



# Probabilidad

Alexa Chávez

## Contents

<b>1</b>	<b>Introducción</b>	<b>6</b>
<b>2</b>	<b>Tipos de Probabilidad y conceptos básicos</b>	<b>7</b>
2.1	Probabilidad Clásica . . . . .	7
2.2	Probabilidad Geométrica . . . . .	7
2.3	Probabilidad Frecuencial . . . . .	7
2.4	Probabilidad Subjetiva . . . . .	8
2.5	Probabilidad Axiomática . . . . .	8
2.6	Experimento determinístico . . . . .	8
2.7	Espacio muestral . . . . .	11
<b>3</b>	<b>Combinaciones, Permutaciones y más</b>	<b>12</b>
3.1	COMBINACIONES . . . . .	12
3.2	Permutaciones . . . . .	13
3.3	Caminos posibles: Rectángulo . . . . .	15
3.4	Generalización de la Pirámide de Pascal . . . . .	17
3.5	Permutaciones Circulares . . . . .	19
3.6	Números de Catalan . . . . .	21
<b>4</b>	<b>Histogramas y graficación en Root</b>	<b>24</b>
4.1	Histograma . . . . .	24
4.2	Fórmula de Stirling . . . . .	25
4.3	Graficando con Root . . . . .	27
4.4	Srinivasa Ramanujan . . . . .	31
4.5	Dado de 3 caras . . . . .	32
<b>5</b>	<b>Tipos de infinitos: Números Álef, Explosión del Challenger, AM y FM</b>	<b>35</b>
5.1	Números Álef . . . . .	35
5.1.1	álef 0 ( $N_0$ ) . . . . .	35
5.1.2	álef 1 ( $N_1$ ) . . . . .	36
5.1.3	Más allá de álef 1 ( $N_1$ ) . . . . .	36
5.1.4	álef 2 . . . . .	37
5.1.5	álef $\omega$ . . . . .	37
5.1.6	Función álef . . . . .	38
5.2	Explosión del Challenger . . . . .	38
5.2.1	Centralización . . . . .	38
5.3	Preguntas: Examen 1 . . . . .	40



5.4	AM y FM . . . . .	41
5.4.1	Qué significa AM? . . . . .	41
5.4.2	Cómo se genera una señal multiples de estéreo FM: . . . . .	42
5.4.3	Proceso del modulador: . . . . .	43
<b>6</b>	<b>Método de mínimos cuadrados, Densidad de Probabilidad</b>	<b>45</b>
6.1	Mínimos cuadrados . . . . .	45
6.2	Densidad de Probabilidad . . . . .	46
6.3	Distribución gaussiana en ROOT . . . . .	46
6.4	Media . . . . .	48
6.4.1	Media muestral . . . . .	49
6.4.2	Media poblacional . . . . .	49
6.5	Medidas de dispersión . . . . .	49
6.5.1	Varianza . . . . .	49
6.5.2	Desviación estándar . . . . .	50
6.6	Graficando en R(Minimos cuadrados) . . . . .	50
6.6.1	Generación de Matriz . . . . .	50
6.6.2	Agregar filas y columnas . . . . .	51
6.6.3	Función "lm" . . . . .	51
6.7	Ejemplos de Mínimos cuadrados . . . . .	53
<b>7</b>	<b>Vector y valor propio, análisis de Componentes pricipales</b>	<b>55</b>
7.1	Vector propio y valor propio . . . . .	55
7.2	Análisis de componentes principales . . . . .	56
7.2.1	Matemáticas del ACP . . . . .	57
7.2.2	Método basado en correlaciones . . . . .	58
7.2.3	Método basado en las covarianzas . . . . .	58
7.3	Demostración $\overline{\Delta u} = \overline{u - \bar{u}} = \bar{u} - \bar{u} = 0$ . . . . .	60
7.4	Demostración $\overline{(\Delta u)^2} = \sum_{i=0}^M P(u_i)(u_i - \bar{u})^2 \geq 0$ . . . . .	61
7.5	Demostración $\overline{(u - \bar{u})^2} = \bar{u^2} - \bar{u}^2$ . . . . .	61
7.6	Demostración $\bar{u^2} \geq \bar{u}^2$ . . . . .	61
7.7	Multiplicidad eigenvectores y eigenvalores . . . . .	62
7.7.1	Multiplicidad Algebraica . . . . .	62
7.7.2	Matriz singular . . . . .	63
7.7.3	Traza de una matriz . . . . .	63
7.7.4	Transpuesta de una matriz . . . . .	64
7.7.5	Ejemplos de Eigenvalores y Eigenvectores . . . . .	65
7.7.6	Otros ejemplos . . . . .	66
7.7.7	Coefficientes de correlación de Pearson . . . . .	70

<b>8</b>	<b>No correlación, Aproximación de <math>\ln(x)</math>, Dato interesante: Napoleón no era bajito</b>	<b>72</b>
8.1	No correlación . . . . .	72
8.2	Aproximación de $\ln(x)$ por series de Taylor . . . . .	72
8.3	Napoleón no era bajito . . . . .	74
<b>9</b>	<b>Distribución Bose-Einsten, Maxwell-Boltzmann, Fermi Dirac</b>	<b>76</b>
9.1	Distribución Bose-Einsten . . . . .	76
9.2	Distribución Maxwell-Boltzmann . . . . .	78
9.3	Distribución Fermi Dirac, . . . . .	79
<b>10</b>	<b>Funciones de Distrubución</b>	<b>81</b>
10.1	Función de distribucion acumulada . . . . .	81
10.2	Distribución Beta . . . . .	81
10.3	Distribución hipergeométrica . . . . .	83
10.4	Distribución de Cauchy . . . . .	84
10.4.1	Función de distribución . . . . .	85
10.4.2	Aplicaciones . . . . .	85
10.5	Distribución Chi-Cuadrado . . . . .	85
10.6	La función Gamma . . . . .	86
10.6.1	Distribución Gamma . . . . .	87
10.6.2	Ejemplo . . . . .	89
10.7	Ley de Benford . . . . .	89
10.8	Distribución Exponencial . . . . .	90
10.8.1	Ejemplos . . . . .	91
10.8.2	Cálculo de variables aleatorias . . . . .	91
<b>11</b>	<b>Suma geométrica</b>	<b>93</b>
11.1	Razón común . . . . .	93
11.2	Suma . . . . .	93
11.3	Fórmula . . . . .	93
11.4	Demostración . . . . .	94
11.5	Convergencia . . . . .	94
11.6	Ejemplos . . . . .	94
11.7	Secuencia infinita . . . . .	95
<b>12</b>	<b>Anexos</b>	<b>96</b>
12.1	Problema: The last sheep . . . . .	96
12.2	Exposición: Redes Neuronales . . . . .	100

## 13 Conclusiones

120

## List of Tables

1	Costos de trabajo. . . . .	53
2	Solución . . . . .	53

## List of Figures

1	La diferencia relativa entre $(\ln x!)y(x \ln x - x)$ tiende a cero al crecer $x$ . . . . .	27
2	Resultados obtenidos mediante la simulación en ROOT (2a) dado lanzado 5000 veces, (2b) dado lanzado 10000 veces, (2c) dado lanzado 100000 veces. . . . .	34
3	Función ejemplo de densidad. La probabilidad de un intervalo, es el área que existe entre la función y el eje de abscisas. . . . .	47
4	Función de distribución Gaussiana con números aleatorios en ROOT, utilizando 10000 valores . . . . .	48
5	Curvas de densidad gamma . . . . .	88
6	Curvas de densidad estandar . . . . .	88

# 1 Introducción

¿Cuál es la probabilidad de que, en un estanque de medusas, cada una de ellas tome cierta posición tal que se repita la misma formación para tomar una foto idéntica a la de la portada?

Para responder, es necesario entender ¿qué es la probabilidad?, ¿qué tipo de probabilidades hay? y ¿Cuál debemos ocupar en ciertas situaciones?

Comenzando, la probabilidad se puede considerar como el análisis por el cual se obtiene la frecuencia de un acontecimiento determinado mediante la realización de un experimento aleatorio, del que se conocen todos los resultados posibles, bajo condiciones suficientemente estables.

Esta recopilación de apuntes, además de contener datos curiosos y notas extrañas, servirá de herramienta para comprender una pequeña parte de lo que abarca el significado de "La Probabilidad".

## 2 Tipos de Probabilidad y conceptos básicos

### 2.1 Probabilidad Clásica

La probabilidad clásica de un evento es la razón entre el número de casos favorables y el número total de casos posibles, siempre y cuando todos los elementos del espacio muestral tengan las mismas posibilidades de ocurrir.

Por lo tanto, la probabilidad clásica de un evento  $A \subset S$  se define como:

$$P(A) = \frac{|A|}{|S|} = \frac{\text{NumDeEventosFavorables}}{\text{NumDePosibilidades}}$$

Existen tres propiedades a destacar:

- $P(S) = \frac{|S|}{|S|} = 1$
- $P(S) = \frac{|A|}{|S|} \geq 0$
- Si A y B son ajenos, entonces:

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

Más generalmente, si  $A_1, A_2, \dots$  son ajenos dos a dos, entonces:

$$P(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k) \text{ cuando } A_1, A_2, \dots$$

### 2.2 Probabilidad Geométrica

La probabilidad geométrica describe la posibilidad de que un punto esté en una parte de un segmento de línea o en una parte de una región. Se define la probabilidad geométrica como:

$$P(A) = \frac{\text{Longitud}(A)}{\text{Longitud}(S)}$$

La definición también puede ser escrita en términos de área o volumen.

### 2.3 Probabilidad Frecuencial

También conocida como probabilidad empírica, este tipo de probabilidad establece que aunque el comportamiento del experimento es aleatorio, eventualmente llegaremos a una regularidad. Notamos a través de gran cantidad de observaciones acumuladas con los diversos juegos de azar una forma general de regularidad que permitió establecer una teoría. Sea un experimento

aleatorio con espacio muestral  $S$  y sea  $ACS$  un evento. Se realizan  $n$  repeticiones del experimento aleatorio y se define  $n(A)$  como el número de veces que ocurre el evento  $A$  en los  $n$  ensayos del experimento. Por lo tanto, se define la probabilidad frecuentista como:

$$P(A) = \lim_{n \rightarrow \infty} \left( \frac{n(A)}{n} \right)$$

La forma de calcular la probabilidad es usar la frecuencia relativa, ya que si se trata de un experimento aleatorio en el cual se repite muchas veces, la frecuencia relativa se acercará mucho a la probabilidad del suceso.

## 2.4 Probabilidad Subjetiva

Se refiere a la probabilidad de ocurrencia de un suceso basado en la experiencia previa, la opinión personal o la intuición del individuo. En este caso después de estudiar la información disponible, se asigna un valor de probabilidad a los sucesos basado en el grado de creencia de que el suceso pueda ocurrir. Un ejemplo muy común es el pronóstico del tiempo, muchos individuos como nosotros realizamos una predicción personal de como serán las condiciones climáticas para el día, basadas más en nuestra experiencia personal pero que muchas veces sustentamos en experiencia de eventos pasados.

## 2.5 Probabilidad Axiomática

Si hacemos un determinado experimento, que tiene un espacio muestral  $S$ , definimos la probabilidad como una función que asocia a cada suceso  $A$  una determinada probabilidad,  $P(A)$ , que cumple las siguientes propiedades:

- $P(A) \geq 0$
- $P(S) = 1$
- $P(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$  cuando  $A_1, A_2, \dots$  son ajenos dos a dos

Los 3 axiomas anteriores se les conoce como 'Axiomas de Kolmogorov'

## 2.6 Experimento determinístico

Un suceso determinista o experimento determinista es un experimento o fenómeno que da lugar a un resultado cierto o seguro, es decir, cuando partiendo de unas mismas condiciones iniciales tenemos la certeza de lo que va



a suceder. LA relación causa-efecto se conoce en su totalidad.

Son los hechos que llegan a suceder con seguridad y precisión aún antes de realizarlos, se puede saber a priori que es lo que sucederá. Por ejemplo, todos los fenómenos que siguen las leyes de la física clásica, como puede ser la caída de un cuerpo. Cuando un experimento o fenómeno no es determinista estamos ante un experimento aleatorio.

Ejemplos:

- Calentar el agua a los 100 C a nivel del mar, el agua se convierte en vapor.
- Calentar una barra de metal, ésta se dilata.
- Al dejar caer un vaso de vidrio desde el primer piso de un edificio, el vaso se rompe.
- Si un número natural es par, el siguiente será impar.
- Si metes las manos al fuego, te quemas.
- Si se avienta un objeto hacia arriba, después de un intervalos de tiempo, el objeto caerá.
- Saber qué día de la semana es mañana.
- Calcular el área de un círculo de radio 4.
- Comparar el peso de un objeto en la Tierra y el peso del mismo objeto en la Luna.
- Obtener la longitud de la línea 3 del metro de la Ciudad de México.

subsectionExperimento aleatorio

Un experimento aleatorio es aquel que bajo el mismo conjunto aparente de condiciones iniciales, puede presentar resultados diferentes, es decir, no se puede predecir o reproducir el resultado exacto de cada experiencia particular. (Ej: Lanzamiento de un dado).

Este tipo de fenómeno es opuesto al fenómeno determinista, en el que conocer todos los factores de un experimento nos hace predecir exactamente el

resultado del mismo. Por ejemplo, conociendo la altura desde la que se arroja un móvil es posible saber exactamente el tiempo que tardará en llegar al suelo en condiciones de vacío.

Ejemplos:

- Lanzar una moneda, al caer el resultado es una de sus dos caras.
- Lanzar un dado, el resultado es alguno de los números en el conjunto  $\{1,2,3,4,5,6\}$
- Extraer dos bolas de una urna con dos bolas blancas (B) y dos bolas negras (N), el resultado puede ser entonces  $\{BB, BN, NN\}$ , es decir dos bolas blancas (BB), una bola blanca y una negra (BN) o dos bolas negras (NN).
- Predecir el sexo de los hijos de las familias con 3 hijos, ordenados del mayor a menor, donde hombre es H y mujer es M, el resultado puede ser  $\{(HHH), (HHM), (HMH), (MHH), (HMM), (MHM), (MMH), (MMM)\}$
- Tomar un foco al azar de un lote de focos, el resultado es que puede o no puede estar defectuoso, el conjunto al que pertenece es defectuoso, no defectuoso
- Tomar una carta de una baraja española y obtener una de las cuatro familias {Oros, Copas, Espadas, Bastos}
- Escoger un número de la lotería y que este sea el ganador, el resultado es un número de un conjunto con una gran cantidad de números.
- Arrojar una bola de boliche para derribar los pinos, el resultado es la cantidad de pinos derribados que pueden ser  $\{0,1,2,3,4,5,6,7,8,9,10\}$
- El tiempo, con respecto a la lluvia, que hará durante 3 días consecutivos, es decir un día lluvioso se llamará L, y un día no lluvioso será N, por lo tanto el resultado es alguna combinación del conjunto  $\{(LLL), (LLN), (LNL), (NLL), (LNN), (NLN), (NNL), (NNN)\}$
- Un foco manufacturado en una planta es expuesta a una prueba de vida y el tiempo de duración de un foco es registrado. En este caso no se conoce cuál será el tiempo de duración del foco seleccionado, pero se puede saber previamente que será un valor entre 0 e infinito.

## 2.7 Espacio muestral

El espacio muestral o espacio de muestreo (denotado  $E$ ,  $S$ ,  $\Omega$  o  $U$ ) consiste en el conjunto de todos los posibles resultados individuales de un experimento aleatorio.

Por ejemplo, si el experimento consiste en lanzar dos monedas, el espacio de muestreo es el conjunto  $\{(cara, cara), (cara, cruz), (cruz, cara) \text{ y } (cruz, cruz)\}$ . Un evento o suceso es cualquier subconjunto del espacio muestral con estructura de  $\sigma$ -álgebra,<sup>1</sup> llamándose a los sucesos que contengan un único elemento sucesos elementales. En el ejemplo, el suceso "sacar cara en el primer lanzamiento", o  $\{(cara, cara), (cara, cruz)\}$ , estaría formado por los sucesos elementales  $\{(cara, cara)\}$  y  $\{(cara, cruz)\}$ .

Para algunos tipos de experimento puede haber dos o más espacios de muestreo posibles. Por ejemplo, cuando se toma una carta de un mazo normal de 52 cartas, una posibilidad del espacio de muestreo podría ser el número (del as al rey), mientras que otra posibilidad sería el palo (diamantes, tréboles, corazones y picas). Una descripción completa de los resultados, sin embargo, especificaría ambos valores, número y palo, y se podría construir un espacio de muestreo que describiese cada carta individual como el producto cartesiano de los dos espacios de muestreo descritos.

Los espacios de muestreo aparecen de forma natural en una aproximación elemental a la probabilidad, pero son también importantes en espacios de probabilidad. Un espacio de probabilidad  $(\Omega, F, P)$  incorpora un espacio de muestreo de resultados,  $\Omega$ , pero define un conjunto de sucesos de interés, la  $\sigma$ -álgebra  $F$ , por la cuál se define la medida de probabilidad  $P$ .

Existen dos tipos de espacios muestrales:

- Espacios muestrales discretos: cuando la cantidad de elementos es finito o contable.
- Espacios muestrales continuos: donde la cantidad de resultados obtenidos pueden ser infinitos.

[1]

### 3 Combinaciones, Permutaciones y más

#### 3.1 COMBINACIONES

Se llama combinaciones de  $m$  elementos tomados de  $n$  en  $n$  ( $m \geq n$ ) a todas las agrupaciones posibles que pueden hacerse con los  $m$  elementos de forma que: No entran todos los elementos, no importa el orden, no se repiten los elementos.

Ejemplos:

1. Si se seleccionan cinco cartas de un grupo de nueve, ¿cuántas combinaciones de cinco cartas habría?

$$C = \binom{9}{5} = \frac{9!}{(5!)(9-5)!} = 126$$

2. José tiene 9 amigos y desea invitarlos a cenar, pero sólo puede invitar a 6 simultáneamente. ¿Cuántos grupos distintos de invitados puede tener?

Queremos saber cuantos grupos distintos podemos formar independientemente del orden en que se elija los invitados.

$$C = \binom{9}{6} = \frac{9!}{(6!)(9-6)!} = 84 \text{ grupos distintos}$$

3. El juego de la Primitiva consiste en acertar 6 números naturales a elegir entre el 1 y el 49. ¿Cuántas posibles combinaciones hay? Si cada combinación nos cuesta 1£ ¿Cuánto nos tendremos que gastar para asegurar que vamos a acertar seguro los 6 números?
4. Entre los 8 candidatos para 2 vacantes de personal, se encuentran 4 hombres y 4 mujeres, ¿de cuántas formas se puede cubrir esta vacante con 2 candidatos cualesquiera de los 8?

$$C = \binom{8}{2} = \frac{8!}{(2!)(8-2)!} = 28$$

Queremos acertar 6 números de 49 posibles, independientemente del orden en que los elijamos.

$$C = \binom{49}{6} = \frac{49!}{(6!)(49-6)!} = 13,983,816£$$

5. ¿De cuántas formas una persona puede tomar 3 libros de una lista de 8 libros?



$$\text{No. de formas} = \binom{8}{3} = \frac{8!}{(3!)(8-3)!} = 56$$

6. ¿De cuántas maneras diferentes se puede formar un comité de 5, de entre un grupo de 62 personas.

$$\text{No. de formas} = \binom{62}{5} = \frac{62!}{(5!)(62-5)!} = 6,471,002$$

7. Entre los 10 deportistas, se encuentran 5 hombres y 5 mujeres, ¿de cuántas formas se puede seleccionar 2 deportistas, donde 1 sea hombre y 1 sea mujer?

$$C = \binom{5}{1} \binom{5}{1} = 5(5) = 25$$

8. ¿Cuántas parejas distintas pueden formarse con cinco individuos?

$$C = \binom{5}{2} = \frac{5!}{(2!)(5-2)!} = 10$$

9. En una carrera compiten 10 corredores y se clasifican los 3 primeros para la fase siguiente. ¿de cuántas maneras diferentes puede producirse la clasificación?

$$\text{No. de formas} = \binom{10}{3} = \frac{10!}{(3!)(10-3)!} = 120 \text{ formas diferentes}$$

10. Un paquete de 10 baterías tiene dos piezas defectuosas, ¿de cuántas maneras se pueden tomar 3 de estas baterías y sacar una de las baterías defectuosas?

$$\text{No. de formas} = \binom{8}{2} \binom{2}{1} = \left(\frac{8!}{(2!)(8-2)!}\right) \left(\frac{2!}{(1!)(2-1)!}\right) = 28(2) = 56$$

### 3.2 Permutaciones

Las permutaciones de  $n$  elementos son las distintas formas en que pueden ordenarse dichos  $n$  elementos.

Ejemplos:

1. Un vendedor quiere visitar 5 ciudades. Si no quiere repetir ciudades, ¿cuántas rutas distintas puede elaborar si puede empezar y acabar en cualquiera de las ciudades? El vendedor puede elegir la primera ciudad que visitará de entre las 5. Elegirá la segunda ciudad que visitará de entre las 4 restantes. Para la tercera ciudad tiene 3 opciones. Para la cuarta, 2. Y para la última.

Así que puede elaborar  $5 * 4 * 3 * 2 * 1 = 5! = 120$  rutas distintas.

2. ¿Cuántos números de tres cifras distintas se pueden formar con los dígitos 5, 6 y 7? El conjunto puede ordenarse de diferentes formas, dando lugar a varias permutaciones:  $\{5,6,7\}$ ,  $\{5,7,6\}$ ,  $\{7,5,6\}$ ,  $\{7,6,5\}$ ,  $\{6,5,7\}$ ,  $\{6,7,5\}$  y  $\{5,6,7\}$ . Calculando de otra manera

$$\text{Hay } P_3 = 3! = 6 \text{ números}$$

3. ¿De cuántas maneras distintas se pueden asignar a 10 profesores las 10 unidades de un curso de economía?

$$P_{10} = 10! = 3,628,800 \text{ maneras distintas}$$

4. ¿De cuántas maneras distintas se pueden colocar 8 libros en un librero con 8 espacios?

$$P_8 = 8! = 40,320 \text{ maneras distintas}$$

5. En el palo de señales de un barco se pueden izar tres banderas rojas, dos azules y cuatro verdes. ¿Cuántas señales distintas pueden indicarse con la colocación de las nueve banderas? Si se repiten los elementos y si importa el orden.

$$PR_9^{2,3,4} = \frac{9!}{2!3!4!} = 1260$$

6. Una mesa presidencial está formada por ocho personas, ¿de cuántas formas distintas se pueden sentar, si el presidente y el secretario siempre van juntos?

Se forman dos grupos: el primero de 2 personas y el segundo de 7 personas, en los dos se cumple que: Si importa el orden no se repiten los elementos.

$$P_2 P_7 = (2!)(7!) = 10,080$$

7. Se ordenan en una fila 5 bolas rojas, 2 bolas blancas y 3 bolas azules. Si las bolas de igual color no se distinguen entre sí, ¿de cuántas formas posibles pueden ordenarse?

$$P_{10}^{5,2,3} = \frac{10!}{(5!)(2!)(3!)} = 2520$$

8. ¿De cuántas formas distintas pueden sentarse ocho personas alrededor de una mesa redonda?

$$PC_8 = P_8 = (8 - 1)! = 7! = 5040$$

9. Doce estudiantes van a ir a Veracruz en tres carros, 3 estudiantes en un carro, 4 estudiantes en el segundo carro, y 5 en el tercer carro. ¿De cuántas formas se pueden acomodar, si cualquiera puede conducir?

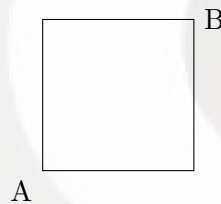
$$P_{3,4,5} = \frac{12!}{(3!)(4!)(5!)} = 27,720$$

10. Con las letras de la palabra *libro*, ¿cuántas ordenaciones distintas se pueden hacer que empiecen por vocal? La palabra empieza por i u o seguida de las 4 letras restantes tomadas de 4 en 4. Si importa el orden. no se repiten los elementos.

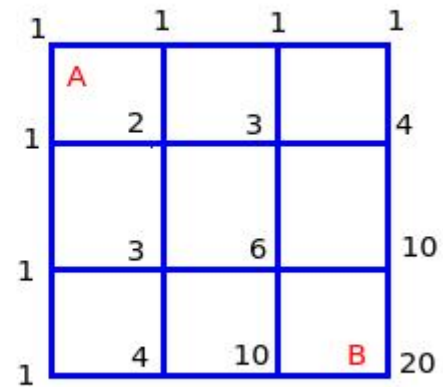
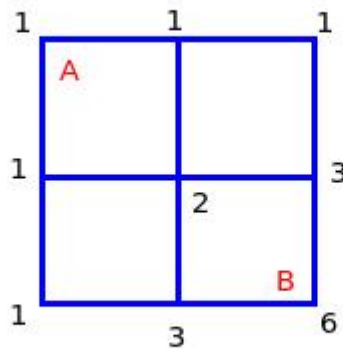
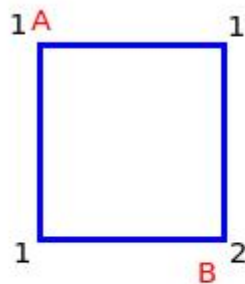
*libro* tiene 2 vocales, cualquiera de las 2 puede estar al principio y las 4 letras restantes despues a esta. Por lo tanto:

$$P_2P_4 = 2(4!) = 2(24) = 48$$

### 3.3 Caminos posibles: Rectángulo



Dado un cuadrado, se desea calcular la cantidad de caminos posibles del vértice A al vértice B. Procedemos a analizar los casos básicos particulares.



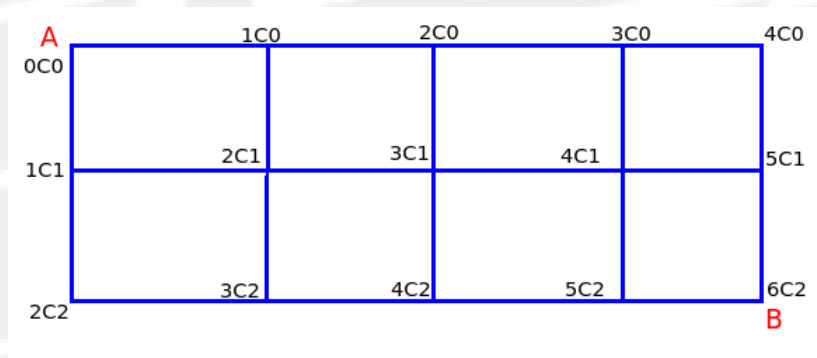
Se puede observar que los v rtices de los cuadros est n relacionados a los valores que contiene los elementos del tri ngulo de Pascal, los cuales son el resultado de combinar la l nea  $n$  con la columna  $p$ . En la siguiente figura se muestra la relaci n entre los valores y las combinaciones del tri ngulo de Pascal.

	p=0	p=1	p=2	p=3	p=4	p=5	p=6	p=7	p=8	p=9
n=0	1									
n=1	1	1								
n=2	1	2	1							
n=3	1	3	3	1						
n=4	1	4	6	4	1					
n=5	1	5	10	10	5	1				
n=6	1	6	15	20	15	6	1			
n=7	1	7	21	35	35	21	7	1		
n=8	1	8	28	56	70	56	28	8	1	
n=9	1	9	36	84	126	126	84	36	9	1

	p=0	p=1	p=2	p=3	p=4	p=5	p=6	p=7	p=8	p=9
n=0	0c0									
n=1	1c0	1c1								
n=2	2c0	2c1	2c2							
n=3	3c0	3c1	3c2	3c3						
n=4	4c0	4c1	4c2	4c3	4c4					
n=5	5c0	5c1	5c2	5c3	5c4	5c5				
n=6	6c0	6c1	6c2	6c3	6c4	6c5	6c6			
n=7	7c0	7c1	7c2	7c3	7c4	7c5	7c6	7c7		
n=8	8c0	8c1	8c2	8c3	8c4	8c5	8c6	8c7	8c8	
n=9	9c0	9c1	9c2	9c3	9c4	9c5	9c6	9c7	9c8	9c9



EL valor que obtenemos en el vertice B es la cantidad de caminos posibles para recorrer de A a B. Dado que los movimientos en un rectángulo son similares a los movimientos que se siguen en el caso del rectángulo, podemos determinar una figura similar, rotulando cada vertice con los valores del triángulo de Pascal.



El total de caminos posibles de A hasta B, de un rectángulo de longitud  $(4+2)$ , es  $6C_2 = (4+2)$  donde  $K=6$  y  $n=2$ . Así el cálculo para el número total de caminos es:

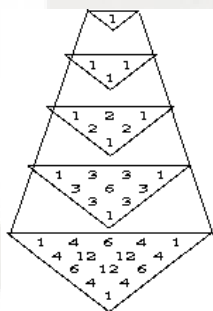
$$\frac{[k-n]!+n!}{[k-n]!n!}$$

### 3.4 Generalización de la Pirámide de Pascal

En vez de considerar las potencias de  $a+b$ , se puede considerar las del trinomio  $a+b+c$ . De esta manera,  $(a+b+c)^n$  es una suma de monomios de la forma  $L_{p,q,r}(a^p)(b^q)(c^r)$  con  $p, q$  y  $r$  positivos,  $p+q+r = n$  y  $L_{p,q,r}$  un número natural que se llama coeficiente trinomial. Los calculos son similares a los del coeficiente binomial, y se dan mediante la siguiente expresión:

$$L_{p,q,r} = \binom{n}{(p,q,r)} = \frac{n!}{p!q!r!}$$

Estos coeficientes se pueden considerar como la analogía tridimensional del triángulo de Pascal. De hecho, a la distribución de estos coeficientes al estilo piramidal se le conoce como *pirámide de Pascal* o *tetraedro de Pascal*. También es infinita, con secciones triangulares y el valor de cada casilla es la suma de los valores de las tres casillas encima de ella.



[3] [4] [2] [5]

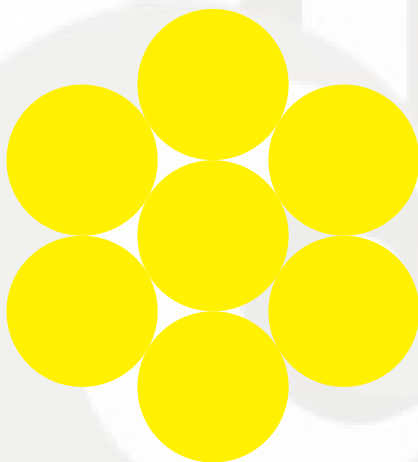
### 3.5 Permutaciones Circulares

Son un caso especial de permutaciones, se utiliza cuando los elementos se han de ordenar en círculo, por ejemplo los comensales de una mesa, de modo que el primer elemento que se sitúe marcará el principio y el final de la muestra. Ejemplos

1. Calcular las permutaciones circulares de 7 elementos

$$P_c^7 = (7 - 1)! = 6! = 720$$

2. De de cuántos modos distintos podemos ubicar las cifras del 1 al 7 en la figura siguiente?



Podemos solucionar este problema como la conjunción de dos sucesos: En primer lugar ubicamos una cifra en el centro (7 posibilidades) y en segundo lugar las otras 6 cifras, las cuales por ordenarse en una circunferencia podrán permutarse de  $(6 - 1)!$  maneras, por lo cual:

$$No.maneras = 7 \times 5! = 7 \times 120 = 840$$

3. De cuántos modos diferentes puede sentarse alrededor de una mesa circular una madre y sus 5 hijos?

$$P_c^6 = (6 - 1)! = 5! = 120$$

4. Si siete personas se reúnen para cenar, Cuántas maneras hay de que se sienten a la mesa si dos insisten en sentarse juntas?

Formamos un bloque con los dos que se quieren sentar juntos, considerándolo como un solo elemento y calculamos el número de permutaciones  $P_C^6$ , internamente en el bloque calculamos el número de permutaciones entre los dos que se sientan juntos mediante  ${}_2P_2$ . Aplicamos el principio multiplicativo.

$${}_2P_2(P_C^6) = 2!(6 - 1)! = 240$$

5. El señor García y su esposa tienen 6 invitados a cenar en su casa, De cuántas formas diferentes podrán sentarse a la mesa los 8 comensales (los 2 anfitriones y los 6 invitados)?

habrá  $(8 - 1)! = 7! = 5,040$  formas en que puedan sentarse a cenar.

6. Juan, Pedro, Jesús y Alberto se reúnen a jugar dominó, De cuántas formas pueden sentarse a la mesa de juego?

$P_C^4 = (4 - 1)! = 3! = 6$  formas diferentes en que puedan sentarse a jugar

7. De cuántas maneras diferentes pueden disponerse circularmente las letras de la A a la Z?

$$P_C^{27} = (27 - 1)! = 26! = 4.03129 \times 10^{26}$$

8. La directiva de deportes consta de 10 integrantes, en cada reunión se sientan en forma circular, si tres de los integrantes siempre conservan el lugar, de cuántas maneras se pueden sentar?

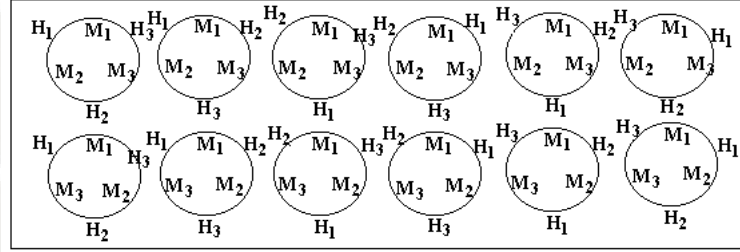
Formamos un bloque con los tres integrantes que siempre conservan su lugar, considerándolo como un solo elemento y calculamos el número de permutaciones  $P_C^6$ , internamente en el bloque calculamos el número de permutaciones entre los dos que se sientan juntos mediante  $P_C^2$ . Aplicamos el principio multiplicativo.

$${}_3P_3(P_C^7) = 3!(7 - 1)! = 4320$$



9. De cuántas formas se pueden sentar 3 parejas de casados alrededor de una mesa circular, si no debe haber dos mujeres juntas ni dos hombres juntos?

$$2!3! = 2(6) = 12$$



El número de formas en que podemos sentar a los 3 mujeres alrededor de una mesa circular, dejando un lugar en medio es  $2!$ . Obsérvese que el primer renglón de círculos, los seis arreglos diferentes tienen a M1 M2 M3 siempre en la misma posición; y en el segundo renglón, los seis arreglos tienen a M1 M3 M2 siempre en la misma posición; por ello son sólo dos arreglos de las tres mujeres, dejando un lugar en medio. Hay  $3! = 6$  formas de sentar a los tres hombres por cada uno de los dos arreglos de mujeres; quedando así en forma alternada.

10. De cuántas formas distintas pueden sentarse ocho personas alrededor de una mesa redonda?

$$P_c^8 = (8 - 1)! = 5040$$

### 3.6 Números de Catalan

Los números de Catalan forman una secuencia de números naturales que aparece en varios problemas de conteo que habitualmente son recursivos. Obtienen su nombre del matemático belga Eugene Charles Catalan (1814-1894). El  $n$ -ésimo número de Catalan se obtiene, aplicando coeficientes binomiales, a partir de la siguiente fórmula:

$$C_n = \frac{1}{n+1} \binom{2n}{n} = \frac{(2n)!}{(n+1)!n!} \quad \text{con } n \geq 0$$

Una expresión alternativa para  $C_n$  es

$$C_n = \binom{2n}{n} - \binom{2n}{n-1} = \frac{(2n)!}{(n+1)!n!} \quad \text{con } n \geq 1$$

Los números de catalan satisfacen la relación de recurrencia siguiente:

$$C_0 = 1 \text{ y } C_{n+1} = \sum_{i=0}^n C_i C_{n-i} \text{ con } n \geq 0$$

$$C_0 = 1 \text{ y } C_{n+1} = \frac{2(2n+1)}{n+2} C_n \text{ con } n \geq 0$$

Existen múltiples problemas de combinatoria cuya solución la dan los números de Catalan. Una aplicación de esta fórmula es la división de polígono es triángulos.

Un polígono convexo es un polígono que cumple que todos sus ángulos interiores miden menos de 180 grados. Según esta definición es evidente que todos los polígonos regulares como el triángulo equilátero, cuadrado, pentágono regular.etc) son convexos.

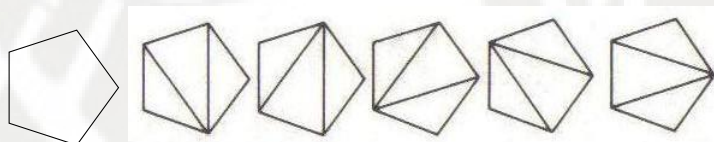
$C_n$  es el nmero de formas distintas de cortar un polígono convexo de  $n + 2$  lados en triángulos conectando vértices con líneas rectas sin que ninguna se corte.

Una triangulación de un polígono es una manera de descomponerlo como una unión disjunta de triángulos, cuyos vértices coinciden con los del polígono. Tomemos el primer polígono regular en lo que a número de lados se refiere, el triángulo equilátero. Está claro que no se puede trazar ninguna diagonal. Esto es, el nmero de formas en las que podemos dividir un tringulo equiltero en triángulos trazando diagonales de la forma descrita antes es 1.


Pasamos al siguiente, el cuadrado. En él podemos trazar dos diagonales que lo dividen en triángulos:



El siguiente es el pentágono. En este caso cada forma de dividirlo en triángulos así consiste en trazar dos diagonales que no se corten. Estas son las 5 formas.



Siguiendo con el mismo procedimiento, el hexágono el número de diagonales a trazar es tres por vez. Dando como resultado 14 formas de dividir el



hexágono, con un heptágono obtendremos 42 formas, con un octágono 132, y así sucesivamente. Al final se ha obtenido una sucesión de números como la siguiente:

1, 2, 5, 14, 42, 132...

## 4 Histogramas y graficación en Root

Cuando se habla de estadística, se suele pensar en un conjunto de datos numéricos presentada de forma ordenada y sistemática. Esta idea es debida a la influencia del entorno, ya que hoy día es casi imposible que cualquier medio de comunicación, periódico, radio, televisión, etc, no nos aborde diariamente con cualquier tipo de información estadística.

Sólo cuando alguien se adentra en un mundo más específico como es el campo de la investigación de las Ciencias Sociales: Medicina, Biología, Psicología,... se empieza a percibir que la Estadística no sólo es algo más, sino que se convierte en la única herramienta que, hoy por hoy, permite dar luz y obtener resultados, y por tanto beneficios, en cualquier tipo de estudio, cuyos movimientos y relaciones, por su variabilidad intrínseca, no puedan ser abordadas desde la perspectiva de las leyes deterministas.

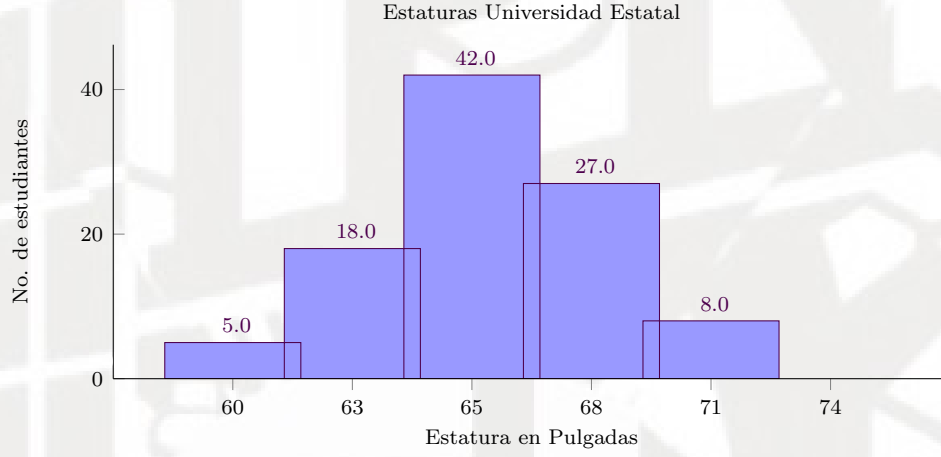
### 4.1 Histograma

El histograma es aquella representación gráfica de estadísticas de diferentes tipos. La utilidad del histograma tiene que ver con la posibilidad de establecer de manera visual, ordenada y fácilmente comprensible todos los datos numéricos estadísticos que pueden tornarse difíciles de entender. Hay muchos tipos de histogramas y cada uno se ajusta a diferentes necesidades como también a diferentes tipos de información.

Los histogramas son utilizados siempre por la ciencia estadística. Su función es exponer gráficamente números, variables y cifras de modo que los resultados se visualicen más clara y ordenadamente. El histograma es siempre una representación en barras y por eso es importante no confundirlo con otro tipo de gráficos como las tortas. Se estima que por el tipo de información brindada y por la manera en que ésta es dispuesta, los histogramas son de especial utilidad y eficacia para las ciencias sociales ya que permiten comparar datos sociales como los resultados de un censo, la cantidad de mujeres y/o hombres en una comunidad, el nivel de analfabetismo o mortandad infantil, etc. Ejemplos:

1. La cantidad de personas que hay de cierto rango de de edad.
2. Las calificaciones de algunos alumnos, y su calificación.
3. Las estaturas en pulgadas de una Universidad estatal se pueden mostrar como histogrma.





## 4.2 Fórmula de Stirling

En matemáticas, la fórmula de Stirling es una aproximación para factoriales grandes. Lleva el nombre en honor al matemático escocés del siglo XVIII James Stirling.

En muchas aplicaciones, particularmente en estadística y teoría de la probabilidad, encontramos necesario tener una aproximación simple a  $n!$  como una función elemental de  $n$ . Tal expresión es dada por el siguiente teorema conocido como fórmula o aproximación de Stirling:

$$\lim_{n \rightarrow \infty} \frac{n!}{(\sqrt{2n\pi}) \left(\frac{n}{e}\right)^n} = 1 \quad (1)$$

Que se escribe frecuentemente como:

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \quad (2)$$

Se puede mejorar la fórmula con un desarrollo asintótico proveniente de la función gamma:

$$\Gamma(n+1) = n! \int_0^\infty t^n e^{-t} dt \quad (3)$$

De la ecuación anterior tomamos el argumento de la integral y aplicamos logaritmo:

$$\log(t^n e^{-t}) = n \log t - t \quad (4)$$

Definimos la variable  $t = (n + \epsilon)$  y sustituimos en la Eq(4).

$$n \log t - t = n \log(n + \epsilon) - (n + \epsilon) \quad (5)$$

Tomamos el termino  $\log(n + \epsilon)$  de la Eq(5) y factorizamos  $n$

$$\log(n + \epsilon) = \log n \left(1 + \frac{\epsilon}{n}\right) = \log n + \log\left(1 + \frac{\epsilon}{n}\right) \quad (6)$$

Para una  $n$  muy grande, podemos asegurar  $\frac{\epsilon}{n} < 1$  y usando los principios de la serie de Taylor para logaritmo tenemos que:

$$\log\left(1 + \frac{\epsilon}{n}\right) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \frac{\epsilon^k}{n^k} \quad (7)$$

Sustituimos el resultado anterior en la Eq(5) y de esta manera obtenemos:

$$n \log(n + \epsilon) - (n + \epsilon) = n \left( \log n + \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \frac{\epsilon^k}{n^k} \right) - n - \epsilon \quad (8)$$

De la ecuación anterior tomamos el segundo miembro de la igualdad y simplificamos.

$$n \log n - n + \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \frac{\epsilon^k}{n^{k-1}} - \epsilon \quad (9)$$

Se puede observar que cuando  $k = 1$  la función se indetermina, por lo que  $k$  inicia con un valor de 2.

$$n \log n - n + \sum_{k=2}^{\infty} \frac{(-1)^{k+1}}{k} \frac{\epsilon^k}{n^{k-1}} - \epsilon \quad (10)$$

Posteriormente se representa el desarrollo de la serie para tener una idea de la función que se va adquiriendo.

$$n \log n - n - \frac{\epsilon^2}{2n} + \frac{\epsilon^3}{3n^2} - \frac{\epsilon^4}{4n^3} \pm \dots \quad (11)$$

La ecuación anterior se obtuvo a partir de  $n \log t - t$ , por lo que sustituyendo en la Eq(4). Obtenemos la siguiente aproximación. Nota: se toma solo la resta del primer termino fraccionario.

$$\log(t^n e^{-t}) \approx n \log n - n - \frac{\epsilon^2}{2n} \quad (12)$$

Se simplifica la expresión al aplicar exponencial a ambos lados de la ecuación.

$$t^n e^{-t} \approx \frac{n^n}{e^n} e^{-\frac{\epsilon^2}{2n}} \quad (13)$$

Si se integra a ambos lados de la igualdad, obtenemos la siguiente expresión.

Se puede observar que llegamos a la equivalencia de  $n!$  de la Eq(3)

$$n! = \int_0^\infty t^n e^{-t} dt \approx \int_{-n}^\infty \frac{n^n}{e^n} e^{-\frac{\epsilon^2}{2n}} d\epsilon \quad (14)$$

Por definición de la operación gamma para  $1/2$  tenemos:

$$\sqrt{\frac{\pi}{p}} = \int_{-\infty}^\infty e^{-px^2} dx \quad (15)$$

La eq(14) es del mismo tipo que la eq(15), y tomando en cuenta que para una  $n$  muy grande  $-n = -\infty$ . Resolviendo la integral obtenemos la fórmula de Stirling

$$n! \approx \frac{n^n}{e^n} \sqrt{2n\pi} \quad (16)$$

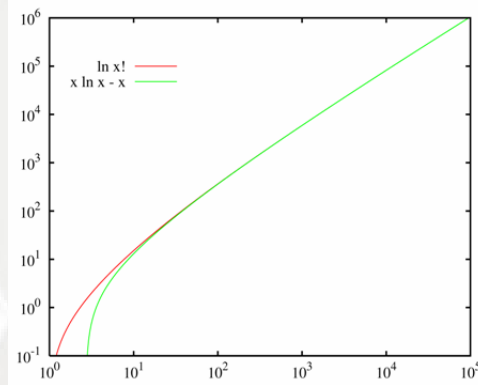


Figure 1: La diferencia relativa entre  $(\ln x!)y(x \ln x - x)$  tiende a cero al crecer  $x$ .

### 4.3 Graficando con Root

Aunque es un poco confuso, el intento se hizo, se logró capturar las siguientes gráficas:

$\ln(x!)$



```
void grafica1() {
    TCanvas *c1 = new TCanvas("c1","ln(x!)",200,10,700,500);

    c1->SetFillColor(30); //Cambia color al fondo
    c1->SetGrid();

    const Int_t n = 150;
    Double_t x[n], y[n],m[n];
    for (Int_t i=1;i<n;i++)
    {
        m[i] = i;
        x[i] = m[i]*x[i-1];
        x[0]=1;
        x[1]=1;
        y[i] = log(x[i]);
    }

    TGraph *gr = new TGraph(n,m,y);
    gr->SetLineColor(2);
    gr->SetLineWidth(3);
    gr->SetMarkerColor(4);
    gr->SetMarkerStyle(1);
    gr->SetTitle("ln(x!)");
    gr->GetXaxis()->SetTitle("EJE X");
}
```

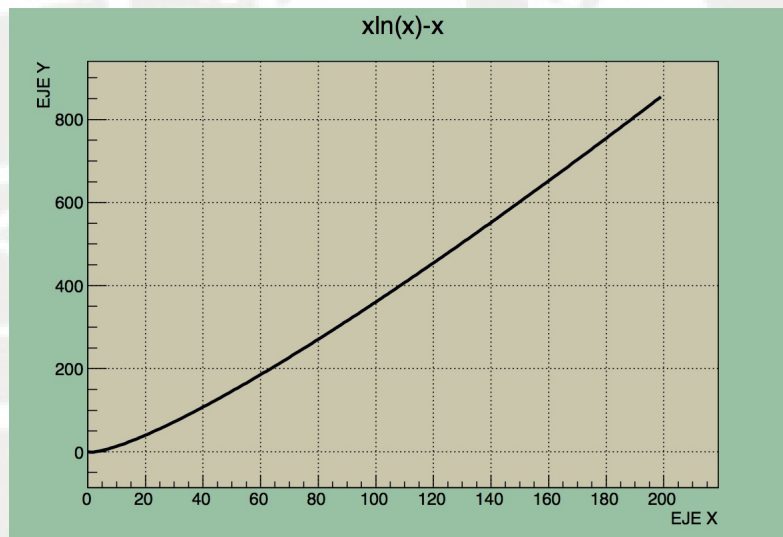
```

gr->GetYaxis()->SetTitle("EJE Y");
gr->Draw();

c1->Update();
c1->GetFrame()->SetFillColor(21);
c1->GetFrame()->SetBorderSize(12);
c1->Modified();
}

```

**xln(x)-x**



```

void grafica2() {

    TCanvas *c1 = new TCanvas("c1","xln(x)-x",200,10,700,500);

    c1->SetFillColor(30); //Cambia color al fondo
    c1->SetGrid();

    const Int_t n = 200;
    Double_t x[n], y[n],m[n];
    for (Int_t i=1;i<n;i++)
    {
        x[i] = i;

```



```

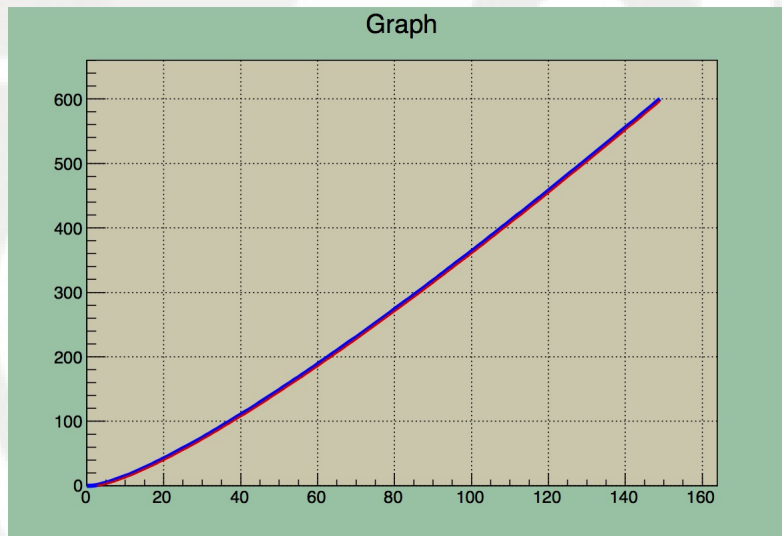
        y[i] = (x[i]*log(x[i]))-x[i];
    }

    TGraph *gr = new TGraph(n,x,y);
    gr->SetLineColor(1);
    gr->SetLineWidth(3);
    gr->SetMarkerColor(4);
    gr->SetMarkerStyle(1);
    gr->SetTitle("xln(x)-x");
    gr->GetXaxis()->SetTitle("EJE X");
    gr->GetYaxis()->SetTitle("EJE Y");
    gr->Draw();

    c1->Update();
    c1->GetFrame()->SetFillColor(21);
    c1->GetFrame()->SetBorderSize(12);
    c1->Modified();
}

```

**Comparación  $\ln(x!)$  con  $x\ln(x)-x$**



```

void grafica3() {

    TCanvas *c1 = new TCanvas("c1","ln(x!) y xln(x)-x",200,10,700,500);

```

```

c1->SetFillColor(30); //Cambia color al fondo
c1->SetGrid();

const Int_t n = 150;
Double_t x[n], y[n], m[n], x2[n], y2[n];
for (Int_t i=1; i<n; i++)
{
    m[i] = i;
    x[i] = m[i]*x[i-1];
    x[0]=1;
    x[1]=1;
    y[i] = log(x[i]);          //Primera funcion

    x2[i] = i;
    y2[i] = (x2[i]*log(x2[i]))-x2[i]; //Segunda funcion
}
TGraph *gr1 = new TGraph(n,m,y); //Grafica primera funcion
gr1->SetLineWidth(4);
gr1->SetLineColor(4);
gr1->Draw("AC");

TGraph *gr2 = new TGraph(n,x2,y2); //Grafica segunda funcion
gr2->SetLineWidth(2);
gr2->SetMarkerStyle(1);
gr2->SetLineColor(2);
gr2->Draw("CP");

c1->Update();
c1->GetFrame()->SetFillColor(21);
c1->GetFrame()->SetBorderSize(12);
c1->Modified();
}

```

#### 4.4 Srinivasa Ramanujan

A los 12 años dominaba la trigonometría, y a los 15 le prestaron un libro con 6.000 teoremas conocidos, sin demostraciones. ésa fue su formación matemática básica. En 1903 y 1907 no aprobó los exámenes universitarios

porque sólo se dedicaba a sus diversiones matemáticas.

En 1912 fue animado a comunicar sus resultados a tres distinguidos matemáticos. Dos de ellos no le respondieron, pero sí lo hizo Godfrey Harold Hardy, de Cambridge. Hardy estuvo a punto de tirar la carta, pero la misma noche que la recibió se sentó con su amigo John Edensor Littlewood a descifrar la lista de 120 fórmulas y teoremas de Ramanujan. Horas más tarde creían estar ante la obra de un genio. Hardy tenía su propia escala de valoración para el genio matemático: 100 para Ramanujan, 80 para David Hilbert, 30 para Littlewood y 25 para sí mismo. Algunas de las fórmulas de Ramanujan le desbordaron, pero escribió ...forzoso es que fueran verdaderas, porque de no serlo, nadie habría tenido la imaginación necesaria para inventarlas. Invitado por Hardy, Ramanujan partió para Inglaterra en 1914 y comenzaron a trabajar juntos. En 1917 Ramanujan fue admitido en la Royal Society de Londres y en el Trinity College, siendo el primer indio que lograba tal honor. De salud muy débil, murió tres años después.

### **Número de Ramanujan**

Se denomina número de Hardy-Ramanujan a todo entero natural que se puede expresar como la suma de dos cubos de dos maneras diferentes. Hardy comenta la siguiente anécdota :

Recuerdo que fui a verle una vez, cuando él ya estaba muy enfermo, en Putney. Había tomado yo un taxi que llevaba el número 1729 y señalé que tal número me parecía poco interesante, y yo esperaba que él no hiciera sino un signo desdenoso. - "No"- me respondió- este es un número muy interesante; es el número más pequeño que podemos descomponer de dos maneras diferentes como suma de dos cubos.

## **4.5 Dado de 3 caras**

Se lanza un dado de 3 caras 5000, 10000 y 100000, los resultados obtenidos se grafican en un histograma con la frecuencia que cae cada valor que contiene el dado.

La simulación se realizó en ROOT mediante la obtención de valores aleatorios del 1 al 3, graficando la frecuencia con la que se obtiene cada valor. El código implementado se muestra a continuación

```
#include<iostream>
#include<string>
#include<fstream>
```

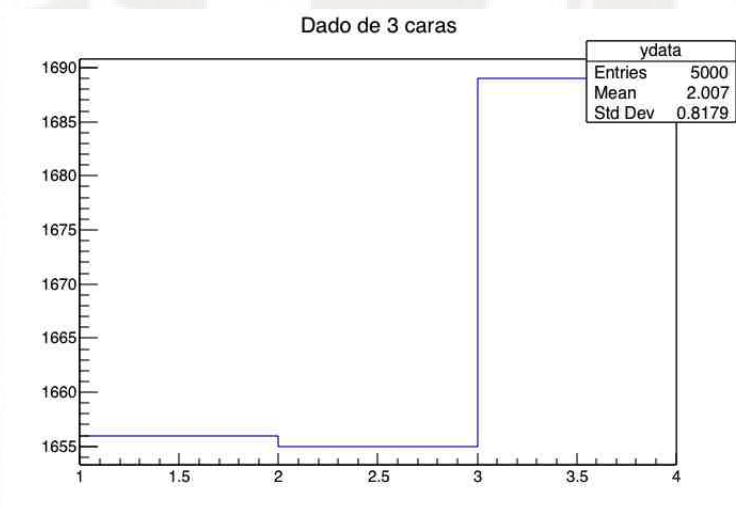
```
using namespace std;

void dados()
{
#include <stdlib.h>
#include <stdio.h>
    int i;
    int y;

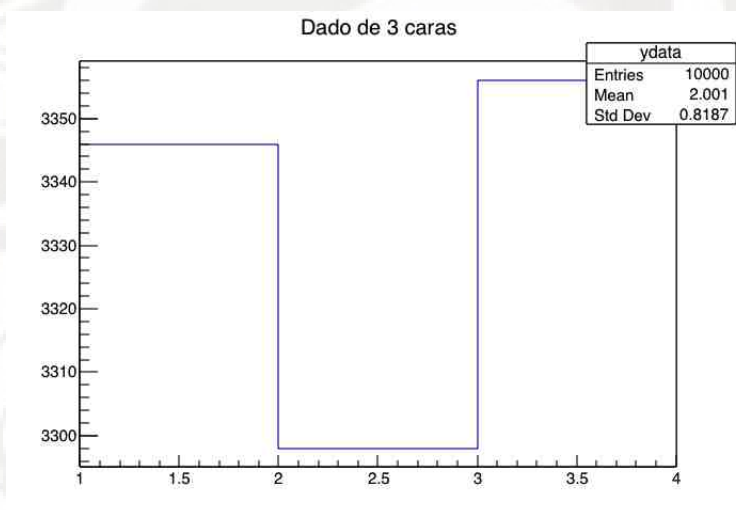
    TH1F* ydata = new TH1F("ydata","Dado_de_3_caras",3,1,4);

    for(i=0; i<10000; i++)
    {
        y=1+rand()%3;

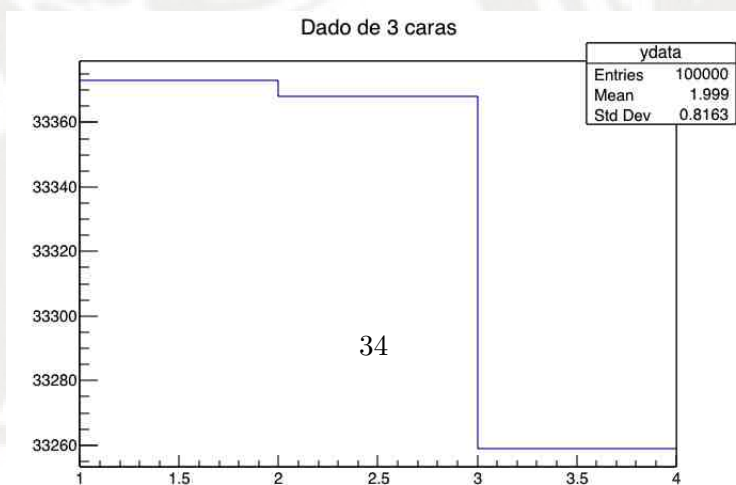
        ydata->Fill(y);
        ydata->Draw();
    }
}
```



(a) 0,0



(b) 0,1



(c) 1,0

Figure 2: Resultados obtenidos mediante la simulación en ROOT (2a) dado

## 5 Tipos de infinitos: Números Álef, Explosión del Challenger, AM y FM

### 5.1 Números Álef

En teoría de conjuntos, álef  $N$ , es un signo empleado para referirse a ciertos números transfinitos que de hecho resultan ser números ordinales iniciales y por tanto números cardinales.

En análisis matemático aparecen frecuentemente álef 0 y álef 1, aunque pueden definirse números transfinitos arbitrariamente grandes más allá de estos dos. El cardinal álef 0 representa la cantidad de elementos de un conjunto infinito del como el de los números naturales, de hecho este cardinal es el número transfinito más pequeño. Georg Cantor, que inauguró la teoría de conjuntos, demostró que existían diferentes tipos de infinitos inconmensurables entre sí, y por tanto no todos los conjuntos infinitos eran equipotentes. De hecho, Cantor demostró que el conjunto de los números reales tenía "más elementos" que los números enteros (si bien ninguno de los dos conjuntos es finito, ambos diferían en su grado de "infinitud"). El número de elementos de la recta real se representó como álef 1.

Puede probarse rigurosamente que dada la clase formada por todos los números ordinales, existe un único isomorfismo (de orden) entre esta clase y la clase de los cardinales transfinitos. Este isomorfismo, denotado como  $N$ , se emplea en teoría de conjuntos para construir cardinales transfinitos arbitrariamente grandes. De hecho, dicho isomorfismo es un epimorfismo (isomorfismo suprayectivo) y, por tanto, matemáticamente todos los cardinales transfinitos resultan ser un cardinal de tipo álef.

#### 5.1.1 álef 0 ( $N_0$ )

El más pequeño de todos los números transfinitos (cardinales), y el más simple de entender conceptualmente es  $N_0$  (se lee como álef sub cero o álef cero), este cardinal es el número de elementos del conjunto de los números naturales. En análisis matemático puede definirse de manera sencilla e intuitiva la clase de conjuntos numerables (conjuntos cuyo cardinal es  $N_0$ ). Cualquier conjunto que pueda ponerse en correspondencia biunívoca con los números naturales es un conjunto numerable. En términos prácticos, esto significa que los elementos de un conjunto numerable pueden "etiquetarse" como 1, 2, 3... de tal manera que a cada elemento de dicho conjunto



le corresponda un número natural (y nada más que un número natural).

Más formalmente, dentro de la teoría axiomática de conjuntos de Zermelo-Fraenkel, el axioma del infinito postula la existencia de un conjunto infinito que puede equipararse fácilmente con los números naturales cuyo cardinal resulta ser  $N_0$ .

### 5.1.2 álef 1 ( $N_1$ )

En matemáticas, se define  $N_1$  como el menor cardinal mayor que  $N_0$ , es decir, el menor cardinal mayor que el cardinal del conjunto de los números naturales. Es decir,  $N_1$  es el sucesor de  $N_0$ , lo cual se escribe  $N_1 = N_0^+$ .

En análisis matemático, se interpreta usualmente al cardinal  $N_1$  como la cantidad de números reales, asumiendo como cierta la hipótesis del continuo. Para justificar esto se parte del teorema de Cantor. Este teorema afirma que el cardinal de  $P(N)$  es mayor que  $N_0$ , donde  $\text{card}(P(N)) = \text{card}(R)$  es el cardinal del conjunto potencia de los números naturales, que es exactamente el mismo que el cardinal de los números reales. Así pues,  $N_0 < \text{card}(R)$ , lo que, considerando que  $\text{card}(R) = 2^{N_0}$ , puede escribirse también así:  $N_0 < 2^{N_0}$ .

En la teoría ZFC, el axioma de elección permite probar que  $N_1 \leq 2^{N_0}$ , mientras que la hipótesis del continuo que afirma que  $N_1 = 2^{N_0}$ , es decir, que el cardinal de los números reales es exactamente  $N_1$ . Sin embargo, los trabajos de Kurt Godel (1938) y Paul Cohen (1963) demostraron que de hecho la hipótesis del continuo es indecidible dentro de la axiomática de Zermelo-Fraenkel (ZF) y por tanto la hipótesis del continuo no puede ser demostrada a partir de ZF (ni desconfirmada dentro de la teoría de conjuntos ordinaria dada por los axiomas ZF).

### 5.1.3 Más allá de álef 1 ( $N_1$ )

El teorema de Cantor sobre el conjunto potencia afirma que para cualquier conjunto  $A$  se cumple que:

$$\text{card}(A) < \text{card}(\mathcal{P}(A))$$

Esto abre la posibilidad de que existan cardinales transfinitos mayores que  $N_1$ . La hipótesis del continuo generalizada de hecho permite ordenar los cardinales transfinitos de manera sencilla, ya que en esencia afirma que:

$$\forall n \geq 0 : (\text{card}(A) = N_n \rightarrow \text{card}(\mathcal{P}(A)) = N_{n+1})$$

#### 5.1.4 álef 2

El cardinal álef 2 ( $\aleph_2$ ) designa, asumiendo como válida la hipótesis del continuo generalizada, el cardinal transfinito del conjunto potencia de los números reales, y por tanto podría adoptarse como definición también  $N_2 = 2^{N_1}$ , por tanto, la cantidad de posibles subconjuntos de números reales sería  $N_2$ . Igualmente aceptando la hipótesis del continuo generalizada, puede demostrarse que  $N_2$  también es el cardinal del conjunto de todas las funciones reales ya que:

$$N_1^{N_1} = 2^{N_1} = N_2$$

Mientras que las funciones continuas tienen cardinal  $N_1$ , ya que

$$N_1^{N_0} = 2^{N_0} = N_1 < N_2.$$

Esto último se debe a que una función continua queda determinada si se especifica su valor sobre los números racionales, que son numerables y por tanto tienen  $N_0$  como cardinal.

El conjunto de todas las funciones reales tiene cardinal  $N_2 = N_1^{N_1}$  sin embargo, el conjunto de las funciones continuas tiene cardinal  $N_1 = N_1^{N_0}$ , ya que una función continua queda especificada si se conoce su valor sobre los números racionales, que son un conjunto numerable. El conjunto de partes de cualquier espacio vectorial real o complejo de dimensión finita tiene también cardinal  $N_2$ .

#### 5.1.5 álef $\omega$

En matemática, se define  $N_\omega$  como el cardinal singular (cardinal no regular) más pequeño de todos. A diferencia de  $N_\omega$ , los primeros cardinales transfinitos como  $N_0, N_1, N_2, \dots$  son todos ellos cardinales regulares. Otra propiedad notoria de  $N_\omega$  es que es un cardinal que no es sucesor ningún otro (a diferencia de lo que pasa con  $N_n$ ,  $n \in \mathbb{N}, n \geq 1$ ), ya que su índice  $\omega$  es un ordinal límite. El hecho de que  $N_\omega$  sea el cardinal singular más pequeño posible significa que es el cardinal más pequeño tal que su cofinalidad es menor que el propio cardinal, es decir:

$$\text{cf}(N_\omega) < N_\omega$$

Dado que el ordinal  $\omega$  coincide con el cardinal  $N_0$  (los dos signos representan el mismo conjunto), técnicamente se podría escribir el cardinal  $N_\omega$  como una aplicación reiterada de la función álef, es decir:

$$N_\omega = N_{N_0}$$

Aunque esa manera de escribirlo no es tan común.

### 5.1.6 Función álef

En teoría de conjuntos, la función álef es el único  $\epsilon$ -isomorfismo entre la clase de los ordinales  $On$  y la clase de los cardinales infinitos  $ICn$  es decir:

$$N : On \rightarrow ICn \subset On$$

Usualmente esta función se designa mediante  $N$  aunque es común escribir su valor sobre un ordinal  $\alpha$  como  $N_N$  más que como  $N(N)$ . Puede demostrarse que esta es una función normal, es decir, es una función monótona creciente y además continua (en el sentido de los ordinales).

## 5.2 Explosión del Challenger

Una dirección de la agencia espacial norteamericana más responsable habría solucionado este problema, que fue detectado -existen pruebas por escrito- en 1978, tres años antes de que se lanzara el primer transbordador espacial. El presidente de la comisión ha llegado a decir, que la NASA prácticamente encubrió las, pruebas de los defectos de diseño que originaron el desastre. "Las juntas de unión de los cohetes no estuvieron bien desde el día en que se fabricaron", afirmó un miembro de la comisión investigadora.

El informe de 200 páginas de la comisión, que ha presidido el ex secretario de Estado William Rogers, no contiene revelaciones espectaculares y huye en principio de denunciar nombres, pero muestra la inevitabilidad de lo sucedido: una explosión a nueve millas de altura sobre el Atlántico que conmocionó a EE UU y puso de manifiesto las limitaciones de la tecnología avanzada de la primera superpotencia. El informe recomendará una completa reorganización de la NASA, sobre todo en el proceso de toma de decisiones, para asegurar que las preocupaciones de los ingenieros y los técnicos lleguen hasta los ejecutivos y sean tenidas en cuenta por éstos.

### 5.2.1 Centralización

El documento recomendará también que los propios cosmonautas y los fabricantes del transbordador espacial aprueben los lanzamientos; una mayor centralización de la agencia espacial para que la sede de Washington controle mejor los centros de Houston, Florida y Alabama, que tienden a convertirse en feudos independientes; criterios más estrictos en el diseño de los cohetes propulsores y, en el control de calidad de la producción de los transbordadores (la NASA, ha reducido en un 70% el número de personas que se ocupan del control de calidad). El informe recomendará también que se

estudie un sistema de escape que permita a los cosmonautas abandonar la nave.

El catastrófico lanzamiento del Challenger se produjo a pesar de que los ingenieros fabricantes de los cohetes propulsores de combustible sólido - los dos grandes cigarros que la nave espacial lleva pegados al depósito de combustible, principal- advirtieron que los famosos O-rings que sellan las juntas de unión de los distintos segmentos del cohete habían perdido elasticidad a causa de las temperaturas bajo cero reinantes en la madrugada previa a la cuenta atrás. Estos anillos sintéticos tienen como misión impedir que los gases de la combustión de los propulsores salgan al exterior. Cincuenta y nueve segundos después del despegue, una lengua de fuego, como una pluma, apareció en una de las juntas de unión del cohete derecho, tal como se vio en las imágenes de televisión. Finalmente, el fuego alcanzó el gigantesco depósito naranja externo de combustible, una verdadera bomba colgada de la panza de la aeronave, que estalló.

El informe explicará que la NASA entró en problema en montañas de papel y siguió adelante. "Vuela y no pasa nada", era la filosofía que condujo al desastre y que ha descrito un miembro de la comisión investigadora el físico Richard Feynman: "Se sugiere que para el vuelo siguiente el riesgo ya no es tan grande y podemos deducir las precauciones un poco, ya que nos salió bien la vez anterior. Es como una ruleta rusa". Fallos reiterados Después del accidente, las juntas de los cohetes propulsores fallaron en las ocho ocasiones en que fueron probadas, en condiciones que simulaban los 3,3<sup>o</sup> de temperatura en Florida la madrugada previa al lanzamiento.

Pero aún queda por saber, y esto es algo que probablemente no descubra el informe, cómo, conociendo todos estos problemas a lo largo de años, la NASA seguía lanzando los transbordadores espaciales. Un ex director de seguridad de la agencia espacial estadounidense lo explica de este modo: "Había presión social. Miles de niños estaban esperando la primera lección desde el espacio. Había presión de la Prensa, y los responsables sintieron que si no lanzaban recibirían informaciones muy desfavorables por los retrasos. Y también existía una presión comercial: el cohete europeo Ariane estaba situando objetos en el espacio a un precio más bajo. Eran presiones sutiles, pero afectaron a la NASA".

### 5.3 Preguntas: Examen 1

- 1 De cuantas formas pueden colocarse los 11 jugadores de un equipo de futbol teniendo en cuenta que el portero no puede ocupar otra posicion distinta que la porteria?

$$P_{10} = 10! = 3628800$$

- 2 En una clase de 35 alumnos se quiere elegir un comite formado por tres alumnos. Cuantos comites diferentes se pueden formar?

$${}_{35}C_3 = (35 * 34 * 33)/(3 * 2 * 1) = 6545$$

- 3 A una reunion asisten 10 personas y se intercambian saludos entre todos. Cuantos saludos se han intercambiado?

$${}_{10}C_2 = (10 * 9)/2 = 45$$

- 4 Cuantas apuestas de Loteria Primitiva de una columna han de rellenarse para asegurarse el acierto de los seis resultados, de 49?

$${}_{49}C_6 = 49!/(49 - 6)! * 6! = 13983816$$

- 5 Se ordenan en una fila 5 bolas rojas, 2 bolas blancas y 3 bolas azules. Si las bolas de igual color no se distinguen entre si, de cuantas formas posibles pueden ordenarse?

$${}_{10}P_3 = 10!/5! * 2! * 3! = 2520$$

- 6 De cuantas formas se pueden sentar 3 parejas de casados alrededor de una mesa circular, si no debe haber dos mujeres juntas ni dos hombres juntos?

$$PC_3 = 2! * 3! = 2 * 6 = 12$$

- 7 De cuantas formas distintas pueden sentarse ocho personas en una fila de butacas?

$$P_8 = 8! = 40320$$

- 8 Con las letras de la palabra libro, cuantas ordenaciones distintas se pueden hacer que empiecen por vocal?

$$P_2! * P_4! = 2 * 4 * 3 * 2 * 1 = 48$$

- 9 Una mesa presidencial esta formada por ocho personas, de cuantas formas distintas se pueden sentar, si el presidente y el secretario siempre van juntos?

$$P_2! * P_7! = 2 * 7! = 2 * 5040 = 10080$$



- 10 De cuantas formas distintas pueden sentarse ocho personas alrededor de una mesa redonda?

$$P_8 = (8 - 1)! = 7! = 5040$$

## 5.4 AM y FM

### Qué significa FM?

La onda portadora se modula variando su frecuencia, sistema que se conoce como FM (frecuencia modulada). Presenta señales libres de interferencias y ruidos, y se efectúa en bandas de alta frecuencia, de 88 a 108 MHz.

#### 5.4.1 Qué significa AM?

La onda portadora se puede modular modificando la amplitud de la onda, sistema que se conoce con el nombre de AM (amplitud modulada). Se utiliza en las emisiones de radio y en la telefonía por onda portadora, entre otros servicios.

El funcionamiento de la radio se basa en la transmisión y recepción de ondas electromagnéticas. Sus antecedentes se remontan a fines del siglo XIX, cuando Heinrich Hertz ideó el oscilador, dispositivo que producía y recibía ondas de este tipo. Más tarde, este aparato fue perfeccionado por el italiano Augusto Righi. El francés Branly continuó con esas investigaciones y en 1890 creó el cohesor, que permitió transformar esas ondas electromagnéticas en impulsos eléctricos. Finalmente, en 1895, el italiano Guglielmo Marconi logró transmitir señales a distancia.

Un sistema de radio consta de varias etapas: la transmisión, en la que el transmisor genera corriente y la traslada a la antena encargada de irradiarla; la modulación de estas vibraciones antes de llegar a la antena; la radiación por la antena, y la recepción de las vibraciones por un dispositivo receptor. El nombre del pionero Hertz sirvió para designar la magnitud de ciclos por segundo, con la que se miden las ondas electromagnéticas. Un kilohercio (KHz) equivale a 1.000 ciclos por segundo, y un megahercio (MHz), a 1 millón de ciclos por segundo. VHF es una banda de frecuencia muy alta comprendida entre 30 y 300 MHz, y se utiliza para transmisiones de televisión, emisiones de radioaficionados y FM.

La FM o frecuencia modulada, vino a cambiar en todo el esquema que ex-



istia en lo que a transmisión y recepción de radio se refiere, libre de los molestos chasquidos que ocurren en AM, fidelidad, excelente. En adelante voy a tratar de explicar como se transmite la FM en estéreo, espero que a mis amigos estudiantes les sea de utilidad. En principio se usan 2 micrófonos, en cambio en AM es necesario únicamente 1, estos se encargan de recibir los sonidos que se transmitirán al espacio y recibidos por el receptor (también se usan los mismos 2 micrófonos para grabar), estos se separan uno del otro y por lo mismo reciben las ondas sonoras de forma diferente.

Asumamos que se va a grabar un concierto, uno de los micrófonos se colocará más cerca de los violines y el otro cerca de otros instrumentos, con esto se logra que, aunque los 2 micrófonos van a captar todos los sonidos, unos serán menos audibles en los altoparlantes o bocinas. Como cada micrófono reciben separadamente los sonidos, estos se pueden procesar por separado durante la transmisión o grabación. Podemos decir que un aparato estereofónico no es más que 2 amplificadores con características idénticas y las señales que recibe se aplican a bocinas separadas, lo cual le da un efecto agradable al oyente.

Cuando se graba en estéreo, esto resulta sencillo de reproducir, no así cuando se trata de transmitir estas señales, en vista de que se necesita transmitir por separado ambas, mismas que se procesan en el receptor para luego escucharlas tal y como se originaron. Después de varios diseños e intentos de un sistema sencillo que fuera compatible con los circuitos del receptor, se llegó a la perfección del sistema ?multiplex estereó de FM?, el cual fue aprobado por la FCC, y mediante el cual se puede transmitir el sonido en estéreo en una sola onda portadora en frecuencia modulada.

Una de las ventajas del multiplexado estereó de FM es que la reproducción del sonido es tan buena en los receptores estereofónicos como en los de FM normal, estos lo reproducen como una señal monofónica de FM.

#### **5.4.2 Cómo se genera una señal multiples de esteréo FM:**

Esto se inicia con las señales de audio que producen los micrófonos, Designándose como L y R (Left y Right) o sea izquierdo y derecho, las cuales se aplican a un circuito que se conoce como ?Matríz?, con lo cual se generan 2 nuevas señales, una de ellas corresponde a la suma instantánea de los valores de las señales L y R, y se le da el nombre de señal  $L + R$ ; la otra es la señal  $L - R$ , y corresponde a la diferencia instantánea de las dos señales básicas.

La señal  $L+R$  se usa para modular en amplitud una sub ? portadora de 38 Khz. , la cual produce como consecuencia bandas laterales de frecuencias superiores e inferiores a los 38 Khz.; esto permite que despues de la modulación se pueda suprimir la frecuencia sub ? portadora central de 38 Khz., con el fin de ahorrar espacio en la onda portadora principal se se transmitirá. Por tal razón solamente las bandas laterales AM resultantes del proceso anterior con la señal  $L+R$  y la portadora de 28 Khz. se aprovechan para modular en frecuencia a la portadora principal en conjunto con la señal  $L + R$ .

Las señales  $L$  y  $R$  siendo de audiofrecuencia tienen un ancho de banda limitado, ya que abarca de 0 á 15 Khz. Las frecuencias superiores a 15 Khz. se eliminan con la ayuda de filtros. Por lo mismo, la señal  $L + R$  que se transfiere a la portadora de Fm tiene un ancho de banda de únicamente 15 Khz. De las 2 bandas laterales que resultan de la modulación de la sub ? portadora auxiliar por la señal  $L+R$ , la inferior se ubica de 23 á 38 Khz. y la superior de 38 á 53 Khz., entonces la información que contiene la otra señal  $L + R$  queda separada por 8 Khz. de la información que contiene la otra señal  $L+R$ , gracias a esto se pueden distinguir y separar fácilmente en el detector del receptor, después que se transmite.

#### 5.4.3 Proceso del modulador:

Para el proceso demodulador en el receptor se necesita que la sub ? portadora de 28 Khz. esté completa, con sus bandas laterales y frecuencia central, tal como estaba en el momento de ser modulada con la señal  $L+R$  en la estación desde donde se transmitió. Como esta se suprimió para ahorrar espacio y potencia del transmisor, es necesario crear otra onda sub ? portadora en el receptor, con la misma fase y frecuencia a la que se quitó, con el fin de reinstalarla al demodulador, simulando que esta llegó por la antena. Para mantener informado y sincronizado al oscilador local del receptor con todo lo relativo a frecuencia y fase de la sub ? portadora usada para la modulación en el transmisor, se envía otra portadora piloto con las señales anteriormente descritas.

Para seguir con el ahorro de espacio en la banda de transmisión, la portadora piloto tiene únicamente 19 Khz. en lugar de 38. De esta forma cabe en una parte del espectro de frecuencias de la señal total, en un punto que no hay ninguna señal de audio. En el receptor se duplicará la frecuencia de

la portadora piloto para obtener la señal de sincronización de 38 KHz, indispensable para el demodulador separador de señal informativa. La portadora piloto también sirve para activar en el receptor un circuito que indica que la transmisión se lleva a cabo en estéreo, el cual se visualiza normalmente por un LED.

Cuando el receptor no está provisto de los circuitos especiales para estéreo, únicamente responde a la señal  $L + R$ , la cual es procesada como una señal monofónica normal. En un receptor diseñado para estéreo la señal  $L+R$  se demodula combinando las 2 bandas laterales AM de la señal  $L+R$  con una onda de 38 KHz. generada en el receptor y luego recuperando la señal original  $L+R$ . A continuación las señales  $L+R$  y  $L-R$  se suman en un circuito matriz el cual es similar al usado en el transmisor, para obtener la señal  $L$  original. También en la matriz se restan las señales  $L+R$  y  $L-R$  para producir la señal  $R$  original. El paso que sigue no es problema, ya que consiste en llevar estas 2 señales  $L$  y  $R$  dos canales de amplificación, como se dijo anteriormente.

## 6 Método de mínimos cuadrados, Densidad de Probabilidad

### 6.1 Mínimos cuadrados

El método de mínimos cuadrados consiste en encontrar la recta que mejor se ajusta. Una recta que mejor se ajusta es una línea recta que es la mejor aproximación del conjunto de datos dado.

Es usada para estudiar la naturaleza de la relación entre dos variables. Una recta que mejor se ajusta puede ser determinada aproximadamente usando el método visual al dibujar una línea recta en una gráfica de dispersión para que tanto el número de puntos arriba de la recta y debajo de la recta sean casi iguales (y la línea pasa a través de tantos puntos como sea posible).

Una forma más precisa de encontrar la recta que mejor se ajusta es el método de mínimos cuadrados. Use los 6 pasos siguientes para encontrar la ecuación de la recta que mejor se ajusta para un conjunto de parejas ordenadas.

1. Calcule la media de los valores de  $x$  y la media de los valores de  $y$ .
2. Realice la suma de los cuadrados de los valores de  $x$ .
3. Realice la suma de cada valor de  $x$  multiplicado por su valor correspondiente  $y$ .
4. Calcule la pendiente de la recta usando la fórmula:

$$m = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} \quad (17)$$

donde  $n$  es el número total de puntos de los datos.

5. Calcule la intercepción en  $y$  de la recta usando la fórmula:

$$b = \bar{y} - m\bar{x} \quad (18)$$

donde son las medias de las coordenadas de  $x$  y  $y$  de los puntos de datos respectivamente.

6. Use la pendiente y la intercepción en  $y$  para formar la ecuación de la recta.

## 6.2 Densidad de Probabilidad

En la teoría de la probabilidad, la función de densidad de probabilidad, función de densidad, o, simplemente, densidad de una variable aleatoria continua describe la probabilidad relativa según la cual dicha variable aleatoria tomará determinado valor. La probabilidad de que la variable aleatoria caiga en una región específica del espacio de posibilidades estará dada por la integral de la densidad de esta variable entre uno y otro límite de dicha región. La función de densidad de probabilidad (FDP o PDF en inglés) es no-negativa a lo largo de todo su dominio y su integral sobre todo el espacio es de valor unitario.

Una función de densidad de probabilidad caracteriza el comportamiento probable de una población en tanto especifica la posibilidad relativa de que una variable aleatoria continua  $X$  tome un valor cercano a  $x$ .

Una variable aleatoria  $X$  tiene densidad  $f$ , siendo  $f$  una función no-negativa integrable de Lebesgue, si:

$$P[a \leq X \leq b] = \int_a^b f(x) dx$$

Por lo tanto, si  $F$  es la función de distribución acumulativa de  $X$ , entonces:

$$F(x) = \int_{-\infty}^x f(u) du$$

y (si  $f$  es continua en  $x$ )

$$f(x) = \frac{d}{dx} F(x)$$

Intuitivamente, puede considerarse  $f(x)dx$  como la probabilidad de  $X$  de caer en el intervalo infinitesimal  $[x, x + dx]$ .

## 6.3 Distribución gaussiana en ROOT

```
#include<iostream>
#include<string>
#include<fstream>
using namespace std;

void gauss()
{

    //1. definir la funcion que nos defina la distribucion
    TF1* myFunc1 = new TF1("func1","gaus");
```

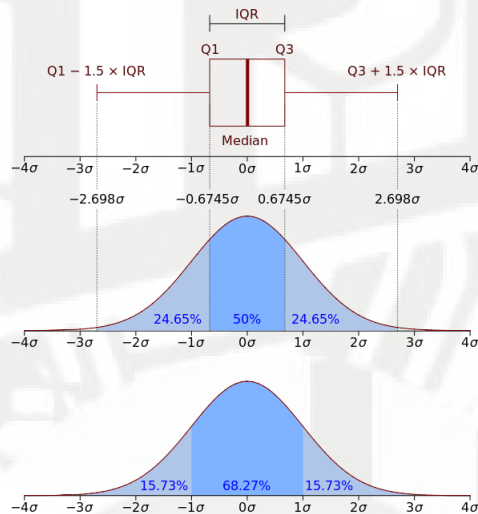


Figure 3: Función ejemplo de densidad. La probabilidad de un intervalo, es el área que existe entre la función y el eje de abscisas.

```
// fijamos los parametros de esta funcion 1: media 2: sigma
myFunc1->SetParameter(0, 1.0); // es la constante que multiplica la fun
myFunc1->SetParameter(1, 6.0); // es la media
myFunc1->SetParameter(2, 2.0); // es el sigma

// 2. definir un histograma
TH1F* myHist1 = new TH1F("Gauss", "Distribucion de valores aleatorios",

// 3. Llenar el histograma con 10000 numeros aleatorios que siguen la d
myHist1->FillRandom("func1", 10000000);

// 4. Algunas opciones graficas de presentacion del histograma
myHist1->GetXaxis()->SetTitle("Distribucion Gaussiana"); // titulo en X
myHist1->SetLineColor(20); // color de la linea
myHist1->SetFillColor(80);

// 5. Dibujar el histograma "e1" es la opcion para que se haga con barr
myHist1->Draw();

}
```



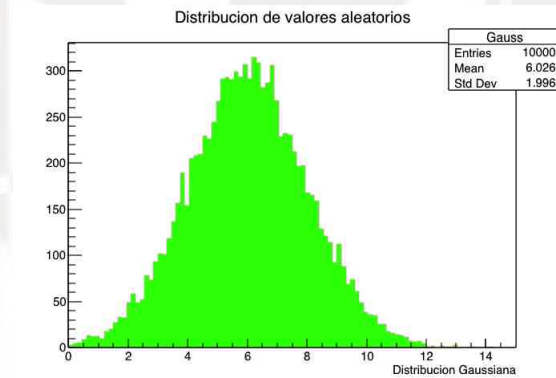


Figure 4: Función de distribución Gaussiana con números aleatorios en ROOT, utilizando 10000 valores

## 6.4 Media

En matemáticas y estadística una media o promedio es una medida de tendencia central que según la Real Academia Española, resulta al efectuar una serie determinada de operaciones con un conjunto de números y que, en determinadas condiciones, puede representar por sí solo a todo el conjunto. Existen distintos tipos de medias, tales como la media geométrica, la media ponderada y la media armónica aunque en el lenguaje común, el término se refiere generalmente a la media aritmética.

La media estadística se usa en estadística para dos conceptos diferentes aunque numéricamente similares:

La media muestral, que es un estadístico que se calcula a partir de la media aritmética de un conjunto de valores de una variable aleatoria. La media poblacional, valor esperado o esperanza matemática de una variable aleatoria. En la práctica dada una muestra estadística suficientemente grande el valor de la media muestral de la misma es numéricamente muy cercano a la esperanza matemática de la variable aleatoria medida en esa muestra. Dicho valor esperado, sólo es calculable si se conoce con toda exactitud la distribución de probabilidad, cosa que raramente sucede en la realidad, por esa razón, a efectos prácticos la llamada media se refiere normalmente a la media muestral.

#### 6.4.1 Media muestral

La media muestral es una variable aleatoria, ya que depende de la muestra, si bien es una variable aleatoria en general con una varianza menor que las variables originales usadas en su cálculo. Si la muestra es grande y está bien escogida, puede tratarse la media muestra como un valor numérico que aproxima con precisión la media poblacional, que caracteriza una propiedad objetiva de la población. Se define como sigue, si se tiene una muestra estadística de valores  $(X_1, X_2, \dots, X_n)$  de valores para una variable aleatoria  $X$  con distribución de probabilidad  $F(x, ?)$  [donde  $?$  es un conjunto de parámetros de la distribución] se define la media muestral  $n$ -ésima como:

$$\bar{X}_n = T(X_1, X_2, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i = \frac{X_1 + X_2 + \dots + X_n}{n}$$

#### 6.4.2 Media poblacional

La media poblacional técnicamente no es una media sino un parámetro fijo que coincide con la esperanza matemática de una variable aleatoria. El nombre "media poblacional" se usa para significar que valor numérico de una media muestral es numéricamente cercano al parámetro media poblacional, para una muestra adecuada y suficientemente grande.

### 6.5 Medidas de dispersión

Así como las medidas de tendencia central nos permiten identificar el punto central de los datos, las Medidas de dispersión nos permiten reconocer que tanto se dispersan los datos alrededor del punto central; es decir, nos indican cuanto se desvían las observaciones alrededor de su promedio aritmético (Media). Este tipo de medidas son parámetros informativos que nos permiten conocer como los valores de los datos se reparten a través de eje  $X$ , mediante un valor numérico que representa el promedio de dispersión de los datos. Las medidas de dispersión más importantes y las más utilizadas son la Varianza y la Desviación estándar (o Típica).

#### 6.5.1 Varianza

Esta medida nos permite identificar la diferencia promedio que hay entre cada uno de los valores respecto a su punto central (Media  $\bar{X}$ ). Este promedio es calculado, elevando cada una de las diferencias al cuadrado (Con el fin de eliminar los signos negativos), y calculando su promedio o media; es decir, sumado todos los cuadrados de las diferencias de cada valor respecto a la media y dividiendo este resultado por el número de observaciones que se

tengan. Si la varianza es calculada a una población (Total de componentes de un conjunto), la ecuación sería:

$$s^2 = \frac{(X_1 - \bar{\mu})^2 + (X_2 - \bar{\mu})^2 + (X_3 - \bar{\mu})^2 + \dots + (X_n - \bar{\mu})^2}{(n-1)} = \frac{\sum_{i=1}^n (X_i - \bar{\mu})^2}{(n-1)}$$

Ecuación de la varianza para Poblaciones - Medidas de Dispersion

Donde  $(\sigma^2)$  representa la varianza,  $(X_i)$  representa cada uno de los valores,  $(\bar{\mu})$  representa la media poblacional y  $(N)$  es el número de observaciones ó tamaño de la población.

### 6.5.2 Desviación estándar

Esta medida nos permite determinar el promedio aritmético de fluctuación de los datos respecto a su punto central o media. La desviación estándar nos da como resultado un valor numérico que representa el promedio de diferencia que hay entre los datos y la media. Para calcular la desviación estándar basta con hallar la raíz cuadrada de la varianza, por lo tanto su ecuación sería:

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

## 6.6 Graficando en R(Minimos cuadrados)

### 6.6.1 Generación de Matriz

Para crear una matriz en R se utiliza la función

```
matrix(data=NA, nrow=1, ncol=1, byrow= FALSE, dimnames=NULL)
```

donde `matrix()` corresponde al nombre de la función y todo lo que está dentro de los paréntesis son los argumentos de dicha función.

Argumentos:

- `data` - Es un vector de datos opcional
- `nrow` - número deseado de filas
- `ncol` - número deseado de columnas
- `byrow` - valor lógico. Si es `valso`(valor por defecto), la matriz se llena por orden columna, de otra manera se llenará primero por filas

- `dimnames` - utilizado para darles nombres a las filas ya las columnas, respectivamente.

*Ejemplo*

```
matrix(c(5,7,13,4,1,7,14,3,11), nrow=3, ncol=3, byrow=FALSE,
dimnames=list(c("blanco", "negro", "rojo"), c("toyota", "audi", "nissan")))
```

	<i>toyota</i>	<i>audi</i>	<i>nissan</i>
<i>blanco</i>	5	4	14
<i>negro</i>	7	1	3
<i>rojo</i>	13	7	11

Se hubiera obtenido el mismo resultado si sólo se escribe uno de los argumentos `nrow=3` o `ncol=3`, no hace falta escribir los dos.

### 6.6.2 Agregar filas y columnas

Para agregar una fila a una matriz ya existente, se debe utilizar la función `rbind()` y se usará la función `cbind()` para añadir columnas a una matriz.

*Ejemplo*

```
A <- -c(8, 5, 7) # con esto se crea un vector
M <- -rbind(M, A) # Agrega A a la matriz como fila
```

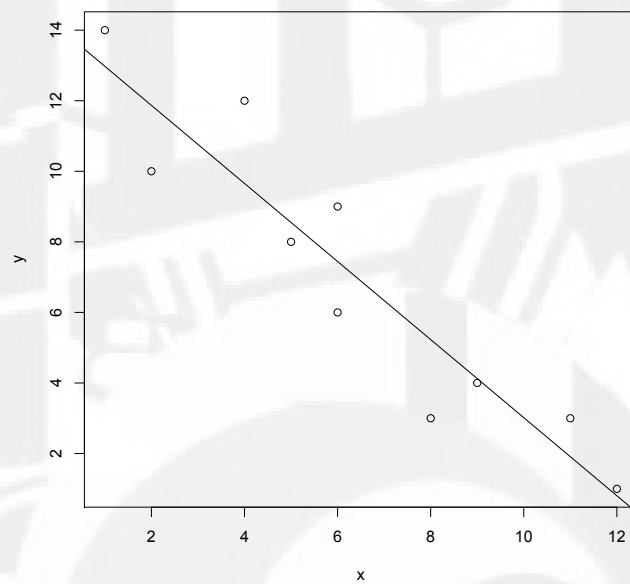
### 6.6.3 Función "lm"

Esta función sirve para el calculo y representación de la recta de mínimos cuadrados. el primer argumento de este comando es una fórmula  $y \sim x$  en la que se especifica cuál es la variable dependiente y cuál es la variable independiente. En el segundo argumento, llamdo *data* especifica cuál es el fichero en el que se encuentran las variables.

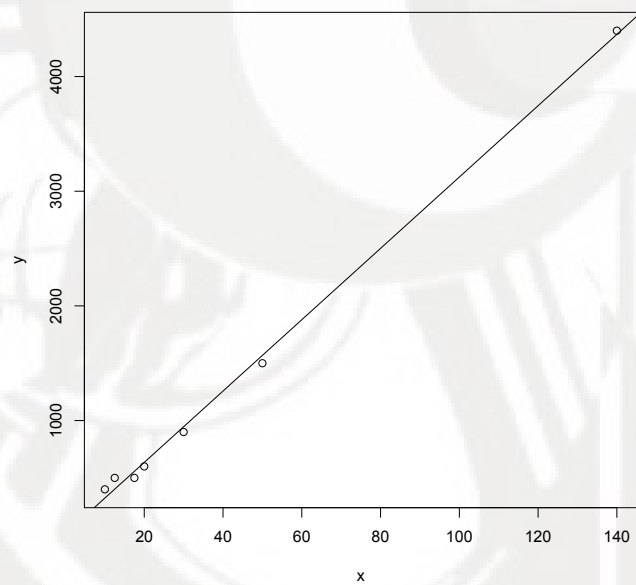
*Ejemplo*

```
x <- -c(8, 2, 11, 6, 5, 4, 12, 9, 6, 1)
y <- -c(3, 10, 3, 6, 8, 12, 1, 4, 9, 14)
plot(x,y,main="Ejemplo 1 de minimos cuadrados")
cor(x, y)
dim <- -lm(y ~ x)
lm(formula = y ~ x)
abline(dim)
summary(dim)
```

**Ejemplo Minimos Cuadrados**



**Ejemplo 2 Minimos Cuadrados**



## 6.7 Ejemplos de Mínimos cuadrados

De acuerdo a la información mostrada a continuación, determina: Cuáles serán los costos en una jornada de trabajo de 40 horas?

MES	COSTO(y)	HORAS(x)
Enero	400	10.00
Febrero	500	12.50
Marzo	500	17.50
Abril	600	20.00
Mayo	1,500	50.00
Junio	900	30.00
Total	4,400	140.00

Table 1: Costos de trabajo.

Encontramos la solución de la siguiente forma:

MES	COSTO(y)	HORAS(x)	(x)(y)	$x^2$
Enero	400	10.00	4000	100
Febrero	500	12.50	6250	156
Marzo	500	17.50	8750	306
Abril	600	20.00	12000	400
Mayo	1,500	50.00	75000	2500
Junio	900	30.00	27000	900
Total	4,400	140.00	133,000	4363

Table 2: Solución

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (19)$$

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \quad (20)$$

$$b = \frac{6(133000) - (140)(4400)}{6(4363) - (140)^2} = \frac{798000 - 616000}{26178 - 19600} = \frac{182000}{6578} = 27.67 \quad (21)$$



$$a = \frac{4400(4363) - (140)(133000)}{6(4363) - (140)^2} = \frac{19197200 - 18620000}{26178 - 19600} = \frac{577200}{6578} = 87.75 \quad (22)$$

$$Y = a + bx \quad (23)$$

$$Y = 87.75 + 27.67(x) \quad (24)$$

Con esta ecuación de mínimo cuadrado se pueden predecir los costos totales aproximados de acuerdo a las horas laboradas.

$$Y = 87.75 + 27.67(40) = 87.75 + 1106.8 = 1194.55 \quad (25)$$

Use el método de mínimos cuadrados para determinar la ecuación de la recta que mejor se ajusta para los datos. Luego grafique la recta.

x	8	2	11	6	5	4	12	9	6	1
y	3	10	3	6	8	12	1	4	9	14

Para mostrar la solución:

Grafique los puntos en un plano coordenado. Calcule las medias de los valores de  $x$  y los valores de  $y$ , la suma de los cuadrados de los valores de  $x$ , y la suma de cada valor de  $x$  multiplicado por su valor correspondiente  $y$ .

Calcule la pendiente

$$m = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{317 - \frac{(64)(70)}{10}}{528 - \frac{(64)^2}{10}} = -1.1 \quad (26)$$

Calcule la intercepción en  $y$ .

Primero, calcule la media de los valores de  $x$  y la media de los valores de  $y$ .

$$\bar{x} = \frac{\sum x}{n} = \frac{64}{10} = 6.4 \quad (27)$$

$$\bar{y} = \frac{\sum y}{n} = \frac{70}{10} = 7.0 \quad (28)$$

x	y	xy	$x^2$
8	3	24	64
2	10	20	4
11	3	33	121
6	6	36	36
5	8	40	25
4	12	48	16
12	1	12	144
9	4	36	81
6	9	54	36
1	14	14	1
$\sum x = 64$	$\sum y = 70$	$\sum xy = 317$	$\sum x^2 = 528$

Use la pendiente y la intercepción en y para formar la ecuación de la recta que mejor se ajusta.

La pendiente de la recta es  $-1.1$  y la intercepción en y es  $14.0$ .

Por lo tanto, la ecuación es  $y = -1.1x + 14.0$ .

## 7 Vector y valor propio, análisis de Componentes principales

### 7.1 Vector propio y valor propio

En álgebra lineal, los vectores propios, autovectores o eigenvectores de un operador lineal son los vectores no nulos que, cuando son transformados por el operador, dan lugar a un múltiplo escalar de sí mismos, con lo que no cambian su dirección. Este escalar  $\lambda$  recibe el nombre valor propio, autovector, valor característico o eigenvalor. A menudo, una transformación queda completamente determinada por sus vectores propios y valores propios. Un espacio propio, autoespacio, eigenespacio o subespacio fundamental asociado al valor propio  $\lambda$  es el conjunto de vectores propios con un valor propio común.

La palabra alemana eigen, que se traduce en español como propio, se usó por primera vez en este contexto por David Hilbert en 1904 (aunque Helmholtz la usó previamente con un significado parecido). Eigen se ha traducido también como inherente, característico o el prefijo auto-, donde se aprecia

el énfasis en la importancia de los valores propios para definir la naturaleza única de una determinada transformación lineal. Las denominaciones vector y valor característicos también se utilizan habitualmente.

Las transformaciones lineales del espacio ¿como la rotación, la reflexión, el ensanchamiento, o cualquier combinación de las anteriores; en esta lista podrían incluirse otras transformaciones? pueden interpretarse mediante el efecto que producen en los vectores. Los vectores pueden visualizarse como flechas de una cierta longitud apuntando en una dirección y sentido determinados.

- Los vectores propios de las transformaciones lineales son vectores que, o no se ven afectados por la transformación o se ven multiplicados por un escalar, y por tanto no varían su dirección.
- El valor propio de un vector propio es el factor de escala por el que ha sido multiplicado.
- Un espacio propio es un espacio formado por todos los vectores propios del mismo valor propio, además del vector nulo, que no es un vector propio.
- La multiplicidad geométrica de un valor propio es la dimensión del espacio propio asociado.
- El espectro de una transformación en espacios vectoriales finitos es el conjunto de todos sus valores propios.

Por ejemplo, un vector propio de una rotación en tres dimensiones es un vector situado en el eje de rotación sobre el cual se realiza la rotación. El valor propio correspondiente es 1 y el espacio propio es el eje de giro. Como es un espacio de una dimensión, su multiplicidad geométrica es uno. Es el único valor propio del espectro (de esta rotación) que es un número real.

## 7.2 Análisis de componentes principales

En estadística, el análisis de componentes principales (en español ACP, en inglés, PCA) es una técnica utilizada para reducir la dimensionalidad de un conjunto de datos. Intuitivamente la técnica sirve para hallar las causas de la variabilidad de un conjunto de datos y ordenarlas por importancia.

Técnicamente, el ACP busca la proyección según la cual los datos queden

mejor representados en términos de mínimos cuadrados. El ACP se emplea sobre todo en análisis exploratorio de datos y para construir modelos predictivos. El ACP comporta el cálculo de la descomposición en autovalores de la matriz de covarianza, normalmente tras centrar los datos en la media de cada atributo.

El ACP construye una transformación lineal que escoge un nuevo sistema de coordenadas para el conjunto original de datos en el cual la varianza de mayor tamaño del conjunto de datos es capturada en el primer eje (llamado el Primer Componente Principal), la segunda varianza más grande es el segundo eje, y así sucesivamente. Para construir esta transformación lineal debe construirse primero la matriz de covarianza o matriz de coeficientes de correlación. Debido a la simetría de esta matriz existe una base completa de vectores propios de la misma. La transformación que lleva de las antiguas coordenadas a las coordenadas de la nueva base es precisamente la transformación lineal necesaria para reducir la dimensionalidad de datos. Además las coordenadas en la nueva base dan la composición en factores subyacentes de los datos iniciales.

Una de las ventajas del ACP para reducir la dimensionalidad de un grupo de datos, es que retiene aquellas características del conjunto de datos que contribuyen más a su varianza, manteniendo un orden de bajo nivel de los componentes principales e ignorando los de alto nivel. El objetivo es que esos componentes de bajo orden a veces contienen el aspecto "más importante" de esa información.

### 7.2.1 Matemáticas del ACP

Supongamos que existe una muestra con  $n$  individuos para cada uno de los cuales se han medido  $m$  variables (aleatorias)  $F_j$ . El ACP permite encontrar un número de factores subyacentes  $p \leq m$  que explican aproximadamente el valor de las  $m$  variables para cada individuo. El hecho de que existan estos  $p$  factores subyacentes puede interpretarse como una reducción de la dimensionalidad de los datos: donde antes necesitábamos  $m$  valores para caracterizar a cada individuo ahora nos bastan  $p$  valores. Cada uno de los  $p$  encontrados se llama componente principal, de ahí el nombre del método.

Existen dos formas básicas de aplicar el ACP:

1. Método basado en la matriz de correlación, cuando los datos no son dimensionalmente homogéneos o el orden de magnitud de las variables aleatorias medidas no es el mismo.

2. Método basado en la matriz de covarianzas, que se usa cuando los datos son dimensionalmente homogéneos y presentan valores medios similares.

### 7.2.2 Método basado en correlaciones

El método parte de la matriz de correlaciones, consideremos el valor de cada una de las  $m$  variables aleatorias  $F_j$ . Para cada uno de los  $n$  individuos tomemos el valor de estas variables y escribamos el conjunto de datos en forma de matriz:

$$(F_j^\beta)_{j=1, \dots, m}^{\beta=1, \dots, n} \quad (29)$$

Obsérvese que cada conjunto

$$\mathcal{M}_j = \{F_j^\beta | \beta = 1, \dots, n\} \quad (30)$$

puede considerarse una muestra aleatoria para la variable  $F_j$ . A partir de los  $m \times n$  datos correspondientes a las  $m$  variables aleatorias, puede construirse la matriz de correlación muestral, que viene definida por:

$$\mathbf{R} = [r_{ij}] \in M_{m \times m} \quad \text{donde} \quad r_{ij} = \frac{\text{cov}(F_i, F_j)}{\sqrt{\text{var}(F_i)\text{var}(F_j)}} \quad (31)$$

Puesto que la matriz de correlaciones es simétrica entonces resulta diagonalizable y sus valores propios  $\lambda_i$  verifican

$$\sum_{i=1}^m \lambda_i = 1 \quad (32)$$

Debido a la propiedad anterior estos  $m$  valores propios reciben el nombre de pesos de cada uno de los  $m$  componentes principales. Los factores principales identificados matemáticamente se representan por la base de vectores propios de la matriz  $R$ . Está claro que cada una de las variables puede ser expresada como combinación lineal de los vectores propios o componentes principales.

### 7.2.3 Método basado en las covarianzas

El objetivo es transformar un conjunto dado de datos  $X$  de dimensión  $n \times m$  a otro conjunto de datos  $Y$  de menor dimensión  $n \times l$  con la menor pérdida de información útil posible utilizando para ello la matriz de covarianza.

Se parte de un conjunto  $n$  de muestras cada una de las cuales tiene  $m$  variables que las describen y el objetivo es que, cada una de esas muestras, se describa con solo  $l$  variables, donde  $l < m$ . Además, el número de componentes principales  $l$  tiene que ser inferior a la menor de las dimensiones de  $X$ .

$$l \leq \min\{n, m\} \quad (33)$$

Los datos para el análisis tienen que estar centrados a media 0 (restándoles la media de cada columna) y/o autoescalados (centrados a media 0 y dividiendo cada columna por su desviación estándar).

$$X = \sum_{a=1}^l t_a p_a^T + E \quad (34)$$

Los vectores  $t_a$  son conocidos como scores y contienen la información de cómo las muestras están relacionadas unas con otras además, tienen la propiedad de ser ortogonales. Los vectores  $p_a$  se llaman loadings e informan de la relación existente entre las variables y tienen la cualidad de ser ortonormales. Al coger menos componentes principales que variables y debido al error de ajuste del modelo con los datos, se produce un error que se acumula en la matriz  $E$ .

El PCA se basa en la descomposición en vectores propios de la matriz de covarianza. La cual se calcula con la siguiente ecuación:

$$\text{cov}(X) = \frac{X^T X}{n-1} \quad (35)$$

$$\text{cov}(X) p_a = \lambda_a p_a \quad (36)$$

$$\sum_{a=1}^m \lambda_a = 1 \quad (37)$$

Donde  $\lambda_a$  es el valor propio asociado al vector propio  $p_a$ . Por último,  $t_a = X p_a$

Esta ecuación la podemos entender como que  $t_a$  son las proyecciones de  $X$  en  $p_a$ , donde los valores propios  $\lambda_a$  miden la cantidad de varianza capturada, es decir, la información que representan cada uno de los componentes principales. La cantidad de información que captura cada componente principal va disminuyendo según su número es decir, el componente principal número uno representa más información que el dos y así sucesivamente.

La aplicación del ACP está limitada por varios supuestos:



- Asunción de linealidad: Se asume que los datos observados son combinación lineal de una cierta base.
- Importancia estadística de la media y la covarianza: el ACP utiliza los vectores propios de la matriz de covarianzas y sólo encuentra las direcciones de ejes en el espacio de variables considerando que los datos se distribuyen de manera gaussiana.

### 7.3 Demostración $\overline{\Delta u} = \overline{u} - \bar{u} = \bar{u} - \bar{u} = 0$

Partiendo de la ecuación

$$\Delta u = u - \bar{u} \quad (38)$$

El promedio de (38) se denota como

$$\overline{\Delta u} = \overline{u - \bar{u}} \quad (39)$$

Sabiendo que el promedio es

$$\bar{u} = \sum_{i=1}^M P(u_i)(u_i) \quad (40)$$

Aplicando el concepto de la ecuación (40) tenemos

$$\overline{\Delta u} = \sum_{i=0}^M P(u_i)(u_i - \bar{u}_i) \quad (41)$$

$$\begin{aligned} \overline{\Delta u} &= \sum_{i=0}^M P(u_i)u_i - \sum_{i=0}^M P(u_i)\bar{u} \\ &= \sum_{i=0}^M P(u_i)u_i - \bar{u} \sum_{i=0}^M P(u_i) \end{aligned} \quad (42)$$

Sabemos que  $\sum_{i=1}^M P(u_i) = 1$ , por lo que sustituyendo en (42)

$$\begin{aligned} \overline{\Delta u} &= \bar{u} - (1)\bar{u} \\ \overline{\Delta u} &= \bar{u} - \bar{u} = 0 \quad \square \end{aligned} \quad (43)$$

#### 7.4 Demostración $\overline{(\Delta u)^2} = \sum_{i=0}^M P(u_i)(u_i - \bar{u})^2 \geq 0$

Partiendo de la ecuación (38), y elevando ambos lados de la ecuación al cuadrado

$$(\Delta u)^2 = (u - \bar{u})^2 \quad (44)$$

$(\Delta u)^2$  es siempre positivo o mayor que 0, entonces

$$(\Delta u)^2 = (u - \bar{u})^2 \geq 0 \quad (45)$$

Asi que, obteniendo el promedio de  $(\Delta u)^2$

$$\begin{aligned} \overline{(\Delta u)^2} &= \overline{(u - \bar{u})^2} \geq 0 \\ \overline{(\Delta u)^2} &= \sum_{i=1}^M P(u_i)(u_i - \bar{u})^2 \geq 0 \quad \square \end{aligned} \quad (46)$$

#### 7.5 Demostración $\overline{(u - \bar{u})^2} = \bar{u}^2 - \bar{u}^2$

Partiendo de la ecuación (44), se puede desarrollar el binomio del lado derecho de la ecuación

$$(\Delta u)^2 = (u - \bar{u})^2 = u^2 - 2u\bar{u} + (\bar{u})^2 \quad (47)$$

Obteniendo el promedio de la función se tiene

$$\begin{aligned} \overline{(u - \bar{u})^2} &= \overline{u^2 - 2u\bar{u} + (\bar{u})^2} \\ &= \bar{u}^2 - 2\bar{u}^2 + \bar{u}^2 \\ \overline{(u - \bar{u})^2} &= \bar{u}^2 - \bar{u}^2 \quad \square \end{aligned} \quad (48)$$

#### 7.6 Demostración $\bar{u}^2 \geq \bar{u}^2$

Dado que  $\overline{(\Delta u)^2} \geq 0$  y  $\overline{(\Delta u)^2} = \bar{u}^2 - \bar{u}^2$  como se mostró anteriormente. Entonces:

$$\bar{u}^2 - \bar{u}^2 \geq 0 \quad (49)$$

Resolviendo la desigualdad se demuestra que:

$$\bar{u}^2 \geq \bar{u}^2 \quad \square \quad (50)$$

## 7.7 Multiplicidad eigenvectores y eigenvalores

### 7.7.1 Multiplicidad Algebraica

La multiplicidad algebraica de un valor propio  $\lambda$  de  $A$  es el orden de  $\lambda$  como cero del polinomio característico de  $A$ ; en otras palabras, si  $\lambda$  es una de las raíces del polinomio, es el número de factores  $(t - \lambda)$  en el polinomio característico tras la factorización. Una matriz  $n \times n$  tiene  $n$  valores propios, contados de acuerdo con su multiplicidad algebraica, ya que su polinomio característico tiene grado  $n$ .

Un valor propio de multiplicidad algebraica 1 recibe el nombre de "valor propio simple".

Por ejemplo, se pueden encontrar exposiciones como la siguiente en artículos de teoría de matrices:

"Los valores propios de una matriz  $A$  son 4,4,3,3,3,2,2,1,"

Lo que significa que la multiplicidad algebraica de 4 es dos, la de 3 es tres, la de 2 es dos y la de 1 es uno. Se emplea este estilo porque la multiplicidad algebraica es la clave de muchas demostraciones matemáticas en teoría de matrices.

Anteriormente se ha definido la multiplicidad geométrica de un valor propio como la dimensión del espacio propio asociado, o el núcleo (espacio propio de los vectores propios del valor propio nulo) de  $\lambda I - A$ . Por ejemplo:

La multiplicidad algebraica también puede entenderse como una dimensión: es la dimensión del espacio propio generalizado (1.er sentido) asociado, que es el núcleo de la matriz  $(\lambda I - A)^k$  para  $k$  suficientemente grande. Es decir, es el espacio de los vectores propios generalizados (1.er sentido), donde un vector propio generalizado es cualquier vector que toma valor 0 si  $\lambda I - A$  se aplica suficientes veces en sucesión.

Cualquier vector propio es un vector propio generalizado, así que cualquier espacio propio está contenido en el espacio propio generalizado asociado. Esto proporciona una demostración simple de que la multiplicidad geométrica es siempre menor o igual a la algebraica. El primer sentido no debe confundirse con el problema de valores propios generalizados tal y como se muestra más adelante.

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

Solo tiene un valor propio  $\lambda = 1$ . El polinomio característico es  $(\lambda - 1)^2$ , as que este valor propio tiene multiplicidad algebraica 2. Sin embargo, el espacio propio asociado es el eje, que normalmente recibe el nombre de eje x, generado por el vector unitario

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

asi que la multiplicidad geometrica es 1.

Los vectores propios generalizados pueden usarse para calcular la forma normal de Jordan de una matriz (comentado mas adelante). El hecho de que los bloques de Jordan en general no son diagonales sino nilpotentes esta directamente relacionado con la distincion entre vectores propios y vectores propios generalizados.

### 7.7.2 Matriz singular

Matriz cuadrada cuyo determinante es igual a cero. Una matriz singular no tiene matriz inversa.

$$\begin{pmatrix} 3 & 2 \\ 6 & 4 \end{pmatrix}$$

$$\text{Determinante de } A = (3 * 4) - (2 * 6) = 0$$

### 7.7.3 Traza de una matriz

Sea una matriz cuadrada A de orden n, se define la traza de la matriz A y se denota por  $tr(A)$  al valor obtenido al sumar todos los elementos de la diagonal principal, es decir

$$tr(A) = \sum_{i=1}^n a_{ii}$$

Veamos algunos ejemplos considerando las siguientes matrices

$$a = \begin{pmatrix} 1 & 5 & 2 \\ 0 & 1 & 2 \\ 3 & 1 & 0 \end{pmatrix}$$

$$b = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 4 & 0 \\ 0 & 0 & -5 \end{pmatrix}$$

$$c = \begin{pmatrix} 2 & 6 & 1 \\ 4 & 3 & 5 \\ 2 & 3 & -5 \end{pmatrix}$$

Es sencillo obtener la traza, ya que basta sumar  $1 + 1 + 0 = 2$

La traza de B  $1 + 4 - 5 = 0$

La traza de C  $2 + 3 - 5 = 0$

#### 7.7.4 Transpuesta de una matriz

Sea A una matriz con m filas y n columnas. La matriz transpuesta, denota con  $A^t$

Esta dado por:

$$(A^t)_{ij} = A_{ji}, 1 \leq i \leq n, 1 \leq j \leq m$$

En donde el elemento  $a_{ji}$  de la matriz original A se convertira en el elemento  $a_{ij}$  de la matriz transpuesta  $A^t$

Ejemplo

$$= \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

### 7.7.5 Ejemplos de Eigenvalores y Eigenvectores

1 Para la matriz A indique cuales vectores son vectores propios

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

$$V1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad V2 = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \quad V3 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad V4 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

Solucion

Debemos multiplicar cada vector por la matriz A y ver si el vector resultante es un multiplo escalar del vector

$$Av_1 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

V1 si es vector propio de A asociado al valor propio 3

$$Av_2 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 8 \\ 7 \end{pmatrix} \neq k \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

v2 no es vector propio de A

$$Av_3 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -1 \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

V3 si es vector propio de A asociado al valor propio -1

$$Av_4 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix} \neq k \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

V4 no es vector propio de A

2 Determine los valores y los vectores propios correspondientes de las matrices:

$$A_1 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, A_2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, A_3 = \begin{pmatrix} 1 & 2 \\ -1 & 2 \end{pmatrix}$$

Solucion:



Para  $A_1$ :

$$PA(\lambda) = \det(A - \lambda I_2) = \det \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$PA(\lambda) = \det \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} = \det \begin{pmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{pmatrix}$$

$$PA(\lambda) = \det(A - \lambda I_2) = \det \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} PA_1(\lambda) = \begin{vmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - 4$$

$$PA_1(\lambda) = \lambda^2 - 2\lambda - 3 = (\lambda - 3)(\lambda + 1)$$

Por tanto los únicos valores propios de  $A_1$  son  $\lambda_1 = 3$  y  $\lambda_2 = -1$   
Vector propio para  $\lambda_1 = 3$

### 7.7.6 Otros ejemplos

Determinar el eigenvalor y eigenvector de  $A = \begin{bmatrix} 3 & 5 \\ -2 & -4 \end{bmatrix}$

Para esta matriz

$$A - \lambda I = \begin{bmatrix} 3 & 5 \\ -2 & -4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3-\lambda & 5 \\ -2 & -4-\lambda \end{bmatrix}$$

Por lo tanto

$$\det(A - \lambda I) = (3 - \lambda)(-4 - \lambda) - 5(-2) = \lambda^2 + \lambda - 2$$

La ecuación característica de  $A$  es  $\lambda^2 + \lambda - 2 = 0$ ; cuando se resuelve para  $\lambda$ , nos da los eigenvalores  $\lambda = 1$  y  $\lambda = -2$ .

Como revisión, utilizamos la Propiedad 7.1: La traza de  $A$  es  $3 + (-4) = -1$ , que es la suma de los eigenvalores.

Los eigenvectores correspondientes de  $\lambda = 1$  son obtenidos por la resolución de la Ec.(1) para  $X = [x_1, x_2]^T$  con el valor de  $\lambda$ . Después tenemos

$$\left( \begin{bmatrix} 3 & 5 \\ -2 & -4 \end{bmatrix} - 1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 5 \\ -2 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

que es equivalente a las ecuaciones lineales

$$\begin{aligned} 2x_1 + 5x_2 &= 0 \\ -2x_1 - 5x_2 &= 0 \end{aligned}$$

La solución a este sistema es  $x_1 = \lambda x_2$  con  $x_2$  arbitrariamente, los eigenvectores corresponden a  $\lambda = 1$  son

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -\lambda x_1 \\ x_2 \end{bmatrix} = X_2 \begin{bmatrix} -5/2 \\ 1 \end{bmatrix}$$

Cuando  $\lambda = -2$  puede ser

$$\begin{aligned} \begin{bmatrix} 3 & 5 \\ -2 & -4 \end{bmatrix} - (-2) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ ó} \\ \begin{bmatrix} 5 & 5 \\ -2 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{aligned}$$

que es equivalente a las ecuaciones lineales

$$\begin{aligned} 2x_1 + 5x_2 &= 0 \\ -2x_1 - 5x_2 &= 0 \end{aligned}$$

La solución a este sistema es  $x_1 = -x_2$  con  $x_2$  arbitrariamente, los eigenvectores corresponden a  $\lambda = -2$  es

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_1 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Determinando los eigenvalores y eigenvectores

$$A = \begin{bmatrix} 5 & 2 & 2 \\ 3 & 6 & 3 \\ 6 & 6 & 9 \end{bmatrix}$$

Para esta matriz

$$A - \lambda I = \begin{bmatrix} 5 & 2 & 2 \\ 3 & 6 & 3 \\ 6 & 6 & 9 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 5-\lambda & 2 & 2 \\ 3 & 6-\lambda & 3 \\ 6 & 6 & 9-\lambda \end{bmatrix}$$

El determinante de esta ultima matriz puede ser obtenida con la expresión de cofactores, que es

$$-\lambda^3 + 20\lambda^2 - 93\lambda + 126 = -(\lambda - 3)^2(\lambda - 14) \quad (51)$$

La ecuación característica de  $A$  es  $-(\lambda - 3)^2(\lambda - 14) = 0$ , que tiene como su solución el eigenvalor  $\lambda = 3$  de la multiplicación de dos y el eigenvalor  $\lambda = 14$  de la multiplicación de uno. Equivale a la Propiedad 7.1.

Los eigenvectores corresponden a  $\lambda = 3$  se obtienen resolviendo por  $X = [x_1, x_2, x_3]^T$  que es el valor de  $\lambda$ :

$$\left( \begin{bmatrix} 5 & 2 & 2 \\ 3 & 6 & 3 \\ 6 & 6 & 9 \end{bmatrix} - 3 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

que es equivalente a las ecuaciones lineales

$$\begin{aligned} 2x_1 + 2x_2 + 2x_3 &= 0 \\ 3x_1 + 3x_2 + 3x_3 &= 0 \\ 6x_1 + 6x_2 + 6x_3 &= 0 \end{aligned}$$

La solución a este sistema es  $x_1 = -x_2 - x_3$  con  $x_2$  y  $x_3$  arbitrariamente, los eigenvectores corresponden a  $\lambda = 3$  son

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_2 - x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Cuando  $\lambda = 14$  se convierte

$$\left( \begin{bmatrix} 5 & 2 & 2 \\ 3 & 6 & 3 \\ 6 & 6 & 9 \end{bmatrix} - 14 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{ó}$$

$$\begin{bmatrix} -9 & 2 & 2 \\ 3 & -8 & 3 \\ 6 & 6 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

que es equivalente a las ecuaciones lineales

$$\begin{aligned} -9x_1 + 2x_2 + 2x_3 &= 0 \\ 3x_1 - 8x_2 + 3x_3 &= 0 \\ 6x_1 + 6x_2 - 5x_3 &= 0 \end{aligned}$$

La solución a este sistema es  $x_1 = \lambda x_3$  y  $x_2 = \lambda x_3$  con  $x_3$  arbitrariamente, los eigenvectores corresponden a  $\lambda = 14$  es

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \lambda x_3 \\ \lambda x_3 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 1/3 \\ 1/2 \\ 1 \end{bmatrix}$$

Determinar los eigenvalores y eigenvectores de

$$A = \begin{bmatrix} 3 & 4 \\ -5 & -5 \end{bmatrix}$$

Para esta matriz

$$A - \lambda I = \begin{bmatrix} 3 & 4 \\ -5 & -5 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 - \lambda & 4 \\ -5 & -5 - \lambda \end{bmatrix}$$

Por lo tanto

$$\det(A - \lambda I) = (3 - \lambda)(-5 - \lambda) - 4(-5) = \lambda^2 + 2\lambda + 5$$

La ecuación característica de  $A$  es  $\lambda^2 + 2\lambda + 5 = 0$ , cuando resolviendo para  $\lambda$ , da dos eigenvalores complejos  $\lambda = -1 + i2$  y  $\lambda = -1 - i2$ . Los eigenvectores corresponden a  $\lambda = 1$  se obtienen resolviendo por  $X = [x_1, x_2]^T$  que es el valor de  $\lambda$ :

$$\begin{aligned} \left( \begin{bmatrix} 3 & 4 \\ -5 & -5 \end{bmatrix} - (-1 + i2) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \text{ó} \\ \begin{bmatrix} 4 - i2 & 4 \\ -5 & -4 - i2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{aligned}$$

que es equivalente a las ecuaciones lineales

$$\begin{aligned} (4 - i2)x_1 + 4x_2 &= 0 \\ -5x_1 + (-4 - i2)x_2 &= 0 \end{aligned}$$

La solución a este sistema es  $x_1 = (-4/5 - i2/5)x_2$  con  $x_2$  arbitrariamente, los eigenvectores corresponden a  $\lambda = -1 + i2$  son

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} (-4/5 - i2/5)x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -4/5 - i2/5 \\ 1 \end{bmatrix}$$

Con  $\lambda = -1 - i2$ , los eigenvalores correspondientes son encontrados en una manera similar

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} (-4/5 + i2/5)x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -4/5 + i2/5 \\ 1 \end{bmatrix}$$

### 7.7.7 Coeficientes de correlación de Pearson

En estadística, el coeficiente de correlación de Pearson es una medida de la relación lineal entre dos variables aleatorias cuantitativas. A diferencia de la covarianza, la correlación de Pearson es independiente de la escala de medida de las variables. En el caso de que se esté estudiando dos variables aleatorias  $x$  y  $y$  sobre una población; el coeficiente de correlación de Pearson se simboliza con la letra  $\rho_{x,y}$ , siendo la expresión que nos permite calcularlo:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sigma_x \sigma_y} \quad (52)$$

Donde:

- $\sigma_{xy}$  es la covarianza de (X,Y)
- $\sigma_x$  es la desviación típica de la variable X
- $\sigma_y$  es la desviación típica de la variable Y

#### *Ejemplo*

Con los datos sobre las temperaturas en dos días diferentes en una ciudad, determinar el tipo de correlación que existe entre ellas mediante el coeficiente de PEARSON.

X	18	17	15	16	14	12	9	15	16	14	16	18	$S_x = 180$
Y	13	15	14	13	9	10	8	13	12	13	10	8	$S_y = 138$

Se calcula la media para X

$$\bar{x} = \frac{\sum x_i}{n} = \frac{180}{12} = 15$$

Se calcula la media para Y

$$\bar{y} = \frac{\sum y_i}{n} = \frac{138}{12} = 11.5$$

Obtenemos los siguientes datos:

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$x^2$	xy	$y^2$
18	13	3	1.5	9	4.5	2.25
17	15	2	3.5	4	7	12.25
15	14	0	2.5	0	0	6.25
16	13	1	1.5	1	1.5	2.25
14	9	-1	-2.5	1	2.5	6.25
12	10	-3	-1.5	9	4.5	2.25
9	8	-6	-3.5	36	21	12.25
15	13	0	1.5	0	0	2.25
16	12	1	0.5	1	0.5	0.25
14	13	-1	1.5	1	-1.5	2.25
16	10	1	-1.5	1	-1.5	2.25
18	8	3	3.5	9	-10.5	12.25
$\Sigma = 180$	$\Sigma = 138$	-	-	$\Sigma = 72$	$\Sigma = 28$	$\Sigma = 63$

Para calcular la correlación:

$$r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}} = \frac{28}{\sqrt{(72)(63)}} = 0.416$$

Lo que quiere decir que existe una correlación moderada entre las variables X y Y



## 8 No correlación, Aproximación de $\ln(x)$ , Dato interesante: Napoleón no era bajito

### 8.1 No correlación

La no correlación es la situación en la que no se puede formar un patrón aparente al trazar puntos de datos para dos variables en un diagrama de dispersión. Muestra que no existe relación entre las dos variables. Un claro ejemplo de cosas que no están relacionadas, es el hecho de que los vuelos estén retrasados en un aeropuerto debido a que haya tráfico vehicular.

### 8.2 Aproximación de $\ln(x)$ por series de Taylor

Una serie de Taylor es una aproximación de funciones mediante una serie de potencias o suma de potencias enteras de polinomios como  $(x - a)^n$  llamados términos de la serie, dicha suma se calcula a partir de las derivadas de la función para un determinado valor o punto a suficientemente derivable sobre la función y un entorno sobre el cual converja la serie. Si esta serie está centrada sobre el punto cero,  $a = 0$ , se le denomina serie de McLaurin.

Esta aproximación tiene tres ventajas importantes:

La derivación e integración de una de estas series se puede realizar término a término, que resultan operaciones triviales;

Se puede utilizar para calcular valores aproximados de funciones;

Es posible calcular la optimidad de la aproximación.

Algunas funciones no se pueden escribir como serie de Taylor porque tienen alguna singularidad.

La serie de Taylor de una función  $f$  real o compleja  $f(x)$  infinitamente diferenciable en el entorno de un número real o complejo  $a$  es la siguiente serie de potencias:

$$f(a) + \frac{f'(a)}{1!}(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \frac{f^{(3)}(a)}{3!}(x - a)^3 + \dots$$

que puede ser escrito de una manera más compacta como la siguiente suma:

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x - a)^n,$$

donde:

$n!$  es el factorial de  $n$

$f^{(n)}(a)$  denota la  $n$ -ésima derivada de  $f$  para el valor  $a$  de la variable respecto de la cual se deriva.

Para el calculo de la serie de Taylor de la función  $\ln(x)$  centrada en  $a = 1$  se procede por calcular las derivadas y a evaluar las funciones en  $a = 1$ . A continuacion se observan

$$\begin{array}{l|l} f(x) = \ln(x) & f(1) = \ln(1) = 0 \\ f'(x) = \frac{1}{x} = x^{-1} & f'(1) = 1 \\ f''(x) = -x^{-2} & f''(1) = -1^{-2} = -1 \\ f'''(x) = 2x^{-3} & f'''(1) = 2(1)^{-3} = 2 \\ f^{(4)}(x) = -6x^{-4} & f^{(4)}(1) = -6(1)^{-4} = -6 \\ \vdots & \vdots \end{array}$$

Sustituyendo en la formula general se obtiene:

$$\begin{aligned} & f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \dots \\ & f(1) + \frac{f'(1)}{1!}(x-1) + \frac{f''(1)}{2!}(x-1)^2 + \frac{f^{(3)}(1)}{3!}(x-1)^3 + \dots \\ & 0 + \frac{1}{1}(x-1) + \frac{-1}{2}(x-1)^2 + \frac{2}{6}(x-1)^3 + \frac{-6}{24}(x-1)^4 + \dots \\ & (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \frac{(x-1)^4}{4} + \dots \end{aligned}$$

O en su forma de sumatoria queda :

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}(x-1)^n}{n}$$

Para comprobar que la aproximación es correcta se procede a realizar un script en R que realice las graficas las aproximaciones usando hasta 5 terminos. El script se muestra a continuación.

```
#serie taylor
x<-seq(0,5,.1)
y<-c(log(x))
y1<-c(x-1)
```

```

y2<-c((x-1) -((1/2)*(x-1)^2))
y3<- c((x-1) -((1/2)*(x-1)^2) +((1/3)*(x-1)^3) )
y4<- c((x-1) -((1/2)*(x-1)^2) +((1/3)*(x-1)^3) -((1/4)*(x-1)^4))
y5<- c((x-1) -((1/2)*(x-1)^2) +((1/3)*(x-1)^3) -((1/4)*(x-1)^4) +((1/5)*(x-1)^5))
plot(x,y,'l')
lines(x,y1,col='blue')
lines(x,y2,col='cyan')
lines(x,y3,col='tomato')
lines(x,y4,col='brown2')
lines(x,y5,col='blue')

```

Se observa que los primeros 2 componentes solo se aproximan al punto donde se centro la serie y que se requieren mas terminos para abarcar mas de la función a la que queremos aproximar.

### 8.3 Napoleón no era bajito

A la gente le encanta hacer generalizaciones: imaginaos, es una forma de saber la catadura de un grupo gigantesco de personas sin tener que tomarnos la molestia de conocerlos individualmente.

Por eso hubo tanta gente que consideraba a los negros intelectualmente inferiores. O a las mujeres. Incluso se usan personajes famosos para reafirmar posturas ideológicas: si Einstein era muy inteligente y ateo, los ateos son más inteligentes. E incluso para derribar esas mismas posturas: eres vegetariano? Pues Hitler también lo fue. Algo similar ocurre con las personas de corta estatura, aunque en este caso la falacia se sustenta, además, en varios equívocos, tanto científicos como históricos. El llamado 'complejo de Napoleón' describe a personas de corta estatura que compensan dicha falta mostrándose agresivos, escandalosos, egomaniacos, etc. Todos tenemos imágenes arquetípicas en la cabeza que confirman esta idea: Sarkozy, por ejemplo. O Tom Cruise, cuando se vuelve tarumba en plena entrevista y se pone a saltar en el sofá.

Sin embargo, no hay pruebas científicas sólidas que sostengan dicha teoría. No se ha reconocido oficialmente como trastorno psiquiátrico, y no parece ocurrir en el reino animal, salvo entre los machos de los peces espada, que el más pequeño iniciaba las peleas el 78 por ciento de las veces.

Por si esto fuera poco, Napoleón, icono del bajito furioso y chulito, ni siquiera era de corta estatura. Esta imagen fue creada, en parte, por el caricaturista británico James Gillray (1757-1815), inspirándose en Los viajes de Gulliver. En la imagen, el rey Jorge III sostiene a Napoleón en la palma

de su mano, mientras le inspecciona con una lupa. En 1821 se llevó a cabo una autopsia de Napoleón Bonaparte y se determinó que su estatura era de 1,69 metros. La estatura media de los varones franceses entre 1800 y 1820 era de 1,64 metros. Y la del inglés medio, 1,68. De modo que Napoleón era más alto que la media. El gran enemigo de Napoleón, Horatio Nelson, por ejemplo, solo medía 1,62 metros.

Lo que sí ocurrió es que Napoleón, tras subir al poder en 1799, impuso requerimientos de estatura al ejército francés. La Guardia Imperial, un mínimo de 1,78 metros. Y los Cazadores Montados, 1,70. Por lo tanto, la mayoría de ocasiones, los soldados que rodeaban a Napoleón eran significativamente más altos que él, por lo que podría dar la impresión de que él era bajito.

Eso no quita que diversas investigaciones sugieran que los bajitos tienen más dificultades para obtener reconocimiento social, tal y como explicamos en Si eres bajito, tendrás menos probabilidades de tener éxito.

## 9 Distribución Bose-Einsten, Maxwell-Boltzmann, Fermi Dirac

### 9.1 Distribución Bose-Einsten

Considerando un gas con un número fijo de partículas  $N$  con energías  $E_r$  que requiere el uso de estadísticas cuánticas para la descripción del mismo. Esto es, que sea un gas de partículas idénticas obedeciendo la estadística de Bose-Einstein (bosones)

$$N = \sum n_i, E_r = \sum n_i E_i \quad (53)$$

donde  $n_i$  corresponde al número de ocupación del estado asumiendo desde cero hasta infinito con energías  $E_1 < E_2 < \dots E_n$

Se utiliza la distribución de Gibbs o distribución macrocanónica. La descripción se hace a través de la función de macro partición del sistema (no confundir con  $Z$  la función de partición para moléculas en un baño térmico) y recordando que el potencial químico está definido como:

$$\mu = -T \frac{\delta S_2}{\delta N_0}$$

$$Z(T, V, \mu) = \sum r e^{-\beta(\mu N - E_{nr})}$$

además con la distribución de probabilidad normalizada (Gibbs)

$$P_r = \frac{e^{-\beta(\mu N - E_r)}}{Z} = p(n_1, n_2, \dots)$$

y sustituyendo ahora en la función de probabilidad

$$\prod_{i=1} \frac{e^{-\beta(\mu - E_i)n_i}}{Z_i} = \prod_{i=1} p_i(n_i) = p_{N_r}$$

lleva a que la distribución de probabilidad depende de un estado  $i$  solamente. Significa que la probabilidad de encontrar  $n_i$  partículas es independiente de los otros números de ocupación  $n_j$  donde  $j \neq i$ . Interpretando este resultado, se concluye que el estado de un sistema de partículas que no interactúan entre sí puede especificarse en términos de los estados individuales. La energía del sistema es la suma de las energías individuales y es posible considerar un estado a la vez.

Además, debe cumplirse que la probabilidad esté normalizada

$$\sum n_i p_i(n_i) = 1$$

para las estadísticas de Bose-Einstein, donde  $n_i = 0, 1, 2, \dots$  la función de macro partición se convierte en una serie geométrica que converge solo si el exponencial es menor que 1, esto ocurre cuando  $\mu < E$

$$Z_i = \frac{1}{1 - e^{-\beta(\mu - E_i)}}$$

ahora calculando el número medio de ocupación ( $n_i$ ) del  $i$ ésimo estado

$$n_i = \frac{1}{1 - e^{-\beta(\mu - E_i)}} = f_{be}(E)$$

distribución BE

generalizando para el número total de partículas en el sistema:

$$N = \sum_i n_i = \sum_i \frac{1}{1 - e^{-\beta(\mu - E_i)}}$$

para BE



## 9.2 Distribución Maxwell-Boltzmann

La estadística de Maxwell-Boltzmann es una función estadística desarrollada para modelar el comportamiento de sistemas físicos regidos por la mecánica clásica. Esta función estadística clásica, formulada originalmente por los físicos J.C. Maxwell y L. Boltzmann, rige la distribución de un conjunto de partículas en función de los posibles valores de energía de los estados que éstas pueden ocupar. Para cada sistema termodinámico, la distribución de Maxwell-Boltzmann no es otra cosa que la aplicación del colectivo canónico de la mecánica estadística, bajo el supuesto no-cuántico de que los números de ocupación de cada estado disponible son pequeños comparados con el número máximo de ocupación.

Esta función es una densidad de probabilidad cuya expresión es:

$$f(\epsilon_i) = A(N; T)e^{(-\epsilon_i)/kT}$$

O de forma más generalizada, puede expresarse como:

$$\frac{N_i}{N} = \frac{g_i}{e^{(\epsilon_i - \mu)/kT}} = \frac{g_i e^{-\epsilon_i/kT}}{Z}$$

En donde:

$A(N; T)$  : es una función dependiente de  $N$ , el número de partículas en el sistema y de  $T$ , la temperatura del sistema en Kelvin.

$N_i$  es el número de partículas en el estado  $i$ .

$\epsilon_i$  es la energía del estado  $i$ -ésimo.

$g_i$  es la degeneración del nivel de energía  $i$ , es decir, el número de estados (excluyendo el estado de partícula libre) con energía  $\epsilon_i$ .

$\mu$  es el potencial químico.

$k$  es la constante de Boltzmann.

$N$  es el número total de partículas:

$$N = \sum_i N_i$$

$Z$  es la función partición:

$$Z = \sum_i g_i e^{-\epsilon_i/kT}$$

$e$  es el número de Euler.

La distribución de Maxwell-Boltzmann se ha aplicado especialmente a la teoría cinética de gases, y otros sistemas físicos, además de en econofísica para predecir la distribución de la renta. En realidad la distribución de

Maxwell-Boltzmann es aplicable a cualquier sistema formado por  $N$  "partículas" o "individuos" que interaccionan estacionariamente entre sí una cierta magnitud  $M$  y cada uno de ellos tiene una cantidad  $m_i$  de la magnitud  $M$  y a lo largo del tiempo se cumple que  $M := m_1 + m_2 + \dots + m_N$ .

### 9.3 Distribución Fermi Dirac,

La distribución de Fermi-Dirac se aplica a los fermiones, partículas con espín semientero, que obedece el principio de exclusión de Pauli. Cada tipo de función de distribución tiene un término de normalización multiplicando el denominador del exponente, que puede ser dependiente de la temperatura. Para el caso de Fermi-Dirac, ese término se suele escribir:

$$e^{\frac{-E_f}{kT}} \quad (54)$$

Donde:

$E_f$  es la energía de Fermi

La importancia de la energía de Fermi se ve más claramente estableciendo  $T = 0$ . En el cero absoluto, la probabilidad es igual a 1 para energías menores que la energía de Fermi y cero para energías mayores que la energía de Fermi. Se puede imaginar todos los niveles hasta la energía de Fermi llenos, pero ninguna partícula tiene una energía mayor. Esto es totalmente coherente con el principio de exclusión de Pauli, donde cada estado cuántico no puede tener más que una sola partícula.

$$f(E) = \frac{1}{e^{(E-E_f)/k_B T} + 1} \quad (55)$$

Donde:

$f(E)$  es la probabilidad que una partícula tenga una energía  $E$ . Revisar la Distribución de Maxwell-Boltzmann para una discusión general del término exponencial.

En valor absoluto cero, los fermiones encontrarán toda la energía disponible sobre un nivel  $E_f$  llamada la energía de Fermi con una y única partícula. Están restringidos por el principio de exclusión de Pauli. A altas temperaturas algunos de ellos son elevados a niveles superiores del nivel de Fermi.

Para temperaturas bajas, éstos estados de energía que corresponden a  $E_f$  tienen una probabilidad de esencialmente 1 y estos debajo de la energía

de Fermi tienen energía de esencialmente zero.  
La diferencia cuántica de las arista partiendo que las partículas son indistinguibles.

$E$  es la energía en el estado  $i$ -ésimo

$E_F$  es el potencial químico

$T$  es la temperatura

$k_B$  la constante de Boltzmann

## 10 Funciones de Distribución

### 10.1 Función de distribución acumulada

La función de distribución acumulada  $F(x)$  de la variable aleatoria discreta  $X$ , cuya distribución de probabilidad es  $p(x)$ , es la probabilidad de que la variable  $X$  sea menor o igual al valor  $x$ . Esto es:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i) \quad (56)$$

#### Ejemplo

1. Divida el rango de la variable en subintervalos:  $x < 0$ ,  $0 \leq 1$ , y  $x \geq 1$ . Esta división es realizada de acuerdo a la partición de la recta real dada en el función de probabilidad.
2. Calcule la función de probabilidad acumulada para un valor  $x$  que se encuentre en el intervalo como la suma de las probabilidades de los valores de la variable menores a  $x$ .

ó  $\boxed{\text{para } x < 0}$

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i) = 0 \quad (57)$$

ó ya que según la definición de la función de probabilidad  $p(x) = 0$  cuando  $x \notin \{0, 1\}$  como es en este caso.

ó  $\boxed{\text{para } 0 \leq x \leq 1}$

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i) = P(X = 0) = 0.2 \quad (58)$$

ó  $\boxed{\text{para } x \geq 1}$

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i) = P(X = 0) + P(X = 1) = 1 \quad (59)$$

### 10.2 Distribución Beta

Es una distribución de probabilidad continua en el intervalo  $(0, 1)$ , Una variable aleatoria se dice que tiene distribución Beta con parámetros  $a > 0$

y  $b > 0$ , si su función de probabilidad es:

$$f(x) = \begin{cases} \frac{1}{\beta(a,b)} x^{a-1} (1-x)^{b-1} & \text{si } 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases} \quad (60)$$

Si  $a = 1$  y  $b = 1$ , entonces  $\beta(1, 1)$  es constante en  $(0, 1)$  por lo que es igual a la distribución uniforme. Se concluye entonces que  $Beta(1, 1) = Unif(0, 1)$ . La función  $\beta(a, b)$  está definida como:

$$\beta(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx \quad (61)$$

con  $a > 0$  y  $b > 0$ . La integral anterior es llamada una *integral beta* y también se relaciona con la función Gamma, dicha relación está expresada como:

$$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad (62)$$

Podemos sustituir la relación *beta-gamma* en la definición de la función de probabilidad, obteniendo lo siguiente:

$$f(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} & \text{si } 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases} \quad (63)$$

### Propiedades de la función Beta

1.  $\beta(a, b) = \beta(b, a)$
2.  $\beta(a, 1) = 1/a$
3.  $\beta(a+1, b) = \frac{a}{a+b} \beta(a, b)$

### Esperanza y Varianza

Por definición, la esperanza de una variable aleatoria  $X$ , está dada por:

$$E(X) = \int_0^1 x f(x) dx \quad (64)$$

Sustituyendo la función de probabilidad se puede obtener una expresión

simple

$$\begin{aligned}
 E(X) &= \int_0^1 x \frac{1}{\beta(a, b)} x^{a-1} (1-x)^{b-1} dx \\
 &= \frac{1}{\beta(a, b)} \int_0^1 x^{(a+1)-1} (1-x)^{b-1} dx \\
 &= \frac{\beta(a+1, b)}{\beta(a, b)}
 \end{aligned} \tag{65}$$

Usando la propiedad 3 de la función Beta:

$$\begin{aligned}
 E(X) &= \frac{\frac{a}{a+b} \beta(a, b)}{\beta(a, b)} \\
 E(X) &= \frac{a}{a+b}
 \end{aligned} \tag{66}$$

De la misma forma obtenemos la varianza, la cual está dada por:

$$\begin{aligned}
 V(X) &= E(X^2) - E^2(X) \\
 &= \frac{(a+1)a}{(a+1+b)(a+b)} - \left( \frac{a}{a+b} \right)^2
 \end{aligned} \tag{67}$$

$$V(X) = \frac{ab}{(a+b+1)(a+b)^2} \tag{68}$$

### 10.3 Distribución hipergeométrica

Suponga que una caja contiene  $b$  bolas blancas y  $r$  rojas, se efectúa  $n$  pruebas de un experimento en el cual se escoge una bola aleatoriamente, se observa su color y se regresa la bola a la caja. Este tipo de experimento se conoce como muestreo con remplazamiento. En tal caso  $X$  es la variable aleatoria para el número de bolas blancas escogidas (éxitos) en  $n$  pruebas, entonces empleando la distribución binomial vemos que la probabilidad de  $x$  éxitos es

$$P(X = x) = \begin{bmatrix} b \\ x \end{bmatrix} \frac{b^x r^{n-x}}{(b+r)^n}$$

ya que  $p = \frac{b}{(b+r)}$   $q = 1 - p = \frac{r}{(b+r)}$

Si modificamos el experimento anterior de modo que el muestreo se sin remplazamiento, es decir las bolas no regresan a la caja después de seleccionarse, entonces

$$P(X = x) = \frac{\begin{bmatrix} b \\ x \end{bmatrix} \begin{bmatrix} r \\ n-x \end{bmatrix}}{\begin{bmatrix} b+r \\ n \end{bmatrix}}$$



Esta es la distribución hipergeométrica. La media y la varianza son

$$\mu = \frac{nb}{b+r} \quad \sigma^2 = \frac{nbr(b+r-n)}{(b+r)^2(b+r-1)}$$

Si consideramos que el número total de bolas blancas y rojas es  $N$ , en tanto que las proporciones de bolas blancas y rojas son  $p$  y  $q = 1-p$  entonces:

$$p = \frac{b}{b+r} = \frac{n}{N} \quad q = \frac{r}{b+r} = \frac{r}{N} \quad ob = Np \quad r = Nq$$

de modo que la probabilidad y la media se convierten en:

$$P(X=x) = \frac{\binom{Np}{x} \binom{Nq}{n-x}}{\binom{N}{n}} \quad \mu = np \quad \sigma^2 = \frac{npq(N-n)}{N-1}$$

Observe que cuando  $N \rightarrow \infty$  o  $N$  es grande comparada con  $n$  la probabilidad se reduce a:

$$P(X=x) = \binom{n}{x} p^x q^{n-x} \quad \mu = np \quad \sigma^2 = npq$$

## 10.4 Distribución de Cauchy

La distribución Cauchy-Lorentz es una distribución de probabilidad continua.

En estadística la distribución de Cauchy (a veces también distribución de Lorentz) es una distribución de probabilidad continua cuya función de densidad es:

$$\begin{aligned} f(x; x_0, \gamma) &= \frac{1}{\pi\gamma \left[ 1 + \left( \frac{x-x_0}{\gamma} \right)^2 \right]} \\ &= \frac{1}{\pi} \left[ \frac{\gamma}{(x-x_0)^2 + \gamma^2} \right] \end{aligned} \quad (69)$$

Donde  $x_0$  es el parámetro de corrimiento que especifica la ubicación del pico de la distribución, y  $y$  es el parámetro de escala que especifica el ancho medio al máximo medio.

En el caso especial donde  $x_0 = 0$  y  $y_0 = 1$  es denominado la distribución estándar Cauchy con la función de densidad de probabilidad.

$$f(x; 0, 1) = \frac{1}{\pi(1 + x^2)}$$

En general la distribución de Cauchy no tiene valor esperado ni varianza.

Sean  $U$  y  $V$  dos variables aleatorias uniformes dentro  $-1$  y  $1$  y  $U^2 + V^2 < 1$ , entonces el número  $U/V$  tiene la distribución Cauchy.

#### 10.4.1 Función de distribución

$$F(x; x_0, \gamma) = \frac{1}{\pi} \arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}$$

y la función inversa de distribución acumulativa para la distribución Cauchy es:

$$F^{-1}(p; x_0, \gamma) = x_0 + \gamma \tan\left[\pi\left(p - \frac{1}{2}\right)\right]$$

#### 10.4.2 Aplicaciones

En el estudio de las partículas. Si  $Z$  es un ángulo aleatorio distribuido uniformemente entre  $-\frac{\pi}{2}$  y  $\frac{\pi}{2}$ ,  $\tan(Z)$  tiene distribución de Cauchy. El cociente de dos variables aleatorias normales estándar independientes tiene también distribución de Cauchy

### 10.5 Distribución Chi-Cuadrado

Sean  $X_1, X_2, \dots, X_v$   $v$  variables aleatorias independientes distribuidas normalmente con media cero y varianza 1. Considérese la variable aleatoria

$$x^2 = X_1^2 + X_2^2 + \dots + X_v^2 \quad (70)$$

donde  $X^2$  se llama chi-cuadrado. Entonces podemos demostrar que para  $x \geq 0$

$$P(X^2 \leq x) = \frac{1}{2^{\frac{v}{2}} \Gamma(\frac{v}{2})} \int_0^x u^{\frac{(v)}{2}-1} e^{-\frac{u}{2}} du \quad (71)$$

y  $P(X^2 \leq x) = 0$  para  $x < 0$ .

La distribución definida anteriormente es la *distribución chi-cuadrado* y  $v$  es el número de grados de libertad. La distribución definida tiene la función de densidad dada por:

$$f(x) = \begin{cases} \frac{1}{2^{\frac{v}{2}} \Gamma(\frac{v}{2})} x^{\frac{v}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Se observa por tanto que la *distribución chi-cuadrado* es un caso especial de la distribución gamma con  $\alpha = \frac{v}{2}$ ,  $\beta = 2$ . Por tanto:

$$\mu = v, \sigma^2 = 2v, M(t) = (1 - 2t)^{-\frac{v}{2}}, \phi(\omega) = (1 - 2i\omega)^{-\frac{v}{2}} \quad (72)$$

Para  $v$  grande ( $v \geq 30$ ) podemos demostrar que  $\sqrt{2x^2} - \sqrt{2v-1}$  está casi distribuida normalmente con media 0 y varianza 1.

Las consideraciones anteriores se sintetizan en el siguiente teorema.

Sean  $X_1, X_2, \dots, X_n$  variables aleatorias independientes normalmente distribuidas con media 0 y varianza 1. Entonces  $X^2 = X_1^2 + X_2^2 + \dots + X_v^2$  tiene distribución chi-cuadrado con  $v$  grados de libertad.

Otros dos teoremas que son útiles en el estudio posterior son:

- Sean  $U_1, U_2, \dots, U_k$  variables aleatorias independientes con distribución chi-cuadrado y  $v_1, v_2, \dots, v_k$  grados de libertad respectivamente. Entonces su suma  $W = U_1 + U_2 + \dots + U_k$  tiene distribución chi-cuadrado con  $v_1 + v_2 + \dots + v_k$  grados de libertad.
- Sean  $V_1$  y  $V_2$  variables aleatorias independientes. Si  $V_1$  tiene distribución chi-cuadrado con  $v_1$  grados de libertad mientras que  $V = V_1 + V_2$  tiene distribución chi cuadrado con  $v$  grados de libertad, donde  $v > v_1$ . Entonces  $V_2$  tiene distribución chi-cuadrado con  $v - v_1$  grados de libertad.

## 10.6 La función Gamma

Para definir la familia de distribuciones gamma, primero se debe de introducir una función que desempeña un importante papel en muchas áreas de las matemáticas:

La función gamma  $\Gamma(\alpha)$  para un  $\alpha > 0$  se define como

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad (73)$$

Las propiedades más importantes de la función gamma son las siguientes:

1. Con cualquier  $\alpha > 1$ ,  $\Gamma(\alpha) = (\alpha - 1)$
2. Con cualquier entero positivo,  $n$   $\Gamma(n) = (n - 1)!$
3.  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

De acuerdo con 73 si:

$$f(x, \alpha) = \begin{cases} x^{\alpha-1}e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (74)$$

Entonces  $f(x, \alpha)$  y  $\int_0^\infty f(x; \alpha)dx = \Gamma(\alpha)/\Gamma(\alpha) = 1$ , así que  $f(x, \alpha)$  satisface las dos propiedades básicas de una función de densidad de probabilidad.

### 10.6.1 Distribución Gamma

Una variable aleatoria tiene la distribución gamma si la función de densidad es

$$f(x) = \begin{cases} \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (\alpha, \beta > 0) \quad (75)$$

donde  $\Gamma(\alpha)$  es la función gamma. La distribución gamma estándar tiene  $\beta = 1$ , así que . La media y la varianza están dadas por

$$\mu = \alpha\beta \quad \sigma^2 = \alpha\beta^2)$$

La función matriz de momentos y la función característica están dadas respectivamente por

$$M(t) = (1 - \beta t)^{-\alpha} \quad \phi(\omega) = (1 - \beta i\omega)^{-\alpha}$$

La distribución exponencial se deriva de considerar  $\alpha = 1$  y  $\beta = 1/\lambda$ . La figura ?? ilustra las gráficas de la función de densidad de probabilidad gamma  $f(x, \alpha, \beta)$  75 para varios pares  $(\alpha, \beta)$  en tanto que en la figura ?? presenta gráficas de la función de densidad de probabilidad gamma estándar. Para la función de densidad de probabilidad estándar cuando  $\alpha \leq 1$ ,  $f(x; \alpha)$  es estrictamente decreciente a medida que  $x$  se incrementa desde 0; cuando  $\alpha > 1$ ,  $f(x, \alpha)$  se eleva desde 0 en  $x = 0$  hasta un máximo y luego decrece. El parámetro  $\beta$  en 75 se llama *parámetro de escala* porque los valores diferentes de uno alargan o comprimen la función de densidad de probabilidad en la dirección  $x$

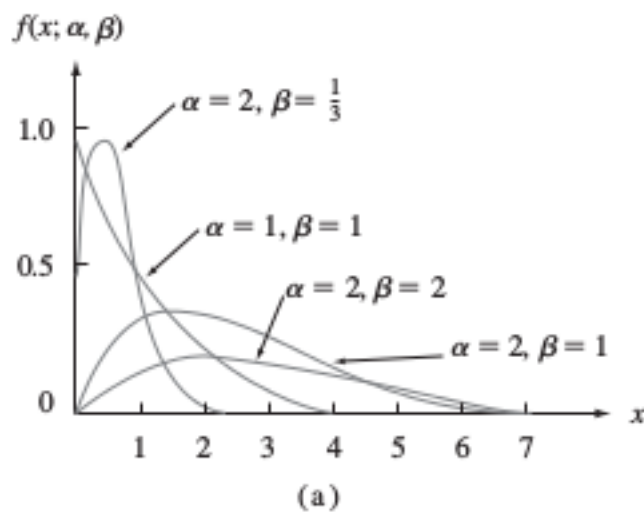


Figure 5: Curvas de densidad gamma

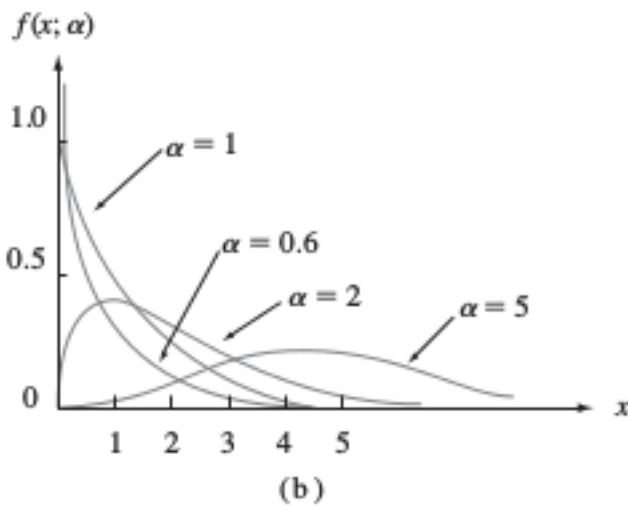


Figure 6: Curvas de densidad estandar

### 10.6.2 Ejemplo

Suponga que el tiempo de reacción  $X$  de un individuo seleccionado al azar de un estímulo tiene una distribución gamma estandar con  $\alpha = 2$ . Como:

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

cuando  $X$  es continua,

$$P(3 \leq X \leq 5) = F(5; 2) - F(3; 2) = 0.960 - 0.801 = 0.159$$

La probabilidad de que el tiempo de reacción sea de mas de 4 s es:

$$P(X > 4) = 1 - P(X < 4) = 1 - F(4; 2) = 1 - 0.908 = 0.092$$

## 10.7 Ley de Benford

También conocida como la ley del primer dígito, asegura que, en los números que existen en la vida real, la primera cifra es 1 con mucha más frecuencia que el resto de los números. Además, según crece este primer dígito, más improbable es que se encuentre en la primera posición. Esta ley se puede aplicar a hechos relacionados con el mundo natural o con elementos sociales.

Por ejemplo; se puede utilizar para encontrar cifras falsificadas en declaraciones fiscales o para corroborar modelos demográficos.

Quien primero se dio cuenta de este fenómeno fue en 1881 el matemático y astrónomo Simon Newcomb. Un día, Newcomb estaba usando un libro de logaritmos y se dio cuenta de que las páginas del libro estaban más viejas y usadas cuanto más cercanas estaban del principio.

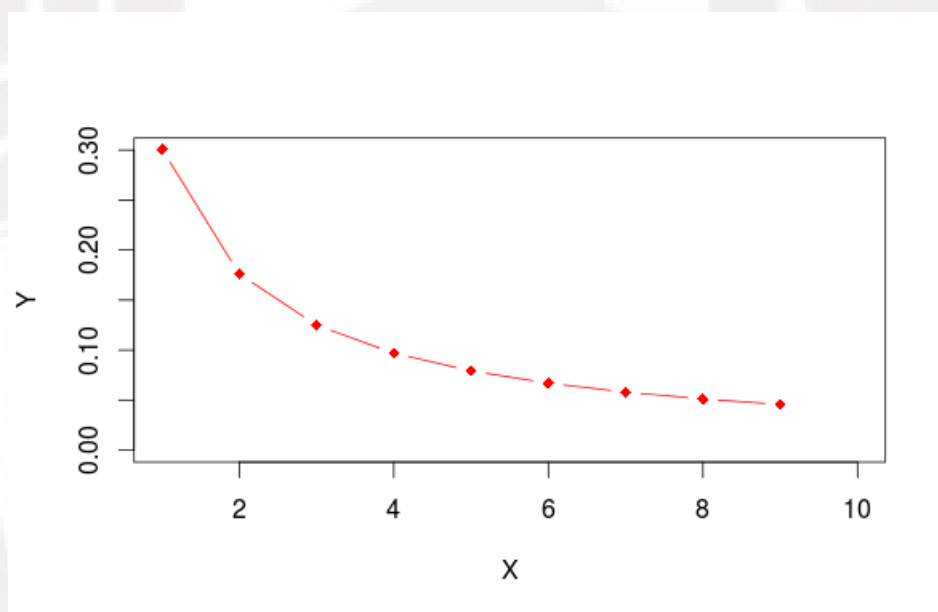
Dedujo que los dígitos iniciales de los números (al menos los utilizados en su trabajo que provenían de la observación de los astros principalmente) no son equiprobables sino que el 1 aparece como dígito inicial más frecuente seguido del 2 etc. Hasta el 9 que es el menos frecuente. Mediante un breve e ingenioso razonamiento, aunque sin presentar realmente un argumento formal ni fórmula matemática, Newcomb enunció verbalmente una relación o ley logarítmica: la ley de probabilidad de ocurrencia de números es tal que las mantisas de sus logaritmos son equiprobables.

El asunto fue rápidamente olvidado hasta 1938, cuando Frank Benford, un

físico de la compañía General Electric, se dio cuenta del mismo patrón. Entusiasmado por el descubrimiento, estudió 20,229 números provenientes de 20 muestras de todo tipo: constantes y magnitudes físicas, longitudes de ríos, estadísticas de béisbol, direcciones de personas... incluso cifras sacadas de portadas de revistas.

A partir de los datos extraídos del mundo real, comprobó que la probabilidad de que un número en una serie de datos comience por el dígito  $d$  es de  $P[d] = \log(1 + 1/d)$  y postuló la llamada "ley de los números anómalos de Benford".

Según dicha ley la probabilidad de que en una serie de muchos datos el primer dígito de un número sea 1 es del 30 por ciento, 17,6 por ciento para un 2, así va decreciendo...



## 10.8 Distribución Exponencial

En estadística la distribución exponencial es una distribución de probabilidad continua con un parámetro  $\lambda > 0$  cuya función  $n$  de densidad es:

$$f(x) = P(x) = \begin{cases} \lambda e^{-\lambda x} & \text{para } x \geq 0 \\ 0 & \text{en caso contrario} \end{cases} \quad (76)$$

ó Su función de distribución acumulada es:



$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{para } x < 0 \\ 1 - e^{-\lambda x} & \text{para } x \geq 0 \end{cases} \quad (77)$$

ó

Donde  $e$  representa el número  $e$ . ó El valor esperado y la varianza de una variable aleatoria  $X$  con distribución exponencial son:

$$E[X] = \frac{1}{\lambda}, \quad V(X) = \frac{1}{\lambda^2} \quad (78)$$

ó La distribución exponencial es un caso particular de distribución gamma con  $k = 1$ . Además la suma de variables aleatorias que siguen una misma distribución exponencial es una variable aleatoria expresable en términos de la distribución gamma.

ó

### 10.8.1 Ejemplos

- La distribución de la longitud de los intervalos de una variable continua que transcurren entre dos sucesos, que se distribuyen según la distribución de Poisson.
- El tiempo transcurrido en un call center hasta recibir la primer llamada del día se podría modelar como una exponencial.
- El intervalo de tiempo entre terremotos (de una determinada magnitud) sigue una distribución exponencial.
- Supongamos una máquina que produce hilo de alambre, la cantidad de metros de alambre hasta encontrar una falla en el alambre se podría modelar como una exponencial.
- En fiabilidad de sistemas, un dispositivo con tasa de fallo constante sigue una distribución exponencial.

### 10.8.2 Cálculo de variables aleatorias

Se pueden calcular una variable aleatoria de distribución exponencial  $x$  por medio de una variable aleatoria de distribución uniforme  $u = U(0, 1)$ :

$$x = -\frac{1}{\lambda} \ln(1 - u) \quad (79)$$

ó o, dado que  $(1 - u)$  es también una variable aleatoria con distribución  $U(0, 1)$  puede utilizarse la versión más eficiente:

$$x = -\frac{1}{\lambda} \ln(u) \quad (80)$$

## 11 Suma geométrica

En matemática, una serie geométrica es una serie en la cual la razón entre los términos sucesivos de la serie permanece constante.

Por ejemplo la serie:

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots + \frac{1}{2^n} \quad (81)$$

es geométrica, pues cada término sucesivo se obtiene al multiplicar el anterior por  $\frac{1}{2}$ .

### 11.1 Razón común

Los términos de una serie geométrica forman una progresión geométrica, es decir que la razón entre términos sucesivos permanece constante.

El comportamiento de los términos depende de la razón común  $r$ :

- Si  $|r| < 1$  los términos decrecen y se acercan a cero en el límite. En tal caso, la serie converge.
- Si  $|r| > 1$  los términos de la serie se incrementan en magnitud. La suma de los términos también aumenta y la serie no tiene suma. La serie diverge.

### 11.2 Suma

La suma de una serie geométrica será finita siempre y cuando los términos se aproximen a cero; a medida que se acercan al cero, las cantidades se vuelven insignificamente pequeñas, permitiendo calcular la suma sin importar el hecho que la serie sea infinita. La suma puede ser obtenida utilizando las propiedades autosimilares de la serie.

### 11.3 Fórmula

Para  $r \neq 1$ , la suma de los primeros  $n$  términos de una serie geométrica es:

$$a + ar + ar^2 + ar^3 + \cdots + ar^{n-1} = \sum_{k=0}^{n-1} ar^k = a \frac{1 - r^n}{1 - r} \quad (82)$$

donde  $a$  es el primer término de la serie y  $r$  la razón común.

## 11.4 Demostración

Sea  $s = a + ar + ar^2 + ar^3 + \cdots + ar^{n-1}$ .

Entonces  $rs = ar + ar^2 + ar^3 + ar^4 + \cdots + ar^n$ .

Entonces  $s - rs = s(1 - r) = a - ar^n$ , luego  $s = a \frac{1-r^n}{1-r}$ .

Ejemplo: Dada la suma de la serie geométrica:

$$s = 1 + \frac{2}{3} + \frac{4}{9} + \frac{8}{27} + \cdots \quad (83)$$

La razón común de esta serie es  $2/3$ . Multiplicando por  $2/3$  cada término, se obtiene:

$$\frac{2}{3}s = \frac{2}{3} + \frac{4}{9} + \frac{8}{27} + \frac{16}{81} + \cdots \quad (84)$$

Esta nueva serie es como la original excepto por el primer término que falta. Restándolas, se obtiene:

$$s - \frac{2}{3}s = 1 \quad (85)$$

, por lo que  $s = 3$ . Una técnica similar puede utilizarse al evaluar cualquier expresión autosimilar.

## 11.5 Convergencia

La serie geométrica real de término inicial  $a \in \mathfrak{R}$  no nulo y de razón  $r \in \mathfrak{R}$  es convergente si y solamente si  $|r| < 1$ . En tal caso, su suma vale:

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r} \quad (86)$$

## 11.6 Ejemplos

Secuencia geométrica finita:

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots + \frac{1}{32768} \quad (87)$$

Escrita en notación sumatoria:

$$\sum_{k=1}^{15} \frac{1}{2^k} \quad (88)$$

## 11.7 Secuencia infinita

Secuencia geométrica infinita:

$$2 + 6 + 18 + 54 + \cdots \quad (89)$$

Escrita en notación sumatoria:

$$\sum_{n=1}^{\infty} (2 * 3^{n-1}) \quad (90)$$



## 12 Anexos

### 12.1 Problema: The last sheep

# Probabilidad, procesos aleatorios e inferencia

Edgar Daniel Garcia Serrano

December 3, 2015

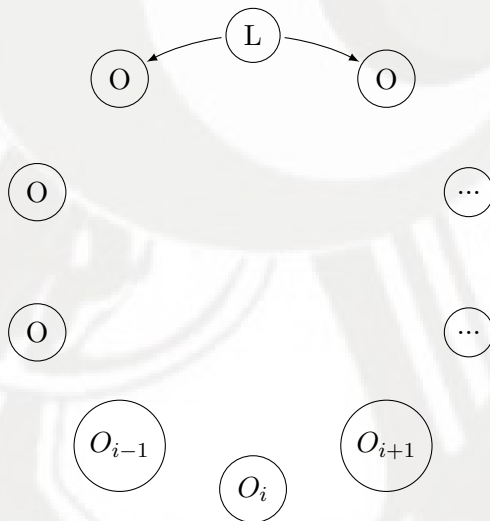
## La última oveja de pie

### Problema

Hay 100 puntos igualmente espaciados alrededor de un círculo. En 99 de los puntos, hay ovejas, y en 1 punto, hay un lobo. En cada paso, el lobo se mueve aleatoriamente ya sea en sentido de la manecillas del reloj, o en contra, sólo se mueve un punto a la vez. Si hay una oveja en ese punto, el lobo se la come. Las ovejas no se mueven. ¿Cuál es la probabilidad que la oveja que esta inicialmente al lado opuesto al lobo sea la última que quede?

### Solución

Nombremos a la oveja que esta inicialmente al lado opuesto del lobo como "i". Si  $i$  es la última oveja que queda, quiere decir que las vecinas de  $i$  son comidas antes que  $i$ . Sin embargo, para que lo anterior se cumpla, el lobo tuvo que comer a una de la vecinas de  $i$ , y luego ir todo el camino para comer al siguiente vecino de  $i$ .



Consideramos entonces el momento justo cuando el lobo se ha comido a la primer vecina de  $i$ . La pregunta entonces se vuelve **¿Cuál es la probabilidad que el lobo se coma a la siguiente vecina de  $i$  antes que a  $i$ ?** Esto se convierte entonces en un problema similar a "*La ruina del jugador (Gambler's ruin)*"

Podemos desplegar el círculo formando una línea recta, y tomando como referencia a la oveja  $i$ , la nombramos como "0", el lobo, que está adyacente a  $i$ , será la posición "1" y la segunda vecina de  $i$  será  $n$ . [1]





Probabilidad de avance  $(+1) \rightarrow p$

Probabilidad de retroceso  $(-1) \rightarrow q = 1 - p$

$P_0 = 0 \rightarrow$  Probabilidad que se coma a  $n$  antes que 0

$P_n = 1 \rightarrow$  Probabilidad que se coma a  $n$  antes que 0

$$P_i = pP_{i+1} + qP_{i-1} \quad (1)$$

$P_i$  es la probabilidad que el lobo se coma a  $n$  antes que 0

Ya que  $p + q = 1$ , entonces:

$$\begin{aligned} (p + q)P_i &= pP_{i+1} + qP_{i-1} \\ pP_i + qP_i &= pP_{i+1} + qP_{i-1} \\ P_{i+1} - P_i &= \frac{q}{p}(P_i - P_{i-1}) \end{aligned} \quad (2)$$

Para el caso  $i = 1$  (Cuando el lobo comienza en 1)

$$P_2 - P_1 = \frac{q}{p}(P_1 - P_0) = \frac{q}{p}P_1$$

Para el caso  $i=2$

$$P_3 - P_2 = \frac{q}{p}(P_2 - P_1) = \left(\frac{q}{p}\right)^2 P_1$$

Para el caso general  $i$

$$P_{i+1} - P_i = \left(\frac{q}{p}\right)^i P_1$$

Sumando todas las ecuaciones(desde 1 hasta  $i$ )

$$P_{i+1} - P_1 = \sum_{k=1}^i (P_{k+1} - P_k) = \sum_{k=1}^i \left(\frac{q}{p}\right)^k P_1$$

Despejando  $P_{i+1}$

$$P_{i+1} = P_1 + P_1 \sum_{k=1}^i \left(\frac{q}{p}\right)^k = P_1 \sum_{k=0}^i \left(\frac{q}{p}\right)^k \quad (3)$$

$$P_{i+1} = \begin{cases} P_1 \left( \frac{1 - \left(\frac{q}{p}\right)^{i+1}}{1 - \left(\frac{q}{p}\right)} \right) & \text{si } p \neq q \\ P_1(i + 1) & \text{si } p = q \end{cases} \quad (4)$$

Suponiendo que  $i = n - 1$

$$P_{i+1} = \begin{cases} P_1 \left( \frac{1 - \left(\frac{q}{p}\right)^n}{1 - \left(\frac{q}{p}\right)} \right) & \text{si } p \neq q \\ P_1(n) & \text{si } p = q \end{cases}$$

Y ya que  $P_n=1$

$$\begin{aligned} P_n &= P_i n \\ 1 &= P_1 n \\ P_1 &= \frac{1}{n} \end{aligned} \quad (5)$$


Para un caso general en el que el lobo comienza en la posición  $i$

$$P_{i+1} = \begin{cases} P_1 \left( \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^n} \right) & \text{si } p \neq q \\ \frac{i}{n} & \text{si } p = q \end{cases} \quad (6)$$

## References

- [1] Columbus University. <http://www.columbia.edu>, 21 de Enero de 2008.

## 12.2 Exposición: Redes Neuronales



¿Qué es una Red Neuronal?  
Perceptrón  
Ventajas  
Desventajas  
Aplicaciones

# Redes Neuronales

Alexa Chávez Álvarez

18 de enero de 2016

## Table of Contents

- 1 ¿Qué es una Red Neuronal?
  - Neurona biológica/artificial
- 2 Perceptrón
- 3 Ventajas
- 4 Desventajas
- 5 Aplicaciones

## ¿Qué es una Red Neuronal?

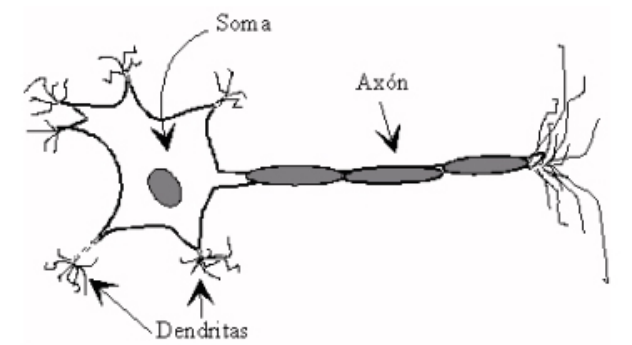
- Procesamiento inspirado en el sistema nervioso
- Sistema de interconexión de neuronas que colaboran
- Modelo computacional capaz de aprender, generalizar u organizar información
- Permite buscar la combinación de parámetros que mejor se ajusta a un determinado problema

¿Qué es una Red Neuronal?  
Perceptrón  
Ventajas  
Desventajas  
Aplicaciones

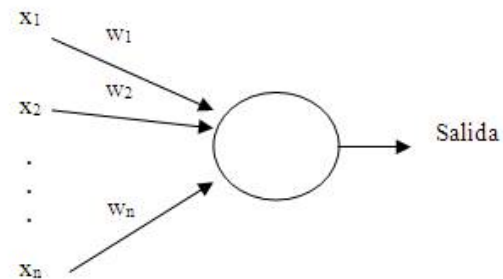
Neurona biológica/artificial

## Neurona biológica/artificial

### Neurona biológica



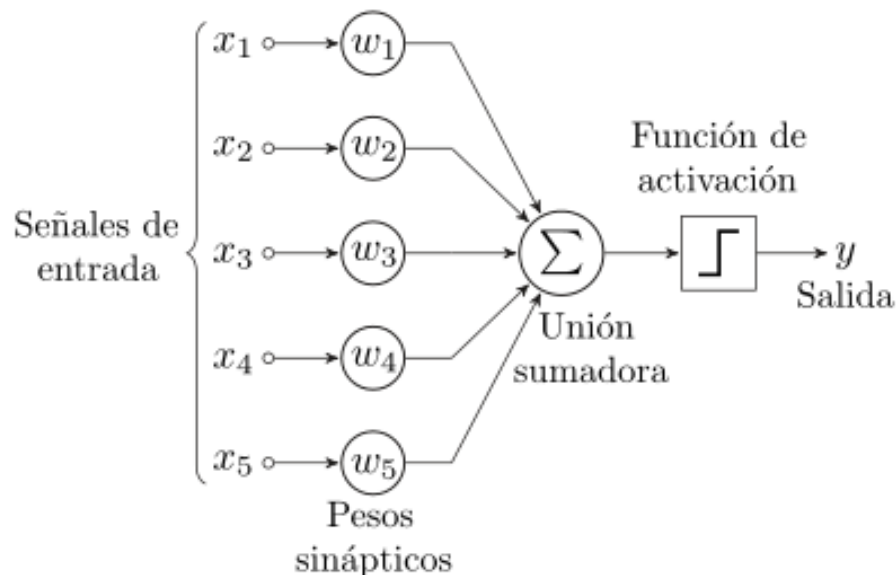
### Neurona artificial





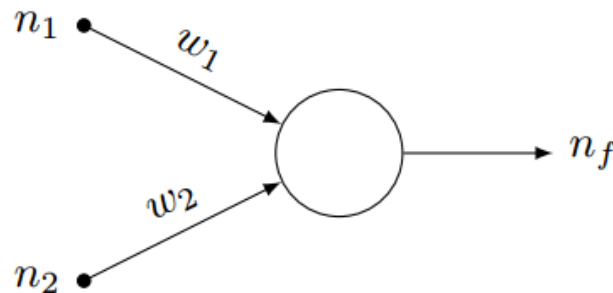
## Perceptrón

La suma ponderada de de las señales de entrada, es comparada con un umbral para determinar la salida de la neurona. Cuando es igual o mayor al umbral, la salida es 1, cuando es menor que el umbral, la salida es 0.



## Ejemplo:

Supongamos que los alumnos de una clase en la que el profesor no ha dicho exactamente cómo va a poner las notas. Sólo han hecho 2 exámenes y se tiene la nota de cada uno de ellos y la final:



Entradas:  $n_1$   $n_2$        $n_f$  : 1 si aprueba, 0 si reprueba

## Ejemplo:

Entonces tenemos que:

$$\text{Si } n_f = 1 \rightarrow (n_1 * w_1) + (n_2 * w_2) > 5 \quad \text{Aprobado} \quad (1)$$

$$\text{Si } n_f = 0 \rightarrow (n_1 * w_1) + (n_2 * w_2) \leq 5 \quad \text{Reprobado} \quad (2)$$

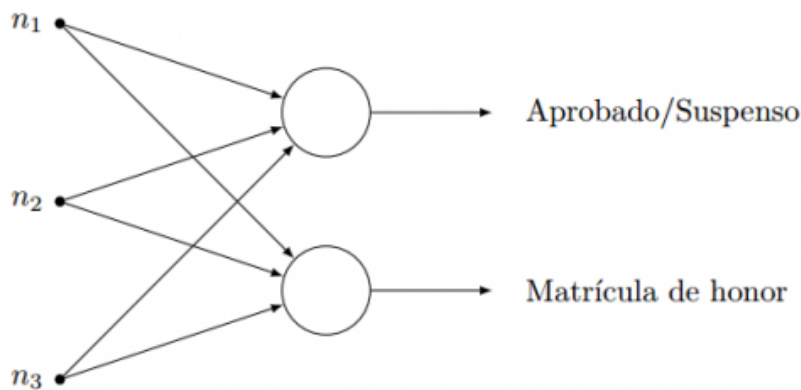
## Ejemplo: Para el entrenamiento

Elegimos 2 pesos aleatorios, por ejemplo  $w_1 = 0.5$  y  $w_2 = 0.5$ , el mismo peso a cada examen y ver qué resultado da para cada alumno, si falla, se ajustarán los pesos para cada alumno.

Ej. Si el segundo examen fue de repaso, y se le asigna un puntaje bajo, el peso  $w_2$  reducirá.

## Ejemplo: Múltiples salidas

Quizás queramos complicarlo más, poniendo más exámenes (más nodos de entrada) o queriendo sacar más resultados.  
Ejemplo: Mostrar quién tiene un buen promedio.



## Ejemplo: Perceptrón multicapa

El ejemplo anterior funciona es demasiado simple. Añadimos un trabajo que había que entregar (T).

2 alumnos  $A_1$  y  $A_2$  que tienen ambos,  $n_1 = n_2 = 10$

$$A_1 \rightarrow T = 4 \quad \text{Aprobado} \quad (3)$$

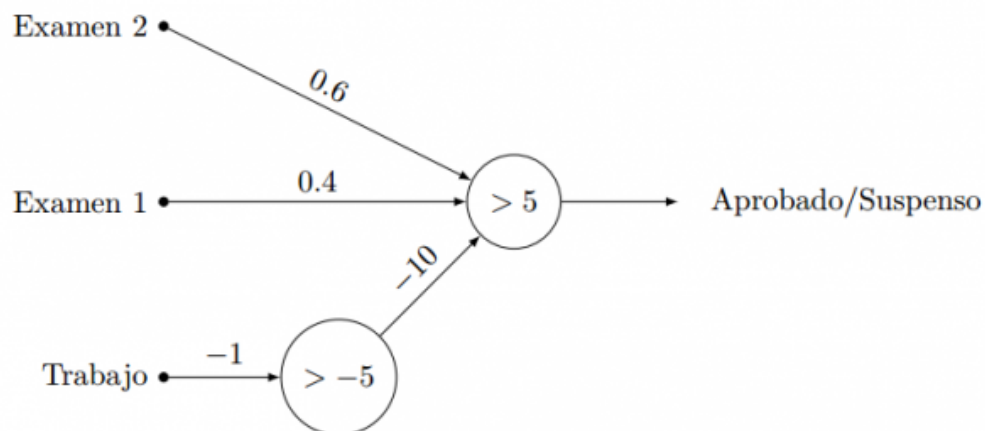
$$A_2 \rightarrow T = 7 \quad \text{Reprobado} \quad (4)$$

Ahora, un tercer alumno  $A_3$

$$A_2 : n_1 = n_2 = 4.99 \text{ pero } T = 10 \quad \text{Reprobado} \quad (5)$$

## Ejemplo: Perceptrón multicapa

Con las capas, añadimos información que antes no existía





## Ejemplo: Perceptrón multicapa

Un psicólogo afirma tener una red neuronal capaz de predecir si un matrimonio fracasará o triunfará, con base en los siguientes parámetros:

$x_1$  = Tamaño del hogar de la pareja expresado en  $m^2$

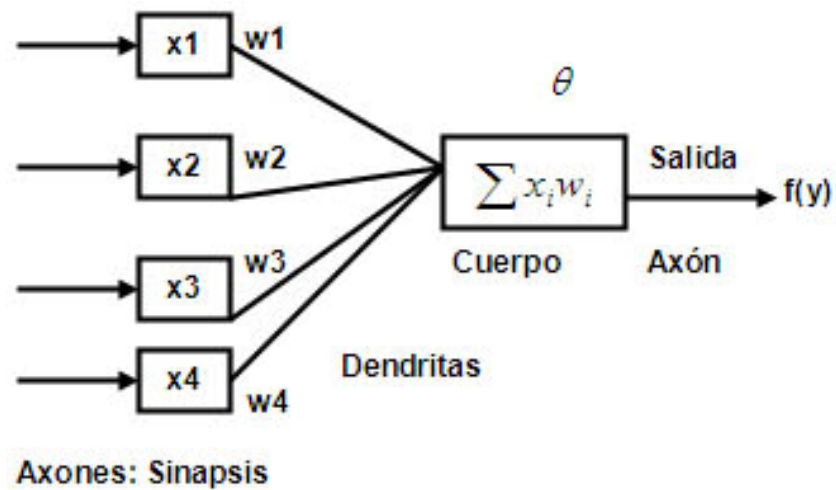
$x_2$  = Cantidad de hijos

$x_3$  = Edad del Hombre

$x_4$  = Edad de la Mujer

¿Qué es una Red Neuronal?  
Perceptrón  
Ventajas  
Desventajas  
Aplicaciones

## Ejemplo: Perceptrón multicapa



## Ejemplo: Perceptrón multicapa

Se cuentan con 50 datos válidos, es decir, conocemos su resultado:

$m^2(x_1)$	$Hijos(x_2)$	$Edad\ El(x_3)$	$Edad\ Ella(x_4)$	$Divorcio(y)$
100	1	37	32	1
80	2	32	20	0
70	0	23	22	1
150	1	29	32	1
60	1	25	20	0
...	...	...	...	...

## Ejemplo: Perceptrón multicapa

Entrenando una RNA:

- Los datos presentados, deberán ser los más representativos
- Se define un rango de salida, o umbral (función de activación)
- Se presentan 1 a 1 los datos a la RN
- Se ajustan los pesos de modo que y sea satisfactoria
- Se corrobora el ajuste de pesos presentando nuevamente los datos
- Test a la RN, con datos diferentes válidos

## ¿Por qué una Red Neuronal?

- Aprendizaje adaptativo: capacidad de aprender con entrenamiento
- Auto-organización: en su etapa de aprendizaje
- Tolerancia a fallos: si se destruye parte de la red, retiene sus capacidades
- Operación en tiempo real: implementación paralela

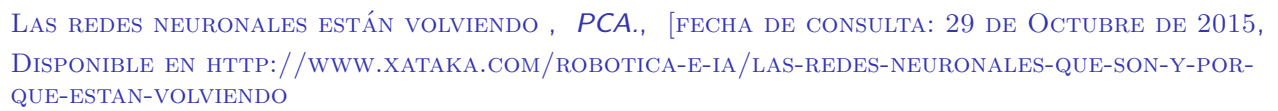
## ¿Por qué NO una Red Neuronal?

- Entrenamiento requiere mucho tiempo en el proceso
- Falta de reglas definitorias
- Muchos factores, poco tiempo
- Red sobreentrenada no sirve

## Actualmente una Red Neuronal...

- Obtención de modelo de la retina
- Analiza tendencias y patrones
- Diagnóstico y tratamiento
- Identificación de falsificaciones





## 13 Conclusiones

La probabilidad mide la frecuencia con la que se obtiene un resultado (o conjunto de resultados) al llevar a cabo un experimento aleatorio, del que se conocen todos los resultados posibles, bajo condiciones suficientemente estables. La teoría de la probabilidad se usa extensamente en áreas como la estadística, la física, la matemática, la ciencia y la filosofía para sacar conclusiones sobre la probabilidad de sucesos potenciales y la mecánica subyacente de sistemas complejos.

La probabilidad constituye un importante parámetro en la determinación de las diversas casualidades obtenidas tras una serie de eventos esperados dentro de un rango estadístico.

La probabilidad de un evento se denota con la letra  $p$  y se expresa en términos de una fracción y no en porcentajes, por lo que el valor de  $p$  cae entre 0 y 1. Por otra parte, la probabilidad de que un evento "no ocurra" equivale a 1 menos el valor de  $p$  y se denota con la letra  $q$ .

La probabilidad es de suma importancia en el momento de llevar a cabo cualquier investigación, debido a que arroja los datos necesarios para describir el comportamiento de los eventos involucrados en el trabajo y permite darle un grado de confianza a los resultados obtenidos. Además, el cálculo de probabilidad tiene como función describir, analizar y predecir.

Nos encontramos en un mundo altamente probabilístico, por lo tanto se tiene la necesidad de obtener un conocimiento de probabilidad claro y conciso, ya que es una herramienta de gran apoyo en cualquier situación científica. Además, actualmente la probabilidad se ha vuelto tan indispensable que múltiples campos como la Economía, Biología, Psicología, Ingeniería, entre otras, basan sus proyectos y resolución de problemas en ésta rama básica.

## References

- [1] Eliana Bottarelli. Probabilidad clásica, empírica y subjetiva, sep 2015.
- [2] SALVADOR ANAYA DEBERNARD. *CARRUSEL MATEMATICO*. LIMUSA, 2000.
- [3] R. Alu Srinivasan Murray R. Spiegel, John J. Schiller. *Schaum's Outline of Probability and Statistics*. McGraw-Hill: New York, Chicago, San Francisco, Lisbon, London, Madrid, Mexico City, Milan, New Delhi, San Juan, Seoul, Singapore, Sydney, Toronto, 2013.
- [4] Unknown. Combinaciones y permutaciones, sep 2015.
- [5] Unknown. Numeros de catakán, oct 2015.