, x_2, x_3, \cdots$, entoces se tiene que

$$\sum_{i=1}^{\infty} f(x_i) = \sum_{i=1}^{\infty} P(X = x_i) = 1$$

4.3 Variables aleatorias continuas

Cuando tenemos una variable aleatoria continua no tiene sentido realizar una suma de las probabilidades de cada uno de los valores que toma, ya que el conjunto es no enumerable por lo que hay que introducir otro concepto.

Sea $f : \mathbb{R} \longrightarrow \mathbb{R}$ una función llamada función de densidad de una variable aleatoria continua, integrable que cumple las propiedades

$$f(x) \geq 0$$

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1$$

además para todo $[a,b]$ se tiene

$$P(a \leq x \leq b) = \int_a^b f(x)\,dx$$

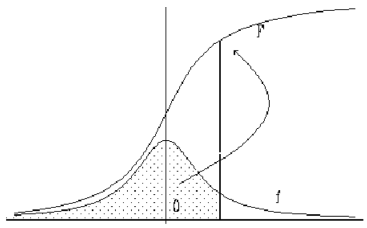
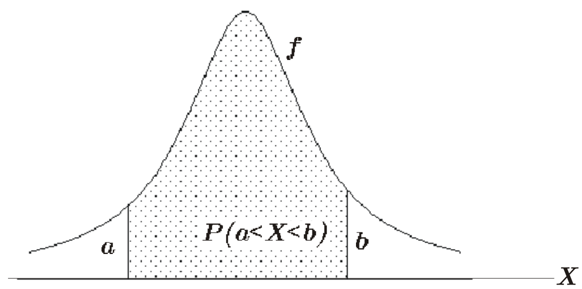


Figura 4.1: Función de distribución F



Función de densidad de $f(x)$

Al observar la 4.3 observamos que $P(a \leq x \leq b)$ es el área bajo la curva de f

Por ser f integrable entonces la la probabilidad en un punto es nula, es decir

$$P(X = a) = P(a \leq x \leq a) = \int_a^a f(x)\,dx = 0$$

Debido a lo anterior se tiene que

- La función de densidad no es única
- $P(a \leq x \leq b) = P(a < x < b)$
- $P(a \leq x \leq b) = P(a \leq x < b) = P(a < x \leq b)$

La función de distribución de una variable aleatoria, F, continua se define de modo que dado $x \epsilon\ \mathbb{R}$, $F(x)$ es la probabilidad de que X sea mayor o igual que x, es decir

$$\begin{array}{lcl} F : \mathbb{R} & \longrightarrow & [0,1] \\ x & \longmapsto & F(x) = P(x \geq X) = \int_{-\infty}^{x} f(t)\,dt \end{array}$$

Proposición 5

ado un intervalo de la forma $(a, b]$, tenemos

$$\begin{aligned} P(X \in (a,b]) &= \int_a^b f(x)\,dx \\ &= \int_{-\infty}^b f(x)\,dx - \int_{-\infty}^a f(x)\,dx \\ &= F(b) - F(a) \end{aligned}$$

Se puede observar que la cantidad $F(b) - F(a)$ representa la masa de probabilidad extendida alo largo del intervalo $(a, b]$. Si dividimos esta cantidad por la longitud del intervalo,

$$\frac{F(b) - F(a)}{b - a}$$

tenemos la masa media de probabilidad por unidad de longitud en $(a, b]$, es decir, su densidad media de probabilidad, ahora si

$$\lim_{a \longrightarrow b} \frac{F(b) - F(a)}{b - a} = F'(b) = f(b)$$

que es la densidad de probabilidad en el punto b

- Si X es una variable continua la función de distribución F es no decreciente, es decir si

$$x_1 < x_2 \Longrightarrow F(x_2) \geq F(x_1)$$

- esta función es absolutamente convergente y se verifica

$$\begin{aligned} F(-\infty) &= \lim_{x \longrightarrow -\infty} F(x) = 0 \\ F(+\infty) &= \lim_{x \longrightarrow +\infty} F(x)\,(1) \end{aligned}$$

Ejemplo: 4.3.5

[] Suponga que la f.d.p de una variable aleatoria X es

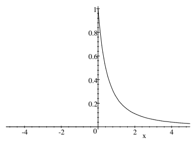
$$f(x) = \begin{cases} 0 & si \quad x \leq 0 \\ \frac{1}{(1+x)^2} & si \quad x > 0 \end{cases}$$

- Represente graficamente la f.d.p
- encuentre la f.d y representela gráficamente

Solución

- Su representación gráfica es

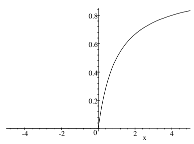
$$f(x) = \begin{cases} 0 & si \quad x \leq 0 \\ \frac{1}{(1+x)^2} & si \quad x > 0 \end{cases}$$



Gráfica del ejemplo

- Si $x > 0$ entonces $F(x) = \int_0^x \frac{1}{(1+t)^2} dt = 1 - \frac{1}{1+x}$ entonces

$$F(x) = \begin{cases} 0 & si \quad x \leq 0 \\ 1 - \frac{1}{1+x} & si \quad x > 0 \end{cases}$$

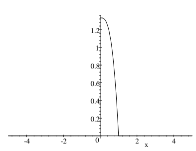


Ejemplo: 4.3.6

[]

- Determine la gráfica de la f.d.p y determinar la f.d

$$f(x) = \begin{cases} \frac{4}{3}\left(1 - x^3\right) & si \quad 0 < x < 1 \\ 0 & si \quad x \geq 1 \\ 0 & si \quad 0 \leq x \end{cases}$$

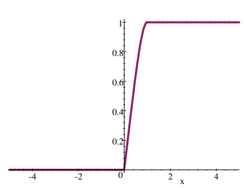


f.d.p de $\frac{4}{3}(1 - x^3)$

- Si $0 < x < 1$ entonces

$$F(x) = 1 - \int_x^1 \frac{4}{3}\left(1 - t^3\right) dt = 1 - \left(\frac{4}{3}x - \frac{1}{3}x^4\right)$$

$$F(x) = \begin{cases} 0 & si \quad x \leq 0 \\ \frac{4}{3}x - \frac{1}{3}x^4 & si \quad 0 < x < 1 \\ 1 & si \quad x \geq 1 \end{cases}$$



f.d del ejemplo 4.6

4.4 Distribuciones de probabilidad conjuntas

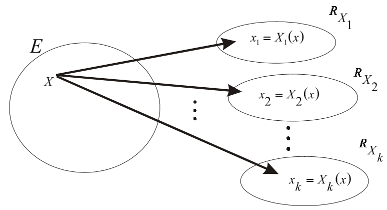


Figura 4.2: vector k-dimensional

En muchas situaciones tratamos problemas donde intervienen más de una variable aleatoria en forma simultánea, por lo que el objetivo de esta sección es el de tratar y formular las distribuciones de probabilidad conjuntas para dos o más variables aleatorias.

Definición 4.6

si S es el espacio muestral de un experimento E, y $X_1, X_2, X_3, \cdots, X_k$ son funciones, cada una de las cuales asignan un número real, $X_1(x), X_2(x), X_3(x), \cdots, X_k(x)$ a cada resultado x, designaremos como $(X_1, X_2, X_3, \cdots, X_k)$ el vector $k-dimensional$, el espacio del rango del vector $(X_1, X_2, X_3, \cdots, X_k)$ es el conjunto de todos los valores posibles del vector aleatorio

En la mayor parte de esta sección trataremos el caso bidimensional es decir el vector $k - dimensional$ es (X_1, X_2)

Definición 4.7

Funciones de probabilidad bivariada

1. Caso discreto: para cada resultado (x_{1i}, x_{2j}) de (X_1, X_2), asociamos un número

$$f(x_{1i}, x_{2j}) = P(X_1 = x_{1i}, X_2 = x_{2j})$$

donde

$$f(x_{1i}, x_{2j}) \geq 0$$
para todo $i \in I, j \in J$ siendo I, J conjuntos de subíndices

y

$$\sum_{i \in I} \sum_{j \in J} f(x_{1i}, x_{2j}) = 1$$

Los valores $((x_{1i}, x_{2j}), f(x_{1i}, x_{2j}))$ para todo $i \in I$ y $j \in J$ forman la distribución de probabilidad de (X_1, X_2)

2. Caso continuo . Si (X_1, X_2) es un vector aleatorio continuo con espacio del rango, R, en $\mathbb{R}^2$, entonces f, la función de densidad conjunta, tiene las siguientes propiedades

$$f(x_1, x_2) \geq 0 \; \forall (x_1, x_2) \in R$$

y

$$\int \int_R f(x_1, x_2)\, dx_1 dx_2 = 1$$

Definición 4.8

Se dice que n variables aleatorias $X_1, X_2, X_3, \cdots, X_k$,tiene una distribución discreta conjunta si el vector aleatorio $X = (X_1, X_2, X_3, \cdots, X_k)$ puede tomar solamente un número finito o una sucesión finita de valores distintos posibles$(x_1, x_2, x_3, \cdots, x_k)$ en $\mathbb{R}^k$. La función de probabilidad conjunta de $X_1, X_2, X_3, \cdots, X_k$ se define entonces como la función f tal que para cualquier punto $x = (x_1, x_2, x_3, \cdots, x_k) \in A \subset \mathbb{R}^k$,

$$f(x) = P(X = x) \quad \forall x \in A$$

donde A es cualquier subconjunto de $\mathbb{R}^k$

$$P(x \in A) = \sum_{x \in A} f(x)$$

Definición 4.9

Distribuciones continuas. se dice que n variables aleatorias $X_1, X_2, X_3, \cdots, X_k$ tienen una distribución conjunta continua si existe una función no negativa f definida sobre $\mathbb{R}^k$ tal que para cualquier subconjunto $A \subset \mathbb{R}^k$

$$\begin{aligned} & P((X_1, X_2, X_3, \cdots, X_k) \in A) \\ & = \int \cdots \int_A f(x_1, x_2, x_3, \cdots, x_k)\, dx_1 dx_2 \cdots dx_k \end{aligned}$$

Definición 4.10

La f.d conjunta de k variables aleatorias
$X_1, X_2, X_3, \cdots, X_k$, se define como la función F cuyo valor en cualquier punto $(x_1, x_2, x_3, \cdots, x_k)$ de un espacio k-dimensional $\mathbb{R}^k$ está dado por la relación

$$F(x_1, x_2, x_3, \cdots, x_k) = P(X_1 \leq x_1, \cdots X_k \leq x_k)$$

Ésta f.d satisface todas las propiedades de la f.d univariada.

En el caso bivariado. Si X y Y son variables aleatorias con f.d.p conjunta f se tiene que

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, du dv$$

Si la distribución conjunta de $X_1, X_2, X_3, \cdots, X_k$ es continua, entonces la f.d.p conjunta f se puede obtener a partir de la f.d conjunta F utilizando la relación

$$f(x_1, x_2, x_3, \cdots, x_k) = \frac{\partial^k F(x_1, x_2, x_3, \cdots, x_k)}{\partial x_1 \partial x_2 \cdots \partial x_k}$$

para todos los puntos $(x_1, x_2, x_3, \cdots, x_k)$ donde exista la derivada.

En el caso bivariado

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

$\forall (x, y)$ donde exista la derivada parcial de 2^o orden

Ejemplo: 4.4.7

[] Supóngase que la variable aleatoria X puede tomar solamente los valores 1,2,3, y que la variable Y puede tomar solamente los valores 1,2,3,4, donde la f.p conjunta de X e Y es como indica la tabla

$X \mid Y$	1	2	3	4
1	0.1	0	0.1	0
2	0.3	0	0.1	0.2
3	0	0.2	0	0

determine los valores de $P(X \geq 2, Y \geq 2)$ y $P(X = 1)$

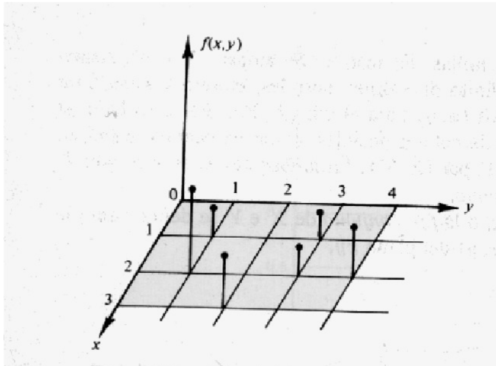


Figura 4.3: f.d.p del ejemplo 4.7

Solución

Sumando$f\left(x,y\right)$ sobre todos los valores de $x\geq 2$ y de $y\geq 2$, se obtiene

$$\begin{aligned}p\left(X\geq 2,Y\geq 2\right) &= f\left(2,2\right)+f\left(2,3\right)+f\left(2,4\right)+\\ &+f\left(3,2\right)+f\left(3,3\right)+f\left(3,4\right)\\ &=0{,}5\end{aligned}$$

$$P\left(X=1\right)=\sum_{y=1}^{4}f\left(1,y\right)=0{,}2$$

Ejemplo: 4.4.8

[] Supóngase que la f.d.p conjunta de X e Y es la siguiente

$$f\left(x,y\right)=\begin{cases} cx^2y & para \quad x^2\leq y\leq 1\\ 0 & \text{en otro caso}\end{cases}$$

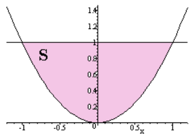
- Determine el valor de la constante c
- $P\left(X\geq Y\right)$

Solución

Sea el conjunto S de puntos (x, y) para los que f$(x, y) > 0$ está representado en la figura 4.8 . puesto que $f(x, y) = 0$ fuera de S, resulta que

$$\int\int_s f(x, y)\, dxdy = \int_{-1}^{1}\int_{x^2}^{1} cx^2 ydydx = \frac{4}{21}c$$

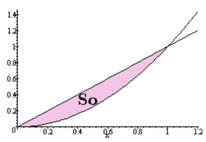
y como esta integral debe ser 1, entonces c=$\frac{21}{4}$



- Sea S_0 el conjunto donde $x \geq y$, entonces

$$\int\int_{S_0} f(x, y)\, dxdy =$$

$$\int_0^1 \int_{x^2}^{x} \frac{21}{4}x^2 ydydx = \frac{3}{20}$$



Fig(b) Ejemplo 4.8

Ejemplo: 4.4.9

[] Supóngase que X e Y son variables aleatorias que solamente pueden tomar valores en los intervalos $0 \leq X \leq 2, 0 \leq Y \leq 2$. Supóngase también que la f.d. conjunta de X e Y, para todos los valores $0 \leq x \leq 2, 0 \leq y \leq 2$, es la siguiente

$$F(x, y) = \frac{1}{16}xy(x+2) \tag{4.1}$$

Solución

Se determinará primero la F_x de la variable aleatoria X y luego la f.d.p conjunta f de X e Y.
El valor de $F(x, y)$ en cualquier punto (x, y) del plano xy que nos representa un par de valores posibles de X e Y se puede calcular a partir 4.1, teniendo en cuenta que $F(x, y) = P(X \leq x, Y \leq y)$. Por tanto, si $x < 0$ o $y < 0$, entonces $F(x, y) = 0$. Si $x > 2$ o $y > 2$, entonces $F(x, y) = 1$, Si $0 \leq x \leq 2, y \geq 2$ entonces $F(x, y) = F(x, 2)$ y resulta de la ecuación 4.1 ya que $F(x, y) = 0$ si $y > 0$

$$F(x, y) = \frac{1}{8}x(x+2)$$

análogamente, si $0 \leq y \leq 2, x > 2$, entonces

$$F(x, y) = \frac{1}{8}y(y+2)$$

La función $F(x, y)$ queda así definida para todo punto del plano xy .
Haciendo $y \longrightarrow \infty$, se determina que la f.d de la variable aleatoria X es

$$F_x = \begin{cases} 0 & si \quad x < 0 \\ \frac{1}{8}x(x+2) & si \quad 2 \geq x \geq 0 \\ 1 & si \quad x > 2 \end{cases}$$

Además, para $0 < x < 2, 0 < y < 2$,

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{1}{8}(x+y)$$

Mientras que si $x < 0, y < 0, x > 2, y > 2, entonces$

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = 0$$

Por tanto, la f.d.p conjunta de X e Y es

$$f(x,y) = \begin{cases} \frac{1}{8}(x+y) & si \quad 0 < x < 2, 0 < y < 2 \\ 0 & \text{en otro caso} \end{cases}$$

4.4.1. Distribuciones marginales

cuando nos interesa conocer la distribución por ejemplo de X_1 solamente entonces es necesario introducir el concepto de una distribución llamada marginal

- En el caso discreto la distribución marginal para X_1 y X_2 es

$$f(x_1) = \sum_{todo\, j} f(x_{1i}, x_{2j})\ \forall i = 1, 2, \cdots$$
$$f(x_2) = \sum_{todo\, i} f(x_{1i}, x_{2j})\ \forall j = 1, 2, \cdots$$

- En el caso continuo la distribución marginal para X_1 y X_2 es

$$f(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2$$
$$f(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1$$

4.4.1.1. Tablas de doble entrada

Consideremos una población de n individuos, desde un punto de vista descriptivo, donde cada uno de ellos presenta dos caracteres que representaremos mediante las variables X e Y donde la variable X tiene k modalidades y la variable Y tiene p modalidades, el problema ahora es tratar de representar toda la información de manera adecuada y fácil de interpretar, por lo que creamos una tabla formada por kp celdas de forma que tenga k filas y p columnas, donde la celda que denotaremos con el subíndice ij representará el número de elementos de la muestra que presentan simultáneamente las modalidades x_i, y_j

$X \vert Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_p	total
x_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1p}	$n_{1\cdot}$
x_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2p}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	n_{11}	n_{ij}	$\cdots$	n_{ip}	$n_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}	$\cdots$	n_{kj}	$\cdots$	n_{kp}	$n_{k\cdot}$
total	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot j}$	$\cdots$	$n_{\cdot p}$	$n_{\cdot\cdot}$

El número de individuos que presentan la modalidad $x_i$es la frecuencia absoluta marginal de x_i y se representa $n_{i\cdot}$ y evidentemente es

$$n_{i\cdot} = \sum_{j=1}^{p} n_{ij}$$

de manera análoga se define la frecuencia absoluta marginal para la modalidad y_j

$$n_{\cdot j} = \sum_{i=1}^{k} n_{ij}$$

El número total de elementos es

$$n = n_{\cdot\cdot} = \sum_{i=1}^{k} n_{i\cdot} = \sum_{j=1}^{p} n_{\cdot j}$$

Llamaremos frecuencia relativa $f_{ij} = \frac{n_{ij}}{n}$

las frecuencia relativas marginales serían

$$f_{i\cdot} = \sum_{j=1}^{p} f_{ij} = \frac{n_{i\cdot}}{n}$$
$$f_{\cdot j} = \sum_{j=1}^{p} f_{ij} = \frac{n_{\cdot j}}{n}$$
$$f_{\cdot\cdot} = 1$$

Ejemplo: 4.4.10

[] Supóngase que X e Y tienen la f.p conjunta dada por la tabla del ejemplo 4.7. La f.p marginal f_x de X se puede determinar sumando los valores de cada fila de esta tabla . De esta manera se obtiene que

$$f_x(1) = 0{,}2$$
$$f_x(2) = 0{,}6$$
$$f_x(3) = 0{,}2$$
$$f_x(x) = 0 \text{ para los restantes valores de } x$$

Ejemplo: 4.4.11

[] Supóngase que la f.p.d conjunta de X e Y es la descrita en el ejemplo 4.8. Obtenga la f.d.p marginal

Solución

Se puede observar en la figura (a) del ej. 4.8 que X no puede tomar ningún valor fuera del intervalo $-1 \leq X \leq 1$ Por tanto, $f_x(x) = 0$ para $x < -1$ o $x > 1$. Además, para $-1 \leq x \leq 1$, se observa en la misma figura que $f(x,y) = 0$, a menos que $x^2 \leq y \leq 1$.Por tanto, para $-1 \leq x \leq 1$

$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)\,dy =$$
$$\int_{x^2}^{1} \left(\frac{21}{4}\right) x^2 y dy = \frac{21}{8} x^2 (1-x)(1+x)\left(x^2+1\right)$$

Se puede observar en la figura (a) del ej. 4.8 que Y no puede tomar ningún valor fuera del intervalo $0 \leq Y \leq 1$. Por tanto, $f_y(y) = 0$ para $y < 0$ o $y > 1$. Además, para $0 \leq y \leq 1$, se observa en la misma fig. Que $-\sqrt{y} \leq x \leq \sqrt{y}$.Por tanto, para $0 \leq y \leq 1$

$$f_y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx =$$
$$= \int_{-\sqrt{y}}^{\sqrt{y}} \left(\frac{21}{4}\right) x^2 y dx = \left(\frac{7}{2}\right) y^{5/2}$$

4.4.1.2. Variables aleatorias independientes

Se dice que dos variables aleatorias son independientes si Para $A, B \subset \mathbb{R}$

$$P(X \in A, Y \in B) = P(X \in A)\,P(Y \in B)$$

Es decir si

$$P(x \geq X, y \geq Y) = P(x \geq X)\, P(y \geq Y)$$
$$F(x, y) = F_x(x)\, F_y(y)$$
$$f(x, y) = f_x(x)\, f_y(y)$$

En el caso discreto la independencia nos indicaría que X, Y son independientes si

$$f_{ij} = f_{i\cdot} \cdot f_{\cdot j}$$

pero cada una de las relaciones siguientes expresa por si sola la independencia

$$\frac{n_{ij}}{n_{i\cdot}} = \frac{n_{\cdot j}}{n_{\cdot\cdot}} = \frac{n_{2j}}{n_{2\cdot}} = \cdots = \frac{n_{kj}}{n_{k\cdot}}$$

Ejemplo: 4.4.12

[] Supóngase que se toman dos medidas independientes X eY de lluvia durante un periodo de tiempo en una localidad y que la f.d.p g de cada medida es la siguiente

$$g(x) = \begin{cases} 2x & si \quad 0 \leq x \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

Se determinará el valor de $P(X + Y \leq 1)$

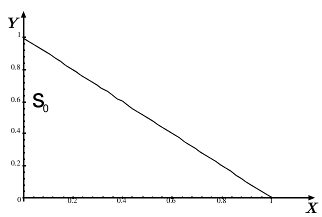
Solución

Puesto que X e Y son independientes y cada una tiene la f.d.p. g, resulta que para cualquier par de valores x e y la f.d.p. conjunta $f(x, y)$ de X e Y está dada por la relación $f(x, y) = g(x)\, g(y)$. Por tanto

$$f(x, y) = \begin{cases} 4xy & si \quad 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

El conjunto S del plano xy en el que $f(x, y) > 0$ y el subconjunto S_0 en el que $x + y \leq 1$ se encuentra representado en la figura 4.9. Por tanto

$$P(X + Y \leq 1) = \int_0^1 \int_0^{1-x} 4xydydx = \frac{1}{6}$$



4.4.2. Distribuciones discretas

4.4.2.1. Distribuciones condicionales discretas

Sean X e Y dos variables aleatorias que tienen distribución discreta conjunta cuya f.p. conjunta es f, definimos f_x, f_y como las f.p marginales de X e Y, respectivamente. Si observamos un valor y de la variable Y, la probabilidad de que la variable aleatoria X tome cualquier valor particular x, está dado por la siguiente probabilidad condicional

$$P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$
$$= \frac{f(x, y)}{f_y(y)}$$

A esta distribución se le denomina distribución condicional de X dado Y y se denota

$$f(x|y) = \begin{cases} \frac{f(x,y)}{f_y(y)} & si \quad f_y(y) > 0 \end{cases}$$

Análogamente se define la distribución de probabilidad condicional de Y dado X

$$f(y|x) = \left\{ \frac{f(x,y)}{f_x(x)} \quad si \quad f_x(x) > 0 \right.$$

La función de densidad conjunta se define

$$F(x|y) = P(x \geq X, Y = y)$$
para un valor fijo de y
$$F(y|x) = P(y \geq Y, X = x)$$
Para un valor fijo de x

Ejemplo: 4.4.13

[] Analice la independencia para los datos tabulados

$X \| Y$	$1 \in (0,2]$	$3 \in (2,4]$	$5 \in (4,6]$	*total*
0	24	4	8	36
1	6	1	2	9
2	12	2	4	18
total	42	7	14	63

4.4.2.2. Distribuciones condicionales continuas

Sean X, Y variables aleatorias continuas con densidad conjunta $f(x,y)$ y las densidades marginales $f_x(x), f_y(y)$, respectrivamente . Entonces la densidad condicional de X dado un $Y = y$ fijo

$$f(x|y) = \begin{cases} \frac{f(x,y)}{fy(y)} & si \quad f_y(y) > 0 \\ 0 & en \quad otro\, caso \end{cases}$$

De manera análoga

$$f(y|x) = \begin{cases} \frac{f(x,y)}{f_x(x)} & si \quad f_x(x) > 0 \\ 0 & en \quad otro\, caso \end{cases}$$

Ejemplo: 4.4.14

[] Supóngase que la f.p conjunta de X e Y es la dada por la tabla del ejemplo 4.7 . Se determinará la f.p condicional de Y dado $X = 2$

Solución

A partir de de la tabla $f_x(x) = 0{,}6$. Por tanto la probabilidad condicional f($y|x$) $= \frac{f(2,y)}{0{,}6}$ Por ejemplo si $Y = 1$ entonces f($1|2$) $= 0{,}5$

Ejemplo: 4.4.15

[] Sea la f.d.p conjunta de X e Y la del ejemplo 4.8 Determinar la f.d.p condicional de Y dado $X = x$

Solución

El conjunto S para el cual $f(x,y) > 0$ se observa en la figura (a) del ejemplo 4.8 . Además la f.d.p marginal se obtuvo en el ejemplo 4.11 y se representa en la fig 4.14 .Se puede observar a partir de esta figura que $f_x(x) > 0$ para $-1 < x < 1$, pero no para $x = 0$.

Por tanto, para cualquier valor concreto de x tal que $-1 < x < 0$ ó $0 < x < 1$, la f.d.p condicional f($y|x$) de Y es la siguiente

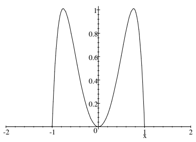


Fig 4.14

$$f(y|x) = \begin{cases} \frac{2y}{1-x^4} & si \quad x^2 \leq y \leq 1 \\ 0 & \quad y > 1 \end{cases}$$

En particular si X=0.5 entonces

$$P(Y \geq 0{,}75|X = 0{,}5)$$
$$= \int_{0{,}75}^{1} f(y|0{,}5)dy = \frac{7}{15}$$

4.4.3. Cambio de variable

4.4.3.1. Funciones de una variable con una distribución discreta

Sea X una variable aleatoria que tiene una f.p discreta f, y sea $Y = h(X)$ otra variable aleatoria definida como función de X, y sea g la f.p discreta de Y, entonces g se puede obtener a partir de f para cualquier valor y de Y, así

$$g(y) = P(Y = y) =$$
$$P(h(X) = y) = \sum_{\{x:h(x)=y\}} f(x)$$

4.4.3.2. Funciones de una variable con una distribución continua

Sea X una variable aleatoria que tiene distribución continua, con una f.d.p de X f, se define para otra variable aleatoria $Y = h(X)$ y para cualquier real y, la f.d $G(y)$ de Y así

$$G(y) = P(y \geq Y)$$
$$P(y \geq h(X)) = \int_{\{x:y=h(x)\}} f(x)\,dx$$

además si la variable Y tiene una distribución continua, su f.d.p se puede obtener

$$g(y) = \frac{dG(y)}{dy}$$

Lo anterior lo podemos formalizar de la siguiente manera

Proposición 6

ea X una variable aleatoria cualquiera. Si realizamos el cambio de variable $Y = h(X)$, tenemos una nueva variable aleatoria de modo que su f.d.p es f y P$(a < x < b) = 1$ y además supóngase que $r(x)$ es continua y estrictamente creciente o estrictamente decreciente en $a < X < b$ si y sólo si $\alpha < Y < \beta$ y sea $X = h^{-1}(Y)$ la función inversa de h para $\alpha < Y < \beta$, entonces la f.d.p de g está dada por la relación

$$g(y) = \begin{cases} g(y) = f\left(h^{-1}(y)\right)\left|\frac{dx}{dy}\right| & si \quad \alpha < Y < \beta \\ 0 & \text{en otro caso} \end{cases}$$

donde se tiene que

$$\frac{dx}{dy} = \frac{1}{h'(h^{-1}(y))}$$

En el caso que la aplicación no sea inyectiva, podemos tener para un y dado ciertos $x_1, x_2, x_3, \cdots, x_n$ tales que $f(x_i) = y$ En este caso

$$g(y) = \sum_{i=1}^{n} f(x_i) \left| \frac{dx_i}{dy} \right|$$

donde

$$\frac{dx_i}{dy} = \frac{1}{h'(x_i)}$$

Ejemplo: 4.4.16

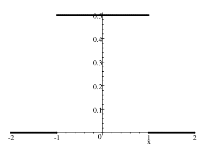
[] Sea X una variable aleatoria continua tal que $Y = X^2$en este caso $h(x) = x^2$ la cual no inyectiva pero si la restringimos a los reales positivos y los reales negativos tenemos que

$$g(y) = f(\sqrt{y}) \frac{1}{2\sqrt{y}} + f(-\sqrt{y}) \frac{1}{2\sqrt{y}}$$

Ejemplo: 4.4.17

[] Supóngase que X tiene una distribución f uniforme sobre el intervalo $(-1, 1)$, así que

$$f(x) = \begin{cases} 0{,}5 & si \quad -1 < x < 1 \\ 0 & si \quad x \geq 1, -1 \geq x \end{cases}$$



Distribución uniforme

Determine la f.d.p para la variable $Y = X^2$

Solución

Como $Y = X^2$ entonces Y debe pertenecer a $[0, 1]$, entonces para todo y $\in$ $(0, 1)$ la f.d G(y) de Y es

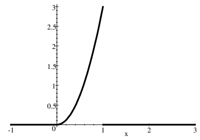
$$G(y) = P(y \geq Y)$$
$$P\left(y \geq X^2\right) = P\left(-y^{0,5} \leq X \leq y^{0,5}\right) =$$
$$\int_{-y^{0,5}}^{y^{0,5}} 0{,}5dx = \sqrt{y} \text{ para } y \in (0, 1)$$

Ahora la f.d.p $g(y)$ de Y es g$(y) = \frac{1}{2\sqrt{y}}$ para $y \in (0, 1)$

Ejemplo: 4.4.18

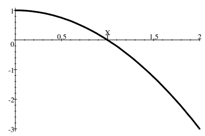
[] Supóngase que X es una v.a cuya f.d.p es

$$f(x) = \begin{cases} 3x^2 & si \quad 0 < x < 1 \\ 0 & si \quad x \geq 1, 0 \geq x \end{cases}$$



Ejemplo 4.18

Obtengase la f.d.p para $Y = 1 - X^2$



$y = 1 - x^2$

Solución

En este ejemplo observamos de la gráfica que $Y = 1 - X^2$es continua y estrictamente decreciente para $0 < x < 1$ y además

$$\int_0^1 3x^2 dx = 1$$

y si $X \in (0,1)$ entonces Y $\in (0,1)$ y $X = (1-Y)^{0,5}$ por tanto para y $\in (0,1)$ $f[h^{-1}(y)] = 3(1-y)$ y $\frac{dx}{dy} = -0{,}5(1-y)^{-0,5}$entonces $g(y) = 1{,}5(1-y)^{0,5}$ así

$$g(y) = \begin{cases} 1{,}5(1-y)^{0,5} & si \quad 0 < y < 1 \\ 0 & si \quad y \geq 1, 0 \geq y \end{cases}$$

4.4.3.3. Funciones de dos o más variables aleatorias

- Sean n variables aleatorias $X_1, X_2, X_3, \cdots, X_n$ las cuales tienen una distribución conjunta discreta cuya f.p conjunta es F y se definen m funciones $Y_1, Y_2, Y_3, \cdots, Y_m$ de estas n variables de la siguiente forma

$$\begin{array}{ccc} Y_1 & = & h_1(X_1, X_2, X_3, \cdots, X_n) \\ Y_2 & = & h_2(X_1, X_2, X_3, \cdots, X_n) \\ \vdots & \vdots & \vdots \\ Y_m & = & h_m(X_1, X_2, X_3, \cdots, X_n) \end{array}$$

para un valor $y = (y_1, y_2, y_3, \cdots, y_m)$ de las m variables $Y_1, Y_2, Y_3, \cdots, Y_m$. Sea

$$A = \{(x_1, x_2, x_3, \cdots, x_n): \\ y_1 = h_1(x_1, x_2, x_3, \cdots, x_n), \cdots, y_m = h_m(x_1, x_2, x_3, \cdots, x_n)\}$$

Entonces podemos determinar la f.p conjunta de $Y_1, Y_2, Y_3, \cdots, Y_m$ en el punto $(y_1, y_2, y_3, \cdots, y_m)$ como

$$g(y_1, y_2, y_3, \cdots, y_m) = \sum_{(x_1, x_2, x_3, \cdots, x_n) \in A} f(x_1, x_2, x_3, \cdots, x_n)$$

- Sean n variables aleatorias $X_1, X_2, X_3, \cdots, X_n$ las cuales tienen una distribución conjunta continua cuya f.d conjunta es F y se definen n funciones $Y_1, Y_2, Y_3, \cdots, Y_n$ de estas n variables de la siguiente forma

$$\begin{array}{ccc} Y_1 & = & h_1(X_1, X_2, X_3, \cdots, X_n) \\ Y_2 & = & h_2(X_1, X_2, X_3, \cdots, X_n) \\ \vdots & \vdots & \vdots \\ Y_n & = & h_n(X_1, X_2, X_3, \cdots, X_n) \end{array}$$

para un valor $y = (y_1, y_2, y_3, \cdots, y_n)$ de las n variables
$Y_1, Y_2, Y_3, \cdots, Y_n$. Sean S el conjunto

$$\{(x_1, x_2, x_3, \cdots, x_n) : P\left((X_1, X_2, X_3, \cdots, X_n) \in I\!R^n\right) = 1\}$$

$$T = \{(y_1, y_2, y_3, \cdots, y_n) : \\ (x_1, x_2, x_3, \cdots, x_n) = h_1^{-1}(y_1, y_2, y_3, \cdots, y_n)\}$$

Donde $(x_1, x_2, x_3, \cdots, x_n) \in S$ Entonces podemos determinar la f.d.p conjunta de $Y_1, Y_2, Y_3, \cdots, Y_n$ en el punto $(y_1, y_2, y_3, \cdots, y_n)$ como

$$g(y) = \begin{cases} f(h^{-1}(y_1, y_2, \cdots, y_n))\,|J|\,, & (y_1, y_2, \cdots, y_n) \in T \\ 0 & \text{en otro caso} \end{cases}$$

donde

$$J = \begin{vmatrix} \frac{\partial h_1^{-1}}{\partial y_1} & \cdots & \frac{\partial h_1^{-1}}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n^{-1}}{\partial y_1} & \cdots & \frac{\partial h_n^{-1}}{\partial y_n} \end{vmatrix}$$

Ejemplo: 4.4.19

[] Supóngase que dos variables aleatorias X_1 y X_2
tienen una distribución continua cuya f.d.p conjunta es

$$f(x_1, x_2) = \begin{cases} 4x_1x_2 & si & \begin{matrix} 0 < x_1 < 1, \\ 0 < x_2 < 1 \end{matrix} \\ 0 & & \text{en otro caso} \end{cases}$$

se determinará la f.d.p conjunta de dos nuevas variables aleatorias

$$Y_1 = \frac{X_1}{X_2} \tag{a}$$

$$Y_2 = X_1X_2 \tag{b}$$

Solución

Al resolver el sistema (a) y (b) obtenemos que

$$X_1 = (Y_1 Y_2)^{0,5}$$
$$X_2 = \left(\frac{Y_2}{Y_1}\right)^{0,5}$$

Sea $S = \{(x_1, x_2) : 0 < x_1 < 1, 0 < x_2 < 1\}$ entonces
$P((X_1, X_2) \in S) = 1$ y además

$$T = \{(y_1, y_2) : y_1 > 0, y_2 > 0, (y_1 y_2)^{0,5} < 1, \left(\frac{y_1}{y_2}\right)^{0,5} < 1\}$$

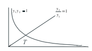
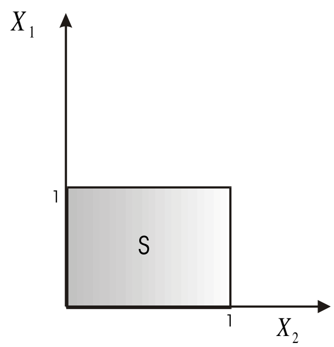


fig 4.19 a



ej 4.19 b

entonces

$$x_1 = h_1^{-1}(y_1, y_2) = (y_1, y_2)^{0,5}$$
$$x_2 = h_2^{-1}(y_1, y_2) = \left(\frac{y_1}{y_2}\right)^{0,5}$$

entonces

$$J = \begin{vmatrix} \frac{1}{2}\left(\frac{y_2}{y_1}\right)^{0,5} & \frac{1}{2}\left(\frac{y_1}{y_2}\right)^{0,5} \\ -\frac{1}{2}\left(\frac{y_2}{y_1^3}\right)^{0,5} & \frac{1}{2}\left(\frac{1}{y_1 y_2}\right)^{0,5} \end{vmatrix} = \frac{1}{2y_1}$$

por tanto

$$g(y_1, y_2) = \begin{cases} 2\left(\frac{y_2}{y_1}\right) & si \quad (y_1, y_2) \in T \\ 0 & \text{en otro caso} \end{cases}$$

4.5 Esperanza matemática o valor esperado

4.5.1. Esperanza de una variable discreta

Definición 4.11

Supóngase que una variable aleatoria X tiene una distribución de probabilidad discreta f. la esperanza matemática de X, se denota E(X) y se define

$$E\left(X\right)=\sum_{I}x_{i}f\left(x_{i}\right)\qquad\forall i\in I \tag{*}$$

Donde I es el conjunto numerable de índices de los valores que puede tomar la variable

Puede de darse el caso de que la serie $\sum_I x_i f\left(x_i\right)$ sea divergente Por lo que en estos casos se dice que la esperanza existe si, y sólo si, la suma de la ecuación (*) es absolutamente convergente es decir si, y sólo si,

$$\sum_{I}\left|x_{i}\right|f\left(x_{i}\right)<\infty \tag{**}$$

Entonces podemos decir que si se cumple (**) E(X) existe y su valor está dado por la ecuación (*) y si (**) no se cumple $E(X)$ no existe

Ejemplo: 4.5.20

[] A un contratista se le asigna el trabajo de determinar la media de los días que tarda un trabajador en terminar un trabajador después de realzar el estudio obtuvo los siguientes datos

x	f(x)
3	$\frac{1}{8}$
4	$\frac{5}{8}$
5	$\frac{2}{8}$
x>5	0

$$E\left(X\right)=3*\frac{1}{8}+4*\frac{5}{8}+5*\frac{2}{8}=\frac{33}{8}$$

4.5.2. Esperanza para una variable continua

Definición 4.12

Sea $f\left(x\right)$ la f.d.p de una variable aleatoria continua, entonces la esperanza matemática se define

$$E\left(X\right)=\int_{-\infty}^{\infty}xf\left(x\right)dx$$

Se dice que esta integral existe, siempre que existan números a y b tales que $\infty<a<b<\infty$ y $P\left(a\leq x\leq b\right)=1$ en este caso se dice que $E\left(X\right)$ existe

Ejemplo: 4.5.21

[] Supóngase que la f.d.p de una variable aleatoria continua es

$$f\left(x\right)=\begin{cases}cx^{2} & si \quad 0<x<1\\ 0 & si \quad x\geq 0\\ 0 & si \quad x\leq 0\end{cases}$$

Determine su esperanza

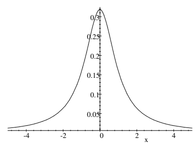
Solución

$E(X) = \int_0^1 cx^3 dx = \frac{1}{4}c$

Ejemplo: 4.5.22

[] Sea la f.d.p de una variable aleatoria continua

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad si \ -\infty < x < \infty$$



Distribución de Cauchy

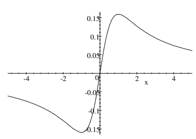
analicemos que

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)} dx = 1$$

pero

$$\int_{-\infty}^{\infty} |x|\, f(x)\,dx = \int_{0}^{\infty} \frac{x}{\pi(1+x^2)} dx = \infty \qquad (+)$$

De la gráfica se observa que la media debería ser cero, pero en la ecuación $(+)$ observamos que $E(X)$ no existe, por lo tanto esta distribución llama da de Cauchy no tiene media



$xf(x)$

4.5.3. Propiedades del valor esperado

Teorema 4.1

Si $Y = aX + b$ donde a y b son constantes entonces

$$E(Y) = aE(X) + b$$

donde a y b son constantes

Prueba

Sea X una v.a con f.d.p f continua entonces

$$E(Y) = E(aX+b) =$$
$$\int_{-\infty}^{\infty} (ax+b)\, f(x)\,dx = a\int_{-\infty}^{\infty} xf(x)\,dx + \int_{-\infty}^{\infty} bf(x)\,dx =$$
$$aE(X) + b$$

En el caso discreto la demostración es similar, pero usando sumatoria en vez de integral ■(q.e.d)

Teorema 4.2 S

$X_1, X_2, X_3, \cdots, X_n$ son variables aleatorias cuyas esperanzas $E(X_i), i = 1, 2, 3, \cdots, n$ existen, entonces

$$E(X_1 + X_2 + X_3 + \cdots + X_n) = \sum_{i=1}^{n} E(X_i)$$

Ejemplo: 4.5.23

[] Sea un experimento de Bernoulli con proporción de obtener un éxito es p, determinar la esperanza al repetir el experimento n veces

Solución

Sea $X_i \rightsquigarrow \mathbb{B}$ (esto indica que X es una variable de Bernoulli), entonces

$$X_i = \begin{cases} 1 & \text{si el resultado es éxito} \\ 0 & \text{si no} \end{cases}$$

por tanto

$$P(X_i = 1) = p$$
$$P(X_i) = 0) = 1 - p$$

entonces

$$E(X_i) = 1p + 0(1 - p) = p$$

Si el experimento se repite n veces tenemos que

$$E\left(\sum_{i=1}^{n} (X_i)\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np$$

Teorema 4.3

Si $X_1, X_2, X_3, \cdots, X_n$ son n variables aleatorias independientes cuyas esperanzas $\mathrm{E}(X_i), i = 1, 2, 3, \cdots, n$ existen, entonces

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i)$$

Ejemplo: 4.5.24

[] Sean X_1, X_2, X_3, variables aleatorias independientes tales que $E(X_i) = 2$ y $E\left(X_i^2\right) = 4$ para $i = 1, 2, 3$, Determine el valor de $E\left(X_1^2 (X_2 - 3X_3)\right)$

Solución

$$\begin{aligned} E\left(X_1^2 (X_2 - 3X_3)^2\right) &= E(X_1^2)E[(X_2 - 3X_3)^2] \\ &= 4E\left(X_2^2 - 6X_2X_3 + 9X_3^2\right) \\ &= 4(4 - 6 * 2 * 2 + 9 * 4) = 64 \end{aligned}$$

Proposición 7

Sea X una v.a cuya f.d.p es f, entonces la esperanza de cualquier función $h(X)$se puede determinar que

$$E\left(h\left(x\right)\right) = E\left(Y\right) = \int_{-\infty}^{\infty} yg\left(y\right) dy = \int_{-\infty}^{\infty} h\left(x\right) f\left(x\right) dx$$

Ejemplo: 4.5.25

[] Determine la esperanza para la $X^{0,5}$ si X tiene una f.d.p como lo determina el ejemplo 4.21

Solución

Del ejemplo 4.21 sabemos que
$E\left(X^{0,5}\right) = \int_0^1 x^{0,5}\left(cx^2\right) dx = 0{,}285\,71c$

Proposición 8

Sean $X_1, X_2, X_3, \cdots, X_n$ variables aleatorias con f.d.p conjunta si existe una relación tal que
$Y = h\left(X_1, X_2, X_3, \cdots, X_n\right)$ se tiene que
$E(Y) = \int \cdots \int_A h(X_1, \cdots, X_n) f(X_1, \cdots, X_n) dx_1 \cdots dx_n,\ A \subset \mathbb{R}^n$

4.6 Varianza

Definición 4.13

Sea X una variable aleatoria con esperanza $\mu = E\left(X\right)$. la varianza de X se denotará $V(X)$ o σ^2 y se define

$$V\left(X\right) = E[(X - E\left(X\right))^2]$$
$$V\left(X\right) = \sum_{i \in I} \left(x_i - \mu\right)^2 f\left(x_i\right) \text{ X es discreta}$$
$$V\left(X\right) = \int_{-\infty}^{\infty} \left(x - \mu\right)^2 f\left(x\right) dx \text{ X es continua}$$

4.6.1. Propiedades de la varianza

- Si b es una constante $V\left(b\right) = 0$ si y solo si $P\left(X = b\right) = 1$

Prueba

V$(b) = E[(b - E\left(b\right))^2] = E[(b - b)^2] = 0$ ■(q.e.d)

- Sea X una v.a y sea a, b contantes entonces

$$V\left(aX + b\right) = a^2 X$$

Prueba

$V\left(aX + b\right) = E[((aX + b) - E\left(aX + b\right))^2] =$
$E[((aX + b) - (a\mu + b))^2] = E[(aX - a\mu)^2]$

$a^2E[(X-\mu)^2] = a^2V(X)$ ■(q.e.d)

- Para cualquier variable aleatoria X,

$$V(X) = E\left(X^2\right) - (E(X))^2$$

- Sea $X_1, X_2, X_3, \cdots, X_n$ variables aleatorias independientes entonces

$$V\left(\sum_{i=1}^{n}(X_i)\right) = \sum_{i=1}^{n} V(X_i)$$

Ejemplo: 4.6.26

[] Consideremos una variable aleatoria discreta con función de probabilidad

$$f(x) = \begin{cases} \frac{c}{4^x} & si \quad x = 1, 2, 3, ... \\ 0 & \text{si no} \end{cases}$$

Obtener

1. El valor de la constante c para que f sea una función de probabiliad
2. calcular $P(X=3)$ y $P(3 \geq X)$
3. Calculese la esperanza y la varianza

Solución

1. $\sum_{x=1}^{\infty} \frac{c}{4^x} = \frac{\frac{1}{4}}{1-\frac{1}{4}}c = \frac{c}{3} = 1 \Longrightarrow c = 3$
   Luego la f.p es
   $$f(x) = \begin{cases} \frac{3}{4^x} & si \quad x = 1, 2, 3, ... \\ 0 & \text{si no} \end{cases}$$
2. $P(X=3) = \frac{3}{4^3} = \frac{3}{64}$ y $P(3 \geq X) = \sum_{x=1}^{3} \frac{3}{4^x} = 0{,}987$
3. $E(X) = \sum_{x=1}^{\infty} \frac{3x}{4^x} = \frac{4}{3}, V(X) = \sum_{x=1}^{\infty} \frac{3(x-\frac{4}{3})^2}{4^x} = \frac{4}{9}$

Solución

Sea X una variable aleatoria con f.d

$$f(x) = \begin{cases} 1 & si \quad 0 \leq x \leq 1 \\ 0 & si \quad x > 1, x < 0 \end{cases}$$

Determinar
1. La media
2. La varianza

1. $E(X) = \int_0^1 x dx = \frac{1}{2}$
2. $V(X) = \int_0^1 (x - 0{,}5)^2 dx = 8.333\,3 \times 10^{-2}$

Ejemplo: 4.6.27

[] Determine la varianza para la variable de Bernoulli definida en el ejemplo 4.23

Solución

tenemos que $V(X) = E\left(X^2\right) - E^2(X)$ por lo que nos que determinar

$$E\left(X_i^2\right) = \sum_{j=1}^{2} x_j^2 f(x_j) = p$$
$$V(X_i) = p - p^2 = p(1-p)$$
$$\sum_{i=1}^{n} V(X_i) = npq$$

Proposición 9

uando estudiamos tablas de doble entrada podemos determinar las medias y varianzas marginales

$$\overline{\overline{x}} = \sum_{i=1}^{m} f_{i\cdot} x_i$$
$$\overline{\overline{y}} = \sum_{j=1}^{k} f_{\cdot j} y_j$$
$$V_X = \sum_{i=1}^{m} f_{i\cdot}(x_i - \overline{\overline{x}})^2$$
$$V_Y = \sum_{j=1}^{k} f_{\cdot j}(y_j - \overline{\overline{y}})^2$$

Proposición 10

uando estudiamos problemas con tablas de doble entrada podemos establecer las medias y varianzas marginales así

$$\overline{\overline{x}} = \frac{1}{n_{\cdot\cdot}} \sum_{i=1}^{m} n_{i\cdot} x_i$$
$$\overline{\overline{y}} = \frac{1}{n_{\cdot\cdot}} \sum_{j=1}^{k} n_{\cdot j} y_j$$
$$V_X = \frac{1}{n_{\cdot\cdot}} \sum_{i=1}^{m} n_{i\cdot}(x_i - \overline{\overline{x}})^2$$
$$V_Y = \frac{1}{n_{\cdot\cdot}} \sum_{j=1}^{k} n_{\cdot j}(y_j - \overline{\overline{y}})^2$$

4.7 Covarianza y correlación

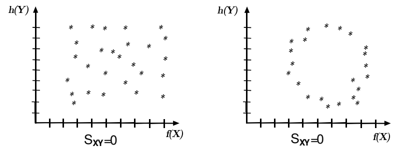


Figura 4.4: fig b

Definición 4.14

[Covarianza]Sean X e Y variables aleatorias que tienen una distribución conjunta y además esperanzas y varianzas $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$ respectivamente . La covarianza de X e Y se denota $cov\,(X,Y)$, y se define

$$cov\,(X,Y) = E[(X-\mu_X)\,(Y-\mu_Y)]$$

$$cov\,(X,Y) = \sum_{i=1}^{m}\sum_{j=1}^{k} f_{ij}x_iy_j - \overline{\overline{x}}\,\overline{\overline{y}}$$

Se dice que esta esperanza existe si $\sigma_X^2 < \infty, \sigma_Y^2 < \infty$
La covarianza puede ser negativa, positiva o cero

Definición 4.15

[Correlación]Si la $cov\,(X,Y)$ existe se denota el coeficiente de correlación $\rho\,(X,Y)$ el cual se define

$$\rho\,(X,Y) = \frac{cov\,(X,Y)}{\sigma_X^2\sigma_Y^2} \qquad si\,\sigma_Y^2, \sigma_X^2 \neq 0$$

4.7.1. Interpretación geométrica de la covarianza

Si consideramos una nube de puntos formados por las parejas de los datos concretos de dos v.a X e Y (x_i, y_j) el centro de gravedad de esta nube de puntos es $\left(\overline{\overline{x}}, \overline{\overline{y}}\right)$, ahora si trasladamos los ejes de tal forma que este punto sea el centro, la nube queda dividida en cuatro cuadrantes los que indica que los puntos que se encuentran en el primer y tercer cuadrante contribuyen positivamente al valor de la covarianza y los que se encuentran en los otros dos cuadrantes contribuyen negativamente. como lo indica la figura.a y si los puntos se reparten con igual proporción la covarianza será nula como en la fig b

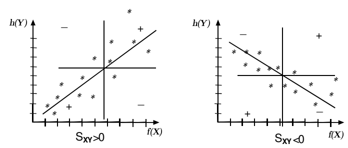


fig a

como indica la fig. b

4.7.2. Interpretación geométrica de la correlación

Es el coseno del ángulo formado por los vectores de las desviaciones con respecto a las medias de X e Y es decir

$$\cos\theta = \frac{\left(\overrightarrow{X} - \overrightarrow{\overline{x}}\right)\left(\overrightarrow{Y} - \overrightarrow{\overline{Y}}\right).}{\left\|\overrightarrow{X} - \overrightarrow{\overline{x}}\right\|\left\|\overrightarrow{Y} - \overrightarrow{\overline{y}}\right\|}$$

Es decir este coeficiente determina la relación lineal entre las desviaciones de las variables y esta dependencia es perfecta cuando $\rho = \pm 1$

4.7.3. Propiedades

1. $-1 < \rho < 1$
2. Para cualesquiera variables X e Y tales que la covarianza exista

   (a).
   $$cov(X,Y) = E(XY) - E(X)E(Y)$$

   (b). Si las variables X e Y son independientes
   $$cov(X,Y) = \rho(X,Y) = 0$$

   (c). $V(X+Y) = V(X) + V(Y) + 2cov(X,Y)$

   (d). $V(X-Y) = V(X) + V(Y) - 2cov(X,Y)$

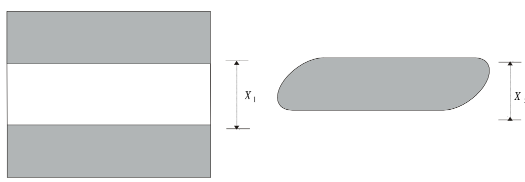
   (e). Si $Y = aX + b$, entonces $cov(X,Y) = a\sigma_X^2$

Ejemplo: 4.7.28

[] Dos piezas, como indica la figura ej.4.29, se pueden ensamblar. Si consideramos que las cotas X_1, X_2 son v.a. independientes y además consideramos que el espacio que queda libre entre las dos es $Y = X_1 - X_2$, donde la función de distribución conjunta para X_1, X_2 es

$$f(x_1,x_2) = \begin{cases} 8\exp(-2x_1 + x_2) & si \quad x_1 \geq 0, x_2 \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

Determine la media $E(Y)$ y la varianza $V(Y)$, si $E(X_1) = 0{,}5$,
$E(X_2) = 0{,}25$



Ej 4.28

Solución

Para determinar la media tenemos que

$$E(Y) = E(X_1 - X_2) = E(X_1) - E(X_2) =$$
$$0{,}5 - 0{,}25 = 0{,}25$$

Para determinar la varianza de Y primero debemos determinar V_{X_1} y V_{X_2}.
Como las variables X_1, X_2 son independientes entonces

$$f(x_1,x_2) = f_{x_1}(x_1)\, f_{x_2}(x_2)$$

además f es fácil de Factoriza entonces

$$f(x_1,x_2) = f_{x_1}(x_1)\, f_{x_2}(x_2)$$
$$[c_1 \exp(-(2x_1))]c_2 \exp(-(4x_2))]$$

y tenemos que

$$E(X_1) = \int_0^\infty c_1 x_1 \exp(-(2x_1))\, dx_1 = \frac{1}{4}c_1 = \frac{1}{2}$$
$$entonces \qquad c_1 = 2 \quad \text{como} \quad c_1 c_2 = 8 \Longrightarrow c_2 = 4$$

Por lo tanto

$$
\begin{aligned}
V(X_1) &= E\left(X_1^2\right) - E^2(X_1) = \\
&= \int_0^\infty 2x_1^2 \exp\left(-(2x_1)\right) dx_1 - \left(\frac{1}{2}\right)^2 = \frac{1}{4} \\
V(X_2) &= \int_0^\infty 4x_2^2 \exp\left(-(2x_2)\right) dx_2 - \left(\frac{1}{4}\right)^2 = \frac{15}{16}
\end{aligned}
$$

Ahora tenemos que X_1 Y X_2 son independientes entonces

$$
\begin{aligned}
V(Y) &= V(X_1) + V(X_2) \\
&= \frac{1}{4} + \frac{15}{16} = \frac{19}{16}
\end{aligned}
$$

Ejemplo: 4.7.29

[] Suponga que X_1 y X_2 son calificaciones codificadas en dos pruebas de inteligencia, y la f.d.p conjunta está dada por

$$
f(x_1, x_2) = \begin{cases} 6x_1^2 x_2 & si \quad 1 \geq x_1 \geq 0, 1 \geq x_2 \geq 0 \\ 0 & en \text{ otro caso} \end{cases}
$$

Encuentre $E(X_2|X_1)$, $E(X_1|X_2)$ y $V(X_1|X_2)$

Solución

La región S está representada en la fig 4.19 . Ahora determinaremos f_{x_1} y f_{x_2}

$$
\begin{aligned}
f_{x_1}(x_1) &= \int_0^1 6x_1^2 x_2 dx_2 = 3x_1^2 \\
f_{x_2}(x_2) &= \int_0^1 6x_1^2 x_2 dx_1 = 2x_2
\end{aligned}
$$

entonces

$$
f(x_1|x_2) = \begin{cases} 3x_1^2 & si \quad 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1 \\ 0 & en \text{ otro caso} \end{cases}
$$

y

$$
f(x_2|x_1) = \begin{cases} 2x_2 & si \quad 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1 \\ 0 & en \text{ otro caso} \end{cases}
$$

Por lo cual

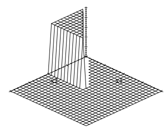
$$
\begin{aligned}
E(X_2|X_1) &= \int_0^1 2x_2^2 dx_2 = \frac{2}{3} \\
E(X_1|X_2) &= \int_0^1 3x_1^3 dx_1 = \frac{3}{4} \\
V(X_1|X_2) &= \int_0^1 3x_1^4 dx_1 - \left(\frac{3}{4}\right)^2 = \frac{3}{80}
\end{aligned}
$$

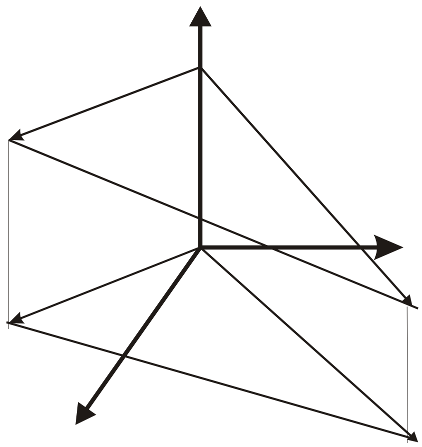
Ejemplo: 4.7.30

[] Sean X_1 y X_2 variables aleatorias que tienen f.d.p conjunta

$$f(x_1, x_2) = \begin{cases} 1 & if \quad -x_2 < x_1 < x_2, 0 < x_2 < 1 \\ 0 & if \quad -x_2 \geq x_1, x_1 \geq x_2, 0 \geq x_2, x_2 \geq 1 \end{cases}$$

determinar $\text{cov}(X_1, X_2)$





Hallemos f_{X_1} y f_{X_2}

$$f(x_1) = \begin{cases} 1 - x_1 & si \quad 0 < x_1 < 1 \\ 1 + x_1 & si \quad -1 < x_1 < 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$f(x_2) = \begin{cases} 2x_2 & si \quad 0 < x_2 < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Por tanto

$$E(X_1) = \int_0^1 x_1(1 - x_1)\, dx_1 + \int_{-1}^0 x_1(1 + x_1)\, dx_1 = 0$$

$$E(X_2) = \int_0^1 x_2(2x_2)\, dx_2 = \frac{2}{3}$$

$$cov(X_1, X_2) = \int_0^1 \int_{-x_2}^{x_2} x_1 x_2 dx_1 dx_2 - 0 = 0$$

4.8 Problemas

1. Supóngase que una variable aleatoria X tiene una distribución discreta con la siguiente f.p.:

$$f(x) = \begin{cases} cx & para \quad x = 1,2,3,4,5 \\ 0 & \text{en otro caso} \end{cases}$$

Determínese el valor de la constante c.

2. Supóngase que se lanzan dos dados equilibrados y sea X el valor absoluto de la diferencia entre los dos números que aparecen. Determínese y represéntese la f.p. de X.

3. Supóngase que se realizan 10 lanzamientos independientes de una moneda equilibrada. Determínese la f.p. del número de caras que se obtienen.

4. Supóngase que una urna contiene 7 bolas rojas y 3 azules. Si se seleccionan 5 bolas aleatoriamente, sin reemplazo, determínese la f.p. del número de bolas rojas que se obtienen.

5. Supóngase que una variable aleatoria X tiene una distribución binomial con parámetros $n = 15$ y $p = 0{,}5$. Calcúlese $P(X < 6)$.

6. Supóngase que una variable aleatoria X tiene una distribución binomial con parámetros $n = 8$ y $p = 0{,}7$. Calcúlese la $P(X > 5)$ utilizando la tabla que se encuentra al final del libro. Sugerencia: Utilícese el hecho de que $P(X \geq 5) = P(Y \leq 3)$, donde Y tiene una distribución binomial con parámetros $n = 8$ y $p = 0{,}3$.

7. SÍ el 10 % de las bolas de una urna son rojas y se selecciona al azar y con reemplazo el 20 %, ¿cuál es la probabilidad de que se obtengan más de tres bolas rojas?

8. Supóngase que una variable aleatoria X tiene una distribución discreta con la siguiente f.p.:

$$f(x) = \begin{cases} \frac{c}{x^2} & para \quad x = 1,2,\cdots, \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre el valor de la constante c.

9. Demuéstrese que no existe un número c tal que la siguiente función sea una f.p.:

$$f(x) = \begin{cases} \frac{c}{x} & para \quad x = 1,2,\cdots, \\ 0 & \text{en otro caso} \end{cases}$$

10. Una variable aleatoria discreta X tiene la función de probabilidad:

$$f(x) = \begin{cases} k\left(\frac{1}{2}\right)^x & para \quad x = 1,2,3 \\ 0 & \text{en otro caso} \end{cases}$$

   a) Determine el valor de k.

   b) Encuentre la función de distribución acumulativa, $F(x)$.

11. La variable aleatoria discreta $N(N = 0, 1, ...)$ tiene probabilidades de ocurrencia de $kr^n(0 < r < 1)$. Encuentre el valor apropiado de k.

12. El servicio postal requiere, en promedio, 2 días para entregar una carta en una ciudad. La varianza se estima como $.4(día)^2$. Si un ejecutivo desea que el 99 por ciento de sus cartas se entreguen a tiempo, ¿con cuánta anticipación las debe depositar en el correo?

13. Dos agentes de bienes raíces, A y B, tienen lotes de terrenos que se ofrecen en venta. Las distribuciones de probabilidad de los precios de venta por lote se muestran en la siguiente tabla.

	Precio					
\$	1000	1050	1100	1150	1200	1350
A	0.2	0.3	0.1	0.3	0.05	0.05
B	0.1	0.1	0.3	0.3	0.1	0.1

Suponiendo que A y B trabajan en forma independiente, calcule

   a) El precio de venta esperado de A y de B.

   b) El precio de venta esperado de A dado que el precio de venta de B es \$1,150.

   c) La probabilidad de que tanto A como B tengan el mismo precio de venta.

14. Demuestre que la función de probabilidad para la suma de los valores obtenidos al lanzar dos dados puede escribirse como

$$f(x) = \begin{cases} \frac{x-1}{36} & si \quad x = 2,3,\cdots,6 \\ \frac{13-x}{36} & si \quad x = 7,8,\cdots,12 \\ 0 & \text{en otro caso} \end{cases}$$

Determine la media y la varianza de X.

15. Una organización de consumidores que evalúa automóviles nuevos reporta regularmente el número de defectos importantes en cada automóvil examinado. Detonemos por X el número de defectos importantes en un automóvil seleccionado al azar de un cierto tipo. La f.d de X es como sigue

$$F(x) = \begin{cases} 0 & si & x < 0 \\ 0{,}06 & si & 0 \leq x < 1 \\ 0{,}19 & si & 1 \leq x < 2 \\ 0{,}39 & si & 2 \leq x < 3 \\ 0{,}67 & si & 3 \leq x < 4 \\ 0{,}92 & si & 4 \leq x < 5 \\ 0{,}97 & si & 5 \leq x < 6 \\ 1 & si & x \geq 6 \end{cases}$$

Calcule las siguientes probabilidades directamente de la fd.

   a) $P(X = 2)$

   b) $P(X > 3)$

   c) $P(2 < X < 5)$

   d) $P(2 \leq X \leq 5)$

16. Una compañía de seguros ofrece a sus tenedores de póliza varias opciones diferentes para el pago de primas. Para un tenedor seleccionado al azar, sea X = número de meses entre pagos sucesivos. La f.d de X es como sigue:

$$F(x) = \begin{cases} 0 & , & x < 1 \\ 0{,}30 & , & 1 \leq x < 3 \\ 0{,}40 & , & 3 \leq x < 4 \\ 0{,}45 & , & 4 \leq x < 6 \\ 0{,}60 & , & 6 \leq x < 12 \\ 1 & , & x \geq 12 \end{cases}$$

   a) ¿Cuál es la f.d.p de X?

   b) Con el solo uso de la f.d, calcule $P(3 < X < 6)$ y $P(4 \geq X)$.

17. El voltaje de una batería nueva puede ser aceptable (A) o no aceptable (I). Cierta linterna de mano necesita dos baterías, así que éstas han de seleccionarse y probarse independientemente hasta encontrar dos aceptables. Supongamos que el 80 % de todas las baterías tiene voltaje aceptable y denotemos por Y el número de baterías que deben ser probadas.

    a) ¿Cuál $P(Y = 2)$?

    b) ¿Cuál es $P(Y = 3)$? (SUGERENCIA: Hay dos resultados diferentes que resultan en Y= 3.)

    c) Para tener $Y = 5$, ¿qué debe ser cierto de la quinta batería seleccionada? Haga una lista de cuatro resultados para los que $Y = 5$ y luego determine $f(5)$.

    d) Utilice el lector el modelo de sus respuestas para las partes de la (a) a la (c) para obtener una fórmula general para $f(y)$.

18. Dos dados no cargados de seis caras se tiran independientemente. Sea $M =$ el máximo de los dos tiros (así que $M(1,5) = 5$,$M(3{,}3) = 3$, etcétera).

    a) ¿Cuál es la f.d.p de M?[SUGERENCIA: Primero determine $f(1)$, luego $f(2)$, etcétera.]

    b) Determine la f.d de M y grafiquela.

19. Una biblioteca se suscribe a dos revistas semanales de noticias, cada una de las cuales se supone que llega por correo el miércoles. En realidad, cada una puede llegar el miércoles, jueves, viernes o sábado. Suponga que las dos llegan independientemente una de la otra y para cada una P(mié.) =0 .4, P(jue.) = 0.3, P(vie.) = 0.2 y P(sáb.) = 0. 1.Sea Y= el número de días después del miércoles que tardan ambas revistas en llegar (por lo que los posibles valores deˆson 0, 1,2 o 3). Calcule la f.d.p de Y. (SUGERENCIA: Hay 16 resultados posibles; Y(M, M) = 0, Y(V. J) = 2, etcétera.]

20. La distribución de probabilidad deˆ, el número de defectos por cada 10 metros de una tela sintética en rollos continuos de ancho uniforme, es

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

    Dibuje la distribución acumulada de X.

21. Una firma de inversiones ofrece a sus clientes bonos municipales que vencen después de diferente número de años. Dada la distribución acumulada de T, el número de años para el vencimiento de un bono seleccionado aleatoriamente, es

$$F(t) = \begin{cases} 0 & , \quad y < 1 \\ \frac{1}{4} & , \quad 1 \le t < 3 \\ \frac{1}{2} & , \quad 3 \le t < 5 \\ \frac{3}{4} & , \quad 5 \le t < 7 \\ 1 & , \quad t \ge 7 \end{cases}$$

    Encuentre

    a) $P(T = 5)$

    b) $P(T > 3)$

    c) $P(1{,}4 < T < 6)$.

22. Una variable aleatoria continua X que puede asumir valores entre $x = 2$ y $x = 3$ tiene una función de densidad $f(x) = 1/2$

    a) Demuestre que el área bajo la curva es igual a 1.

    b) Encuentre $P(2 < X < 2{,}5)$

    c) Encuentre $P(X \le 1{,}6)$.

23. Una variable aleatoria continua X que puede tomar valores entre $x = 2$ y $x = 5$ tiene una función de densidad $f(x) = 2(1 + x)/27$. Encuentre

    a) $P(X < 4)$

    b) $P(3 < X < 4)$.

24. La proporción de personas que contestan una cierta encuesta enviada por correo es una variable aleatoria continua .y que tiene la f.d.p

$$f(x) = \begin{cases} \frac{2(x+2)}{5} & , \quad 0 < x < 1 \\ 0 & \quad \text{en otro caso} \end{cases}$$

    a) Demuestre que $P(0 < X < 1) = 1$

    b) Encuentre la probabilidad de que más de 0.25 pero menos de 0.5 de las personas en contacto responderán a este tipo de encuesta.

25. De una caja que contiene 4 monedas de 1000 pesos y 2 de 500, se seleccionan 3 de ellas al azar sin reemplazo. Determine la distribución de probabilidad para el total T de las 3 monedas. Exprese gráficamente la distribución de probabilidad como un histograma.

26. De una caja que contiene 4 pelotas negras y 2 verdes, se seleccionan 3 de ellas en sucesión con reemplazo. Encuentre la distribución de probabilidad para el número de pelotas verdes.

27. Encuentre la distribución de probabilidad para el número de discos de jazz cuando 4 discos se seleccionan al azar de una colección que consiste de 5 discos de jazz, 2 de música clásica y 3 de polka. Exprese el resultado por medio de una fórmula.

28. Encuentre una fórmula para la distribución de probabilidad de la variable aleatoria X que representa el resultado de un solo lanzamiento de un dado.

29. Un embarque de 7 televisores contiene 2 aparatos defectuosos. Un hotel realiza una compra aleatoria de 3 de ellos. Si X es el número de unidades defectuosas que se compran, encuentre la distribución de probabilidad de X. Expres e los resultados gráficamente como un histograma de probabilidad.

30. De un paquete de cartas se sacan tres en sucesión sin reemplazo. Encuentre la distribución de probabilidad para el número de cartas de espadas.

31. Suponga que X y Y tienen la siguiente distribución de probabilidad conjunta:

$f(x,y)$		$x=2$	$x=4$
y	1	0.10	0.13
	2	0.20	0.30
	3	0.10	0.15

    a) Encuentre la distribución marginal de X

    b) Encuentre la distribución marginal de Y

32. Dada la función de densidad conjunta

$$f(x,y) = \begin{cases} \frac{6-x-y}{8}, & 0 < x < 2, \quad 2 < y < 4 \\ 0 & \quad \text{en otro caso} \end{cases}$$

    Encuentre $P(1 < Y < 3|X = 2)$

33. Si X y Y representan las duraciones, en años, de dos componentes en un sistema electrónico. Si la función de densidad conjunta de estas variables es

$$f(x,y) = \begin{cases} e^{-(x+y)} & x > 0, \quad y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre $P(0 < X < 1|Y = 2)$.

34. Determine si las dos variables aleatorias del , ejercicio 31 son dependientes o independientes.

35. La cantidad de kerosene, en miles de litros, en un tanque al principio del día es una cantidad aleatoria Y, de la cual una cantidad aleatoria X se vende durante el día. Suponga que el tanque no se rellena durante el día, de tal forma que $x \leq y$, y suponga que la función de densidad conjunta de estas variables es

$$f(x,y) = \begin{cases} 2, & 0 < x < y, \quad 0 < y < 1 \\ 0, & \text{en otro caso} \end{cases}$$

a) Determine si X y Y son independientes.

b) Encuentre $P(1/4 < X < 1/2|Y = 3/4)$.

36. Una vinatería cuenta con instalaciones para atender a clientes que llegan en automóvil y a quienes llegan caminando. En un día seleccionado aleatoriamente, sean X y Y, respectivamente, los periodos de tiempo que se utilizan para cada caso y suponga que la función de densidad conjunta para estas dos variables aleatorias es

$$f(x,y) = \begin{cases} \frac{2}{3}(x+2y), & 0 < x < 1, \quad 0 < y < 1 \\ 0, & \text{en otro caso} \end{cases}$$

a) Encuentre la densidad marginal de X

b) Encuentre la densidad marginal de Y

c) Encuentre la probabilidad de que las instalaciones para quienes lleguen en automóvil se utilicen menos de la mitad del tiempo.

37. Una compañía dulcera distribuye cajas de chocolates con una mezcla de tres tipos de chocolate: cremas, de chiclosos y envinados. Suponga que el peso de cada caja es de 1 kilogramo, pero los pesos individuales de las cremas, de los chiclosos y de los envinados varían de una caja a otra. Para una caja seleccionada aleatoriamente, X y Y representan los pesos de las cremas y de los chiclosos, respectivamente, y suponga que la función de densidad conjunta de estas variables es

$$f(x,y) = \begin{cases} 24xy, & 0 < x < 1, \ x + y \leq 1, \quad 0 < y < 1 \\ 0, & \text{en otro caso} \end{cases}$$

a) Encuentre la probabilidad de que en una caja determinada el peso de los chocolates envinados sea más de 0.5 del peso total

b) Encuentre la densidad marginal para el peso de las cremas.

c) Encuentre la probabilidad de que el peso de los chocolates de chiclosos en una caja sea menos de 1/8 de kilogramo, si se sabe que las cremas constituyen 3/4 del peso.

38. Un distribuidor de aparatos electrodomésticos vende tres modelos diferentes de congeladores verticales con capacidad de 13.5, 15.9 y 19.1 pies cúbicos de espacio de almacenaje, respectivamente. Sea X = la cantidad de espacio de almacenaje comprado por el siguiente cliente que va a comprar un congelador. Supongamos que X tiene f.d.p

x	13.5	15.9	19.1
$f(x)$	0.2	0.5	0.3

a) Calcule $E(X)$, $E(X^2)$ y

b) Si el precio de un congelador que tiene una capacidad de X pies cúbicos es $25X - 8{,}5$, ¿cuál es el precio esperado pagado por el siguiente cliente que va a comprar un congelador?

c) ¿Cuál es la varianza del precio $25X - 8{,}5$ pagado por el siguiente cliente?

d) Suponga que mientras la capacidad nominal de un congelador es X, la capacidad real es $h(X) = X - 0{,}01X^2$. ¿Cuál es la capacidad real esperada del congelador comprado por el siguiente comprador?

39. Suponga que el número de plantas de un tipo particular que se encuentra en una región rectangular (llamada cuadrante por ecologistas) de cierta región geográfica es una v.a. X con f.d.p

$$f(x) = \begin{cases} \frac{c}{x^3} & , \quad x = 1, 2, 3, \cdots \\ 0 & , \quad \text{en otro caso} \end{cases}$$

¿Es $E(X)$ finita? Justifique su respuesta (ésta es otra distribución a la que los expertos en estadística llamarían de cola gruesa).

40. Una pequeña farmacia solicita ejemplares de una revista de noticias para su estante cada semana. Sea X = demanda de la revista, con f.d.p

x	1	2	3	4	5	6
$f(x)$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{2}{15}$

Suponga que el propietario de la farmacia paga en realidad \$0.25 por cada ejemplar de la revista de” precio a clientes es de \$1.00. Si las revistas que quedan el fin de semana no tienen valor de recuperación, ¿es mejor pedir tres o cuatro ejemplares de la revista? (SUGERENCIA: Para tres y cuatro ejemplares solicitados, exprese el ingreso neto como función de la demanda X y luego calcule el ingreso esperado.)

41. Sea X el daño incurrido (en \$) en cierto tipo de accidente durante un año dado. Los posibles valores de X son 0, 1 000, 5 000 y 10 000, con probabilidades 0.8, 0.1, 0.08 y 0.02 respectivamente. Una compañía particular ofrece una póliza deducible de \$500. Si la compañía desea que su utilidad esperada sea de \$100, ¿qué prima debe cobrar?.

42. Los n candidatos para un trabajo han sido clasificados como 1, 2, 3,. .., n. Sea X = el grado de un candidato seleccionado al azar, de modo que$X$ tiene una f.d.p de

$$f(x) = \begin{cases} \frac{1}{n} & , \quad 1, 2, 3, \cdots, n \\ 0 & \quad \text{en otro caso} \end{cases}$$

(ésta recibe el nombre de distribución discreta uniforma). Calcule $E(X)$ y $V(X)$. [SUGERENCIA: La suma de los primeros n enteros positivos es $n(n+1)/2$, mientras que la suma de su cuadrados es $n(n+1)(2n+1)/6$.]

43. Sea X = resultado cuando un dado no cargado se lanza una vez Si antes de hacer rodar el dado se ofrece al tirador ya sea (1/3.5) dólares o $h(X) = 1/X$ dólares, ¿aceptaría la cantidad garantizada o jugaría? [NOTA: No es generalmente cierto que $1/E(X) = E(1/X)$.]

44. Una compañía proveedora de productos químicos tiene actualmente en existencia 100 libras desierto producto que vende a clientes en lotes de 5 libras. Sea X = número de lotes ordenados por un cliente seleccionado al azar, y suponga que X tiene una f.d.p de:

x	1	2	3	4
$f(x)$	0.2	0.4	0.3	0.1

Calcule $E(X)$ y $V(X)$. Luego calcule el número esperado de libras sobrantes después de embarcar el pedido del siguiente cliente, y la varianza del número de libras restantes. (SUGERENCIA: El número de libras restantes es una función lineal de X.]

45. Considere la f.d.p para tiempo total de espera Y de dos autobuses es

$$f(y) = \begin{array}{lcl} \frac{1}{25}y & , & 0 \leq y < 5 \\ \frac{2}{5} - \frac{1}{25}y & , & 5 \leq y \leq 25 \\ 0 & , & \text{en otro caso} \end{array}$$

a) Calcule y trace la f.d. de Y. [SUGERENCIA: Considere separadamente $0 \leq y < 5$ y$5 \leq y \leq 10$ al calcular $F(y)$. Una figura de la f.d.p podría ser útil.]

b) Obtenga una expresión para el $(100p)$avo percentil. (SUGERENCIA: Considere separadamente $0 < p < 0{,}5$ y $0{,}5 << 1$.)

c) Calcule $E(Y)$ y $V(Y)$. ¿Cómo se comparan éstas con el tiempo de espera y varianza supuestos para un solo autobús cuando el tiempo es uniformemente distribuido en [0, 5]?

46. Un ecologista desea marcar una región circular de muestreo de 10 m de radio. Sin embargo, el radio de la región resultante es en realidad una variable aleatoria R con f.p.d

$$f(r) = \left\{ \begin{array}{lcl} \frac{3}{4}\left[1-(10-r)^2\right] & , & 9 \leq r \leq 11 \\ 0 & , & \text{en otro caso} \end{array} \right|$$

¿Cuál es el área esperada de la región circular resultante?

47. La demanda semanal de gas propano (en miles de galones) de una distribuidora en particular es una v.a. X con f.d.p

$$f(x) = \left\{ \begin{array}{lcl} 2\left(1-\frac{1}{x^2}\right) & , & 1 \leq x \leq 2 \\ 0 & , & \text{en otro caso} \end{array} \right.$$

a) Calcule la f.d de X.

b) Obtenga una expresión para el $(100p)$avo percentil ¿Cuál es el valor de $\tilde{X}$?

c) Calcule $E(X)$ y $V(X)$.

d) Si 1500 galones están en existencia al principio de semana y no se recibe nuevo suministro durante la semana, ¿cuánto de los 1500 galones se espera que queden al fin de la semana?

48. Si la temperatura a la que un cierto compuesto es una variable aleatoria con valor n desviación estándar de $2°C$, ¿cuáles son la temperatura media y la desviación estándar medidas ^oF, $^oF = 1{,}8°C + 32$)

49. 24. Haga que X tenga la f.d.p de Pareto

$$f(x) = \left\{ \begin{array}{lcl} \frac{k\theta^k}{x^{k+1}} & ; & x \geq \theta \\ 0 & , & x < \theta \end{array} \right.$$

a) Si $k > 1$, calcule $E(X)$.

b) ¿Qué se puede decir acerca de $E(X)$ si $k = 1$?

c) Si $k > 2$, demuestre que $V(X) = k\theta^2 (k-1)^{-2} (k-2)^{-1}$

d) Si $k = 2$, ¿qué se puede decir acerca de $V(X)$

e) ¿Qué condiciones para k son necesarias para asegurar que $E(X^n)$ sea finita?

50. Sea X la temperatura en $°C$ a la que tiene lugar cierta reacción química, y sea Y la temperatura en $°F$

a) Si la mediana de la distribución X es $\tilde{X}$, demuestre que. $1{,}8\tilde{X} + 32$ es la mediana de la distribución Y

b) Más generalmente, si $Y = aX + b$, ¿cómo se relaciona cualquier percentil particular de la distribución Y con el correspondiente percentil de la distribución X?

51. Al recordar la definición de σ^2 para una sola v.a. X, escriba una fórmula que sea correcta para calcular la varianza de una función $h(X,Y)$ de dos variables aleatorias. [SUGERENCIA: Recuerde que la varianza es sólo un valor esperado especial.]

52. Utilice las reglas de valor esperado para demostrar que

a) $Cov(aX + b, cY + d) = acCov(X,Y)$.

b) Utilice la parte (a) junto con las reglas de varianza y desviación estándar para demostrar que $Corr(aX + b, cY + d) = Corr(X,Y)$ cuando$a$ y c tienen el mismo signo.

c) ¿Qué sucede si a y c tienen signos opuestos?

53. Demuestre que si $Y = aX + b$, $(a \neq 0)$, entonces $Corr(X,Y) = +1$ o -1. ¿Bajo qué condiciones será $\rho = \pm 1$?

54. Supóngase que la f.d.p. conjunta de dos variables aleatorias X e Y es la siguiente:

$$f(x,y) = \left\{ \begin{array}{ccc} cy^2, & 0 \leq x \leq 2, & 0 \leq y \leq 1, \\ 0 & , & \text{en otro caso} \end{array} \right.$$

Determínese

a) El valor de la constante c

b) $P(X + Y > 2)$

c) $P(Y < 1/2)$

d) $P(X < 1)$

e) $P(X = 3Y)$.

55. Supóngase que la f.d.p. conjunta de dos variables aleatorias X e Y es la siguiente:

$$f(x,y) = \left\{ \begin{array}{ccc} c\left(x^2 + y^2\right) & si & 0 \leq y \leq 1 - x^2 \\ 0 & , & \text{en otro caso} \end{array} \right.$$

Determínese

a) El valor de la constante c

b) $P(0 < X < 1/2)$

c) $P(Y < X + 1)$

d) $P(Y = X^2)$.

56. Supóngase que se selecciona aleatoriamente un punto (X,Y) de la región S en el plano xy que contiene todos los puntos (x,y) tales que $x \geq 0$, $y \geq 0$ y $4y + x \leq 4$.

   a) Determínese la f.d.p. conjunta de .X e Y.

   b) Supóngase que S_0 es un subconjunto de la región S con área α y determínese $P[(X,Y) \in S_0]$.

57. Supóngase que se selecciona un punto (X,Y) del cuadrado S en el plano xy que contiene todos los puntos (x,y) tales que $0 \leq x \leq 1$ y $0 < y < 1$. Supóngase que la probabilidad de que el punto seleccionado sea el vértice $(0,0)$ es 0,1; la probabilidad de que sea el vértice $(1,0)$ es 0,2; la probabilidad de que sea el vértice $(0,1)$ es 0,4; y la probabilidad de que sea el vértice $(1,1)$ es 0,1. Supóngase también que si el punto seleccionado no es uno de los cuatro vértices del cuadrado, entonces será un punto interior y se seleccionará de acuerdo con una f.d.p. constante en el interior del cuadrado. Determínese

   a) $P(X < 1/4)$

   b) $P(X + Y < 1)$.

58. Supóngase que X e Y son variables aleatorias tales que (X,Y) puede pertenecer al rectángulo del plano xy que contiene todos los puntos (x,y) para los cuales $0 < x < 3$ y $0 \leq y \leq 4$. Supóngase también que la f.d. conjunta de X e Y en cualquier punto (x,y) de este rectángulo es la siguiente:

$$F(x,y) = \frac{1}{156} XY \left(X^2 + Y\right)$$

   Determínese

   a) $P(l < X \leq 2, 1 \leq Y \leq 2)$

   b) $P(2 \leq X < 4, 2 < Y < 4)$

   c) La f.d. de Y

   d) La f.d.p. conjunta de X e Y

   e) $P(Y < X)$.

59. Supóngase que X e Y tienen una distribución discreta conjunta cuya f.p. conjunta se define como sigue:

$$f(x,y) = \begin{cases} \frac{1}{30}(x+y), & x = 0,1,2; \quad y = 0,1,2,3, \\ 0 & , \quad \text{en otro caso} \end{cases}$$

   a) Determínese las f.p. marginales de X e Y,

   b) ¿Son independientes .X e Y?

60. Supóngase que X e Y tienen una distribución continua conjunta cuya f.d.p. conjunta se define como sigue:

$$f(x,y) = \begin{cases} \left(\frac{3}{2}\right)y^2, & 0 \leq x \leq 2, \quad 0 \leq y \leq 1, \\ 0 & , \quad \text{en otro caso} \end{cases}$$

   a) Determínese las f.d.p. marginales de X e Y.

   b) ¿Son independientes X e Y?

   c) ¿Son independientes los sucesos $\{X < 1\}$ y $\{Y \geq 1/2\}$?

61. Supóngase que la f.d.p. conjunta de X e Y es la siguiente:

$$f(x,y) = \begin{cases} \left(\frac{15}{4}\right)x^2 & , \quad 0 \leq y \leq 1 - x^2 \\ 0 & , \quad \text{en otro caso} \end{cases}$$

   a) Determínese las f.d.p. marginales de .X e Y

   b) ¿Son independientes X e Y ?

62. Un establecimiento tiene tres teléfonos públicos. Para $i = 0,1,2,3$, sea p_i, la probabilidad de que exactamente i teléfonos estén ocupados un lunes cualquiera a las 8 de la noche, y supóngase que $p_0 = 0{,}1$, $p_1 = 0{,}2$, $p_2 = 0{,}4$ y $p_3 = 0{,}3$. Sean X e Y el número de teléfonos ocupados a las 8 de la noche en dos lunes independientes. Determínese:

   a) La f.p. conjunta de X e Y.

   b) $P(X = Y)$.

   c) $P(X > Y)$.

63. Supóngase que la f.d.p. conjunta de dos variables aleatorias X e Y es la siguiente:

$$f(x,y) = \begin{cases} c\left(x + y^2\right), & 0 \leq x \leq 1, \quad 0 \leq y \leq 1 \\ 0 & , \quad \text{en otro caso} \end{cases}$$

   Determínese

   a) La f.d.p. condicional de X para cualquier valor concreto de Y.

   b) $P(Y < 1/2 | X = 1/2)$.

64. Supóngase que la f.d.p. conjunta de dos variables aleatorias X e Y es la siguiente:

$$f(x,y) = \begin{cases} c\,sen\,x, & 0 \leq x \leq \frac{\pi}{2}, \quad 0 \leq y \leq 3 \\ 0 & , \quad \text{en otro caso} \end{cases}$$

   Determínese

   a) La f.d.p. condicional de Y para cualquier valor dado de X

   b) $P(1 < Y < 2 | X = 0{,}73)$.

65. Supóngase que la f.d.p. conjunta de dos variables aleatorias X e Y es la siguiente:

$$f(x,y) = \begin{cases} \frac{3}{16}(4 - 2x - y), & x > 0, \; y > 0, \quad 2x + y < 4, \\ 0 & , \quad \text{en otro caso} \end{cases}$$

   Determínese

   a) La f.d.p. condicional de Y para cualquier valor dado de X.

   b) $P(y > 2 | X = 0{,}5)$.

66. Supóngase que la calificación X de una persona en una prueba de aptitud de matemáticas es un número entre 0 y 1, y que su calificación Y en una prueba de aptitudes musicales es también un número entre 0 y 1. Supóngase, además, que en la población de todos los estudiantes de bachillerato de Colombia, las calificaciones X e Y se distribuyen de acuerdo con la siguiente f.d.p. conjunta:

$$f(x,y) = \begin{cases} \frac{2}{5}(2x + 3y), & 0 \leq x \leq 1, \quad 0 \leq y \leq 1 \\ 0 & , \quad \text{en otro caso} \end{cases}$$

   a) ¿Qué proporción de estudiantes de bachillerato obtienen una calificación mayor que 0.8 en la prueba de matemáticas?

   b) Si la calificación en la prueba de música de un estudiante es 0.3, ¿cuál es la probabilidad de que su calificación en la prueba de matemáticas sea mayor que 0.8?

c) Si la calificación en la prueba de matemáticas de un estudiante es 0.3, ¿cuál es la probabilidad de que su calificación en la prueba de música sea mayor que 0.8?

67. Supóngase que una variable aleatoria X puede tomar cada uno de los siete valores $-3, -2, -1, 0, 1, 2, 3$ con la misma probabilidad. Determínese la f.p. de $Y = X^2 - X$.

68. Supóngase que la f.d.p. de una variable aleatoria X es la siguiente:

$$f(x) = \begin{cases} \frac{1}{2}x & , \quad 0 < x < 2 \\ 0 & , \quad \text{en otro caso} \end{cases}$$

Además,

a) supóngase que $Y = .Y(2 - X)$. Determínese la f.d. y la f.d.p. de Y.

b) Supóngase $Y = 4 - X^3$. Determínese la f.d.p. de Y

c) Supóngase $Y = aX + b$ $(a \neq 0)$. Demuéstrese que la f.d.p. de Y es la siguiente:

$$g(y) = \frac{1}{|a|} f\left(\frac{y-b}{a}\right) \text{ para } -\infty < y < \infty$$

69. Supóngase que la f.d.p. de $.X$ es la siguiente:

$$f(x) = \begin{cases} e^{-x} & , \quad x > 0 \\ 0 & , \quad x \leq 0 \end{cases}$$

Determínese la f.d.p. de $Y = X^{\frac{1}{2}}$.

70. Supóngase que X_1 y X_2 tienen una distribución conjunta continua cuya f.d.p. conjunta es la siguiente:

$$f(x_1, x_2) = \begin{cases} x_1 + x_2, & 0 < x_1 < 1, \quad 0 < x_2 < 1 \\ 0 & , \quad \text{en otro caso} \end{cases}$$

Determínese

a) la f.d.p. de $Y = X_1 X_2$.

b) la f.d.p. conjunta de $Z = \frac{X_1}{X_2}$

71. Sean X e Y variables aleatorias cuya f.d.p. conjunta es la siguiente:

$$f(x, y) = \begin{cases} 2(x + y) & si \quad 0 \leq x \leq y \leq 1 \\ 0 & , \quad \text{en otro caso} \end{cases}$$

Determínese la f.d.p. de $Z = X + Y$.

72. Supóngase que X_1 y X_2 son variables aleatorias i.i.d.(identicamente distribuidas e independientes) y que la f.d.p. de cada una de ellas es la siguiente:

$$f(x) = \begin{cases} e^{-x} & si \quad x > 0, \\ 0 & , \quad \text{en otro caso} \end{cases}$$

Determínese la f.d.p. de $Y = X_1 - X_2$.

73. Suponga que $X_1, X_2, \cdots, X_n$ constituyen una muestra aleatoria de tamaño n de una distribución

$$f(x) = \begin{cases} c & si \quad 0 \leq x \leq 1 \\ 0 & , \quad \text{en otro caso} \end{cases}$$

si $Y = \text{máx}\{X_1, X_2, \cdots, X_n\}$. Determínese el valor más pequeño de n tal que $P(Y \geq 0{,}99) \geq 0{,}95$. (a $f(x)$ se le llama distribución uniforme)

74. Sea W el rango de una muestra aleatoria de n observaciones de una distribución uniforme sobre el intervalo $(0, 1)$. Determínese el valor de $P(W > 0{,}9)$.

75. Supóngase que una variable aleatoria X tiene una distribución uniforme sobre el intervalo $0 < x < 1$. Demuéstrese que la esperanza de $1/X$ no existe.

76. Supóngase que X e Y tienen una distribución conjunta continua cuya f.d.p. conjunta la siguiente:

$$f(x, y) = \begin{cases} 12y^2 & , \quad 0 \leq y \leq x \leq 1 \\ 0 & , \quad \text{en otro caso} \end{cases}$$

Determínese el valor de $E(XY)$.

77. Supóngase que se selecciona al azar un punto de un bastón de longitud unidad y que se rompe en dos trozos por ese punto. Determínese el valor esperado de la longitud del trozo más grande.

78. Supóngase que se libera una partícula del origen del plano xy y pasa al semi-plano en que $x > 0$. Supóngase que la partícula se mueve en línea recta y que el ángulo entre el semieje x positivo y esta línea es α, el cual puede ser positivo o negativo. Supóngase, por último, que el ángulo α tiene una distribución uniforme sobre intervalo $(-\pi/2, \pi/2)$. Sea Y la ordenada del punto en que la partícula corta la recta vertical $x = 1$. Demuéstrese que la distribución de Y es una distribución de Cauchy.

79. Supóngase que las variables aleatorias $X_1, X_2, \cdots, X_n$ constituyen una muestra aleatoria(es decir todas la variables son i.i.d) de tamaño n (de una distribución uniforme sobre el intervalo $(0, 1)$. Sea $Y_1 = \text{mín}\{X_1, ..., X_n\}$ y sea $Y_n = \text{máx}\{.X_1, ..., Xn_n\}$. Determínese $E(Y_1)$ y $E(Y_n)$

80. Para cualquier par de números a y b tales que $a < b$, determínese la varianza de la distribución uniforme sobre el intervalo (a, b).

81. Supóngase que X es una variable aleatoria cuyas $E(X) = \mu$ y $Var(X) = \sigma^2$. Demuéstrese que $E[X(X-1)] = \mu(\mu - 1) + \sigma^2$.

82. Sea X una variable aleatoria cuya $E(X) = \mu$ y $Var(X) = \sigma^2$ y sea c cualquier constante. Demuéstrese que

$$E[(X - c)^2] = (\mu - c)^2 + \sigma^2$$

83. Supóngase que X e Y son variables aleatorias independientes con varianzas finitas tales que $E(X) = E(Y)$. Demuéstrese que

$$E[(X - Y)^2] = V(X) + Var(Y).$$

84. Supóngase que$X$ e Y son variables aleatorias independientes con $V(X) = V(Y) = 3$. Determínese los valores de

a) $V(X - Y)$,

b) $V(2X - 3Y + 1)$.

85. Construyase un ejemplo de una distribución cuya media exista, pero no su varianza.

86. Supóngase que es igualmente verosímil que un valor observado de X provenga de una distribución continua cuya f.d.p. es f que de una cuya f.d.p. es g. Supóngase que $f(x) > 0$ para $0 < x < 1$ y $f(x) = 0$ en otro caso y supóngase también que $g(x) > 0$ para $2 < x < 4$ y$g(x) = 0$ en otro caso. Determínese, la media y la mediana de la distribución de X,

87. Supóngase que la variable aleatoria X tiene una distribución continua cuya f.d.p. es la siguiente:

$$f(x) = \begin{cases} 2x & , \quad 0 < x < 1 \\ 0 & \quad \text{en otro caso} \end{cases}$$

Determínese el valor de d que minimiza

a) $E[(X-d)^2]$ y

b) $E(|X-d|)$.

88. Supóngase que la calificación X de una persona en un examen concreto es un número del intervalo $0 \leq X \leq 1$ y que X tiene una distribución continua cuya f.d.p. es la siguiente:

$$f(x) = \begin{cases} x + \frac{1}{2} & , \quad 0 \leq x \leq 1 \\ 0 & \quad \text{en otro caso} \end{cases}$$

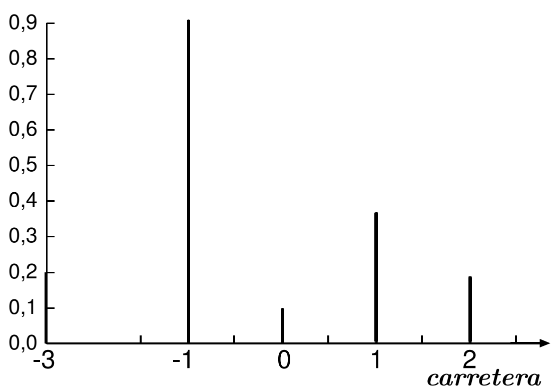
Determínese el valor de d que minimiza

a) $E[(X-d)^2]$ y

b) $E(|X-d|)$.

89. Supóngase que la distribución de una variable aleatoria X es simétrica respecto al punto $x = 0$ y que $E(X^4) < \infty$. Demuéstrese que$E[(X-d)^4]$ se minimiza con el valor $d = 0$.

90. Supóngase que puede ocurrir un incendio en cualquiera de cinco puntos a lo largo de una carretera. Estos puntos se localizan en $-3-1, 0, 1$ y 2 en la 90. Supóngase también que la probabilidad de que cada uno de estos puntos sea la localización del próximo incendio que ocurra a lo largo de la carretera es como se representa en la 90.¿En qué punto a lo largo de la carretera debería esperar un coche de bomberos para minimizar el valor esperado del cuadrado de la distancia que debe viajar hasta el siguiente incendio?



Gráfica del ejercicio 90

91. Demuéstrese que si $Var(X) < \infty$ y $Var(Y) < \infty$, entonces $Cov(X,Y)$ es finita. Sugerencia: Considerando la relación $[(X-\mu_X) \pm (Y-\mu_Y)]^2 \geq 0$, demuestre que

$$|(X-\mu_X)(Y-\mu_Y)| \leq \frac{1}{2}[(X-\mu_X)^2 + (Y-\mu_Y)^2].$$

92. Supóngase que X tiene una distribución uniforme en el intervalo $(-2,2)$ y que $Y = X^6$. Demuéstrese que X e Y son no correlacionadas.

93. Supóngase que la distribución de una variable aleatoria$X$ es simétrica respecto al punto $x = 0$, que $0 < E(X^4) < \infty$ y que $Y = X^2$. Demuéstrese que X e Y son no correlacionadas.

94. Para cualesquiera variables aleatorias X eY y constantes a, b, c y d, Demuéstrese que

$$Cov(aX+6, cY+d) = acCov(X,Y)$$

95. Sean X e Y variables aleatorias tales que $0 < \sigma_X^2 < \infty$ y $0 < \sigma_Y^2 < \infty$. Supóngase que $U = aX + b$ y $V = cY + d$ donde $a \neq 0$ y $c \neq 0$. Demuéstrese que $\rho(U,V)= \rho(X,Y)$ si $ac > 0$ y que $\rho(U,V) = -\rho(X,Y)$ si $ac < 0$.

96. Sean X, Y y Z tres variables aleatorias tales que $Cov(X,Z)$ y $Cov(Y,Z)$ existen y sean a, b y c constantes cualesquiera. Demue que

$$Cov(aX + bY + c, Z) = aCov(X,Z) + bCov(Y,Z).$$

97. Supóngase que $X_1, \cdots, X_n$ y $Y_1, \cdots, Y_n$ son variables aleatorias tales que $Cov(X_i, Y_i)$ existen para $i = 1, ..., n$ y $j = 1, ..., m$; y supóngase que $a_1, ..., a_m$ y $b_1, ..., b_n$ son constantes. Demuéstrese que

$$Cov\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j Cov(X_i, Y_j)$$

98. Considérese una función de utilidad U para la que $U(0) = 0$ y $U(100) = 1$. Supóngase que a una persona que tiene esta función de utilidad se muestra indiferente entre aceptar un juego cuya ganancia será 0 dólares con probabilidad 1/3 ó 100 dólares con probabilidad 2/3 y aceptar 50 dólares seguros. ¿ Cuál es el valor $U(50)$?

99. Considérese una función de utilidad U para la que $U(0) = 5$, $U(1) = 8$ y $U(2) = 1$ Supóngase que una persona que tiene esta función de utilidad se muestra indiferente entre los juegos X e Y para los cuales las distribuciones de probabilidad de las ganancias son como sigue:

100. $P(X=-1) = 0{,}6$, $P(X=0) = 0{,}2$, $P(X=2) = 0{,}2$, $P(Y=0) = 0{,}9$, $P(Y=1) = 0{,}1$. ¿Cuál es el valor de $U(-1)$?

101. Supóngase que una persona debe aceptar un juego X con la forma siguiente: $P(X=a) = p$ y $P(X = 1-a) = 1-p$, donde p es un número tal que $0 < p < 1$. Supóngase también que la persona puede elegir y fijar el valor de a $(0 \leq a \leq 1)$ utilizado en este juego. Determine el valor de a que la persona elegiría si su función de utilidad es $U(x) = \log x$ para $x > 0$

CAPÍTULO

5 Distribuciones

Contenido Del Capítulo

Introducción

Para realizar predicciones en estadística es necesario establecer un modelo probabilístico y para ello hay que realizar algo más que un histograma una tabulación o una fórmula para representar una variable aleatoria ya que que con frecuencia , las observaciones que se generan en diferentes experimentos estadísticos presentan el mismo comportamiento , por lo que las variables aleatorias asociadas a estos experimentos pueden describirse por la misma función de distribucíon ,ya sea discreta o continua, por esto a continuación presentaremos las funciones de distribucción más importantes.

5.1 Algunas distribuciones discretas importantes

5.1.1. Distribución uniforme

Esta distribución de probabilidad discreta es la más sencilla de todas y es aquella para la cual la variable aleatoria asociada a ella asume cada uno de sus valores con idéntica probabilidad es decir

Definición 5.1

Sea X una variable aleatoria que asume los valores $x_1, x_2, x_3, \cdots x_k$, con iguales probabilidades, entonces

$$f(x,k) = \begin{cases} \frac{1}{k} & si \quad x = x_1, x_2, x_3, \cdots x_k \\ 0 & \text{si no} \end{cases}$$

Teorema 5.1

Sea Xuna v.a discreta con distribución de probabilidad uniforme, entonces

$$\mu = E(X) = \frac{\sum_{i=1}^{k} x_i}{k}$$
$$\sigma^2 = V(X) = \frac{\sum_{i=1}^{k} (x_i - \mu)^2}{k}$$

y se denota $X \rightsquigarrow \mathbb{U}_d\left(\mu, \sigma^2\right)$

Ejemplo: 5.1.1

Se selecciona a un empleado de un grupo de 10 para supervisar un cierto proyecto, escogiendo aleatoriamente una placa de una caja que contiene 10 placas numeradas del 1 al 10. Encuentre la función de probabilidad de X que representa el número de la placa que se saca y

1. ¿Cuál es la probabilidad de que el número que se saque sea menor que 4 ?

   a) Encuentre la media y la varianza de X

   Como cada placa tiene la misma posibilidad de salir entonces $X \rightsquigarrow \mathbb{B}_d\left(\mu, \sigma^2\right)$ la cual es

$$f(x,10) = \begin{cases} \frac{1}{10} & si \quad x = 1, 2, 3, \cdots 10 \\ 0 & \text{si no} \end{cases}$$

Solución

1. $P(4 \geq X) = \sum_{i=1}^{3} f(x_i) = (1+1+1)\frac{1}{10} = \frac{3}{10}$

   a) $E(X) = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{11}{2}, V(X) = \frac{\sum_{i=1}^{10} (x_i - 5{,}5)^2}{10} = 8.25$

5.1.2. Distribución de Bernoulli

Si realizamos un experimento una sola vez y observamos si cierto suceso ocurre o no decimos que este experimento es de Bernoulli y decimos que la variable aleatoria X la podemos definir

$$X = \begin{cases} 0 & si \quad \text{no ocurre el suceso} \\ 1 & si \quad \text{el suceso ocurre} \end{cases}$$

y su función de probabilidad

$$f(x,p) = \begin{cases} q & si \quad x = 0 \\ p & si \quad x = 1 \\ 0 & \text{en otro caso} \end{cases}$$

y su función de distribución es

$$F(x,p) = \begin{cases} 0 & si \quad x < 0 \\ q & si \quad 1 > x \geq 0 \\ 1 & si \quad x \geq 1 \end{cases}$$

En el ejemplo 4.23 se determinó que $E(X) = p$ y $V(X) = pq$ en el ejemplo 4.27 una variable de Bernoulli con media p y varianza pq se denota $X \rightsquigarrow \mathbb{B}(p, pq)$

5.1.3. Distribución binomial

Se dice que una variable aleatoria X tiene una distribución binomial de parámetros n, p si $X = \sum_{i=1}^{n} X_i$ donde cada $X_i \rightsquigarrow \mathbb{B}(p, pq)$ y se denota $B(np, npq)$

Definición 5.2

Supóngase que se realiza un experimento de Bernoulli n veces , donde en todas ellas, la probabilidad de éxito es la misma (p), y queremos calcular el número de éxitos, X obtenidos en el total de n pruebas, entonces su función de probabilidad es

$$f(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & si \quad x = 0, 1, 2, \cdots, n \\ 0 & \text{en otro caso} \end{cases}$$

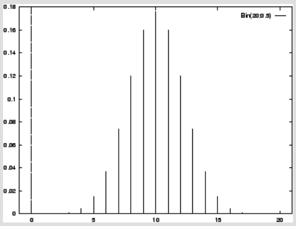


figura 5.1 distribución binomial

Por tanto, su función de distribución es

$$F(x) = \begin{cases} 0 & si \quad x < 0 \\ \sum_{k=0}^{[x]} \binom{n}{k} p^k q^{n-k} & si \quad n \geq x \geq 0 \\ 1 & si \quad x \geq n \end{cases}$$

La esperanza y la varianza se calcularon en los ejemplos 4.23 y 4.27

$$E(X) = np, V(X) = npq$$

Ejemplo: 5.1.2

Supóngase que la probabilidad de que un cierto experimento tenga éxito es 0.4 y sea X el número de éxitos que se obtienen en 15 realizaciones independientes del experimento. Utilice la tabla de la distribución binomial para determinar el valor P($9 \geq X \geq 6$)

Solución

Tenemos que X tiene una distribución binomial con $p = 0{,}4$ y $n = 15$ por tanto se busca en la tabla estos parámetros y se suman los valores entre 6 y 9

$$\begin{aligned} P(9 \geq X \geq 6) &= 0{,}2066 + 0{,}1771 + 0{,}1181 + 0{,}0612 \\ &= 0.\,563 \end{aligned}$$

Ejemplo: 5.1.3

Un cierto sistema electrónico contiene diez componentes. Supóngase que la probabilidad de que un componente individual falle es 0.2 y que los componentes fallan independientemente unos de otros. Dado de que al menos uno de los componentes ha fallado .¿Cuál es la probabilidad de que fallen al menos dos de los componentes?

Solución

$p = 0{,}2$ y $n = 9$ Entonces $P(1 \geq X \geq 0) = 0{,}4362$

5.1.4. Distribución geométrica

Consideremos una sucesión de v.a independientes de Bernoulli

$$X_1, X_2, X_3, \cdots$$
$$donde \quad X_i \rightsquigarrow B\left(np, npq\right) \quad \forall i = 1, 2, 3, \cdots$$

Una v.a. X tiene una distribución geométrica o de fracaso , si esta es la suma del número de fracasos obtenidos hasta la aparición del primer éxito en la sucesión $\{X_i\}_i$

$$\begin{array}{cccccccc}
X_1 & X_2 & X_3 & \cdots & & X & & f \\
 & & & \vdots & & & & \\
\downarrow & \downarrow & \downarrow & & & \downarrow & & \downarrow \\
1 & 0 & 0 & \cdots & \Longrightarrow & 0 & & f(0) = p \\
0 & 1 & 0 & \cdots & \Longrightarrow & 1 & & f(1) = qp \\
0 & 0 & 1 & \cdots & \Longrightarrow & 2 & & f(2) = qqp \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots
\end{array}$$

Siguiendo este proceso se obtiene que la función de probabilidad es

$$f(x) = \begin{cases} pq^x & si \quad x = 0, 1, 2, \cdots \\ 0 & \text{en otro caso} \end{cases}$$

Tenemos que

$$\sum_{k=0}^{\infty} f(k) = \sum_{k=0}^{\infty} pq^k =$$
$$p\sum_{k=0}^{\infty} q^k = p\frac{1}{1-q} = 1$$

ya que es una serie geométrica por ser 1$\geq q \geq 0$, por lo que nos queda determinar su esperanza y su varianza

$$E(X) = p\sum_{k=0}^{\infty} kq^k = \frac{p}{(p-1)^2} = \frac{pq}{(q-1)^2} = \frac{q}{p}$$
$$E\left(X^2\right) = p\sum_{k=0}^{\infty} k^2q^k = p\frac{q^2+q}{(1-q)^3} = \frac{q(q+1)}{p^2}$$
$$V(X) = \frac{q(q+1)}{p^2} - \left(\frac{q}{p}\right)^2 = \frac{q}{p^2}$$

Ejemplo: 5.1.4

Un matrimonio quiere tener una hija, y por ello deciden tener hijos hasta el nacimiento de una hija. Calcular la probabilidad de que una pareja acabe teniendo tres hijos o más

Solución

Vamos a suponer que la probabilidad de tener una niña es de $\frac{2}{3}$. Sea X el número de hijos varones que tienen antes de tener una niña. por lo tanto $X \rightsquigarrow G\left(\frac{1}{2}, \frac{3}{4}\right), entonces$

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) \\ &= 1 - p - qp \\ &= 1 - \frac{2}{3} - \frac{2}{9} = \frac{1}{9} \end{aligned}$$

5.1.5. Distribución hipergeométrica

Cuando El interés que se tiene es el de determinar la probabilidad de seleccionar x éxitos de los k posibles resultados o artículos considerados éxitos donde hay $n-x$ resultados considerados fracasos de los $N-k$ posibles resultados, entonces se realiza un experimento llamado hipergeométrico.

Un experimento hipergeométrico es aquel que sigue los siguientes pasos

1. Se tiene una muestra aleatoria de tamaño n se selecciona sin reemplazo un total de N resultados o artículos
2. k artículos del total de N pueden clasificarse como éxitos y $N-k$ como fracasos.

Entonces el número X de éxitos en un experimento hipergeométrico recibe el nombre de variable aleatoria hipergeométrica y la distribucón de probabilidad se llama distribución hipergeométrica

Definición 5.3

La distribución de probabilidad de la variable aleatoria hipergeométrica X, el número de éxitos en una muestra aleatoria de tamaño n seleccionada de N resultados posibles, de los cuales k son considerados éxitos y $N-k$ como fracasos es

$$f(x)=\begin{cases}\dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} & si \quad x=1,2,3,\cdots,n\\ 0 & \text{en otro caso}\end{cases}$$

La media y la varianza de una variable hipergeométrica son

$$E(X)=\frac{nk}{N}$$
$$V(X)=\frac{N-n}{N-1}n\frac{k}{N}\left(1-\frac{k}{N}\right)$$

Cuando $N\longrightarrow\infty$ observamos que $HG(N,n,p)\longrightarrow B(np,npq)$

$E(X)=np$ y $V(X)=npq\frac{N-n}{N-1}$ el cual no es la misma varianza de la binomial por lo que al factor $\frac{N-n}{N-1}$ se le llama de corrección para la población finita

Ejemplo: 5.1.5

En un departamento de inspección de envíos, se reciben en forma periódica lotes de ejes de bombas. Los lotes contienen 100 unidades y el siguiente plan de muestreo de aceptación se utiliza. Se selecciona una muestra aleatoria de 10 unidades sin reemplazo. El lote se acepta si la muestra no tiene más de un artículo defectuoso : supóngase que se recibe un lote si éxito una probabilidad de que un artículo sea defectuoso es 0.05 .¿Cuál es la probabilidad de que sea aceptado?

Solución

Tenemos que determinar $P(1\geq X)$ entonces $k=0.05*100$

$$P(1\geq X)=\sum_{x=0}^{1}\frac{\binom{5}{x}\binom{95}{10-x}}{\binom{100}{10}}=0.923\,14:$$

5.1.6. Distribucción de Poisson

Sea X una variable aleatoria con una distribución discreta y supóngase que el valor de X debe ser un entero no negativo. Se dice que X tiene una distribución de Poisson con media λ si la f.p de X es

$$f(x)=\begin{cases}\frac{e^{-\lambda}\lambda^{x}}{x!} & si \quad x=1,2,3,\cdots\\ 0 & \text{en otro caso}\end{cases}$$

del cálculo tenemos que $e^{\lambda}=\sum_{x=0}^{\infty}\frac{\lambda^{x}}{x!}$ por tanto $\sum_{x=0}^{\infty}f(x)=\sum_{x=0}^{\infty}\frac{e^{-\lambda}\lambda^{x}}{x!}=e^{-\lambda}e^{\lambda}=1$
$E(X)=\sum_{x=0}^{\infty}x\frac{e^{-\lambda}\lambda^{x}}{x!}=\lambda\sum_{x=0}^{\infty}\frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}=\lambda$

se deja como ejercicio comprobar que
$\mathrm{E}(X(X-1)) = \lambda^2 = E\left(X^2\right) - \lambda$

$\mathrm{V}(X) = \lambda$

Se denota $X \rightsquigarrow P(\lambda, \lambda)$

Ejemplo: 5.1.6

Cierta enfermedad tiene una probabilidad muy baja de ocurrir, $p = 10^{-6}$. Calcular la probabilidad de que en una ciudad con 500000 habitantes haya más de 3 personas con dicha enfermedad. Calcular el número de habitantes que la padecen.

Solución

Si consideramos la v.a X que contabiliza el número de personas que padecen la enfermedad, es claro que sigue un modelo binomial, pero que puede ser aproximado por un modelo de Poisson, tal que $\lambda = np = 5$, es decir $X \rightsquigarrow P(5,5)$, *entonces*

$$\begin{aligned}P(X > 3) &= 1 - P(3 \geq X)\\ &= 1 - \sum_{i=0}^{3} \frac{\exp(-5)\,5^i}{i!} = 1 - \frac{118}{3}e^{-5}\\ &= 0.73497\end{aligned}$$

5.1.6.1. Proceso de poisson

Al definir un proceso de Poisson, consideramos una colección de sucesos relativos al tiempo, las cuales generalmente se les denomina llegadas.
Al escoger una variable aleatoria X de interés como el número de llegadas que ocurren en un intervalo de tiempo $[0,t]$, obtenemos una variable discretizada con $R_X = \{0,1,2,\cdots\}$ y que para determinar su distribución hay que suponer dos condiciones las cuales solo tienen sustento empírico
La primera condición es que el número de llegadas durante un intervalo de tiempo son representadas por variables aleatorias independientes, la segunda es que exista una constante $c > 0$ tal que cumpla los siguientes postulados

1. La probabilidad de que ocurrirá con exactitud una llegada en un intervalo con duración $\triangle t$, es aproximadamente $c\triangle t$. donde c se le denomina tasa de llegada media
2. La probabilidad de que ocurrirá exactamente cero llegadas en el intervalo es aproximadamente $1 - c\Delta t$
3. La probabilidad de que ocurra dos o más llegadas en le intervalo es igual a una cantidad $\circ(\Delta t)$, donde $\circ$ es una función de t con la propiedad que

$$\lim_{t \longrightarrow \infty} \frac{\circ(t)}{t} = 0$$

Considerando estos supuestos entonces

$$\lambda = c\Delta t \text{ y } X \rightsquigarrow P(c\Delta t, c\Delta t)$$

Ejemplo: 5.1.7

Supóngase que un comerciante determina que el número de pedidos para un cierto aparato doméstico en un periodo de tiempo particular tiene una distribución de Poisson con parámetro λ. Al le gustaría determinar el nivel de existencias a para el principio del periodo de manera que haya una probabilidad de al menos 0.95 de surtir a todos los clientes que pidan el aparato durante el periodo, pues no desea devolver pedidos ni volver a surtir el almacén durante ese periodo.

Solución

Si X representa el número de pedidos, entonces

$$P(a \geq X) \geq 0{,}95 \Longleftrightarrow 0{,}5 \geq P(X > a)$$

por lo tanto

$$0{,}5 \geq \sum_{x=(a+1)}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{x!}$$

De acuerdo con la tabla $7 = a + 1 \Longrightarrow a = 6$

Ejemplo: 5.1.8

Un dispositivo electrónico de conmutación ocasionalmente funciona mal y puede ser necesario reemplazarlo. Se sabe que el dispositivo es satisfactorio si, en promedio, no comete más de 0.20 errores por hora. Se selecciona un periodo particular de 5 horas como prueba del dispositivo. Si no ocurre más de un error, el dispositivo se considera satisfactorio.¿Cuál es la probabilidad de que un dispositivo se considere insatisfactorio de acuerdo con la prueba?

Solución

Sea X el número de fallas del dispositivo en 5 horas entonces podemos suponer que está variable sigue un proceso de Poisson tal que $\lambda = 5\,(0{,}20) = 1$ entonces

$$P\,(X \geq 1) = 1 - P\,(X < 1) = 0{,}63$$

5.1.7. Distribucción Multinomial

Supóngase que una población contiene artículos de $k \geq 2$ tipos distintos y que la proporción de artículos del tipo i en la población es p_i, $i = 1, 2, 3, \cdots, k$.Supóngase que $p_i > 0$ y que $\sum_{i=1}^{k} p_i = 1$. Además, supóngase que es seleccionan al azar con reemplazo n artículos de la población y sea X_i el número de artículos seleccionados que son del tipo i. Se dice entonces que el vector aleatorio $\overrightarrow{X} = (X_1, \cdots, X_k)$ tiene una distribución multinomial con parámetro $p\,(p_1, \cdots, p_k)$ es

$$f\left(\overrightarrow{x}\right) = \begin{cases} \frac{n!}{x_1! \cdots x_2!} & si \quad x_1, \cdots x_k \in \mathbb{Z}^+ \\ 0 & \text{si no} \end{cases}$$

Donde $\sum_{i=1}^{k} x_i = n$

5.1.7.1. Medias, varianzas y covarianzas

$$E\,(X_i) = np_i, V\,(X_i) = np_i q_i$$
$$V\,(X_i + X_j) = n\,(p_i + p_j)\,(1 - p_i - p_j)$$
$$cov\,(X_i, X_j) = -np_i p_j \quad \forall i, j = 1, 2, \cdots, k$$

Ejemplo: 5.1.9

Se fabican lápices mecánicos por medio de un proceso que implica una gran cantidad de mano de obra en las operaciones de ensamble. Éste es un trabajo altamente repetitivo y hay un pago de incentivo. La inspección final ha revelado que el 85 % del producto es bueno y el 10 % defectuoso pero que puede reelaborarse, y el 5 % es defectuoso y se desecha. Estos porsentajes se mantienen constantes todo el tiempo. Se selecciona una muestra aleatoria de 20 artículos . Si se designa

$X_1 =$ número de artículos buenos

$X_2 =$ número de artículos defectuosos que se pueden rescatar

$X_3 =$ número de artículos que se desechan

Determinar la probabilidad que 18 sean buenos y que 2 se puedan rescatar

Solución

Nos piden la probabilidad

$$P(18,2,0) = \frac{20!}{(18!)(2!)(0!)}(0{,}85)^{18}(0{,}1)^2(0{,}05)^0$$
$$= 0.101\,93$$

5.2 Algunas distribuciones continuas

5.2.1. Distribución uniforme

Se dice que una v.a X tiene una distribución uniforme continua en un intervalo $[a,b]$ si su función densidad es

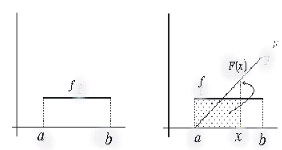
$$f(x) = \begin{cases} \frac{1}{b-a} & si \quad x \in [a,b] \\ 0 & \text{en otro caso} \end{cases}$$

y se denota $X \rightsquigarrow U_c\left(\mu, \sigma^2\right)$ donde μ es la media y σ^2 es la varianza. Se puede determinar su función de distribución como

$$F(x) = \begin{cases} 0 & si \quad x < a \\ \frac{x-a}{b-a} & si \quad x \in [a,b] \\ 1 & si \quad x \geq b \end{cases}$$

Es fácil calcular su media y su varianza se le deja como ejercicio al lector

$$E(X) = \frac{b+a}{2}$$
$$V(X) = \frac{(b-a)^2}{12}$$



Distribución uniforme

Ejemplo: 5.2.10

Se elige un punto al azar en el intervalo [0,10]. Supóngase que deseamos encontrar la probabilidad de que el punto esté entre 1.5 y 3.5 si la función de densidad de la v.a aleatoria X es

$$f(x) = \begin{cases} \frac{1}{10} & si \quad x \in [0,10] \\ 0 & \text{si no} \end{cases}$$

Tenemos que

$$P(3,5 \geq X \geq 1,5) = \frac{2}{10}$$

5.2.2. Distribución exponencial

La distribución exponencial es el equivalente continuo por así decirlo de la distribución geométrica discreta, Esta describe un proceso en el que.

Nos interesa saber el tiempo hasta que ocurre determinado evento, sabiendo que el tiempo que pueda ocurrir desde cualquier instante dado t, hasta que ello ocurra en un instante t_f, no depende del tiempo trascurrido anteriormente en el que no pasó nada

Si X es una v.a se dice que tiene una distribución exponencial si

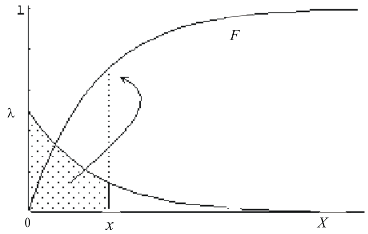
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & si \quad x > 0 \\ 0 & \text{si no} \end{cases}$$

Y su función de distribución es

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & si \quad x > 0 \\ 0 & \text{si no} \end{cases}$$

Su esperanza y su varianza son

$$E(X) = \frac{1}{\lambda}$$
$$V(X) = \frac{1}{\lambda^2}$$



Distribución exponencial

Ejemplo: 5.2.11

Se sabe que un componente electrónico tiene una vida útil representada por una densidad exponencial con tasa de falla de 10^{-5} fallas por hora.Determinar la fracción de tales componentes que podrían fallar antes de la vida media

Solución

Sea T el tiempo que tarda en dañarse el componente, entonces $\lambda = 10^5$

$$P\left(\frac{1}{\lambda} \geq T\right) = \int_0^{\frac{1}{\lambda}} \lambda e^{-\lambda x} dx = 1 - e^{-1}$$

Ejemplo: 5.2.12

Se ha comprado que el tiempo de vida de cierto tipo de marca pasos sigue una distribución exponencial con media de 16 años. ¿uál es la probabilidad de que a una persona a la que se le ha implantado este marca paso se le deba reimplantar otro antes de 20 años?. Si el marca paso lleva funcionando correctamente 5 años en un paciente, ¿Cuál es la probabilidad de que haya que cambiarlo antes de 25 años?.

Solución

Sea T la variable aleatoria que mide la duración de un marca pasos en una persona. Tenemos que $\lambda = \frac{1}{16}$, entonces

$$P(20 \geq T) = \int_0^{20} f(t)\, dt = F(20) = 0{,}7135$$

ahora tenemos

$$P(25 \geq T | T \geq 5) = \frac{P(25 \geq T \geq 5)}{P(T \geq 5)}$$

Por lo tanto

$$
\begin{aligned}
P\left(25 \geq T \geq 5\right) &= F\left(25\right) - F\left(0\right) = 0{,}522 \\
P\left(T \geq 5\right) &= F\left(+\infty\right) - F\left(5\right) = 0{,}7316 \\
P\left(25 \geq T|T \geq 5\right) &= 0{,}7135
\end{aligned}
$$

5.2.3. Distribución normal o Gaussiana

Se dice que una v.a tiene una distribución normal de parámetros μ y σ^2 lo que se denota $X \rightsquigarrow N\left(\mu, \sigma^2\right)$ si su función de densidad es

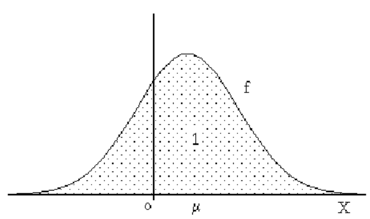
$$f\left(x\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad si \quad x \in I\!R$$

$$F\left(x\right) = P\left(x \geq X\right) = \int_{-\infty}^{x} f\left(x\right) dx$$

La media y la varianza son

$$
\begin{aligned}
E\left(X\right) &= \mu \\
V\left(X\right) &= \sigma^2
\end{aligned}
$$

- Estas variables tienen una propiedad importante que se llama reproductividad, es decir la suma de variables aleatorias e independientes normales también es normal
- Una propiedad de esta distribución es que es simétrica con respecto a la media
- Si $X \rightsquigarrow N\left(0,1\right)$ se dice que X tiene una distribución normal estándar y su f.d se denota $\varphi\left(X\right)$ y su función de distribución $\Phi\left(X\right)$
- Los puntos de inflexión están en $x = \mu \pm \sigma$
- $$\lim_{x \longrightarrow +\infty} f\left(x\right) = 0 \;\text{ y }\; \lim_{x \longrightarrow -\infty} f\left(x\right) = 0$$



Distribución Normal

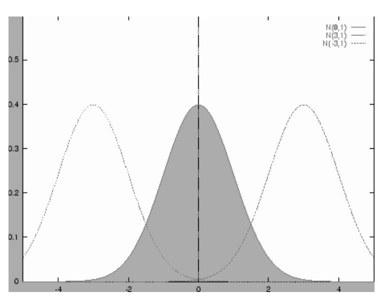


Figura 5.1: Distribuciones Normales con igual varianza y diferente esperanza

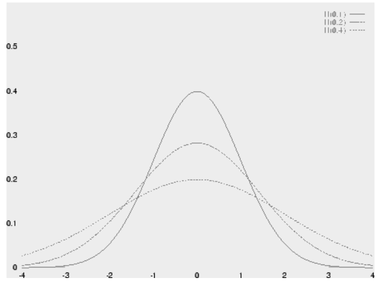


Figura 5.2: Distribuciones Normales con igual media pero diferente varianza

Proposición 11

ean $a, b \in \mathbb{R}$, entonces

$$X \rightsquigarrow N\left(\mu, \sigma^2\right) \Longrightarrow$$
$$Y = aX + b \rightsquigarrow N\left(a\mu + b, (a\sigma)^2\right)$$

Este resultado lo utilizaremos para estandarizar una variable normal
Si $X \rightsquigarrow N\left(\mu, \sigma^2\right)$ entonces $Z = \frac{X-\mu}{\sigma} \rightsquigarrow N(0,1)$

Ejemplo: 5.2.13

Supongamos que cierto fenómeno pueda ser representado mediante una v.a. aleatoria $X \rightsquigarrow N(45, 81)$, y queremos calcular la probabilidad de que X tome un valor entre 39 y 48,

Solución

Para resolver este problema debemos usar la tabla que aparece al final del capítulo, pero antes hay que estándarizar la variable de tal manera que

$$Z = \frac{X-\mu}{\sigma}$$
$$= \frac{X-45}{\sqrt{81}}$$

de modo que si
$48 \geq X \geq 39 \Longleftrightarrow \frac{1}{3} \geq \frac{X-45}{9} \geq \frac{-2}{3}$, entonces
$P(48 \geq X \geq 39) = \Phi\left(\frac{1}{3} \geq Z \geq \frac{-2}{3}\right) = 0,378$

Cola de la derecha
$Z \leq \frac{1}{3}$

Ejemplo: 5.2.14

Sugún un estudio de la altura de los varones de cierta ciudad es una variable aleatoria X, que podemos considerar que se distribuye normalmente con media 175cm y varianza 10cm. Determinar un intervalo para el cual el 50 % de los habitantes de esa ciudad están comprendidos en él.

Solución

Existen infinitas soluciones para este problema en la gráficas siguientes podemos analizar tres de las tantas posibilidades

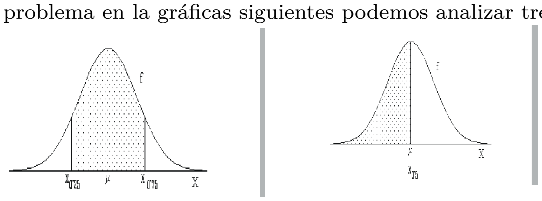


fig 5.07

$X \geq x_{0,5}$ y que $x_{0,75} \geq X \geq x_{0,25}$ en los dos casos hay que estandarizar las variables, en el primer caso es obvio que el intervalo es $(-\infty, 175]$ en el segundo caso usaremos la tabla para determinar

$$
\begin{aligned}
z_{0,75} &= 0,675 \\
&= \frac{x_{0,75} - 175}{10} \Longrightarrow \\
x_{0,75} &= 181{,}75 \\
z_{0,25} &= -0{,}675 \\
&= \frac{x_{0,25} - 175}{10} \Longrightarrow \\
x_{0,25} &= 168,25
\end{aligned}
$$

por lo que el intervalo es [168.5;181.75] de los dos resultados escogemos el segundo por ser de menor longitud y ser simétrico con respecto a la media

5.3 Distribución Gamma

Muchos problemas de estadística se resuelven con la distribución Normal como estudiaremos a partir del próximo capítulo, pero en ocasiones los problemas nos indican que la distribución de la variable de interés no es acampanada y simétrica , por lo que tenemos que usar distribuciones sesgadas y estas son precisamente las familias de distribuciones Gamma y Beta.

Antes de definir la distribución Gamma definiremos una función muy utilizada en matemáticas llamada función Gamma y de la cual la distribución Gamma toma su nombre.

Definición 5.4

[Función Gamma]Para cualquier positivo α, el valor $\Gamma(\alpha)$ se define por la integral

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \tag{5.1}$$

se puede demostrar que esta integral existe para cualquier $\alpha > 0$

5.3.1. Propiedades de la función Gamma

Teorema 5.2

Si $\alpha > 1$, entonces $\Gamma(\alpha) = (\alpha - 1)\,\Gamma(\alpha - 1)$

Lema 5.5

Si n es un entero positivo se tiene que $\Gamma(n) = (n-1)!$, $\Gamma\left(n+\frac{1}{2}\right) = \left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\left(n-\frac{5}{2}\right)\cdots\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)$

Lema 5.6

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

Definición 5.7

[Distribución Gamma]Sea X una variable aleatoria continua se dice que tiene una distribución Gamma Si su $f.d.p$ es

$$f(x) = \begin{cases} \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{para } x > 0 \\ 0 & \text{para } x \leq 0 \end{cases} \tag{5.2}$$

Calculemos la esperanza y la varianza de XAntes de seguir con esta distribución definiremos una función llamada generadora de momentos

Definición 5.8

Sea X una variable aleatoria la función generadora de momentos se define

$$\psi(t) = E\left(e^{tX}\right), \text{ para cada real } t \tag{5.3}$$

resulta que $\psi'(0) = E(X)$, y en general $\psi'(t)$ existe para todos los valores de t en un intervalo alrededor de $t=0$ y además

$$\psi^{(n)}(0) = \left[\frac{d^n}{dt^n} E\left(e^{tX}\right)\right]_{t=0} = E(X^n) \tag{5.4}$$

De esta forma la función generadora de momentos para nuestro caso sería

$$\psi(t) = \int_0^{\infty} e^{tx} f(x)\, dx = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\beta-t)^{\alpha}} = \beta^{\alpha}(\beta-t)^{-\alpha}, \quad t < \beta \tag{5.5}$$

Determinemos

$$\begin{aligned} E(X^n) &= \int_0^{\infty} x^n f(x)\, dx = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha+n-1} e^{-\beta x} dx \\ &= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma(\alpha+n)}{\beta^{\alpha+n}} = \frac{\Gamma(\alpha+n)}{\beta^n \Gamma(\alpha)} \\ &\frac{\alpha(\alpha+1)(\alpha+2)\cdots(\alpha+n-1)}{\beta^n} \end{aligned} \tag{5.6}$$

es decir

$$\begin{aligned} E(X) &= \frac{\alpha}{\beta}, \quad E\left(X^2\right) = \frac{\alpha(\alpha+1)}{\beta^2} \\ V(X) &= \frac{\alpha(\alpha+1)}{\beta^2} - \left(\frac{\alpha}{\beta}\right)^2 = \frac{\alpha}{\beta^2} \end{aligned} \tag{5.7}$$

Si X tiene una distribución Gamma se denota $X \rightsquigarrow \Gamma\left(\frac{\alpha}{\beta}, \frac{\alpha}{\beta^2}\right)$

Teorema 5.3

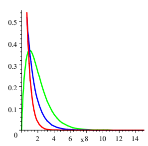
Si las variables aleatorias $X_1, X_2, \cdots X_k$ son independientes y $X_i \rightsquigarrow \Gamma\left(\frac{\alpha_i}{\beta}, \frac{\alpha_i}{\beta^2}\right)$, para todo $i = 1, 2, 3, \cdots, k$ entonces $\sum_{i=1}^{k} X_i \rightsquigarrow \Gamma\left(\frac{\sum_{i=1}^{k}\alpha_i}{\beta}, \frac{\sum_{i=1}^{k}\alpha_i}{\beta^2}\right)$

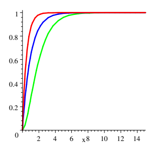
ahora estamos en condiciones de determinar la $f.d.a$ de X

$$F(x) = \begin{cases} 1 - \int_x^{\infty} \frac{\lambda}{\Gamma(\alpha)} (\lambda t)^{\alpha-1} e^{-\lambda t} dt & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{5.8}$$

si α es un entero positivo esta integral puede integrarse por partes obteniendo

$$F(x) = 1 - \sum_{k=0}^{\alpha-1} \frac{e^{-\lambda x} (\lambda x)^k}{k!}, \quad x > 0 \tag{5.9}$$


$f.d.p$ de la Gamma


$f.d.a$ de la Gamma

que es la suma de los términos de Poisson con media λx con lo que nos queda que podemos utilizar la distribución de Poisson para calcular la distribución Gamma $f(y)$ haciendo $y = \lambda x, \ \beta = \lambda^{-1}$

Ejemplo: 5.3.15

Sea X una variable aleatoria tal que $X \rightsquigarrow Exp\left(\frac{1}{\lambda}, \frac{1}{\lambda^2}\right)$, esta variable se puede representar con una Gamma haciendo $\alpha = 1, \beta = \lambda^{-1}$

Ejemplo: 5.3.16

Un sistema de control opera como muestra la figura . al principio la unidad 1 está en línea, en tanto que las unidades 2 , 3 y 4 están en espera. Cuando la unidad 1 falla entra en funcionamiento la unidad 2 hasta que falla y entonces se activara la unidad 3, cuando esta falla entre en funcionamiento la unidad 4 el interruptor de decisión se supone perfecto, por lo que la vida del sistema puede representarse como la suma de las vidas de los subsistemas. Así $X = X_1 + X_2 + X_3 + X_4$. Si la vidas de los subsistemas son independientes entre sí, y cada subsistema tiene una vida $X_j, \ j = 1, 2, 3, 4$, con $f.d.p$

$$g(x_j) = \begin{cases} \frac{1}{1000} e^{-\frac{x_j}{1000}} & x_j \geq 0 \\ 0 & x_j < 0 \end{cases} \tag{5.10}$$

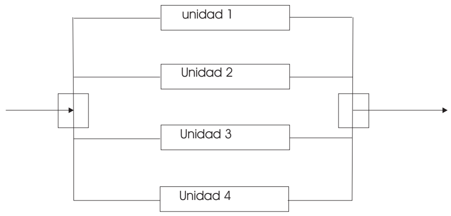
Calcule la probabilidad de que el sistema operará al menos x horas

Solución. Entonces X tendrá una $f.d.p$ Gamma con $\alpha = 4$ y $\beta = 0{,}001^{-1}$ de acuerdo con el teorema5.5 y el ejemplo anterior, decir

$$f(x) = \begin{cases} \frac{0{,}001}{3!} (0{,}001x)^3 e^{-0{,}01x} & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{5.11}$$

por lo que nos queda

$$P\left(\infty > x\right) = \sum_{k=0}^{3} e^{,0{,}001x}\left(0{,}001x\right)^{k}/k! \tag{5.12}$$



Ejemplo: 5.3.17

Suponga que el tiempo de reacción X a cierto estímulo en un individuo seleccionado al azar tiene una distribución estándar $(\beta = 1)$ con $\alpha = 2$s calcular la probabilidad de que el tiempo de reacción sea mayor de 4

Solución entonces tenemos que

$$P\left(X > 4\right) = \sum_{k=0}^{1} \frac{e^{-4}\left(4\right)^{k}}{k!} = 5e^{-4} = 9.\,157\,8 \times 10^{-2} \tag{5.13}$$

Ejercicio. 5.3.1

S uponga que el tiempo X de supervivencia en semanas, de un ratón macho seleccionado al azar y expuesto a 240 rads de radiación gamma, tiene una distribución Gamma con $\alpha = 8$ y $\beta = 15$. El tiempo esperado de supervivencia es $E\left(X\right) = \left(8\right)\left(15\right) = 120$ semanas y la varianza es $V\left(X\right) = \left(8\right)\left(15\right)^{2}\ semanas^{2}$. Cuál es la probabilidad de que un ratón sobreviva entre 60 y 120 semanas

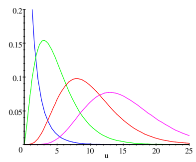
Solución tenemos que calcular

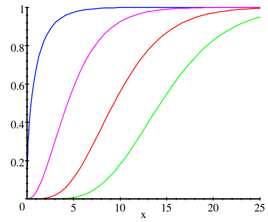
$$\begin{aligned} P\left(60 < X < 120\right) &= P(X < 120) - P\left(X < 60\right) \\ &= -\sum_{k=0}^{7} \frac{e^{-(120/15)}\left(120/15\right)^{k}}{k!} + \sum_{k=0}^{7} \frac{e^{-(60/15)}\left(60/15\right)^{k}}{k!} = 0{,}495\,91 \end{aligned} \tag{5.14}$$

Ejercicio. 5.3.2

p ara una distribución gamma con $\beta = 2$ y $\alpha = \frac{v}{2}$, donde v es un entero positivo está distribución así definida se le denomina chi-cuadrada y se denota χ^{2}, y su $f.d.p$ es

$$f\left(X^{2}\right) = \begin{cases} \dfrac{1}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)} x^{\frac{v}{2}-1} e^{-\frac{x^{2}}{2}} & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{5.15}$$





5.4 Distribución Gamma

Muchos problemas de estadística se resuelven con la distribución Normal como estudiaremos a partir del próximo capítulo, pero en ocasiones los problemas nos indican que la distribución de la variable de interés no es acampanada y simétrica , por lo que tenemos que usar distribuciones sesgadas y estas son precisamente las familias de distribuciones Gamma y Beta.

Antes de definir la distribución Gamma definiremos una función muy utilizada en matemáticas llamada función Gamma y de la cual la distribución Gamma toma su nombre.

Definición 5.9

[Función Gamma]Para cualquier positivo α, el valor $\Gamma(\alpha)$ se define por la integral

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx, \tag{5.16}$$

se puede demostrar que esta integral existe para cualquier $\alpha > 0$

5.4.1. Propiedades de la función Gamma

Teorema 5.4

Si $\alpha > 1$, entonces $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$

Lema 5.10

Si n es un entero positivo se tiene que $\Gamma(n) = (n-1)!$, $\Gamma\left(n + \frac{1}{2}\right) = \left(n - \frac{1}{2}\right)\left(n - \frac{3}{2}\right)\left(n - \frac{5}{2}\right)\cdots\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)$

Lema 5.11

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

Definición 5.12

[Distribución Gamma]Sea X una variable aleatoria continua se dice que tiene una distribución Gamma Si su $f.d.p$ es

$$f(x) = \begin{cases} \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{para } x > 0 \\ 0 & \text{para } x \leq 0 \end{cases} \tag{5.17}$$

Calculemos la esperanza y la varianza de XAntes de seguir con esta distribución definiremos una función llamada generadora de momentos

Definición 5.13

Sea X una variable aleatoria la función generadora de momentos se define

$$\psi(t) = E\left(e^{tX}\right), \text{ para cada real } t \tag{5.18}$$

resulta que $\psi'(0) = E(X)$, y en general $\psi'(t)$ existe para todos los valores de t en un intervalo alrededor de $t = 0$ y además

$$\psi^{(n)}(0) = \left[\frac{d^n}{dt^n} E\left(e^{tX}\right)\right]_{t=0} = E(X^n) \tag{5.19}$$

De esta forma la función generadora de momentos para nuestro caso sería

$$\psi(t) = \int_0^\infty e^{tx} f(x)\, dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\beta - t)^\alpha} = \beta^\alpha (\beta - t)^{-\alpha}, \quad t < \beta \tag{5.20}$$

Determinemos

$$\begin{aligned} E(X^n) &= \int_0^\infty x^n f(x)\, dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha+n-1} e^{-\beta x} dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+n)}{\beta^{\alpha+n}} = \frac{\Gamma(\alpha+n)}{\beta^n \Gamma(\alpha)} \\ &\frac{\alpha(\alpha+1)(\alpha+2)\cdots(\alpha+n-1)}{\beta^n} \end{aligned} \tag{5.21}$$

es decir

$$\begin{aligned} E(X) &= \frac{\alpha}{\beta}, \quad E\left(X^2\right) = \frac{\alpha(\alpha+1)}{\beta^2} \\ V(X) &= \frac{\alpha(\alpha+1)}{\beta^2} - \left(\frac{\alpha}{\beta}\right)^2 = \frac{\alpha}{\beta^2} \end{aligned} \tag{5.22}$$

Si X tiene una distribución Gamma se denota $X \rightsquigarrow \Gamma\left(\frac{\alpha}{\beta}, \frac{\alpha}{\beta^2}\right)$

Teorema 5.5

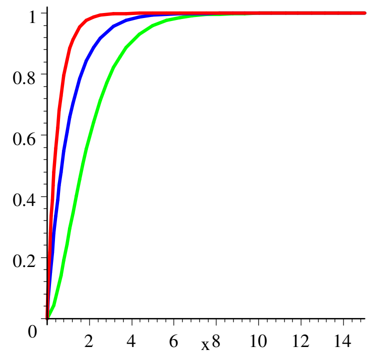
Si las variables aleatorias $X_1, X_2, \cdots X_k$ son independientes y $X_i \rightsquigarrow \Gamma\left(\frac{\alpha_i}{\beta}, \frac{\alpha_i}{\beta^2}\right)$, para todo $i = 1, 2, 3, \cdots, k$ entonces $\sum_{i=1}^{k} X_i \rightsquigarrow \Gamma\left(\frac{\sum_{i=1}^{k}\alpha_i}{\beta}, \frac{\sum_{i=1}^{k}\alpha_i}{\beta^2}\right)$

ahora estamos en condiciones de determinar la $f.d.a$ de X

$$F(x) = \begin{cases} 1 - \int_x^\infty \frac{\lambda}{\Gamma(\alpha)} (\lambda t)^{\alpha-1} e^{-\lambda t} dt & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{5.23}$$

si α es un entero positivo esta integral puede integrarse por partes obteniendo

$$F(x) = 1 - \sum_{k=0}^{\alpha-1} \frac{e^{-\lambda x} (\lambda x)^k}{k!}, \quad x > 0 \tag{5.24}$$



$f.d.a \quad de \quad la$ Gamma

que es la suma de los términos de Poisson con media λx con lo que nos queda que podemos utilizar la distribución de Poisson para calcular la distribución Gamma $f(y)$ haciendo $y = \lambda x, \;\; \beta = \lambda^{-1}$

Ejemplo: 5.4.20

Sea X una variable aleatoria tal que $X \rightsquigarrow Exp\left(\frac{1}{\lambda}, \frac{1}{\lambda^2}\right)$, esta variable se puede representar con una Gamma haciendo $\alpha = 1, \beta = \lambda^{-1}$

Ejemplo: 5.4.21

Un sistema de control opera como muestra la figura . al principio la unidad 1 está en línea, en tanto que las unidades 2 , 3 y 4 están en espera. Cuando la unidad 1 falla entra en funcionamiento la unidad 2 hasta que falla y entonces se activara la unidad 3, cuando esta falla entre en funcionamiento la unidad 4 el interruptor de decisión se supone perfecto, por lo que la vida del sistema puede representarse como la suma de las vidas de los subsistemas. Así $X = X_1 + X_2 + X_3 + X_4$. Si la vidas de los subsistemas son independientes entre sí, y cada subsistema tiene una vida $X_j, \; j = 1, 2, 3, 4$, con $f.d.p$

$$g(x_j) = \begin{cases} \frac{1}{1000} e^{-\frac{x_j}{1000}} & x_j \geq 0 \\ 0 & x_j < 0 \end{cases} \tag{5.25}$$

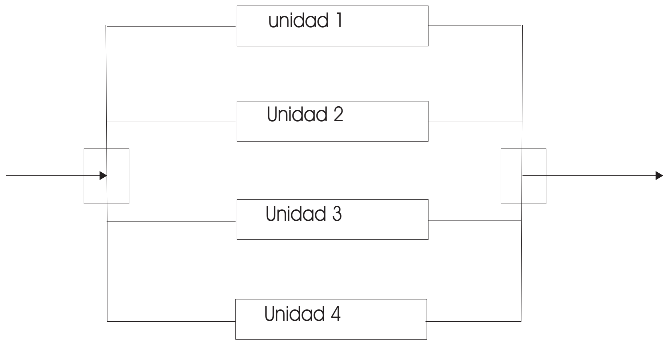
Calcule la probabilidad de que el sistema operará al menos x horas

Solución. Entonces X tendrá una $f.d.p$ Gamma con $\alpha = 4$ y $\beta = 0{,}001^{-1}$ de acuerdo con el teorema5.5 y el ejemplo anterior, decir

$$f(x) = \begin{cases} \frac{0{,}001}{3!} (0{,}001x)^3 e^{-0{,}01x} & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{5.26}$$

por lo que nos queda

$$P(\infty > x) = \sum_{k=0}^{3} e^{,0{,}001x} (0{,}001x)^k / k! \tag{5.27}$$



Ejemplo: 5.4.22

Suponga que el tiempo de reacción X a cierto estímulo en un individuo seleccionado al azar tiene una distribución estándar $(\beta = 1)$ con $\alpha = 2$s calcular la probabilidad de que el tiempo de reacción sea mayor de 4

Solución entonces tenemos que

$$P\left(X > 4\right) = \sum_{k=0}^{1} \frac{e^{-4}\left(4\right)^{k}}{k!} = 5e^{-4} = 9.1578 \times 10^{-2} \tag{5.28}$$

Ejercicio. 5.4.3

S uponga que el tiempo X de supervivencia en semanas, de un ratón macho seleccionado al azar y expuesto a 240 rads de radiación gamma, tiene una distribución Gamma con $\alpha = 8$ y $\beta = 15$. El tiempo esperado de supervivencia es $E\left(X\right) = \left(8\right)\left(15\right) = 120$ semanas y la varianza es $V\left(X\right) = \left(8\right)\left(15\right)^{2}$ $semanas^{2}$. Cuál es la probabilidad de que un ratón sobreviva entre 60 y 120 semanas

Solución tenemos que calcular

$$\begin{aligned} P\left(60 < X < 120\right) &= P(X < 120) - P\left(X < 60\right) \\ &= -\sum_{k=0}^{7} \frac{e^{-(120/15)}\left(120/15\right)^{k}}{k!} + \sum_{k=0}^{7} \frac{e^{-(60/15)}\left(60/15\right)^{k}}{k!} = 0{,}49591 \end{aligned} \tag{5.29}$$

Ejercicio. 5.4.4

p ara una distribución gamma con $\beta = 2$ y $\alpha = \frac{v}{2}$, donde v es un entero positivo está distribución así definida se le denomina chi-cuadrada y se denota χ^{2}, y su $f.d.p$ es

$$f\left(X^{2}\right) = \begin{cases} \dfrac{1}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)} x^{\frac{v}{2}-1} e^{-\frac{x^{2}}{2}} & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{5.30}$$

5.5 Problemas

1. La presión de aire de un neumático seleccionado al azar, instalado en un automóvil nuevo, está normalmente distribuida con valor medio de 31 Ib/pulg^2 y desviación estándar de 2 Ib/pulg^2

   a) ¿Cuál es la probabilidad de que la presión de un neumático seleccionado al azar exceda de 30.5 lb/pulg^2?

   b) ¿Cuál es la probabilidad de que la presión de un neumático seleccionado al azar se encuentre entre 30.5 y 31.5 Ib/pulg^2? y ¿entre 30 y 32 Ib/pulg^2?

   c) Suponga que un neumático se considera con presión baja si su presión está debajo de 30.4 Ib/pulg^2. ¿Cuál es la probabilidad de que al menos uno de los cuatro neumáticos de un automóvil se encuentre desinflado?(SUGERENCIA Si A ={al menos un neumático está desinflado}, ¿cuál es el complemento de A?)

2. La dureza Rockwell de un metal se determina al golpear con un punto acerado (herramienta) la superficie del metal y luego se mide la profundidad de penetración del punto Suponga que la dureza Rockwell de cierta aleación está normalmente distribuida con media de 70 y desviación estándar de 3 (suponga que la dureza Rockwell se mide en una escala continua)

   a) Si un espécimen es aceptable sólo si su dureza está entre 67 y 75, ¿cuál es la probabilidad de que un espécimen seleccionado al azar tenga una dureza aceptable ?

   b) Si la escala aceptable de dureza fue $(70 - c, 70 + c)$, ¿para qué valor de c tendría una dureza aceptable el 95 % de todos los especímenes?

   c) Si la escala aceptable es como en la parte (a) y la dureza de cada diez especímenes seleccionados al azar se determina independientemente, ¿cuál es el número esperado de especímenes aceptable entre los diez?

   d) ¿Cuál es la probabilidad de que a lo sumo ocho de diez especímenes seleccionados independientemente tengan una dureza menor de 73.84 (SUGERENCIA Y = el número entre los diez especímenes con dureza menor de 73.84 es una variable binomial, ¿cuál es p?

3. La distribución del peso de paquetes enviados de cierto modo es normal con valor medio de 10 libras y desviación estándar de 2 libras. El servicio de paquetería desea establecer un valor de peso c, más allá del cual habrá cargo extra ¿Cuál valor de c es tal que 99 % de todos los paquetes pesen por lo menos 1 libra abajo del peso con cargo extra?

4. Suponga que la tabla de la distribución normal del Apéndice contiene Φ(z) sólo para z >0 Explique cómo se podría calcular todavía

   a) $P(-1{,}72 \leq Z \leq -0{,}55)$.

   b) $P(-1{,}72 \leq Z \leq 0{,}55)$.
   ¿es necesario poner Φ(z) en la tabla para z negativa?, ¿que propiedad de la curva normal estándar justifica su respuesta?

5. Represente por X el número de páginas de texto de una tesis de doctorado en matemáticas seleccionado al azar. Aun cuando X puede tomar sólo valores enteros positivos, suponga que está distribuida normalmente en forma aproximada con valor esperado de 90 y desviación estándar de 15 ¿Cuál es la probabilidad de que una tesis seleccionada al azar contenga

   a) A lo sumo 100 páginas ?

   b) Entre 80 y 110 páginas ?

6. Haga que X tenga una distribución binomial con parámetros $n = 25$ y p calcule cada una de las siguientes probabilidades para los casos $p = 0{,}5, 0{,}6$ y.8

   a) $P(15 < X < ,20)$

   b) $P(X \leq 15)$

   c) $P(20 < X)$

7. Suponga que el 10 % de todos los ejes de acero producidos por cierto proceso están fuera de especificaciones, pero se pueden volver a trabajar (en lugar de tener que enviarlos la chatarra) Considere una muestra aleatoria de 200 ejes y denote por X el número entre ellos que estén fuera de especificaciones y se puedan volver a trabajar ¿Cuál es la probabilidad (aproximada) de que X sea

   a) A lo sumo 30?

   b) Menosde30?

   c) Entre 15 y 25 (inclusive)?

8. Suponga que sólo el 40 % de todos los automovilistas de cierto estado usan con regularidad su cinturón de seguridad Se selecciona al azar una muestra de 500 automovilistas ¿Cuál es la probabilidad de que X este
   entre 180 y 230 (inclusive) de los automovilistas de la muestra utilice su cinturón con regularidad?

9. Si X es una *va* normal con media 80 y desviación estándar 10, calcule las siguientes probabilidades mediante estandanzación

   a) $P(X \leq 80)$

   b) $P(X < 100)$

   c) $P(65 < X < 100)$

   d) $P(70 < X)$

   e) $P(|X - 80| < 10)$

   f) $P(85 < X \leq 95)$

10. Suponga que la fuerza que actúa sobre una columna que ayuda a sostener un edificio está normalmente distribuida con media de 15.0 kips y desviación estándar de 1.25 kips ¿Cuál es la probabilidad de que la fuerza

    a) Sea a lo sumo 17 kips?

    b) Se encuentre entre 12 y 17 kips?

    c) Difiera de 15.0 kips en a lo sumo 2 desviaciones estándar?

11. Suponga que el tiempo de revelado para un tipo particular de papel fotográfico cuando se expone a una fuente luminosa durante 5 s está normalmente distribuido con una media de 25 s y una desviación estándar de 1.3 s ¿Cuál es la probabilidad de que

    a) Una impresión en particular necesite más de 26.5 s para revelarse?

    b) El tiempo de revelado sea por lo menos 23 s?

    c) El tiempo de revelado difiera del tiempo esperado en más de 2.5 s?

12. Un tipo particular de tanque de gasolina para un automóvil compacto está diseñado para contener 15 galones Suponga que la capacidad real X de un tanque escogido al

azar de este tipo esté normalmente distribuido con media de 15 galones y desviación estándar de 2 galones

a) ¿Cuál es la probabilidad de que un tanque seleccionado al azar contenga a lo sumo 14.8 galones?

b) ¿Cuál es la probabilidad de que un tanque seleccionado al azar contenga entre 14.7 y 15.1 galones?

c) Si el automóvil en el que se instala un tanque seleccionado al azar recorre exactamente 25 millas por galón, ¿cuál es la probabilidad de que el automóvil pueda recorrer 370 millas sin reabastecerse?

13. Hay dos máquinas para cortar corchos destinados para usarse en botellas de vino La primera produce corchos con diámetros que están normalmente distribuidos con media de 3 cm y desviación estándar de 1 cm La segunda máquina produce corchos con diámetros que tienen una distribución normal con media de 3.04 cm y desviación estándar de 0.02 cm Los corchos aceptables tienen diámetros entre 2.9 cm y 3.1 cm ¿Cuál máquina tiene más probabilidad de producir un corcho aceptable?

a) Si una distribución normal tiene $\mu = 25$ y $\sigma = 5$, ¿cuál es el 91avo percentil de la distribución?

b) ¿Cuál es el sexto percentil de la distribución de la parte (a)?

c) El ancho de una linea grabada en un circuito integrado esta normalmente distribuido con media de 3.000 μm y desviación estándar de 0.150 ¿Que valor separa al 10 % mas ancho de todas las líneas del otro 90 %?

14. El dispositivo automático de apertura de un paracaídas militar de carga se ha diseñado para abrir el paracaídas cuando éste se encuentre a 200 m de altura sobre el suelo Supongamos que la altitud de apertura en realidad tiene una distribución normal con valor medio de 200 m y desviación estándar de 30 m Habrá daño al equipo si el paracaídas se abre a una altitud de menos de 100 m Cuál es la probabilidad de que haya daño a la carga en al menos uno de cinco paracaídas lanzado independientemente?

15. La lectura de temperatura de un termopar puesto en un medio de temperatura constante esta normalmente distribuida con media μ, la temperatura real del entorno, y desviación estándar σ ¿Cual tendría que ser el valor de σ para asegurar que el 95 % de todas las lecturas se encuentren dentro de 0.1^o de μ?

16. Se sabe que la distribución de resistencia para resistores de cierto tipo es normal, 10 % de todos los resistores tienen una resistencia que excede los 10.256 ohms y 5 % tiene una resistencia menor de 9.671 ohms ¿Cuales son los valores de la media y de la desviación estándar de la distribución de resistencia?

17. Si el diámetro de un cojinete está normalmente distribuido, ¿cuál es la probabilidad de que el diámetro de un cojinete seleccionado al azar esté

a) Dentro de 1.5σ de su valor medio?

b) A más de 2 5σ de su valor medio?

c) Entre 1 y 2 σ de su valor medio?

18. Suponga que sólo 20 % de los automovilistas se detienen por completo en un crucero, donde hay un semáforo con luz roja intermitente en todas las direcciones cuando no haya otros automóviles visibles ¿Cuál es la probabilidad de que, de 20 automovilistas seleccionados al azar lleguen al crucero en estas condiciones,

a) a lo sumo 5 se detengan por completo?

b) exactamente 5 se detengan por completo?

c) por lo menos 5 se detengan por completo?

d) ¿Cuántos, de los siguientes 20 automovilistas, espera el lector que se detengan por completo?

19. Si el 90 % de todos los solicitantes para cierto tipo de hipoteca no llenan correctamente el formato de solicitud en la primera remisión, ¿cuál es la probabilidad de que entre 15 de estos solicitantes seleccionados al azar

a) por lo menos 12 no la llenen a la primera remisión?

b) entre 10 y 13 inclusive no la llenen a la primera remisión?

c) a lo sumo 2 llenen correctamente sus formatos antesde la remisión inicial

20. Un tipo particular de raqueta de tenis se fabrica en tamaños mediano y extragrande El 60 % de todos los clientes de cierta tienda buscan el tamaño extragrande

a) Entre diez clientes seleccionados al azar, que desean este tipo de raqueta, ¿cuál es la probabilidad de que por lo menos seis busquen el tamaflo extragrande?

b) Entre diez clientes seleccionados al azar, ¿cuál es la probabilidad de que el número que buscan el tamaño extragrande esté dentro de 1 σ del valor medio?

c) La tienda tiene actualmente siete raquetas de cada modelo ¿Cuál es la probabilidad de que los siguientes diez clientes que buscan esta raqueta puedan comprar el modelo que buscan, de entre la existencia actual

21. Veinte por ciento de todos los teléfonos de cierto tipo se remiten para repararse cuando todavía está vigente su garantía De éstos, 60 % se pueden reparar y el otro 40 % debe sustituirse con aparatos nuevos Si una compaflía compra diez de estos teléfonos, (,cuál es la probabilidad de que exactamente se cambien dos dentro del periodo de garantía

22. Un lote muy grande de componentes ha llegado a un distribuidor El lote se puede clasificar como aceptable sólo si la proporción de componentes defectuosos es a lo sumo 10. El distribuidor determina seleccionar al azar 10 componentes y aceptar el lote sólo si el número de componentes defectuosos en la muestra es a lo sumo 2.

a) ¿Cuál es la probabilidad de que el lote sea aceptado cuando la proporción real de piezas defectuosas es 0.1, 0.05, 0.10, 0.20 y 0.25 ?

b) Denotemos por p la proporción real de piezas defectuosas del lote Una gráfica de P(lote aceptado) como función de p, con p en el horizontal y P(lote aceptado) en el vertical, se llama *curva característica de operación* para el plan de muestreo de aceptación del lote. Utilice los resultados de la parte (a) para trazar esta curva para $0 \leq p \leq 1$.

c) Repita las partes (a) y (b) con 1 sustituyendo a 2 en el plan de muestreo de aceptación del lote

d) Repita las partes (a) y (b) con 15 sustituyendo a 10 en el plan de muestreo de aceptación del lote.

e) ¿Cual de los tres planes de muestreo, el de la parte (a) (c) o (d) parece mas satisfactorio, y por que?

23. Un reglamento. que requiere que un detector de humo (pueda instalar en todas las casas prefabricadas ha estado en vigor en una ciudad durante un año. El departamento de bomberos está preocupado porque muchas casas siguen sin detectores. Sea $p =$ la verdadera proporción de las casas que tienen detectores y supongamos que se inspecciona al azar una muestra de 25 casas. Si la muestra indica fuertemente que menos del 80 % tienen detector el departamento de bomberos hará una campaña para que el programa de inspecciones sea obligatorio Debido a lo costoso del programa, el departamento prefiere no pedir tales inspecciones a menos que la evidencia muestral apoye con argumentos sólidos esta necesidad Denotenotemos por X el número de casas con detectores entre las 25 en que se hace muestreo Considere rechazar la afirmación de que $p \geq 0.8$ si $x < 15$, donde x es el valor observado de X.

    a) ¿Cuál es la probabilidad de que la petición sea rechazada cuando el valor real de p es 0.8

    b) ¿Cuál es la probabilidad de no rechazar la petición cuando p =0.7? o ¿.Cuándo $p = 6$?

    c) ¿Cómo cambian las ¡¡probabilidades de error¿¿de las partes (a) y (b) si el valor de 15 de la regla de decisión se sustituye por 14?

24. Un geólogo ha recolectado 10 especímenes de roca basáltica y 10 de granito Si el geólogo instruye a un asistente de laboratorio para que seleccione al azar 15 de los especímenes para analizarlos,¿cuál es la f.d.p del número de especímenes de basalto seleccionados para analizarlos? Cuál es la probabilidad de que todos los especímenes de uno de los dos tipos de roca sean seleccionados para análisis?

25. Un director de personal que entrevista a 11 ingenieros para cuatro vacantes ha programado seis entrevistas para el primer día y cinco para el segundo día de entrevistas. Suponga que los candidatos son entrevistados al azar

    a) ¿Cuál es la probabilidad de que x de los mejores cuatro candidatos sean entrevistados el primer día?

    b) ¿Cuántos de los mejores cuatro candidatos pueden esperar ser entrevistados el primer día?

26. Veinte parejas de individuos que juegan un torneo de *bridge* han sido sembrados como $1, \cdots, 20$ En la primera parte del torneo, los 20 se dividen al azar en 10 parejas este-oeste y 10 parejas norte-sur

    a) ¿Cuál es la probabilidad de que x de las mejores 10 parejas terminen Jugando este-oeste?

    b) ¿Cuál es la probabilidad de que todos los de las mejores cinco parejas terminen Jugando en la misma dirección?

    c) Si hay *2n* parejas, (,cuál es la f.d.p de $X =$ el número entre las mejores n parejas que terminan Jugando este-oeste, y cuáles son $E(X)$ y V(X)

27. Ha sido puesta en vigor la segunda etapa de una alerta de *smog* en cierta zona del condado de Los Ángeles, en la que hay 50 empresas industriales Un inspector visitará 10 de ellas seleccionadas al azar para verificar las violaciones de los reglamentos

    a) Si 15 de las empresas violan en realidad al menos un reglamento, ¿cuál es la f.d.p del número de empresas visitadas por el inspector que están violando al menos un reglamento?

    b) Si hay 500 empresas en la zona, de las cuales 150 están en violación, aproxime la f.d.p de la parte (a) por una f.d.p más sencilla

    c) Para $X =$ el número entre las 10 visitadas que estén en violación, calcule *E(X) y V(X)* tanto para la f.d.p exacta como para la f.d.p aproximada de la parte (b)

28. Suponga que $p =$ P(nace niño) =0. 5. Una pareja desea tener exactamente dos niñas en su familia Tendrán hijos hasta que se satisfaga esta condición.

    a) ¿Cuál es la probabilidad de que la familia tenga-x hijos(hombres)?

    b) ¿Cuál es la probabilidad de que la familia tenga cuatro hijos?

    c) ¿Cuál es la probabilidad de que la familia tenga a lo sumo cuatro hijos?

    d) ¿Cuántos hijos (hombres) se esperaría que tenga esta familia?, ¿Cuántos hijos se esperara que tenga esta **familia?**

29. Una familia decide tener hijos hasta que tenga tres del mismo sexo Si se supone que $P(B) =$ *P[G) =0.5* ¿cual es la f.d.p de $X =$ número de hijos de la familia?

30. Tres hermanos y sus esposas deciden tener hijos hasta que cada familia tenga dos niñas ¿Cuál es la f.d.p de $X =$ el número total de niños (hombres) nacidos de los hermanos?. ¿Cuál es $E(X)$ y cómo se compara con el número esperado de hijos (hombres) nacidos de cada hermano?

31. El individuo A tiene un dado rojo y B tiene un dado verde (ambos no cargados) Si cada uno de ellos tira su dado hasta obtener ¡¡dobles¿¿ (1 -1, ..., 6-6), (,cuál es la f.d.p de $X =$ el número total de veces que se tira un dado? ¿Cuáles son $E(X)$ y V(X)

32. Haga que X tenga una distnbución de Poisson con parámetro λ= 5 calcule las siguientes probabilidades

    a) $P(X < 8)$

    b) $P(X = 8)$

    c) $P(9 \geq X)$

    d) $P(5 < X \leq 8)$

    e) $P(5 < X < 8)$

33. Suponga que el número X de tornados observados en una región particular durante un periodo de 1 año tiene una distnbución de Poisson con λ= 8.

    a) Calcule $P(X < 5)$

    b) Calcule $P(6 < X < 9)$.

    c) Calcule $P(10 \geq X)$

    d) ¿Cuántos tornados se puede esperar que se observen durante un periodo de un año, y cuál es la desviación estándar del número de tornados observados?

34. Sea $X =$ el número de automóviles de un año y modelo particular que en algún momento en el futuro sufrirán una falla grave en el mecanismo de dirección, que ocasionará pérdida completa de control a alta velocidad Suponga que X tiene una distribución de Poisson con parámetro λ= 10

    a) ¿Cuál es la probabilidad de que a lo sumo 10 automóviles sufran dicha falla?

b) ¿Cuál es la probabilidad de que entre 10 y 15 (inclusive) automóviles sufran dicha falla

c) ¿Cuáles son $E(X)$ y $V(X)$?

35. Considere escribir en un disco de computadora y luego enviar el escrito por un certificador que cuenta **el** número de pulsos faltantes Suponga que este número X tiene una distnbución de Poisson con parámetro λ= 2. (Sugendo en ¡¡Average Sample Number for SermCunailed Sampling Using the Poisson Distribution¿¿, *J Quahty Technology,* 1983, pp 126-129)

    a) ¿Cuál es la probabilidad de que un disco tenga exactamente un pulso faltante?

    b) ¿Cuál es la probabilidad de que un disco tenga al menos dos pulsos faltantes?

    c) Si dos discos se seleccionan independientemente, ¿cuál es la probabilidad de que ninguno contenga un pulso faltante ?

36. Un artículo de *Los Angeles Times* (3 de diciembre de 1993) reporta que. de cada 200 personas, una lleva el gene defectuoso que ocasiona cáncer de colon hereditario En una muestra de 1 000 personas, ¿cuál es la distribución aproximada del número de quienes llevan este gene?.Utilice esta distribución para calcular la probabilidad aproximada de que

    a) Entre 4 y 7 (inclusive) lleven el gene

    b) Por lo menos 8 lleven el gene.

37. Se envía un aviso a todos los propietarios de cierto tipo de automóvil, solicitándoles llevar su automóvil a un distribuidor para comprobar la presencia de un defecto particular de fabricación Suponga que sólo 0.05 % de tales automóviles tienen el defecto Considere una muestra aleatoria de 10 000 automóviles.

    a) ¿Cuáles son el valor esperado y la desviación estándar del número de automóviles de la muestra que tiene el defecto?

    b) ¿Cuál es la probabilidad (aproximada) de que por lo menos 10 automóviles en los que se efectuó el muestreo tengan el defecto ?

    c) ¿Cuál es la probabilidad (aproximada) de que ninguno de los automóviles en los que se efectuó el muestreo tengan el defecto?

38. Suponga que aviones pequeños llegan a cierto aeropuerto según un proceso de Poisson, con tasa $\lambda = 8$ aviones por hora, de modo que el número de llegadas durante un periodo de t horas es una *va* de Poisson con parámetro c = 8t

    a) ¿Cuál es la probabilidad de que exactamente 5 aviones pequeños lleguen durante un periodo de una hora? Por lo menos 5?, ¿Por lo menos 10?

    b) ¿Cuáles son el valor esperado y la desviación estándar del número de aviones pequeños que lleguen durante un periodo de 90 min?

    c) ¿Cuál es la probabilidad de que por lo menos 20 aviones pequeños lleguen durante un periodo de 2 1/2 h (,De que a lo sumo 10 lleguen durante este periodo?

39. El número de infracciones expedidas por el lector de un parquímetro, por violaciones de estacionamiento. puede modelarse mediante un proceso de Poisson con una tasa de cinco infracciones por hora. ¿Cuál es la probabilidad de que exactamente cinco infracciones se expidan durante una hora particular?

40. Un experimento consta de cuatro ensayos de Bernoulli independientes con probabilidad de éxitos en cada uno de ellos. La variable aleatoria A'es el número de éxitos. Enumere la distribución de probabilidad de X.

41. Se planean seis misiones espaciales independientes a la luna. La probabilidad estimada de éxito de cada misión es 0.95. ¿Cuál es la probabilidad de que al menos cinco de las misiones planeadas tengan éxito?

42. La Compañia XYZ ha planeado presentaciones de ventas a una docena de clientes importantes. La probabilidad de recibir un pedido como un resultado de tal presentación se estima en 0.5. ¿Cuál es la probabilidad de recibir cuatro o más pedidos como resultado de las reuniones?

43. Un corredor de bolsa llama a sus 20 más importantes clientes cada mañana. Si la probabilidad de que efectúe una transacción como resultado de dichas llamadas es de uno a tres, ¿cuáles son las posibilidades de que maneje 10 o más transacciones?

44. ¿Un proceso de producción que manufactura transistores genera, en promedio, una fracción de 2 por ciento de piezas defectuosas. Cada dos horas se toma del proceso una muestra aleatoria de tamaño 50. Si la muestra contiene más de dos piezas defectuosas el proceso debe interrumpirse. Determine la probabilidad de que el proceso será interrumpido por medio del esquema de muestreo indicado.

45. Se sabe que el proceso de producción de luces de un tablero de automóvil de indicador giratorio produce uno por ciento de luces defectuosas. Si este valor permanece invariable, y se selecciona al azar una muestra de 100 luces, encuentre $P(p\leq \mathbf{0},03)$, donde p es la fracción de defectos de la muestra.

46. Suponga que una muestra aleatoria de tamaño 200 se toma de un proceso que tiene una fracción de defectos de 0.07. Cuál es la probabilidad de que p exceda la fracción verdadera de defectos en una desviación estándar?, ¿En dos desviaciones estándar? y ¿,En tres desviaciones estándar?

47. Un agente de bienes raíces estima que la probabilidad de vender una casa es .10.1 día de hoy tiene que ver cuatro clientes Si tiene éxito en las primeras tres visita cual es la probabilidad de que su cuarta visita no sea exitosa?

48. Suponga que se van a realizar cinco experimentos de laboratorio idénticos independientes Cada experimento es en extremo sensible a las condiciones ambientales, y solo hay una probabilidad p de que se tenmnará con éxito Grafique, como una función *de* la probabilidad de que el quinto experimento sea el primero que falle. Obtenga matemáticamente el valor de p que maximice la probabilidad de que el quinto ensayo sea el primer experimento no exitoso.

49. La probabilidad de que un submarino hunda un barco enemigo con un disparo de sin torpedos es 8 Si los disparos son independientes, determine la probabilidad de hundimiento dentro de los primeros dos disparos, y dentro de los primeros tres ,

50. En Atlanta la probabilidad de que ocurra una tormenta en cualquier dia durante la primavera es 050 Suponiendo independencia, ¿cual es la probabilidad de que la primera tormenta ocurra el cinco de abril Suponga que la primavera empieza el 1 de marzo ?

51. Una compradora recibe lotes pequeños ($N = 25$) de un dispositivo de alta precisión Pretende rechazar el lote el 95 por ciento de las veces si contienen hasta siete unidades defectuosas Suponga que la compradora decide que la presencia de una unidad defectuosa en la muestra, es suficiente para provocar el rechazo Qué tan grande debe ser el tamaño de su muestra?

52. Se estima que el número de automóviles que pasa por un cruce particular por hora es de 25 Obtenga la probabilidad de que menos de 10 vehículos crucen durante cualquier intervalo de una hora Suponga que el núiriero de vehículos sigue una distribución de Poisson.

53. Las llamadas llegan a un tablero de control telefónico de modo tal que el numero de llamadas por hora sigue una distnbución de Poisson con media 10 El equipo disponible puede manejar hasta 20 llamadas sin que se sobrecargue ¿Cuál es la probabilidad de que ocurra dicha sobrecarga?

54. El número de células de sangre por unidad cuadrada visible bajo el microscopio sigue una distribucion de Poisson con media 4 Encuentre la probabilidad de que más de 5 de tales células de sangre sean visibles para el observador.

55. Una compaflía grande de seguros ha descubierto que 2 por ciento de la población de Estados Unidos esta lesionada como resultado de algún tipo de accidente particular Esta compañía tiene 15,000 asegurados que están protegidos contra tal accidente Cuál es la probabilidad de que tres o menos reclamos se entablen en relación con esas pólizas de seguro durante el siguiente año?, ¿Cinco o mas reclamos?

56. Cuadrillas de mantenimiento llegan a un almacén de herramientas solicitando una pieza de repuesto particular de acuerdo con una distnbución de Poisson con parámetro λ= 2 Tres de estas piezas de repuesto por lo general se tienen disponibles Si ocurren mas de tres solicitudes, las cuadrillas deben desplazarse a una distancia considerable hasta los almacenes centrales

    a) En un f.d.a cualquiera, cuál es la probabilidad de que se realice un viaje a los almacenes centrales?

    b) Cual es la demanda media esperada para piezas de repuesto?

    c) ¿Cuantas piezas de repuesto deben transportarse si el almacen de herramientas tiene que dar servicio a las cuadnllas que llegan el 90 por ciento de las veces?

    d) Cuál es el numero esperado de cuadnllas atendidas dianamente en el almacén de herramientas?

    e) ¿Cuál es el numero esperado de cuadrillas que realizaran el viaje a los almacenes generales?

57. **En una ciudad cualquiera el** consumo diario de agua (en millones de litros) sigue aproximada-mente una distribución gamma con α= 2 y $\beta = 3$. Si la capacidad diaria para esta ciudad es de 9 millones dc litros de agua, ¿cuál es la probabilidad de que en un determinado día el suministro de agua sea inadecuado?

58. Suponga que el tiempo, en horas, que toma reparar una bomba es una variable aleatoria X que tiene una distribución gamma con parámetros $\alpha = 2$ y β= 1/2. ¿Cuál es la probabilidad de que en el siguiente servicio

    a) tome cuando mucho 1 hora reparar la bomba?

    b) al menos se requieran 2 horas para reparar la bomba?

59. En una ciudad cualquiera el consumo de energía eléctrica, en millones de kilowatts-hora, es una variable aleatoria X que tiene una distribución gamma con media μ= 6 y $\sigma^2 = 12$,

    a) Encuentre los valores de α y β.

    b) Encuentre la probabilidad de que en un determinado día el consumo de energía eléctrica se exceda por 12 kilowatts-hora.

60. El tiempo que transcurre antes de que una persona sea atendida en una cafetería es una variable aleatoria que tiene una distnbución exponencial con una media de 4 minutos. Cuál es la probabilidad de que una persona sea atendida antes dc que transcurran 3 minutos en al menos 4 de los 6 días siguientes?

Índice alfabético