

The background of the entire page is a complex, abstract 3D geometric pattern. It consists of numerous white cubes of varying sizes and orientations, arranged in a way that creates a sense of depth and movement. The cubes are set against a solid black background, which makes the white structure stand out prominently. The pattern appears to be a digital or mathematical construction, possibly related to the book's theme of probability and inference.

APUNTES

**Probabilidad, Procesos Aleatorios
e Inferencia**

Aldo Nicolás Alejandro Miranda Aguilar

Apuntes de Probabilidad, Procesos Aleatorios e Inferencia

Aldo Nicolás Alejandro Miranda Aguilar
Centro de Investigación en Computación
Instituto Politécnico Nacional
<http://www.cic.ipn.mx>

Statistically, the probability of any
one of us being here is so small
that the mere fact of our existence
should keep us all in a state of
contented dazzlement

Lewis Thomas

Table of Contents

Table of Contents

iii

1	Introducción	1
1.1	Teoría de conjuntos	1
1.1.1	Preliminares	1
1.1.2	Conjuntos, elementos	1
1.2	Práctica: Dibujar el triángulo de Pascal con TIKZ	8
1.3	Espacio muestral	8
1.4	Práctica: Instalar R y practicar con el capítulo 3 de [1]	9
1.5	Árbol de probabilidad	10
1.6	Práctica: Hacer diagramas de árbol con TIKZ	11
1.7	El principio del palomar	13
1.7.1	Enunciado	13
1.7.2	Demostración	13
1.7.3	Enunciado general	13
1.8	Reseña: ¿Qué es contar?	14
1.9	Resumen: El hombre anumerico	15
1.9.1	Números grandes y probabilidades pequeñas	15
1.9.2	Sangre, montañas y hamburguesas	15
1.9.3	Los números colosales y los 400 de Forbes	15
1.9.4	Arquímedes y los números prácticamente infinitos	16
1.9.5	La regla del producto y los vales de Mozart	16
1.9.6	Los helados de tres sabores y el truco de Von Neumann	16
1.9.7	Julio César y tú	16
1.10	Práctica: Simular un dado de tres caras en ROOT	17
1.11	Antinomia de Russell	20
1.12	La paradoja del barbero	21
1.13	make	22
2	El Teorema de Bayes	23
2.1	Introducción	23
2.2	Bayesianos contra Frecuentistas	24
2.2.1	¿Importa tanto la diferencia?	24

2.2.2	Y entonces, ¿quién tiene razón?	25
2.3	El problema de Monty Hall	26
2.3.1	Explicación intuitiva	27
2.3.2	Explicación matemática	27
2.4	La falacia del apostador	28
2.5	Problema de las ovejas y el lobo	29
3	Variable aleatoria	31
3.1	Introducción	31
3.1.1	Tipos de variables aleatorias	31
3.1.2	Distribución de probabilidad de una variable aleatoria	32
3.1.3	Función de densidad de una v.a. continua	32
3.2	Teorema del límite central	33
3.3	Práctica: Realizar un histograma en ROOT que se llene con la distribución de Gauss.	34
3.4	Mínimos cuadrados	35
3.5	Función de densidad de probabilidad	35
3.6	Cuota estadística (Odds)	36
3.7	Curtosis	39
3.8	Skewness	40
3.8.1	Coefficiente de asimetría de Fisher	40
3.8.2	Coefficiente de asimetría de Pearson	41
3.9	Programa: Congruencia Zeller	42
3.10	Is Economics the Next Physical Science?	43
3.10.1	Física social	43
3.10.2	La búsqueda de la universalidad	45
3.10.3	Memorias distantes	45
3.10.4	Modelar el comportamiento de los agentes	46
3.10.5	Racionalidad imperfecta	46
3.10.6	Juegos minoritarios	46
3.10.7	Métodos de entropía	47
3.11	Curva ROC	48
3.11.1	El espacio ROC	49
3.11.2	Cómo se puede interpretar una curva ROC	51
3.12	Delta de Dirac	52
3.12.1	Definición como distribución de densidad	53
3.12.2	Definición como límite de sucesiones de funciones	53
3.12.3	Propiedades	53
3.13	El Hombre Anumérico: Capítulo 2 "Probabilidad y coincidencia"	54
3.13.1	Algunos cumpleaños y un cumpleaños determinado	54
3.13.2	Encuentros fortuitos	54
3.13.3	Un timo bursátil	55
3.13.4	Valores esperados: de los análisis de sangre al juego del chuck-a-luck	55

3.13.5	Eligiendo cónyuge	55
3.13.6	Las coincidencias y la ley	56
3.13.7	Monedas no trucadas y ganadores o perdedores en el juego de la vida	56
3.13.8	La racha de suerte y el manitas	57
3.14	Matrices semejantes	57
3.14.1	Introducción	57
3.14.2	Aplicación: potencias de una matriz	59
3.15	Ejercicios de cadenas de Markov	59
3.16	Matrices dispersas	61
3.16.1	Introducción	61
3.16.2	Lista enlazada	61
3.16.3	Aplicaciones	62
3.17	Cómo modelar de crecimiento viral	63
3.17.1	¿Qué es un producto viral?	63
3.17.2	El modelo más simple posible	64
3.18	Teoremas de punto fijo	65
3.18.1	En el análisis matemático	65
3.18.2	En álgebra y matemáticas discretas	66
3.18.3	Lista de teoremas de punto fijo	67
3.19	Clasificación de la inteligencia artificial	67
3.19.1	Inteligencia artificial convencional (IA suave) . . .	68
3.19.2	Inteligencia artificial computacional (IA dura) . . .	68
3.20	Teorema de los cuatro colores	68
4	Apéndices	72
4.1	Examen propuesto	72
4.2	Examen resuelto	78
	Bibliography	85

Introducción

1.1 Teoría de conjuntos

1.1.1 Preliminares

Este capítulo trata de algunas de las ideas y conceptos elementales de la teoría de conjuntos que son necesarios para una introducción moderna a la teoría de la probabilidad.

1.1.2 Conjuntos, elementos

Se llama conjunto a una lista o colección bien definida de objetos; los objetos comprendidos en un conjunto son llamados sus elementos o miembros.

Escribimos

$p \in A$ si p es un elemento del conjunto A

Si cada elemento de A pertenece también a un conjunto B , esto es, si $p \in A$ implica $p \in B$, entonces se dice que A es subconjunto de B , o que está contenido en B ; esto se denota por

$$A \subset B \text{ o } B \supset A$$

Dos conjuntos son iguales si cada uno está contenido en el otro; esto es,

$$A = B \text{ si y sólo si } A \subset B \text{ y } B \subset A$$

Las negaciones de $p \in A$, $A \subset B$ y $A = B$ se escriben $p \notin A$, $A \not\subset B$ y $A \neq B$ respectivamente.

Especificamos un conjunto particular, o por la lista de sus elementos o estableciendo las propiedades que caracterizan dichos elementos. Por ejemplo,

$$A = \{1, 3, 5, 7, 9\}$$

significa que A es el conjunto formado por los números 1, 3, 5, 7 y 9; y

$$B = \{x : x \text{ es un numero primo}, x < 15\}$$

significa que B es el conjunto de los números primos menores que 15.

A menos que se establezca otra cosa, todos los conjuntos en una in-

vestigación se suponen subconjuntos de un conjunto fijo llamado conjunto universal denotado (en este capítulo) por U . También usamos el símbolo \emptyset para indicar el conjunto vacío o nulo, esto es, el conjunto que no contiene elementos; este conjunto se considera como un subconjunto de cualquier otro conjunto. Así para cualquier conjunto A , tenemos $\emptyset \subset A \subset U$.

Ejemplo 1.1: Los conjuntos A y B anteriores pueden también escribirse como

$$A = \{x : x \text{ es un número impar}, x < 10\} \text{ y } B = \{2, 3, 5, 7, 11, 13\}$$

Nótese que $9 \in A$ pero $9 \notin B$, y $11 \in B$ pero $11 \notin A$; mientras que $3 \in A$

y $3 \in B$, y $6 \notin A$ y $6 \notin B$

Ejemplo 1.2: Usamos los símbolos especiales siguientes:

\mathbb{N} = conjunto de los enteros positivos: 1, 2, 3, \dots

\mathbb{Z} = conjunto de los enteros: $\dots, -2, -1, 0, 1, 2, \dots$

\mathbb{R} = conjunto de los números reales.

Así tenemos $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{R}$.

Ejemplo 1.3: Los intervalos sobre la línea real, definidos a continuación, aparecen muy frecuentemente en matemáticas. Aquí a y b son números reales con $a < b$.

Intervalo abierto de a a b = $(a, b) = \{x : a < x < b\}$

Intervalo cerrado de a a b = $[a, b] = \{x : a \leq x \leq b\}$

Intervalo abierto-cerrado de a a b = $(a, b] = \{x : a < x \leq b\}$

Intervalo cerrado-abierto de a a b = $[a, b) = \{x : a \leq x < b\}$

Los intervalos abierto-cerrado y cerrado-abierto son también llamados intervalos semi-abiertos.

Ejemplo 1.4: En estudios de población, el conjunto universal está formado por todas las personas del mundo.

Ejemplo 1.5: Sea $C = \{x : x^2 = 4, x \text{ es impar}\}$. Entonces $C = \emptyset$, o sea que C es el conjunto vacío.

Teorema 1.1: Sean A , B y C unos conjuntos. Además: (i) $A \subset A$; (ii) si $A \subset B$ y $B \subset A$ entonces $A = B$; y (iii) si $A \subset B$ y $B \subset C$ entonces $A \subset C$.

Hacemos énfasis en que a $A \subset B$ no excluye la posibilidad de $A = B$. Sin embargo, si $A \subset B$ pero $A \neq B$, entonces decimos que A es un subconjunto propio de B . (Algunos autores usan el símbolo \subsetneq para un subconjunto y \subset solamente para un subconjunto propio.)

OPERACIONES CON CONJUNTOS

Sean A y B conjuntos arbitrarios. La reunión de A y B expresada por $A \cup B$, es el conjunto de elementos que pertenecen a A o a B :

$$A \cup B = \{x : x \in A \text{ o } x \in B\}$$

Aquí "o" se usa en el sentido de y/o.

La intersección de A y B , expresada por $A \cap B$, es el conjunto de elementos que pertenecen a A y B :

$$A \cap B = \{x : x \in A \text{ y } x \in B\}$$

Si $A \cap B = \emptyset$, esto es, si A y B no tienen elementos en común, entonces se dice que A y B son disyuntos.

La diferencia de A y B o el complemento relativo de B con respecto a A , expresada por $A \setminus B$, es el conjunto de elementos que pertenecen a A pero no a B :

$$A \setminus B = \{x : x \in A, x \notin B\}$$

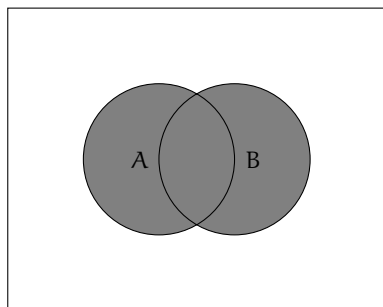
Obsérvese que $A \setminus B$ y B son disyuntos, esto es, $(A \setminus B) \cap B = \emptyset$.

El complemento absoluto o, simplemente, complemento de A , expresado por A^c , es el conjunto de elementos que no pertenecen a A :

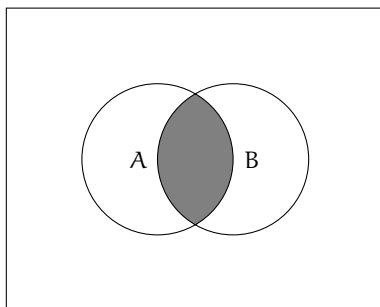
$$A^c = \{x : x \in U, x \notin A\}$$

O sea que A^c es la diferencia entre el conjunto universal U y el conjunto A .

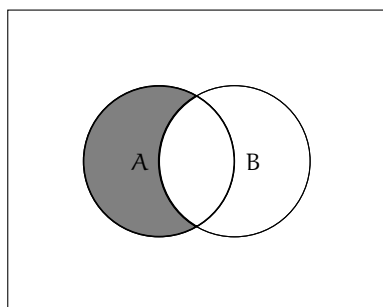
Ejemplo 1.6: Los diagramas siguientes, llamados diagramas de Venn, ilustran las operaciones anteriores. Aquí los conjuntos están representados por simples superficies planas y U , el conjunto universal, por la superficie total del rectángulo.



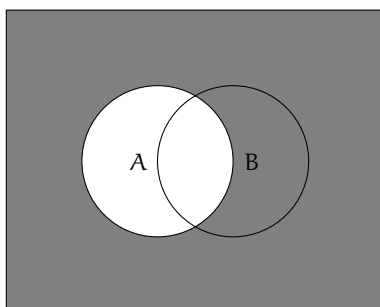
$A \cup B$ sombreado



$A \cap B$ sombreado



$A \setminus B$ sombreado



A^c sombreado

Ejemplo 1.7: Sean $A = \{1, 2, 3, 4\}$ y $B = \{3, 4, 5, 6\}$ donde $U = \{1, 2, 3, \dots\}$. Entonces

$$\begin{aligned} A \cup B &= \{1, 2, 3, 4, 5, 6\} & A \cap B &= \{3, 4\} \\ A \setminus B &= \{1, 2\} & A^c &= \{5, 6, 7, \dots\} \end{aligned}$$

LEYES DEL ÁLGEBRA DE CONJUNTOS	
Leyes de Idempotencia	
1a. $A \cup A = A$	1b. $A \cap A = A$
Leyes Asociativas	
2a. $(A \cup B) \cup C = A \cup (B \cup C)$	2b. $(A \cap B) \cap C = A \cap (B \cap C)$
Leyes conmutativas	
3a. $A \cup B = B \cup A$	3b. $A \cap B = B \cap A$
Leyes distributivas	
4a. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	4b. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
Leyes de identidad	
5a. $A \cup \emptyset = A$	5b. $A \cap U = A$
6a. $A \cup U = U$	6b. $A \cap \emptyset = \emptyset$
Leyes de complemento	
7a. $A \cup A^c = U$	7b. $A \cap A^c = \emptyset$
8a. $(A^c)^c = A$	8b. $U^c = \emptyset, \emptyset^c = U$
Leyes de De Morgan	
9a. $(A \cup B)^c = A^c \cap B^c$	9b. $(A \cap B)^c = A^c \cup B^c$

Los conjuntos de las operaciones anteriores satisfacen las diferentes leyes o identidades que se relacionan en la tabla inferior (tabla 1). Para el efecto, establecemos el:

Teorema 1.2: Conjuntos que satisfacen las leyes de la tabla 1

Observación: Cada una de las leyes anteriores provienen de una ley lógica análoga. Por ejemplo,

$$A \cap B = \{x : x \in A \text{ y } z \in B\} = \{z : z \in B \text{ y } z \in A\} = B \cap A$$

Aprovechamos aquí el hecho de que la proposición compuesta “p y q”, escrito $p \wedge q$, es lógicamente equivalente a la proposición compuesta “q y p”, esto es, $q \wedge p$.

La relación entre contenenencia y conjunto contenido en otro y las anteriores operaciones con conjuntos lleva al:

Teorema 1.3: Cada una de las condiciones siguientes es equivalente a $A \subset B$:

(i) $A \cap B = A$

(iii) $B^c \subset A^c$

(v) $B \cup A^c = U$

(ii) $A \cup B = B$

(iv) $A \cap B^c = \emptyset$

CONJUNTOS FINITOS Y CONTABLES

Los conjuntos pueden ser finitos o infinitos. Un conjunto es finito si está vacío o si consta exactamente de n elementosen donde n es un entero positivo; de otra manera es infinito.

Ejemplo 1.8: Sea M el conjunto de los días de la semana; esto es,

$M = \{\text{lunes, martes, miércoles, jueves, viernes, sábado, domingo}\}$

Entonces M es finito.

Ejemplo 1.9: Sea $P = \{x : x \text{ es un río en la Tierra}\}$. Aunque puede ser difícil contar el número de ríos de la Tierra P es un conjunto finito.

Ejemplo 1.10: Sea Y el conjunto de los enteros pares (positivos), esto es, $Y = \{2, 4, 6, \dots\}$. Entonces Y es un conjunto infinito.

Ejemplo 1.11: Sea I el intervalo unidad de los números reales, esto es $I = \{x : 0 \leq x \leq 1\}$. Entonces I es un conjunto infinito.

Un conjunto es contable si es finito o si sus elementos pueden ser ordenados en forma de sucesión, en cuyo caso se dice que es contablemente infinito; de lo contrario el conjunto es no contable. El conjunto del ejemplo 1.10 es contablemente infinito, mientras se puede comprobar que el conjunto del ejemplo 1.11 es no contable.

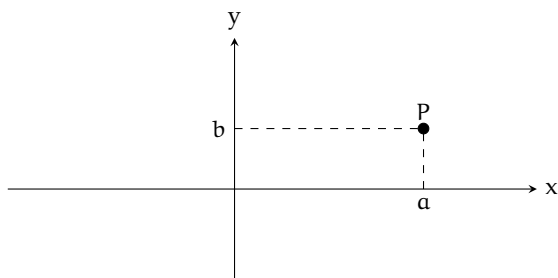
CONJUNTO PRODUCTO

Sean A y B dos conjuntos. El conjunto producto de A y B , expresado por $A \times B$, está formado por todas las parejas ordenadas (a, b) donde $a \in A$ y $b \in B$:

$$A \times B = \{(a, b) : a \in A, b \in B\}$$

El producto de un conjunto por sí mismo, $A \times A$, se denota A^2

Ejemplo 1.12: El lector está familiarizado con el plano cartesiano $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ indicado abajo.



Aquí cada punto P representa una pareja ordenada (a, b) de números reales y viceversa.

Ejemplo 1.13: Sean $A = \{1, 2, 3\}$ y $B = \{a, b\}$. Entonces

$$A \times B = \{(1, a), (1, b), (2, a), (2, b), (3, a), (3, b)\}$$

El concepto de conjunto de producto se extiende a un número finito de conjuntos en forma natural. El conjunto producto de los conjuntos A_1, A_2, \dots, A_m , escrito $A_1 \times A_2 \times \dots \times A_m$ es el conjunto de todas las m -tuplas ordenadas (a_1, a_2, \dots, a_m) donde $a_i \in A_i$ para cada i .

CLASES DE CONJUNTOS

Con frecuencia los elementos de un conjunto son a su vez conjuntos. Por ejemplo, en un conjunto de líneas cada línea es un conjunto de puntos. Para aclarar estos casos, se acostumbra usar para dicho conjunto la palabra *clase* o *familia*. Las palabras subclase y subfamilia tienen significados análogos a subconjunto.

Ejemplo 1.14: Los miembros de la clase $\{\{2, 3\}, \{2\}, \{5, 6\}\}$ son los conjuntos $\{2, 3\}$, $\{2\}$ y $\{5, 6\}$

Ejemplo 1.15: Considérese un conjunto A . El conjunto potencia de A , expresado por $\mathbb{P}(A)$, es la clase de todos los subconjuntos de A . En particular, si $A = \{a, b, c\}$, entonces

$$\mathbb{P}(A) = \{A, \{a, b\}, \{a, c\}, \{b, c\}, \{a\}, \{b\}, \{c\}, \emptyset\}$$

En general, si A es finito y tiene n elementos, entonces $\mathbb{P}(A)$ tendrá 2^n elementos.

Una partición de un conjunto X es una subdivisión de X entre subconjuntos no vacíos de X tales que a cada $a \in X$ pertenece a un único subconjunto. Los subconjuntos de una partición son llamados células.

Ejemplo 1.16: Considérense las siguientes clases de subconjuntos de $X = \{1, 2, \dots, 8, 9\}$:

- (i) $\{\{1, 3, 5\}, \{2, 6\}, \{4, 8, 9\}\}$
- (ii) $\{\{1, 3, 5\}, \{2, 4, 6, 8\}, \{5, 7, 9\}\}$
- (iii) $\{\{1, 3, 5\}, \{2, 4, 6, 8\}, \{7, 9\}\}$

Entonces (i) no es una partición de X puesto que $7 \in X$ pero 7 no pertenece a ninguna célula. Además, (ii) no es una partición de X puesto que $5 \in X$ y 5 pertenece a ambas $\{1, 3, 5\}$ y $\{5, 7, 9\}$. Por otra parte, (iii) es una partición de X puesto que cada elemento de X pertenece a una célula exactamente.

Cuando hablamos de clases de conjuntos con índices $\{A_i : i \in I\}$ o simplemente $\{A_i\}$, significamos que hay un conjunto A_i , asignado a cada elemento $i \in I$. El conjunto I se denomina conjunto de índices y se dice que los conjuntos A_i tienen por índice I . Cuando el conjunto de índices es el conjunto \mathbb{N} de los enteros positivos, la clase con índices $\{A_1, A_2, \dots\}$ se denomina sucesión de conjuntos. Por la reunión de conjuntos de los A_i denotados por $\bigcup_{i \in I} A_i$ (o simplemente $\bigcup_i A_i$), se entiende el conjunto de elementos que pertenecen por lo menos a uno de los A_i ; y por la intersección de los A_i , denotada por $\bigcap_{i \in I} A_i$ (o simplemente $\bigcap_i A_i$), se entiende el conjunto de elementos que pertenecen a cada A_i .

También escribimos

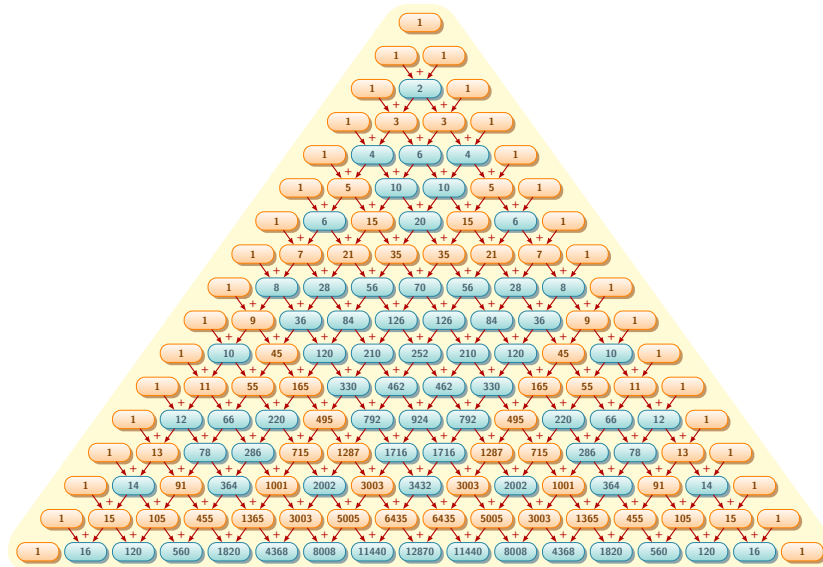
$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \dots \quad \text{y} \quad \bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \dots$$

para la reunión e intersección, respectivamente, de una sucesión de conjuntos.

Definición: Una clase no vacía \mathcal{A} de subconjuntos de U se denomina un álgebra (σ -álgebra) de conjuntos si:

- (i) el complemento de algún conjunto de \mathcal{A} pertenece a \mathcal{A} ; y (ii) la reunión de un número finito (contable) de conjuntos de \mathcal{A} pertenece a \mathcal{A} ; esto es, si \mathcal{A} es cerrada para complementos y reuniones finitas (contables).

1.2 Práctica: Dibujar el triángulo de Pascal con TIKZ



1.3 Espacio muestral

En la teoría de probabilidades, el espacio muestral o espacio de muestreo (denotado E , S , Ω 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 σ -álgebra, 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 (Ω, F, P) incorpora un espacio de muestreo de resultados, Ω , pero define un conjunto de sucesos de interés, la σ -álgebra F , por la cuál se define la medida de probabilidad P .

1.4 Práctica: Instalar R y practicar con el capítulo 3 de [1]

```
> 1+2+3
[1] 6
> 1+2*3
[1] 7
> (1+2)*3
[1] 9
> c(0, 1, 1, 2, 3, 5, 8)
[1] 0 1 1 2 3 5 8
> 1:50
 [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
[33] 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
> c(1, 2, 3, 4) + c(10, 20, 30, 40)
[1] 11 22 33 44
> c(1, 2, 3, 4, 5) + c(10, 100)
[1] 11 102 13 104 15
Warning message:
In c(1, 2, 3, 4, 5) + c(10, 100) :
  longitud de objeto mayor no es m'ultiplo de la longitud de uno menor
> 1 / c(1, 2, 3, 4, 5)
[1] 1.0000000 0.5000000 0.3333333 0.2500000 0.2000000
> "Hello_world."
[1] "Hello_world."
> c("Hello_world", "Hello_R_interpreter")
[1] "Hello_world" "Hello_R_interpreter"
> # Here is an example of a comment at the beginning of a line
> 1 + 2 + # and here is an example in the middle
+ 3
[1] 6
> exp(1)
[1] 2.718282
> cos(3.141593)
[1] -1
> log2(1)
[1] 0
> log(x=64, base=4)
[1] 3
> log(64,4)
[1] 3
> 2 ^ 10
[1] 1024
> 3 == 4
[1] FALSE
> x <- 1
> y <- 2
> z <- c(x,y)
> # evaluate z to see what's stored as z
```

```

> z
[1] 1 2
> y <- 4
> z
[1] 1 2
> b <- c(1,2,3,4,5,6,7,8,9,10,11,12)
> b
[1] 1 2 3 4 5 6 7 8 9 10 11 12
> # let's fetch the 7th item in vector b
> b[7]
[1] 7
> # fetch items 1 through 6
> b[1:6]
[1] 1 2 3 4 5 6
> # fetch only members of b that are congruent to zero (mod 3)
> # (in non-math speak, members that are multiples of 3)
> b[b %% 3 == 0]
[1] 3 6 9 12
> # fetch 1, 6, 11
> b[c(1,6,11)]
[1] 1 6 11
> b[b %% 3 == 0]
[1] FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE TRUE
> b[b %% 3 == 0]
[1] 3 6 9 12

```

1.5 Árbol de probabilidad

Un diagrama de árbol es una herramienta que se utiliza para determinar todos los posibles resultados de un experimento aleatorio. En el cálculo de la probabilidad se requiere conocer el número de objetos que forman parte del espacio muestral, estos se pueden determinar con la construcción de un diagrama de árbol.

El diagrama de árbol es una representación gráfica de los posibles resultados del experimento, el cual consta de una serie de pasos, donde cada uno de estos tiene un número finito de maneras de ser llevado a cabo. Se utiliza en los problemas de conteo y probabilidad.

Para la construcción de un diagrama en árbol se partirá poniendo una rama para cada una de las posibilidades, acompañada de su probabilidad. Cada una de estas ramas se conoce como rama de primera generación.

En el final de cada rama de primera generación se constituye a su vez, un nudo del cual parten nuevas ramas conocidas como ramas de segunda generación, según las posibilidades del siguiente paso, salvo si el nudo representa un posible final del experimento (nudo final).

Hay que tener en cuenta que la construcción de un árbol no depende de tener el mismo número de ramas de segunda generación que salen de cada rama de primera generación y que la suma de probabilidades de las ramas de cada nudo ha de dar 1.

Existe un principio sencillo de los diagramas de árbol que hace que éstos sean mucho más útiles para los cálculos rápidos de probabilidad: multiplicamos las probabilidades si se trata de ramas adyacentes (contiguas), el ejemplo de alumna de la primera facultad, o bien las sumamos si se trata de ramas separadas que emergen de un mismo punto, el ejemplo de encontrar un alumno.

1.6 Práctica: Hacer diagramas de árbol con TIKZ

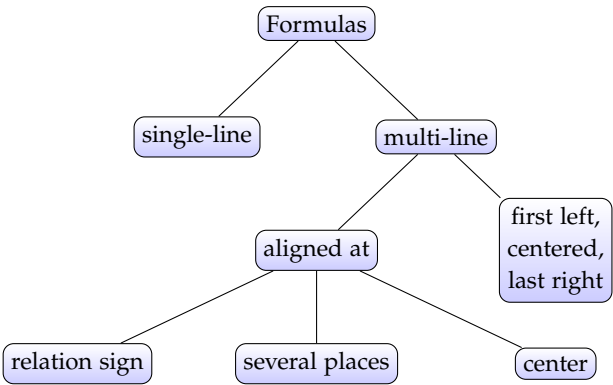


FIGURE 1.1 Árbol básico

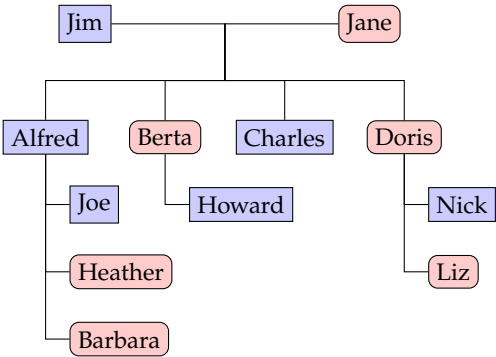
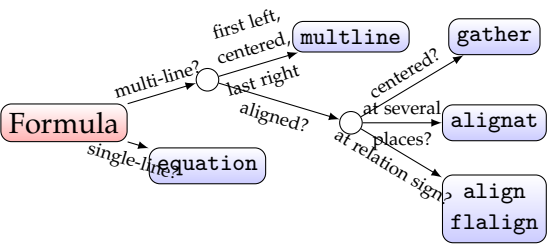


FIGURE 1.2 Árbol genealógico



Árbol de decisión

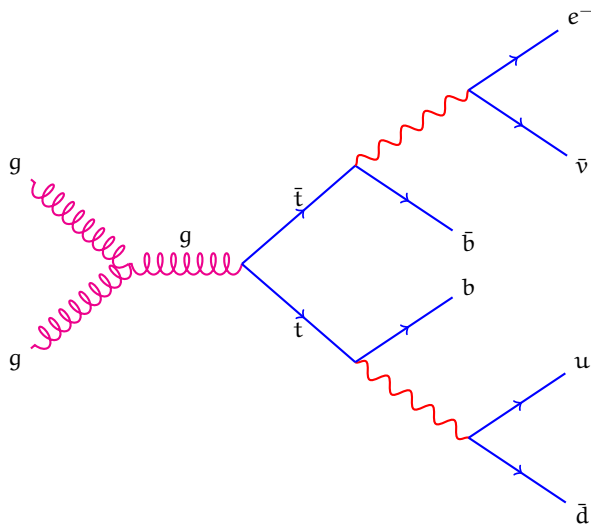
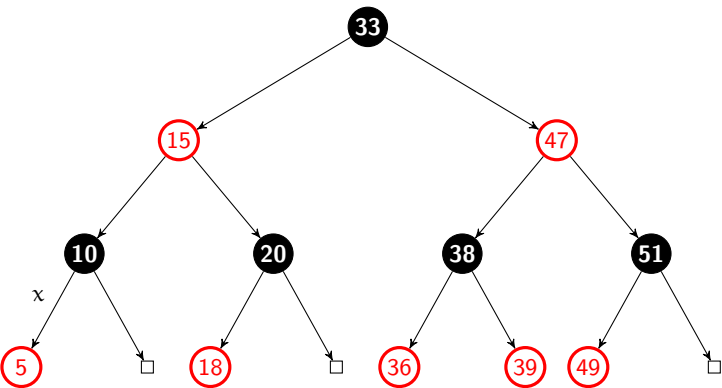


Diagrama de Feynman



Árbol rojo-negro

1.7 El principio del palomar

El principio del palomar, también llamado principio de Dirichlet o principio de las cajas, establece que si n palomas se distribuyen en m palomares, y si $n \geq m$, entonces al menos habrá un palomar con más de una paloma. Otra forma de decirlo es que m huecos pueden albergar como mucho m objetos si cada uno de los objetos está en un hueco distinto, así que el hecho de añadir otro objeto fuerza a volver a utilizar alguno de los huecos. A manera de ejemplo: si se toman trece personas, al menos dos habrán nacido el mismo mes.

El primer enunciado del principio se cree que proviene de Dirichlet en 1834 con el nombre de Schubfachprinzip ("principio de los cajones"). No debe confundirse con otro principio sobre funciones armónicas, también con el nombre de este autor.

1.7.1 Enunciado

Principio de distribución, del palomar o del cajón de la paloma de Dirichlet. Sean m , n y p tres números naturales. Si se desean colocar $np + m$ palomas en n cajas, alguna caja debe contener al menos $p + 1$ palomas.

1.7.2 Demostración

Si cada caja contiene como mucho p objetos, el número total de objetos que podemos colocar es $np < np + 1 \leq np + m$.

En su versión más simple, este principio dice que no puede existir una aplicación inyectiva entre un conjunto de m elementos y otro de n elementos, si $m > n$. Equivalentemente, si se desean colocar m objetos en n cajas, con $m > n$, al menos una caja debe contener al menos 2 objetos.

1.7.3 Enunciado general

Una versión generalizada de este principio dice que, si n objetos discretos deben guardarse en m cajas, al menos una caja debe contener no menos de

$\lceil \frac{n}{m} \rceil$ objetos, donde $\lceil \dots \rceil$ denota la función techo.

Además existirá otra caja que contendrá no más de

$\lfloor \frac{n}{m} \rfloor$ objetos, donde $\lfloor \dots \rfloor$ denota la función suelo.

Como ejemplo de aplicación en una ciudad de más de un millón de habitantes habrá como mínimo 2733 personas que hayan nacido el mismo día del año, ya que:

$\lceil 1000000/366 \rceil = \lceil 2732,24\dots \rceil = 2733$ Donde se ha tenido en cuenta que existen 366 posibilidades para la fecha de aniversario de una persona contando la existencia de años bisiestos.

1.8 Reseña: ¿Qué es contar?

Podría parecer una pregunta simple, en principio. Sin embargo, hasta el día de hoy este proceso es capaz de plantear cuestionamientos interesantes para la ciencia.

Contar, según una referencia rápida como Wikipedia significa es: "Un proceso de abstracción que nos lleva a otorgar un número cardinal como representativo de un conjunto". Una definición razonable en términos generales, pero también acotada al hecho de que sólo los seres humanos somos capaces de contar; inclusive la misma definición podría suponer que debemos tener la abstracción de lo que son los números cardinales para poder contar.

A pesar de lo anterior, hay evidencia de que el acto de contar, de alguna manera, está presente en seres que anteriormente la ciencia ha descartado que puedan llegar a poseer la habilidad del conteo. Por ejemplo:

- Se ha demostrado que animales como ardillas o loros pueden ser entrenados para contar.
- Algunos insectos pueden reconocer números en patrones visuales similares
- Por supuesto, los primates avanzados como los chimpancés han demostrado tener habilidad para aprender a contar e incluso interactuar con nuestras máquinas
- Las cigalas son unos insectos que pasan las primeras etapas de su vida bajo tierra, sin embargo, cuando emergen a la superficie, lo hacen en periodos de años que coinciden con los números primos, usualmente 13 y 17.
- Estudios recientes indican que las hormigas son capaces de encontrar el camino de vuelta a su hormiguero en distancias de hasta 50 metros. También se observó que hay una relación entre la longitud de las patas de los insectos y este mecanismo de conteo.

Como mencionamos anteriormente, era de creencia común pensar que los seres humanos comenzamos a contar a partir de que desarrollamos un sistema de escritura. Sin embargo, hay contraejemplos como Quipus de los antiguos Incas. Estos artefactos eran capaces de registrar objetos, intervalos de tiempo, planos de construcción e incluso los patrones de danza.

En la opinión de quien escribe, me resulta bastante interesante el hecho

de que estemos descubriendo cualidades en los seres, mal llamados *inferiores* que pueden ayudarnos a encontrar respuestas a retos tecnológicos que tenemos en la actualidad o por el lado del tema de los Quipus que estemos volteando a técnicas perdidas en el tiempo que si las redescubrimos y adaptamos a nuestro entorno, pueden ayudarnos a encontrar paradigmas tecnológicos más eficientes.

1.9 Resumen: El hombre anumerico

1.9.1 Números grandes y probabilidades pequeñas

La facilidad con que la gente se desenvuelve con los números va de la del aristócrata a la de Ramanujan, pero la triste realidad es que la mayoría está más próxima al aristócrata.

1.9.2 Sangre, montañas y hamburguesas

En una columna sobre anumerismo en Scientific American, el informático Douglas Hofstadter cita el caso de la Ideal Toy Company, que en el envoltorio del cubo de Rubik afirmaba que el cubo admitía más de tres mil millones de configuraciones distintas. Si uno lo calcula, obtiene que las configuraciones posibles son más de $4 \cdot 10^{19}$, un 4 seguido de 19 ceros. La frase del envoltorio es cierta, las configuraciones

posibles son, en efecto, más de tres mil millones. La subestimación que supone esa cifra es, sin embargo, un síntoma de un omnipresente anumerismo que encaja muy mal en una sociedad tecnológicamente avanzada. Es como si en la entrada del Lincoln Túnel hubiera un rótulo anunciando: Nueva York, más de 6 habitantes; o como si McDonald se vanagloriara de haber vendido más de 120 hamburguesas.

1.9.3 Los números colosales y los 400 de Forbes

El tema de los cambios de escala ha sido uno de los pilares de la literatura mundial, desde la Biblia hasta los liliputienses de Swift, y desde Paul Bunyan hasta el colosal Gargantúa de Rabelais. Siempre me ha chocado, sin embargo, la inconsistencia que han mostrado los distintos autores en su empleo de los números grandes. Se dice que el niño Gargantúa se tomaba la leche de 17.913 vacas. De joven fue a

estudiar a París montado en una yegua que abultaba como seis elefantes y llevaba colgadas del cuello las campanas de Nôtre Dame a modo de cascabeles. En el camino de vuelta a casa, fue atacado a cañonazos desde un castillo y se sacó las bombas del pelo con un rastrillo de 300 metros de longitud. Para hacerse una ensalada cortaba lechugas del tamaño de un nogal y devoraba media docena de peregrinos que se habían refugiado en la arboleda. ¿Pueden apreciar las inconsistencias internas de este cuento?

1.9.4 Arquímedes y los números prácticamente infinitos

La arquimedianidad es una propiedad fundamental de los números (llamada así por el matemático griego Arquímedes), según la cual se puede rebasar cualquier número, por grande que sea, agregando repetidas veces cualquier número menor, por pequeño que este sea. Aunque esta propiedad sea en principio evidente, a veces la gente se resiste a aceptar sus consecuencias, como ese alumno mío que sostenía que el cabello humano no crece a razón de kilómetros por hora.

1.9.5 La regla del producto y los vales de Mozart

Este es quizás un buen momento para insistir en lo que dije al principio, que el lector anumérico puede saltarse tranquilamente los trozos más difíciles que vaya encontrando de vez en cuando. En las siguientes secciones puede que haya algunos. Del mismo modo, el lector anumérico puede saltarse tranquilamente los trozos triviales con que se encuentre. (Claro que cualquiera puede saltarse tranquilamente cualquier parte del libro, pero preferiría que esto sólo ocurriera con párrafos aislados).

1.9.6 Los helados de tres sabores y el truco de Von Neumann

Las heladerías Baskin-Robbins anuncian helados de treinta y un sabores distintos. El número de helados posibles de tres sabores distintos es por tanto

$$31 \times 30 \times 29 = 26.970$$

cualquiera de los treinta y un sabores puede estar encima, cualquiera de los treinta restantes puede estar en el centro y cualquiera de los veintinueve restantes debajo. Si no nos importa el orden en que están los sabores del helado, sino que sólo nos interesa saber cuántos posibles helados de tres sabores hay, dividiremos 26.970 entre 6, con los que obtendremos 4.495 helados distintos. El motivo de esta división es que hay $6 = 3 \times 2 \times 1$ maneras distintas de ordenar los tres sabores en un helado de, por ejemplo, fresa, vainilla y chocolate:

FVC, FCV, VFC, VCF, CVF y CFV.

1.9.7 Julio César y tú

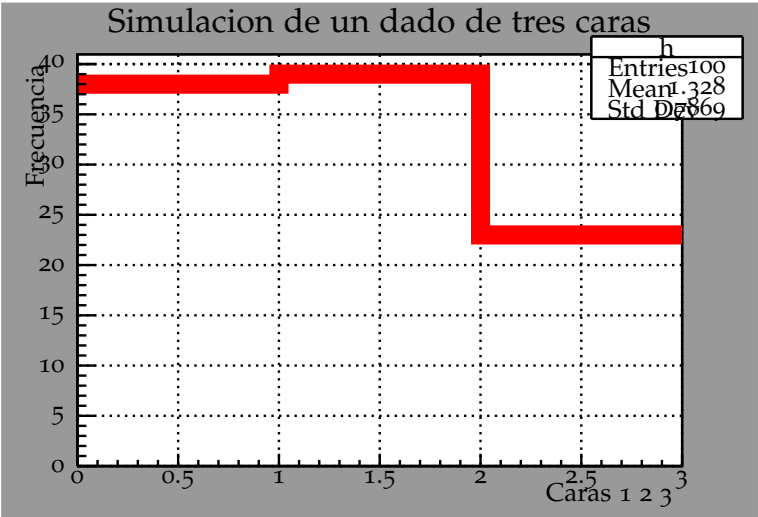
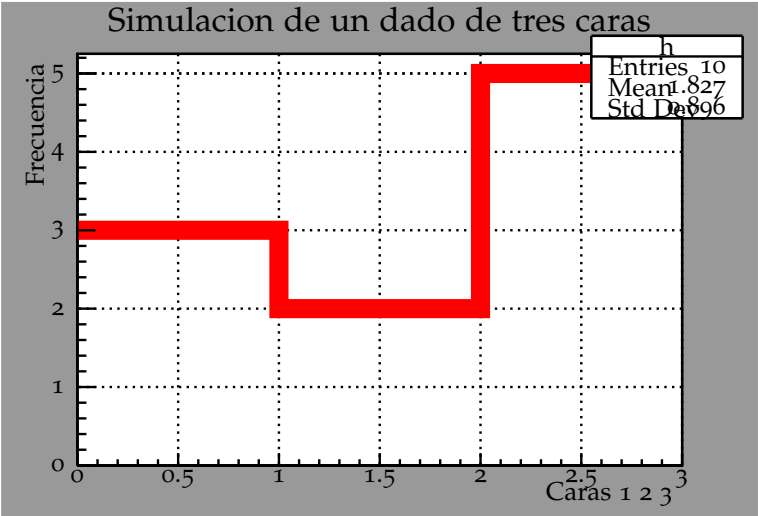
Para terminar, daremos otras dos aplicaciones de la regla del producto, la primera un tanto deprimente y la segunda, esperanzadora. La primera es la probabilidad de no sufrir ninguna enfermedad, accidente u otra desgracia de cierta lista que enumeraré. No morir en un accidente de automóvil es seguro en un 99 por ciento, mientras que un 98 por ciento de nosotros se salvará de morir en un accidente doméstico. Tenemos una probabilidad del 95 por ciento de librarnos de una enfermedad pulmonar; un 90 por ciento de la locura; un 80 por ciento del cáncer, y un 75

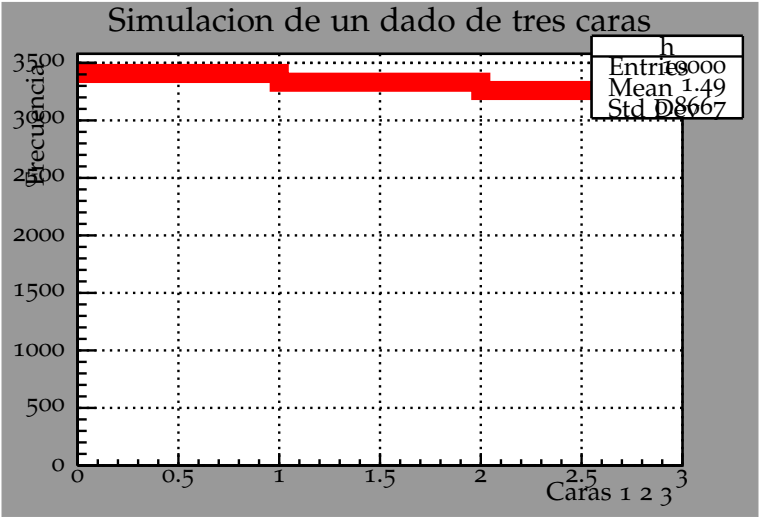
por ciento de una enfermedad del corazón. He tomado sólo estas cifras a modo de ejemplo, pero se pueden hacer estimaciones muy precisas para una amplia gama de posibles calamidades. Y aunque la probabilidad de librarse de cada una de estas enfermedades o accidentes por separado es alentadora, la de salvarse de todas no lo es. Si suponemos que, en general, estas desgracias son independientes, y multiplicamos todas las probabilidades citadas, el producto se hace en seguida inquietantemente pequeño: la probabilidad de no padecer ninguna desgracia de esta corta lista que he citado es menor del 50 por ciento. Resulta pues preocupante que algo tan inofensivo como la regla del producto pueda intensificar en tal medida nuestra mortalidad.

1.10 Práctica: Simular un dado de tres caras en ROOT

dice3.C

```
void dice3(){
    TCanvas *c1 = new TCanvas("c1", "Simulacion de un dado de tres caras", 200, 100, 700, 500);
    TH1F *h1 = new TH1F("h", "histograma", 3, 0, 3);
    h1->SetLineColor(2);
    h1->SetLineWidth(8);
    h1->SetMarkerColor(50);
    h1->SetMarkerStyle(50);
    h1->SetTitle("Simulacion de un dado de tres caras");
    h1->GetXaxis()->SetTitle("Caras 1,2,3");
    h1->GetYaxis()->SetTitle("Frecuencia");
    TF1 fun1("fun1", "2", 0, 3);
    c1->SetFillColor(15);
    c1->SetGrid();
    h1->SetMinimum(0);
    for(int i=0; i<10000; i++){
        h1->Fill(fun1.GetRandom());
    }
    h1->Draw();
    c1->Update();
    c1->GetFrame()->SetFillColor(10);
    c1->GetFrame()->SetBorderSize(20);
    c1->Modified();
}
```





1.11 Antinomia de Russell

"Bertrand Russell construyó una contradicción dentro del sistema mismo de la lógica elemental, que es precisamente análoga a la contradicción primeramente desarrollada en la teoría cantoriana de las clases infinitas. La antinomia de Russell puede ser enunciada del modo siguiente. Las clases parecen ser de dos tipos: las que no se contienen a sí mismas como miembros y las que sí se contienen. Una clase será llamada "normal" si, y solamente si, no se contiene a sí misma como miembro; en otro caso se la llamará "no normal". Un ejemplo de clase normal es la clase de los matemáticos, ya que, evidentemente, la clase misma no es un matemático y, por tanto, no es un miembro de sí misma. Un ejemplo de clase no normal es la clase de todas las cosas pensables, ya que la clase de todas las cosas pensables es, a su vez, pensable y, por consiguiente, un miembro de sí misma. Sea "N", por definición, la clase de todas las clases normales. Preguntamos si N mismo es una clase normal. Si N es normal, es un miembro de sí misma (pues, por definición, N contiene a todas las clases normales); pero, en ese caso, N es no normal, porque, por definición, una clase que se contiene a sí misma es no normal. Por otra parte, si N es no normal, es un miembro de sí misma (por la definición de no normal); pero, en ese caso, N es normal, porque, por definición, los miembros de N son las clases normales. En resumen, N es normal si, y solamente si, N es no normal. De lo que se desprende que la afirmación "N es normal" es verdadera y falsa a la vez. Esta fatal contradicción se produce como consecuencia de utilizar sin espíritu crítico una noción aparentemente diáfana de clase. Posteriormente fueron encontrándose otras paradojas, construidas todas por medio de familiares y aparentemente convincentes modos de razonamiento. Los matemáticos acabaron comprendiendo que, en la tarea de desarrollar sistemas consistentes, la familiaridad y la claridad intuitiva son soportes harto débiles en que apoyarse."

1.12 La paradoja del barbero

Hace muchos años, en un lejano reino, había pocas personas que su oficio fuera ser barbero. Para solucionar el problema, el rey dictaminó que los barberos solo podían afeitarse a las personas que no podían afeitarse por sí mismas.

Uno de esos barberos, era el único en su comarca y le entró la siguiente duda: “Como barbero no puedo afeitar al barbero de mi comarca, que soy yo, porque entonces podría afeitarme a mí mismo. Pero entonces, algún barbero debe de afeitarme, pero como soy el único que hay, entonces no me puedo afeitar

¿Qué puede hacer nuestro pobre barbero para hacer desaparecer sus barbas? La verdad que está en una situación realmente complicada. Tan complicada, que es una paradoja y no tiene solución.

Para ver qué pasa lo primero que hacemos es definir qué es un conjunto. Podemos imaginarnos que un conjunto es un cajón muy grande donde metemos cosas. Por ejemplo, el conjunto de las frutas sería un cajón en el que metemos todas las manzanas, todas las naranjas, todas las peras...

Un subconjunto, sería tomar un conjunto más pequeño dentro de nuestro conjunto. Por ejemplo, en nuestro cajón gigante, poner dentro una caja en la que metemos la fruta roja. Ahí pondríamos las manzanas, las cerezas,...

Por último, los elementos serían por separado cada uno de los objetos o números que hay en un conjunto. En nuestro ejemplo sería una manzana o una naranja concretas.

Vamos a ver un ejemplo más numérico. Tomamos un conjunto X compuesto por los números naturales: $X = \{1, 2, 3, 4, 5, \dots\}$. Un subconjunto M serían por ejemplo los números pares: $M = \{2, 4, 6, \dots\}$. Y algunos elementos del conjunto M (que también lo serían de X) son por ejemplo el 4, el 6 ó el 48, por citar algunos. Pues una vez entendido esto, existen dos tipos de conjuntos: los normales y los singulares.

Los conjuntos normales son aquellos que no se contienen a sí mismos. En cambio, los singulares son aquellos que sí se contienen a sí mismo.

¿Cómo se contiene un conjunto a sí mismo? Supongamos que creamos el conjunto de todo lo que no es un lápiz. Como mi conjunto NO es un lápiz, entonces está dentro del propio conjunto.

Además, estas propiedades son excluyentes. Un conjunto es o normal o singular. Ni hay otra opción, ni pueden ser los dos a la vez.

¿Y qué tiene que ver todo esto con nuestro amigo el barbero? Pues que él está en una posición difícil. Si pertenece al conjunto de los que no se pueden afeitar, él razona que sí se podría afeitar. Y si pertenece al conjunto de los que sí se pueden afeitar, entonces razona que no se podría afeitar. Por lo tanto, el conjunto al que pertenece el barbero es un conjunto normal y singular a la vez, que hemos visto que es imposible.

Esta paradoja hizo temblar a toda la comunidad matemática, puesto que la base más elemental de las matemáticas nace a raíz de la teoría de conjuntos. El problema se solucionó excluyendo los conjuntos singulares, algo un poco trampa en mi opinión, pero que de momento funciona.

1.13 make

En el contexto del desarrollo de software, Make es una herramienta de gestión de dependencias, típicamente, las que existen entre los archivos que componen el código fuente de un programa, para dirigir su recompilación o "generación" automáticamente. Si bien es cierto que su función básica consiste en determinar automáticamente qué partes de un programa requieren ser recompiladas y ejecutar los comandos necesarios para hacerlo, también lo es que Make puede usarse en cualquier escenario en el que se requiera, de alguna forma, actualizar automáticamente un conjunto de archivos a partir de otro, cada vez que éste cambie.

Make es muy usada en los sistemas operativos tipo Unix/Linux. Por defecto lee las instrucciones para generar el programa u otra acción del fichero makefile. Las instrucciones escritas en este fichero se llaman dependencias.

La herramienta make se usa para las labores de creación de fichero ejecutable o programa, para su instalación, la limpieza de los archivos temporales en la creación del fichero, todo ello especificando unos parámetros iniciales (que deben estar en el makefile) al ejecutarlo.

Además de ser éste su objetivo principal, es utilizado para automatización de otras tareas como la creación de documentos del formato docbook, mantenimiento del sistema, simplemente usando o creando makefiles que hagan estas tareas.

El Teorema de Bayes

2.1 Introducción

El teorema de Bayes parte de una situación en la que es posible conocer las probabilidades de que ocurran una serie de sucesos A_i .

A esta se añade un suceso B cuya ocurrencia proporciona cierta información, porque las probabilidades de ocurrencia de B son distintas según el suceso A_i que haya ocurrido.

Conociendo que ha ocurrido el suceso B , la fórmula del teorema de Bayes nos indica como modifica esta información las probabilidades de los sucesos A_i .

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

Ejemplo: Si seleccionamos una persona al azar, la probabilidad de que sea diabética es 0.03. Obviamente la probabilidad de que no lo sea es 0.97. Si no disponemos de información adicional nada más podemos decir, pero supongamos que al realizar un análisis de sangre los niveles de glucosa son superiores a 1.000 mg/l, lo que ocurre en el 95% de los diabéticos y sólo en un 2% de las personas sanas.

¿Cuál será ahora la probabilidad de que esa persona sea diabética?

La respuesta que nos da el teorema de bayes es que esa información adicional hace que la probabilidad sea ahora 0,595. Vemos así que la información proporcionada por el análisis de sangre hace pasar, la probabilidad inicial de padecer diabetes de 0,03, a 0,595. Evidentemente si la prueba del análisis de sangre hubiese sido negativa, esta información modificaría las probabilidades en sentido contrario. En este caso la probabilidad de padecer diabetes se reduciría a 0,0016.

2.2 Bayesianos contra Frecuentistas

El edificio de la estadística descansa en el concepto de probabilidad. Y, sin embargo, no existe un consenso universal sobre cómo interpretar este concepto.

Existen dos formas de interpretar la idea de probabilidad: la interpretación clásica (o frecuentista) y la interpretación bayesiana.

La interpretación clásica, mayoritaria por lo menos hasta ahora, define la probabilidad en términos de experimentación. Si repites un experimento un número infinito de veces y compruebas que en 350 de cada 1.000 ocasiones se ha producido un determinado resultado, un frecuentista diría que la probabilidad de ese resultado es del 35%. Basándose en esta definición, un frecuentista afirma que es posible asociar a cada evento una probabilidad de obtener un valor VERDADERO del mismo.

La aproximación clásica se basa por lo tanto en estudiar la probabilidad "real" de las cosas, tratando de determinar hasta qué punto una medición realizada sobre un conjunto de experimentos se aproxima a la probabilidad real que subyace. Es por ello que un frecuentista definiría probabilidad como una expresión matemática que predice en qué medida es verosímil que ciertos eventos ocurran basándose en el patrón observado hasta este momento.

Por el contrario, la interpretación bayesiana se basa en un conocimiento limitado de las cosas. Afirma que sólo asocias una probabilidad a un evento porque hay incertidumbre sobre el mismo, es decir, porque no conoces todos los hechos. En realidad, un evento dado, o bien ocurrirá (probabilidad=100%) o bien no ocurrirá (probabilidad=0%). Cualquier otra cosa es una aproximación que hacemos del problema a partir de nuestro conocimiento incompleto del mismo. El enfoque bayesiano se basa por lo tanto en la idea de refinar predicciones a partir de nuevas evidencias. Un bayesiano definiría probabilidad como la expresión matemática que mide el nivel de conocimiento que tenemos para hacer una predicción. Por lo tanto, para un bayesiano, estrictamente hablando es incorrecto decir "predigo que hay un 30% de probabilidades de que ocurra el evento P", sino que debería decir "basándome en el conocimiento actual que tengo, tengo un 30% de certeza de que P ocurrirá".

2.2.1 ¿Importa tanto la diferencia?

La dos formas de entender una probabilidad son más relevantes de lo que pudiera parecer. Es muy diferente pensar en la probabilidad como el número de veces en que un experimento arrojará un resultado, a con-

cebirla como el grado de conocimiento que tengo sobre dicho resultado. En varios problemas, ambos enfoques pueden ser equivalentes. En otros, no tanto.

La aproximación bayesiana hace que algunas cosas sean más claras. Emplear una visión clásica lleva a muchas personas a confundir el significado de las estadísticas. Por ejemplo, si un estudio muestra que de cada 100 fumadores 10 van a desarrollar cáncer de pulmón, una interpretación clásica del dato puede llevarnos pensar que si fumo, tengo un 10% de probabilidad de enfermar. Y esto no es cierto. Lo que realmente nos dice el estudio es que el dato (10 de cada 100 fumadores desarrollan cáncer) puede considerarse como un buen punto de partida para hacer una predicción sobre las opciones de desarrollar cáncer si fumo, pero un bayesiano dejaría muy claro que se trata de un conocimiento muy incompleto, y que por lo tanto no es de mucha utilidad a no ser que disponga de más información para incrementar mi certeza.

Como contrapartida, el razonamiento bayesiano puede llevarnos al absurdo. Un bayesiano cree que puedes hacer un análisis probabilístico prácticamente de cualquier cosa, simplemente alineando los factores que pueden influir en ella y combinando adecuadamente tu conocimiento sobre esos factores. Esta idea ha sido empleada frecuentemente por auténticos obsesos del pensamiento bayesiano para asignar una probabilidad a la existencia de Dios, a que la tierra haya sido visitada por alienígenas, a que los artistas hayan incluido mensajes secretos en sus pinturas, etc.

Los frequentistas no son víctimas de este problema. Un frequentista considera que las probabilidades sólo tienen sentido en el ámbito de experimentos repetibles. Bajo este punto de visto, es absurdo asignar una probabilidad a la existencia de Dios, porque la existencia de un creador del universo no es un problema que podamos repetir y observar numerosas veces.

Los frequentistas no creen que sea posible asignar una probabilidad a absolutamente cualquier cosa, algo que a menudo es ridículo.

2.2.2 Y entonces, ¿quién tiene razón?

Probablemente ambos tienen razón. O ambos están equivocados. Hay escenarios en los que la idea de una probabilidad fija basada en un modelo repetible y observable, en un experimento controlado, es demasiado simple. Y otros casos en que la idea de una probabilidad que mide el grado de conocimiento de una cosa es igualmente simplista.

Como en tantas otras cosas de la ciencia ? y de la vida ? no existe un enfoque universalmente válido. La visión frecuentista ha aportado un rigor y un sistema de trabajo a muchos ámbitos de la estadística, empezando por el cálculo del error muestral. La visión bayesiana, por el contrario, ha mostrado ser una buena herramienta para hacer predicciones sobre sucesos complejos, tomando datos de múltiples fuentes de forma simultánea. También es la base de uno de los algoritmos más exitosos en el análisis de utilidad de estudios tipo Conjoint.

En cualquier caso, la polarización de las ideas no suele ser una buena consejera para tomar buenas decisiones. Y en el momento de emplear un enfoque frecuentista o bayesiano de un problema, no deberíamos olvidar esta máxima.

2.3 El problema de Monty Hall

El Problema de Monty Hall es un problema de probabilidad que está inspirado por el concurso televisivo estadounidense Let's Make a Deal (Hagamos un trato). , famoso entre 1963 y 1986. Su nombre proviene del presentador, Monty Hall.

En este concurso, el concursante escoge una puerta entre tres, y su premio consiste en lo que se encuentra detrás. Una de ellas oculta un coche, y tras las otras dos hay una cabra. Sin embargo, antes de abrirla, el presentador, que sabe donde esta el premio, abre una de las otras dos puertas y muestra que detrás de ella hay una cabra. Ahora tiene el concursante una última oportunidad de cambiar la puerta escogida ¿Debe el concursante mantener su elección original o escoger la otra puerta? ¿Hay alguna diferencia?

¿Cuál sería la opción correcta?

1. Quedarse con la puerta inicial
2. Cambiar a la otra puerta
3. Es irrelevante cambiar o no cambiar

A primera vista parece obvio que da igual (opción 3). La intuición nos dice que ahora, quitando una puerta sin premio, la puerta que nosotros escogimos tiene un 50% de tener una cabra y por tanto da igual cambiar que no hacerlo. Pero no sería una paradoja o problema si fuera tan trivial, ¿verdad?.

Suceso	Descripción
A	El jugador selecciona la puerta que contiene el coche en su selección i
B	El jugador selecciona una puerta que contiene una cabra en su selecció
G	El jugador gana el coche.

2.3.1 Explicación intuitiva

Trataremos de verlo de esta forma:
Si no cambiamos las posibilidades de ganar son de $\frac{1}{3}$, ya que escogemos una vez sin tener informacion y luego no cambiamos, de modo que el hecho de que el presentador abra una puerta no cambia nuestras probabilidades aunque parezca lo contrario.
Sin embargo si cambiamos:
Escogemos puerta con cabra → Presentador muestra la otra cabra → cambiamos y GANAMOS
Escogemos puerta con coche → Presentador muestra la otra cabra → cambiamos y PERDEMOS
y dado que hay 2 cabras y 1 coche las posibilidades de ganar son de $\frac{2}{3}$.

2.3.2 Explicación matemática

Lo explicaremos matemáticamente, con probabilidades condicionadas. Esta es la forma más rigurosa pero probablemente la que peor se entiende.

Definimos cuidadosamente los siguientes sucesos . Asumimos que hay dos tipos de jugador, los que nunca cambian de puerta y los que cambian siempre; en este caso la pregunta se limita a ver que tipo de jugador tiene la mayor probabilidad de ganar el coche.

Estamos interesados en calcular $P(G)$ para cada tipo de jugador.

Para calcular $P(G)$, basta con notar que $G = (G \cap A) \cup (G \cap B)$ ya que $A \cap B = \emptyset$ y $A \cup B = \Omega$ (esto es equivalente a decir que $\{A, B\}$ es una partición de Ω)

$$P(G) = P((G \cap A) \cup (G \cap B)) = P(G \cap A) + P(G \cap B) = P(\frac{G}{A})P(A) + P(\frac{G}{B})P(B)$$

En cualquier caso, dado que no tenemos ninguna razón para pensar lo contrario, diremos que $P(A) = \frac{1}{3}$ y $P(B) = \frac{2}{3}$ pues hay un coche y dos cabras.

Ahora debemos definir que tipo de jugador estamos estudiando.

- Jugador que nunca se cambia.

En este caso $P(G|A) = 1$ y $P(G|B) = 0$ pues el jugador se queda con su selección inicial.

Por lo tanto $P(G) = \frac{1}{3}$.

- Jugador que siempre se cambia.

En este caso $P(G|A) = 0$ y $P(G|B) = 1$ pues el jugador se cambia a la única puerta cerrada que queda (y sabemos que como el presentador sabe donde está el coche, siempre mostrará una cabra).

Por lo tanto $P(G) = \frac{2}{3}$.

Claramente la mejor estrategia es cambiar siempre, pues la probabilidad efectiva de ganar es el doble de la correspondiente al jugador que no cambia nunca.

2.4 La falacia del apostador

La falacia del apostador o falacia de Montecarlo es una falacia lógica por la que se cree erróneamente que los sucesos pasados afectan a los futuros en lo relativo a actividades aleatorias, como en muchos juegos de azar. Puede comprender las siguientes ideas equivocadas:

- Un suceso aleatorio tiene más probabilidad de ocurrir porque no ha ocurrido durante cierto período
- Un suceso aleatorio tiene menos probabilidad de ocurrir porque ha ocurrido durante cierto período
- Un suceso aleatorio tiene más probabilidad de ocurrir si no ocurrió recientemente
- Un suceso aleatorio tiene menos probabilidad de ocurrir si ocurrió recientemente

Las anteriores son ideas equivocadas que surgen cotidianamente en razonamientos sobre probabilidades, muchos de los cuales se han estudiado con gran profundidad. Mucha gente pierde dinero apostando debido a su creencia errónea en esta falacia.

Sencillamente, las probabilidades de que algo suceda la próxima vez no están necesariamente relacionadas con lo que ya sucedió, especialmente en muchos juegos de azar. Esto suele resumirse en la frase "Los dados (o moneda) no tiene memoria", pues su naturaleza es la misma, independiente del número de tiros y resultados previos.

La falacia del apostador se puede ejemplificar considerando el experimento de lanzar una moneda sin truco al aire. Los resultados de cada lanzamiento son estadísticamente independientes y la probabilidad de

obtener cara en un solo lanzamiento es de exactamente $\frac{1}{2}$, luego entonces, la probabilidad de obtener dos caras consecutivas en dos lanzamientos es $\frac{1}{4}$, la probabilidad de que salgan tres caras seguidas es de $\frac{1}{8}$ y en general, si decimos que A_i es el evento número i en el que sale cara entonces tenemos,

$$P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i) = \frac{1}{2^n}$$

Ahora supongamos que hemos tirado la moneda cuatro veces y nos han salido cuatro caras consecutivamente, si en el siguiente lanzamiento obtenemos cara de nuevo tendremos cinco caras seguidas; la probabilidad de que cinco caras seguidas salgan es de $\frac{1}{32}$, por lo que una persona haciendo un razonamiento como el de la falacia del apostador podría concluir que la probabilidad de que en el siguiente lanzamiento es más probable que el resultado sea cruz, sin embargo, esto no es correcto porque el evento en el que ocurren las cinco caras consecutivamente y el evento donde ocurren cuatro caras seguidas de una cruz tienen la misma probabilidad de $\frac{1}{32}$. Dado que nuestros primeros cuatro lanzamientos resultaron cara, la probabilidad de que el siguiente lanzamiento sea cara es,

$$P(A_5|A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_5) = \frac{1}{2}$$

2.5 Problema de las ovejas y el lobo

There are 100 equally spaced points around a circle. At 99 of the points, there are sheep, and at 1 point, there is a wolf. At each time step, the wolf randomly moves either clockwise or counterclockwise by 1 point. If there is a sheep at that point, he eats it. The sheep don't move. What is the probability that the sheep who is initially opposite the wolf is the last one remaining?

Solución: Supongamos que la oveja que está justamente enfrente del lobo es la última en sobrevivir, entonces ambas vecinas de esta oveja tienen que ser devoradas antes. Sin embargo, también es cierto que, después de devorar a la primera vecina, el lobo debe ir sobre todo el círculo para devorar a la otra vecina. Contrario a lo que se podría percibir, la probabilidad de que el lobo haga lo antes mencionado o se coma a nuestra oveja de una vez es la misma, dado que en cada intervalo de tiempo

el lobo puede igualmente moverse hacia la derecha o la izquierda, lo mismo que cuando se lanza una moneda. Dicho lo anterior, la probabilidad de que la oveja que comienza justamente frente al lobo, o de hecho cualquiera de las ovejas en el círculo, de ser la última sobreviviente es de $\frac{1}{99}$

Variable aleatoria

3.1 Introducción

Una variable aleatoria o variable estocástica es una función que asigna un valor, usualmente numérico, al resultado de un experimento aleatorio. Por ejemplo, los posibles resultados de tirar un dado dos veces: (1, 1), (1, 2), etc. o un número real (p.e., la temperatura máxima medida a lo largo del día en una ciudad concreta).

Los valores posibles de una variable aleatoria pueden representar los posibles resultados de un experimento aún no realizado, o los posibles valores de una cantidad cuyo valor actualmente existente es incierto (p.e., como resultado de medición incompleta o imprecisa). Intuitivamente, una variable aleatoria puede tomarse como una cantidad cuyo valor no es fijo pero puede tomar diferentes valores; una distribución de probabilidad se usa para describir la probabilidad de que se den los diferentes valores. En términos formales una variable aleatoria es una función definida sobre un espacio de probabilidad.

Las variables aleatorias suelen tomar valores reales, pero se pueden considerar valores aleatorios como valores lógicos, funciones o cualquier tipo de elementos (de un espacio medible). El término elemento aleatorio se utiliza para englobar todo ese tipo de conceptos relacionados. Un concepto relacionado es el de proceso estocástico, un conjunto de variables aleatorias ordenadas (habitualmente por orden o tiempo).

3.1.1 Tipos de variables aleatorias

Para comprender de una manera más amplia y rigurosa los tipos de variables, es necesario conocer la definición de conjunto discreto. Un conjunto es discreto si está formado por un número finito de elementos, o si sus elementos se pueden enumerar en secuencia de modo que haya un primer elemento, un segundo elemento, un tercer elemento, y así sucesivamente⁵ (es decir, un conjunto infinito numerable sin puntos

de acumulación). Para variables con valores en \mathbb{R} las variables aleatorias se clasifican usualmente en:

Variable aleatoria discreta: una v.a. es discreta si su recorrido es un conjunto discreto. La variable del ejemplo anterior es discreta. Sus probabilidades se recogen en la función de cuantía.

Variable aleatoria continua: una v.a. es continua si su recorrido es un conjunto no numerable. Intuitivamente esto significa que el conjunto de posibles valores de la variable abarca todo un intervalo de números reales. Por ejemplo, la variable que asigna la estatura a una persona extraída de una determinada población es una variable continua ya que, teóricamente, todo valor entre, pongamos por caso, 0 y 2,50 m, es posible.

Las definiciones anteriores pueden generalizarse fácilmente a variables aleatorias con valores sobre \mathbb{R}^n o \mathbb{C}^n . Esto no agota el tipo de variables aleatorias ya que el valor de una variable aleatoria puede ser también una partición, como sucede en el proceso estocástico del restaurante chino o el conjunto de valores de una variable aleatoria puede ser un conjunto de funciones como el proceso estocástico de Dirichlet.

3.1.2 Distribución de probabilidad de una variable aleatoria

La distribución de probabilidad de una v.a. X , también llamada función de distribución de X es la función $F_X(x)$, que asigna a cada evento definido sobre X una probabilidad dada por la siguiente expresión:

$$F_X(x) = P(X \leq x)$$

Y de manera que se cumplan las siguientes tres condiciones:

$$\lim_{x \rightarrow -\infty} F(x) = 0 \text{ y } \lim_{x \rightarrow \infty} F(x) = 1$$

Es continua por la derecha.

Es monótona no decreciente.

La distribución de probabilidad de una v.a. describe teóricamente la forma en que varían los resultados de un experimento aleatorio. Intuitivamente se trataría de una lista de los resultados posibles de un experimento con las probabilidades que se esperarían ver asociadas con cada resultado.

3.1.3 Función de densidad de una v.a. continua

La función de densidad de probabilidad (FDP) o, simplemente, función de densidad, representada comúnmente como $f(x)$, se utiliza con el propósito

de conocer cómo se distribuyen las probabilidades de un suceso o evento, en relación al resultado del suceso.

La FDP es la derivada (ordinaria o en el sentido de las distribuciones) de la función de distribución de probabilidad $F(x)$, o de manera inversa, la función de distribución es la integral de la función de densidad:

$$F(x) = \int_{-\infty}^x f(t) dt$$

La función de densidad de una v.a. determina la concentración de probabilidad alrededor de los valores de una variable aleatoria continua.

3.2 Teorema del límite central

El teorema del límite central o teorema central del límite indica que, en condiciones muy generales, si S_n es la suma de n variables aleatorias independientes y de varianza no nula pero finita, entonces la función de distribución de S_n "se aproxima bien" a una distribución normal (también llamada distribución gaussiana, curva de Gauss o campana de Gauss). Así pues, el teorema asegura que esto ocurre cuando la suma de estas variables aleatorias e independientes es lo suficientemente grande.

Sea $X_1, X_2, X_3, \dots, X_n$ un conjunto de variables aleatorias, independientes e idénticamente distribuidas, con media μ y varianza $0 < \sigma^2 < \infty$ Sea

$$S_n = X_1 + \dots + X_n \text{ entonces } \lim_{n \rightarrow +\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right) = \Phi(z)$$

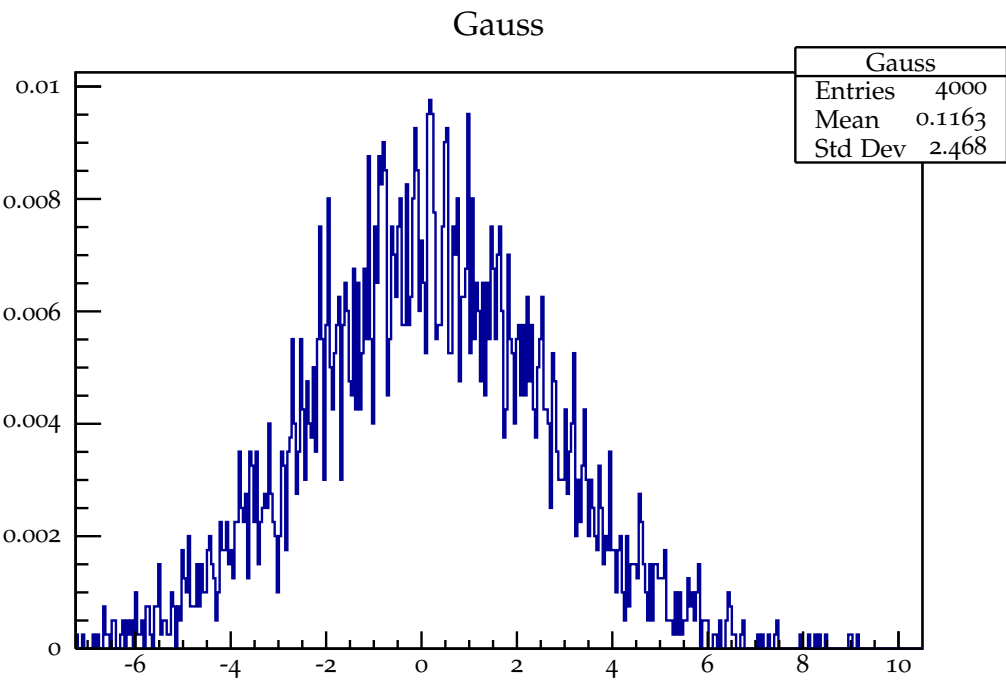
Propiedades

- El teorema del límite central garantiza una distribución normal cuando n es suficientemente grande.
- Existen diferentes versiones del teorema, en función de las condiciones utilizadas para asegurar la convergencia. Una de las más simples establece que es suficiente que las variables que se suman sean independientes, idénticamente distribuidas, con valor esperado y varianza finitas.
- La aproximación entre las dos distribuciones es, en general, mayor en el centro de las mismas que en sus extremos o colas, motivo por el cual se prefiere el nombre "teorema del límite central" ("central"

califica al límite, más que al teorema).

- Este teorema, perteneciente a la teoría de la probabilidad, encuentra aplicación en muchos campos relacionados, tales como la inferencia estadística o la teoría de renovación.

3.3 Práctica: Realizar un histograma en ROOT que se llene con la distribución de Gauss.



3.4 Mínimos cuadrados

Mínimos cuadrados es una técnica de análisis numérico enmarcada dentro de la optimización matemática, en la que, dados un conjunto de pares ordenados (variable independiente, variable dependiente) y una familia de funciones, se intenta encontrar la función continua, dentro de dicha familia, que mejor se aproxime a los datos (un “mejor ajuste”), de acuerdo con el criterio de mínimo error cuadrático.

En su forma más simple, intenta minimizar la suma de cuadrados de las diferencias en las ordenadas (llamadas residuos) entre los puntos generados por la función elegida y los correspondientes valores en los datos. Específicamente, se llama mínimos cuadrados promedio (LMS) cuando el número de datos medidos es 1 y se usa el método de descenso por gradiente para minimizar el residuo cuadrado. Se puede demostrar que LMS minimiza el residuo cuadrado esperado, con el mínimo de operaciones (por iteración), pero requiere un gran número de iteraciones para converger.

Desde un punto de vista estadístico, un requisito implícito para que funcione el método de mínimos cuadrados es que los errores de cada medida estén distribuidos de forma aleatoria. El teorema de Gauss-Márkov prueba que los estimadores mínimos cuadráticos carecen de sesgo y que el muestreo de datos no tiene que ajustarse, por ejemplo, a una distribución normal. También es importante que los datos a procesar estén bien escogidos, para que permitan visibilidad en las variables que han de ser resueltas (para dar más peso a un dato en particular, véase mínimos cuadrados ponderados).

La técnica de mínimos cuadrados se usa comúnmente en el ajuste de curvas. Muchos otros problemas de optimización pueden expresarse también en forma de mínimos cuadrados, minimizando la energía o maximizando la entropía.

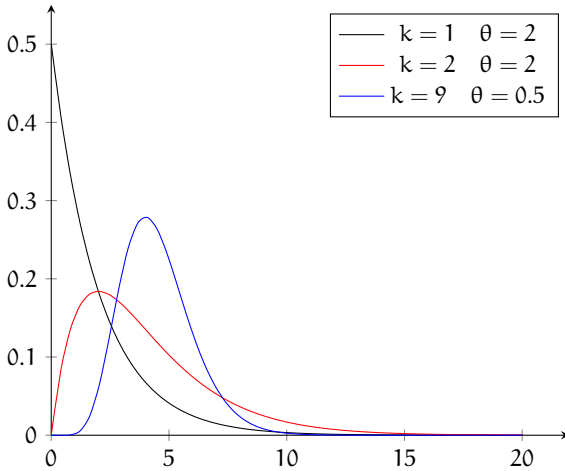
3.5 Función de densidad de probabilidad

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 . Se define como el cociente entre la probabilidad de X de tomar un valor en el intervalo $[x, x + dx]$ y dx , siendo dx un infinitesimal. La mayoría de las funciones de densidad de probabilidad requieren uno o más parámetros para especificarlas totalmente. Recíprocamente respecto de la definición ya desarrollada, pueden hacerse las siguientes consideraciones.

La probabilidad de que una variable aleatoria continua X quede ubicada entre los valores a y b está dada por el desenvolvimiento en el intervalo de la FDP; de los valores comprendidos en el rango entre a y b .

3.6 Cuota estadística (Odds)

En estadística, la cuota es el inverso de la probabilidad, de forma que cuanto mayor es la cuota menor es la probabilidad teórica de que aparezca ese resultado.¹

La cuota a favor de un evento o proposición se calcula mediante la fórmula

$$\frac{p}{(1 - p)}$$

dónde p es la probabilidad del evento o proposición. La cuota en contra del mismo evento se calcula mediante la fórmula

$$\frac{(1 - p)}{p}$$

Por ejemplo, al escoger al azar un día de la semana (7 días), la cuota asociada a favor de un domingo cualquiera es

	Casos	No casos	
Presente	a	b	Total de expuestos(mi)
Ausente	c	d	Total de no expuestos(n)
	Total de casos(ni)	Total de no casos(no)	Población total(n)

$$\frac{(1/7)}{(1-(1/7))} = \frac{1/7}{6/7} = \frac{1}{6}$$

no 1/7, como podría parecer. La cuota en contra del mismo domingo es 6/1 = 6; eso significa que es 6 veces más probable que no sea domingo a que sí sea domingo. Estas cuotas son relativas a la probabilidad de suceder.

Generalmente las cuotas no son determinadas por el público en general a causa de la confusión natural que se tiene con la probabilidad de un suceso expresada de forma fraccionaria. De esta forma, la probabilidad al escoger al azar un domingo respecto a todos los días de la semana es de uno a siete (1/7). Un corredor de apuestas (para sus propios propósitos) utilizará las cuotas en formato uno contra seis, expresada generalmente por la mayoría de personas como 6 a 1, 6-1, o 6/1 (leído como seis a uno) donde la primera cifra representa el número de días contrarios al éxito del suceso, y la segunda cifra corresponde al los días a favor del éxito del suceso (cuotas a favor). En otras palabras, un evento con m a n de cuotas en contra, tendrá una probabilidad de n/(m + n), mientras que un evento de m a n de cuotas a favor, la probabilidad será de m/(m + n). La teoría de la probabilidad expresa que las cuotas juegan un papel más natural o más conveniente que las probabilidades.

Para algunos juegos de azar, puede ser también la forma más conveniente para que el apostante entienda mejor cuantas ganancias tendrá si su selección es la correcta: al individuo le pagaran seis por cada unidad apostada. Por ejemplo, una apuesta ganadora de \$10 se pagará a 6 x \$10 = \$60, retornando también los \$10 originales de la apuesta.

La razón de momios (RM), razón de oportunidades o razón de probabilidades -en inglés, odds ratio (OR)- es una medida estadística utilizada en estudios epidemiológicos transversales y de casos y controles, así como en los metaanálisis. En términos formales, se define como la posibilidad de que una condición de salud o enfermedad se presente en un grupo de población frente al riesgo de que ocurra en otro. En epidemiología, la comparación suele realizarse entre grupos humanos que presentan condiciones de vida similares, con la diferencia de que uno se encuentra expuesto a un factor de riesgo (mi) mientras que el otro carece de esta característica (mo). Por lo tanto, la razón de momios o de posibilidades es una medida de tamaño de efecto.

Para mostrar la definición de la razón de momios, se puede hacer referencia a la tabla estándar de 2x2.

En un estudio transversal, la razón de momios es el cociente entre la odds de exposición observada en los casos (enfermos) a/b y la odds de exposición del grupo control c/d :

$$OR = \frac{a/b}{c/d}$$

En un estudio de casos y controles, la razón de momios es el cociente entre la oportunidad de enfermedad del grupo expuesto (o en el grupo tratado) a/b y la oportunidad de enfermedad del grupo no expuesto (o no tratado) c/d (los casos serían los enfermos al final del estudio y los controles los no enfermos):

$$OR = \frac{a/b}{c/d}$$

Tanto la fórmula de la RM de un estudio de cohorte como la de uno de casos y controles puede leerse como:

$$OR = \frac{a.d}{b.c}$$

Lo que en definitiva no es más que una forma de expresar la proporción de veces que un suceso ocurra frente a que no ocurra. De tal manera que un $OR = 2,5$ debemos leerlo como 2,5:1, mejor aún dado que un efecto aparece ante la presencia de otra variable es de 2,5 veces más que si esta variable no está presente.

Si se piensa que el $OR = 1$, esto significa que la cantidad de veces que el evento ocurra ante la presencia de otra variable vs. las veces que ocurra en ausencia de esa variable, o sea 1:1. Lo que es lo mismo que decir que aparecerá tantas veces cuando la variable esté presente como cuando la variable no se presente.

Pensado esto así, no siempre es fácil traducirlo en probabilidades, una forma mucho más fácil de entender para quienes hablamos el castellano. De modo tal que podemos transformar el OR en probabilidades a partir de la fórmula:

$$\text{Probabilidad} = \frac{RM}{RM + 1}$$

En este caso si el OR fue de 2,5 entonces aplicando la fórmula la probabilidad es de 0,714, o lo que es igual del 71,4%. Mientras que en el caso del $OR = 1$, la probabilidad es del 50%, es decir que existen en este último caso las mismas probabilidades que el evento ocurra estando o no la otra variable en estudio presente.

Existe cierta confusión al homologarlo con el riesgo relativo, que se usa en estudios prospectivos o de cohorte, el cual en realidad está comparando dos tasas de incidencia (o probabilidades acumuladas), una con el factor de exposición presente y otra con el factor de exposición ausente.

3.7 Curtosis

En teoría de la probabilidad y estadística, la curtosis es una medida de la forma. Así, las medidas de curtosis tratan de estudiar la proporción de la varianza que se explica por la combinación de datos extremos respecto a la media en contraposición con datos poco alejados de la misma. Una mayor curtosis implica una mayor concentración de datos muy cerca de la media de la distribución coexistiendo al mismo tiempo con una relativamente elevada frecuencia de datos muy alejados de la misma. Esto explica una forma de la distribución de frecuencias con colas muy elevadas y con un centro muy apuntado.

Un coeficiente de apuntamiento o de curtosis es el basado en el cuarto momento con respecto a la media y se define como:

$$\beta_2 = \frac{\mu_4}{\sigma^4}$$

donde μ_4 es el 4º momento centrado o con respecto a la media y σ es la desviación estándar.

No obstante, está más extendida la siguiente definición del coeficiente de curtosis:

$$g_2 = \frac{\mu_4}{\sigma^4} - 3$$

donde al final se ha sustraído 3 (que es la curtosis de la Normal) con objeto de generar un coeficiente que valga o para la Normal y tome a ésta como referencia de apuntamiento:

Tomando, pues, la distribución normal como referencia, una distribución puede ser:

- más apuntada y con colas menos anchas que la normal - leptocúrtica.
- menos apuntada y con colas más anchas que la normal - platicúrtica.
- la distribución normal es mesocúrtica.

En la distribución normal se verifica que $\mu_4 = 3\sigma^4$, donde μ_4 es el momento de orden 4 respecto a la media y σ la desviación típica.

Así tendremos que:

- Si la distribución es leptocúrtica $\beta_2 > 3$ y $g_2 > 0$
- Si la distribución es platicúrtica $\beta_2 < 3$ y $g_2 < 0$
- Si la distribución es mesocúrtica $\beta_2 = 3$ y $g_2 = 0$

Otra forma de medir la curtosis se obtiene examinando la fórmula de la curtosis de la suma de variables aleatorias. Si Y es la suma de n variables aleatorias estadísticamente independientes, todas con igual distribución X , entonces $\text{Kurt}[Y] = \frac{\text{Kurt}[X]}{n}$, complicándose la fórmula si la curtosis se hubiese definido como $\frac{\mu_4}{\sigma^4}$.

3.8 Skewness

Las medidas de asimetría son indicadores que permiten establecer el grado de simetría (o asimetría) que presenta una distribución de probabilidad de una variable aleatoria sin tener que hacer su representación gráfica.

Como eje de simetría consideramos una recta paralela al eje de ordenadas que pasa por la media de la distribución. Si una distribución es simétrica, existe el mismo número de valores a la derecha que a la izquierda de la media, por tanto, el mismo número de desviaciones con signo positivo que con signo negativo. Decimos que hay asimetría positiva (o a la derecha) si la "cola" a la derecha de la media es más larga que la de la izquierda, es decir, si hay valores más separados de la media a la derecha. Diremos que hay asimetría negativa (o a la izquierda) si la "cola" a la izquierda de la media es más larga que la de la derecha, es decir, si hay valores más separados de la media a la izquierda.

3.8.1 Coeficiente de asimetría de Fisher

En teoría de la probabilidad y estadística, la medida de asimetría más utilizada parte del uso del tercer momento estándar. La razón de esto es que nos interesa mantener el signo de las desviaciones con respecto a la media, para obtener si son mayores las que ocurren a la derecha de la media que las de la izquierda. Sin embargo, no es buena idea tomar el momento estándar con respecto a la media de orden 1. Debido a que una simple suma de todas las desviaciones siempre es cero. En efecto, si por ejemplo, los datos están agrupados en k clases, se tiene que:

$$\sum_{i=1}^k f_i(x_i - \mu) = \sum_{i=1}^k f_i x_i - \mu \sum_{i=1}^k f_i = \mu - \mu = 0$$

en donde x_i representa la marca de la clase i -ésima y f_i denota la frecuencia relativa de dicha clase. Por ello, lo más sencillo es tomar las desviaciones al cubo.

El coeficiente de asimetría de Fisher, representado por γ_1 , se define como:

$$\gamma_1 = \frac{\mu_3}{\sigma^3},$$

donde μ_3 es el tercer momento en torno a la media y σ es la desviación estándar.

Si $\gamma_1 > 0$, la distribución es asimétrica positiva o a la izquierda.

Si $\gamma_1 < 0$, la distribución es asimétrica negativa o a la derecha.

Si la distribución es simétrica, entonces sabemos que $\gamma_1 = 0$. El recíproco no es cierto: es un error común asegurar que si $\gamma_1 = 0$ entonces la dis-

tribución es simétrica (lo cual es falso).

3.8.2 Coeficiente de asimetría de Pearson

Sólo se puede utilizar en distribuciones uniformes, unimodales y moderadamente asimétricas. Se basa en que en distribuciones simétricas la media de la distribución es igual a la moda.

$$A_p = \frac{\mu - \text{moda}}{\sigma},$$

donde μ es el momento central de orden 1, que corresponde a la media aritmética de la variable X .

Si la distribución es simétrica, $\mu = \text{moda}$ y $A_p = 0$. Si la distribución es asimétrica positiva la media se sitúa por encima de la moda y, por tanto, $A_p > 0$.

Coeficiente de asimetría de Bowley

Está basado en la posición de los cuartiles y la mediana, y utiliza la siguiente expresión:

$$A_B = \frac{Q_{3/4} + Q_{1/4} - 2Me}{Q_{3/4} - Q_{1/4}}$$

En una distribución simétrica el tercer cuartil estará a la misma distancia de la mediana que el primer cuartil. Por tanto $A_B = 0$.

Si la distribución es positiva o a la derecha, $A_B > 0$.

La asimetría resulta útil en muchos campos. Muchos modelos simplistas asumen una distribución normal, esto es, simétrica en torno a la media. La distribución normal tiene una asimetría cero. Pero en realidad, los valores no son nunca perfectamente simétricos y la asimetría de la distribución proporciona una idea sobre si las desviaciones de la media son positivas o negativas. Una asimetría positiva implica que hay más valores distintos a la derecha de la media.

Las medidas de asimetría, sobre todo el coeficiente de asimetría de Fisher, junto con las medidas de apuntamiento o curtosis se utilizan para contrastar si se puede aceptar que una distribución estadística sigue la distribución normal. Esto es necesario para realizar numerosos contrastes estadísticos en la teoría de inferencia estadística.

3.9 Programa: Congruencia Zeller

Makefile

```
zeller: main.cpp zeller.cpp
    gcc -o zeller main.cpp zeller.cpp -I. -lstdc++
```

zeller.h

```
/******
 * zeller.h - Header with support routines for Zeller's
 *           congruence formula
 * Author: Aldo Miranda-Aguilar
 *
 * March 2016
 *
 *****/
#ifndef ZELLER_H
#define ZELLER_H

int zeller(int d, int m, int y);

#endif
```

zeller.cpp

```
/******
 * zeller.cpp - Implementation of Zeller's
 *           congruence formula
 * Author: Aldo Miranda-Aguilar
 *
 * March 2016
 *
 *****/
#include <cmath>

int zeller(int d, int m, int y) {
    if (m == 1 || m == 2) {
        m = m + 12;
        y = y - 1;
    }
    return (d + (int)floor((13 * (m + 1)) / 5) + y %100 +
            (int)floor((y%100)/ 4) + (int)floor(((int)floor(y/100))/4) +
            5*(int)floor(y/100)) % 7;
}
```

main.cpp

```
#include <iostream>
#include <string>
#include "zeller.h"

using namespace std;

int main(int argc, char** argv) {
    int m, d, y, daynum;
    char ans;
    do {
        cout << "Ingresa una fecha (mm dd aaaa): \n";
        cin >> d >> m >> y;
        daynum = zeller(m, d, y);

        switch (daynum) {
            case 2:
                cout << m << "/" << d << "/" << y << " LUNES \n";
                break;
            case 3:
                cout << m << "/" << d << "/" << y << " MARTES \n";
                break;
            case 4:
                cout << m << "/" << d << "/" << y << " MIERCOLES \n";
                break;
            case 5:
```

```

        cout << m << "/" << d << "/" << y << "JUEVES\n";
        break;
    case 6:
        cout << m << "/" << d << "/" << y << "VIERNES\n";
        break;
    case 0:
        cout << m << "/" << d << "/" << y << "SABADO\n";
        break;
    case 1:
        cout << m << "/" << d << "/" << y << "DOMINGO\n";
        break;
    }
    cout << "?Otra_fecha?(s/n)\n";
    cin >> ans;
} while ((ans == 's') || (ans == 'S'));

return 0;
}

```

3.10 Is Economics the Next Physical Science?

En la última década o así, los físicos han comenzado a hacer la investigación académica en economía. Tal vez un centenar de personas están ahora involucrados activamente en un campo emergente a menudo llamado econofísica, y dos nuevas revistas y conferencias frecuentes se dedican al campo. Al menos diez libros se han escrito recientemente sobre econofísica en general o sobre subtemas específicos. departamentos de física en todo el mundo están concediendo tesis de doctorado de investigación en economía, y en Europa varios profesores de los departamentos de física se especializan en econofísica. La consultora internacional McKinsey y Co patrocina un nuevo premio de investigación anual, el Premio Joven Científico-for Socio y Econofísica. Es toda esta actividad sólo una moda, o es que ocurra algo más sustancial?

Si los físicos quieren hacer investigación en economía, ¿por qué no acaba de obtener títulos en economía en el primer lugar? ¿Por qué no la econophysicist rediseñar, encontrar puestos de trabajo en los departamentos de economía, y publicar en revistas económicas tradicionales? Tal vez el crecimiento de econofísica es sólo un fenómeno temporal, impulsado por una generación de físicos que han hecho malas elecciones de carrera. ¿Hay alguna razón por la cual los miembros de los departamentos de economía deben prestar atención a los métodos de la física? ¿Qué ventaja, en su caso, se confiere por una formación en física? Y lo más importante, ¿cómo econofísica difieren de la economía, y qué contribución única que puede hacer?

3.10.1 Física social

La participación de los físicos en la ciencia social tiene una larga historia, que se remonta al menos a Daniel Bernoulli, que se presentó en 1738 la idea de utilidad para describir las preferencias de las personas. Pierre-Simon Laplace, en su *Essai philosophique sur les probabilités* (1812),

señalaron que los eventos que pueden parecer aleatorios e impredecibles, tales como el número de cartas en la oficina de la letra muerta de París, pueden ser bastante predecible y puede ser demostrado obedecer las leyes simples. Las ideas de Laplace se amplificaron aún por Adolphe Quetelet, que era un estudiante de Joseph Fourier y que estudió la existencia de patrones en conjuntos de datos que van desde la frecuencia de los diferentes métodos para cometer un asesinato con el tamaño del pecho de los hombres de Escocia. Fue Quetelet que, en 1835, acuñó el término "física social". Las analogías a la física juegan un papel importante en el desarrollo de la teoría económica a través del siglo 19, y algunos de los fundadores de la teoría económica neoclásica, incluyendo Irving Fisher, fueron originalmente entrenados como los físicos; Fisher era un estudiante de Willard Gibbs. En 1938, Ettore Majorana describió proféticamente tanto las oportunidades y las dificultades en la aplicación de métodos estadísticos en la física a las ciencias sociales.

La gama de temas que han sido abordados por los físicos abarca muchas áreas diferentes de la economía. Finanzas está particularmente bien representado (ver el artículo de José Pimbley, *HOY FÍSICA*, enero de 1997, página 42); temas de muestra incluyen la observación empírica de regularidades en los datos del mercado, la dinámica de la formación de los precios, la comprensión de las burbujas y los pánicos, los métodos de fijación de precios para las opciones y otros derivados, y la construcción de portafolios óptimos. temas más amplios en la economía incluyen la distribución del ingreso, la aparición del dinero, y las implicaciones de la simetría y de escala para el funcionamiento del mercado. Creemos que la unión de los métodos de la física y la economía, y la colaboración entre físicos y economistas, puede añadir valor a la ciencia económica. Sin embargo, la sobreventa ese punto de vista tiene sus peligros; econofísica está lejos de ser bien establecida.

A pesar de la larga historia de la asociación de los campos de los, la contribución sustancial de la física a la economía se encuentra todavía en una etapa temprana, y creemos que es descabellado predecir lo que en última instancia puede lograr. Casi con toda seguridad, aspectos "físicos" de las teorías del orden social serán no simplemente recapitular las teorías existentes en la física, aunque ya parece que hay solapamientos. El desarrollo de las sociedades y las economías puede estar supeditada a los accidentes de la historia y en cada giro gira en torno a los aspectos complejos de la conducta humana.

No obstante, en huelga regularidades empíricas sugieren que al menos un poco de orden social no es históricamente contingente, y es quizás predecible a partir de primeros principios. El papel de los mercados como mediadores de la comunicación y la computación distribuida, subyacentes en los procesos colectivos de formación de precios y la asignación de recursos, y la aparición de las instituciones sociales que apoyan esas funciones, son fenómenos económicos quintaesencia. Sin embargo, las nociones de comunicación mercados o las capacidades computa-

cionales, y la forma en que las diferencias en los *capacities* representan la estabilidad y la sucesión histórica de los mercados, naturalmente, puede ser parte del mundo físico con sus dinámicas sociales humanas.

Mercados y otras instituciones económicas traen consigo conceptos de eficiencia o la optimización de los deseos humanos satisfactorias. Aunque intuitivamente atractiva, tales ideas han demostrado ser difícil de formalizar, incluso si se ha hecho algún progreso. Al igual que con la mayoría de las nuevas áreas de investigación física, esperamos que los objetivos finales de una economía física se declararán en retrospectiva, de los éxitos en la identificación, medición, modelado, y en algunos casos la predicción de regularidades empíricas.

3.10.2 La búsqueda de la universalidad

Los economistas suelen ser mejor entrenados que los físicos en el análisis estadístico, por lo que los físicos podrían parecer tener poco que aportar en esa zona. Sin embargo, las diferencias en las metas y la filosofía son importantes. La física es impulsado por la búsqueda de leyes universales. En parte debido a la extrema complejidad de los fenómenos sociales, que la búsqueda se ha abandonado en gran medida en nuestro mundo posmoderno, donde filosofías relativistas de la ciencia disfrutaron de la aceptación generalizada de forma preocupante propagación. Hoy en día, el trabajo en las ciencias sociales es en gran medida centrado en la documentación de diferencias.

Aunque la tendencia es mucho menos evidente en economía, un papel típico de la economía financiera, por ejemplo, podría estudiar la diferencia entre la Nueva York y las bolsas de valores NASDAQ, o el efecto de cambiar el tamaño de la señal, o el más pequeño cambio, de los precios. Los físicos han entrado en el campo, tal vez ingenuamente, con nuevos ojos y nuevas hipótesis, y han analizado los datos económicos con el objetivo de encontrar regularidades generalizados; hacen hincapié en lo que podría ser común a todos los mercados en lugar de lo que los hace diferentes. Dicho trabajo ha sido motivado de manera oportunista por la existencia de grandes conjuntos de datos como principales historias completas de transacciones de cambio que abarcan años y que a veces contienen cientos de millones de eventos.

3.10.3 Memorias distantes

La coexistencia de la oferta y la demanda, con la eficiencia del mercado, la memoria a largo crea una *zle puz-* aún sin resolver. procesos de larga memoria son altamente predecibles. Desde la entrada de los compradores tiende a subir los precios, y la entrada de los vendedores tiende a empujar hacia abajo, la larga memoria de oferta y demanda

sugiere que el cambio de precios también debe ser la memoria de largo. Pero eso sería incompatible con la eficiencia informativa. Para evitar tales incompatibilidades, los agentes del mercado deben ajustar de alguna manera colectiva su comportamiento: Podrían, por ejemplo, crear una respuesta asimétrica de los precios de manera que con un exceso de compradores, la respuesta de los precios de las órdenes de compra es menor de lo que es vender pedidos. ¿Cómo y por qué viene que la asimetría acerca sigue siendo un misterio, tal vez relacionada con la causa de la volatilidad en clúster.

3.10.4 Modelar el comportamiento de los agentes

La diferencia fundamental entre un sistema físico y una economía es una economía que está habitado por personas, que tienen interacciones estratégicas. Debido a que la gente piensa y planificar, y luego tomar decisiones basadas en sus planes, que son mucho más complicados que los átomos (véase el artículo de George Ellis, *HOY FÍSICA*, julio de 2005, página 49). Como consecuencia de ello, las técnicas matemáticas y la filosofía de modelado en la economía han ido distanciando de los de la física. A pesar de esta divergencia es claramente necesaria, muchos físicos podrían argumentar que la brecha es más amplia de lo que debería ser.

3.10.5 Racionalidad imperfecta

En los últimos 20 años, la economía ha comenzado a cuestionar los supuestos de la elección racional y mercados perfectos mediante el modelado de imperfecciones tales como la información asimétrica, la estructura del mercado incompleto, y la racionalidad limitada. Varias nuevas escuelas de pensamiento han surgido. Los economistas del comportamiento intentan tomar la psicología humana en cuenta mediante el estudio de las elecciones reales de las personas en los entornos económicos idealizadas.

3.10.6 Juegos minoritarios

Como otra alternativa a la elección racional, econophysicists han desarrollado modelos muy simplificados de interacción estratégica para capturar la esencia de la conducta colectiva en un mercado financiero. problema bar El Farol de Brian Arthur ofrece una alternativa a la teoría de juegos convencional. El Farol es un bar en Santa Fe que a menudo se llena de gente. En el juego de Arthur, los agentes deciden cada día ya sea para ir a escuchar música. Si la barra tiene espacio, son felices;

si es demasiado lleno de gente, se sienten decepcionados. El juego está construido de manera que sólo una minoría de las personas puede ser feliz, lo que conduce a un fenómeno análogo a la frustración, algunos deseos son necesariamente insatisfactos. Como un resultado adicional de la estructura del juego, un astronómicamente gran número de equilibrios puede emerger.

3.10.7 Métodos de entropía

Hacienda no es la única área de la economía en la que los físicos están activos. La economía, como la física, distingue a los sistemas abiertos de los cerrados, y esa distinción da lugar a diferentes nociones de equilibrio. Mercados considerados simplemente como conductos para los bienes producidos o consumidos en otros lugares son descritos por las teorías de equilibrio parcial que se especifican en gran medida por las condiciones de contorno de sistema abierto. Los mercados financieros están abiertos en este sentido. Los economistas también tratan de determinar el equilibrio general de sociedades en su conjunto, teniendo en cuenta no sólo el comercio, sino la producción, el consumo, y, en cierta medida en la regulación, el gobierno.

La economía y la física en conjunto han experimentado un crecimiento en la comprensión de la relajación al equilibrio, cuando son posibles equilibrios, y si ellos son únicos. En ambos campos, modelos mecánicos se utilizaron primero, seguido por explicaciones estadísticas. Un trabajo reciente ha demostrado qué subconjunto de los problemas de decisión económica tiene una estructura idéntica a la de la termodinámica clásica y ha hablado de la aparición de un principio fenomenológico equivalente a la entropía de maximización. Más problemas de equilibrio general, por lo general considerados por los economistas corresponden a problemas físicos con muchos equilibrios, como granular, cristalina, o la relajación de histéresis. La idea de que los equilibrios corresponden a conjuntos estadísticamente más probables de configuraciones ha conducido a intentos de definir la formación de precios en términos estadísticos. Como se discutió en la caja, la idea relacionada de que la distribución del ingreso parece consistente con las diversas formas de la maximización de la entropía refunde el problema de la desigualdad de ingresos comprensión e interpretación de lo mucho que realmente habla de las fuerzas sociales que afectan a la renta.

En el otro extremo del espectro macro-micro, ideas de la mecánica estadística podrían hacer contribuciones prácticas a los problemas en la microestructura del mercado. Modelos de estilo Física sugieren, por ejemplo, que la volatilidad de precios podrían reducirse si las reglas del mercado se cambian para crear incentivos para la "paciente" órdenes de operaciones que no son inmediatamente transactados.^{9,10} Un problema práctico relacionado se refiere a la estrategia óptima para los creadores

de mercado, es decir, agentes que compran y venden de forma simultánea y obtener un beneficio tomando la diferencia. Aunque los mercados son cada vez más electrónica, los fabricantes automáticos del mercado están siendo diseñados de una manera más o menos ad hoc. Una teoría para la creación de mercado basado en los métodos de la mecánica estadística podría resultar en menores costos de transacción y los mercados en general más eficientes.

Varias ideas fundamentales de la física son de origen económico. Un prejuicio de que los libros deben equilibrar probable era responsable de la contabilidad de James Joule para el contenido de energía de calor antes de que el principio de conservación de la energía fue bien apoyado por los datos. El concepto de una moneda informa vista del papel de la energía en sistemas complejos de los científicos, particularmente en la bioquímica. La comprensión de la dinámica y la mecánica estadística de la agencia se compromete a ampliar el alcance conceptual de la física.

3.11 Curva ROC

En la Teoría de detección de señales una curva ROC (acrónimo de Receiver Operating Characteristic, o Característica Operativa del Receptor) es una representación gráfica de la sensibilidad frente a (1 - especificidad) para un sistema clasificador binario según se varía el umbral de discriminación. Otra interpretación de este gráfico es la representación de la razón o ratio de verdaderos positivos (VPR = Razón de Verdaderos Positivos) frente a la razón o ratio de falsos positivos (FPR = Razón de Falsos Positivos) también según se varía el umbral de discriminación (valor a partir del cual decidimos que un caso es un positivo). ROC también puede significar Relative Operating Characteristic (Característica Operativa Relativa) porque es una comparación de dos características operativas (VPR y FPR) según cambiamos el umbral para la decisión. En español es preferible mantener el acrónimo inglés, aunque es posible encontrar el equivalente español COR. No se suele utilizar ROC aislado, debemos decir "curva ROC" o "análisis ROC". Sobre la historia del acrónimo ROC consultar Swets (1996).

El análisis de la curva ROC, o simplemente análisis ROC, proporciona herramientas para seleccionar los modelos posiblemente óptimos y descartar modelos subóptimos independientemente de (y antes de especificar) el coste de la distribución de las dos clases sobre las que se decide. La curva ROC es también independiente de la distribución de las clases en la población (en diagnóstico, la prevalencia de una enfermedad en la población). El análisis ROC se relaciona de forma directa y natural con el análisis de coste/beneficio en toma de decisiones diagnósticas.

La curva ROC se desarrolló por ingenieros eléctricos para medir la eficacia en la detección de objetos enemigos en campos de batalla mediante pantallas de radar, a partir de lo cual se desarrolla la Teoría de Detección de Señales (TDS). El análisis ROC se aplicó posteriormente en medicina, radiología, psicología y otras áreas durante varias décadas. Sólo recientemente ha encontrado aplicación en áreas como aprendizaje automático (o machine learning en inglés), y minería de datos (data mining en inglés).

Un modelo de clasificación (clasificador o Clasificadores (matemático) o Diagnóstico) es una función que permite decidir cuáles de un conjunto de instancias están relacionadas o no por pertenecer a un mismo tipo o clase. El resultado del clasificador o del diagnóstico puede ser un número real (valor continuo), en cuyo caso el límite del clasificador entre cada clase debe determinarse por un valor umbral (por ejemplo para determinar si una persona tiene hipertensión basándonos en una medida de presión arterial), o puede ser un resultado discreto que indica directamente una de las clases.

Consideremos un problema de predicción de clases binario, en la que los resultados se etiquetan positivos (p) o negativos (n). Hay cuatro posibles resultados a partir de un clasificador binario como el propuesto. Si el resultado de una exploración es p y el valor dado es también p, entonces se conoce como un Verdadero Positivo (VP); sin embargo si el valor real es n entonces se conoce como un Falso Positivo (FP). De igual modo, tenemos un Verdadero Negativo (VN) cuando tanto la exploración como el valor dado son n, y un Falso Negativo (FN) cuando el resultado de la predicción es n pero el valor real es p. Un ejemplo aproximado de un problema real es el siguiente: consideremos una prueba diagnóstica que persiga determinar si una persona tiene una cierta enfermedad. Un falso positivo en este caso ocurre cuando la prueba predice que el resultado es positivo, cuando la persona no tiene realmente la enfermedad. Un falso negativo, por el contrario, ocurre cuando el resultado de la prueba es negativo, sugiriendo que no tiene la enfermedad cuando realmente sí la tiene.

Definamos un experimento a partir de P instancias positivas y N negativas. Los cuatro posibles resultados se pueden formular en una Tabla de contingencia (o Matriz de confusión) 2x2 como sigue:

3.11.1 El espacio ROC

La tabla de contingencia puede proporcionar varias medidas de evaluación (ver caja de terminología). Para dibujar una curva ROC sólo son necesarias las razones de Verdaderos Positivos (VPR) y de falsos pos-

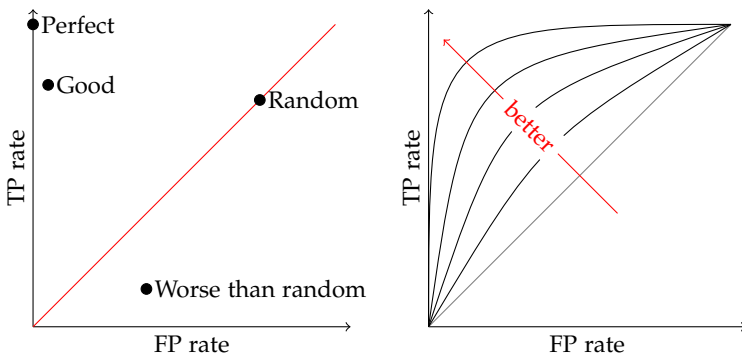
		Valor en la realidad		total
		p	n	
Prediction outcome	p'	Verdaderos Positivos	Falsos Positivos	P'
	n'	Falsos Negativos	Verdaderos Negativos	N'
total		P	N	

itivos (FPR). La VPR mide hasta qué punto un clasificador o prueba diagnóstica es capaz de detectar o clasificar los casos positivos correctamente, de entre todos los casos positivos disponibles durante la prueba. La FPR define cuántos resultados positivos son incorrectos de entre todos los casos negativos disponibles durante la prueba.

Un espacio ROC se define por FPR y VPR como ejes x e y respectivamente, y representa los intercambios entre verdaderos positivos (en principio, beneficios) y falsos positivos (en principio, costes). Dado que VPR es equivalente a sensibilidad y FPR es igual a 1-especificidad, el gráfico ROC también es conocido como la representación de sensibilidad frente a (1-especificidad). Cada resultado de predicción o instancia de la matriz de confusión representa un punto en el espacio ROC.

El mejor método posible de predicción se situaría en un punto en la esquina superior izquierda, o coordenada (0,1) del espacio ROC, representando un 100% de sensibilidad (ningún falso negativo) y un 100% también de especificidad (ningún falso positivo). A este punto (0,1) también se le llama una clasificación perfecta. Por el contrario, una clasificación totalmente aleatoria (o adivinación aleatoria) daría un punto a lo largo de la línea diagonal, que se llama también línea de no-discriminación, desde el extremo inferior izquierdo hasta la esquina superior derecha (independientemente de los tipos de base positivas y negativas). Un ejemplo típico de adivinación aleatoria sería decidir a partir de los resultados de lanzar una moneda al aire, a medida que el tamaño de la muestra aumenta, el punto de un clasificador aleatorio de ROC se desplazaría hacia la posición (0.5, 0.5).

La diagonal divide el espacio ROC. Los puntos por encima de la diagonal representan los buenos resultados de clasificación (mejor que el azar), puntos por debajo de la línea de los resultados pobres (peor que al azar). Nótese que la salida de un predictor consistentemente pobre simplemente podría ser invertida para obtener un buen predictor.



3.11.2 Cómo se puede interpretar una curva ROC

La curva ROC se puede usar para generar estadísticos que resumen el rendimiento (o la efectividad, en su más amplio sentido) del clasificador. A continuación se proporcionan algunos:

- El punto de inserción de la curva ROC con la línea convexa a la línea de discriminación.
- El área entre la curva ROC y la línea de convexo-paralela discriminación.
- El área bajo la curva ROC, llamada comúnmente AUC (Área Bajo la Curva). También se puede encontrar denominada A' ("a-prima"), o el estadístico c' (c-statistic).
- Índice de sensibilidad o d' (d-prima, por cierto siempre minúscula). Es la distancia entre la media de la distribución de actividad en el sistema bajo condiciones de sólo ruido y su distribución bajo condiciones de sólo señal, dividido por su desviación típica, bajo el supuesto de que ambas distribuciones son normales con la misma desviación típica. Bajo estos supuestos, se puede probar que la forma de la curva ROC sólo depende de este parámetro d' .

El indicador más utilizado en muchos contextos es el área bajo la curva ROC o AUC. Este índice se puede interpretar como la probabilidad de que un clasificador ordenará o puntuará una instancia positiva elegida aleatoriamente más alta que una negativa. Se puede demostrar que el área bajo la curva ROC es equivalente a la Prueba de Mann-Whitney, una prueba no paramétrica aplicada a dos muestras independientes, cuyos datos han sido medidos al menos en una escala de nivel ordinal. Se trata de una prueba estadística virtualmente idéntica a la realización de una

prueba paramétrica ordinaria T de dos muestras en los datos después de haber ordenado las muestras combinadas. Es también equivalente a la Prueba de los signos de Wilcoxon. También se ha demostrado la relación del área bajo la curva ROC con el Coeficiente de Gini, con la siguiente fórmula $G_1 + 1 = 2(A \cup C)$, donde:

$$G_1 = 1 - \sum_{k=1}^n (X_k - X_{k-1})(Y_k + Y_{k-1})$$

Otra forma básica de calcular AUC es usando un promedio de una serie de aproximaciones trapezoidales. Sin embargo, se ha comentado que este indicador, en general, reducir la curva ROC en varios metros, hace perder información sobre el patrón de intercambios del algoritmo discriminador en cuestión.

La comunidad de aprendizaje automático utiliza el estadístico AUC para la comparación de modelos. En otras áreas de ingeniería se prefiere la medida del área entre la curva ROC y la línea de no-discriminación. Finalmente en Psicofísica se utiliza preferentemente d' .

En ocasiones puede ser más útil mirar a una región específica de la curva ROC más que a toda la curva. Es posible calcular áreas parciales bajo la curva, o AUC parciales. Por ejemplo, nos podríamos concentrar en la región de la curva con razones de falsos positivos más bajas, que es a menudo el interés principal de las pruebas de Detección precoz (o (medicine)—screening en la población.

3.12 Delta de Dirac

La delta de Dirac es una función generalizada que viene definida por la siguiente fórmula integral:

$$\int_{-\infty}^{\infty} \delta(x - a)f(x) dx = f(a) \quad \left[\text{e.g.} \int_{-\infty}^{\infty} \delta(x) dx = 1 \right]$$

La delta de Dirac no es una función estrictamente hablando, puesto que se puede ver que requeriría tomar valores infinitos. A veces, informalmente, se define la delta de Dirac como el límite de una sucesión de funciones que tiende a cero en todo punto del espacio excepto en un punto para el cual divergería hacia infinito; de ahí la "definición convencional" dada por la también convencional fórmula aplicada a las funciones definidas a trozos:

$$\delta(x) = \begin{cases} \infty, & x = 0 \\ 0, & x \neq 0 \end{cases};$$

Es frecuente que en física la delta de Dirac se use como una distribución de probabilidad idealizada; técnicamente, de hecho, es una distribución (en el sentido de Schwartz).

En términos del análisis dimensional, esta definición de $\delta(x)$ implica que $\delta(x)$ posee dimensiones recíprocas a dx .

3.12.1 Definición como distribución de densidad

$$\int_a^b f(x)\delta(x-x_0) dx = \begin{cases} f(x_0) & \text{si } a < x_0 < b \\ 0 & \text{si } x_0 < a \text{ o } x_0 > b \end{cases}$$

3.12.2 Definición como límite de sucesiones de funciones

La delta de Dirac se define como "límite distribucional" de una sucesión de funciones que convergen puntualmente a la función cero en todos los puntos de su dominio excepto uno. Se dice que una sucesión de funciones $f_n(x)$ converge distribucionalmente cuando:

$$\left[\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f_n(x)\phi(x) dx \right] \rightarrow d(\phi)$$

Donde ϕ es una función perteneciente a un espacio vectorial de funciones, y d es un funcional continuo del espacio vectorial dual (el conjunto de esos elementos continuos es un subespacio vectorial del dual, conocido como espacio dual topológico del espacio original de funciones. La delta de Dirac centrada se puede definir como el límite distribucional del funcional dado por $d(\phi) = \phi(0)$, es decir, el límite en el sentido de las distribuciones de una sucesión de funciones tales que:

$$\left[\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f_n(x)\phi(x) dx \right] \rightarrow \phi(0)$$

Algunos ejemplos posibles de sucesión de funciones que cumpla lo anterior son:

$$f_n(x) = \begin{cases} n & \|x\| < \frac{1}{2n} \\ 0 & \|x\| \geq \frac{1}{2n} \end{cases} \quad f_n(x) = \frac{n}{\sqrt{\pi}} e^{-n^2 x^2}$$

$$f_n(x) = \frac{1}{\pi n^2 x^2 + 1} \quad f_n(x) = \frac{\sin nx}{\pi x}$$

3.12.3 Propiedades

Estas propiedades se pueden demostrar multiplicando ambos miembros de cada igualdad por una función $f(x)$ e integrando teniendo en cuenta que la función delta no puede formar parte del resultado a menos que esté dentro de una integral.

$$\delta(x) = \delta(-x)f(x)\delta'(x) = -f'(x)\delta(x)\delta'(x) = -\delta'(-x)x^n\delta(x) = 0 \quad \forall n > 0, x \in \mathbb{R}$$

En coordenadas esféricas se tiene:

$$\delta(\mathbf{r} - \mathbf{r}_0) = \begin{cases} \frac{1}{r^2 \sin \theta} \delta(r - r_0) \delta(\theta - \theta_0) \delta(\phi - \phi_0) & x_0, y_0, z_0 \neq 0 \\ \frac{1}{2\pi r^2 \sin \theta} \delta(r - r_0) \delta(\theta - \theta_0) & x_0 = y_0 = 0, z_0 \neq 0 \\ \frac{1}{4\pi r^2} \delta(r - r_0) & x_0 = y_0 = z_0 = 0 \end{cases}$$

3.13 El Hombre Anumérico: Capítulo 2 "Probabilidad y coincidencia"

3.13.1 Algunos cumpleaños y un cumpleaños determinado

Esta sección nos relata anécdotas relativas a las coincidencias que mucha gente toma como acontecimientos sumamente afortunados o cosas del destino; pero que en realidad, al tratarse de casos generales, cuentan con una probabilidad razonable de ocurrir. Tal es el caso de escoger el mínimo número de personas que necesitamos juntar para que al menos dos de ellas hayan nacido el mismo día. El razonamiento numérico expuesto por el autor resulta muy claro, se puede concluir que, con frecuencia tomamos sucesos fortuitos, totalmente independientes, como totalmente improbables o con una probabilidad muy baja de ocurrir; cosa que no es del todo cierta, solo basta hacer un análisis de conteo relativamente simple para descubrirlo. También nos damos cuenta que el truco para tener una probabilidad de buen tamaño es plantear un suceso general en vez de uno específico.

3.13.2 Encuentros fortuitos

En éste parrafo, el autor escribe sobre las conexiones interpersonales que la gente de Estados Unidos tiene. Se aprecia que, la probabilidad de encontrarse con alguien con el cual se tiene un conocido en común es más alta de lo que se piensa; inclusive, si el par de personas no tiene un conocido en común, indirectamente están conectados a través de no más de dos intermediarios.

Las suposiciones en que basamos la deducción anterior se pueden relajar un tanto. Quizás el adulto medio conozca menos de 1.500 personas o, lo que es más probable, la mayoría de la gente que conoce vive cerca y no está dispersa por todo el país. Incluso en este caso, menos favorable, es inesperadamente alta la probabilidad de que dos personas escogidas al azar estén unidas por una cadena de como mucho dos intermediarios.

Como 2.200.000 es considerablemente mayor que 500.000, podemos estar seguros de que por lo menos dos personas tienen el mismo número de cabellos; esto es, que alguno de los buzones recibirá por lo menos dos habitantes de Filadelfia.

3.13.3 Un timo bursátil

Los asesores de bolsa están en todas partes y es muy probable encontrar alguno que diga cualquier cosa que uno esté dispuesto a oír. Normalmente son enérgicos, parecen muy expertos y hablan una extraña jerga de opciones de compra y de venta, cupones de cero y cosas por el estilo. A la luz de mi humilde experiencia, la mayoría no tiene mucha idea de lo que está hablando, pero cabe esperar que algunos sí.

Antes de la radio, la televisión y el cine, los músicos, los atletas, etcétera, podían hacerse un público local de leales, pues eran lo mejor que la mayoría de esas personas iba a ver en su vida. Los públicos de ahora nunca quedan satisfechos de las figuras locales, ni siquiera en las zonas rurales, y exigen talentos de primera línea. Se puede decir en este sentido que, con los grandes medios de comunicación, los públicos han salido ganando, y los artistas perdiendo.

3.13.4 Valores esperados: de los análisis de sangre al juego del chuck-a-luck

La explicación del chuck-a-luck que se da para atraer a la gente puede ser muy persuasiva. El que apuesta elige un número de 1 a 6 y el encargado lanza tres dados. Si el número elegido sale en los tres dados, el jugador cobra 3 dólares; si sale en dos de los dados, cobra 2 dólares y si sale en uno de los tres dados, sólo cobra 1 dólar. Únicamente en el caso de que el número escogido no salga en ninguno de los dados tendrá que pagar el jugador, y sólo 1 dólar. Con tres dados distintos, el apostador tiene tres posibilidades a su favor; además, a veces gana más de 1 dólar, que es lo máximo que puede perder cada vez.

3.13.5 Eligiendo cónyuge

Es muy posible que, al recordar amores pasados, alguien que enfoque sus romances con el corazón se lamente de las oportunidades perdidas y que piense que nunca jamás volverá a amar así. Otra persona más práctica, que se decida por un enfoque más realista, seguramente estará interesada por el siguiente resultado probabilístico.

En la obtención del mejor sistema se emplea la idea de probabilidad condicional (que presentaremos en el próximo capítulo) y también hay que calcular un poco. El sistema en sí, no obstante, se describe muy fácilmente. Diremos que un pretendiente es un novio si es mejor que todos los candidatos anteriores. María debería rechazar aproximadamente el primer 37% de los candidatos que probablemente vaya a conocer y luego aceptar al primer novio que le salga de entre los pretendientes posteriores (si es que le sale alguno, claro).

Para grandes valores de N , la probabilidad de que aplicando esta regla del 37 por ciento María encuentre a su hombre ideal, es también

aproximadamente del 37 por ciento.

3.13.6 Las coincidencias y la ley

En 1964 una mujer rubia peinada con una cola de caballo robó el bolso a otra mujer en Los ángeles. La ladrona huyó a pie, pero posteriormente alguien la reconoció cuando montaba en un coche amarillo conducido por un negro con barba y bigote. Las investigaciones de la policía acabaron por encontrar a una mujer rubia con cola de caballo que regularmente frecuentaba la compañía de un negro de barba y bigote que tenía un coche amarillo. No había ninguna prueba fehaciente que relacionara a la pareja con el delito, ni testigos que pudieran identificar a ninguno de los dos. Se estaba de acuerdo, no obstante, en los hechos citados.

En determinados contextos, la improbabilidad es algo que no sorprende. Cada mano de bridge es muy improbable. También lo son las manos de póker y los billetes de lotería. En el caso de la pareja californiana, la improbabilidad es más significativa. Sin embargo, el razonamiento correcto es el de su abogado defensor.

3.13.7 Monedas no trucadas y ganadores o perdedores en el juego de la vida

Imaginemos a dos jugadores, Pedro y Pablo, que juegan a cara o cruz, tirando una moneda al aire una vez por día. En un momento dado, diremos que Pedro va ganando si hasta aquel momento han salido más caras que cruces, y en caso contrario es Pablo quien va ganando. En cualquier momento, tanto Pedro como Pablo tienen la misma probabilidad de ir ganando, pero sea quien sea el que vaya ganando, este es el que tiene mayor probabilidad de haber estado ganando más rato. Si han tirado la moneda cien veces y acaba ganando Pedro ¡es considerablemente mayor la probabilidad de que este haya estado por delante más del 90 por ciento del tiempo, pongamos, que la de que lo haya estado entre el 45 y el 55 por ciento! Y análogamente, si acaba ganando Pablo, la probabilidad de que este haya estado ganando más del 96 por ciento del tiempo es mucho menor que la de que lo haya estado entre el 48 y el 52 por ciento.

La moneda no sabe nada, no obstante, de medias ni de bandas elásticas, y si ha salido cara 519 veces y cruz 481, es tan probable que la diferencia entre caras y cruces aumente como que disminuya. Y esto es cierto a pesar de que la proporción de caras tienda a $1/2$ a medida que aumenta el número de tiradas.

Los comentaristas tienen siempre un reparto habitual de personajes a los que recurrir para explicar cualquier recuperación o cualquier descenso. Siempre tienen a mano la realización de las plusvalías, el déficit federal, o cualquier otra cosa para explicar los giros a la baja, y el au-

mento de los beneficios de las sociedades, el aumento de los tipos de interés o lo que sea para explicar los giros alcistas.

3.13.8 La racha de suerte y el manitas

Los grupos, series y pautas que presentan las sucesiones aleatorias son hasta cierto punto predecibles. Las sucesiones de caras y cruces de una longitud dada, pongamos veinte tiradas, tienen generalmente cierto número de series de caras consecutivas. Diremos que una sucesión de veinte tiradas de una moneda que diera diez caras seguidas y diez cruces tiene sólo una serie de caras, mientras que una sucesión de veinte tiradas que diera alternativamente cara y cruz tiene diez series de caras. Es muy improbable que esas dos sucesiones hayan sido generadas al azar.

Se puede pulir más el razonamiento para tratar el caso de que la media de aciertos del jugador sea distinta del 50 por ciento, y parece que valen resultados parecidos. Un jugador que marca el 65 por ciento de sus tiros, pongamos, marca tantos del mismo modo que marca caras en una moneda trucada que cae cara en el 65 por ciento de las veces que la tiramos; es decir, cada tiro es independiente del anterior. Siempre he tenido la sospecha de que cosas como rachas de suerte o manitas o un equipo que siempre remonta no eran más que exageraciones de los periodistas deportivos, sin otra intención que tener algo de que hablar. Seguramente tales expresiones signifiquen algo, pero demasiado a menudo sólo son fruto de un intento mental por descubrir un significado donde no hay más que probabilidad.

Primero hay que conocer aproximadamente la improbabilidad del hecho y, una vez conocida, se puede usar esta información junto con la fórmula de Poisson para tener una idea bastante aproximada de, por ejemplo, cuántos años pasarán sin que haya muertos por coz de caballo, en qué porcentaje de los años venideros habrá una de tales muertes, en qué porcentaje habrá dos, etc. De modo análogo, se puede predecir el porcentaje de los años en los que no habrá precipitaciones de lluvia en un desierto, una precipitación, dos, etcétera.

3.14 Matrices semejantes

3.14.1 Introducción

Si A es una matriz real cuadrada de dimensión n , el problema de la diagonalización de A consiste en:

- hallar una matriz regular P
- y hallar una matriz diagonal D tal que

(3.1)

$$A = PDP^{-1}$$

Cuando dos matrices cuadradas A y D cualesquiera verifican una relación del tipo anterior, diremos que ambas matrices son semejantes. Supuesto que A sea diagonalizable, veamos quienes han de ser estas matrices P y D . Escribamos P fijándonos en sus columnas, es decir $P = (x_1 | x_2 | \cdots | x_n)$ siendo x_i la i -ésima columna de P y llamaremos $\lambda_1, \lambda_2, \dots, \lambda_n$ a los números que aparecen en la diagonal de la matriz D . Entonces, reescribiendo la relación entre A , P y D de la siguiente forma

$$A = PDP^{-1} \Rightarrow AP = PD \quad (3.2)$$

y fijándonos en la i -ésima columna de las dos igualdades obtenemos:

$$Ax_i = \lambda_i x_i, 1 < i < n \quad (3.3)$$

Cuando se encuentra un número real λ y un vector no nulo x que verifican la relación $Ax = \lambda x$ diremos que:

- λ es un valor propio, autovalor o eigenvalor de la matriz A
- x es un vector propio, autovector o eigenvector de A , asociado al eigenvalor λ

Por lo tanto, encontrar la matriz D equivale a encontrar n valores propios (no necesariamente distintos), que serán los que conformen la diagonal de D , y encontrar la matriz P equivale a encontrar n vectores propios linealmente independientes (para que P sea regular) asociados a sus respectivos valores propios que conformarán las columnas de P . Así, las columnas de P formarían una base de \mathbb{R}^n compuesta únicamente por vectores propios de la matriz A . Es fundamental, pues, hallar los valores propios de A y los vectores propios asociados. Como los primeros hacen que el sistema $Ax = \lambda x$ tenga solución x distinta de la solución cero, la matriz de coeficientes $A - \lambda I$ (donde I denota la matriz identidad de orden n) debe tener determinante no nulo. Este determinante $\det(A - \lambda I)$ es un polinomio en λ de grado n y se denomina polinomio característico de A . Por lo tanto los valores propios de A serán los ceros del polinomio característico de A al que se suele notar como

$$p_A(\lambda) = \det(A - \lambda I) \quad (3.4)$$

Por otro lado, el conjunto de vectores propios de A asociados a un mismo valor propio λ forman un subespacio vectorial de \mathbb{R}^n que se llama subespacio propio asociado al valor propio λ . Para concluir si una matriz A es o no diagonalizable bastará pues averiguar si hay suficientes valores propios para construir D y si hay suficientes vectores propios linealmente independientes asociados; esta información nos la dará la dimensión de los subespacios propios y queda recogida en el siguiente resultado.

3.14.2 Aplicación: potencias de una matriz

En ocasiones es necesario calcular potencias elevadas de una matriz dada o incluso raíces. En estos casos es útil disponer de una matriz semejante a la dada a la que sea fácil calcularle esas potencias o raíces para disminuir el coste computacional. La simplificación se basa en que:

$$(3.5) \quad A = PDP^{-1} \Rightarrow A^m = PD^mP^{-1}$$

Si la matriz D es diagonal, sus potencias se calculan simplemente haciendo la misma potencia (incluso fraccionaria) de cada uno de los elementos de la diagonal, lo que conlleva menos cómputo que hacerlo para una matriz arbitraria.

3.15 Ejercicios de cadenas de Markov

1. Una agencia de arriendo de vehículos ha definido la variable aleatoria X_t como el número de automóviles disponibles en la agencia al empezar la semana $t+1$. Sea D_t una variable aleatoria que representa la demanda por automóviles la semana t . La agencia utiliza una política de reorden (s, S) con $s=1$ y $S=3$. No se acepta demanda pendiente. Sea $X_0 = 3$ y suponga que la variable aleatoria D_t tiene distribución de Poisson con $\lambda = 1$.

- a) Obtenga los valores de la variable X_t
- b) Exprese a través de una fórmula de recurrencia la relación entre x_t y X_{t+1}
- c) Encuentre la matriz P (valores numéricos)
- d) Suponga ahora que el costo incurrido es un valor fijo de \$ 110.000 por orden más un valor variable de \$ 25.000 por automóvil. Encuentre el costo esperado de inventario.

Solución

$$a) \Omega_{X_t} = \{S, s\} = \{1, 2, 3\}$$

b)

$$X_{t+1} = \begin{cases} S = 3 & \text{si } X_t - D_{t+1} < s \\ X_t & \text{si } X_t - D_{t+1} \geq s \end{cases}$$

c)

$$P = \begin{bmatrix} P(D_t = 0) & 0 & P(D_t > 0) \\ P(D_t = 1) & P(D_t = 0) & P(D_t > 1) \\ P(D_t = 2) & P(D_t = 1) & P(D_t > 2) \end{bmatrix}$$

$$d) C = f_1^{(n)} \cdot [25 \cdot 1 + 110 \cdot P(D_t > 0)] + f_2^{(n)} \cdot [25 \cdot 2 + 110 \cdot P(D_t > 1)] + f_3^{(n)} \cdot [25 \cdot 3 + 110 \cdot P(D_t > 2)]$$

2. Se quiere construir un modelo markoviano para estimar la dinámica de vida de un cultivo de ovas de salmón en la etapa de maduración de la ova. En un estanque con agua dulce, se ponen N ovas fecundadas. Después de 8 días, las ovas que han sobrevivido se convierten en pequeños salmones, los cuales se traspasan a otros estanques para seguir su evolución. En este proceso de Maduración, algunas ovas se mueren. Se sabe que la distribución de probabilidades de una ova en esta etapa de desarrollo, es exponencial con media μ . Sea X_n el número de ovas vivas al inicio del día n .
 - a) Diga cual es el rango de la variable
 - b) Obtenga la regla de transición
 - c) Obtenga la matriz P

Solución Sea X_n : número de ovas vivas al inicio del día n . Parten N ovas vivas, pero van muriendo día a día.

3. Juan y Pedro tienen 2 monedas cada uno. Se disponen a enfrentar un juego en que, en cada oportunidad, cada jugador lanza una moneda de sus monedas. Si ambas coinciden, gana Juan y se queda con la moneda de Pedro. En caso contrario, gana Pedro. El juego termina cuando uno de los jugadores gana las 4 monedas.
 - a) Obtenga la distribución de probabilidades del número de jugadas necesarias hasta que Juan logre tener 3 monedas por primera vez.
 - b) Explique como obtendría la distribución de probabilidades del número de jugadas hasta que el juego termina.
4. Considere un cultivo que contiene inicialmente un solo glóbulo rojo. Después de una cantidad de tiempo el glóbulo rojo muere y es reemplazado por dos nuevos glóbulos rojos o bien por dos glóbulos blancos. Las probabilidades de estos eventos son $\frac{1}{4}$ y $\frac{3}{4}$ respectivamente. Subsecuentemente, cada glóbulo rojo se reproduce de la misma forma. Por otra parte, cada glóbulo blanco muere después de una unidad de tiempo sin reproducirse. Se desea calcular la probabilidad de que el cultivo se extinga en algún momento. Formule para tal efecto un modelo detallado e indique con precisión como lo utilizaría para obtener la probabilidad pedida.

3.16 Matrices dispersas

3.16.1 Introducción

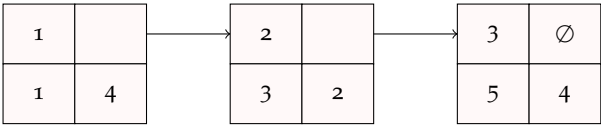
Las matrices dispersas son ampliamente usadas en la computación científica, especialmente en la optimización a gran escala, análisis estructural y de circuitos, dinámica de fluidos computacionales y en general en solución numérica de ecuaciones diferenciales parciales; otras áreas de interés en donde se puede aplicar la representación dispersa son la teoría de grafos, teoría de redes, la combinatoria, los métodos numéricos, entre otros. Dada su frecuente aparición existen múltiples paquetes de software que permiten su implementación y operación como son Sparspak, el Yale Sparse Matrix package, algunas rutinas de la librería de subrutinas de Harwell, además del paquete de software Matlab, entre otros.

Dado que este tipo de matrices ocurren tan naturalmente a través de los años se han desarrollado distintos métodos para representarlas en un computador, de manera que sean más eficientes en su computación y almacenamiento.

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \end{bmatrix}$$

3.16.2 Lista enlazada

Es una estructura de datos dinámica que consiste en tener una lista enlazada, en la cual se almacena un elemento no nulo de la matriz en cada nodo de la lista enlazada. La información que se almacenaría sería el valor numérico de cada elemento de la matriz, y la posición que ocupa ese dato dentro de la matriz, es decir, el número de fila y columna en la que se encuentra.

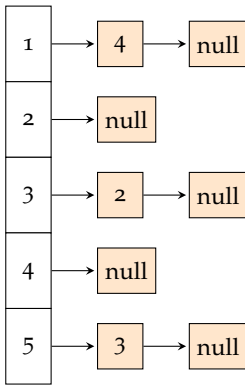


En la Figura se observa una representación de la matriz dispersa de la Figura 1 en una lista enlazada, en donde cada nodo almacena un elemento de la matriz, se puede observar que se tienen punteros de un nodo al siguiente, además de la información de la fila y la columna en la que se encuentra en dicha matriz.

Por ejemplo en el nodo 1 se tiene el valor numérico 1 que se encuentra en la posición 1,4 dentro de la matriz. 2 se tiene el valor numérico 2 que

se encuentra en la posición 3,2 dentro de la matriz 3 se tiene el valor numérico 3 que se encuentra en la posición 5,4 dentro de la matriz. Esta forma de almacenamiento es muy ineficiente por el alto consumo de memoria aleatoria. Su única ventaja es la agregación de nuevos datos a la matriz, pero debido a que la gran mayoría de veces las matrices dispersas en la computación científica son estáticas, éste método no sirve de mucho.

El formato comprimido tiene una eficiencia equiparable al formato coordenado, pero tiene una ventaja y es su menor uso de memoria, ya que la idea de este formato es almacenar el índice de fila implícitamente. Para explicar este formato se debe retomar el concepto de las listas enlazadas:



En la figura se muestra una estructura de datos con las listas enlazadas. En esta este formato en lugar de guardar todos los elementos de la matriz en una lista, se reagrupan por fila, por lo tanto se tendrá una lista enlazada por cada fila de la matriz.

Este formato es un poco más complejo pero tiene 2 ventajas muy importantes. La primera es el poder localizar un elemento más fácilmente, por ejemplo si se quiere conocer los datos de la fila 3, simplemente se va al puntero de la fila 4 y ahí se localizan los datos de esa fila. La segunda ventaja que se tiene es que no es necesario almacenar el índice de la fila por lo que se requerirá menos memoria.

3.16.3 Aplicaciones

Entre las aplicaciones en las que se pueden usar este tipo de matrices se encuentran, la teoría de redes, la cuales tienen una baja densidad de datos significantes o de conexiones. La representación de grafos es otra aplicación de las matrices dispersas, ya que los grafos en estas matrices ocuparan menos memoria debido a que solo se representaran los enlaces existentes en el grafo.

En el procesamiento vectorial se pueden aplicar matrices dispersas, debido a que existen casos especiales en operaciones vectoriales binarias como el empaquetamiento y desempaquetamiento, los cuales actúan como máscara y el vector resultado tiene un tamaño diferente al de los operandos. En la matemática combinatoria también se pueden aplicar las matrices dispersas, debido a que el fin de la combinatoria es estudiar las diferentes agrupaciones de los objetos en donde es prescindible la naturaleza de estos mas no el orden, por lo tanto se encuentran frecuentemente problemas en los que son muy útiles este tipo de matrices.

Las matrices dispersas pueden convertirse en herramientas computacionales muy útiles, pero su uso indiscriminado puede empeorar el consumo de memoria y los tiempos de procesamiento. Existen distintas implementaciones de matrices dispersas algunas de ellas tratadas en este documento. Cada representación tiene sus ventajas y desventajas con respecto al consumo de memoria o el tiempo de procesamiento debido a esto no existe una representación universal que permita resolver todos los posibles problemas y cada implementación debe hacerse teniendo en cuenta que objetivos se quieren lograr.

3.17 Cómo modelar de crecimiento viral

3.17.1 ¿Qué es un producto viral?

Los productos adquieren usuarios a través de una variedad de canales, como la prensa, la publicidad, y las asociaciones. Tal vez el canal más fascinante es el propio usuario.

Un producto viral deriva gran parte de su crecimiento a partir de sus usuarios actuales reclutar nuevos usuarios. Un usuario podría contratar a otro a través de una simple invitación ("Confirmar este producto, es fresco / útil / entretenido!"), O directamente a través de la utilización del producto ("Quiero enviar dinero en PayPal!").

Uno de los ejemplos más famosos de la viralidad es YouTube. Antes de que el sitio se convirtió en el gran destino que es hoy, lo más probable es encontrar un vídeo de YouTube embebido en el sitio de noticias o blog personal. Te lo mira, y al final, usted será invitado a enviar por correo electrónico el vídeo a gente que conoces, y también estaría dado código para incrustar el vídeo en su propio sitio. Si no lo hizo el intercambio de fantasía en ese momento, YouTube podría sugerir otros vídeos que te pueden gustar, y tal vez sería buena idea compartir uno de esos. Entonces la gente que ha compartido con tendrían ver el video, y tal vez que habrían compartirlo con la gente que conocen. Y así, dando vueltas y vueltas este "ciclo viral" fue, durante el cual adquirió YouTube usuarios a un ritmo sin precedentes.

¿Cómo podemos predecir cómo nuestro producto viral llevará a cabo?
¿Cuánto tiempo se tarda en adquirir los usuarios 1M? Vamos a llegar a

10 millones de usuarios? ¿Qué pasa con los usuarios 100M?

Para responder a preguntas como éstas, que necesitamos para construir un modelo de viralidad.

3.17.2 El modelo más simple posible

Digamos que empezamos con 5.000 usuarios. ¿Cuántos usuarios nuevos reclutarán a estos usuarios iniciales?

Bueno, algunos usuarios tendrán gusto de nuestro producto, mientras que otros no les gusta a ella. Algunos usuarios invitar a muchas personas, mientras que otros no invitar a nadie en absoluto. Algunos usuarios pueden invitar a la gente después de un día, mientras que otros pueden tardar una semana.

Vamos a barrer todas estas incertidumbres de distancia, y decir que, en promedio, 1 de cada 5 de nuestros usuarios reclutará con éxito un nuevo usuario en su primer mes. En otras palabras, nuestro factor viral es $1/5 = 0,2$, y nuestros primeros 5.000 usuarios reclutará otros $5.000 * 0,2 = 1.000$ usuarios en el mes 1. Estos 1.000 usuarios serán entonces reclutar otros $1.000 * 0,2 = 200$ usuarios en el mes 2, que luego reclutar otros $200 * 0,2 = 40$ usuarios en meses 3, y así sucesivamente.

Adquirimos los usuarios a una tasa decreciente hasta que tengamos 6.250 usuarios. ¿Qué pasa si nuestro factor viral es 0,4 en lugar? Una vez más, adquirimos los usuarios a una tasa decreciente. Pero esta vez, nuestro crecimiento colea en alrededor de 8.300 usuarios.

Ahora, ¿qué ocurre si nuestro factor viral es 1,2 en lugar? Esta vez, los usuarios adquieren a un ritmo cada vez mayor. De hecho, con unas matemáticas sencillas, podemos mostrar lo siguiente:

Con los usuarios iniciales x , y con el factor viral v menor que 1, que adquirirá los usuarios a una tasa decreciente hasta que tengamos los usuarios $x / (1-v)$. Con el factor viral mayor que 1, adquirimos usuarios a una tasa aparentemente cada vez mayor. Así que es simple, todo lo que tenemos que hacer es conseguir nuestro factor viral más de 1, y estamos listos, ¿verdad?

En primer lugar, hay un montón de malo en nuestro modelo. Por ejemplo, a medida que adquirimos usuarios cada vez más, eventualmente quedará sin nuevas personas para invitar! No explorar esta idea más lejos aquí, pero para aprender más acerca efecto, veo Facebook marketing viral.

En segundo lugar, el crecimiento viral cierto es increíblemente raro. Me tomó un tiempo para apreciar esto: muy pocos productos han sufrido un factor viral más de 1 por un período de tiempo significativo. Pero si no hay que apostar por un factor viral superior a 1, lo que debemos usar en nuestro modelo?

A partir de las discusiones con otros empresarios, inversores, y los hackers de crecimiento, he aprendido lo siguiente: para un producto de con-

sumo de Internet, un factor viral sostenible de 0,15 a 0,25 es bueno, 0,4 es grande, y alrededor de 0,7 es excepcional.

Sin embargo, ya hemos demostrado que con nuestro factor viral cuando es menor que 1, adquirimos los usuarios a una tasa decreciente hasta que no más crecemos. Eso no es un resultado que nadie quiere, por lo que lo que falta en esta foto?

Nos faltan todos los demás canales con los que podemos adquirir los usuarios: prensa, tiendas de aplicaciones, dirigen el tráfico, la comercialización de entrada, la publicidad pagada, asociaciones de integración, las asociaciones de promoción cruzada, marketing en buscadores, optimización de motores de búsqueda, apoyo de famosos, calle- presionante esquina, y todo lo que podemos pensar. Vamos a revisar nuestro modelo y tenerlas en cuenta.

3.18 Teorémas de punto fijo

En las matemáticas, un teorema de punto fijo es el resultado diciendo que una función F tendrá por lo menos un punto fijo (un punto x para el que $F(x) = x$), bajo algunas condiciones en F que se pueden exponer en términos generales. Los resultados de este tipo se encuentran entre los más útiles generalmente en matemáticas.

3.18.1 En el análisis matemático

El teorema del punto fijo de Banach da un criterio general que garantiza que, si se cumple, el procedimiento de iteración de una función se obtiene un punto fijo. Por el contrario, el punto fijo de Brouwer teorema es un no resultado constructivo: se dice que cualquier continua función de la cerrada bola unidad n -dimensional espacio euclidiano a sí mismo debe tener un punto fijo, pero no describe cómo encontrar el punto fijo (Véase también el lema de Sperner).

Por ejemplo, el coseno función es continua en $[-1,1]$ y mapea en $[-1, 1]$, y por lo tanto debe tener un punto fijo. Esto es evidente cuando se examina un gráfico dibujado de la función coseno; el punto fijo se produce cuando la curva coseno $y = \cos(x)$ corta la recta $y = x$. Numéricamente, el punto fijo es aproximadamente $x = 0.73908513321516$ (por lo tanto $x = \cos(x)$ para este valor de x).

El teorema del punto fijo de Lefschetz (y el teorema del punto fijo Nielsen) de la topología algebraica es notable, ya que da, en cierto sentido, una manera de contar los puntos fijos.

Hay una serie de generalizaciones a Banach teorema del punto fijo y complementaria; Estos se aplican en PDE teoría. Ver teoremas de punto fijo en espacios de dimensión infinita.

El teorema collage en compresión fractal demuestra que, para muchas imágenes, existe una relativamente pequeña descripción de una función

que, cuando iterativamente aplicar a cualquier imagen de partida, rápidamente converge en la imagen deseada.

3.18.2 En álgebra y matemáticas discretas

El teorema de Knaster-Tarski afirma que cualquier función de orden de preservación en un retículo completo tiene un punto fijo, y de hecho un punto fijo más pequeño. El teorema tiene aplicaciones en la interpretación abstracta, una forma de análisis de programas estática.

Un tema común en el cálculo lambda es encontrar puntos fijos de las expresiones lambda dado. Cada expresión lambda tiene un punto fijo, y un combinador de punto fijo es una "función", que toma como entrada una expresión lambda y produce como salida un punto fijo de dicha expresión. Un combinador de punto fijo importante es el combinador Y utilizado para dar recursivas definiciones.

En la semántica denotativa de lenguajes de programación, un caso especial del teorema de Knaster-Tarski se utiliza para establecer la semántica de las definiciones recursivas. Mientras que el teorema de punto fijo es aplicado a la "misma" función (desde un punto de vista lógico), el desarrollo de la teoría es bastante diferente.

La misma definición de función recursiva se puede dar, en la teoría de la computabilidad, aplicando el teorema de recursión de Kleene. Estos resultados son teoremas no equivalentes; el teorema de Knaster-Tarski es un resultado mucho más fuerte que lo que se utiliza en la semántica denotativa. Sin embargo, a la luz de la tesis de Church-Turing su significado intuitivo es el mismo: una función recursiva puede ser descrito como el punto mínimo fijo de un cierto funcional, funciones de mapeo a funciones.

La técnica anterior de la iteración una función para encontrar un punto fijo también se puede utilizar en la teoría de conjuntos; el punto fijo lema para las funciones normales establece que toda función continua estrictamente creciente de los ordinales a los ordinales tiene uno (y de hecho muchos) puntos fijos.

Cada operador de cierre en un conjunto parcialmente ordenado tiene muchos puntos fijos; estos son los "elementos cerrados" con respecto al operador de la clausura, y son la razón principal por la que el operador de cierre se definió en el primer lugar.

Cada involución en un conjunto finito con un número impar de elementos tiene un punto fijo; más en general, para cada una involución en un conjunto finito de elementos, el número de elementos y el número de puntos fijos tienen la misma paridad. Don Zagier utilizar estas observaciones para dar una prueba de una sola frase del teorema de Fermat sobre sumas de dos cuadrados, por que describe dos involuciones en el mismo conjunto de triples de números enteros, uno de los cuales puede ser fácilmente demostrado que tienen sólo un punto fijo y el otro de los cuales tiene un punto fijo para cada representación de un primer dado

(congruente con 1 mod 4) como una suma de dos cuadrados. Desde la primera involución tiene un número impar de puntos fijos, también lo hace la segunda, y por lo tanto siempre existe una representación de la forma deseada.

3.18.3 Lista de teoremas de punto fijo

- Teorema del punto fijo de Atiyah-Bott
- Teorema del punto fijo de Banach
- Teorema del punto fijo de Borel
- Teorema del punto fijo de Browder
- Teorema del punto fijo de Brouwer
- Teorema del punto fijo de Caristi
- Lema Diagonal , también conocido como el lema de punto fijo, para la producción de oraciones autorreferenciales de la lógica de primer orden
- Teorema del punto fijo de Kakutani
- Teorema del punto fijo de Kleene
- Teorema de Knaster-Tarski
- Teorema del punto fijo de Lefschetz
- Teorema del punto fijo de Nielsen
- Teorema del punto fijo de Poincaré-Birkhoff
- Teorema del punto fijo de Schauder
- Teorema del punto fijo de Tychonoff

3.19 Clasificación de la inteligencia artificial

La inteligencia artificial (IA) es un área multidisciplinaria, que a través de ciencias como las ciencias de la computación, la matemática, la lógica y la filosofía, estudia la creación y diseño de sistemas capaces de resolver problemas cotidianos por sí mismos, utilizando como paradigma la inteligencia humana.

La IA se divide en dos escuelas de pensamiento:

- La inteligencia artificial convencional
- La inteligencia computacional

3.19.1 Inteligencia artificial convencional (IA suave)

Se conoce también como IA simbólico-deductiva. Está basada en el análisis formal y estadístico del comportamiento humano ante diferentes problemas:

Razonamiento basado en casos: Ayuda a tomar decisiones mientras se resuelven ciertos problemas concretos y, aparte de que son muy importantes, requieren de un buen funcionamiento. Sistemas expertos: Infieren una solución a través del conocimiento previo del contexto en que se aplica y ocupa de ciertas reglas o relaciones. Redes bayesianas: Propone soluciones mediante inferencia probabilística. Inteligencia artificial basada en comportamientos: Esta inteligencia contiene autonomía y puede auto-regularse y controlarse para mejorar. Smart process management: Facilita la toma de decisiones complejas, proponiendo una solución a un determinado problema al igual que lo haría un especialista en la dicha actividad.

3.19.2 Inteligencia artificial computacional (IA dura)

Inteligencia Computacional (IC) es una rama de la inteligencia artificial centrada en el estudio de mecanismos adaptativos para permitir el comportamiento inteligente de sistemas complejos y cambiantes. Se presenta como una alternativa a la GOFAI ("Good Old-Fashioned Artificial Intelligence"), tratando de no confiar en algoritmos heurísticos tan habituales en la Inteligencia Artificial más tradicional. Dentro de la Inteligencia Computacional podemos encontrar técnicas como las Redes Neuronales, Computación Evolutiva, Swarm Intelligence, Sistemas Inmunes Artificiales o Sistemas difusos. También se relaciona con técnicas como los Fractales, Teoría del Caos, Wavelets, Autómata celular, etc.

La Inteligencia Computacional combina elementos de aprendizaje, adaptación, evolución y Lógica difusa para crear programas que son, en cierta manera, inteligentes. La investigación en Inteligencia Computacional no rechaza los métodos estadísticos, pero muy a menudo aporta una vista complementaria. Las Redes Neuronales son una rama de la inteligencia computacional muy relacionada con el aprendizaje automático.

3.20 Teorema de los cuatro colores

En teoría de grafos, el teorema de los cuatro colores (o teorema de la minimalidad cromática) es un teorema sobre la coloración de grafos que establece lo siguiente:

Dado cualquier mapa geográfico con regiones continuas, este puede ser coloreado con cuatro colores diferentes, de forma que no queden regiones adyacentes con el mismo color.

Asumiendo que las regiones adyacentes comparten no solo un punto, sino todo un segmento de borde (frontera) en común.

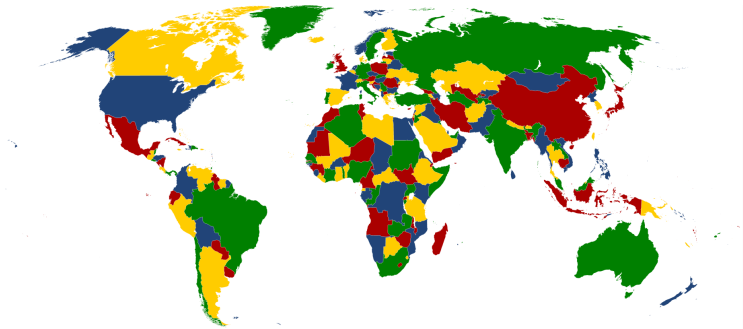
Tres colores son suficientes para mapas simples, pero en algunos casos es necesario un cuarto color adicional, esto es, cuando una región a colorear queda encerrada por un número impar de regiones que se tocan formando un ciclo. El teorema de los cinco colores, cuya demostración es corta y elemental, establece que cinco colores son suficientes para colorear un mapa y fue probado en el siglo XIX por Heawood.¹ Una serie de pruebas falsas y falsos contraejemplos han aparecido desde el primer enunciado del teorema de los cuatro colores en 1852.

En primer lugar, todas las esquinas y puntos en común que pertenecen a tres o más países, deben ser ignoradas. Sin esta restricción, los mapas extraños (utilizando las regiones del área finita pero perímetro infinito) pueden requerir más de cuatro colores.

Una versión más simple del teorema utiliza la teoría de grafos. El conjunto de las regiones de un mapa se puede representar de manera más abstracta como un grafo simple no dirigido asociando un vértice para cada región y una arista para cada par de regiones que comparten un segmento de borde. Esta representación del mapa con vértices y aristas es un grafo dual y el problema de colorear países se cambia por la coloración del grafo. Este grafo es plano, o sea, que se puede dibujar en el plano sin cruce de aristas mediante la colocación de cada vértice en un lugar elegido arbitrariamente dentro de la región a la que corresponde. Con la terminología de la teoría de grafos, el teorema de cuatro colores establece que:

Si G es un grafo plano, entonces $\chi(G) \leq 4$

Es decir, los vértices de cada grafo plano pueden ser coloreados con un máximo de cuatro colores de modo que no existan dos vértices adyacentes con el mismo color. $\chi(G)$ corresponde al número cromático.



CHAPTER FOUR

Apéndices

4.1 Examen propuesto

PRIMER EXAMEN DE PROBABILIDAD, PROCESOS ALEATORIOS E INFERENCIA

Elaborado por: Aldo Nicolás Alejandro Miranda Aguilar

Nombre: _____

1. Un estudiante responde al azar a 4 preguntas de verdadero o falso.

- Escriba el espacio muestral.
- Escriba el suceso responder "falso" a una sola pregunta.
- Escriba el suceso responder "verdadero" al menos a 3 preguntas.
- Escriba la unión de estos dos sucesos, la intersección y la diferencia entre el 2 y el 1.
- La colección formada por estos 5 sucesos, más el suceso seguro y el suceso imposible Constituyen un σ -álgebra?

Solution:

- El espacio muestral es el conjunto de todos los sucesos elementales. Los sucesos elementales son cada uno de los resultados posibles del experimento aleatorio, indescomponibles en otros más simples. Como el experimento consiste en responder al azar a cuatro preguntas, cada uno de los posibles patrones de respuesta constituirá un suceso elemental.

$$E = \{(V, V, V, V), (V, V, V, F), (V, V, F, V), (V, F, V, V), (F, V, V, V), (V, V, F, F), (V, F, V, F), (V, F, F, V), (F, V, V, F), (F, V, F, V), (F, F, V, V), (V, F, F, F), (F, V, F, F), (F, F, V, F), (F, F, F, V), (F, F, F, F)\}$$

- El Suceso responder falso a una sola pregunta será el subconjunto del espacio muestral formado por todos los sucesos elementales en que solo hay una respuesta falso, lo llamaremos A y será:

$$A = \{(V, V, V, F) \cup (V, V, F, V) \cup (V, F, V, V) \cup (F, V, V, V)\}$$

- El suceso responder verdadero al menos a 3 preguntas, lo llamaremos B y será:

$$B = (V, V, V, F) \cup (V, V, F, V) \cup (V, F, V, V) \cup (F, V, V, V) \cup (V, V, V, V)$$

- Observando los sucesos elementales que los componen se deducen inmediatamente los siguientes resultados:

$$A \cup B = B; A \cap B = A; B - A = \{(V, V, V, V)\}$$

- La colección formada por el suceso A, el B, la unión de ambos, su intersección, y su diferencia, más el suceso seguro y el suceso imposible, no constituye un σ -álgebra. Para demostrarlo basta comprobar que se incumple una de las dos condiciones. Por ejemplo, el suceso A incumple la segunda porque su contrario no pertenece a la colección

2. La prevalencia de la diabetes es del 4%. La glucemia basal diagn stica correctamente el 95% de los diab ticos, pero da un 2% de falsos positivos. Diagnosticada una persona Cu l es la probabilidad de que realmente sea diab tica?

Solution: Sea D el suceso de tener diabetes, \overline{D} el suceso de no tenerla y $Gl+$ el suceso de dar positivo en la prueba de la glucemia basal. Los datos del problema nos dicen que:

$$P(D) = 0.04; P(\overline{D}) = 0.96; P(Gl+|D) = 0.95; P(Gl+|\overline{D}) = 0.02$$

Entonces el teorema de Bayes, escrito en los t rminos de este problema nos dice que:

$$P(D|Gl+) = \frac{P(Gl+|D)P(D)}{P(Gl+|D)P(D) + P(Gl+|\overline{D})P(\overline{D})}$$

sustituyendo:

$$P(D|Gl+) = \frac{0.95(0.04)}{0.95(0.04) + 0.02(0.96)} = \frac{0.038}{0.038 + 0.0192} = 0.664$$

3. En un sistema de comunicaci n, los mensajes se codifican en base 3, es decir, con los tres s mbolos 0, 1 y 2. La probabilidad de emitir cualquiera de los tres s mbolos es la misma y se ha observado que el 0 y el 2 nunca se confunden, es decir, la probabilidad de recibir 0 si se ha emitido 2 (respectivamente, recibir 2 si se ha emitido 0) es cero. Adem s, sea cual sea el s mbolo que se emite, la probabilidad de recibir ese mismo s mbolo es 0,9 y finalmente se sabe que cuando se emite 1, la probabilidad de recibir 0 y la probabilidad de recibir 2, es la misma. Calcular:

- (a) La probabilidad de recibir 1
(b) La probabilidad de que el s mbolo emitido haya sido 0 si se recib  0

Solution:

- (a) Se definen los sucesos:

E_0 = "se emite el s mbolo 0",
 E_1 = "se emite el s mbolo 1",
 E_2 = "se emite el s mbolo 2",
 R_0 = "se recibe el s mbolo 0",
 R_1 = "se recibe el s mbolo 1" y
 R_2 = "se recibe el s mbolo 2"

Con esta notaci n, se tiene por los datos del enunciado que

$$P(E_0) = P(E_1) = P(E_2) = \frac{1}{3}$$

$$P(R_0|E_2) = P(R_2|E_0) = 0$$

$$P(R_0|E_0) = P(R_1|E_1) = P(R_2|E_2) = 0.9$$

$$P(R_0|E_1) = P(R_2|E_1)$$

Por tanto, la probabilidad de recibir 1, usando el teorema de la probabilidad total, es

$$\begin{aligned} P(R_1) &= P(R_1 \cap (E_0 \cup E_1 \cup E_2)) \\ &= P(R_1|E_0)P(E_0) + P(R_1|E_1)P(E_1) + P(R_1|E_2)P(E_2) \\ &= (1 - P(R_0|E_0) - P(R_2|E_0)) \cdot \frac{1}{3} + 0.9 \cdot \frac{1}{3} + (1 - P(R_0|E_2) - P(R_2|E_2)) \cdot \frac{1}{3} \\ &= (1 - 0.9 - 0) \cdot \frac{1}{3} + 0.9 \cdot \frac{1}{3} + (1 - 0 - 0.9) \cdot \frac{1}{3} = 0.3667 \end{aligned}$$

(b) Para resolver este apartado se utiliza el teorema de Bayes, de forma que:

$$\begin{aligned} P(R_0|E_0) &= \frac{P(R_0|E_0)P(E_0)}{P(R_0|E_0)P(E_0) + P(R_0|E_1)P(E_1) + P(R_0|E_2)P(E_2)} \\ &= \frac{0.9 \cdot \frac{1}{3}}{0.9 \cdot \frac{1}{3} + P(R_0|E_1) \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} \end{aligned}$$

Dado que $P(R_0|E_1) = P(R_2|E_1) = x$, se tiene que al ser $P(R_1|E_1) = 0.9$, entonces $x + 0.9 + x = 1$ y por tanto $x = 0.05$. Por último

$$P(E_0|R_0) = \frac{0.9 \cdot \frac{1}{3}}{0.9 \cdot \frac{1}{3} + 0.05 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = 0.9474$$

4. Decir si es verdadera o falsa la siguiente afirmación. En caso de que sea verdadera, demostrarlo y en caso de que sea falsa dar un contraejemplo:

"Si $B \subset A$ con $0 < P(A) < 1$, entonces B y A no pueden ser independientes"

Solution:

$$\text{Si } B \subset A$$

$$\Rightarrow A \cap B = B$$

$$\Rightarrow P(A \cap B) = P(B)$$

Por lo tanto, para que $P(A \cap B) = P(A) \cdot P(B)$, tendría que ser necesario que $P(A) = 1$.

Por lo que podemos decir que la afirmación es **verdadera**.

5. De cuántas formas diferentes se pueden cubrir los puestos de presidente, vicepresidente y tesorero de un club de fútbol sabiendo que hay 12 posibles candidatos?

Solution:

NO entran todos los elementos

Sí importa el orden

NO se permiten repeticiones

Por lo tanto

$$\# \text{ de formas} = \text{Permutaciones}(12, 3) = \frac{12!}{(12-3)!} = 1320$$

6. Una persona cruza dos semáforos para ir al trabajo. La probabilidad de que cada uno de ellos esté rojo es de 0.4; de que esté amarillo, 0.2, y de que esté verde, 0.4. Elabore el árbol de probabilidad y calcule la probabilidad de que uno esté en verde y el otro en rojo.

Solution:

Se aplica la regla de la suma o de la probabilidad total.

$$P(RV) + P(VR) = 0.4 \cdot 0.4 + 0.4 \cdot 0.4 = 0.32$$

7. Un estudiante para aprobar un examen que consta de 10 preguntas, debe contestar 7 de ellas. De cuántas maneras puede hacer la selección para aprobar el examen?

Solution:
NO entran todos los elementos
NO importa el orden
NO se permiten repeticiones

Por lo tanto el resultado lo obtenemos con las combinaciones de 7 en 10

$$\binom{10}{7} = \frac{10!}{7!(10-7)!} = 120$$

8. Demuestre que $\binom{n}{r}$ es simétrico para $0 \leq r \leq n$

Solution: Para que $\binom{n}{r}$ sea simétrico se debe cumplir que:

$$\binom{n}{r} = \binom{n}{n-r}$$

Entonces construimos con $r = (n - r)$

$$\begin{aligned} \binom{n}{n-r} &= \frac{n!}{(n-r)!(n-(n-r))!} \\ &= \frac{n!}{(n-r)!(n-n+r)!} \\ &= \frac{n!}{r!(n-r)!} \\ &= \binom{n}{r} \end{aligned}$$

9. El portero titular de un equipo de fútbol para 8 de cada 10 penaltis, mientras que el suplente solo para 5. el portero suplente juega, por termino medio, 15 minutos en cada partido (90 minutos). Si se lanza un penalti y no se para cuál es la probabilidad de que estuviera jugando el portero titular?

Solution:

Se consideran los sucesos:

P = el portero para un penalti

T = juega el portero titular

S = juega el portero suplente ($S = T^c$)

Con probabilidades:

$$\begin{aligned} P(S) &= \frac{15}{90} = \frac{1}{6} \\ P(T) &= 1 - P(S) = 1 - \frac{1}{6} = \frac{5}{6} \\ P(P|T) &= \frac{8}{10} = \frac{4}{5} \\ P(P^c|T) &= 1 - \frac{4}{5} = \frac{1}{5} \\ P(P|S) &= \frac{5}{10} = \frac{1}{2} \\ P(P^c|S) &= 1 - \frac{1}{2} = \frac{1}{2} \end{aligned}$$

para calcular $P(T|P^c)$ se aplica el teorema de Bayes, con los sucesos T y S como sistema completo de sucesos:

$$P(T|P^c) = \frac{P(P^c|T)P(T)}{P(P^c|T)P(T) + P(P^c|S)P(S)} = \frac{\frac{1}{5} \cdot \frac{2}{6}}{\frac{1}{5} \cdot \frac{2}{6} + \frac{1}{2} \cdot \frac{1}{6}} = \frac{10}{15} \approx 0.66667$$

10. Cuántas diagonales se pueden trazar dentro de un pentágono regular?

Solution: Una diagonal la podemos definir como la recta que va desde un vértice a a otro vértice b dentro del pentágono, tal que b no sea vecino de a y $a \neq b$, por lo tanto, podemos decir que para cada vértice del polígono podemos escoger otros dos vértices con los cuales formar una diagonal y entonces podemos decir que el número de vértices esta dado por

$$\binom{5}{2}$$

Sin embargo, hay que tomar en cuenta que la diagonal ab es la misma que la diagonal ba , por lo que tenemos que dejar de contar estas repeticiones dividiendo entre 2, entonces

$$\# \text{ de diagonales} = \frac{1}{2} \cdot \binom{5}{2} = \frac{1}{2} \cdot 10 = 5$$

4.2 Examen resuelto

1. ¿Cuántos números de 4 cifras diferentes se pueden formar (base 10)? ¿Y cuántos de 4 cifras sin importar si se repiten?

Solution: Se calculan las permutaciones posibles, en el primer caso sin repetición y en el segundo caso con repetición. Tomando en cuenta que el primer dígito no debe de ser cero, entonces el número de cifras está dado por:

$$9 \cdot P(9, 3) = 9 \cdot \frac{n!}{(n-r)!} = 9 \cdot \frac{9!}{(9-3)!} = 4536$$

Para el segundo problema, teniendo en cuenta la restricción del primer dígito diferente de cero, tenemos que:

$$9 \cdot n^{(r-1)} = 9 \cdot 10^3 = 9000$$

2. Una familia de 12 individuos se reúne para festejar el año nuevo. Al sonar las 12 campanadas se comienzan a abrazar para felicitarse, siempre de dos en dos. Si nadie olvidó abrazar a ninguno de sus parientes, ¿cuántos abrazos se realizaron en total?

Solution: Se calculan el número de combinaciones posibles entre los 12 individuos, tomándolos de dos en dos.

$${}^{12}C_2 = \frac{n!}{r!(n-r)!} = \frac{12!}{2!(12-2)!} = 66$$

3. Se tiene un espacio muestral E formado por los sucesos S_1 , S_2 y S_3 tales que

$$S_1 \cup S_2 \cup S_3 = E$$

$$S_1 \cap S_3 = \emptyset$$

$$S_2 \cap S_3 = \emptyset$$

donde

$$P(S_1) = \frac{1}{3}; \quad P(S_1 \cup S_2) = \frac{5}{7}; \quad P(S_3) = \frac{2}{7}$$

Determinar si están bien definidas las probabilidades.

Solution: Para determinar si están bien definidas las probabilidades se tienen que comprobar los dos primeros axiomas de Kolmogorov.

Para el primer axioma observamos que las probabilidades de los eventos son no negativas.

Para el segundo axioma tenemos lo siguiente:

$$P(E) = 1$$

Y como $S_1 \cup S_2$ y S_3 son disjuntos, podemos aplicar el tercer axioma de la siguiente manera.

$$P(E) = P(S_1 \cup S_2) + P(S_3) = \frac{5}{7} + \frac{2}{7} = 1$$

$$P = \frac{1,098,240}{2,598,960} = 0.42256903$$

5. Se tienen tres bolsas numeradas del uno al tres. Cada una contiene 10 pelotas distribuidas de la siguiente forma:

- 5 azules, 5 rojas
- 3 azules, 3 rojas, 4 blancas
- 3 azules, 7 rojas

Se extrae una pelota de la primera bolsa y se mete en la segunda, posteriormente se extrae una pelota de la segunda bolsa y se introduce en la tercera, y por último se extrae una pelota de la tercera bolsa.

¿Cuál es la probabilidad de que la pelota extraída de la tercera bolsa sea azul?

Solution: Todas las posibles combinaciones en las bolsas las podemos observar en el siguiente árbol del probabilidades

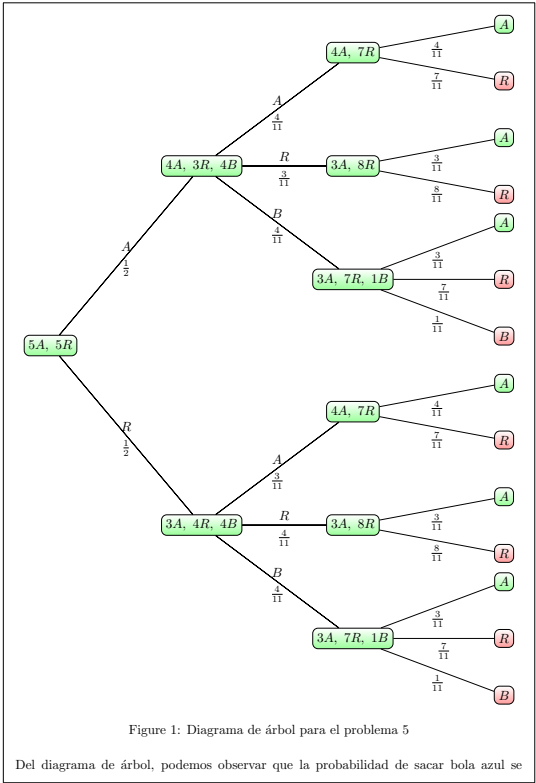


Figure 1: Diagrama de árbol para el problema 5

Del diagrama de árbol, podemos observar que la probabilidad de sacar bola azul se

obtiene de la siguiente manera:

$$P(A) = \frac{1}{2} \cdot \frac{4}{11} \cdot \frac{4}{11} + \frac{1}{2} \cdot \frac{3}{11} \cdot \frac{3}{11} + \frac{1}{2} \cdot \frac{4}{11} \cdot \frac{3}{11} + \frac{1}{2} \cdot \frac{3}{11} \cdot \frac{4}{11} + \frac{1}{2} \cdot \frac{4}{11} \cdot \frac{3}{11} + \frac{1}{2} \cdot \frac{3}{11} \cdot \frac{3}{11} =$$

$$P(A) = \frac{73}{242}$$

6. Un producto esta compuesto de cuatro piezas. Las probabilidades de que alguna de estas piezas sea defectuosa es 1 por millar para la pieza uno, 3 por millar para la pieza dos, 12 por millar para la pieza 3 y 9 por millar para la pieza 4.

Calcular la probabilidad de que el producto tenga alguna pieza defectuosa.

Solution: Para calcular la probabilidad requerida se necesita determinar la probabilidad de que sean defectuosas la primera pieza, la segunda pieza, la tercera pieza, la cuarta pieza, la primera y la segunda piezas, la primera y la tercera piezas, ... Tomando todas las combinaciones posibles entre las diferentes piezas.

Alternativamente, se puede calcular la probabilidad del complemento del evento de interés (algunas piezas defectuosas), es decir, la probabilidad de que ninguna pieza sea defectuosa, y como estos eventos son complementarios, restamos el resultado obtenido a la unidad, y de esta forma obtenemos el resultado que nos interesa.

Definimos el evento *alguna pieza defectuosa* como D y el evento *ninguna pieza defectuosa* como \overline{D} . También definimos los eventos D_1, D_2, D_3, D_4 para los casos en que cada una de las piezas es defectuosa, y los eventos $\overline{D}_1, \overline{D}_2, \overline{D}_3, \overline{D}_4$ para los casos en que cada una de las piezas no es defectuosa.

$$P(\overline{D}) = P(\overline{D}_1) \cdot P(\overline{D}_2) \cdot P(\overline{D}_3) \cdot P(\overline{D}_4)$$

$$P(\overline{D}) = \frac{999}{1000} \cdot \frac{997}{1000} \cdot \frac{988}{1000} \cdot \frac{991}{1000} = 0.975194505$$

$$P(D) = 1 - P(\overline{D}) = 1 - 0.975194505 = 0.024805494$$

7. Se tienen 3 dados trucados y 1 normal. En los dados trucados la probabilidad de sacar 6 es la mitad que la de cualquiera de los otros cinco números, los cuales tienen la misma probabilidad de salir.

Se elige un dado al azar, se lanza y se obtiene un 6. ¿Cuál es la probabilidad de que el dado lanzado haya sido el dado normal?

Solution: En los dados trucados, la probabilidad de obtener un seis se puede calcular de la siguiente manera:

$$P(\text{total}) = P(1) + P(2) + P(3) + P(4) + P(5) + P(6)$$

Haciendo $P(1) = P(2) = P(3) = P(4) = P(5) = p$ y $P(6) = \frac{1}{2}p$ tenemos:

$$P(\text{total}) = p + p + p + p + p + \frac{1}{2}p$$

Despejando p :

$$p = \frac{2}{11}$$

Y para la probabilidad de obtener 6:

$$P(6) = \frac{1}{2}p = \frac{1}{11}$$

Ahora calculando la probabilidad condicional de que el dado lanzado haya sido un dado normal sabiendo que se obtuvo un seis $P(\text{normal}/6)$.

$$P(\text{normal}/6) = \frac{P(6/\text{normal}) \cdot P(\text{normal})}{P(6/\text{normal}) \cdot P(\text{normal}) + P(6/\text{trucado}) \cdot P(\text{trucado})}$$

$$P(\text{normal}/6) = \frac{\frac{1}{6} \cdot \frac{1}{4}}{\frac{1}{6} \cdot \frac{1}{4} + \frac{1}{11} \cdot \frac{3}{4}} = \frac{\frac{1}{24}}{\frac{1}{24} + \frac{3}{44}} = \frac{\frac{1}{24}}{\frac{23}{66}} = \frac{11}{29}$$

8. En el año 2010 se estimaba que 220,000 personas vivían con SIDA en México. En ese año México tenía una población de 117.9 millones de habitantes.

Existe una prueba, ELISA, la cual si se efectúa tres meses después del contagio de SIDA, tiene un 97% de certeza para detectar la enfermedad. Esta prueba pueda dar un falso positivo en el 1% de los casos.

Si una persona se realiza esta prueba tres meses después de un posible contagio y da positivo, ¿Cuál es la probabilidad de que esté enferma?

Solution: Basándonos en los datos estadísticos proporcionados para el 2010, podemos calcular que la probabilidad de tener SIDA en ese año era:

$$P(\text{enfermo}) = P(E) = \frac{220,000}{117,900,000} = \frac{11}{5,895}$$

Calculando la probabilidad de dar positivo en la prueba ELISA:

$$P(Po) = P(Po/E)P(E) + P(Po/S)P(S) = 0.97 \cdot \frac{11}{5,895} + 0.01 \cdot \frac{5,884}{5,895} = 0.011791348$$

Y calculando la probabilidad de estar enfermo al realizar la prueba tres meses después de un posible contagio:

$$P(E/Po) = \frac{P(Po/E) \cdot P(E)}{P(Po)} = \frac{0.97 \cdot \frac{11}{5,895}}{0.011791348} = \frac{1067}{6951} = 0.153503093$$

9. En una oficina hay tres empleados y todos al llegar dejan sus llaves encargadas con el guardia de la entrada. Uno de los llaveros tiene 2 llaves, el segundo tiene 3 llaves y el último tiene 5 llaves.

Todos los empleados tienen una llave del baño, el cual está ubicado en el pasillo que lleva a la oficina.

Uno de los empleados está enfermo del estómago y en un determinado momento sale corriendo al baño y le pide al guardia que le de las primeras llaves que encuentre.

Al llegar al baño prueba con la primera llave que puede tomar y, para su fortuna, abre la puerta.

Determinar la probabilidad de que el llavero que le dió el guardia haya sido el que tiene 3 llaves.

Solution:

$$P(3/A) = \frac{P(A/3) \cdot P(3)}{P(A/3) \cdot P(3) + P(A/2) \cdot P(2) + P(A/5) \cdot P(5)}$$

$$P(3/A) = \frac{\frac{1}{3} \cdot \frac{1}{3}}{\frac{1}{3} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{5} \cdot \frac{1}{3}} = \frac{\frac{1}{9}}{\frac{1}{9} + \frac{1}{6} + \frac{1}{15}} = \frac{10}{31}$$

10. Exprese la probabilidad de la unión de tres eventos en términos de las probabilidades de cada uno de los eventos y de sus intersecciones.

Solution: Se tienen los conjuntos S_1 , S_2 y S_3 .

La probabilidad de la unión de estos conjuntos se calcula de la siguiente manera:

$$\begin{aligned} P(S_1 \cup S_2 \cup S_3) &= \\ &= P(S_1 \cup (S_2 \cup S_3)) = \\ &= P(S_1) + P(S_2 \cup S_3) - P(S_1 \cap (S_2 \cup S_3)) = \end{aligned}$$

$$\begin{aligned}
&= P(S_1) + P(S_2) + P(S_3) - P(S_2 \cap S_3) - P((S_1 \cap S_2) \cup (S_1 \cap S_3)) = \\
&= P(S_1) + P(S_2) + P(S_3) - P(S_2 \cap S_3) - P(S_1 \cap S_2) - P(S_1 \cap S_3) + P(S_1 \cap S_2 \cap S_3)
\end{aligned}$$

Bibliography

- [1] J. Adler. *R in a Nutshell: A Desktop Quick Reference*. O'Reilly Media, 2010.
- [2] D.P. Bertsekas and J.N. Tsitsiklis. *Introduction to Probability*. Athena Scientific books. Athena Scientific, 2002.
- [3] J.K. Blitzstein and J. Hwang. *Introduction to Probability*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2014.
- [4] S. Lipschutz. *Probabilidad*. Serie de Compendios Schaum. McGraw-Hill, 1975.
- [5] J. Schiller, R.A. Srinivasan, and M. Spiegel. *Schaum's Outline of Probability and Statistics, 4th Edition*. Schaum's Outline Series. McGraw-Hill Education, 2012.