Unidad 5. Principios de probabilidad, desarrollo y aplicación de distribución normal de datos

5.1 Conceptos principales de la probabilidad

La probabilidad es la medida asociada con un suceso *A* y denotado por Pr(*A*) que toma un valor tal que:

$0 \leq \Pr(A) \leq 1$. Es la expresión cuantitativa de las posibilidades de que ocurra un suceso.

A mayor probabilidad Pr(*A*), mayor posibilidad de que el suceso ocurra. Si un suceso no puede ocurrir, Pr(*A*) = 0; si es seguro que ocurra Pr(*A*) = 1.

Los valores numéricos pueden ser asignados en caso simples mediante uno de los siguientes dos métodos:

a) Si el espacio muestral puede ser divido en $n(n \geq 2)$ posibles subconjuntos equiprobables y el suceso *A* está asociado a $r(0 \leq r \leq n)$ de ellos, entonces Pr(*A*) = *r/n.*
b) Si un experimento puede ser repetido un gran número de veces, *n*, y en *r* casos el suceso *A* ocurre, entonces se denota como *r*/*n* a la frecuencia relativa de *A.* Si este valor tiende a un límite cuando $n \rightarrow \infty$, este límite es Pr(*A*).

5.1.1 Teoría de la probabilidad

Analiza la noción de azar, desarrollando modelos matemáticos que permiten tratar con los fenómenos aleatorios.

En el estudio de la probabilidad se dan dos corrientes o escuelas:

a) Escuela subjetiva: Interpreta la probabilidad como una medida de la confianza, de la fe o de la intuición que se derivan de juicios subjetivos.
b) Escuela objetiva: Se basa en la suposición de que la ocurrencia o no ocurrencia de un fenómeno aleatorio es totalmente independiente de aspectos subjetivos, en esta escuela se basan los estudios científicos.

5.1.2 Principios de la probabilidad

a) Principio de indiferencia: También conocida como *principio de razón insuficiente*, plantea que cuando los eventos de un fenómeno aleatorio son simétricos, no existe razón para preferir a uno sobre otro. Es decir, todos los fenómenos son igualmente probables. Se dice que los eventos son simétricos, legales o insesgados cuando no existe razón para aumentar la probabilidad de alguno de ellos.

b) Principio de la frecuencia: cuando un fenómeno aleatorio se repite un número grande de veces tiende a presentarse una regularidad estadística que permite proyectar el comportamiento del fenómeno en el futuro.

5.1.3 Definiciones de Probabilidad:

a) Definición a priori: La probabilidad de un evento *X* es igual al número de eventos favorables, dividido entre el número total de eventos posibles.

Sea: P(*X*) ≡ La probabilidad del evento *X*
f ≡ Número de veces que puede aparecer *X*
u ≡ Número de veces que ***no puede*** aparecer *X*

$$\text{Entonces: } P(X) = \frac{f}{f+u}$$

Sea: n ≡ el número total de eventos posibles

Entonces: $n = f + u$

$$y\ P(X) = \frac{f}{n}$$

Ejemplo:

x ≡ Reyes en una baraja española
P(x)≡ Probabilidad de hallar un rey
f = 4
u = 36
n = 40

$$P(X) = \frac{4}{4+36} = \frac{4}{40} = 0.1$$

b) Definición a posteriori: conocida también como *definición de frecuencia.* En *una serie real* de pruebas, la probabilidad de un evento es igual al número de veces que ocurrió dicho evento, dividido entre el número total de ensayos.

Sea: P(*X*) ≡ La probabilidad del evento *X*
f ≡ Número de veces que ***ocurrió*** *X*
u ≡ Número de veces que ***no ocurrió*** *X*

$$\text{Entonces: } P(X) = \frac{f}{f+u}$$

Sea: n ≡ el número total de ensayos ejecutados

Entonces: $n = f + u$

$$y\ P(X) = \frac{f}{n}$$

Ejemplo:

x ≡	Sol
P(x)≡	300
f =	153
u =	147
n =	40

$$P(X) = \frac{153}{153 + 147} = \frac{153}{300} = 0.51$$

Como puede observarse, ambas definiciones usan la misma fórmula. La diferencia es que la *a priori* se calcula antes de ejecutar el fenómeno (por eso es conocida también como *probabilidad teórica*) mientras que la *posteriori* se calcula después (denominándose *probabilidad empírica*).

5.1.4 Axiomas básicos de la probabilidad

Axioma

Planteamiento es indemostrable, básico, elemental y que es evidente por sí mismo.

Teorema

Planteamiento o proposición compleja que tiene que ser demostrable por medio de axiomas u otros teoremas. Es una proposición demostrable con base a axiomas.

Corolario:

Planteamiento que se deduce de un teorema. Son los pasos para demostrar un teorema.

Como toda teoría matemática, la teoría de la probabilidad tiene conceptos básicos sobre los que descansa todo su desarrollo. Estos conceptos, conocidos como axiomas, son planteamientos simples e indemostrables.

Axiomas de la teoría de la probabilidad

Axioma 1 o *Axioma de los límites de la probabilidad*: Sea P(x) la probabilidad de cualquier evento, entonces: $0 \leq P(x) \leq 1$ la probabilidad de cualquier evento se mueve entre 0 y 1.

Observaciones: $\nexists$ $P(x) \in R^-$, $\nexists$ $P(x) > 1$ — Axioma de los límites de la probabilidad

Si $P(x) = 1$ Se llama *evento seguro*

Si $P(x) = 0$ Se llama *evento imposible*

$100\ P(x)$ Es un valor porcentual de la probabilidad

Axioma 2 o *axioma del espacio muestral*: Sea P(s) la probabilidad del espacio muestral, entonces: Sea P(s)=1. La probabilidad de los eventos del espacio muestral y es 1.

Observación: $P(s) = P(X_1)+ P(X_2)+\ldots+ P(X_n)$

Axioma 3: la probabilidad de ocurrencia de uno o más eventos posibles, es igual a la suma de eventos posibles.

$$P\ (X_1\ UX_2\ UX\ldots\ UX_n) = P(X_1)+ P(X_2)+\ldots+ P(X_n)$$

Observación: nótese que $P\ (X_1\ UX_2\ UX\ldots\ UX_n)$ es un evento compuesto, formado por los eventos simples X_1, X_2, …, X_n

Reglas de adición y multiplicación de probabilidades

Existen dos reglas para realizar las operaciones aritméticas básicas con probabilidades. Su aplicación es la base de cualquier cálculo de probabilidades.

Regla de adición: la probabilidad de que ocurra uno de dos o más eventos posibles es igual a la suma de estos eventos.

Ejemplo uno: ¿cuál es la probabilidad de obtener un 5 o un 6 al lanzar un dado.

$$P(5\ U\ 6) = P(5)+P(6)$$

$$c= 1/6 +1/6$$

$$P(5\ U\ 6) = 2/6 = 1/3 = 0.3333 = 33.33\%$$

Ejemplo dos: ¿Cuál es la probabilidad de encontrar una *figura (jocker, reina o rey)* en una baraja americana de 52 cartas?

Sea: $J \equiv$ Jocker

$Q \equiv$ Reina

$K \equiv$ Rey

$$P(J\ U\ Q\ U\ K) = P(J)+P(Q)+P(K)$$

$$= 4/52+4/52+4/52$$

$$P(J\ U\ Q\ U\ K) = 12/52 = 0.2308 = 23.08\%$$

Ejemplo tres: se tiene una botella con 14 canicas blancas, 32 rojas y 27 negras

a) ¿Cuál es la probabilidad de encontrar roja?

b) ¿Blanca o negra?
c) ¿Roja o negra?
d) ¿Blanca, o roja o negra?

B =	14
R =	32
N =	27

a) P(R) = f/f+w = 32/(32+41) = 32/73 = 43.84%
b) P(B U N) = P(B)+P(N) = 17/73 +27/73 = 41/73 = 56.16%
c) P(R U N) = P(R)+P(N) = 32/73 +27/73 = 59/73 = 0.8082 = 80.82%
d) P(B U R U N) = P(S) = 1 = 14/73+32/73+27/73 = 73/73 = 1 = 100%

Ejemplo cuatro: en las escuelas secundarias del Estado de México se diagnosticó la aptitud escolar de los alumnos que ingresan, y se encontró lo siguiente:

Excelente	2426
Bueno	6131
Aceptable	15603
Malo	7121
Inaceptable	2104
Total	32785

Suponiendo que para ingresar a la universidad no existiera examen de admisión ¿cuál sería el riesgo de aceptar alumnos cuya aptitud sea inferior a la aceptable? ¿Correría ese riesgo?

Lo que se pide es que se estime la probabilidad de encontrar alumnos con aptitud mala o inaceptable, lo que se puede representar:

P(M U I)

Por la regla de la suman: *P(M U I) = P(M)+ P(I)*

Por la definición de probabilidad: *P(M) = M/n = 7121/32785*

P(I) = I/n = 2104/32785

Entonces: *P(M U I)= 7121/32785 + 2104/32785 = 9225/32785 = 28.14%*

Por lo tanto, el riesgo de aceptar alumnos con aptitud inferior a lo aceptable es de: 28.14%.

Aceptamos por convención, que el riesgo máximo que si puede correr es del 15%, entonces es inaceptable el riesgo de no aplicar examen de admisión.

5.1.5 Diagramas de árbol

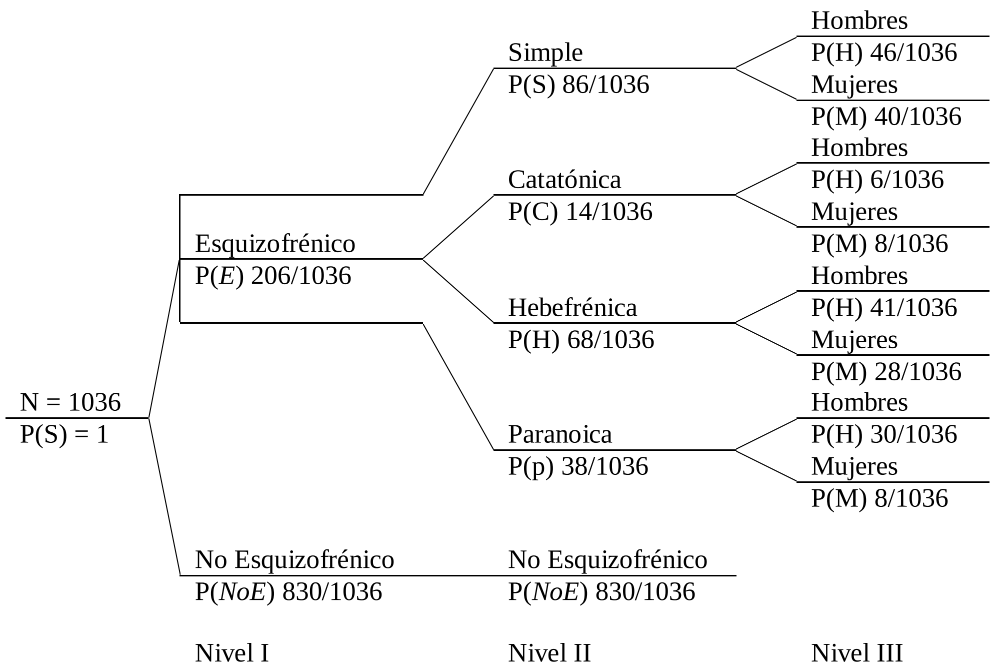
Esto diagramas sirven para representar el evento sucesivos, y sus probabilidades de ocurrencia. Para construirlos deben observarse los siguientes puntos:

1. Siempre inician en N, tal que P(S) = 1
2. De N (quince conoce como *tronco)* pueden desprenderse dos o más eventos o ramas. Cada rama tiene P(xi) tal que ΣP(xi) = P(S) = 1

3. A su vez, de cada rama pueden desprenderse 2 o más ramas, tal que la suma de la probabilidad de un conjunto de ramas es igual a la probabilidad de la rama que lo originó.
4. En todos los casos, la suma de un nivel de ramas es P(S) = 1. En todos los casos el denominador es N.

Ejemplo: En cierto hospital para enfermos mentales existe 1036 internos de los cuales 206 fueron diagnosticados como esquizofrénicos. De los esquizofrénicos 86 presentan esquizofrenia es simple, 14 catatónica, 68 hebefrénica y 38 paranoica. Represente el diagrama de árbol:

Total	1036
Esquizofrénicos	206
Simple	86
Catatónica	14
Hebefrénica	68
Paranoica	38



Reglas de la multiplicación: La probabilidad de que ocurra un evento B, después de ocurrir un evento A, ser representa: P(A| B) probabilidad de que ocurra A y después B.

Si el evento B es independiente del evento A, se aplica la regla de la multiplicación, que consiste en multiplicar P(A) P(B).

Ejemplo uno: ¿Cuál es la probabilidad de que al lanzar dos veces una moneda caiga primero cara y después cruz?

Observaciones:

1. Nótese que existen dos eventos:

$A \equiv$ Cara en el primer lanzamiento
$B \equiv$ Cruz en el segundo lanzamiento

Entonces, lo que plantea el problema es $P(A|B)$

2. A y B son eventos independientes, pues el resultado de un segundo lanzamiento (B) no es afectado por el primero (A), por tanto puede aplicarse la regla de la multiplicación.
   Analizaremos gráficamente el problema.

$S \equiv$ Espacio muestral
$H \equiv$ Cara
$T \equiv$ Cruz

$$S = \{HH, HT, TT, TH\} |\ P(X) = 1/4$$

La regla de la multiplicación indica la probabilidad de que ocurra un evento seguido de otro.

$$P(A|B) = P(A)\ P(B) = (1/2)\ (1/2) = (1/4)$$

Ejemplo dos: Si se lanza tres veces una moneda ¿cuál es la probabilidad de que ocurran cruz, cara y cruz?

$A \equiv$ T
$B \equiv$ H
$C \equiv$ T

$$P(A|B|C) = P(A)\ P(B)\ P(C) = (1/2)\ (1/2)\ (1/2) = (1/8)$$
$$S = \{HHH, HHT, TTT, TTH, THH, HTH, THT\} |\ P(X) = 1/8$$

Ejemplo tres: Se tiene una botella con cinco canicas rojas y tres azules ¿cuál es la probabilidad de que primero aparezca roja y después azul?

$R \equiv$ 5
$A \equiv$ 3

$$P(R|A) = P(R)\ P(A) = (5/8)\ (3/8) = (15/64) = 23.43\%$$

Ejemplo cuatro: de los alumnos que ingresaron a una escuela, 321 provienen de preparatoria de Toluca y quince de preparatorias de fuera de Toluca. En un grupo se tomó una muestra aleatoria de tres alumnos; el primero fue de Toluca, el segundo de fuera y el tercero de fuera. ¿es razonable suponer que el grupo se formó aleatoriamente?

$T \equiv$ 321
$F \equiv$ 15
T, F, F
P(T, F, F)

Si el grupo fuera aleatorio esperaríamos:

$$P(T|F) = (321/336)\ (15/336) = 0.0426 = 4.26\%$$

Apliquemos las reglas de la multiplicación:

$$P(T|F|F) = P(T)\ P(F)\ P(F) = (321/336)\ (15/336)(15/336)$$
$$= 0.0019 = 0.19\%$$

entonces:

Probabilidad P_O: $P_o <$ En conclusión: no es razonable

observada:	0.19%	P_e	suponer que el grupo se formó
Probabilidad esperada:	P_e: 4.26%		aleatoriamente

5.1.6. El triángulo de Pascal

1	**1**										
1	**2**	**1**									
1	**3**	**3**	**1**								
1	**4**	**6**	**4**	**1**							
1	**5**	**10**	**10**	**5**	**1**						
1	**6**	**15**	**20**	**15**	**6**	**1**					
1	**7**	**21**	**35**	**35**	**21**	**7**	**1**				
1	**8**	**28**	**56**	**70**	**56**	**28**	**8**	**1**			
1	**9**	**36**	**84**	**126**	**126**	**84**	**36**	**9**	**1**		
1	**10**	**45**	**120**	**210**	**252**	**210**	**120**	**45**	**10**	**1**	
1	**11**	**55**	**165**	**330**	**462**	**462**	**330**	**165**	**55**	**11**	**1**

1.1.7. Teorema de *Bayes* (probabilidad condicional)

Dados dos eventos A y B mutuamente dependientes, que tiene:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}$$

Donde: $P(\bar{A}) = 1 - P(A)$

Ejemplo uno: supóngase que a una persona elegir al azar de la ciudad de Toluca se le aplica un test para medir neurosis, y que el diagnóstico es positivo ¿cuál es la probabilidad de que el sujeto sea realmente neurótico?

Para poder calcular esta probabilidad, obtenemos los siguientes datos del departamento de salud mental:

1. La probabilidad de que un sujeto neurótico obtenga un resultado positivo con el que es de 0.98, esto es: P(Diagnóstico | Error) = 0.05
2. Existe el 0.01 en neuróticos en la ciudad. Esto es, para elegir aleatoriamente a un neurótico: P(Neurótico) = 0.01
   Análisis: denotemos que

   P(A) ≡ probabilidad de hallar un neurótico

   P(A) = 0.01

   P(B) ≡ diagnóstico positivo acertado

   P(B) = 0.98

   $P(\bar{A}) =$ Error de diagnóstico

   $P(\bar{A}) =$0.05

P(Diagnóstico acertado) = $1 - P(A)$ = 1- 0.05 = 0.95

P(Diagnóstico negativo | No neurosis)

P(Diagnóstico acertado)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}$$

$$P(A|B) = \frac{(0.98)(0.01)}{(0.98)(0.01) + (0.048)\left(\overline{0.05}\right)}$$

$$P(A|B) = 0.8033 = 80.33\%$$

P(Neurótico diagnóstico positivo) = 80.43%

Ejemplo dos: en cierta facultad 60% de los estudiantes son mujeres, de las mujeres 1% reprueban por falta de asistencia y de los hombres 4% reprueban por falta de asistencia. Si seleccionamos una boleta de calificaciones al azar y el resultado es “reprobado por falta de asistencia” ¿cuál es la probabilidad de que corresponda a una mujer?

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}$$

A ≡ Mujer
B ≡ Reprobado por falta de asistencia
Ā ≡ Hombre

P(A) = 0.60
P(B|A) = 0.01
P(B|Ā) = 0.04
P(Ā) = 0.40

$$P(A|B) = \frac{0.01 \times 0.60}{0.01 \times 0.60 + 0.04 \times 0.40}$$

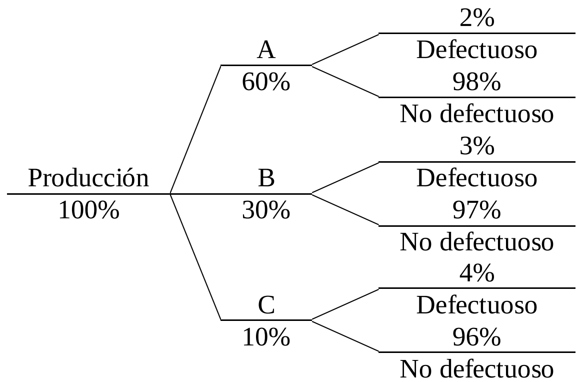
$$P(A|B) = 0.2727 = 27.27\%$$

Ejemplo tres: en una fábrica existen tres departamentos, el del producto A, el del producto B y el del producto C. Y se producen en respectivamente el 60%, el 30% y el 10% del total de los artículos. Los porcentajes de artículos defectuosos son respectivamente 2%, 3% y 4%. En las ventas regresaron artículo defectuoso ¿cuál es la probabilidad de que haya sido producido por el departamento C?

Nota: obsérvese que esta es una generalización del teorema de Bayes.

Departamentos	Producción	Art. defectuosos
A	60%	2%
B	30%	3%

C	10%	4%



Sea	X≡	Artículo defectuoso
	P(A)=	0.60
	P(B)=	0.30
	P(C)=	0.10
	P(A|X)=	0.02
	P(B|X)=	0.03
	P(C|X)=	0.04

La generalización del teorema de Bayes es:

$$P(Ai|B) = \frac{P(Ai)P(B|Ai)}{P(A_1)P(A_1|B) + P(A_2)P(A_2|B) + \cdots P(A_n)P(A_n|B)}$$

Para nuestro caso:

$$P(C|X) = \frac{P(C)P(X|C)}{P(A)P(A|X) + P(B)P(B|X) + P(C)P(C|X)}$$

$$P(C|X) = \frac{0.10 \times 0.04}{0.60 \times 0.02 + 0.30 \times 0.03 + 0.10 \times 0.04}$$

$$P(C|X) = 0.16 \ \approx 16\%$$

Conclusión: la probabilidad de que el artículo defectuoso devuelto haya sido producido por el departamento C es del 16%.

5.1.7. Combinaciones y permutaciones

Definición

Sea *X* un elemento de los números Z^+

Sea $X \epsilon Z^+$, entonces, el factorial de *X* que denotaremos *X!* está definido por:

$$X! = X(X-1)(X-2)\ldots 1$$

Ejemplo: 5! = 5 x 4 x 3 x 2 x 1 = 120

Observaciones:

1	$\nexists\ X! \mid X \in Z$	$\nexists\ (-5)!$
2	$\nexists\ X! \mid X \in R$	$\exists\ (2.5)!$
3	$\nexists\ X! \mid X \in Q$	$\exists\ (1/4)!$

Definiciones:

1) 0! = 1
2) 1! = 1

Combinaciones

Una combinación es un arreglo de elementos que considera el orden.

Un arreglo personal muestra, por tanto el número de muestras diferentes de tamaño *n* que pueden obtenerse de una población de tamaño *N*, con remplazo y considerando el orden, es:

$$NCn = N^n$$

N ≡ Tamaño de la población		Consideraciones:
n ≡ Tamaño de la muestra		1. **¡Error! No se encuentra el origen de la referencia.**
C≡ Combinaciones		2. Importa el orden

Ejemplo uno:

W = {1, 2, 3, 4}
N = 4
n = 2

¿Cuántas muestras diferentes se pueden sacar?

(1,1) (1,2) (1,3) (1,4) (2,1) (2,2) (2,3) (2,4) (3,1) (3,2) (3,3) (3,4) (4,1) (4,2) (4,3) (4,4)

$$NCn = N^n = 4C2 = 4^2 = 16$$

Ejemplo dos: Se tienen ocho libros, y se van prestar a parejas de estudiantes, deseamos considerar el orden (a qué estudiante se prestó) ¿de cuántas formas es posible prestar los libros?

$$NCn = N^n$$
$${}_8C_2 = 8^2$$
$${}_8C_2 = 64$$

Ejemplo tres : Se tiene una batería de seis pruebas psicométricas, que serán aplicadas a cuatro personas. Un investigador desea estudiar cómo afecta el orden de aplicación a los resultados de los test ¿Cuántas aplicaciones diferentes tendrá que hacer?

Test:	1, 2, 3, 4, 5, 6
Muestras:	(1, 2, 3, 4) (1, 2, 3, 5) (1, 2, 3, 6)
${}_6C_4$ =	6^4
${}_6C_4$ =	1296

5.6.3. Permutaciones

Una permutación es un arreglo en el que se importa el orden y no existe reemplazamiento. Las permutaciones quedan definidas por:

$$NP_n = \frac{N!}{(N-n)!}$$

$n \leq N$

$n = N \Rightarrow M - n = 0$

$0! = 1$

Así, el número de muestras diferentes que pueden obtenerse de una población, sin reemplazo y con orden se obtiene por la fórmula anterior:

$$4P_2 = \frac{4!}{(4-2)!}$$

	1	2	3	4
1	~~1,1~~	1,2	1,3	1,4
2	2,1	~~2,2~~	2,3	2,4
3	3,1	3,2	~~3,3~~	3,4
4	4,1	4,2	4,3	~~4,4~~

¿(1,2) = (2,1-)? El orden importa, no es lo mismo

5.7 Binomio

El número de muestras diferentes que pueden obtenerse de una población, sin reemplazo y sin orden está definido por el binomio factorial:

$$\binom{N}{n} = \frac{N!}{(N-n)!\,n!} \qquad \text{Binomio de Newton}$$

$$\binom{4}{2} = \frac{4!}{(4-2)!\,2!} = 6$$

	1	2	3	4
1	~~1,1~~	1,2	1,3	1,4
2	~~2,1~~	~~2,2~~	2,3	2,4

3	~~3,1~~	~~3,2~~	~~3,3~~	3,4
4	~~4,~~	~~4,2~~	~~4,3~~	~~4,4~~

En resumen:

1) Con orden
   Con reemplazo $$NC_n = N^n$$
2) Con orden
   Sin reemplazo $$NP_n = \frac{N!}{(N-n)!}$$
3) Sin orden
   Sin reemplazo $$\binom{N}{n} = \frac{N!}{(N-n)!\,n!}$$

5.8 Números aleatorios

En probabilidad y estadística casi siempre tratamos con muestras aleatorias. El tipo de muestreo más usado es el muestreo elemental, donde todos y cada uno de los elementos tienen exactamente la misma posibilidad de ocurrencia, es decir, son *equiprobables*.

Observemos algunas situaciones prácticas en que sea necesario muestrear aleatoriamente a partir de una población dada. Por ejemplo, si la población no es muy grande, podemos usar papeles doblados o canicas con la identificación del espacio muestral; colocándolos en una urna, revolverlos (aleatorizando) y extraer los elementos hasta completar *n*. Este procedimiento parece razonable. Sin embargo, aunque sería demasiado complejo, puede demostrarse que de hecho no es un muestreo completamente aleatorio (se le llama *pseudoaleatorio*), basta poner por ejemplo, que es un procedimiento usado por las loterías y estudios estadísticos muy detallados han demostrado que no es completamente aleatorio.

Un procedimiento bastante mejor y mucho más sencillo de realizar, es el uso de *tablas de números aleatorios*. Llamaremos a estas tablas y al programa computarizado *generadores de aleatorios*. Estos han sido diseñados de tal forma que los dígitos que aparecen son equiprobables en su ocurrencia y además independientes (pues existe reemplazo). Los pasos para generar aleatorios con una tabla o con una máquina son:

1. Determinar *N* y *n*
2. Determinar cuánto dígitos tiene *N*
3. Lista y numerar todos los elementos de la población, agregando ceros a la izquierda para que todos los números de identificación de los elementos de la población tengan el mismo número de dígitos que *N*; el primer número siempre es 0.
4. Si se usa una tabla deberá definirse un punto inicial cualquiera (al principio, en el centro, al final, etc.) y una ruta que también podrá ser cualquiera. Si se usa una calculadora, bastará tomar el primer número que parezca y leer dígitos de izquierda a derecha. A través de una computadora en el programa Excel tiene la posibilidad de generar números aleatorios.

5. Considerando el punto dos vemos que:
   5.1 Si $N \leq 10$ se tomará un dígito entre 0 – 9
   5.2 Si $N \leq 11 \leq 100$ se tomarán dos dígitos entre 00 – 99
   5.3 Si $N \leq 101 \leq 1000$ se tomarán tres dígitos entre 000 – 999
   5.4 Se generaliza para cualquier N
6. Los primeros números n que aparezcan se tomarán como elementos muestrales, considerando:
   6.1 Si es con remplazo, se aceptan números repetidos
   6.2 Si es sin reemplazo, se eliminen números repetidos

Puede afirmarse que las calculadoras y las computadoras son generadores de aleatorios bastante mejor que cualquier tabla, ya que el número de combinaciones es prácticamente infinito, mientras que en cualquier tabla es necesariamente finito.

Quizá el problema más serio en la práctica sea listar todos los elementos de la población, pues existen algunas extremadamente grandes, o aún desconocida. En estos casos conviene estudiar técnicas avanzadas de muestreo.

5.9 La ley de los grandes números

Establece un concepto básico en probabilidad y estadística que tiene aplicaciones a múltiples áreas (una de ellas es la teoría del muestreo).
Formalmente, la ley de los grandes números puede establecerse como sigue:

Sea: PO ≡ Probabilidad observada (a posteriori)
PT ≡ Probabilidad teórica (a priori)
ξ ≡ Cualquier número suficientemente pequeño

Entonces: $P(|P_0 - P_1 > \xi) \rightarrow 0$ cuando $n \rightarrow \infty$ que podríamos leer:
"La probabilidad de que la probabilidad observada y quiere en forma absoluta de la probabilidad teórica en una cantidad mayor que ξ, tiende a cero cuando el número de ensayos tiende a infinito".
Ejemplo: Se lanzaron tres series de una moneda y se obtuvo:

Número de lanzamientos	Frecuencias f(H)	f(T)	$P_0(H)$	$P_T(H)$	$\lvert P_0(H)\ P_T(H)\rvert$	ξ	
10	7	3	7/10	1/2 – 5/10	2/10 = 0.20	0.05	$\lvert P_0 - P_T\rvert > \xi$
50	40	10	40/50	1/2 – 25/50	15/50 = 0.30	0.05	$\lvert P_0 - P_T\rvert > \xi$
500	264	236	264/500	1/2 – 250/500	14/500 = 0.03	0.05	$\lvert P_0 - P_T\rvert < \xi$

Número de lanzamientos	Frecuencias f(H)	f(T)	$P_0(T)$	$P_T(T)$	$\lvert P_0(T)\ P_T(H)\rvert$	ξ	
10	4	6	6/10	1/2 – 5/10	1/10 = 0.10	0.05	$\lvert P_0 - P_T\rvert > \xi$
50	26	24	24/50	1/2 – 25/50	1/50 = 0.02	0.05	$\lvert P_0 - P_T\rvert < \xi$
500	175	325	325/500	1/2 – 250/500	25/500 = 0.15	0.05	$\lvert P_0 - P_T\rvert > \xi$

Obsérvese que se contradice la ley que los grandes números, pues a mayor número de ensayo, mayor número de error, por lo tanto se concluye que las series de lanzamientos no son legales.
Con palabras menos rigurosos, podríamos decir que la ley de los grandes números establece que:

a) La probabilidad a posteriori (observada) tiende a ser igual a la a priori (teórica) de la cuando *n* tiende a cero.
b) El error tiende a cero cuando *n* tiende a crecer.

5.9.1. Desigualdad de Tchebyser

Hay que recordar dos conceptos básicos: el recíproco multiplicativo y la estandarización de una variable.

Recíproco multiplicativo

Sea *X*, un número real cualquiera, entonces el recíproco de *X* es:

$$\frac{1}{X} \text{ y es que tal que } X\left(\frac{1}{X}\right) = 1$$

5.9.2 Estandarización de una variable:

Si *X* es una variable con media (μ) y desviación estándar (σ), entonces el valor estandarizado de *X* que notamos por *z*, está dada por:

$$z = \left|\frac{x - y}{\sigma}\right|$$

La desigualdad de Tchebyser establecer una relación fundamental entre la media, la desviación estándar y el concepto de probabilidad. Quizá la mayor importancia de este principio resida en el hecho de que es independiente de la forma en que se distribuye la variable de interés.

Informa operativas, la desigualdad de Tchebyser puede plantearse como sigue:

Sea: X ≡ Probabilidad observada (a posteriori)

μ ≡ Probabilidad teórica (a priori)

σ ≡ Cualquier número suficientemente pequeño

k ≡ Un número real positivo cualquiera

Entonces:

$$P\left(\left|\frac{X-\mu}{\sigma}\right| \geq k\right) \leq \frac{1}{k^2}$$

Que podemos leer como: “la probabilidad de que una variable aleatoria estandarizada tenga un valor absoluto mayor o igual a cualquier número positivo, es siempre menor o igual que el recíproco del cuadrado de dicho número”.

Ejemplo: En una escuela existe una media de 105 y una desviación estándar de 8 ¿cuál es la probabilidad de hallar un alumno con C.I. mayor o igual a 115?

X ≥ 115

μ = 105

σ ≡ 8

$$P(|11.25| \geq k) \leq \frac{1}{k^2}$$

$$P\left(\left|\frac{X-\mu}{\sigma}\right| \geq 1.25\right) \leq \frac{1}{1.25^2}$$

$$P\left(\left|\frac{X-\mu}{\sigma}\right| \geq 1.25\right) \leq 0.64$$

En otras palabras, existe una probabilidad no mayor de 64% de hallar alumnos que se alejen diez unidades de C.I., en Valor absoluto de su media. Es decir, la probabilidad de hallar alumnos, 95≥C.I>115 es no mayor al 64%.

Ejemplo: En cierta fábrica, el promedio de inasistencias al mes en los obreros este 2.8, con una desviación estándar de 0.17 ¿cuál es la probabilidad de que un obrero elegido aleatoriamente presente siete inasistencias o más?

X ≥ 7

μ = 2.8

σ ≡ 0.17

$$P\left(\left|\frac{X-\mu}{\sigma}\right| \geq k\right) \leq \frac{1}{k^2}$$

$$P\left(\left|\frac{7-2.8}{0.17}\right| \geq k\right) \leq \frac{1}{k^2}$$

$$P\left(\left|\frac{4.2}{0.17}\right| \geq k\right) \leq \frac{1}{k^2}$$

$$P(|24.7| \geq k) \leq \frac{1}{k^2}$$

$$P\left(\left|\frac{X-\mu}{\sigma}\right| \geq 24.7\right) \leq \frac{1}{24.7^2}$$

$$P\left(\left|\frac{X-\mu}{\sigma}\right| \geq 24.7\right) \leq \frac{1}{1.64\mathrm{E}-0.3}$$

$$P\left(\left|\frac{X-\mu}{\sigma}\right| \geq 24.7\right) \leq 0.0016$$

La probabilidad de hallar un obrero que presiente siete o más inasistencias o ninguna, no es mayor del 0.16%.

5.10 Teorema del límite central

Describe la distribución de la media de una muestra aleatoria proveniente de una población con varianza finita. Cuando el tamaño de la muestra es lo suficientemente grande, la distribución de las medias sigue aproximadamente una distribución normal.

La distribución normal, es la distribución más poderosa en la inferencia estadística. Aunque existen muchos otros tipos de distribuciones, muchas de ellas presentan funciones cuyo cálculo es demasiado difícil, o aún, imposible.

El teorema del límite central permite calcular la función de distribución de cualquier muestra a partir de la $\overline{N}$, y como sabemos, ésta ha sido estandarizada y ya existen tablas para el cálculo de áreas.

Así, el teorema del límite central juega un papel tan importante que ha sido llamado *la columna vertebral de la estadística Inferencial*.

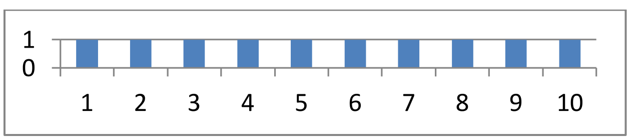
Sin embargo, en matemáticas pocos teoremas son tan complejos como el teorema del límite central. A continuación se presenta una forma simplificada de este teorema para tratar de comprender su significado, aunque sólo sea aproximado.

Si una población cualquiera tiene media μ y varianza σ^2 finitas, entonces la distribución muestra de la media $\bar{X}$ calculada de n observaciones aleatorias independientes, se aproximaba a una distribución normal con varianza $S^2 = \sigma^2/n$ y media $\bar{X} = \mu$ cuando n tiende a crecer. Cuando n es suficientemente grande, la distribución normal es una buena aproximación para la distribución muestral.

Ejemplo uno: Cierta escuela empleado 10 maestros. La variable de interés que se denotará por X es el número de años de experiencia docente de cada profesor. Se midió X y se obtuvo:

Maestro	X		X	f
1	6		1	1
2	1		2	1
3	2	La	3	1
4	9	distribución	4	1
5	5	de la	5	1
6	8	población	6	1
7	4	será:	7	1
8	3		8	1
9	10		9	1
10	7		10	1

Y su representación gráfica será:



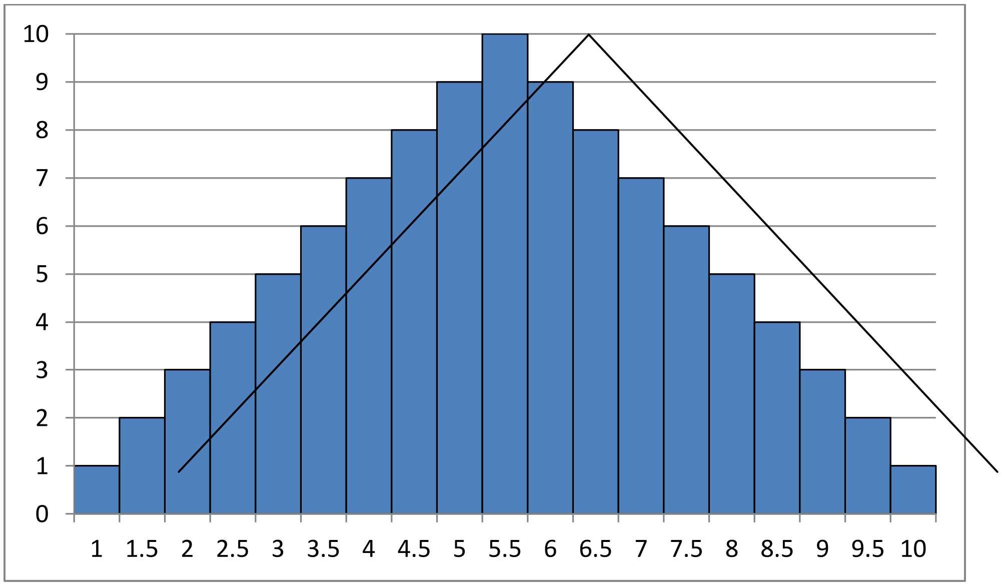
Vemos que la población muestra una distribución uniforme.

Veamos ahora las muestras de tamaños dos que puede extraerse de esta población en muestreo con remplazo (para cumplir la condición de independencia que establece el teorema del límite central).

	1	2	3	4	5	6	7	8	9	10
1	1-1 1	1-2 1.5	1-3 2	1-4 2.5	1-5 3	1-6 3.5	1-7 4	1-8 4.5	1-9 5	1-10 5.5
2	2-1 1.5	2-2 2	2-3 2.5	2-4 3	2-5 3.5	2-6 4	2-7 4.5	2-8 5	2-9 5.5	2-10 6
3	3-1 2	3-2 2.5	3-3 3	3-4 3.5	3-5 4	3-6 4.5	3-7 5	3-8 5.5	3-9 6	3-10 6.5
4	4-1 2.5	4-2 3	4-3 3.5	4-4 4	4-5 4.5	4-6 5	4-7 5.5	4-8 6	4-9 6.5	4-10 7
5	5-1 3	5-2 3.5	5-3 4	5-4 4.5	5-5 5	5-6 5.5	5-7 6	5-8 6.5	5-9 7	5-10 7.5
6	6-1 3.5	6-2 4	6-3 4.5	6-4 5	6-5 5.5	6-6 6	6-7 6.5	6-8 7	6-9 7.5	6-10 8
7	7-1 4	7-2 4.5	7-3 5	7-4 5.5	7-5 6	7-6 6.5	7-7 7	7-8 7.5	7-9 8	7-10 8.5
8	8-1 4.5	8-2 5	8-3 5.5	8-4 6	8-5 6.5	8-6 7	8-7 7.5	8-8 8	8-9 8.5	8-10 9
9	9-1 5	9-2 5.5	9-3 6	9-4 6.5	9-5 7	9-6 7.5	9-7 8	9-8 8.5	9-9 9	9-10 9.5
10	10-1 5.5	10-2 6	10-3 6.5	10-4 7	10-5 7.5	10-6 8	10-7 8.5	10-8 9	10-9 9.5	10-10 10

Al construir la distribución de frecuencias de $\bar{X}$:

$\bar{X}$	f
1	1
1.5	2
2	3
2.5	4
3	5
3.5	6
4	7
4.5	8
5	9
5.5	10
6	9
6.5	8
7	7
7.5	6
8	5
8.5	4
9	3
9.5	2
10	1



Conclusión: la distribución de la población es uniforme, pero sus medias muestrales se distribuyen en forma muy aproximada a la normal.

5.11 Distribución binomial

Características generales: la distribución binomial es una distribución discreta en probabilidad, que fue desarrollada por J. Bernoulli. La distribución binomial solo aceptados eventos (momios) posibles, uno de los cuales se denomina arbitrariamente *éxito* (p) y el otro *fracaso* (q). Es importante notar que cualquier espacio muestral, puede reducirse a un espacio binomial. En la distribución binomial el experimento se repite n veces, de forma independiente, y se calcula la probabilidad de un determinado número de éxitos

Para considerarse una distribución binomial debe cumplir con tres requisitos:

1) Que la probabilidad solo pueda ser fracaso o éxito.
2) Los resultados de cada ensayo son independientes entre sí.
3) En cada ensayo la probabilidad sea la misma

Así, puede plantearse la siguiente relación: $P(S) = 1$

Por definición:

$b =$ Distribución binomial

$S(b) = \{p, q\}$

$\Rightarrow p + q = 1$

$p = 1 - q$

$q = 1 - p$

Cálculo de probabilidad de binomiales

Sea:	b ≡	Probabilidad binomial		p ≡	Éxito		n ≡	Número de ensayos (tamaño de la muestra)
	k ≡	Número de éxitos		q ≡	Fracaso			

Entonces:

$$b(K|n,p) = \binom{n}{k}(p^k)(q^{n-k})$$

Pero se sabe que:

$$\binom{n}{k} = \left(\frac{n!}{(n-k)!\,k!}\right)(p^k)(q^{n-k})$$

Factorial *n*!

Señala la multiplicación de una serie de números que desciende:

n	$n!$		
1	1	= 1	= 1
2	2 x 1	= 2 x **1!**	= 2
3	3 x 2 x 1	= 3 x **2!**	= 6
4	4 x 3 x 2 x 1	= 4 x **3!**	= 24
5	5 x 4 x 3 x 2 x 1	= 5 x **4!**	= 120
6	etc.	etc.	

Ejemplo: en cierta colonia existe un 40% de familias que ven cierto programa de t.v. Si tomamos una muestra de 30 familias ¿Cuál es la probabilidad de que exactamente 15 vean el programa de t.v.?

Análisis:

Se define:

p ≡ Probabilidad de hallar una familia que vea el programa
p = 0.40
q ≡ Probabilidad de hallar una familia que no vea el programa
p = 1 – p = 1 – 0.40 = 0.60
k ≡ Número de éxitos
k = 15
n = 30

Sustituyendo:

$$b(15|30{,}0.40) = \left(\frac{30!}{(30-15)!\,15!}\right)(0.40^{15})(0.60^{30-15})$$

$$b(15|30{,}0.40) = 0.0783$$

Conclusión: La probabilidad de que en la colonia estudiada se hallen exactamente 15 familias que vean el programa de t.v. en una muestra de 30, es de 7.83%,

$$p(k = 15) = 7.83$$

$$p(k \geq 15) = p(k = 15) + p(k = 16) + \ldots + p(k = 30)$$

5.11.1 Tabla de distribución binomial

Puede observar en el ejemplo anterior que el cálculo directo de probabilidades binomiales, puede volverse demasiado laborioso. Sin embargo, existen tablas calculadas para diferentes tamaños de n, $p \wedge k$.

Ejemplo: En una escuela, existe un 25% de alumnos hombres. Si se extrae una muestra aleatoria de 13 alumnos ¿Cuál es la probabilidad de que exactamente 4 sean hombres?

Análisis:

Definimos:

p ≡	Probabilidad de hallar un alumno hombre
p =	0.25
q ≡	Probabilidad de no hallar un alumno hombre
p =	1 – p = 1 – 0.25 = 0.75
k ≡	Número de éxitos
k =	4
n =	13

Sustituyendo:

$$b(4|13{,}0.25) = \left(\frac{13!}{(13-4)!\,4!}\right)(0.25^{4})(0.75^{13-4})$$

$$b(4|13{,}0.25) = \left(\frac{13\ *12\ *11*10*9*8}{(9*8*7*6*5)!\,4!}\right)(0.25^{4})(0.75^{13-4})$$

$$b(4|13{,}0.25) = 0.2097$$

Conclusión: Existe una probabilidad de 20.97% de hallar exactamente cuatro hombres en la muestra de trece alumnos.

Ejemplo 2: En una fábrica 40% del personal son obreros. Si elegimos aleatoriamente a 18 empleados de la fábrica ¿Cuál es la probabilidad de que exactamente 10 sean obreros?

Análisis:

Definimos:
p = 0.40
p = 0.60
k = 10
n = 18

Sustituyendo:

$$b(10|18{,}0.40) = \left(\frac{18!}{(18-10)!\,10!}\right)(0.40^{10})(0.60^{18-10})$$

$$b(10|18{,}0.40) = 0.0771$$

Conclusión: Existe una probabilidad de 7.71% de hallar exactamente 10 obreros en la muestra de 18 empleados.

5.12 Distribución de Poisson

Es una de las más importantes distribuciones de *variable discreta*. Se aplica en la modelización de situaciones en las que interesa determinar el número de hechos de cierto tipo que se *pueden producir en un intervalo de tiempo o de espacio*, bajo presupuestos de aleatoriedad y ciertas circunstancias restrictivas.

En este tipo de experimentos los éxitos buscados son expresados por unidad de área, tiempo, pieza, etc. :
- Número de defectos de una tela por m^2.

- Número de aviones que aterrizan en un aeropuerto por día, hora, minuto, etc.
- Número de bacterias por cm^2 de cultivo.
- Número de llamadas telefónicas a un conmutador por hora, minuto, etc.
- Número de llegadas de embarcaciones a un puerto por día, mes, etc.

En esta distribución el número de éxitos que ocurren por unidad de tiempo, área o producto *es totalmente al azar* y que *cada intervalo de tiempo es independiente de otro intervalo dado*, así como cada área es independiente de otra área dada y cada producto es independiente de otro producto dado.

Esta distribución se puede hacer derivar de un proceso experimental de observación en el que tengamos las siguientes características:

a. Se observa la realización de hechos de cierto tipo durante un cierto periodo de tiempo o a lo largo de un espacio de observación

b. Los hechos a observar tienen naturaleza aleatoria; pueden producirse o no de una manera no determinística.

c. La probabilidad de que se produzcan un número *x* de éxitos en un intervalo de amplitud t no depende del origen del intervalo (aunque sí de su amplitud).

d. La probabilidad de que ocurra un hecho en un intervalo infinitésimo es prácticamente proporcional a la amplitud del intervalo.

e. La probabilidad de que se produzcan dos o más hechos en un intervalo infinitésimo es un infinitésimo de orden superior a dos. En consecuencia, en un intervalo infinitésimo podrán producirse 0 o 1 hecho pero nunca más de uno

f. Si en estas circunstancias aleatorizamos de forma que la variable aleatoria *X* signifique o designe el *número de hechos que se producen en un intervalo de tiempo o de espacio*, la variable *X* se distribuye con una distribución de parámetro l.

g. El parámetro de la distribución es, en principio, el factor de proporcionalidad para la probabilidad de un hecho en un intervalo infinitésimo. Se le suele designar como *parámetro de intensidad*, aunque se corresponde con el número medio de hechos que cabe esperar que se produzcan en un intervalo unitario (media de la distribución); y que coincide con la varianza de la distribución.

h. Por otro lado es evidente que se trata de un modelo discreto y que el campo de variación de la variable será el conjunto de los números naturales, incluido el cero: $x \in [1, 2, 3, 4, \ldots]$.

A partir de las hipótesis del proceso, se obtiene una ecuación diferencial de definición del mismo que puede integrarse con facilidad para obtener la función de cuantía de la variable "número de hechos que ocurren en un intervalo unitario de tiempo o espacio

$$P = (x = k) = \frac{\mu^k e^{-\mu}}{k!}$$

μ = media o promedio de éxitos por unidad de tiempo, área o producto
e = 2.718

x = variable que nos denota el número de éxitos que se desea que ocurra

Requisitos a cubrir para determinar si se habrá de utilizar la distribución de Poisson

1. Aleatoriedad
2. Los sucesos que ocurren en el intervalo deben ser independientes