

Matemáticas Aplicadas

Juliho Castillo Colmenares

Índice general

Capítulo 1. Matemáticas Discretas	5
1. Matrices	5
2. Teoría general de grafos	8
3. Digrafos	15
Capítulo 2. Probabilidad	21
1. Fundamentos de probabilidad	24
2. Probabilidad condicional	28
3. Análisis combinatorio	29
4. Variables aleatorias	34
5. Funciones de probabilidad discretas	34
6. Funciones de distribución para variables aleatorias discretas	34
7. Variable Aleatorias Continuas	36
8. Interpretación gráfica	37
9. Distribución conjunta de probabilidad	38
10. Variables Aleatorias Independientes	40
11. Distribución Condicional	42
12. Definición de Esperanza Matemática	44
13. Varianza y Desviación Estándar	47
14. Covarianza	48
15. La Distribución Binomial	50
16. Distribución Normal	52
17. Relación entre las distribuciones binomial y normal	56
18. La Distribución de Poisson	57
19. Relación entre las Distribuciones Binomiales y de Poisson	59
20. Distribución multinomial	60
21. Problemas Resueltos	60

Matemáticas Discretas

1. Matrices

Las matrices son arreglos rectangulares de números que nos ayudan a codificar información. Por ejemplo:

$$\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}$$

puede ser útil para codificar los coeficientes del sistema de ecuaciones:

$$\begin{cases} a_{1,1}x + a_{1,2}y = b_1 \\ a_{2,1}x + a_{2,2}y = b_2 \end{cases}$$

En general, una matriz tiene la forma

$$(A) \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ \vdots & & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}$$

Los subíndices de cada elemento $a_{i,j}$ denotan la posición del mismo: i es el número del *renglón* (contando de arriba a abajo), mientras que j es el número de la columna (contando de izquierda a derecha).

Podemos extraer renglones y columnas de la matriz (A): El i -ésimo renglón es

$$R_i = (a_{i,1} \quad \cdots \quad a_{i,n})$$

mientras que la j -ésima columna será

$$C_j = \begin{pmatrix} a_{j,1} \\ \vdots \\ a_{j,m} \end{pmatrix}$$

Diremos que la matriz (A) tiene dimensión $m \times n$.

Si existe un conjunto de números F , tal que todos los elementos $a_{i,j}$ de la matriz pertenecen a dicho conjunto, diremos que la matriz tiene coeficientes en F .

OBSERVACIÓN 1.1. Para que las operaciones entre matrices estén bien definidas, es necesario que la suma, resta y multiplicación entre elementos de F también esté bien definida. Por esto generalmente F se elige como \mathbb{R} o \mathbb{Z} .

La colección de todas las matrices de dimensión $m \times n$ con coeficientes en F se denotará por

$$M_{m,n}(F).$$

DEFINICIÓN 1.1. Las matrices de dimensión $m \times 1$ se conocen como *vectores columna*, mientras que las de dimensión $1 \times n$ se conocen como *vectores renglón*.

La colección $M_{m,1}(F)$ de todos los vectores columna con coeficientes comunmente se denota por F^m . Mientras que la colección $M_{1,n}(F)$ de todos los vectores columna con coeficientes comunmente se denota por F^{n*} .

1.1. Operaciones elementales. Por brevedad, la matriz (A) se denota por $A = [a_{i,j}]$.

En el caso de los vectores renglones y columnas, podemos omitir el subíndice fijo

$$R = [R_{1,j}] = [R_j], \quad C = [C_{i,1}] = [C_i].$$

Si $B = [b_{i,j}]$ es otra matriz de dimensión $m \times n$, la suma se define como

$$A + B = [a_{i,j} + b_{i,j}].$$

De manera similar, la resta se define como

$$A - B = [a_{i,j} - b_{i,j}].$$

EJEMPLO 1.1.

$$\begin{pmatrix} 1 & -1 & 0 \\ 2 & 3 & -4 \end{pmatrix} + \begin{pmatrix} 7 & 0 & -1 \\ 2 & -1 & 5 \end{pmatrix} =$$

$$\begin{pmatrix} 1 & -1 & 0 \\ 2 & 3 & -4 \end{pmatrix} - \begin{pmatrix} 7 & 0 & -1 \\ 2 & -1 & 5 \end{pmatrix} =$$

Observe que para que la *suma y resta* tenga sentido, ambas matrices deben tener exactamente las *mismas dimensiones*.

Después de ver la facilidad para definir la suma y resta, uno se ve tentado a definir la multiplicación de la misma forma. Pero tal definición es poco útil en las aplicaciones.

Por esta razón, desarrollaremos el concepto de multiplicación, a fin de poder aplicar esta operación en la resolución de Ejemplos.

1.2. Multiplicación.

DEFINICIÓN 1.2. Sean $R = [R_j]$ un vector renglón y $C = [C_i]$ un vector columna, ambos de longitud n .

El *producto renglón-columna* se define como

$$(RC) \quad RC = (R_1 \quad \cdots \quad R_n) \begin{pmatrix} C_1 \\ \vdots \\ C_n \end{pmatrix} = \sum_{i=1}^n R_i C_i.$$

EJEMPLO 1.2. Considere

$$R = (1 \quad 0 \quad -1), \quad C = \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}.$$

Calcule RC .

EJEMPLO 1.3. Reescriba la siguiente ecuación, utilizando el *producto renglón-columna*:

$$2x - 3y + z = 0.$$

DEFINICIÓN 1.3. Sea $A = [a_{i,j}] \in M_{m \times n}$ y $B = [b_{j,k}] \in M_{n \times l}$. Definimos su producto como

$$(AB) \quad AB = (R_i C_k)$$

donde R_i es el i -ésimo renglón de A y C_k es la k -ésima columna de B .

OBSERVACIÓN 1.2. ■ Para que esta multiplicación tenga sentido, los renglones de A y las columnas de B deberán tener la misma longitud n .

- La matriz resultante tendrá dimensión $m \times l$.
- A menos que $m = l$, el producto BA podría no estar definido.
- Aun cuando BA estuviera bien definido, el producto de matrices no es *conmutativo*, es decir, generalmente tendremos que

$$AB \neq BA.$$

EJEMPLO 1.4. Encuentre el producto AB de las siguientes matrices

$$A = \begin{pmatrix} 0 \end{pmatrix}$$

$$B = \begin{pmatrix} 0 & -1 \end{pmatrix}$$

Solución:

$$AB = \begin{pmatrix} 0 & 0 \end{pmatrix}$$

EJEMPLO 1.5. Encuentre el producto AB de las siguientes matrices

$$A = \begin{pmatrix} 0 & -1 \\ -1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$B = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Solución:

$$AB = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

EJEMPLO 1.6. Encuentre el producto AB de las siguientes matrices

$$A = \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix}$$

$$B = \begin{pmatrix} -1 & 0 & 0 \end{pmatrix}$$

Solución:

$$AB = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

EJEMPLO 1.7. Encuentre el producto AB de las siguientes matrices

$$A = \begin{pmatrix} 6 \\ -9 \\ -10 \end{pmatrix}$$

$$B = \begin{pmatrix} -5 \end{pmatrix}$$

Solución:

$$AB = \begin{pmatrix} -30 \\ 45 \\ 50 \end{pmatrix}$$

EJEMPLO 1.8. Encuentre el producto AB de las siguientes matrices

$$A = \begin{pmatrix} 2 \end{pmatrix}$$

$$B = \begin{pmatrix} -1 & 1 & -3 \end{pmatrix}$$

Solución:

$$AB = \begin{pmatrix} -2 & 2 & -6 \end{pmatrix}$$

EJEMPLO 1.9. Encuentre el producto AB de las siguientes matrices

$$A = \begin{pmatrix} -1 & -3 \\ -7 & -1 \end{pmatrix}$$

$$B = \begin{pmatrix} -7 \\ -4 \end{pmatrix}$$

Solución:

$$AB = \begin{pmatrix} 19 \\ 53 \end{pmatrix}$$

EJEMPLO 1.10. Rescriba el siguiente sistema de ecuación en forma matricial y encuentre su solución:

$$\begin{cases} -x - 3y = 19 \\ -7x - y = 53 \end{cases}$$

2. Teoría general de grafos

En matemáticas, la *teoría de grafos* estudia estructuras matemáticas usadas para modelar relaciones por pares entre objetos.

2.1. Definición de grafo. Concepto de grafo Un *grafo* G consiste de:

- (a) Un conjunto V cuyos elementos son llamados *vértices*, puntos o nodos de G .
- (b) Un conjunto E de pares (no ordenados) de distintos vertices, a los que llamaremos *aristas* de G .

Denotaremos un grafo por $G(V, E)$ cuando querramos enfatizar los componentes del mismo.

OBSERVACIÓN 2.1. Debido a una ambigüedad en la traducción del inglés al español, en ocasiones, a un grafo también se le conoce como *gráfica*, que se puede confundir con el concepto de teoría de conjuntos. En este material, a veces utilizaremos *gráfica*, pero debe entenderse como un grafo.

Multigrafos Consideremos la figura 1 (b). Las aristas e_4 y e_5 son llamadas *aristas multiples* ya que conectan los mismos extremos, mientras que la arista e_6 es llamada *bucle* ya que conecta un vértice consigo mismo.

Tales diagramas son llamados *multigrafos*; la definición formal de grafo no admite aristas multiples ni bucles.

OBSERVACIÓN 2.2. Sin embargo, algunos textos utilizan “grafos” para referirse a lo que nosotros llamaremos multigrafos, mientras que ocupan “grafo simple” para lo que nosotros llamaremos grafos.

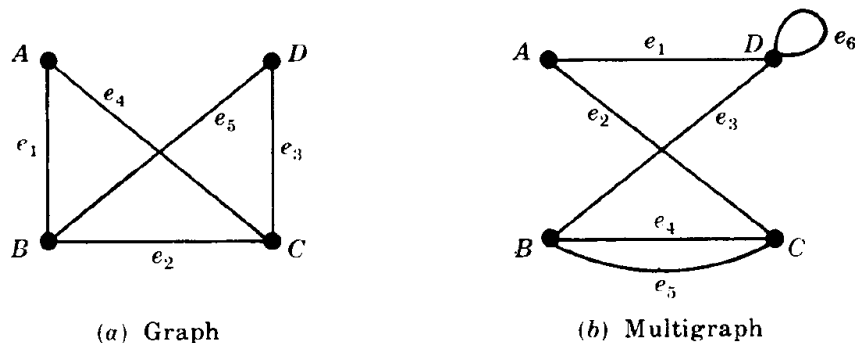


FIGURA 1. Grafos y multigrafos

Grado de un vértice El *grado* de un vértice v en un grafo G , denotado por $\deg(v)$, es igual al número de aristas en G que contienen a v , es decir, que *inciden* en v .

Dado que cada arista incide en dos vértices diferentes, tenemos el siguiente resultado simple pero importante:

TEOREMA 2.1. *La suma de los grados de los vértices en un grafo G es el doble del número de aristas.*

EJEMPLO 2.1. En el grafo de la figura 1(a), tenemos que

$$\deg(A) = 2, \deg(B) = 3, \deg(C) = 3, \deg(D) = 2.$$

La suma de los grados es igual a 10, que es dos veces el número de aristas.

DEFINICIÓN 2.1. Diremos que un vértice es *par* o *impar* de acuerdo a la paridad de su grado.

En el ejemplo anterior, tanto A como D son vértices pares, mientras que B y C son impares.

OBSERVACIÓN 2.3. Diremos que un vértice de grado cero está *aislado*.

Gráfos finitos y triviales Diremos que un grafo es *finito* si tiene un número finito de vértices y un número finito de aristas.

Observe que un número finito de vértices implica un número finito de aristas; pero no lo contrario no es necesariamente cierto.

Diremos que un grafo con un único vértice, sin aristas, es *trivial*.

OBSERVACIÓN 2.4. A menos que se indique de otra manera, sólo trataremos con grafos finitos.

2.2. Subgrafos y grafos homeomorfos e isomorfos. Ahora, discutiremos relaciones de equivalencia entre grafos.

Subgrafos Consideremos un grafo $G(V, E)$. Diremos que otro grafo $H(V', E')$ es un *subgrafo* de G si los vértices y aristas de H están contenidos en los vértices y aristas de G , es decir,

$$V' \subset V, E' \subset E.$$

En particular:

- (a) Un subgrafo $H(V', E')$ de $G(V, E)$ es llamado subgrafo *inducido* por sus vértices V' si el conjunto de aristas E' contiene todas las aristas en G cuyo extremos pertenecen a los vértices en H .
- (b) Si v es un vértice en G , entonces $G - v$ es el subgrafo de G obtenido al borrar v de G y todas las aristas en G que inciden en v .
- (c) Si e es una arista en G , entonces $G - e$ es el subgrafo de G obtenido borrando la arista e en G .

Grafos isomorfos Dos grafos $G(V, E)$ y $G^*(V^*, E^*)$ son llamados *isomorfos* si existe una función biyectiva $f : V \rightarrow V^*$ tal que: $\{u, v\}$ es una arista de G si y solo si $\{f(u), f(v)\}$ es una arista de G^* .

La idea es que estos grafos son equivalentes, aún cuando sus representaciones pueden lucir muy diferentes.

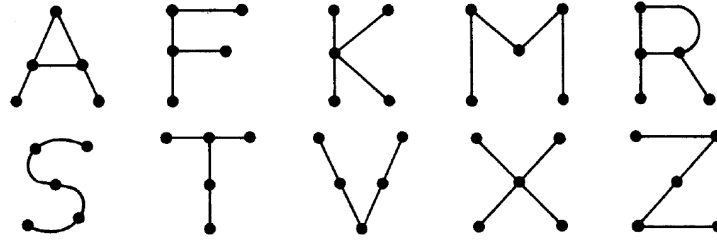


FIGURA 2. Grafos isomorfos.

Grafos homeomorfos Dado un grafo G , podemos obtener un nuevo grafo dividiendo una arista de G con vértices adicionales.

Dos grafos G y G^* son llamados *homeomorfos* si pueden obtenerse de gráficas isomorfas a través de este método.

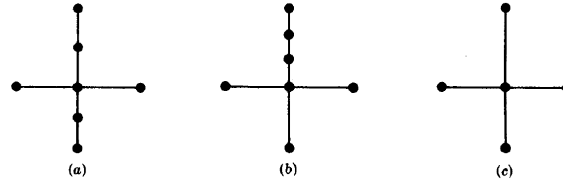


FIGURA 3. Grafos homomorfos

Los grafos (a) y (b) son homeomorfos, ya que se pueden obtener añadiendo vértices al grafo (c).

2.3. Caminos y conexidad. Un *camino* en un (multi)grafo G consiste en una sucesión alternante de vértices y arista de la forma

$$v_0, e_1, v_1, \dots, e_{n-1}, v_{n-1}, e_n, v_n$$

donde cada arista e_i contiene los vértices v_{i-1} y v_i .

OBSERVACIÓN 2.5. Observe que en grafo, podemos simplificar la notación para un camino, indicando sólo los vértices que recorre:

$$v_0, v_1, \dots, v_n.$$

Diremos que el camino es *cerrado* si $v_n = v_0$. En otro caso, diremos que el camino conecta v_0 con v_n .

Un *camino simple* es aquel en el cual todos los vértices son distintos. Mientras que un camino en el que todas las aristas son distintas se llama *paseo*.

La *longitud* de un camino es igual a número de aristas en la sucesión que lo define.

Un *ciclo* es un camino cerrado de *longitud* al menos 3, en el que todos los vértices son distintos, excepto el inicial v_0 y el final v_n .

Un ciclo de longitud k es llamado k -*ciclo*.

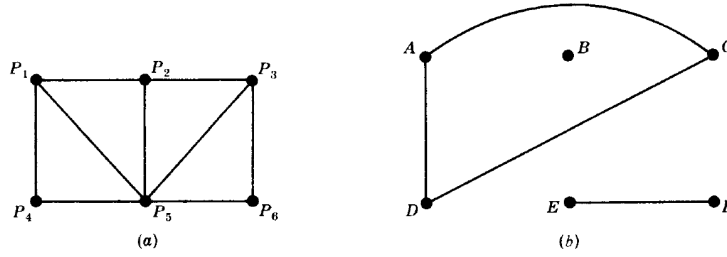


FIGURA 4. Conexidad en grafos

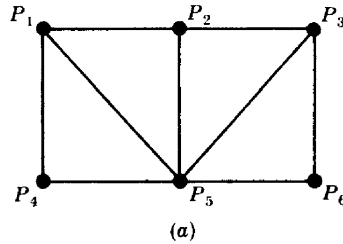
EJEMPLO 2.2. Consideremos el grafo 4(a). Considere las siguientes sucesiones

$$\alpha = (P_4, P_1, P_2, P_5, P_1, P_2, P_3, P_6),$$

$$\beta = (P_4, P_1, P_5, P_2, P_6)$$

$$\gamma = (P_4, P_1, P_5, P_2, P_3, P_5, P_6)$$

$$\delta = (P_4, P_1, P_5, P_3, P_6).$$



α es un camino de P_4 a P_6 , pero no es un paseo.

β no es un camino, ya que no existe alguna arista $\{P_2, P_6\}$.

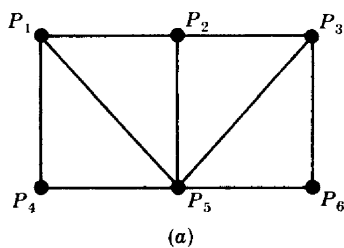
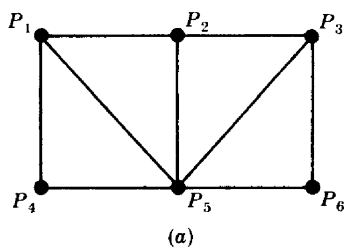
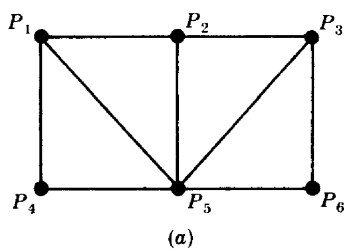
γ es un paseo, pero no es un camino simple.

δ es un camino simple de P_4 a P_6 , pero no es el camino más corto, es decir, con el menor número de aristas. ¿Cuál es el camino más corto?

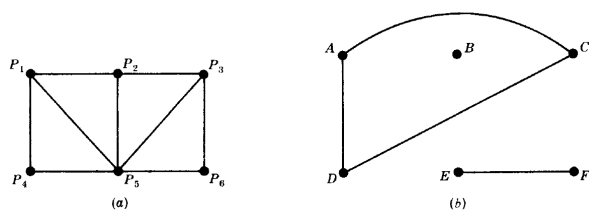
Eliminando aristas innecesarias, no es difícil ver que cualquier camino de u a v puede ser reemplazado por un camino simple.

Formalmente:

TEOREMA 2.2. *Existe un camino del vértice u a v si y solo si existe un camino simple de u a v .*



Conexidad y componentes conexas Un grafo G es conexo si existe un camino entre cualesquiera dos vértices. Por ejemplo, el grafo 4(a) es conexo, pero no así el grafo 4(b).



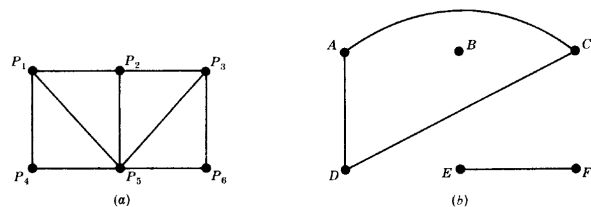
Consideremos un grafo G . Un subgrafo conexo H de G es llamado *componente conexa* de G si H no está contenido de manera propia en cualquier otro grafo conexo de G .

Por ejemplo, el grafo 4(b) tiene tres componentes conexas.

OBSERVACIÓN 2.6. Formalmente, permitiendo que un vértice u esté conectado consigo mismo, la relación

$$u \text{ está conectado con } v$$

es una relación de equivalencia en el conjunto de vértices del grafo G , y las clases de equivalencia de esta relación son las componentes conexas de G .



Distancia y diámetro Consideremos un grafo conexo G . La distancia entre dos vértices u y v en G , denotada por $d(u, v)$, es la longitud del camino más corto entre u y v . El diámetro de G , escrito $\text{diam}(G)$, es la distancia máxima entre cualesquiera dos puntos en G .

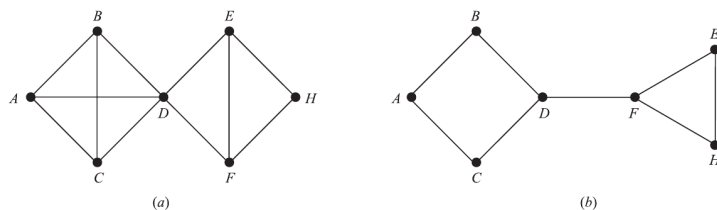


FIGURA 5. Distancia y diámetro

Por ejemplo, en el grafo 5(a), el diámetro es 3, mientras que en el (b), el diámetro es 4.

Puntos de corte y puentes Sea G un grafo conexo. Un vértice v en G es llamado *punto de corte* si $G - v$ es desconexo. Una arista e en G es llamada *punto de corte* si $G - e$ es desconexo.

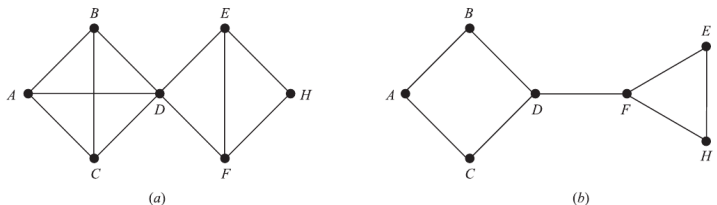


FIGURA 6. Puntos de corte y puentes

2.4. Grafos transitables y eulerianos. Un multigrafo es llamado *transitable* si existe un *paseo* (un camino donde todas las aristas son diferentes), que incluye *todos los vértices y todas las aristas*.

Tal paseo será llamado *paseo transitable*.

OBSERVACIÓN 2.7. De manera equivalente, un paseo transitable es un camino en el que todos los vértices se transitan *al menos* una vez, pero las aristas *exactamente* una vez.

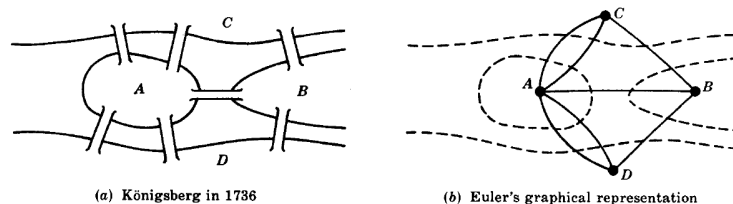


FIGURA 7. Puentes de Königsberg y su representación

PROPIEDAD 2.1. *Cualquier grafo conexo y finito con exactamente dos vértices impares es transitable. Un paseo transitable puede comenzar en alguno de los vértices impares y terminar en el otro vértice impar.*

Un grafo G es llamado *grafo Euleriano* si existe un *paseo transitable cerrado*. A tal paseo le llamaremos *paseo Euleriano*.

TEOREMA 2.3 (Euler). *Un grafo conexo y finito es Euleriano si y solo si cada vértice tiene grado par.*

Grafos hamiltonianos En la definición de grafos Eulerianos se enfatizó pasar por todas las aristas.

Ahora, nos enfocaremos en visitar todos los vértices.

Un *circuito Hamiltoniano* es un grafo G es un camino cerrado que visita cada vértice en G *exactamente* una vez.

Si G admite un circuito Hamiltoniano, entonces G es llamado un *grafo Hamiltoniano*.

OBSERVACIÓN 2.8. En la definición de circuito Hamiltoniano, cuando decimos que el camino *visita* cada vértice exactamente una vez significa que, aunque el vértice inicial tiene que ser el mismo que el final, todos los demás vértices intermedios deben ser distintos.

OBSERVACIÓN 2.9. Un **paseo Euleriano** atraviesa **cada una de las aristas** exactamente una vez, pero los vértices se pueden repetir, mientras que un **circuito Hamiltoniano** visita **cada uno de los vértices** exactamente una vez, pero las aristas pueden repetirse.

TEOREMA 2.4. *Sea G un grafo conexo con n vértices. Entonces G es Hamiltoniano si $n \geq 3$ y $n \leq \deg(v)$ para cada vértice v en G .*

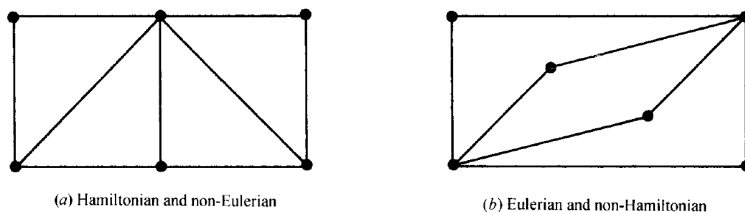


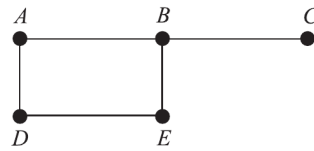
FIGURA 8. Circuitos Eulerianos y Hamiltonianos

2.5. Matriz de adyacencia. Supongamos que G es un grafo con m vértices y que estos han sido ordenados:

$$v_1, v_2, \dots, v_m.$$

Entonces, la *matriz de adyacencia* $A = (a_{i,j})$ del grafo G es la matriz de dimensión $m \times m$ definida por:

$$a_{i,j} = \begin{cases} 1 & v_i \text{ es adyacente a } v_j \\ 0 & \text{en otro caso} \end{cases}$$



(a)

$$\begin{array}{c} A \quad B \quad C \quad D \quad E \\ A \quad \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \end{bmatrix} \\ B \quad \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \end{bmatrix} \\ C \quad \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \end{bmatrix} \\ D \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \end{bmatrix} \\ E \quad \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

(b)

FIGURA 9. Matriz de adyacencia

3. Digrafos

Los *grafos dirigidos* o *digrafos* son grafos en los que las aristas tienen una dirección.

3.1. Grafos dirigidos. Un grafo dirigido $G = G(V, E)$ consiste de:

1. Un conjunto $V = V(G)$ cuyos elementos son llamados *vértices*;
2. un conjunto $E = E(G)$ de *pares ordenados* ordenados de vértices llamados *arcos* o *aristas dirigidas*.

Supongamos que $e = (u, v)$ es un arco en el digrafo G . Entonces, la siguiente terminología es usada:

- e comienza en u y termina en v ;
- u es el origen o punto inicial de e , mientras que v es el destino o punto final de e .
- v es un sucesor de u ;
- u es adyacente a v y v es adyacente desde u .

Si $u = v$, e es llamado un *bucle*.

Si las aristas o los vértices de un digrafo están etiquetados con algún tipo de dato, diremos que es un *digrafo etiquetado*.

De manera similar a un grafo, un digrafo será finito si el conjunto de vértices y el de aristas es finito.

EJEMPLO 3.1. Consideremos el siguiente digrafo. Las aristas e_2 y e_3 son llamados *paralelos*, ya que ambos comienzan en B y terminan en A . La arista e_7 es un *bucle*.

EJEMPLO 3.2.

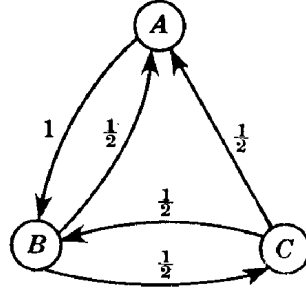
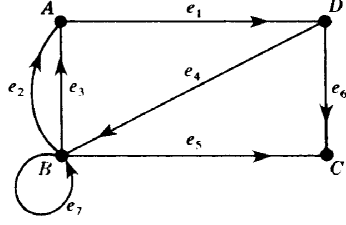


FIGURA 10. Proceso estocástico

3.2. Matriz de adyacencia. Ahora, sólo consideraremos *digrafos simples* $G(V, E)$, es decir, sin aristas paralelas. Entonces E es simplemente una relación en V .

De manera inversa, si R es una relación en V , entonces $G(V, R)$ es un digrafo simple.

En unidades anteriores, ya hemos construido digrafos asociados a relaciones de orden parcial, llamados diagramas de Hasse.

Supongamos que G es un digrafo simple con m vértices, y supongamos que los vértices de G han sido ordenados y son llamados v_1, v_2, \dots, v_m .

Entonces la *matriz de adyacencia* $A = (a_{i,j})$ de G es la una matriz de dimensión $m \times m$ definida de la siguiente manera

$$a_{i,j} = \begin{cases} 1 & \exists e \in E : e = (v_i, v_j) \\ 0 & \text{en otro caso} \end{cases}$$

OBSERVACIÓN 3.1. Las matrices de adyacencia de un mismo grafo dependen del orden en que se enumeren los vértices. Sin embargo, dos matrices de adyacencia de un mismo grafo están relacionadas por operaciones elementales: cambiar el orden de columnas y renglones.

EJEMPLO 3.3. Sea G el siguiente digrafo

La matriz identidad $I_m = (I_{i,j})$ de dimensión $m \times m$ se define como

$$I_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j, \end{cases}$$

es decir, es matriz cuadrangular con 1's en la *diagonal principal*, y ceros en cualquier otra entrada.

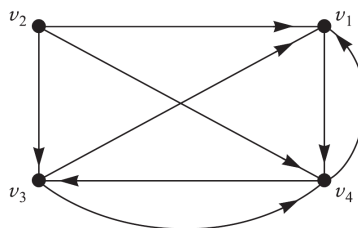


FIGURA 11. Construya su matriz de adyacencia del digrafo anterior.

EJEMPLO 3.4.

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

La propiedad principal de una matriz identidad I_m es que es neutra respecto a la multiplicación de matrices, es decir, para cualquier otra matriz $A \in M_n$:

$$AI_n = I_n A = A.$$

La potencia n -ésima de una matriz $A \in M_n$ se define de manera recursiva como

$$A^n = \begin{cases} I_n & n = 0 \\ AA^{n-1} & n \in \mathbb{N}. \end{cases}$$

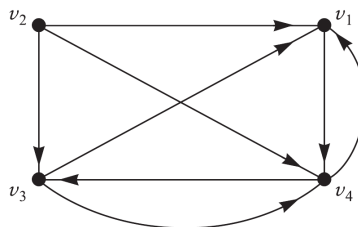
Es decir,

$$A^0 = I, A^1 = A, A^2 = AA, \dots$$

Definamos $a_k(i, j)$ como la entrada en la posición i, j de A^k .

PROPIEDAD 3.1. Sea A la matriz de adyacencia de un grafo G . Entonces $a_k(i, j)$ es igual al número de caminos de longitud k que van de v_i a v_j .

Ejemplo Consideremos nuevamente el grafo



Recordemos que su matriz de adyacencia es

$$(AD) \quad A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

Entonces

$$A^2 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 2 & 0 & 1 & 2 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 2 \end{pmatrix} \quad A^3 = \begin{pmatrix} 1 & 0 & 0 & 2 \\ 3 & 0 & 2 & 3 \\ 2 & 0 & 1 & 2 \\ 2 & 0 & 2 & 1 \end{pmatrix}$$

$$A^4 = \begin{pmatrix} 2 & 0 & 2 & 1 \\ 5 & 0 & 3 & 5 \\ 3 & 0 & 2 & 3 \\ 3 & 0 & 1 & 4 \end{pmatrix}$$

Observe que $a_2(4, 1) = 1$, de manera que existe un solo camino de longitud 2 de v_4 a v_1 . De manera similar, como $a_3(2, 3) = 2$, entonces existen dos caminos de longitud 3 de v_2 a v_3 .

OBSERVACIÓN 3.2. Si definimos

$$B_r = \sum_{i=1}^r A^i,$$

entonces la entrada i, j de esta matriz nos indicará el número de caminos de longitud a lo más r de v_i a v_j .

En nuestro ejemplo, considerando A dado por (AD), tenemos que

$$(1) \quad B_4 = \begin{pmatrix} 4 & 0 & 3 & 4 \\ 11 & 0 & 7 & 11 \\ 7 & 0 & 4 & 7 \\ 7 & 0 & 4 & 7 \end{pmatrix}$$

¿Existe alguna manera de llegar al vertice v_2 desde el vértice v_1 , sin importar la longitud del camino?

3.3. Matriz de accesibilidad. Sea $G = G(V, E)$ un grafo simple dirigido con m vértices v_1, \dots, v_m . La *matriz de accesibilidad* de G es la matriz m -cuadrangular $P = (p_{ij})$ definida de la siguiente manera:

$$p_{ij} = \begin{cases} 1 & \text{existe un camino de } v_i \text{ a } v_j \\ 0 & \text{en otro caso} \end{cases}$$

PROPIEDAD 3.2. Sea A la matriz de adyacencia de un grafo G con m vértices. Entonces la matriz de accesibilidad y

$$(2) \quad B_m = \sum_{i=1}^m A^i$$

tienen exactamente las mismas entradas no nulas.

DEFINICIÓN 3.1. Un digrafo es *fuertemente conexo* si para cualquier par de vértices u, v existe al menos un camino de u a v y otro de v a u .

PROPIEDAD 3.3. Sea $A \in M_m$ la matriz de adyacencia de un grafo G . Entonces, las siguientes proposiciones son equivalentes:

1. G es fuertemente conexo;
2. la matriz de accesibilidad P no tiene entradas nulas;
3. la matriz B_m , dada por (2), no tiene entradas nulas.

EJEMPLO 3.5. Para encontrar la matriz de accesibilidad asociada a la matriz de adyacencia A , dada por (AD), basta sustituir las entradas no nulas en la matriz B_4 , dada por (1), por 1's :

$$P = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

Probabilidad

1. Fundamentos de probabilidad

1.1. Experimentos aleatorios.

EJEMPLO 1.1. Si lanzamos una moneda, el resultado del experimento será “águila” (*tail* en inglés) que simbolizaremos por T o 0 o “sol” (*head* en inglés) simbolizado por H o 1, es decir, uno de los elementos del conjunto $\{T, H\}$ (o bien $\{0, 1\}$.)

EJEMPLO 1.2. Si lanzamos un dado, el resultado del experimento resultará en uno de los números del conjunto $\{1, 2, 3, 4, 5, 6\}$.

EJEMPLO 1.3. Si lanzamos una moneda dos veces, existen cuatro posibles resultados:

$$\{HH, HT, TH, TT\}.$$

EJEMPLO 1.4. Si estamos haciendo tornillos con una máquina, el resultado del experimento es que un tornillo puede salir defectuoso. Entonces cuando el tornillo este fabricado pertenecerá al conjunto

$$\{\text{defectuoso}, \text{no defectuoso}\}$$

EJEMPLO 1.5. Si un experimento consiste en medir la *vida útil* de una bombilla eléctrica producida por una compañía, entonces el resultado del experimento es tiempo t medido en horas en algún intervalo

$$0 \leq t \leq T,$$

donde T es el tiempo de vida máximo de una bombilla.

1.2. El espacio muestral. Un conjunto S que consiste de todos los posibles resultados de un experimento aleatorio es llamado *espacio muestral*, y cada posible resultado es llamado un *punto muestral*.

Usualmente existirá más de un espacio muestral que describe un experimento, pero usualmente, sólo uno provee la mayor información.

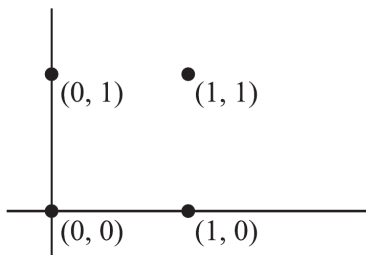
EJEMPLO 1.6. Si lanzamos un dado, un posible espacio muestral está dado por $\{1, 2, 3, 4, 5, 6\}$, mientras que otro está dado por $\{\text{par}, \text{impar}\}$.

EJEMPLO 1.7. Si lanzamos una moneda dos veces seguidas un posible espacio muestral esta dado en el ejemplo 1.3, mientras que otro esta dado por

$$\{(0, 0), (0, 1), (1, 0), (1, 1)\}.$$

Tipos de espacio muestral

- *Finito*: tiene un número finito de puntos.



- *Infinito numerable*: Tiene tantos puntos como los números naturales \mathbb{N} (es decir, podemos numerar el espacio).
- *Infinito no numerable*: Tiene tantos puntos como la recta real \mathbb{R} . Por ejemplo, el intervalo $0 < x < 1$.

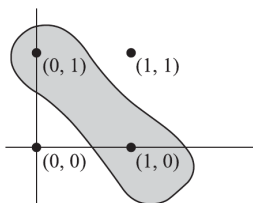
Si el espacio muestral es finito o infinito numerable, diremos que es *discreto*. Si es infinito no numerable, diremos que es *continuo*.

1.3. Eventos. Un *evento* es un subconjunto A de un espacio muestral S , es decir, un subconjunto de todos los posibles resultados de un experimento.

Si el resultado es un elemento de A , diremos que A *ha ocurrido*.

Un evento que consiste de un único punto de S es llamado a veces *evento elemental o simple*.

EJEMPLO 1.8. Si lanzamos una moneda dos veces, el evento de que obtengamos *exactamente* un águila es un subconjunto del espacio muestral:



Como eventos particulares, podemos considerar todo el espacio muestral S como el *evento cierto o seguro* y el conjunto vacío \emptyset como el *evento imposible*.

Operaciones entre eventos Supongamos que A, B son dos eventos.

- $A \cup B$ es el evento “*ocurre A o B o ambos*”, y también es llamado *unión de A con B* .
- $A \cap B$ es el evento “*ocurre A y B* ”, y también es llamado *intersección de A con B* .
- A' es el evento “*no ocurre A* ”, y también es llamado *negación de A* .
- $A - B = A \cap B'$ es el evento “*ocurre A pero no B* ”, y también es llamado *diferencia de A menos B* . Observe que $A' = S - A$.

Si $A \cap B = \emptyset$, entonces diremos que A y B son *disjuntos o mutuamente excluyentes*.

DEFINICIÓN 1.1. Si A_1, A_2, \dots es una colección de eventos tales que $A_i \cap A_j = \emptyset$ siempre que $i \neq j$, entonces diremos que son *eventos mutuamente excluyentes*

DEFINICIÓN 1.2. Si A_1, A_2, \dots son eventos mutuamente excluyentes diremos que $A_1 \cup A_2 \cup \dots$ es la *unión disjunta* de tales eventos y en ese caso escribiremos

$$A_1 \sqcup A_2 \sqcup \dots$$

Si

$$S = A_1 \sqcup A_2 \sqcup \dots$$

diremos que A_1, A_2, \dots es una *partición de S* .

EJEMPLO 1.9. Respecto al experimentos de lanzar una moneda dos veces, consideremos el evento A que consiste en *obtener al menos un sol*, mientras que el evento B consiste en que *el segundo lanzamiento sea un águila*.

Entonces $A = \{TH, HT, HH\}$, $B = \{HT, TT\}$ y por tanto

- (a) $A \cup B = \{HT, TH, HH, TT\} = S$
- (b) $A \cap B = \{HT\}$
- (c) $A' = \{TT\}$
- (d) $A - B = \{TH, HH\}$

1.4. El Concepto de Probabilidad. A cualquier evento en un espacio muestral se le puede asignar un número entre $0 = 0\%$ y $1 = 100\%$ que representa su *probabilidad* de ocurrir.

Enfoque clásico Si un evento puede ocurrir en h diferentes maneras de un total de n posibles resultados, todos igualmente plausibles, entonces la probabilidad del evento es h/n .

EJEMPLO 1.10. Supongamos que queremos conocer la probabilidad de que un sol aparezca en un solo volado. Desde que hay dos maneras diferentes *igualmente probables* en que la moneda caiga, y de esas dos maneras un sol sólo puede hacerlo de una manera, razonamos que su probabilidad es $1/2$.

OBSERVACIÓN 1.1. Aquí suponemos que la moneda no está cargada.

Enfoque frecuencial Si después de n repeticiones de un experimento, donde n es suficientemente grande, se observa que un evento ocurre en h ocasiones, entonces diremos que la probabilidad del evento es h/n . Esta es también llamada *probabilidad empírica* del evento.

EJEMPLO 1.11. Si lanzamos una moneda 1000 veces y obtenemos sol 532 veces, estimamos que la probabilidad resultantes es $532/1000 = 0.532$.

OBSERVACIÓN 1.2. Ambos enfoque tienen sus inconvenientes:

- (a) En el caso clásico, la expresión “*igualmente probable*” es vaga;
 - (b) mientras que en el enfoque frecuencial, “*un número muy grande*” no es preciso.
- Por estas razones, los matemáticos han desarrollado un *enfoque axiomático* de la probabilidad.

1.5. Los Axiomas de la probabilidad. Supongamos que tenemos un espacio muestral S . Supongamos que C es la colección de todos los eventos en S . Diremos que $P : C \rightarrow \mathbb{R}$ es una función de probabilidad si satisface las siguientes propiedades:

AXIOMA 1.1. Para cada evento A , se tiene que

$$(3) \quad P(A) \geq 0.$$

AXIOMA 1.2. La probabilidad del evento cierto S es

$$(4) \quad P(S) = 1.$$

AXIOMA 1.3. Para cualquier cantidad numerable de eventos mutuamente excluyentes A_1, A_2, \dots tenemos que

$$(5) \quad P(A_1 \sqcup A_2 \sqcup \dots) = P(A_1) + P(A_2) + \dots$$

En particular, para dos eventos mutuamente excluyentes A_1, A_2 ,

$$(6) \quad P(A_1 \sqcup A_2) = P(A_1) + P(A_2)$$

1.6. Algunos teoremas importantes en probabilidad.

TEOREMA 1.1. Si $A_1 \subset A_2$, entonces $P(A_1) \leq P(A_2)$ y

$$P(A_2 - A_1) = P(A_2) - P(A_1).$$

TEOREMA 1.2. Para cada evento A ,

$$(7) \quad 0 \leq P(A) \leq 1,$$

es decir, la probabilidad siempre se encuentra entre 0% y 100%.

TEOREMA 1.3. El evento imposible tiene probabilidad cero, es decir,

$$(8) \quad P(\emptyset) = 0.$$

TEOREMA 1.4. La probabilidad de un evento complementarios está dada por

$$(9) \quad P(A') = 1 - P(A)$$

TEOREMA 1.5. Si $A = A_1 \sqcup \dots \sqcup A_N$ es la unión disjunta de eventos mutuamente excluyentes entonces

$$(10) \quad P(A) = P(A_1) + \dots + P(A_N).$$

En particular, si $S = A_1 \sqcup \dots \sqcup A_N$ entonces

$$(11) \quad P(A_1) + \dots + P(A_N) = 1.$$

TEOREMA 1.6. Si A, B, C son dos eventos no necesariamente excluyentes, entonces

$$(12) \quad P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

$$(13) \quad P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

$$(14) \quad - P(A \cap B) - P(B \cap C) - P(C \cap A)$$

$$(15) \quad + P(A \cup B \cup C).$$

TEOREMA 1.7. Para cualesquiera eventos A, B ,

$$(16) \quad P(A) = P(A \cap B) + P(A \cap B').$$

TEOREMA 1.8. Si A_1, A_2, \dots, A_N es una partición del espacio muestral S , es decir, $S = A_1 \sqcup A_2 \sqcup \dots \sqcup A_N$ entonces para cualquier evento A

$$(17) \quad P(A) = P(A \cap A_1) + P(A \cap A_2) + \dots + P(A \cap A_N).$$

1.7. Asignación de probabilidades. Si un espacio muestral consiste en una cantidad *finita* de posibles resultados a_1, \dots, a_N , entonces por el teorema 1.5,

$$(18) \quad P(A_1) + \dots + P(A_n) = 1$$

donde A_1, \dots, A_n son *conjuntos elementales* o *eventos simples* dados por $A_i = \{a_i\}$.

Se sigue que uno puede escoger de manera arbitraria cualesquiera números no negativos como probabilidades de estos eventos simples, siempre que se satisfaga (18).

En particular, si suponemos *probabilidades iguales* para todos los eventos, entonces

$$(19) \quad P(A_k) = \frac{1}{n}, \quad k = 1, 2, \dots, n,$$

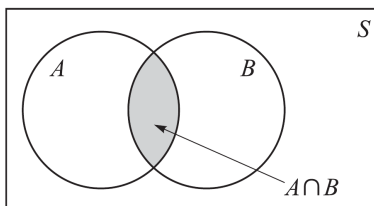
y si A es un evento formado por la unión disjunta de h eventos simples, entonces

$$(20) \quad P(A) = \frac{h}{n}.$$

OBSERVACIÓN 1.3. Esto es equivalente al *enfoque clásico*. Pero podemos usar el *enfoque frecuencial* para asignar dichas probabilidades.

EJEMPLO 1.12. Un solo dado se lanza. Encuentre la probabilidad de que obtengamos un 2 o un 5.

Sean A, B dos eventos tales que $P(A) > 0$.



Denotaremos por $P(B|A)$ la probabilidad de B dado que A haya ocurrido y diremos que es la *probabilidad condicional* de B dado A .

2. Análisis combinatorio

El *análisis combinatorio* es una manera sofisticada de contar.

2.1. Principio fundamental del conteo y diagramas de árbol. Si una tarea se puede realizar en n formas diferentes y otra en m formas diferentes, entonces las dos tareas se pueden realizar en $n \times m$ formas diferentes.

[t]

EJEMPLO 2.1.

1. Si una persona tiene 2 camisas y 4 corbatas, ¿de cuántas formas puede combinarlas?
2. Construya un diagrama de árbol para representar todas estas opciones.

2.2. Permutaciones. Si tenemos n objetos distintos y queremos ordenarlos tendremos

$$n \times (n-1) \times \dots \times 2 \times 1$$

formas diferentes de hacerlo.

DEFINICIÓN 2.1 (n factorial).

$$(21) \quad n! = \begin{cases} 1 & n = 0 \\ n \times (n-1)! & n > 0 \end{cases}$$

Si tenemos n objetos distintos y queremos arreglar r de estos en una línea, entonces tendremos una *permutación* de n en r dada por

$$(22) \quad P_r^n = n \times (n-1) \times \dots \times (n-r+1)$$

o de manera equivalente

$$(23) \quad P_r^n = \frac{n!}{(n-r)!}$$

[t]

EJEMPLO 2.2. ¿Cuántas permutaciones de longitud 3 se pueden formar con las letras A, B, C, D, E, F, G ?

[t]

EJEMPLO 2.3. Encuentre el número de permutaciones diferentes de las 11 letras de la palabra *MISSISSIPPI*.

2.3. Combinaciones. En una *permutación*, uno está interesado en el orden de los objetos. Así abc y bca son permutaciones diferentes. Pero en algunos problemas, uno está interesado sólo en elegir objetos sin importar su orden. Tales selecciones se llaman *combinaciones*. Por ejemplo, abc y bca representan la misma combinación.

El número de combinación C_r^n al elegir r objetos de una colección de n diferentes está dada por el *número combinatorio*

$$(24) \quad C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Algunas fórmulas combinatorias

$$(25) \quad \binom{n}{r} = \frac{P(n, r)}{r!}$$

$$(26) \quad \binom{n}{r} = \binom{n}{n-r}$$

$$(27) \quad \binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$$

[t]

EJEMPLO 2.4. En una baraja inglesa, ¿cuántas formas hay de escoger dos cartas del mismo palo?

TEOREMA 2.1 (Teorema del binomio).

$$(28) \quad (x + y)^n = \sum_{r=0}^n \binom{n}{r} x^r y^{n-r}$$

Aproximación de Stirling

$$(29) \quad n! \approx \sqrt{2\pi n} (n^n e^{-n})$$

3. Probabilidad condicional

3.1. Definición.

DEFINICIÓN 3.1 (Probabilidad condicional).

$$(30) \quad P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$(31) \quad P(A \cap B) = P(A)P(B|A)$$

OBSERVACIÓN 3.1. La probabilidad condicional satisface todos los axiomas de una función de probabilidad. Podemos pensar $P(\cdot|A)$ como la función de probabilidad que se obtiene al reemplazar el espacio muestral S por A .

[t]

EJEMPLO 3.1. Encontrar la probabilidad de que una solo lanzamiento de un dado resulte en un número menor que 4 si

- (a) no hay más información;
- (b) se sabe que el lanzamiento resultó en un número impar.

3.2. Teoremas sobre Probabilidad Condicional.

TEOREMA 3.1. Para cualesquiera tres eventos A_1, A_2, A_3 , tenemos que

$$(32) \quad P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

TEOREMA 3.2. Si $S = A_1 \sqcup \dots \sqcup A_N$, entonces

$$(33) \quad P(A) = P(A_1)P(A|A_1) + \dots + P(A_N)P(A|A_N)$$

3.3. Eventos independientes. Si $P(B|A) = P(B)$, i.e., la probabilidad de que B ocurra no está afectada por la ocurrencia de A , entonces diremos que A y B son independientes.

DEFINICIÓN 3.2. A y B son eventos independientes si y solo si

$$(34) \quad P(A \cap B) = P(A)P(B).$$

La definición se puede generalizar a más de dos eventos. Por ejemplo, diremos que A_1, A_2, A_3 son eventos independientes si

$$(35) \quad k \neq j \rightarrow P(A_j \cap A_k) = P(A_j)P(A_k), \quad j, k = 1, 2, 3$$

$$(36) \quad P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3).$$

TEOREMA 3.3 (Teorema de Bayes). Si $S = A_1 \sqcup A_2 \sqcup \dots \sqcup A_N$, entonces

$$(37) \quad P(A_k|A) = \frac{P(A_k)P(A|A_k)}{\sum_j P(A_j)P(A|A_j)}$$

[t]

EJEMPLO 3.2. Demuestre el teorema de Bayes.

[t]

EJEMPLO 3.3. La caja I contiene 3 canicas rojas y 2 azules, mientras que la caja II contiene 8 canicas rojas y 8 azules. Una moneda se lanza: Si cae un sol, se escoge una moneda de la caja I y si cae águila, de la caja II . Encuentre la probabilidad de obtener una canica roja.

[t]

EJEMPLO 3.4. Supongamos que en el problema anterior, quien lanza la moneda no revela si ha caído águila o sol (de manera que la caja de la que se obtiene la canica no se revela) pero revela que una canica roja se ha obtenido. ¿Cual es la probabilidad de haber obtenido un sol?

Índice general

4. Variables aleatorias

Supongamos que a cada punto del espacio muestral se le asigna un número. Entonces hemos definido una *función* en el espacio muestral. Esta función es llamada *variable aleatoria* (o *variable estocástica*) o de manera más precisa *función aleatoria*.

Usualmente, las variables aleatorias se denotan por letras mayúsculas como X o Y . En general, una variable aleatoria tiene algún significado físico, geométrico, económico, financieros, etc.

[t]

EJEMPLO 4.1. Supongamos que una moneda se lanza dos veces de manera que el espacio muestral es $\{HH, HT, TH, TT\}$. Digamos que X representa el número de soles (H) que obtenemos.

5. Funciones de probabilidad discretas

Sea X una variable aleatoria discreta. Supongamos que los valores que puede tomar son x_1, \dots, x_k , arreglados en algún orden dado. Supongamos también que esos valores tienen alguna probabilidad dada por

$$(38) \quad P(X = x_k) = f(x_k).$$

Función de probabilidad

$$(39) \quad P(X = x) = \begin{cases} f(x) & x = x_k \\ 0 & \text{en otro caso} \end{cases}$$

En general, $f(x)$ será una función de probabilidad si

$$\begin{cases} f(x) \geq 0 \\ \sum_x f(x) = 1. \end{cases}$$

EJEMPLO 5.1. Encuentre la función de probabilidad correspondiente a la variable aleatoria X del ejemplo 4.1.

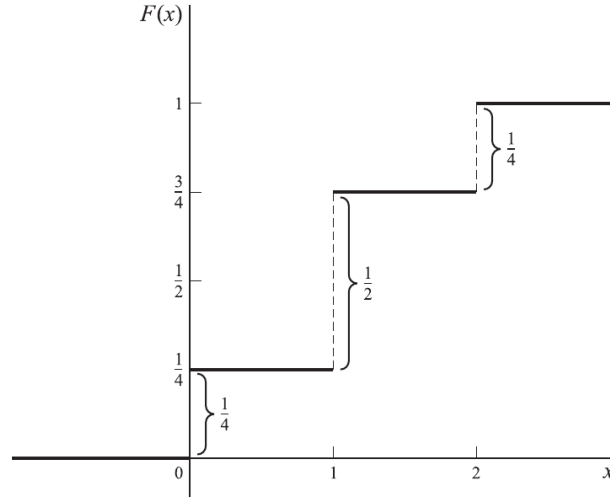
EJEMPLO 5.2. Suponga que un par de dados se lanzan. Sea X la variable aleatoria dada por la suma de los puntos. Encuentre la distribución de probabilidad de X .

EJEMPLO 5.3. Encuentre la distribución de probabilidad de niños y niñas en familias con 3 hijos, suponiendo la misma probabilidad para niños y niñas.

6. Funciones de distribución para variables aleatorias discretas

La función de distribución de una variable discreta X se obtiene de la función de probabilidad a través de la siguiente fórmula

$$(40) \quad F(x) = P(X \leq x) = \sum_{u \leq x} f(u).$$



Si X toma sólo un número finito de valores x_1, \dots, x_n entonces la función de distribución está dada por

$$(41) \quad F(x) = \begin{cases} 0 & -\infty < x < 0 \\ f(x_1) & x_1 \leq x < x_2 \\ f(x_1) + f(x_2) & x_2 \leq x < x_3 \\ \vdots & \vdots \\ f(x_1) + \dots + f(x_n) & x_n \leq x < \infty \end{cases}$$

EJEMPLO 6.1. Encuentre la función de distribución para la variable aleatoria X del ejemplo 5.1 y obtenga su gráfica.

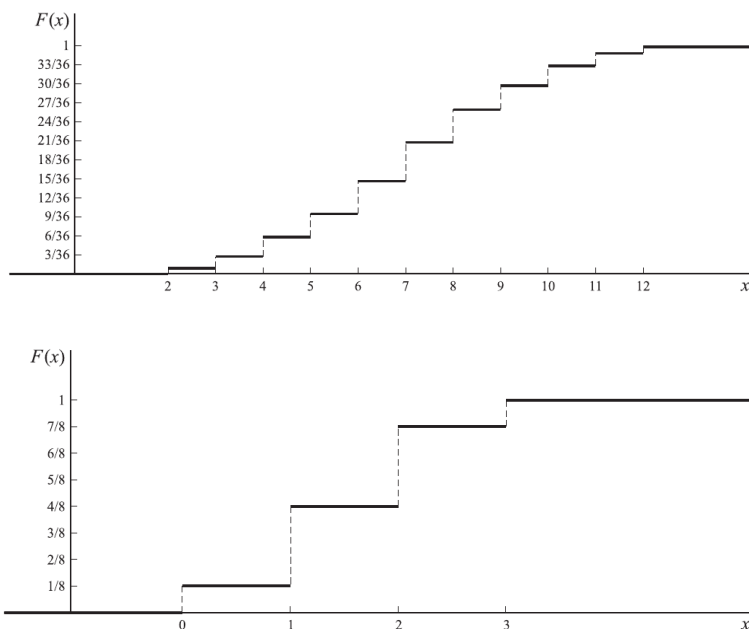
- OBSERVACIÓN 6.1. ■ Los saltos en la función de distribución están determinados por el valor de la función de probabilidad.
- Este tipo de funciones se conoce como *función escalonada*. Debe observarse que son *continuas por la derecha*.
 - La función de distribución es *monótonamente creciente*.

La función de probabilidad se puede obtener a partir de la función de distribución con la siguiente fórmula

$$(42) \quad f(x) = F(x) - \lim_{u \rightarrow x^-} F(u).$$

- EJEMPLO 6.2. (a) Encuentre la función de distribución $F(x)$ para la variable aleatoria del problema resuelto 5.2;
- (b) grafique esta función de distribución.

- EJEMPLO 6.3. (a) Encuentre la función de distribución $F(x)$ para la variable aleatoria del problema resuelto 5.3;
- (b) grafique esta función de distribución.



7. Variable Aleatorias Continuas

Una variable aleatoria no discreta X se llama *absolutamente continua* (o simplemente *continua*) si su función de distribución puede ser representada como

$$(43) \quad F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du, \quad -\infty < x < \infty.$$

La función f usualmente se llama *densidad de probabilidad* y debe satisfacer las siguientes propiedades:

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x) dx = 1$.

La probabilidad de que X se encuentre entre dos valores a y b está dada por

$$(44) \quad P(a < x < b) = \int_a^b f(x) dx.$$

[t]

$$(45) \quad P(X = a) = 0.$$

Por tanto, en (44) podemos reemplazar cualquier signo $<$ por \leq .

[t]

EJEMPLO 7.1. (a) Encuentre la constante c tal que la función

$$(46) \quad f(x) = \begin{cases} cx^2 & 0 < x < 3 \\ 0 & \text{en otro caso} \end{cases}$$

sea una función de probabilidad.

(b) Calcule $P(1 < X < 2)$.

[t]

EJEMPLO 7.2. Encuentre la distribución de probabilidad para la variable aleatoria del ejemplo 7.1 y utilícela para calcular $P(1 < x \leq 2)$.

La probabilidad de que X se encuentre entre x y $x + \Delta x$ esta dada por

$$(47) \quad P(x \leq X \leq x + \Delta x) = \int_x^{x+\Delta x} f(u)du,$$

de manera que si $\Delta x \approx 0$, tendremos que

$$(48) \quad P(x \leq X \leq x + \Delta x) \approx f(x)\Delta x.$$

También podemos deducir de (43), al diferenciar de ambos lados, que

$$(49) \quad \frac{dF(x)}{dx} = f(x)$$

en todos aquellos puntos en que $f(x)$ sea continua. Es decir, la derivada de la función de distribución es la función de densidad.

OBSERVACIÓN 7.1. Existen variables aleatorias que no son discretas ni continuas. Por ejemplo

$$(50) \quad F(x) = \begin{cases} 0 & x < 1 \\ \frac{x}{2} & 1 \leq x < 2 \\ 1 & x \leq 2. \end{cases}$$

EJEMPLO 7.3. Una variable aleatoria X tiene función de densidad

$$(51) \quad f(x) = \frac{c}{x^2 + 1}, \quad -\infty < x < \infty.$$

(a) Encuentre el valor de c ;

(b) encuentre la probabilidad de que

$$\frac{1}{3} < X^2 < 1.$$

EJEMPLO 7.4. Encuentre la función de distribución correspondiente a la función de densidad del problema resuelto 7.3

EJEMPLO 7.5. La función de distribución para una variable aleatoria X es

$$(52) \quad F(x) = \begin{cases} 1 - e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Encuentre

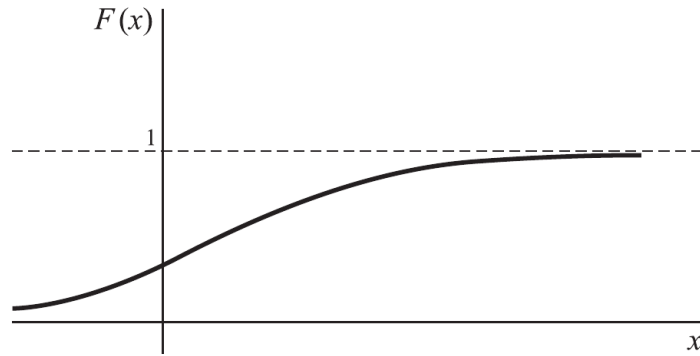
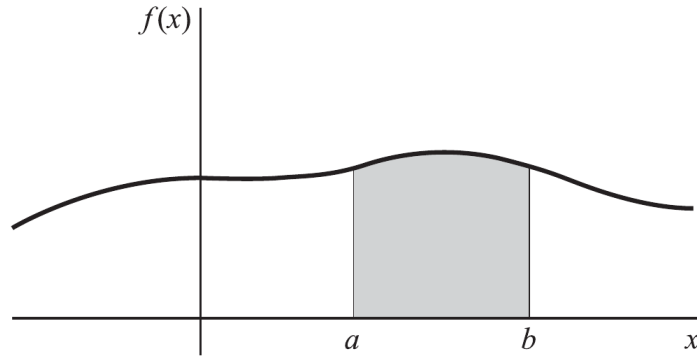
(a) la función de densidad;

(b) la probabilidad de que $X > 2$;

(c) la probabilidad que $-3 < X \leq 4$.

8. Interpretación gráfica

La distribución $F(x) = P(X \leq x)$ es monótonamente creciente de 0 a 1...
...y el área bajo dicha curva es igual a 1.



9. Distribución conjunta de probabilidad

Las ideas anteriores se generalizan fácilmente a dos o más variables.

Caso discreto Si X y Y son ambas variables aleatorias discretas, definimos la *función de probabilidad conjunta* de X y Y por

$$(53) \quad P(X = x, Y = y) = f(x, y)$$

donde

1. $f(x, y) \geq 0$;
2. $\sum_k \sum_y f(x, y) = 1$.

Supongamos que X sólo toma uno de los m valores x_1, \dots, x_m , mientras que Y sólo toma uno de los n valores y_1, \dots, y_n .

Entonces la probabilidad del evento $X = x_j, Y = y_k$ está dada por

$$(54) \quad P(X = x_j, Y = y_k) = f(x_j, y_k)$$

Una función de probabilidad conjunta para X y Y puede ser representada por una *tabla de probabilidad conjunta* como la siguiente:

La probabilidad de $X = x_j$ se obtiene de la siguiente manera

$$(55) \quad P(X = x_j) = f_X(x_j) = \sum_{k=1}^n f(x_j, y_k).$$

$\begin{array}{c} Y \\ \diagdown \\ X \end{array}$	y_1	y_2	\dots	y_n	Totals ↓
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	\dots	$f(x_1, y_n)$	$f_1(x_1)$
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	\dots	$f(x_2, y_n)$	$f_1(x_2)$
\vdots	\vdots	\vdots		\vdots	\vdots
x_m	$f(x_m, y_1)$	$f(x_m, y_2)$	\dots	$f(x_m, y_n)$	$f_1(x_m)$
Totals →	$f_2(y_1)$	$f_2(y_2)$	\dots	$f_2(y_n)$	1 ← Grand Total

De manera similar, la probabilidad de $Y = y_k$ se obtiene de la siguiente manera

$$(56) \quad P(Y = y_k) = f_Y(y_k) = \sum_{j=1}^m f(x_j, y_k).$$

Nos referiremos a $f_X(x)$ y $f_Y(y)$ como *funciones de probabilidad marginal* de X y Y respectivamente.

Observe que

$$(57) \quad \sum_{j=1}^m f_X(x_j) = 1, \sum_{k=1}^n f_Y(y_k) = 1,$$

lo cual se puede reescribir como

$$(58) \quad \sum_{j=1}^m \sum_{k=1}^n f(x_j, y_k) = 1.$$

La *función de distribución conjunta* está definida por

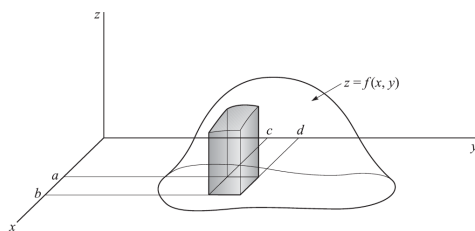
$$(59) \quad F(x, y) = P(X \leq x, Y \leq y) = \sum_{u \leq x} \sum_{v \leq y} f(u, v)$$

Caso continuo El caso en el que ambas variables son continuas es obtenido de manera análoga reemplazando las sumas por integrales.

La *función de probabilidad conjunta* (o de manera más común *función de densidad conjunta*) de X y Y está definida por

1. $f(x, y) \geq 0$;
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$.

Gráficamente $z = f(x, y)$ representa una *superficie de probabilidad* tal que el



volumen bajo la superficie es igual a 1.

$$(60) \quad P(a < X < b, c < Y < d) = \int_a^b \int_c^d f(x, y) dy dx.$$

A cada evento A corresponde una región \mathcal{R}_A del plano xy tal que

$$(61) \quad P(A) = \iint_{\mathcal{R}_A} f(x, y) dx dy.$$

La *función de distribución conjunta* de X y Y en este caso está definida por

$$(62) \quad F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du.$$

Se sigue que

$$(63) \quad \frac{\partial^2 F}{\partial x \partial y} = f(x, y)$$

$$(64) \quad P(X \leq x) = F_X(x) = \int_{-\infty}^x \int_{-\infty}^{\infty} f(u, v) dv du$$

$$(65) \quad P(Y \leq y) = F_Y(y) = \int_{-\infty}^{\infty} \int_{-\infty}^y f(u, v) dv du$$

Diremos que $F_X(x), F_Y(y)$ son las *funciones de distribución marginal*, o simplemente *funciones distribuciones*, de X y Y , respectivamente.

Las derivadas de (64) y (65) con respecto a x y y son llamadas *funciones de densidad marginal*, o simplemente las *funciones de densidad*, de X y Y están dados por

$$(66) \quad f_X(x) = \int_{-\infty}^{\infty} f(x, v) dv, \quad f_Y(y) = \int_{-\infty}^{\infty} f(u, y) du.$$

10. Variables Aleatorias Independientes

Supongamos que X y Y son variables aleatorias discretas. Si los eventos $X = x$ y $Y = y$ son eventos independientes para todo x, y , entonces diremos que X, Y son v.a's independientes.

En tal caso,

$$(67) \quad P(X = x, Y = y) = P(X = x)P(Y = y),$$

o de manera equivalente

$$(68) \quad f(x, y) = f_X(x)f_Y(y).$$

De manera inversa, si para todo x, y la función de probabilidad conjunta $f(x, y)$ pueden ser expresada como el producto de funciones de probabilidad marginal $f_X(x)f_Y(y)$, entonces X, Y son independientes.

Si no pueden expresarse de dicha manera, entonces X, Y son dependientes.

Si X, Y son v.a's continuas, diremos que son *independientes* si los eventos $X \leq x$ y $Y \leq y$ son independientes para todo x, y .

En tal caso, escribiremos

$$(69) \quad P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$$

o de manera equivalente

$$(70) \quad F(x, y) = F_X(x)F_Y(y)$$

donde $F_X(x)$ y $F_Y(y)$ son las funciones de distribución marginal de X, Y respectivamente.

De manera inversa, si para todo x, y la función de probabilidad conjunta $f(x, y)$ pueden ser expresada como el producto de funciones de probabilidad marginal $F_X(x)F_Y(y)$, entonces X, Y son independientes.

Si no pueden expresarse de dicha manera, entonces X, Y son dependientes.

Para v.a's independientes continuas, también es cierto que la función de densidad conjunta $f(x, y)$ es el producto de funciones $f_X(x)f_Y(y)$ y estas son las funciones de densidad marginal de X, Y respectivamente.

EJEMPLO 10.1. La función de probabilidad conjunta de dos variables discretas X, Y está dada por

$$(71) \quad f(x, y) = \begin{cases} c(2x + y) & 0 \leq x \leq 2, 0 \leq y \leq 3 \\ 0 & \text{en otro caso.} \end{cases}$$

- (a) Encuentre el valor de la constante c ;
- (b) encuentre $P(X = 2, Y = 1)$;
- (c) encuentre $P(X \geq 1, Y \leq 2)$.

$X \backslash Y$	0	1	2	3	Totals ↓
0	0	c	$2c$	$3c$	$6c$
1	$2c$	$3c$	$4c$	$5c$	$14c$
2	$4c$	$5c$	$6c$	$7c$	$22c$
Totals →	$6c$	$9c$	$12c$	$15c$	$42c$

EJEMPLO 10.2. Encuentre las funciones de probabilidad marginal para X y Y en el problema resuelto 10.1.

EJEMPLO 10.3. Muestre que las variables aleatorias del problema resuelto 10.1 son dependientes.

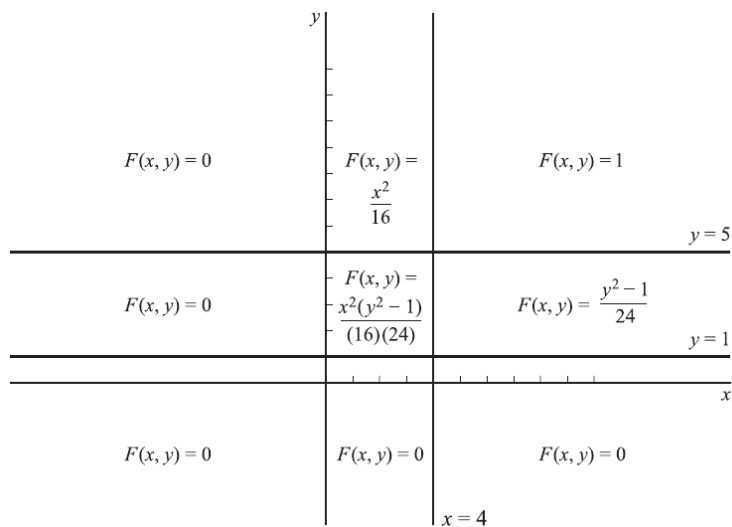
EJEMPLO 10.4. La función de densidad conjunta de dos variables aleatorias continuas X y Y es

$$(72) \quad f(x, y) = \begin{cases} cxy & 0 < x < 4, 1 < y < 5 \\ 0 & \text{en otro caso.} \end{cases}$$

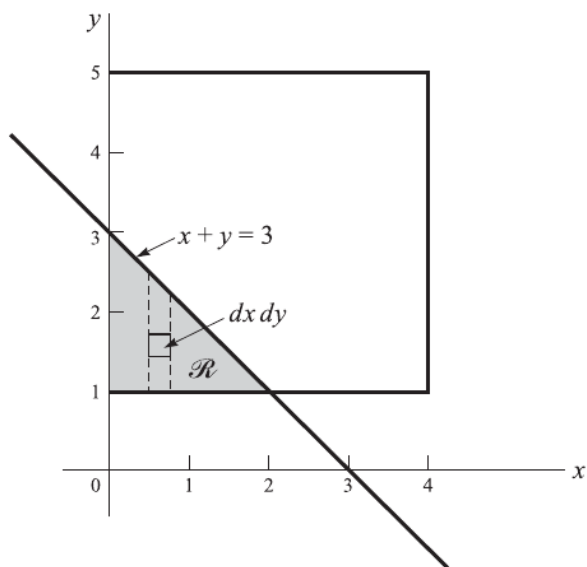
- (a) Encuentre el valor de c ;
- (b) encuentre $P(1 < X < 2, 2 < Y < 3)$;
- (c) encuentre $P(X \geq 3, Y \leq 2)$.

EJEMPLO 10.5. Encuentre las funciones de probabilidad marginal de las v.a's X, Y del problema resuelto 10.4.

EJEMPLO 10.6. Encuentre la función de distribución conjunta para las v.a's del problema resuelto 10.4.



EJEMPLO 10.7. En el problema resuelto 10.4, encuentre $P(X + Y < 3)$.



11. Distribución Condicional

Nosotros ya sabemos que si $P(A) > 0$,

$$(73) \quad P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

Si X, Y son v.a's discretas y tenemos los eventos $A = \{X = x\}$, $B = \{Y = y\}$, entonces (73) se convierte

$$(74) \quad P(Y = y|X = x) = \begin{cases} \frac{f(x, y)}{f_X(x)} & 0 < f_X(x) \\ 0 & \text{en otro caso} \end{cases}$$

$f(x, y) = P(X = x, Y = y)$ es la función de probabilidad conjunta, mientras que $f_X(x)$ es la función de probabilidad marginal para X .

Definimos la *función de probabilidad condicional de Y dado X* como

$$(75) \quad f(y|x) = \begin{cases} \frac{f(x, y)}{f_X(x)} & 0 < f_X(x) \\ 0 & \text{en otro caso} \end{cases}$$

De manera similar, definimos la *función de probabilidad condicional de X dado Y* como

$$(76) \quad f(x|y) = \begin{cases} \frac{f(x, y)}{f_Y(y)} & 0 < f_Y(y) \\ 0 & \text{en otro caso} \end{cases}$$

Estas ideas son fácilmente extensibles al caso donde X, Y son v.a's continuas.

Por ejemplo, la *función de densidad condicional de Y dado X* es

$$(77) \quad f(y|x) = \begin{cases} \frac{f(x, y)}{f_X(x)} & 0 < f_X(x) \\ 0 & \text{en otro caso} \end{cases}$$

donde $f(x, y)$ es la función de densidad conjunta de X y Y y $f_X(x)$ es la función de densidad marginal de X .

Usando (77) podemos por ejemplo encontrar que la probabilidad que Y se encuentre entre c y d dado que $X = x$ es

$$(78) \quad P(c < Y < d|X = x) = \int_c^d f(y|x) dy.$$

EJEMPLO 11.1. Para la distribución del problema resuelto 10.1, encuentre

- (a) $f(y|2)$; y
- (b) $P(Y = 1|X = 2)$

EJEMPLO 11.2. Si X y Y tienen función de densidad conjunta

$$(79) \quad f(x, y) = \begin{cases} \frac{3}{4} + xy & 0 < x < 1, 0 < y < 1 \\ 0 & \text{en otro caso,} \end{cases}$$

encuentre

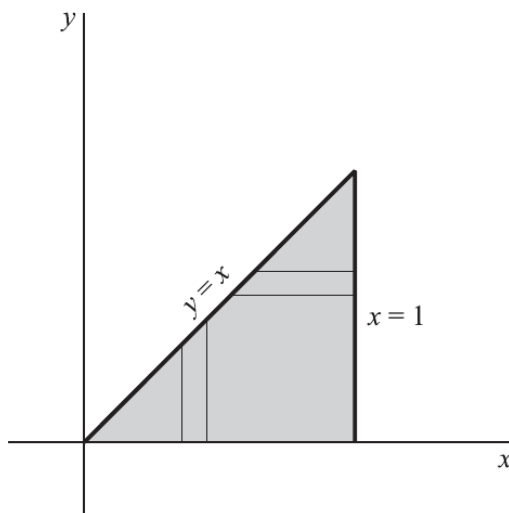
- (a) $f(y|x)$;
- (b) $P(Y > \frac{1}{2}|X = \frac{1}{2})$.

EJEMPLO 11.3. La función de densidad conjunta de las variables aleatorias X y Y está dada por

$$(80) \quad f(x, y) = \begin{cases} 8xy & 0 \leq x \leq 1, 0 \leq y \leq x \\ 0 & \text{en otro caso.} \end{cases}$$

Encuentre

- (a) la densidad marginal de X ;
- (b) la densidad marginal de Y ;
- (c) la densidad condicional de X ;
- (d) la densidad condicional de Y .



EJEMPLO 11.4. Determine si las v.a's del problema resuelto 11.3 son independientes.

12. Definición de Esperanza Matemática

Para una variable aleatoria discreta X que toma valores x_1, \dots, x_n , la *esperanza matemática* se define como

$$(81) \quad E(X) = \sum_{j=1}^n x_j P(X = x_j) =: \sum x P(X = x),$$

o de manera equivalente

$$(82) \quad E(X) = \sum_{j=1}^n x_j f(x_j) =: \sum x f(x),$$

donde $f(x) = P(X = x)$.

Como un caso especial, cuando $f(x) \equiv \frac{1}{n}$, obtenemos la *media aritmética*:

$$(83) \quad E(X) = \frac{\sum_{i=1}^n x_i}{n}.$$

EJEMPLO 12.1. Sea X el número que se obtiene al lanzar un dado. Entonces, cada cara x tiene la misma probabilidad

$$f(x) = \frac{1}{6}$$

de caer.

$$\text{Por tanto, } E(X) = (1) \left(\frac{1}{6}\right) + \dots + (6) \left(\frac{1}{6}\right) = \frac{1 + \dots + 6}{6} = 3.5$$

Caso Discreto Numerable En el caso en que X tome un cantidad (infinita) numerable de valores x_1, x_2, \dots , definimos

$$E(X) = \sum_{i=1}^{\infty} x_i f(x_i),$$

siempre y cuando dicha *serie* converja.

Caso Continuo Para una variable aleatoria continua X que tenga función de densidad $f(x)$, la esperanza de X se define como

$$(84) \quad E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

siempre y cuando dicha *integral* converja.

La esperanza de X es llamada a menudo *media* de X y es denotada por μ_x , o simplemente μ , cuando la variable aleatoria subyacente se sobreentiende.

La media o esperanza de X da un único valor que representa el promedio de los valores de X , y por esta razón decimos que es una *medida de tendencia central*.

EJEMPLO 12.2. Supongamos que un juego se juega con un dado único que se suponen justos. En este juego, un jugador gana \$20 si sale un 2; \$40 con un 4; \$30 con un 6; y no gana ni pierde con cualquier otra cara. Encuentre la suma esperada de dinero que ganaría.

$$\begin{aligned} \mu &= \$20 \left(\frac{1}{6} \right) + \$40 \left(\frac{1}{6} \right) + \$60 \left(\frac{1}{6} \right) \\ &= \frac{\$20 + \$40 + \$60 + 3 \times \$0}{6} \\ &= \$15 \end{aligned}$$

EJEMPLO 12.3. La función de densidad de una variable aleatoria X está dada por

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre el valor esperado de X .

$$\begin{aligned} \mu &= E(X) \\ &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_0^2 x \left(\frac{1}{2}x \right) dx \\ &= \frac{1}{6} x^3 \Big|_0^2 \\ &= \frac{1}{6} (2)^3 - \frac{1}{6} (0)^3 \\ &= \frac{4}{3} \end{aligned}$$

12.1. Funciones de Variables Aleatorias. Sea X una variable aleatoria discreta con función de probabilidad $f(x)$. Entonces $Y = g(X)$ es una variable aleatoria discreta con función de probabilidad

$$(85) \quad h(y) = P(g(X) = y) = \sum_{\{x|g(x)=y\}} g(x)f(x)$$

Entonces, en el caso discreto.

$$(86) \quad E(g(X)) = \sum_x g(x)f(x)$$

De manera similar, en el caso continuo

$$(87) \quad E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

EJEMPLO 12.4. Si X es la variable aleatoria del ejemplo 12.3, encuentre $E(3X^2 - 2X)$.

En este caso, $g(x) = 3x^2 - 2x$.

Recordemos que

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2 \\ 0 & \text{en otro caso} \end{cases}$$

$$\begin{aligned} E(3X^2 - 2X) &= E(g(X)) \\ &= \int_{-\infty}^{\infty} (3x^2 - 2x) f(x)dx \\ &= \int_0^2 (3x^2 - 2x) \left(\frac{1}{2}x\right) dx \\ &= \int_0^2 \frac{3}{2}x^3 - x^2 dx \\ &= \left. \frac{3}{8}x^4 - \frac{1}{3}x^3 \right|_0^2 \\ &= \frac{10}{3} \end{aligned}$$

12.2. Algunos temas sobre esperanza matemática. Linealidad

TEOREMA 12.1. Si c, d son constantes y X, Y son variables aleatorias, entonces

$$(88) \quad E(cX + dY) = cE(X) + dE(Y)$$

Esperanza e independencia

TEOREMA 12.2. Si X, Y son variables aleatorias independientes, entonces

$$(89) \quad E(XY) = E(X)E(Y)$$

13. Varianza y Desviación Estándar

Ya vimos que la espereza matemática de una variable aleatoria X es una medida de tendencia central y que generaliza a la *media aritmética* μ .

OBSERVACIÓN 13.1. Por esta razón, de aquí en adelante definiremos

$$\mu = \mu_X = E(X).$$

Otra cantidad de gran importancia es la *varianza* que se define como

$$(90) \quad \sigma_X^2 = \text{Var}(X) = E\left((X - \mu_X)^2\right)$$

La *desviación estándar* se definirá como

$$(91) \quad \sigma_X = \sqrt{\text{Var } X}$$

OBSERVACIÓN 13.2. Si la variable aleatoria X se sobreentiende del contexto, omitiremos el subíndice correspondiente, es decir,

$$\mu = \mu_X, \sigma = \sigma_X, \sigma^2 = \sigma_X^2.$$

Si X es una variable aleatoria discreta, la varianza está dada por

$$(92) \quad \sigma^2 = E\left((X - \mu)^2\right) = \sum (x - \mu)^2 f(x),$$

siempre y cuando esta suma converja.

En el caso de que todas las probabilidades sean iguales y la variable aleatoria X sea finita tenemos

$$(93) \quad \sigma^2 = \frac{(x_1 - \mu)^2 + \dots + (x_n - \mu)^2}{n}$$

EJEMPLO 13.1. Como vimos anteriormente, si X es la cara obtenida al lanzar un dado, entonces $\mu_X = 3.5$.

La varianza de X es

$$\begin{aligned} \sigma^2 &= \frac{(1 - 3.5)^2 + \dots + (6 - 3.5)^2}{6} \\ &= \frac{17.5}{6} \\ &\approx 2.916 \end{aligned}$$

Si X es una variable aleatoria continua con función de densidad $f(x)$, entonces la varianza está dada por

$$(94) \quad \sigma^2 = E\left((X - \mu)^2\right) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

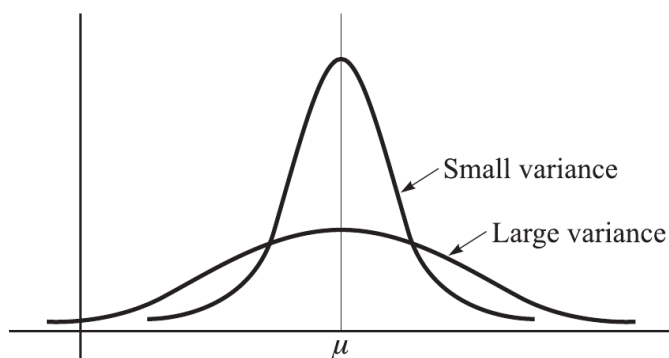
siempre y cuando la integral converja.

Tanto la varianza como la desviación estándar es una *medida de dispersión*.

EJEMPLO 13.2. Encuentre la varianza y la desviación estándar de la variable aleatoria del ejemplo 12.3.

Recordemos que esta variable aleatoria X tiene densidad de probabilidad

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2 \\ 0 & \text{en otro caso} \end{cases}$$



$$\begin{aligned}
 \sigma^2 = \text{Var}(X) &= \int_0^2 \left(x - \frac{4}{3}\right)^2 f(x) dx \\
 &= \int_0^2 \frac{1}{2} \left(x - \frac{4}{3}\right)^2 x dx \\
 &= \left. \frac{1}{8} x^4 - \frac{4}{9} x^3 + \frac{4}{9} x^2 \right|_0^2 \\
 &= \frac{2}{9} \\
 &\approx 0.2
 \end{aligned}$$

Algunos teoremas sobre Varianza

$$(95) \quad \sigma^2 = E(X^2) - \mu^2$$

$$(96) \quad \text{Var}(cX) = c^2 \text{Var}(X)$$

$$(97) \quad \sigma^2 = \min_a \left\{ E\left((X - a)^2\right) \right\}$$

Si X, Y son independientes

$$(98) \quad \text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$$

Variables Aleatorias Estandarizadas Sea X una variable aleatoria con media μ y desviación estándar $\sigma > 0$. Diremos que la *variable aleatoria estandarizada* asociada está dada por

$$(99) \quad X^* = \frac{X - \mu}{\sigma}.$$

$$(100) \quad E(X^*) = 0, \text{Var}(X^*) = 1.$$

14. Covarianza

Los resultados dados anteriormente para una variable aleatoria pueden extenderse a dos variables.

$$\begin{aligned}
 (101) \quad \mu_X &= E(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy \\
 \mu_Y &= E(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy
 \end{aligned}$$

$$\begin{aligned}
 (102) \quad \sigma_X^2 &= E((X - \mu_X)^2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x, y) dx dy \\
 \sigma_Y^2 &= E((Y - \mu_Y)^2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_Y)^2 f(x, y) dx dy
 \end{aligned}$$

Covarianza

$$(103) \quad \sigma_{XY} = \text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y))$$

$$(104) \quad \sigma_{XY} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y) dx dy$$

Caso Discreto

$$\begin{aligned}
 (105) \quad \mu_X &= \sum_x \sum_y x f(x, y) \\
 \mu_Y &= \sum_x \sum_y y f(x, y)
 \end{aligned}$$

$$(106) \quad \sigma_{XY} = \sum_x \sum_y (x - \mu_X)(y - \mu_Y) f(x, y)$$

$$(107) \quad \sigma_{XY} = E(XY) - E(X)E(Y) = E(XY) - \mu_X \mu_Y$$

Si X, Y son independientes, entonces

$$(108) \quad \sigma_{XY} = \text{Cov}(X, Y) = 0$$

$$(109) \quad \text{Var}(X \pm Y) = \text{Var}(X) \pm 2 \text{Cov}(X, Y) + \text{Var}(Y).$$

De manera equivalente,

$$(110) \quad \sigma_{X \pm Y}^2 = \sigma_X^2 \pm 2\sigma_{XY} + \sigma_Y^2$$

Coefficiente de correlación de Pearson

$$(111) \quad \rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

TEOREMA 14.1.

$$(112) \quad |\sigma_{XY}| \leq \sigma_X \sigma_Y$$

$$(113) \quad |\rho| \leq 1.$$

Propiedades de la correlación

$$1. \quad -1 \leq \rho \leq 1$$

2. $\rho \approx 0$: *Correlación débil*, prácticamente no existe una correlación lineal.
3. $|\rho| \approx 1$: *Correlación fuerte*, la correlación está dada prácticamente por una función afín $y = mx + b$.
4. $\rho > 0$: *Correlación positiva*, en la medida que una crece, la otra también crece.
5. $\rho < 0$: *Correlación negativa*, en la medida que una crece, la otra decrece.

EJEMPLO 14.1. Sean X, Y variables aleatorias discretas con densidad de probabilidad conjunta

$$(114) \quad f(x, y) = \begin{cases} \frac{2x+y}{42} & 0 \leq x \leq 2, 0 \leq y \leq 3 \\ 0 & \text{en otro caso.} \end{cases}$$

Encuentre los siguientes estadísticos:

- | | | |
|--------------------|----------------------------------|--------------------------------------|
| (a) $\mu_X = E(X)$ | (e) $E(Y^2)$ | (i) σ_Y |
| (b) $\mu_Y = E(Y)$ | (f) $\sigma_X^2 = \text{Var}(X)$ | (j) $\sigma_{XY} = \text{Cov}(X, Y)$ |
| (c) $E(XY)$ | (g) σ_Y | (k) ρ |
| (d) $E(X^2)$ | (h) $\sigma_Y^2 = \text{Var}(Y)$ | |

EJEMPLO 14.2. Sean X, Y variables aleatorias continuas con densidad de probabilidad conjunta

$$(115) \quad f(x, y) = \begin{cases} \frac{1}{210}(2x+y) & 2 < x < 6, 0 < y < 5 \\ 0 & \text{en otro caso.} \end{cases}$$

Encuentre los siguientes estadísticos:

- | | | |
|--------------------|----------------------------------|--------------------------------------|
| (a) $\mu_X = E(X)$ | (e) $E(Y^2)$ | (i) σ_Y |
| (b) $\mu_Y = E(Y)$ | (f) $\sigma_X^2 = \text{Var}(X)$ | (j) $\sigma_{XY} = \text{Cov}(X, Y)$ |
| (c) $E(XY)$ | (g) σ_Y | (k) ρ |
| (d) $E(X^2)$ | (h) $\sigma_Y^2 = \text{Var}(Y)$ | |

15. La Distribución Binomial

Si p es la probabilidad de que en un solo ensayo ocurra un evento (llamada la probabilidad de éxito) y $q = 1 - p$ es la probabilidad de que este evento no ocurra en un solo ensayo (llamada probabilidad de fracaso), entonces la probabilidad de que el evento ocurra exactamente x veces en N ensayos (es decir, que ocurran x éxitos y $N - x$ fracasos) está dada por

$$(116) \quad f(x) = P(X = x) = \binom{N}{x} p^x q^{N-x}$$

donde $x = 0, 1, \dots, N$.

EJEMPLO 15.1. La probabilidad de obtener exactamente dos caras en seis lanzamientos de una moneda es

$$\binom{6}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{6-2} = \frac{15}{64}$$

empleando (116) con $N = 6, x = 2, p = q = \frac{1}{2}$.

EJEMPLO 15.2. Calcule la probabilidad de obtener al menos 4 caras en 6 lanzamientos de una moneda.

En lo subsecuente, daremos por hecho que hemos importado los siguientes paquetes:

- `scipy.stats`
- `numpy` como `np`

```
[fragile, allowframebreaks]statsBinom.py
from scipy import stats
import numpy as np
import matplotlib.pyplot as plt

#Consideremos 6 experimentos con p de éxito 1/2
p=0.5
N=6
binDist = stats.binom(N,p)
#probabilidad de obtener dos éxitos
print binDist.pmf(2)
##0.234375
#probabilidad de obtener al menos 4 éxitos
print sum(binDist.pmf(np.arange(4,6+1)))
##0.34375
```

EJEMPLO 15.3. Desarrolle $(p + q)^4$.

```
[fragile, allowframebreaks]coefBinom.py

from scipy import stats
import numpy as np

#coeficientes de (p+q)^4
p=.5
N=4
binomDist = stats.binom(N,p)
binDistExmp = binomDist.pmf(np.arange(5))
print binDistExmp*2**N
##[ 1.  4.  6.  4.  1.]
```

Propiedades de la distribución binomial Supongamos que realizamos N experimentos con probabilidad éxito p y de fracaso $q = 1 - p$.

$$(117) \quad \mu = Np$$

$$(118) \quad \sigma^2 = Npq$$

```
[fragile,allowframebreaks]histBinom.py
import numpy as np
import matplotlib.pyplot as plt

#Ejemplo de distribución binomial
N,p=100, 0.5
```

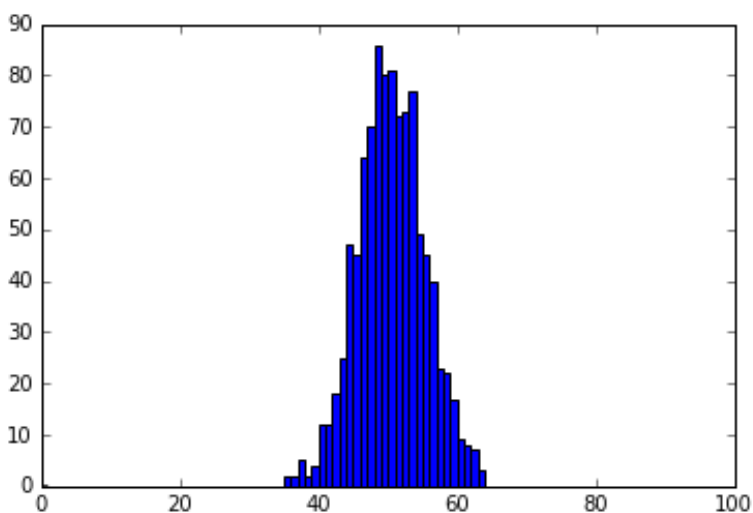
```

s = np.random.binomial(N,p,1000)

miHist = np.histogram(s, bins = np.arange(100+1))
print miHist[0]
print miHist[1]
print np.mean(s)
print N*p
print np.var(s)
print N*p*(1-p)

plt.hist(s, bins = np.arange(100+1))
plt.show()

```



16. Distribución Normal

Una de las distribuciones de probabilidad continua más importantes es la *distribución normal*, también llamada *distribución gaussiana*, que se define mediante la función de densidad

$$(119) \quad f_{a,b}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x-a)^2}{b^2}}$$

donde a, b son parámetros específicos para cada v.a. X .

Propiedades de la distribución normal Si la v.a. X tiene la función de densidad dada por (122), con parámetros a, b entonces

$$(120) \quad a = \mu_X$$

$$(121) \quad b = \sigma_X$$

Si una variable aleatoria normal X tiene función de densidad

$$(122) \quad f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}},$$

escribiremos $X \sim N(\mu, \sigma^2)$.

Variable aleatoria normalizada

$$(123) \quad Z = \frac{X - \mu}{\sigma}$$

$$(124) \quad \mu_Z = 0$$

$$(125) \quad \sigma_Z = 1$$

Forma Estándar

$$(126) \quad f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

En este caso, diremos que Z está *normalmente distribuida*.

[fragile, allowframebreaks]distribucionNormal.py

```
import scipy.integrate as integrate
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.patches import Polygon

def fn(x,m=0,s=1):
    return np.exp(-(x-m)**2/(2*s**2))/(s*np.sqrt(2*np.pi))
x1 = np.arange(-4,4,0.1)
plt.plot(x1, fn(x1))
plt.show()

for s in np.arange(1,4+1):
    result = integrate.quad(lambda x:fn(x),-s,s)
    print result

for s in np.arange(1,4+1):
    result = integrate.quad(lambda x:fn(x),-s,s)

    a, b = -s, s # integral limits
    x = np.arange(-4,4,0.01)
    y = fn(x)

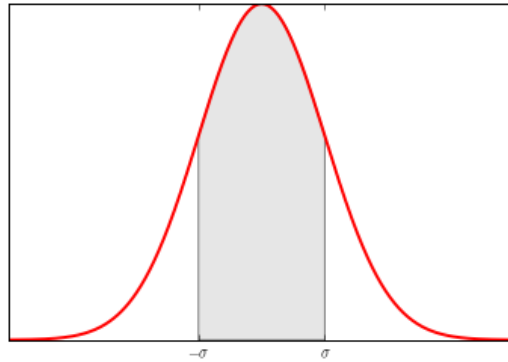
    fig, ax = plt.subplots()
    plt.plot(x, y, 'r', linewidth=2)
    plt.ylim(ymin=0)

    # Make the shaded region
    ix = np.linspace(a, b)
    iy = fn(ix)
    verts = [(a, 0)] + list(zip(ix, iy)) + [(b, 0)]
    poly = Polygon(verts, facecolor='0.9', edgecolor='0.5')
    ax.add_patch(poly)

    ax.set_xticks((a, b))
    ax.set_xticklabels(('$_{-\sigma}$', '$_{\sigma}$'))
    ax.set_yticks([])
```

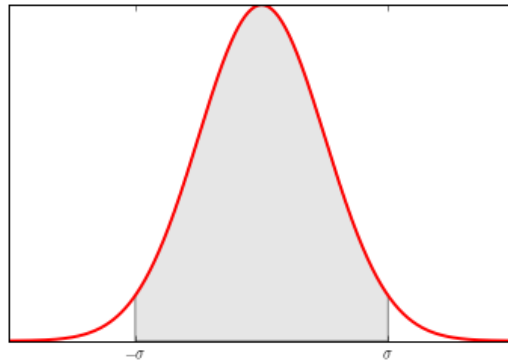
```
plt.show()
print result
```

```
[fragile]
```



```
#(0.682689492137086, 7.579375928402476e-15)
```

```
[fragile]
```



```
#(0.9544997361036417, 1.8403548653972355e-11)
```

```
[fragile]
```

```
#(0.9973002039367399, 1.1072256503105314e-14)
```

```
[fragile]
```

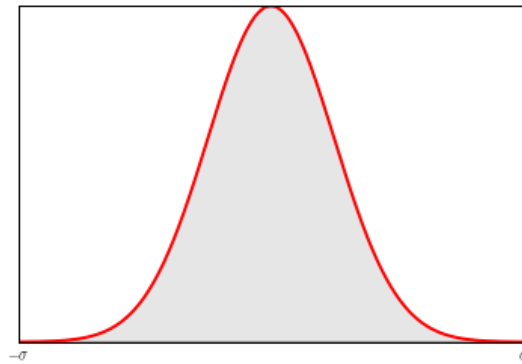
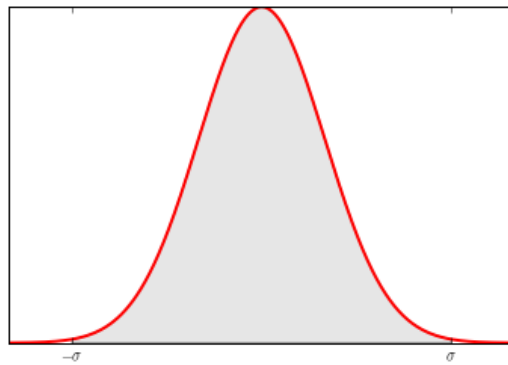
```
#(0.9999366575163339, 4.838904125482879e-12)
```

```
[fragile, allowframebreaks]normalCDF.py
```

```
from scipy import stats
import numpy as np
import matplotlib.pyplot as plt
```

```
mu = 3.5
```

```
sigma = 0.76
```



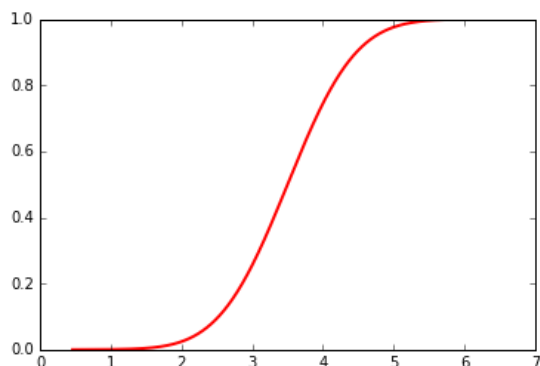
```
nd = stats.norm(mu, sigma)

x = np.arange(mu - 4*sigma, mu + 4*sigma, 0.01)
y = nd.cdf(x)

fig, ax = plt.subplots()
plt.plot(x, y, 'r', linewidth=2)
plt.ylim(ymin=0)

for k in range(1,5):
    print nd.cdf(mu+k*sigma)-nd.cdf(mu-k*sigma)

#0.682689492137
#0.954499736104
#0.997300203937
#0.999936657516
```



17. Relación entre las distribuciones binomial y normal

Si $N \sim \infty, p, q \gg 0$, y X es una distribución binomial con parámetros N, p entonces

$$(127) \quad \frac{X - Np}{\sqrt{Npq}} \sim N(0, 1).$$

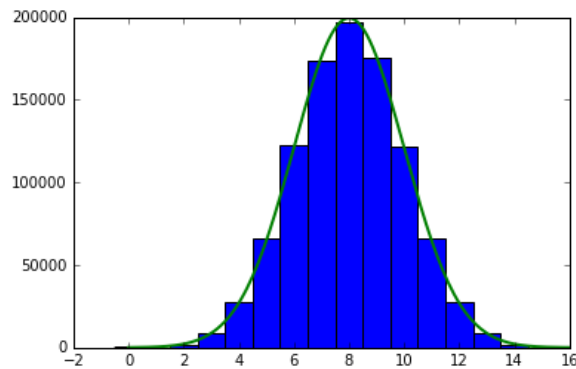
EJEMPLO 17.1. Consideremos el experimento de lanzar 16 veces una moneda. Repitamos 1,000,000 dicho experimento. Compruebe que dicho experimento se puede modelar por una variable aleatoria con distribución $N(\mu = 8, \sigma^2 = 4)$

[fragile, allowframebreaks]relBinomNormal.py

```
import numpy as np
import matplotlib.pyplot as plt

def fn(x,m=0,s=1):
    C = 1/(s*np.sqrt(2*np.pi))
    return C*np.exp(-(x-m)**2/(2*s**2))

N,p=30, 0.5
R = 1000000
q=1-p
mB = N*p
sB = np.sqrt(N*p*q)
X = np.random.binomial(N,p,R)
myBins = np.arange(-0.5,N+0.5,1)
plt.hist(X, bins = myBins)
x = np.arange(mB-4*sB,mB+4*sB+0.1,0.1)
y = R*fn(x, m=mB, s=sB)
plt.plot(x,y,lw=2)
plt.ylim(ymin=0)
plt.show()
```



18. La Distribución de Poisson

Distribución de Poisson Diremos que una variable aleatoria *discreta* X tiene distribución de Poisson si su función de probabilidad está dada por:

$$(128) \quad f(n) = \frac{\lambda^n e^{-\lambda}}{n!}, \quad n = 0, 1, 2, \dots$$

En este caso, $\mu_X = \sigma^2 = \lambda$.

En teoría de probabilidad y estadística, la distribución de Poisson es una distribución de probabilidad discreta que expresa, a partir de una frecuencia de ocurrencia media, la probabilidad de que ocurra un determinado número de eventos durante cierto período de tiempo. Concretamente, se especializa en la probabilidad de ocurrencia de sucesos con probabilidades muy pequeñas, o sucesos raros.

Wikipedia: Distribución de Poisson

EJEMPLO 18.1. El número de personas por día que llegan a una sala de urgencias tiene una distribución de Poisson con media 5. Hallar la probabilidad de que cuando mucho lleguen tres por día y la probabilidad de que por lo menos lleguen 8 personas por día.

```
[fragile, allowframebreaks]distPoisson.py
from scipy import stats
import numpy as np
import matplotlib.pyplot as plt

def f(x, mu=1):
    return stats.poisson.pmf(x, mu)

def F(x, mu=1):
    return stats.poisson.cdf(x, mu)

x1 = np.arange(0,100+1)
plt.plot(x1, f(x1, mu=5), 'bo')
plt.show()

s = np.random.poisson(5,365)
```

```

M = np.max(s)
myBins = np.arange(0,M+1)
plt.hist(s, bins = myBins)
plt.show()

```

```

print F(3, mu=5)
print 1 - F(7, mu=5)

```

```

for k in range(12+1):
    print k, F(k, 5)
"""
0 0.00673794699909
1 0.0404276819945
2 0.124652019483
3 0.265025915297
4 0.440493285065
5 0.615960654833
6 0.762183462973
7 0.86662832593
8 0.931906365278
9 0.968171942694
10 0.986304731402
11 0.994546908087
12 0.997981148373
"""

```

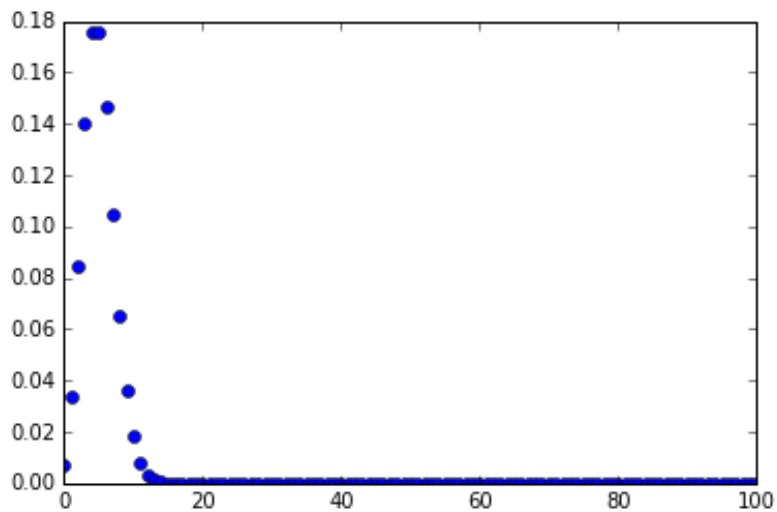


FIGURA 1. Distribución de Poisson

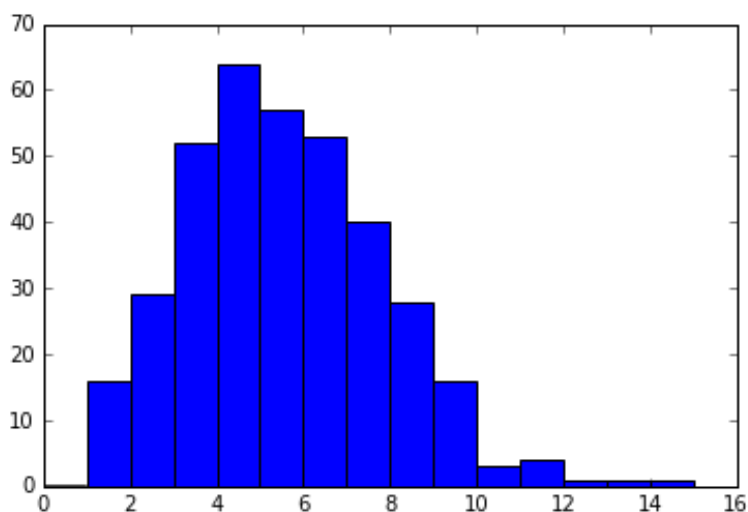


FIGURA 2. Histograma de pacientes en sala de urgencias durante un año con media $\lambda = 5$

19. Relación entre las Distribuciones Binomiales y de Poisson

Si en la función de probabilidad binomial, N es muy grande pero $p \approx 0$, esto modela un *evento raro*. En la práctica esto significa $N \gg 50, Np \ll 5$.

En este caso, la distribución Binomial con parámetros N, p se aproxima a una Poisson con parámetro $\lambda = Np$.

[fragile, allowframebreaks]relBinomPoisson.py

```
from scipy import stats
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl

mpl.style.use("ggplot")

fig, ax = plt.subplots(1, 1)

def fP(x, mu=1):
    return stats.poisson.pmf(x, mu)

def fB(x, N=30, p=0.5):
    return stats.binom(N,p).pmf(x)

N_=50
p_=5./N_
mu_ = N_*p_
x1 = np.arange(0,20+1)
ax.plot(x1, fP(x1, mu=mu_), 'bo', label="Poisson")
ax.plot(x1, fB(x1, N=N_, p=p_), 'ro', label="Binomial")
```

```
legend = ax.legend(loc='upper center', shadow=True)
plt.show()
```

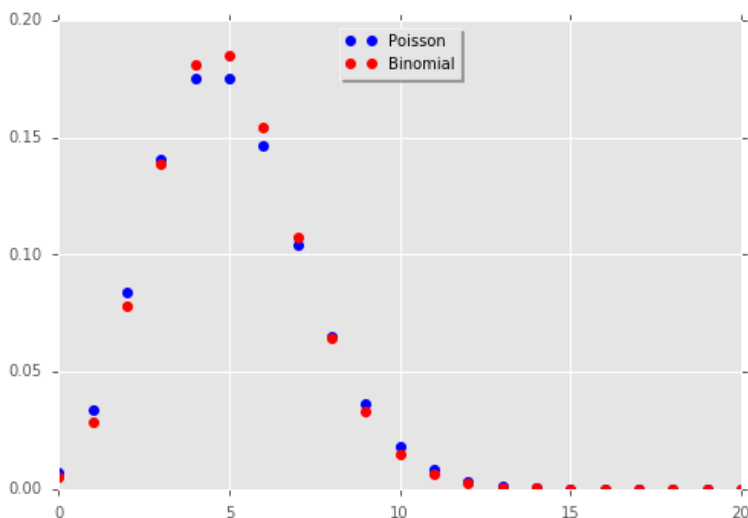


FIGURA 3. Comparación entre distribuciones Binomial y Poisson para eventos raros.

20. Distribución multinomial

Si los eventos E_1, \dots, E_k pueden ocurrir con probabilidades p_1, \dots, p_k respectivamente, entonces la probabilidad de que ocurran X_1, \dots, x_k veces respectivamente esta dado por la *distribución multinomial*

$$(129) \quad f(x_1, \dots, x_k) = \frac{x_1! \dots x_k!}{(x_1 + \dots + x_k)!} p_1^{x_1} \dots p_k^{x_k}.$$

EJEMPLO 20.1. Si un dado se lanza 12 veces, encontrar la probabilidad de obtener cada uno de los números 1, 2, 3, 4, 5, 6 exactamente dos veces.

21. Problemas Resueltos

Percentil Diremos que $x = P_q$ es el percentil q , $0 \leq q \leq 100$ de la distribución $F(x)$ si $F(P_q) = q\%$. En el caso de que q sea un valor realizable de $F(x)$, podemos “despejar”

$$(130) \quad P_q = F^{-1}\left(\frac{q}{100}\right).$$

A tal función se le llama *distribución inversa*.

Cuartiles En la literatura se definen conceptos similares. Por ejemplo, el primer *cuartil* corresponde al percentil 25; el segundo cuartil al percentil 50; y así sucesivamente.

Combinaciones

EJEMPLO 21.1. Encuentre

- (a) $5!$
 (b) $\binom{8}{3}$

utilizando Python.

```
[fragile, allowframebreaks]combinaciones.py
import math
import scipy.special

print math.factorial(5)
print scipy.special.binom(8,3)
```

Distribución Binomial

EJEMPLO 21.2. Supóngase que 15% de la población es zurda. Encontrar la probabilidad de que en un grupo de 50 individuos haya:

- (a) cuando mucho 10 zurdos;
 (b) por lo menos 5 zurdos;
 (c) entre 3 y 6 zurdos;
 (d) exactamente 5 zurdos.

```
[fragile, allowframebreaks]solvedBinom.py
from scipy import stats

#7.2 N=50, p=15%
def f(x):
    return stats.binom(50,.15).pmf(x)
def F(x):
    return stats.binom(50,.15).cdf(x)
#(a) P(X<=10)
print sum([f(x) for x in range(0,10+1)])
##0.8800826828
print F(10)
##0.8800826828
#(b) P(X>=5)
print 1-sum([f(x) for x in range(0,4+1)])
##0.887894791945
print 1-F(4)
##0.887894791945
#(c) P(3<=X<=6)
print sum([f(x) for x in range(3,6+1)])
##0.3471108697
print F(6)-F(2)
##0.3471108697
#(d) P(X=5)
print f(5)
##0.3471108697
```

Distribución Normal

EJEMPLO 21.3. En un examen final de matemáticas, la media fue 72 y la desviación estándar fue 15. Determinar las puntuaciones estándar de los estudiantes que obtuvieron:

- (a) 60;
- (b) 93;
- (c) 72.

EJEMPLO 21.4. Con los datos del problema 21.3, encontrar las calificaciones que corresponden a las siguientes puntuaciones estándar:

- (a) -1 ;
- (b) 1.6.

EJEMPLO 21.5. Supóngase que la cantidad de juegos en que participan los beisbolistas de la liga mayor durante su carrera se distribuye normalmente con media de 1500 juegos y desviación estándar 350 juegos. Emplear **Python** para responder las siguientes preguntas:

- (a) ¿Qué porcentaje participa en menos de 750 juegos?;
- (b) ¿qué porcentaje participa en más de 2000 juegos?;
- (c) encontrar el *percentil* 90 de la cantidad de juegos en los que participan en su carrera.

```
[fragile, allowframebreaks]solvedNorm.py
from scipy import stats

mu = 1500
sigma = 350
nd = stats.norm(mu, sigma)

def F(x):
    return nd.cdf(x)

#a
print F(750)
##0.3471108697

#b
print 1-F(2000)
##0.0765637255098

def inverseF(x):
    return nd.ppf(x)

#c
print inverseF(.90)
##1948.54304794
```

Eventos raros

EJEMPLO 21.6. Si la probabilidad de que un individuo tenga una reacción adversa por la inyección de determinado suero es 0.001, determinar la probabilidad de que de 2000 individuos:

- (a) exactamente 3;
 - (b) más de 2
- sufran una reacción adversa.

```
[fragile, allowframebreaks]eventosRaros.py
from scipy import stats
#7.28
#a
N = 2000
p = 0.001
print stats.binom(N,p).pmf(3)
##0.180537328032
print stats.poisson(N*p).pmf(3)
##0.180447044315
#(b)
print 1-stats.binom(N,p).cdf(2)
##0.32332356124
print 1-stats.poisson(N*p).cdf(2)
##0.323323583817
```