

Apuntes y Tareas

Probabilidad e Inferencia



Luis Norberto Zúñiga Morales
Centro de Investigación en Computación

Contents

1	Tarea 1	6
1.1	Experimentos deterministas	6
1.2	Experimentos aleatorios	7
1.3	Espacios muestrales	8
1.3.1	El espacio muestral de arrojar una moneda	8
1.3.2	El espacio muestral de una mano de póquer	8
1.3.3	El espacio de nuestra lotería	9
1.3.4	Tiro con arco	9
1.3.5	La democracia en un país	9
1.4	Algunas definiciones	9
1.5	Principio del Palomar	12
1.6	Expansiones del binomio de Newton	12
2	Tarea 2	17
2.1	Ejemplos de permutación	17
2.2	Ejemplos de combinación	18
2.3	El problema de los caminos posibles en un rectángulo	20
2.4	Pirámide de Pascal	21
2.5	El problema de los caminos posibles en un cubo	22
2.6	Reseñas	22
2.6.1	Primates Count	22
2.6.2	Cigada-Generated Prime Bumpers	22
2.6.3	Ant Odometer	23
2.6.4	Quipu	23
2.6.5	Magic Squares	23
2.7	Demostraciones	23
3	Tarea 3	24
3.1	Telómeros	24
3.2	Circuitos lógicos	24
3.3	Permisos en archivos y carpetas en Unix	25
4	Tarea 4	26
4.1	Paradojas y antinomias	26
4.2	Paradojas en las Matemáticas	27
4.2.1	Paradoja de Newcomb	27
4.2.2	Paradoja de Bertrand	28
4.2.3	La paradoja de Monty Hall	29
4.3	Demostraciones	29
5	Tarea 5	30
5.1	Paradoja del Falso Positivo	30
5.2	Falacia del Apostador	32
5.3	Paradoja de Borel - Kolmogorov	33
5.4	Problema 1.2	34

5.5	Histograma en ROOT	36
6	Tarea 6	36
6.1	Teorema del límite central	36
6.2	Teoría de la Ruina del Jugador	36
6.3	Histograma de números generados aleatoriamente según una distribución normal . .	38
6.4	Mínimos cuadrados	39
6.5	El hombre anumérico	41
7	Tarea 7	42
7.1	Problema de los 4 tornillos	42
7.2	Problemas Regla de Bayes	45
7.3	Grafos de divisibilidad	48
8	Tarea 8	50
8.1	Tensor de Levi-Civita	50
8.2	La máquina enigma	51
8.3	Esperanza matemática	52
8.4	Un pájaro en mano vale más que 2.48 volando	53
8.5	Tensor covariante y contravariante	54
9	Tarea 9	56
9.1	Ventaja	56
9.2	El problema del caballero De Meré	59
10	Tarea 10	60
10.1	Congruencia de Zeller	60
11	Tarea 11	62
11.1	Momentos	62
11.2	Función generadora de momentos	62
11.3	Sesgo	63
11.4	Curtosis	64
12	Tarea 12	65
12.1	Π	65
12.2	Algoritmo de Gauss-Legendre	66
12.3	Algoritmo de Borwein	67
12.4	Enteros largos	67
12.5	Variable aleatoria	70
12.5.1	Distribución y esperanza de una variable aleatoria finita	71
12.5.2	Varianza y desviación estándar	71
12.5.3	Distribución conjunta	72
12.5.4	Variables aleatorias independientes	73
12.5.5	Funciones de una variable aleatoria	74
12.5.6	Variables aleatorias discretas en general	74
12.5.7	Variables aleatorias continuas	75

12.5.8	Función de distribución acumulada	76
12.5.9	Desigualdad de Tchebycheff y Ley de los Grandes Números	77
13	Tarea 13	77
13.1	Is Economics the Next Physical Science?	77
13.2	Función de distribución de probabilidad	78
14	Tarea 14	79
14.1	Análisis ROC	79
15	Tarea 15	81
15.1	Algunas demostraciones para valores medios	81
15.2	Una demostración útil	83
15.3	Aplicaciones de caminatas aleatorias	83
15.4	Caminata aleatoria	84
16	Tarea 16	84
16.1	Función Gamma	84
16.2	Otras demostraciones para distribuciones de probabilidad para N grande	85
16.3	Simulación de Movimiento Browniano	89
17	Tarea 17	89
17.1	Función delta de Dirac	89
17.2	Caminata aleatoria (parte 2)	96
17.3	Caminata aleatoria en dos dimensiones	96
18	Tarea 18	96
18.1	Relación entre la distribución normal y la Gaussiana	96
18.2	Graficar puntos de mathematica en Tikz	98
18.3	Delta de dirac en coordenadas polares	99
19	Tarea 19	99
19.1	Distribuciones de probabilidad	99
19.1.1	Disitribución de Bernoulli	99
19.1.2	Distribución binomial	100
19.1.3	Distribución uniforme	101
19.1.4	Distribución exponencial	102
19.1.5	Distribución normal	103
19.1.6	Distribución de Poisson	104
19.1.7	Distribución hipergeométrica	106
19.1.8	Distribución de Cauchy	107
19.1.9	Distribución gamma	108
19.1.10	Distribución beta	109
19.2	Procesos NP	111
19.3	El hombre anumérico capítulo 2	112
19.4	Fork	114
19.5	Problemas selectos	114

20 Tarea 20	117
20.1 Eigenvalores y Eigenvectores	117
20.2 Matrices similares	119
21 Tarea 21	120
21.1 Matrices poco densas	120
21.2 Cadenas de Markov	122
21.3 Teorema del punto fijo	127
21.3.1 Teorema del punto fijo de Banach	127
21.3.2 Teorema del punto fijo de Brouwer	127
22 Tarea 22	128
22.1 Clasificación de la inteligencia artificial	128
22.2 Problema de los 4 colores	130
22.3 Experimentos de Mendel	131

1 Tarea 1

1.1 Experimentos deterministas

1. Calentar una barra de hierro: si se somete una barra de hierro a una fuente de calor, con el tiempo se dilatará al llegar a cierta temperatura. Si dejamos enfriar la barra y la sometemos nuevamente al mismo procedimiento se observará el mismo resultado anterior.
2. Tirar una piedra desde una lugar elevado: aunque un poco más exigente debido a que hay que vigilar aspectos como el que no se presente una ráfaga de viento cuando la piedra cae, al repetir el experimento se observarán los mismos resultados una y otra vez.
3. Introducir un cuerpo a un recipiente con agua: si introducimos el mismo cuerpo, aún en recipientes con distinto volumen de agua, el volumen desplazado por el cuerpo será el mismo.
4. Lanzar una piedra (tiro parabólico): suponiendo que de alguna forma se es capaz de mantener el mismo ángulo con que se lanza una piedra y se usa la misma fuerza en cada lanzamiento, podemos predecir y observar la distancia que recorre y el punto donde cae.
5. Cálculo del modulo de Young de un material: mediante ensayos de tracción del material se puede calcular el modulo para éste. El valor será el mismo, al considerar que el material es exactamente el mismo.
6. Cálculo del coeficiente de dilatación de materiales: de igual manera, si se observa una cambio en la longitud ΔL de un material debido a cambio de temperatura ΔT y al considerar que esta relación es proporcional, la razón o proporción del cambio de L esta dada por

$$\Delta L = \alpha L \Delta T$$

Si calculamos de manera experimental el valor de α para cierto material, al repetir el experimento encontraremos el mismo valor para el coeficiente de expansión lineal.

7. Reacciones químicas: en la química, por ejemplo, si al carbonato de calcio se le comunica calor este se descompone en óxido de calcio y dióxido de carbono. De manera general, ya existen procesos para obtener sustancias a través de distintas reacciones químicas que siempre generan el mismo resultado.
8. Velocidad de un cuerpo: si se mantiene la aceleración constante de un cuerpo, al medir su velocidad al medir un intervalo de tiempo y distancia. Si se repite este experimento bajo las mismas condiciones, el resultado será el mismo.
9. Probabilidad crítica de percolación: el fenómeno de percolación se da cuando el parámetro p o probabilidad de percolación excede una cantidad llamada probabilidad crítica p_c . A pesar de la naturaleza aleatoria del fenómeno, si $p < p_c$ no se generará el clúster infinito; por otro lado, si $p > p_c$ siempre se va a generar dicho clúster infinito. Es decir, dependiendo del valor de p que se maneje, se presentará o no el clúster.
10. Decaimiento de temperatura en un cuerpo: la ley del enfriamiento de Newton nos dice que la velocidad de enfriamiento de un cuerpo a temperatura T en un ambiente a temperatura T_f es proporcional a la diferencia entre la temperatura instantánea del cuerpo y la del ambiente [8], es decir,

$$\frac{dT(t)}{dt} = -k(T - T_f)$$

al resolver la ecuación diferencial usando $T(0) = T_0$ como condición inicial, obtenemos un decaimiento exponencial de la temperatura

$$T(t) = T_f + (T_0 - T_f)e^{-kt}$$

este resultado puede obtener experimentalmente al calentar un cuerpo o sustancia y se logra observar como la temperatura del objeto decae de dicha manera.

1.2 Experimentos aleatorios

1. Muestras de sangre: el mismo nombre lo indica, en una primera de sangre se pueden observar ciertos resultados (colesterol, urea, etc.), pero el resultado va a diferir de muestra en muestra.
2. Tirar un dado de n caras: al tirar n veces un dado justo, es imposible predecir que número resultará o que sucesión de números ocurrirá.
3. Barajar cartas: al suponer que una baraja de n cartas esta bien barajada, cualquier combinación de cartas es posible en una mano, por lo que es imposible predecir que mano tendrá uno al inicio. Adems, el número de maneras en que estas 52 cartas pueden estar ordenadas es $52!$, número cercano a 10^{67} .
4. Jalar una carta en específico de una mazo de cartas: extendiendo el ejemplo anterior, contando cartas se puede calcular la probabilidad de que la carta que jale de las m restantes, con $m < n$, sea una en específico. Aún así, predecir cual saldrá es imposible, al final una probabilidad no deja de ser nada más que eso, a menos que solo quede una.
5. Estimar el precio de una acción de tu empresa favorita al día siguiente: resulta fácil y atractivo pensar en que se pueda predecir con modelos matemáticos el precio de una acción, pero al depender de muchos factores y ser muy sensible a ruidos externos, solo se pueden dar estimaciones o intervalos del precio en el día que sigue.
6. Observar botellas en una linea de llenado: si se analiza la cantidad de líquido que se introduce en una botella en una linea de llenado, se observará que habrá aquellas que tengan más o menos producto debido a la máquina que las llena. De manera más clara, cada botella tendrá un cierto volumen l más/menos cierta cantidad s , es decir, $l \pm s$.
7. Observar el movimiento de un grano de polen moverse en un líquido: el biólogo escocés Robert Brown observó que partículas de polen al introducirlas en agua, se desplazaban sin ningún patrón reconocible: se movían de manera aleatoria debido al choque incesante entre las moléculas del agua y las partículas de polen.
8. Observación del número de accidentes viales: para una compaía de seguros, es de vital importancia obtener una cantidad razonable de cuanto dinero arriesga al asegurar vehículos. En cierto intervalo fijo de tiempo la cantidad de accidentes ocurridos diferirá y a lo más se puede estimar un resultado, más no calcular éste de manera determinista.

9. Estimación del clima de un determinado día: las predicciones del clima se realizan en base a modelos probabilísticos, pero el computo solo permite obtener estimaciones. El evento real, o el que sucede, evidentemente difiere de pronóstico a pronóstico, a excepción de casos donde el clima dependa de factores críticos como huracanes que propician lluvias o frentes fríos que propician temperaturas bajas.
10. Teoría de colas: la teoría de colas estudia todo lo relacionado con las filas; el que una persona llegue a un lugar y se forme ciertamente no depende de nosotros y por ejemplo, en un intervalo de una hora podemos observar que llegan n personas. Sin embargo, si repetimos el mismo ejercicio en la misma hora del siguiente día el número de personas variará indudablemente. Aunque el fenómeno en general tiene solución para preguntas como si es más efectivo una sola fila para varias cajas o una fila para cada caja o determinar tiempos de atención, se basa en un fenómeno meramente aleatorio, lo cual ejemplifica que un fenómeno de esta índole puede arrojar resultados deterministas.

1.3 Espacios muestrales

Se denomina espacio muestral al conjunto de todos los posibles resultados de un experimento aleatorio E , llamando a cada uno de sus elementos integrantes suceso o comportamiento elemental.

+

1.3.1 El espacio muestral de arrojar una moneda

En este evento, el espacio muestral esta conformado por los posibles resultados de arrojar una moneda, es decir

$$E = \{H, T\}$$

donde H representa cara (o Heads) y T cruz (o Tails).

Dado el caso donde se tiren dos monedas, ahora el espacio muestral esta conformado de la siguiente manera

$$E = \{HH, HT, TH, TT\}$$

donde la primera letra representa el resultado de la primera moneda y la segunda representa el valor del segundo tiro.

1.3.2 El espacio muestral de una mano de póquer

Por como se juega, en el póquer se empieza con una mano de cinco cartas de un mazo de 52 cartas (13 cartas para cada uno de los cuatro palos). El espacio muestral esta conformado por todas las posibles combinaciones de estas 5 de 52 cartas, la cardinalidad de este conjunto esta dada por

$$|E| = \frac{52!}{47!} = 311875200$$

Lo cual es un número relativamente grande, anotar todas las posibles manos no tiene sentido ya que lleva a una pérdida de tiempo.

1.3.3 El espacio de nuestra lotería

Cada cierto tiempo, los alumnos se reúnen para celebrar un sorteo en el que se rifa ya sea una bolsa acumulada de dinero o un producto en específico. El sorteo se hace mediante la impresión de boletos numerados de tres dígitos. De alguna manera aleatoriamente se eligen los tres dígitos ganadores y felizmente el boleto ganador se lleva el premio.

En este ejemplo el espacio muestral está formado por todos los boletos que se imprimen. Suponiendo que se empieza desde el boleto 001 y termina hasta el boleto 999

$$E = \{001, \dots, 999\}$$

Si suponemos ahora que, por alguna razón se eliminan todos los boletos que tengan cierto dígito, por ejemplo el 7, el espacio muestral está dado por todos los boletos que no tengan éste dígito, resultando en $9^3 = 729$ boletos.

1.3.4 Tiro con arco

Supongamos que practicamos el tiro con arco. Aunque el resultado del tiro en sí no es aleatorio, podemos elegir el ángulo de tiro de manera aleatoria. Si tenemos una diana que, para comodidad nuestra podemos representarla mediante la ecuación $x^2 + y^2 = 1$, donde una unidad equivale a un metro, podemos ubicar dos eventos posibles en nuestro ejercicio.

Al tirar aleatoriamente flechas a la diana, el espacio muestral lo podemos considerar como todas aquellas que pegaron en la diana, $A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ y todas aquellas que fallaron, $F = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 > 1\}$, es decir, $E = A \cup F$.

1.3.5 La democracia en un país

Supongamos que en cierto país su forma de asignar a los representantes populares es mediante una votación. Dicho esto, el territorio se divide en 252 distritos electorales y en la actualidad hay 5 partidos políticos que compiten en los comicios de este año.

Si consideramos nuestro espacio como todos los resultados posibles en que los 252 distritos pueden repartirse, el espacio muestral E está formado por $5^{252} \approx 1.381787 \times 10^{176}$

1.4 Algunas definiciones

Un espacio de probabilidad [?] es una tripla (Ω, F, P) , donde

- i) Ω es el conjunto de todos los posibles resultados de un experimento aleatorio, también llamado espacio muestral.

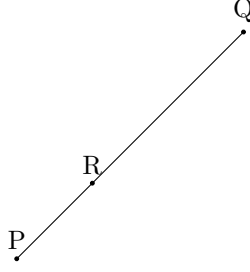


Figure 1: Recta \overline{PR}

- ii) F es una familia de subconjuntos de Ω que tiene una estructura de un campo σ
 - a) $\emptyset \in F$
 - b) Si $A \in F$ entonces, $A^c \in F$
 - c) Si $A_1, A_2, \dots \in F \implies \cup_{i=1}^{\infty} A_i \in F$
- iii) P es una función que asocia un número $P(A)$ a cada conjunto $A \in F$ del siguiente modo
 - a) $0 \leq P(A) \leq 1$
 - b) $P(\Omega) = 1$
 - c) Para una sucesión A_1, A_2, \dots de conjuntos disjuntos de dos en dos,

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} (P(A_i))$$

A los elementos de F se les llama eventos y al mapeo P se le llama medida de probabilidad.

En la probabilidad clásica, la probabilidad de un evento se define como la razón de los eventos favorables de dicho evento y el total de casos posibles. Es decir, sea E nuestro espacio muestral

$$P(A) = \frac{|A|}{|E|}$$

notando que existe el supuesto que todos los eventos tienen el mismo peso, lo que hace que todos sean igualmente probables.

En la probabilidad geométrica se considera la probabilidad de que un punto o conjunto de puntos se encuentren en la región de intereés. Por ejemplo, si buscamos que un punto x se encuentre en el segmento de recta \overline{PR} en la figura 1 esta dado por

$$P(A) = \frac{\|\overline{PR}\|}{\|\overline{PQ}\|}$$

La probabilidad frecuentista se basa en que la probabilidad de un evento esta determinada por la razón entre la frecuencia con la que ocurre el evento de intererés y el número de veces que se repitió el experimento. Dicho de otra forma

$$P(A) = \frac{A}{B}$$

donde A es el número de veces que se observa el evento de interés y B es el número de veces que se repite el experimento.

Los axiomas de probabilidad son las condiciones mnimas que deben verificarse para que una función que se define sobre un evento o conjunto de eventos sea una función de probabilidad.

Dichos axiomas son los siguientes

1. $P(A) \geq 0$
2. $P(E) = 1$
3. Si A_1, A_2, \dots son eventos mutuamente excluyentes (disjuntos de dos a dos)

$$P(A_1 \cup A_2 \cup \dots) = \sum P(A_i)$$

De manera mas formal, del espacio muestral Ω se toma una σ -álgebra (subconjuntos del espacio muestral) y al función de probabilidad P es una medida sobre dicha σ -álgebra cuya medida total sea 1.

La probabilidad subjetiva recae sobre aquellos eventos cuya ocurrencia esta sujeta a valores empíricos y subjetivos propios de aquel que intenta calcular la probabilidad de dicho evento.

La probabilidad lógica combina la teoría de la probabilidad para que, en conjunto con la lógica deductiva se pueda explotar dicho evento. La probabilidad lógica se basa en los siguientes axiomas:

1. Dados p y q hay un sólo valor de q/p
2. Los posibles valores de q/p son todos los nmeros reales en el intervalo $(0, 1)$
3. Si p implica q , entonces $q/p = 1$.
4. Si p no implica q , entonces $q/p = 0$.
5. La probabilidad de q y r dado p es la probabilidad de q dado p multiplicada por la probabilidad de r dado p (axioma conjuntivo).
6. La probabilidad de q o r dado p es la probabilidad de q dado p ms la probabilidad de r dado p menos la probabilidad de q y r dado p (axioma disyuntivo).

Un proceso estocástico con espacio de estados S es una colección de variables aleatorias $X_t, t \in T$ definido en el mismo espacio de probabilidad (Ω, F, P) . El conjunto T se le conoce como conjunto de estados, el cual puede ser discreto si $T = \mathbb{N} = \{0, 1, 2, \dots\}$ o bien si T es incontable, se dice que el proceso tiene parámetros continuos. Notemos que decir X_t hace referencia al estado o posición del proceso al tiempo t .

Para toda $\omega \in \Omega$ fija el mapeo

$$t \rightarrow X_t(\omega)$$

definido en el conjunto de parámetro T es llamado una trayectoria, realización o camino muestra.

De manera general, un proceso estocástico es todo aquel cuya probabilidad ya depende del tiempo (o en este caso, estados) por el que transita. Como ejemplos tenemos la caminata aleatoria, cadenas de Markov, procesos de Poisson, etc.

Es un valor de la variable estadística X que representa el número de veces que se observa dicha variable en una muestra N . Si se descompone al experimento en A_1, A_2, \dots con frecuencia absoluta X_1, X_2, \dots respectivamente, se tiene que la suma de todas las frecuencias absolutas debe ser igual al tamaño de la muestra N , es decir

$$\sum_{i=1}^n X_i = N$$

La frecuencia relativa esta dada por la razón entre la frecuencia absoluta del i -ésimo evento observado y el tamaño de la muestra N , es decir

$$f_i = \frac{X_i}{N} = \frac{X_i}{\sum_i X_i}$$

Al ordenar por los valores de la muestra de menor a mayor, la frecuencia acumulada se refiere al total de la suma de frecuencias absolutas que se acumula hasta cierto valor x_i de la muestra. Además, si se divide entre el tamaño de la muestra se obtiene la frecuencia relativa acumulada.

- Riesgo: Contingencia o proximidad de un daño.
- Peligrosidad: Cualidad de peligroso, que tiene riesgo o puede ocasionar daño.
- Amenaza: Sust. Dar indicios de ir a sufrir algo malo o dañino.
- Vulnerabilidad: Cualidad de vulnerable, que puede ser herido o recibir lesión, física o moralmente.

1.5 Principio del Palomar

Si m palomas ocupan n nidos y $m > n$, entonces al menos un nido tiene dos o más palomas descansando en él [5].

1.6 Expansiones del binomio de Newton

Para $n = 2$ se tiene que

$$(x + y)^2 = x^2 + 2xy + y^2$$

$n = 3$

$$(x + y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$$

$$n = 4$$

$$(x + y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4$$

$$n = 5$$

$$(x + y)^5 = x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5$$

$$n = 6$$

$$(x + y)^6 = x^6 + 6x^5y + 15x^4y^2 + 20x^3y^3 + 15x^2y^4 + 6xy^5 + y^6$$

$$n = 7$$

$$(x + y)^7 = x^7 + 7x^6y + 21x^5y^2 + 35x^4y^3 + 35x^3y^4 + 21x^2y^5 + 7xy^6 + y^7$$

$$n = 8$$

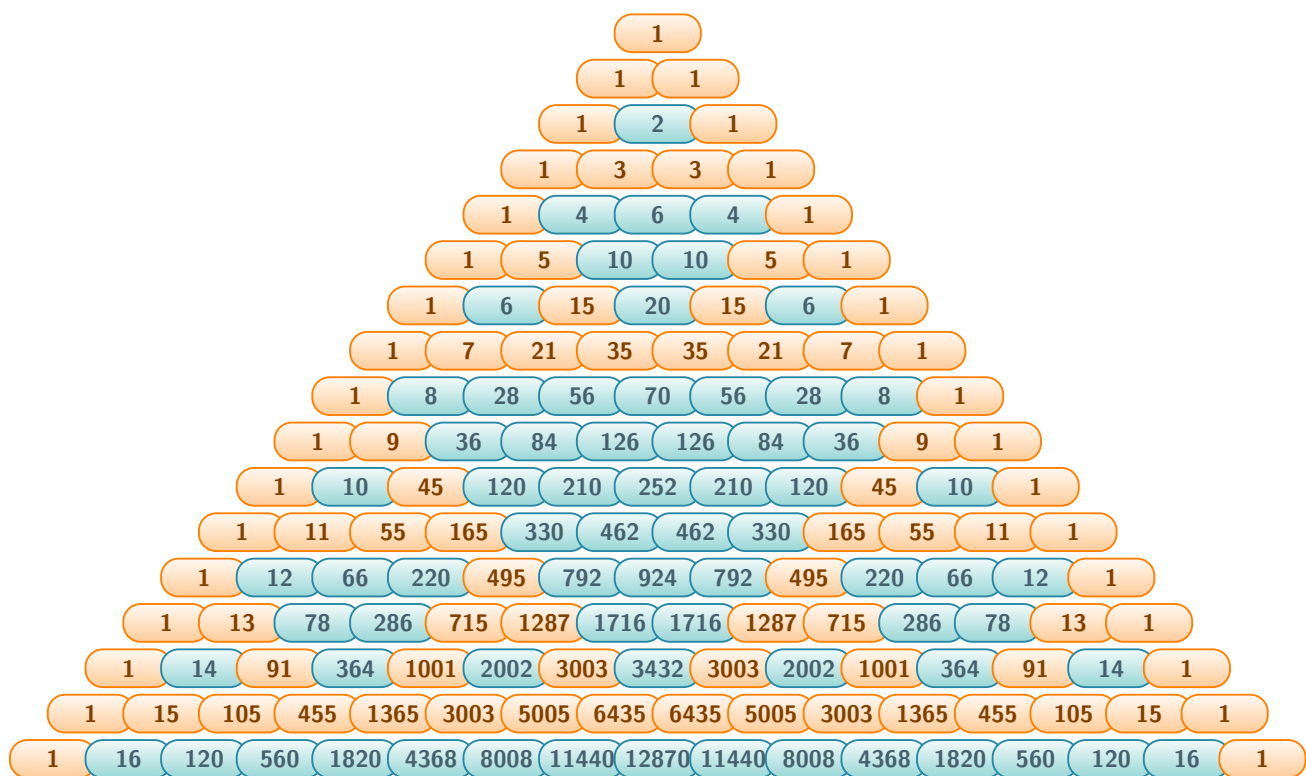
$$(x + y)^8 = x^8 + 8x^7y + 28x^6y^2 + 56x^5y^3 + 70x^4y^4 + 56x^3y^5 + 28x^2y^6 + 8xy^7 + y^8$$

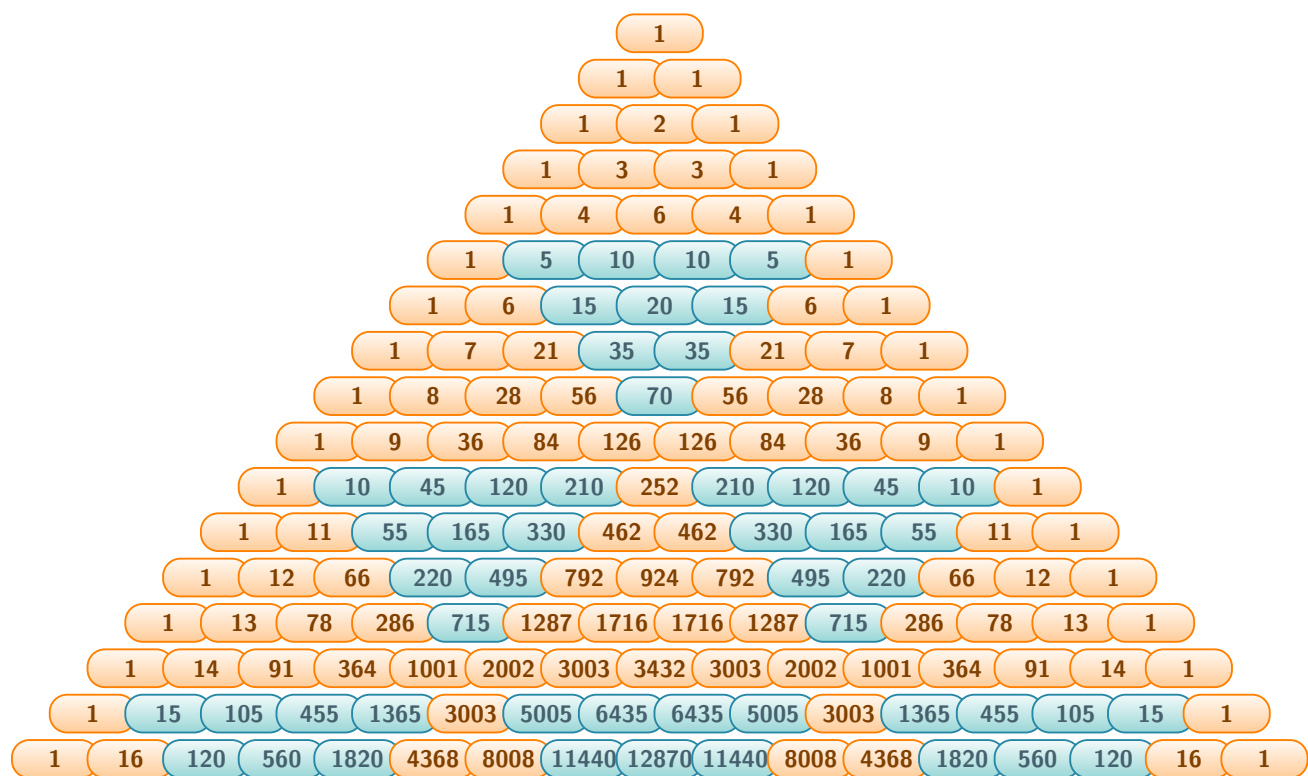
$$n = 9$$

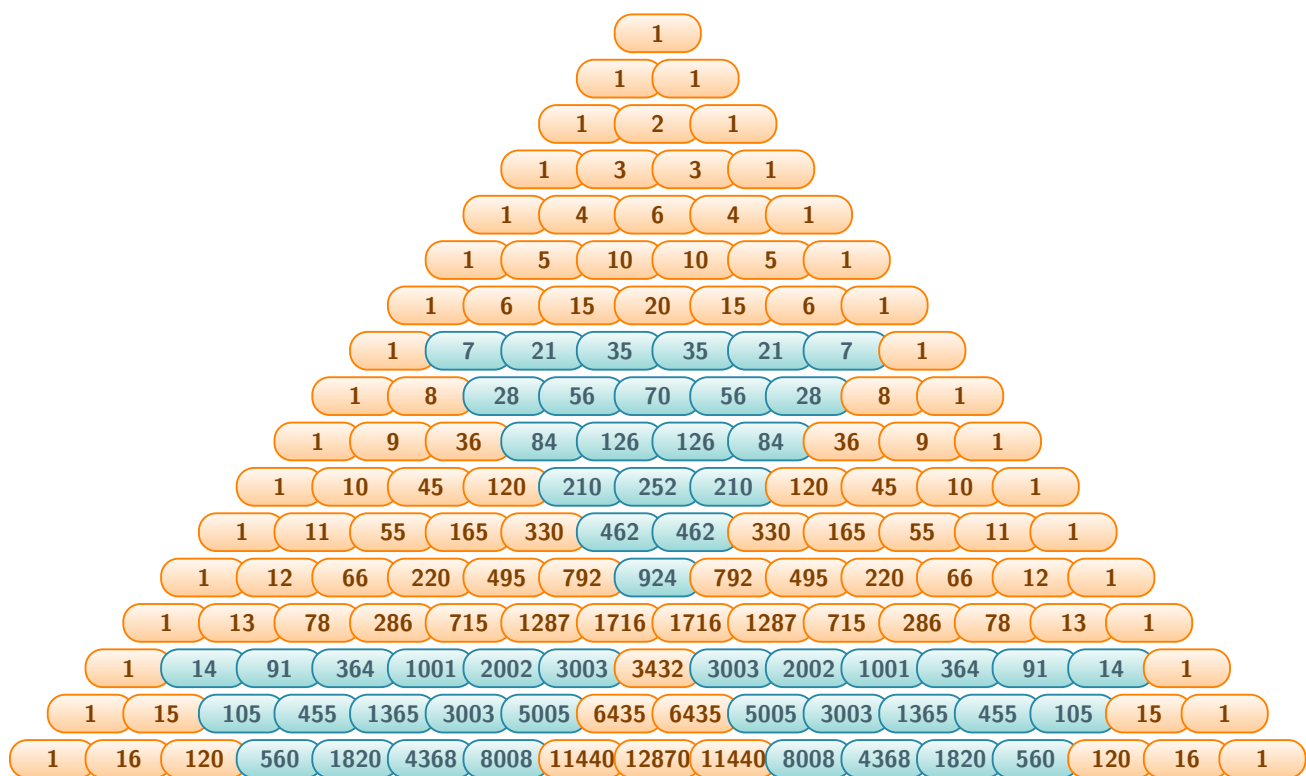
$$(x + y)^9 = x^9 + 9x^8y + 36x^7y^2 + 84x^6y^3 + 126x^5y^4 + 126x^4y^5 + 84x^3y^6 + 36x^2y^7 + 9xy^8 + y^9$$

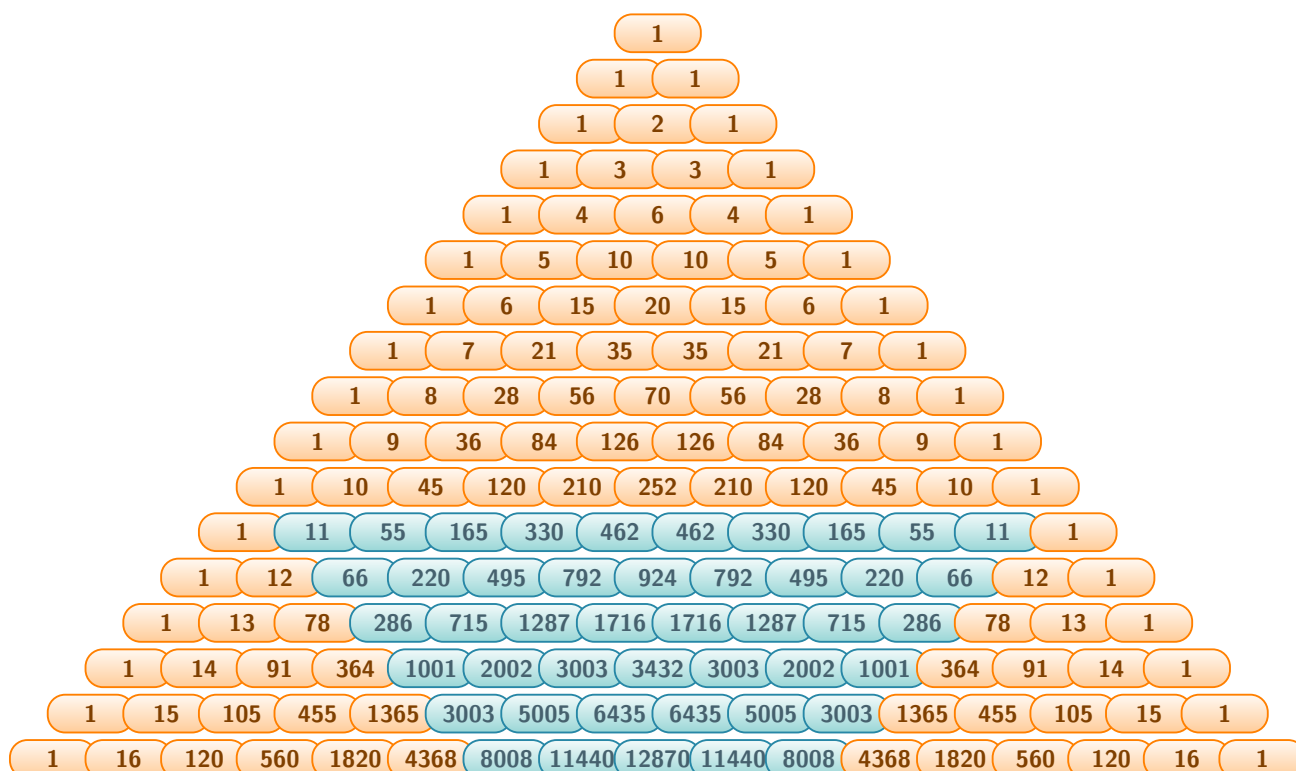
$$n = 10$$

$$(x+y)^{10} = x^{10} + 10x^9y + 45x^8y^2 + 120x^7y^3 + 210x^6y^4 + 252x^5y^5 + 210x^4y^6 + 120x^3y^7 + 45x^2y^8 + 10xy^9 + y^{10}$$









2 Tarea 2

2.1 Ejemplos de permutación

1. Cuántos números de 5 cifras diferentes se puede formar con los dígitos: 1, 2, 3, 4, 5?

Respuesta: Como son diferentes, en este caso solo se pueden armar en tal $5! = 120$ números distintos

2. De cuántas formas pueden sentarse 6 personas en una fila de sillas?

Respuesta: Como importa el orden al momento de sentarse, el resultado es $6! = 720$

3. De la palabra *payasos*, encuentre todas las posibles palabras que se pueden formar utilizando todas las letras.

Respuesta: En total se pueden hacer $7! = 5040$ palabras distintas (claro, algunas sin sentido alguno).

4. De la palabra *payasos*, encuentre las que empiecen con una vocal.

Respuesta: En la palabra *payasos* hay 3 vocales, por lo que el total de palabras está dado por $3 \cdot 6!$.

5. Cuántos números de 5 cifras impares se pueden formar en total?

Respuesta: Existen 5 dígitos impares en total, por lo que el número de cifras posibles que se pueden formar es $5^5 = 3125$.

6. De las cifras anteriores, encuentre cuantas son mayores a 70000.

Respuesta: En este caso, solo existen dos casos posibles: si el número empieza con 7 o 9. De este modo, la cantidad de números posibles es $2 \cdot 5^4 = 1250$.

7. De entre 8 hombres y 5 mujeres se debe escoger un comité de 5 personas que reciba al Papa. Las reglas para acomodar a las personas es alternar un hombre y una mujer. De cuantas formas se pueden acomodar el comité si siempre debe empezar en un hombre.

Respuesta: Para este problema, debemos alternarlos 8 hombres y las 5 mujeres:

$$8 \cdot 5 \cdot 7 \cdot 4 \cdot 6 = \frac{8!}{(8-3)!} \cdot \frac{5!}{(5-2)!} = 1680$$

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

Respuesta: Se pueden sentar de $(8-1)! = 7! = 5040$ formas distintas.

9. Calcular el número de cadenas de 5 letras que se pueden hacer con el alfabeto (27 letras) que empiecen con una vocal y terminen con la letra k .

Respuesta: Para este caso, existen 5 vocales por lo que el número de cadenas alfabéticas que se pueden hacer es dado por

$$5 \cdot 27^3 \cdot 1 = 98415$$

10. Calcular cuantos enteros positivos n se pueden formar con los dígitos 3, 4, 4, 5, 5, 6 y 7 si queremos que n sea mayor 5,000,000.

Respuesta: Estamos restringidos a que el número empiece con 5, 6 o 7: 4 posibles casos y los demás se pueden distribuir, es decir

$$4 \cdot 6! = 4 \cdot 720 = 2880$$

2.2 Ejemplos de combinación

1. En una clase de 26 alumnos se quiere elegir un comité formado por tres alumnos. Cuántos comités diferentes se pueden formar?

Respuesta: El número de comités que se pueden armar es

$$\binom{26}{3} = \frac{26!}{23!3!} = 2600$$

2. Calcular el número de formas en que un ejecutivo puede elegir a 4 de 11 empleados para un ascenso.

Respuesta: La respuesta esta dada por:

$$\binom{11}{4} = \frac{11!}{7!4!} = 330$$

3. Calcular cuantas diagonales tiene un pentágono.

Respuesta: Para resolver este problemas, podemos considerar que cada vértice se va a unir a otro, es decir de cinco vertices vamos a elegir dos, pero evitando las cinco aristas que unen a éstos vértices. Es decir:

$$5c2 - 5 = \binom{5}{2} - 5 = 10 - 5 = 5$$

4. Calcular cuantos triángulos se puede formar con los vértices de un pentágono.

Respuesta: De manera similar al anterior, un triángulo se forma cuando se unen tres vértices, es decir

$$5c3 = \binom{5}{3} = 10$$

triángulos distintos.

Un grupo formado por cinco hombres y siete mujeres, forma un comité de 4 hombres y 3 mujeres. De cuntas formas puede formarse, si:

5. Puede pertenecer cualquier hombre o mujer.

Respuesta: $\binom{5}{4} \cdot \binom{7}{3} = 5 \cdot 35 = 175$

6. Un hombre en específico debe pertenecer al comité

Respuesta: Como restringimos a que un hombre en particular forme parte del comité, se tiene que

$$1 \cdot \binom{4}{3} \cdot \binom{7}{3} = 4 \cdot 35 = 140$$

7. Tres mujeres no pueden formar parte del comité

Respuesta: En este caso, los casos posibles para las mujeres se reducen de siete a cuatro, por lo que

$$\binom{5}{3} \cdot \binom{4}{3} = 10 \cdot 4 = 40$$

8. Un inversionista desea vender siete opciones de inversión de su portafolio: cuatro puts y tres bonos. Su portafolio consta de 13 bonos y 12 puts.

Respuesta: Se deben elegir cuatro de 13 puts y tres de 12 bonos, por lo que el resultado es

$$\binom{13}{4} \cdot \binom{12}{3} = 715 \cdot 220 = 157300$$

9. En una caja de 12 focos hay dos focos defectuosos. Calcular el número de formas en que se pueden seleccionar tres focos incluyendo los dos defectuosos.

Respuesta: De los tres a elegir solo podemos elegir uno de los 10 buenos y obligatoriamente a los dos defectuosos, es decir

$$\binom{10}{1} \cdot \binom{2}{2} = 10 \cdot 1 = 10$$

10. El número de disposiciones de la palabra TALLAHASSEE es igual a

$$\frac{11!}{3!2!2!1!1!} = 831600$$

Si queremos calcular cuantas de ellas no tienen la letra A de manera adyacente, podemos considerar las disposiciones que no contienen la letra A:

$$\frac{8!}{1!2!2!1!1!} = 5040$$

formas de arreglar las letras restantes. Entre esas 8 letras existen 9 espacios en los que puede introducirse la letra A, es decir existen $\binom{9}{1} = 9$ formas para elegir esos lugares. En total, existen $5040 \cdot 9 = 45360$ formas de acomodar las letras de TALLAHASSEE sin que haya dos letras A adyacentes.

2.3 El problema de los caminos posibles en un rectángulo

El problema se define como sigue: cuantos caminos hay de A a B que tengan longitud mínima (se puede apreciar un ejemplo en la siguiente figura).

Del nodo de origen en $(0, 0)$ se puede ir en dos direcciones (arriba y a la izquierda) y en cada una se puede ir de dos maneras distintas. Al llegar hasta el punto B se requieren todos los caminos posibles a los puntos B_1 y B_2 , por lo que se puede ver la semejanza con el triángulo de Pascal (en el caso de un cuadrado, si lo giramos 45 resulta el triángulo de Pascal).

Un camino mínimo se puede realizar haciendo $n + k - 1$ movimientos: $k - 1$ son hacia arriba y n hacia los lados en dirección a B, por lo que el número de soluciones [7] esta dado por

$$\binom{n + k - 1}{k - 1}$$

De manera general, se tiene la formula para el desarrollo del trinomio para el término $a^{n-m}b^{m-k}c^k$ [6]

$$\sum_{m=0}^n \sum_{k=0}^m \binom{n}{m} \binom{m}{k} a^{n-m} b^{m-k} c^k$$

donde

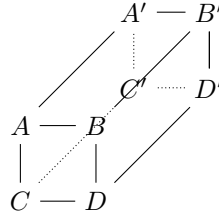
$$\binom{n}{m} \binom{m}{k} = \frac{n!}{(n-m)!(m-k)!k!} \quad (1)$$

2.5 El problema de los caminos posibles en un cubo

De manera análoga al problema del camino en el rectángulo, en el cubo para ir de la arista C a la arista B' se parece a la pirámide de Pascal solo que el vértice al que queremos llegar es la suma de todos los caminos posibles anteriores. En la imagen de abajo se encuentra el caso más elemental cuando $n = 3$.

Éste resultado es el coeficiente del termino ABC del desarrollo del trinomio. De (2) con $m = n - 1$ y $k = n - 2$ se tiene que el número de caminos posibles de mínima distancia es

$$\binom{n}{n-1} \binom{n-1}{n-2} = \frac{n!}{(n-2)!}$$



2.6 Reseñas

2.6.1 Primates Count

Entre los expertos en conducta animal existe el debate acerca de como se puede concebir el sentido de conteo por los animales y si algunos de ellos poseen algún sentido de los números. H. Kalmus escribe que algunos animales como las ratas, pericos o ardillas pueden ser entrenados para contar; por otro lado, algunos animales son capaces de distinguir números a través patrones visuales o seales acústicas.

2.6.2 Cigada-Generated Prime Bumbers

Las cigadas son insectos voladores que pasan la mayor parte de su vida bajo tierra con un periodo de apareamiento para morir en poco tiempo. Lo sorprendente es que ésta salida esta sincronizada

en periodos de aos múltiplos de 13 y 17. Se ha especulado que la razón de este comportamiento es para evitar depredadores cuyos ciclos de vida son cortos ya que si fuera, por ejemplo, múltiplo de 12 las criaturas con ciclos de vida de 2, 3 o 6 aos acabarían mas fácilmente con la población de insectos.

2.6.3 Ant Odometer

Las hormigas son insectos con un comportamiento social muy estudiado. La hormiga del Sahara, *Cataglyphis fortis* viaja distancias enormes en terreno arenoso en busca de comida. Las observaciones más interesantes es que éstas hormigas son capaces de regresar a su hormiguero a través de comportamientos como juzgar la luz para orientarse y aparentemente con capaces de contar los pasos que realizan al viajar.

2.6.4 Quipu

Los Incas usaban bancos de memoria hechos cuerdas y nudos para almacenar números. A pesar de no tener un sistema de escritura ya poseían un sistema lógico-numérico en los quipus, un mapeo de números a objetos de la vida real, como planes de construcción, danzas, etc. Este hecho dispersa la noción de que una civilización las matemáticas florecen hasta que se desarrolla un sistema de escritura.

2.6.5 Magic Squares

Un cuadrado mágico consta de N^2 celdas llena de números enteros diferentes, donde la suma de las filas, columnas y diagonales principales es constante. Albrecht Drer creo un cuadrado mágico de tamaño 4×4 en 1514. En él, el valor constante es igual a 34 y las esquinas suman al igual que el cuadrado interno de tamaño 2×2 suman 34.

2.7 Demostraciones

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

Demostración

Por definición

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

sumando y restando n al termino de $k!$

$$\frac{n!}{(n-k)!(k+n-n)!} = \frac{n!}{(n-k)!(n-(n-k))!} = \binom{n}{n-k}$$

- $\binom{n}{0} = 1$

Demostración

Por definición

$$\binom{n}{0} = \frac{n!}{n!0!} = \frac{n!}{n!1} = 1$$

3 Tarea 3

3.1 Telómeros

Los telómeros son estructuras especializadas situadas en los extremos de los cromosomas, que los protegen de posibles fusiones y de su degradación, con lo que se garantiza la estabilidad de los cromosomas y viabilidad de las células. En ausencia de un mecanismo compensatorio de elongación los telómeros se vuelven cada vez mas cortos en cada procesos de división celular[9].

La telomerasa es una enzima que activa la elongación de los telómeros durante la división celular, sin embargo, ésta es suprimida después del nacimiento en la mayoría de las células somáticas. Éste comportamiento provoca que el individuo, con el paso del tiempo, no pueda regenerar tejidos y lleva al envejecimiento del organismo (enfermedades relacionadas con la edad)

Los estudios hechos se basan en la longitud media del telómero, medida usualmente en células sanguíneas, lo que llevo a observar una relación entre mayor riesgo de padecer enfermedades por edad, tiempo de vida e inclusive el riesgo de padecer diferentes tipos de cáncer[9].

3.2 Circuitos lógicos

Un circuito lógico es un dispositivo que tiene una o más entradas y una sola salida, donde cada entrada solo puede tener el valor de 0 o 1; los datos son procesados dentro del circuito y dan un valor de salida con valor de 0 o 1.

Los circuitos lógicos son conformados por circuitos elementales denominados compuertas lógicas:

- Compuertas básicas: OR, AND, NOT
- Compuertas derivadas: NOR, NAND

La compuerta OR con dos entradas X y Y tiene como salida

$$A = X + Y$$

Los resultados se resumen en la siguiente tabla:

La compuerta AND con entradas X y Y tiene como salida

$$A = X \cdot Y$$

donde el producto se define como

X	Y	$X + Y$
0	0	0
0	1	1
1	0	1
1	1	1

Table 1: Salidas para $X + Y$

X	Y	$X \cdot Y$
0	0	0
0	1	0
1	0	0
1	1	1

Table 2: Salidas para $X \cdot Y$

La compuerta NOT con entrada X tiene como salida

$$X = \overline{X}$$

Las compuertas NOR y NAND son compuertas derivadas compuestas por dos compuertas básicas. La compuerta NOR es una compuerta OR seguida de una compuerta NOT. Una compuerta NAND equivale a una compuerta AND seguida de una compuerta NOT. Si las entradas son X y Y , las salidas resultantes son

$$A = \overline{X + Y}$$

$$A = \overline{X \cdot Y}$$

3.3 Permisos en archivos y carpetas en Unix

El comando CHMOD permite colocar permisos de acceso para los archivos. En general son tres tipos de personas que potencialmente pueden acceder a los archivos: dueño (OWNER), grupo (GROUP) y cualquier otra persona (ANYBODY).

Al usar CHMOD tienes 3 grupos de números, lo cual representa igualmente 3 grupos de usuarios:

1. Owner o User, el cual de manera general
2. Group, un grupo de usuarios configurado
3. World o Anyone, que es básicamente cualquier usuario

Los permisos disponibles son:

- 0 = sin permisos
- 1 = permiso de ejecutar únicamente

X	\overline{X}
0	1
1	0

Table 3: Salidas para \overline{X}

X	Y	$\overline{X + Y}$	$\overline{X \cdot Y}$
0	0	1	1
0	1	0	1
1	0	0	1
1	1	0	0

Table 4: Salidas para $\overline{X \cdot Y}$ y $\overline{X + Y}$

- 2 = permiso de escribir solamente
- 3 = escribir y ejecutar
- 4 = permiso de leer
- 5 = leer y ejecutar
- 6 = leer y escribir
- 7 = leer, escribir y ejecutar

Por ejemplo, al usar CHMOD 650:

- El primer número, 6, da al OWNER los permisos de lectura y escritura sobre el archivo.
- El segundo número, 5, le da al GROUP los permisos de lectura y ejecución.
- El tercer número, 0, le da a todos los demás ningún permiso sobre el archivo.

La secuencia -rwxr-r- indica los permisos otorgados en el archivo. El primer - indica que es un archivo. Las siguientes tres letras, rwx muestran que el dueño del archivo (owner) puede leer, escribir, y ejecutar. Los siguientes tres, r-, muestran que los permisos del grupo son de lectura solamente, y los siguientes tres elementos, r- muestran que los permisos para todo mundo son slo de lectura igualmente.

4 Tarea 4

4.1 Paradojas y antinomias

Una paradoja es una figura retórica que consiste en el uso de expresiones que envuelven una contradicción. Las paradojas se pueden clasificar en dos grupos: en función de su veracidad y por el área de conocimiento que engloban.

Una antinomia (del *anti* y *nomos*) significa contradicción entre dos tesis que se excluyen mutuamente, pudiendo cada una de ellas ser igualmente demostrada de una manera convincente por vía lógica. Esta noción de antinomia desempeña un papel importante en el sistema filosófico de Kant, que enumera las cuatro antinomias siguientes:

1. • *Tesis*: El mundo tiene un principio en el tiempo y el espacio
 • *Antítesis*: El mundo es infinito en el tiempo y el espacio
2. • *Tesis*: Todo en el mundo se compone de lo simple
 • *Antítesis*: No hay nada simple, todo es complejo
3. • *Tesis*: En el mundo existen causas libres.
 • *Antítesis*: No existe ninguna libertad, todo es naturaleza (o sea, necesidad).
4. • *Tesis*: En la serie de las causas universales hay algún ser necesario
 • *Antítesis*: En esta serie no hay nada necesario, todo es casual

A su juicio, la tesis y la antítesis son igualmente demostrables. Es decir, en una antinomia las dos proposiciones se reconocen igualmente verdaderas pero se excluyen por ser contradictorias la una con la otra.

En conclusión, la antinomia se compone de dos argumentos en teoría válidos, plausibles y que se excluyen el uno al otro. Una paradoja no es sólo una contradicción, es una par de contradicciones a las que se llega mediante razonamientos lógicos impecables.

4.2 Paradojas en las Matemáticas

4.2.1 Paradoja de Newcomb

La paradoja de Newcomb es un estudio de un juego entre dos jugadores, de los cuales uno es capaz de predecir el futuro. El caso se formula de la siguiente manera:

En este juego hay dos participantes: un oráculo capaz de predecir el futuro y un jugador normal. Al jugador se le presentan dos cajas: una abierta que contiene \$1,000 y otra cerrada que puede contener \$1.000.000 o bien \$0. El jugador debe decidir si prefiere recibir el contenido de ambas cajas o sólo el de la caja cerrada.

La complicación consiste en que anteriormente, el oráculo ha predicho lo que va a escoger el jugador. Si vaticina que el jugador se llevará sólo la caja cerrada, pondrá \$1,000,000 dentro de esa caja. Si vaticina que el jugador se llevará las dos cajas, dejará vacía la caja cerrada. El jugador conoce el mecanismo del juego pero no la predicción, que ya ha sido realizada.

La pregunta es que caja debe elegir el jugador. Los pagos del juego se resumen en la siguiente tabla:

Si el oráculo acierta el 100% de las veces, si el jugador se lleva la caja cerrada, obtendrá \$1.000.000. Si el jugador se lleva ambas cajas, la caja cerrada estará vacía, por lo que sólo se llevará \$1.000. Según este razonamiento, el jugador debe escoger siempre la caja cerrada.

	El oráculo predice la caja cerrada	El oráculo predice ambas cajas
El jugador escoge la caja cerrada	\$1,000,000	\$0
El jugador escoge ambas cajas	\$1,001,000	\$1,000

Table 5: Tabla de pagos

Pero en el momento en el que el jugador se acerca a las cajas para hacer su elección, su contenido ya está definido. La caja cerrada o tiene algo o no lo tiene, pero es demasiado tarde para cambiar su contenido. El jugador debe llevarse el contenido de ambas cajas, ya que tenga lo que tenga la caja cerrada obtendrá \$1.000 ms, porque de todos modos se llevará la abierta. Según este razonamiento, el jugador debe escoger siempre llevarse las dos cajas.

Como se puede observar, hay dos conceptos de teoría de decisión en conflicto. El principio de utilidad esperada (basado en la probabilidad de cada desenlace) opta por tomar sólo la caja vacía. El principio de dominancia establece que, si una estrategia es siempre mejor, no importa cuál es la situación, hay que escogerla sobre las demás.

El problema consiste en que la gente se divide casi a la mitad sobre cuál es la solución al problema, con un gran porcentaje que cree que la otra mitad está equivocada.

La paradoja de Newcomb se considera una paradoja porque lleva a una autocontradicción. La causalidad inversa está definida en el problema, por lo que no puede haber libre albedrío. Al mismo tiempo, el libre albedrío está definido en el problema, de otro modo, el jugador no estaría realizando una verdadera elección.

4.2.2 Paradoja de Bertrand

La paradoja de Russell o paradoja del barbero demuestra que la teoría de conjuntos por Cantor y Frege es contradictoria.

La paradoja consiste en considerar si el conjunto de los conjuntos que no forman parte de sí mismos forma parte de sí mismo. Si no forma parte de sí mismo, pertenece al tipo de conjuntos que no forman parte de sí mismos y por lo tanto forma parte de sí mismo. Es decir, formará parte de sí mismo sólo si no forma parte de sí mismo.

De manera formal, se expresa como

$$M = \{x : x \notin x\}$$

o bien,

$$\forall x, x \in M \Leftrightarrow x \notin x$$

como x es un conjunto, se puede sustituir a x por M

$$M \in M \Leftrightarrow M \notin M$$

lo cual lleva a una contradicción.

4.2.3 La paradoja de Monty Hall

También llamada la paradoja de las tres puertas, se puede adaptar al entorno nacional de la siguiente manera:

Supongamos que tenemos la suerte de estar en Familia con Chabelo en la sección donde podemos elegir de entre tres posibles puertas. Detrás de una hay un coche y en las dos restantes hay dos cabras. Chabelo conoce que hay detrás de cada puerta.

Ya en mente nuestra elección, Chabelo nos la pide y a continuación procede a abrir una de las dos puertas restantes que tiene una cabra (ya que el sabe donde está cada premio). A continuación, se le da la opción al concursante de cambiar la puerta. ¿Debe o no cambiarse de puerta?

Como la respuesta correcta parece contradecir conceptos básicos de probabilidad, se puede considerar como una paradoja. La respuesta se basa en suposiciones que no son obvias y que no se encuentran expresadas en el planteamiento del problema, por lo que también se puede considerar como una pregunta con trampa.

La probabilidad de que el jugador escoja la puerta que lleva al carro es igual a $1/3$ y su complemento, $2/3$ es la probabilidad de que la puerta lleve a una cabra. Un error es suponer que, cuando el presentador abre una puerta se sigue que ambas tienen la misma probabilidad de contener el carro, ya que se abre después de se anuncia la puerta que el jugador eligió.

Si el jugador escoge en su primera opción la puerta que contiene el coche (con una probabilidad de $1/3$), entonces el presentador puede abrir cualquiera de las otras dos puertas. Por otro lado jugador pierde el coche si cambia cuando se le ofrece la oportunidad.

Si el jugador escoge una cabra en su primera oportunidad (con una probabilidad de $2/3$), el presentador sólo tiene la opción de abrir una puerta, y esta es la única puerta restante que contiene una cabra. En ese caso, la puerta restante tiene que contener el coche, por lo que cambiando lo gana.

Se concluye que, si el jugador mantiene su elección original gana si escogió originalmente el coche (con probabilidad de $1/3$), mientras que si cambia, gana si escogió originalmente una de las dos cabras (con probabilidad de $2/3$). Por lo tanto, el concursante debe cambiar su elección si quiere maximizar la probabilidad de ganar el coche

4.3 Demostraciones

1. Primera ley de De Morgan

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

Demostración

Sea x un elemento arbitrario del conjunto universo U

$$x \in (A \cup B)' \Leftrightarrow x \notin (A \cup B) \Leftrightarrow x \notin A \wedge x \notin B \Leftrightarrow x \in A' \cap B'$$

2. Segunda ley de De Morgan

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

Demostración

Sea x un elemento arbitrario del conjunto universo U

$$x \in (A \cap B)' \Leftrightarrow x \notin (A \cap B) \Leftrightarrow x \notin A \vee x \notin B \Leftrightarrow x \in A' \cup B'$$

3. Si $A \subset B$ y $B \subset A$, entonces $A = B$.

Demostración.

Por definición, si $A \subset B$ se sigue que si $x \in A$ implica $x \in B$. De igual modo, si $B \subset A$ se sigue que si $x \in B$ implica $x \in A$.

Esto significa que, $x \in B \Leftrightarrow x \in A$ lo cual se cumple sólo cuando $A = B$.

4. Ley aditiva de probabilidad

La probabilidad de unión de dos conjuntos se define como

$$P(S_1 \cup S_2) = P(S_1) + P(S_2) - P(S_1 \cap S_2)$$

Para el caso donde se tienen tres conjuntos:

$$P(S_1 \cup S_2 \cup S_3) = P(S_1 \cup (S_2 \cup S_3))$$

aplicando la definición de suma considerando como un solo elemento lo que engloba el paréntesis

$$P(S_1 \cup (S_2 \cup S_3)) = P(S_1) + P(S_2 \cup S_3) - P(S_1 \cap (S_2 \cup S_3))$$

sustituyendo la expresión por la definición de suma en dos conjuntos y aplicando distributividad

$$P(S_1) + P(S_2 \cup S_3) - P(S_1 \cap (S_2 \cup S_3)) = P(S_1) + P(S_2) + P(S_3) - P(S_2 \cap S_3) - P((S_1 \cap S_2) \cup (S_1 \cap S_3))$$

volviendo a aplicar la definición de probabilidad de una unión se llega a que

$$P(S_1 \cup S_2 \cup S_3) = P(S_1) + P(S_2) + P(S_3) - P(S_2 \cap S_3) - P(S_1 \cap S_2) - P(S_1 \cap S_3) - P(S_1 \cap S_2 \cap S_3)$$

Como se puede observar, a la probabilidad de cada conjunto se le resta la probabilidad de cada intersección posible entre cada uno de ellos.

5 Tarea 5

5.1 Paradoja del Falso Positivo

Un paciente empieza a mostrar signos de una enfermedad extremadamente rara y mortal: solo un 0.1% de la población le afecta. El paciente busca un diagnóstico fidedigno por lo que visita a un doctor para que se lo realice. El resultado es desgarrador: da positivo.

El doctor, de manera seria le comenta al paciente que la prueba es 99% segura, por lo que lo más seguro es que muera. De algún modo, la prueba que es 99% segura terminó con un falso positiva, cuya probabilidad de ocurrencia es muy baja.

La pregunta que surge es si realmente esta probabilidad es en realidad *baja* y si lo es, que tan *bajo* es bajo. Antes de continuar, hagamos unos cuantos cálculos.

Supongamos que nuestra población es de 100,000 personas, el porcentaje de incidencia de la enfermedad es de 0.1% y que la precisión de la prueba es del 99%.

# Personas	Enfermas	Sanas	Total
Test positivo	99	999	1,098
Test negativo	1	98,901	98,902
Total	100	99,900	100,000

Table 6: Cálculos en la población

¿Qué se quiere mostrar con ésta tabla? Con estos datos, tenemos cerca de $0.1\% \cdot 100,000 = 100$ personas que en realidad están enfermas. Aplicando la prueba en ellas, se obtienen 99 positivos verdaderos y 1 falso positivo.

Por otro lado, se tienen 99,000 personas que no están enfermas; al aplicarles la prueba resultan 98,901 resultados negativos mientras que 999 de éstas personas sanas obtienen un resultado positivo.

En total, se obtienen 1,098 resultados positivos (los falsos y los verdaderos) pero solamente 9% de ellos resultan verdaderos positivos. En conclusión, el análisis lleva a que practicar una prueba así arroja un 91% de certeza de no estar enfermo.

De manera muy específica, cuando una condición (en éste caso, la enfermedad mortal) con bajo índice que ocurrencia en una población es muy posible obtener falsos positivos al realizar pruebas para detectar dicha condición a pesar de que *tan* precisa sea dicha prueba. Es erróneo pensar que el grado de precisión del resultado individual es igual a la precisión de la prueba.

Usando probabilidad condicional puede resultar un poco más sencillo de analizar. $P(enfermo) = 1\% = 0.01$ y $P(sano) = 99\% = .99$.

El considerar un falso positivo es hablar de un resultado positivo en la prueba cuando en realidad se esta sano, es decir

$$P(positivo|sano) = 1\%$$

de manera análoga, el obtener un resultado negativo dado que no se esta enfermo

$$P(negativo|sano) = 99\%$$

Suponiendo que al aplicar la prueba a una persona enferma, se tiene

$$P(negativo|enfermo) = 1\%$$

y

$$P(positivo|enfermo) = 99\%$$

En base a esto, se pueden calcular las probabilidades de los siguientes eventos. La fracción de personas sanas *y* que dan negativo

$$P(sano \cap negativo) = P(sano) \cdot P(negativo|sano) = 0.99 \cdot 0.99 = 0.9801$$

La fracción de individuos en el grupo que están enfermos *y* dan positivo:

$$P(enfermo \cap positivo) = P(enfermo) \cdot P(positivo|enfermo) = 0.01 \cdot 0.99 = 0.0099$$

La fracción de individuos en el grupo que dan falso positivo:

$$P(sano \cap positivo) = P(sano) \cdot P(positivo|sano) = 0.99 \cdot 0.01 = 0.0099$$

La fracción de individuos en el grupo que dan falso negativo:

$$P(enfermo \cap negativo) = P(enfermo) \cdot P(negativo|enfermo) = 0.01 \cdot 0.01 = .0001$$

Los que dan positivo en toda la población

$$P(positivo) = P(sano \cap positivo) + P(enfermo \cap positivo) = 0.0099 + 0.0099 = 0.0198$$

Por lo que, la probabilidad de que una persona tenga realmente la enfermedad dado que la prueba resultó positiva es

$$P(enfermo|positivo) = \frac{P(enfermo \cap positivo)}{P(positivo)} = \frac{.0099}{0.0198} = 0.5$$

Concluyendo que es totalmente distinto el considerar $P(enfermo|positivo)$ y $P(positivo|enfermo)$, además, la mitad de los casos resultan en falsos positivos ($P(sano|positivo)$).

5.2 Falacia del Apostador

También conocida como falacia de Monte Carlo o falacia de la madurez de las posibilidades, es un concepto erróneo en el que la gente cree que los eventos pasados afectan a los futuros en relación a actividades aleatorias, frecuentemente observada en juegos de azar. Los posibles pensamientos de ésta falacia son:

- 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 probabilidades de que algo suceda en la siguiente realización del experimento no está necesariamente relacionado con lo que ya sucedió.

El ejemplo más sencillo para analizar este caso es el de tiros sucesivos de una moneda. Suponiendo que la moneda es justa, la probabilidad de que salga cara o cruz es igual a 1/2 (excluyendo que

caiga de canto). De manera general, la probabilidad de que salgan n caras de manera consecutiva es

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

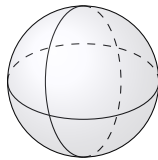
Si, al realizar el experimento resultan 5 tiros consecutivos en los que sale cara. Uno podría pensar que la probabilidad de que el siguiente tiro sea cara es igual a $1/64$. Este es al paso falaz del razonamiento ya que esta probabilidad se puede calcular *antes de* que se realice el experimento de lanzar las 6 caras consecutivas. Los resultados después de lanzar ya no son desconocidos para nosotros y por lo tanto, la probabilidad de que salga cara o cruz en el siguiente intento es exactamente la misma: $1/2$.

Razonar que es más probable que el próximo lanzamiento será cruz en vez de cara debido a los anteriores lanzamientos es la falacia: la idea de que una racha de suerte pasada influye de alguna forma en las posibilidades futuras.

Por otro lado, también existe la falacia del apostador inversa. Consiste en la conclusión en base a la ocurrencia de un resultado de un proceso aleatorio improbable, de que éste ha sido practicado muchas veces con anterioridad. Usando el clásico ejemplo de los dados, si un tiro resulta en un doble 6, es erróneo pensar que el dado ha sido tirado muchas veces antes de observar dicho resultado.

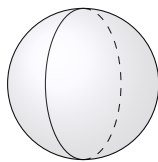
5.3 Paradoja de Borel - Kolmogorov

La paradoja consiste en lo siguiente: supongamos que una variable tiene distribución uniforme en una esfera unitaria.

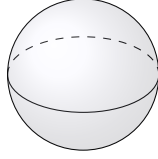


Para elegir un punto en la esfera (el contorno) se necesitan dos coordenadas: latitud y longitud.

La primera coordenada es la longitud λ , que se elige de manera uniforme entre $[-\pi, \pi]$, debido a que el radio tiene un valor de 1.



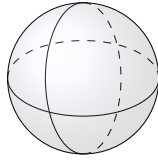
La segunda coordenada es la latitud ϕ entre $[-\frac{\pi}{2}, \frac{\pi}{2}]$ con densidad $\frac{1}{2} \cos \phi$.



¿Cuál es la distribución condicional de la variable en un gran círculo?

Desde el ecuador de la variable

$$f(\lambda|\phi = 0) = \frac{1}{2\pi}$$



Para una línea de longitud $\lambda = 0$

$$f(\phi|\lambda = 0) = \frac{1}{2} \cos \phi$$

Es decir, una es uniforme en el círculo y la otra no; ambas se refieren al mismo gran círculo en diferentes sistemas de coordenadas.

5.4 Problema 1.2

Se lanzan dos dados. Se define como A al suceso "la suma de los dos números es múltiplo de tres" y como B al suceso "sale al menos un seis". Encontrar las probabilidades de los sucesos $A \cup B$, $A \cap B$, $A^* \cup B$ y $(A \cup B)^*$.

Los posibles resultados para los dos tiros de dados son los siguientes

$$\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

En el caso del evento A , existen cuatro posibles resultados: 3, 6, 9, 12. Observemos que para obtener 3 existen dos formas, para el 6 existen 5 formas, para el 9 existen cuatro y para el 12 existe una forma de obtenerlo. Por lo que

$$P(A) = 2/36 + 5/36 + 4/36 + 1/36 = 12/36$$

Por otro lado, para el suceso B , existen 11 maneras de que al tirar salga un 6 en al menos uno de los dados

$$P(B) = 11/36$$

Notemos que existen tres casos que se repiten en ambos sucesos: $P(3, 6)$, $P(6, 3)$ y $P(6, 6)$. Así

$$P(A) \cap P(B) = 3/36$$

y finalmente

$$A \cup B = P(A) + P(B) - P(A \cap B) = 12/36 + 11/36 - 3/36 = 20/36$$

Por otro lado

$$P(B) = P(A \cap B) + P(A^* \cap B)$$

por lo que

$$P(A^* \cap B) = P(B) - P(A \cap B) = 11/36 - 3/36 = 8/36$$

Finalmente

$$P(A \cup B)^* = 1 - (A \cup B) = 1 - 20/36 = 16/36$$

5.5 Histograma en ROOT

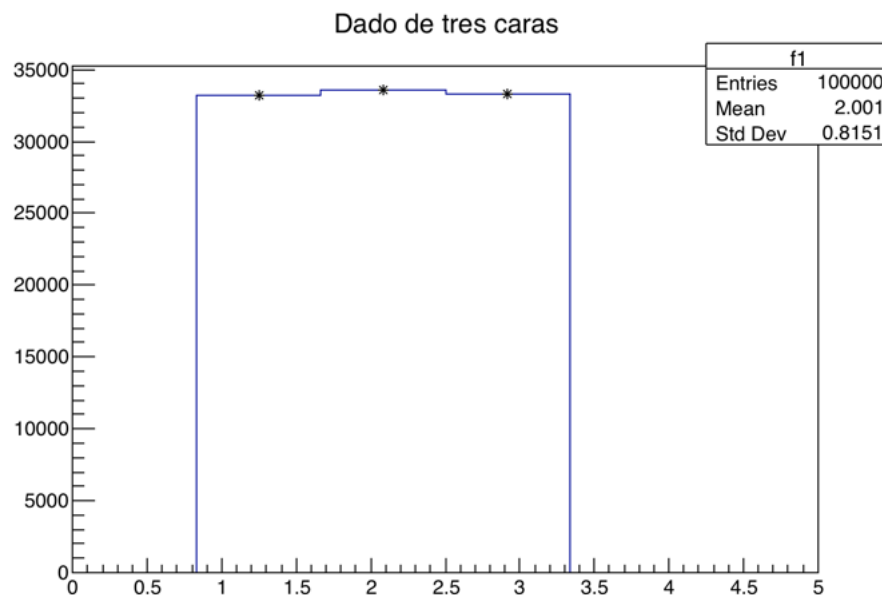


Figure 2: Histograma de un dado de 3 caras

6 Tarea 6

6.1 Teorema del límite central

El teorema establece que la distribución de \bar{x} , la media de una muestra aleatoria de una población con varianza finita tiene una distribución aproximadamente normal cuando el tamaño de la muestra es grande, independientemente de la forma de la distribución de la población.

Muchos procedimientos estadísticos comunes requieren que los datos sean aproximadamente normales, pero el teorema del límite central le permite aplicar estos procedimientos útiles a poblaciones que son marcadamente no normales.

6.2 Teoría de la Ruina del Jugador

En todos aquellos juegos de azar cuya marcha se basa en que los participantes deben apostar alguna cantidad de dinero para poder seguir jugando, es previsible que tarde o temprano alguno de los participantes terminará quedándose con el dinero de todos sus oponentes, o que a éstos se les irá

agotando el dinero y ya no podrán seguir participando en el juego, casos en los cuales el juego puede considerarse terminado tanto para los unos como para los otros.

Existe el interés por calcular las probabilidades que un jugador enfrentado a otro tiene de terminar a largo plazo en una situación en la cual ya no dispone de más dinero para seguir participando en el juego, situación que se conoce como el estado de ruina del jugador.

Para poder calcular esta probabilidad es necesario considerar el capital inicial que tiene cada jugador (en inglés *bankroll*). Consideremos un juego que consiste en lanzar una moneda justa (cada jugador tiene un 50% de ganar en cada tiro) y el jugador que pierde transfiere una moneda al ganador, hasta que uno de los dos se queda sin monedas.

El juego evidentemente termina si se imponen una restricción al número de tiros que se pueden realizar, pero el juego terminará con éstas reglas con probabilidad 1. Esto se puede confirmar debido a que cualquier posible cadena de resultados es posible, por más pequeña que sea su probabilidad de ocurrir.

Si el jugador uno tiene n_1 monedas y el jugador dos tiene n_2 monedas, la probabilidad P_1 y P_2 de los jugadores uno y dos, respectivamente, de terminar sin monedas esta dada por

$$P_1 = \frac{n_2}{n_1 + n_2}$$

$$P_2 = \frac{n_1}{n_1 + n_2}$$

En el caso general, sean q_z la probabilidad de ruina y p_z la probabilidad de ganar. Usando caminatas aleatorias, q_z es la probabilidad de que una partícula empezando en z sea absorbida en 0 y p_z se absorbida en a .

Después de la primera ronda, la fortuna del jugador es $z - 1$ o $z + 1$, por lo que

$$q_z = pq_{z+1} + qq_{z-1} \quad (2)$$

además

$$q_0 = 1, q_a = 0 \quad (3)$$

por lo que q_z satisface para $z = 1, 2, \dots, a - 1$.

Supongamos que $p \neq q$. Al considerar las condiciones de frontera (3) se obtiene [6]

$$q_z = \frac{(q/p)^a - (q/p)^z}{(q/p)^a - 1} \quad (4)$$

Para $q = p = 1/2$ la Eq. (2) resulta [6]

$$q_z = 1 - \frac{z}{a} \quad (5)$$

6.3 Histograma de números generados aleatoriamente según una distribución normal

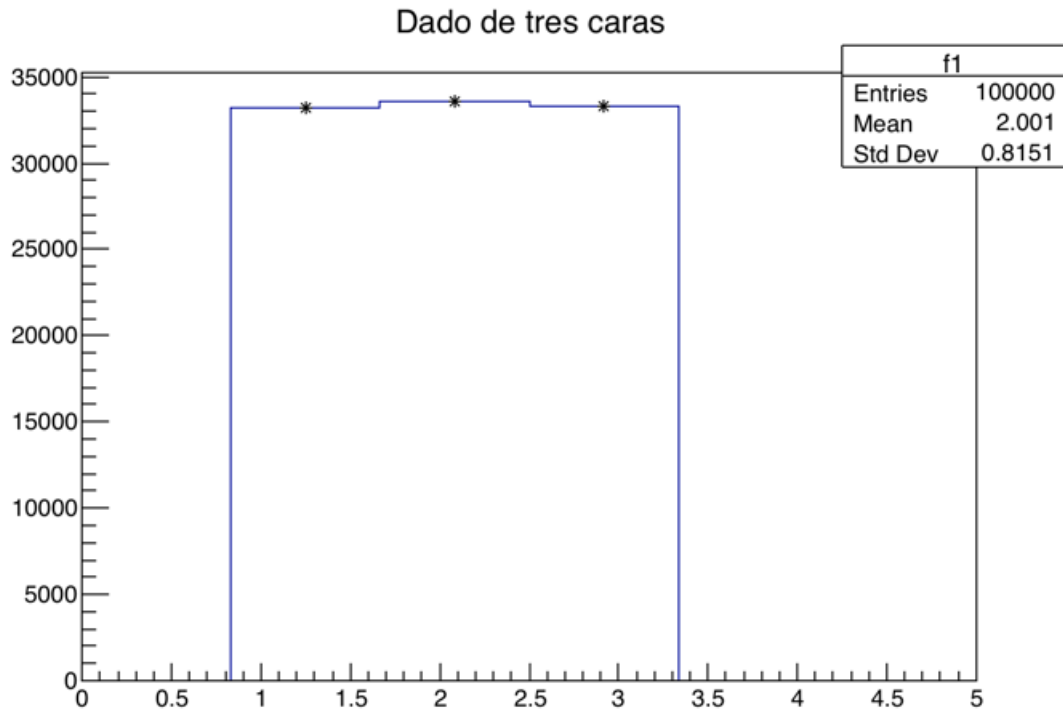


Figure 3: Histograma de números generados aleatoriamente según una distribución normal

6.4 Mínimos cuadrados

Al obtener datos acerca de un evento de interés, se busca encontrar alguna tendencia entre ellos representada mediante alguna fórmula matemática. El modelo más simple se reduce a una sola variable, es decir, una función que mapea de un conjunto X a un conjunto Y . Se desprende de éste caso que el modeo más sencillo no trivial es considerar que los datos se ajustan (o tienen una tendencia) lineal de la forma

$$y = mx + b$$

con un total de n pares ordenados (las medidas o datos) $(x_1, y_1), \dots, (x_n, y_n)$

Sea $\{(x_k, y_k)\}_{k=1}^n$ un conjunto de n pares con abscisas diferentes, y sea $\{f_j(x)\}_{j=1}^m$ un conjunto de m funciones linealmente independientes (es decir, forma un espacio vectorial de funciones). Se busca encontrar una función $f(x)$ que sea combinación lineal de dicho espacio, es decir

$$f(x) = c_1 f_1(x) + \dots + c_m f_m(x) = \sum_{j=1}^m c_j f_j(x)$$

Buscamos los m coeficientes y en concreto, una función $f(x)$ que sea la mejor aproximación a los n pares de datos empleando el criterio del mínimo error cuadrático medio de la función $f(x)$ con respecto a los n puntos.

La i -ésima observación tiene asociada una error e_i

$$y_i = \alpha + \beta x_i + e_i$$

El error cuadrático medio es

$$\begin{aligned} E_{cm} &= \sqrt{\frac{\sum_{k=1}^n (e_k)^2}{n}} \\ &= \sqrt{\frac{1}{n} \sum_{k=1}^n (y_k - f(x_k))^2} \\ &= \sqrt{\frac{1}{n} \sum_{k=1}^n (y_k - \sum_{j=1}^m c_j f_j(x_k))^2} \end{aligned}$$

Minimizar el error cuadrático medio es equivalente a minimizar el error cuadrático, definido como el radicando del error cuadrático medio, esto es:

$$E_c = \sum_{k=1}^n (y_k - \sum_{j=1}^m c_j f_j(x_k))^2$$

Los c_j que minimizan E_{cm} también minimizan E_c y se pueden calcular derivando e igualando a cero de la siguiente manera

$$\frac{\partial E_c}{\partial c_i} = \sum_{k=1}^n 2(y_k - \sum_{j=1}^m c_j f_j(x_k))(-f_i(x_k)) = 0$$

donde $i = 1, 2, \dots, m$.

Se obtiene un sistema de m ecuaciones con m incógnitas. Operando con ellas:

$$\sum_{k=1}^n \left(\sum_{j=1}^m c_j f_j(x_k) f_i(x_k) \right) = \sum_{k=1}^n y_k f_i(x_k)$$

para $i = 1, 2, \dots, m$; por otro lado

$$\sum_{j=1}^m \left(\sum_{k=1}^n f_i(x_k) f_j(x_k) \right) c_j = \sum_{k=1}^n y_k f_i(x_k)$$

para $i = 1, 2, \dots, m$.

Al desarrollar la suma para la i -ésima ecuación del sistema de m ecuaciones:

$$\left(\sum_{k=1}^n f_i(x_k) f_1(x_k) \right) c_1 + \dots + \left(\sum_{k=1}^n f_i(x_k) f_m(x_k) \right) c_m = \sum_{k=1}^n y_k f_i(x_k)$$

que se puede expresar de forma matricial de la siguiente manera:

$$\begin{pmatrix} (f_1, f_1)_d & (f_1, f_2)_d & \dots & (f_1, f_m)_d \\ \vdots & \vdots & \vdots & \vdots \\ (f_m, f_1)_d & (f_m, f_2)_d & \dots & (f_m, f_m)_d \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} (f_1, y)_d \\ \vdots \\ (f_m, y)_d \end{pmatrix}$$

Siendo $(a, b)_d$ el producto escalar discreto definido como

$$(h(x), g(x))_d = \sum_{k=1}^n h(x_k) g(x_k)$$

La resolución de dicho sistema permite obtener, para cualquier base de funciones derivables localmente, la función $f(x)$ que se aproxime al conjunto de puntos antes mencionado.

Por otro lado, se pueden calcular los estimadores de la recta de regresión especificada

$$Y_t = \beta_1 + \beta_2 X_t + U_t$$

a una recta de regresión estimada

$$\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_t$$

Los estimadores se obtienen aplicando el criterio de minimización de la suma cuadrática de los errores

$$\sum_{t=1}^n \hat{u}_t^2 = S = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^n (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_t)^2$$

Para minimizar S se deriva la función respecto de cada uno de los parámetros

$$\frac{\partial S}{\partial \hat{\beta}_1} = -2 \sum_{t=1}^n (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_t) = 0$$

$$\frac{\partial S}{\partial \hat{\beta}_2} = -2 \sum_{t=1}^n (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_t) = 0$$

Operando y reagrupando términos se obtendría la expresión analítica de los estimadores mínimo cuadráticos de la regresión lineal simple:

$$\begin{aligned}\hat{\beta}_1 &= \bar{Y} - \hat{\beta}_2 \bar{X} \\ \hat{\beta}_2 &= \frac{Cov(Y, X)}{var(X)}\end{aligned}$$

6.5 El hombre anumérico

A primera vista, resulta inconcebible la idea de que existan seres humanos que no puedan asimilar algunos conceptos básicos de las matemáticas. Uno no espera que al preguntarles acerca de temas tan comunes (en el sentido de que viven todos los días rodeados de aplicaciones prácticas) como la probabilidad y toda la jerga que conlleva o estimaciones respondan sin error o profundizen de manera magistral en el tema.

Se espera que ese sentido común de las cantidades y matemáticas esté activo en cada ser humano, que sea capaz de refutar ideas mediante razonamientos lógicos y le permita vivir una vida bajo la luz de éste sentido común. Más sin embargo, no todo en la vida es color de rosa.

Poseer un sentido común como el antes descrito puede llegar a ser privilegiado, pero en realidad cualquier persona puede desarrollarlo mediante la unión de la práctica y el conocimiento general: la cultura que va de los temas mundanos, cotidianos, hasta los más específicos que ahondan en ramas más lejanas del tronco principal. No hace mucho escuchaba en un podcast que es tanto el trauma emocional que causan las matemáticas que de inmediato pensar en algo que se relacione con ellas causa tanto desagrado como mirar un svástica.

Y no solo se trata de encontrar datos curiosos o triviales, si no de entender magnitudes temporales o físicas como "grande" o "pequeño". Entender conceptos como el principio de la multiplicación para contar eventos que aterrizan en la vida real y calcular probabilidades, comprender los resultados y sacar conclusiones basadas en los hechos y aprender a estimar y rechazar enunciados.

Todo mundo es libre de creer en lo que quiera, ya sea en relación a la divinidad y su influencia en la causalidad o cualquier otro fenómeno que uno puede considerar poco probable su aparición. Sin embargo, algo que he notado durante mi corta vida es que los sucesos que aparentemente "tienen poca probabilidad de suceder" acontecen con una frecuencia mayor de lo que uno esperaría: los dichosos eventos de colas. Sin embargo es prudente y de gente grande el no aceptar las coincidencias como actos milagrosos, actos unidos por un fino hilo rojo que intentan demostrar la pequeñez del hombre ante el mundo y sus extraas leyes que nos rigen.

La pseudociencia es otro de los aspectos que vuelven más vulnerables a los humanos. Ciertamente es que no todos tienen el digno acceso a una educación de calidad, y en general, las dificultades socioeconómicas de cada uno es un factor que impacta en como se forma intelectual y culturalmente una persona. Sin embargo, no es una ley y no se puede generalizar.

Podemos ver en la televisión anuncios de gente como Walter Mercado que presumía de sus habilidades para leer los astros e interpretar sus seales que rigen la vida de las personas. ¿Cuánta gente

cree en éstas cosas tanto como los cultos creen que la tierra es redonda? No lo sé con exactitud, pero si algo es continuamente publicitado por un medio de comunicación masivo, ciertamente tiene con que mantenerse. No es un insulto no menospreciar a la gente, es un llamado de observar y dar una explicación racional basada en hechos fidedignos.

No se insulta a la religión o a las creencias que cada quien tenga, pero el admitir que nuestros actos estan guiados por finos hilos dorados cuyo origen es el creador puede ser igual de creíble que aceptar cálculos par explicar ciertas condiciones de la vida diaria. ¿Cómo debemos regir la vida? Con nuestro sentido común, entender y aprender, escuchar y nunca estar cerrado a nuevos conocimientos.

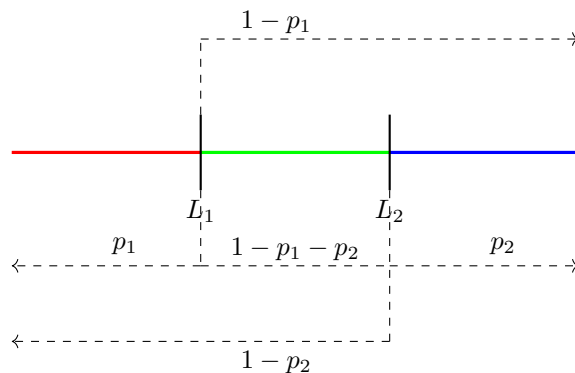
7 Tarea 7

7.1 Problema de los 4 tornillos

Una máquina que fabrica tornillos se desajusta de forma que produce tornillos de longitud indeterminada. Se sabe que la probabilidad de que un tornillo tenga una longitud menor a L_1 es p_1 y superior a L_2 es p_2 , donde $L_2 > L_1$.

Se extraen cuatro tornillos al azar, calcular:

- Probabilidad de que los cuatro tornillos tengan longitud superior a L_1
- Probabilidad de que tres de ellos tengan longitudes entre L_1 y L_2
- Probabilidad de que, a lo sumo, tres sean de longitud inferior a L_2



Para el inciso a) podemos utilizar un diagrama de árbol para encontrar la probabilidad de interés, marcada de color verde en la Fig.(4), y el resultado es

$$(1 - p_1)^4$$

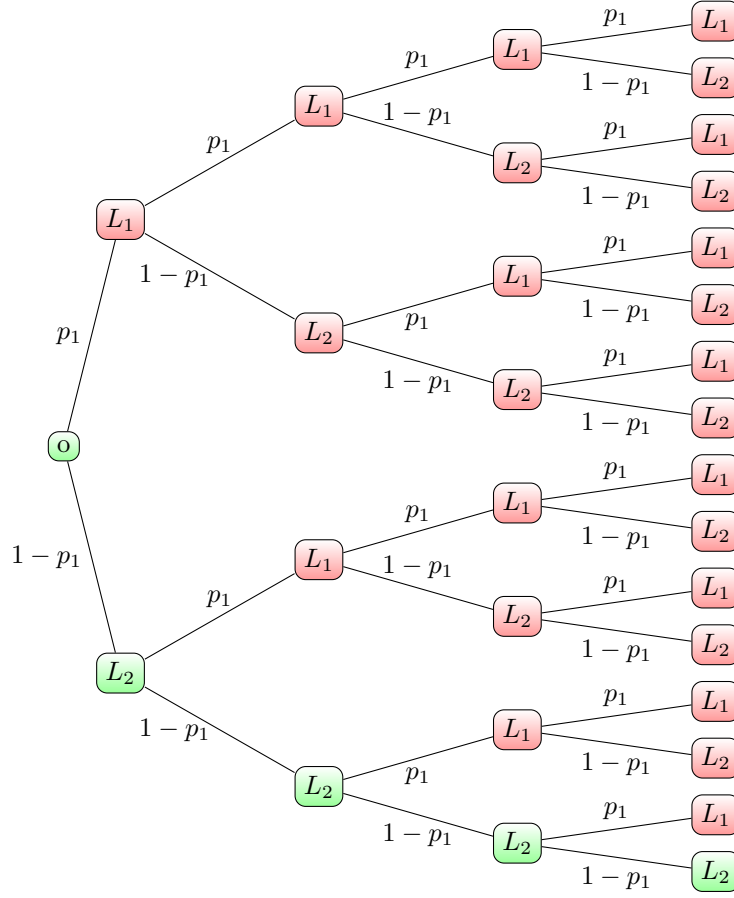


Figure 4: Los 4 tornillos tienen longitudes superiores a L_1

Para (b) podemos considerar un argumento de conteo. El número de formas de dividir n objetos distintos en k grupos distintos que contienen n_1, \dots, n_k objetos, respectivamente, donde cada objeto aparece en exactamente un grupo y $\sum_{i=1}^k n_i = n$ es[5]

$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1! n_2! \dots n_k!}$$

En este caso, se tiene que elegir 3 tornillos en el grupo que se encuentra entre L_1 y L_2 y el tornillo que sobra puede ser menor a L_1 o mayor a L_2 , por lo que, el coeficiente multinomial en este caso es

$$\binom{4}{1, 3} = \frac{4!}{1!3!} = 4$$

Como existen dos opciones para el tornillo que sobra, se tiene que multiplicar la probabilidad de que los tres tornillos se encuentren en el intervalo entre L_1 y L_2 por p_1 y el otro caso por p_2 y al

final sumar ambos resultados, obteniendo

$$\begin{aligned} P(3 \text{ tornillos entre } L_1 y L_2) &= 4p_1(1 - p_1 - p_2)^3 + 4p_2(1 - p_1 - p_2)^3 \\ &= 4(1 - p_1 - p_2)^3(p_1 + p_2) \end{aligned} \quad (6)$$

Para c) es mas fácil considerar el caso complementario a que a lo sumo tres de los tornillos sean inferiores a L_2 , esto es, considerar que los 4 tornillos que se eligieron tienen una longitud inferior a L_2 , cuya probabilidad se puede obtener del árbol en la Fig. (5), resultando

$$P(\text{a lo sumos 3 inf. a } L_2) = 1 - (1 - p_2)^4$$

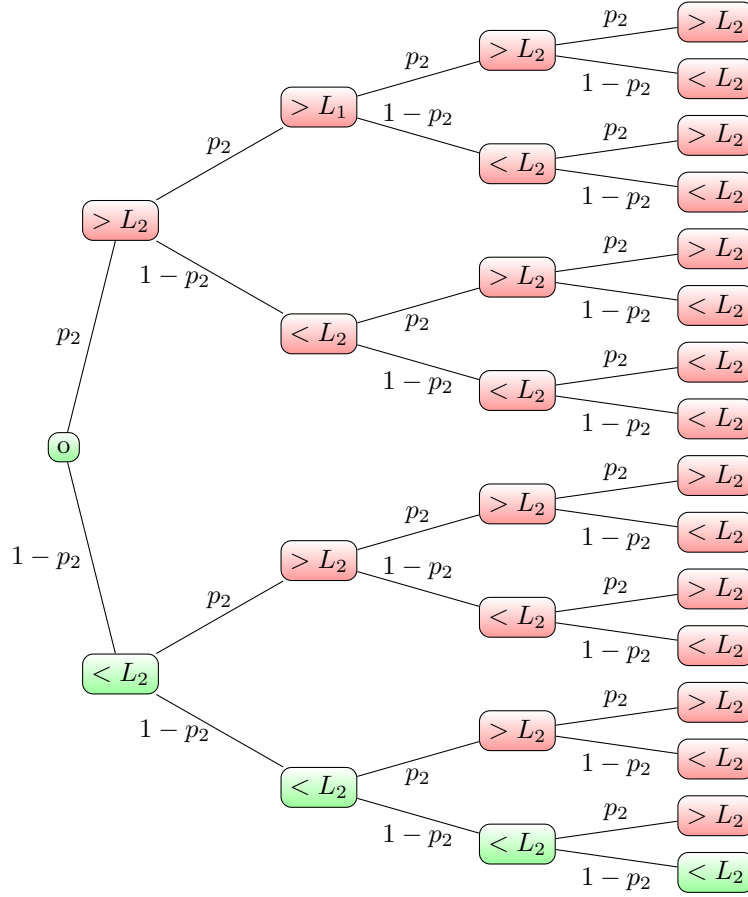


Figure 5: Los 4 tornillos tienen longitudes inferiores a L_2

7.2 Problemas Regla de Bayes

1. Una población de electores contiene 40% de republicanos y 60% de demócratas. Se publica que 30% de los republicanos y 70% de los demócratas están a favor de un tema de elección. Se encuentra que una persona seleccionada al azar de esta población está a favor del tema en cuestión. Encuentre la probabilidad condicional de que esta persona sea un demócrata.

Solución

Para este ejercicio nos preguntan la probabilidad condicional $P(\text{Demócrata}|\text{Favor})$, es decir, que sea un demócrata dado que la persona que se eligió está a favor. Las probabilidades se resumen en el árbol de la Fig. (6) obteniendo:

$$P(\text{Demócrata}|\text{Favor}) = \frac{(0.7) \cdot (0.6)}{(0.7) \cdot (0.6) + (0.4) \cdot (0.3)} \approx 0.77$$

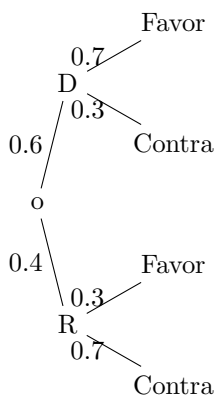


Figure 6: Árbol para el primer problema

2. Una prueba de diagnóstico para una enfermedad es tal que (correctamente) detecta la enfermedad en 90% de los individuos que en realidad tienen la enfermedad. También, si una persona no tiene la enfermedad, la prueba reportará que él o ella no la tiene con probabilidad 0.9. Sólo 1% de la población tiene la enfermedad en cuestión. Si una persona es seleccionada al azar de la población y la prueba de diagnóstico indica que tiene la enfermedad, ¿cuál es la probabilidad condicional de que tenga, en realidad, la enfermedad?

Solución

Se debe proceder con cautela para resolver este problema (y en general, cualquier otro que se plantee en la vida). El evento que sucede o del que se depende es que la prueba resulta positivo, y por consiguiente, el evento que está sujeto a que resultó positivo el examen es que esta realmente enferma esta persona. Lo antes dicho se puede representar de la siguiente forma:

$$P(\text{Enfermo}|\text{Positivo})$$

Del árbol de la Fig. (7) se obtiene que

$$P(\text{Enfermo}|\text{Positivo}) = \frac{(0.9) \cdot (0.01)}{(0.9) \cdot (0.01) + (0.1) \cdot (0.99)} \approx 0.083 \approx 8.33\%$$

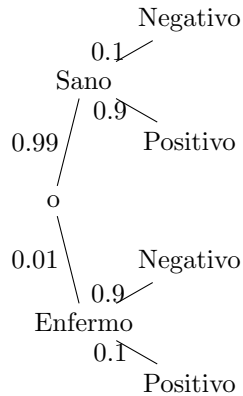


Figure 7: Árbol para el segundo problema

- Un médico cirujano se especializa en cirugías estéticas. Entre sus pacientes, el 20% se realizan correcciones faciales, un 35% implantes mamarios y el restante en otras cirugías correctivas. Se sabe además, que son de género masculino el 25% de los que se realizan correcciones faciales, 15% implantes mamarios y 40% otras cirugías correctivas. Si al seleccionar un paciente al azar resulta que es de género masculino, determine la probabilidad que se haya realizado una cirugía de implantes mamarios.

Solución

El problema consiste en que, dado que se seleccionó un hombre, determinar la probabilidad de que este se haya realizado una cirugía de implantes mamarios. Sean

F = cirugía facial
M = implantes mamarios
O = otras cirugías
H = hombre

Usando la fórmula de la regla de Bayes.

$$\begin{aligned} P(M|H) &= \frac{P(H|M) \cdot P(M)}{P(F)P(F|H) + P(M)P(M|H) + P(O)P(O|H)} \\ &= \frac{0.35 \cdot 0.15}{0.2 \cdot 0.25 + 0.35 \cdot 0.15 + 0.45 \cdot 0.40} \\ &\approx 0.19 \end{aligned}$$

- Sean tres urnas con las siguientes composiciones de bolas blancas y negras:

$U_1 : [3 \text{ blancas y } 2 \text{ negras}]$

$U_2 : [4 \text{ blancas y } 2 \text{ negras}]$

$U_3 : [1 \text{ blancas y } 4 \text{ negras}]$

Calcule:

- a) Probabilidad de extraer bola blanca
- b) Probabilidad de que una bola negra extraída proceda de la segunda urna

Solución

Podemos dar por hecho el que las tres urnas son equiprobables

$$P(U_1) = P(U_2) = P(U_3) = 1/3$$

Por el teorema de la probabilidad total

$$\begin{aligned} P(\text{blanca}) &= P(\text{blanca}|U_1)P(U_1) + P(\text{blanca}|U_2)P(U_2) + \\ &\quad + P(\text{blanca}|U_3)P(U_3) \\ &= \frac{3}{5} \cdot \frac{1}{3} + \frac{4}{6} \cdot \frac{1}{3} + \frac{1}{5} \cdot \frac{1}{3} = \frac{22}{45} \end{aligned}$$

Por el teorema de Bayes

$$P(U_2|\text{negra}) = \frac{P(\text{negra}|U_2)P(U_2)}{P(\text{negra})}$$

donde $P(\text{negra})$ se puede determinar usando probabilidad total o bien, $P(\text{negra}) = 1 - P(\text{blanca})$ por lo que

$$\begin{aligned} P(U_2|\text{negra}) &= \frac{P(\text{negra}|U_2)P(U_2)}{(1 - P(\text{blanca}))} \\ &= \frac{\frac{2}{6} \cdot \frac{1}{3}}{(1 - \frac{22}{45})} \\ &= \frac{5}{23} \end{aligned}$$

5. Se tienen dos lotes con piezas procedentes de un mismo proceso productivo. Del primer lote (L_1) se sabe que un 1% son defectuosas, en tanto que del segundo (L_2) lo son el 5%. El primer lote tiene el doble de piezas que el segundo.

Se elige aleatoriamente una pieza de esos lotes:

- (a) Calcule la probabilidad de que la pieza elegida sea defectuosa.
- (b) Si la pieza es defectuosa, calcule la probabilidad de que proceda del segundo lote.

Solución

a) Primero calculemos la probabilidad de elegir de un lote en concreto:

$$\begin{aligned} P(L_1) + P(L_2) &= 1 \\ P(L_1) &= 2P(L_2) \end{aligned} \tag{7}$$

lo que lleva a que $P(L_1) = 2/3$ y $P(L_2) = 1/3$.

Del teorema de probabilidad total:

$$\begin{aligned} P(D) &= P(D|L_1)P(L_1) + P(D|L_2)P(L_2) \\ &= 1/100 \cdot 2/3 + 5/100 \cdot 1/3 = 7/300 \end{aligned}$$

(b) Por el teorema de Bayes

$$\begin{aligned} P(L_2|D) &= \frac{P(D|L_2)P(L_2)}{P(D|L_1)P(L_1) + P(D|L_2)P(L_2)} \\ &= \frac{5/100 \cdot 1/3}{7/300} = 5/7 \end{aligned}$$

7.3 Grafos de divisibilidad

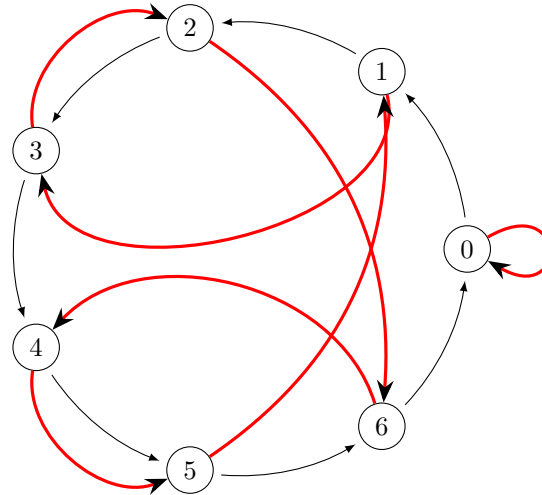


Figure 8: Grafo para divisibilidad entre 7

En general, usando un grafo de divisibilidad se puede saber si cualquier número es divisible entre n , por supuesto, considerando la divisibilidad entre números enteros. El primer caso que consideraremos es la divisibilidad entre 7, como se muestra en la Fig. (7.3). El algoritmo es el siguiente (usando el grafo):

1. Usando el primer dígito de izquierda a derecha se recorren d_1 lugares correspondientes a las flechas negras.
2. En el nodo en que se para, se sigue la flecha roja que lleva al nodo correspondiente
3. Al nodo que se llega desde la flecha roja, se vuelve a aplicar el proceso usando el segundo dígito recorriendo d_2 lugares en dirección de las flechas negras.
4. Se repite el proceso hasta llegar al último dígito del número en cuestión. El último nodo que nos lleva la última flecha roja representa el resto del número modulo 7. Si es cero, resulta que es divisible entre 7.

¿Cuál es el "truco"? El método funciona porque se cumple que:

$$(10 * a + b) \mod 7 = ((10 * a) \mod 7 + b) \mod 7$$

El grafo está hecho de modo que las flechas blancas apuntan a $n * 10 \mod 7$, siendo n el número del nodo del círculo en que se encuentre. La flecha negra acarrea el dígito multiplicado por una potencia de 10^n y la flecha roja lleva el residuo (o el módulo del número), después se suma el siguiente dígito, no es más que una forma distinta de ver la división que enseñan en la primaria.

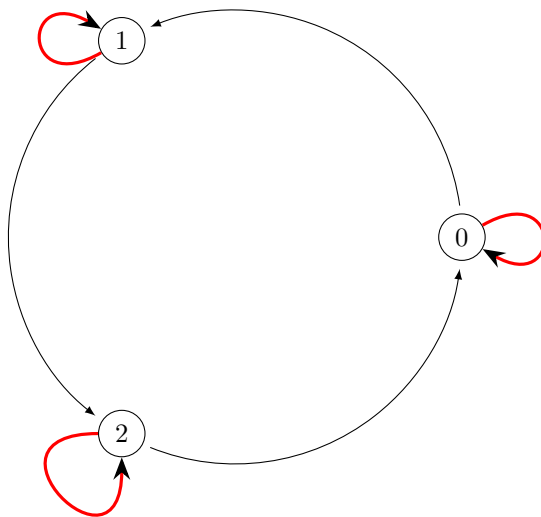


Figure 9: Grafo para divisibilidad entre 3

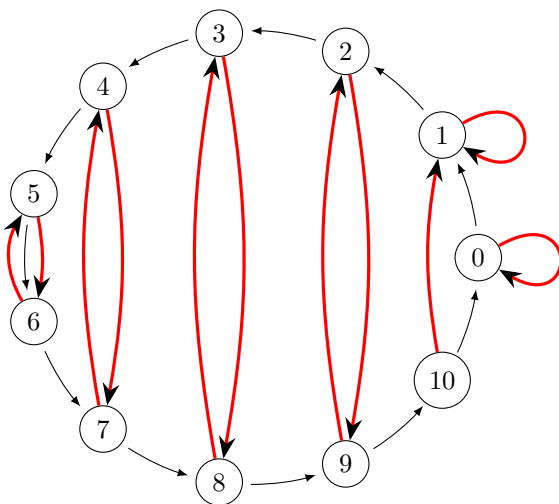


Figure 10: Grafo para divisibilidad entre 11

8 Tarea 8

8.1 Tensor de Levi-Civita

En matemáticas, y en particular en cálculo tensorial, se define el símbolo de Levi-Civita, también llamado el símbolo de permutación o tensor de Levi-Civita, como sigue:

$$\epsilon_{ijk} = \begin{cases} +1 & \text{si } (i, j, k) \text{ es } (1, 2, 3), (2, 3, 1) \text{ o } (3, 1, 2) \\ -1 & \text{si } (i, j, k) \text{ es } (3, 2, 1), (1, 3, 2) \text{ o } (2, 1, 3) \\ 0 & \text{si } i = j, j = k \text{ o } k = i \end{cases}$$

El producto vectorial de dos vectores se puede obtener del cálculo de la determinante de la siguiente matriz usando como menores la primera fila de la matriz

$$a \times b = \begin{vmatrix} e1 & e2 & e3 \\ a1 & a2 & a3 \\ b1 & b2 & b3 \end{vmatrix}$$

Usando la notación de Levi-Civita, se puede escribir más fácilmente como:

$$a \times b = c, \text{ donde } c_i = \sum_{j,k=1}^3 \epsilon_{ijk} a_j b_k$$

Por ejemplo, para el dos vectores \vec{a} , \vec{b} de tres componentes

$$\vec{a} = (a_1, a_2, a_3)$$

$$\vec{b} = (b_1, b_2, b_3)$$

utilizando la notación de Einstein

$$\begin{aligned}(a \times b)_1 &= \epsilon_{123} = a_2 b_3 - a_3 b_2 \\(a \times b)_2 &= \epsilon_{2,1,3} = -(a_1 b_3 - b_1 a_3) \\(a \times b)_3 &= \epsilon_{3,1,2} = a_1 b_2 - b_1 a_2\end{aligned}$$

8.2 La máquina enigma

A simple vista se ve igual que una antigua máquina de escribir, con la diferencia de que era un híbrido mecánico eléctrico. El proceso no es tan complicado de explicar: primero, se introduce la letra del mensaje a cifrar y después se ilumina otra letra en otro teclado que se ilumina, denotando que esa es la letra con la que se cifra la letra original.

Por ejemplo, el operador al oprimir la letra "T" en el teclado, se crea una seal eléctrica que se enlaza a la letra "T" en el *plugboard*. En esta parte, las letras están conectadas con otras, lo que provoca que se cambie la letra original por aquella con la que se esta enlazada. O puede no estar enlazada con otra letra, por lo que letra original pasa tal cual al siguiente componente.

La nueva letra (o la original, según sea el caso) se manda al rotor estático que no hace nada a la seal, solo convierte los cables en contactos. La salida del rotor estática es conectada a la entrada del primer rotor.

Cada rotor tiene un anillo interno de contactos y otro anillo externo de contactos cuyo fin es revolver la seal que se manda. El anillo externo conecta a dos rotores y al anillo interno. El anillo interior de puede rotar en relación a su correspondiente anillo externo y todo el rotor se puede rotar en relación al rotor estático. ¿Cuántas formas existen de acomodar 5 rotores en 3 espacios disponibles? Existen $5 \cdot 4 \cdot 3 = 60$ formas distintas de acomodar los 5 rotores y, además existen 26^3 posiciones iniciales para cada rotor. En cuanto a los anillos de cada rotor, existen $26^2 = 676$ formas de posicionar los anillos en una máquina de 3 rotores.

Ya que pasaba por cada rotor, llegaba a uno de los dos reflectores que mandaba la letra recibida transformada en otra letra. La salida vuelve a pasar por los rotores, volviendo a revolver la letra en su camina de regreso al *plugboard*.

Para terminar el viaje, la letra resultante en el paso anterior se ilumina en el teclado luminoso de la máquina, resultando ser el cifrado de la letra que originalmente se introdujo.

Las máquinas enigma constaban de 10 cables para conectar pares de letras, por lo que resulta natural preguntar cuantas maneras existen de conectar parejas de letras en la máquina. En total, existen 150,738,274,937,250 configuraciones. Una manera de ver el resultado es como sigue: supongamos que los 10 cables son de distinto color, por lo que podemos diferenciarlos. Existen $C(26,2)$ maneras de escoger un par de letras para el primer cable y $C(24,2)$ para el segundo cable, etc., es decir

$$\prod_{i=0}^{10} \binom{26-2i}{2} = \frac{26!}{6! \cdot 2^{10}}$$

Pero la máquina enigma no tenia los cables coloreados, por lo que hay dividir entre el resultado de

permutar los 10 cables ($10!$), es decir

$$\frac{26!}{6! \cdot 10! \cdot 2^{10}}$$

De manera abstracta, se tiene que para elegir m pares de n objetos la fórmula

$$\frac{n!}{((n-2m)!m!2^m)}$$

Finalmente, existen

$$60 \cdot 17,576 \cdot 676 \cdot 150,738,274,937,250$$

formas en que se puede configurar una máquina enigma.

8.3 Esperanza matemática

La distribución de probabilidad para una variable es un modelo teórico para la distribución empírica de datos asociados con una población real. Si el modelo es una representación precisa de la naturaleza, las distribuciones teóricas y empíricas son equivalentes. En consecuencia, se trata de hallar la media y la varianza de una variable aleatoria y por tanto adquirir medidas numéricas descriptivas, parámetros, para la distribución de probabilidad $p(x)$.

Sea x una variable aleatoria discreta con valores posibles $x_1, x_2 \dots x_n$ y sus probabilidades representadas por la función de probabilidad $p(x_i)$ la esperanza se calcula como:

$$\mathbb{E}[X] = x_1p(X = x_1) + \dots + x_np(X = x_n) = \mathbb{E}[X] = \sum_{i=1}^n x_ip(x_i)$$

Para una variable aleatoria absolutamente continua, la esperanza se calcula mediante la integral de todos los valores y la función de densidad $f(x)$:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)dx$$

Sus propiedades:

1. Si X es siempre positiva, entonces siempre lo es $E(X)$.
2. La esperanza matemática de una constante es igual a esa misma constante, es decir, si c es una constante, entonces $\mathbb{E}[c] = c$.
3. Si X está delimitada por dos números reales, a y b , tal que: $a < X < b$, entonces también lo está su media: $a < \mathbb{E}(X) < b$
4. Linealidad. Si existe $\mathbb{E}(X)$ y se considera $Y = a + bX$, entonces $\mathbb{E}(Y) = \mathbb{E}(a + bX) = a + b\mathbb{E}(X)$

8.4 Un pájaro en mano vale más que 2.48 volando

En cualquier parte de nuestra vida nos hemos topado con la experiencia de participar en aquellos juegos que nos permiten incrementar nuestro capital: las loterías. Y realmente no se puede culpar a nadie por haber participado en alguna edición de éstas ya que el premio puede resultar muy llamativo.

Sin embargo, poniendo un poco más de seso al asunto, la probabilidad de ganar una lotería es muy remota, no imposible en el estricto sentido de la palabra pero, si se puede definir de alternativamente la palabra imposible, sería *ganar la lotería*.

En realidad nuestro cerebro no está condicionado para procesar (y mucho menos entender a primera vista) números tan pequeños (¿infinitesimales?). Resulta natural pensar que en algún momento, a alguien le debe tocar el premio gordo, lo cual no es incorrecto, le pregunta es cuanto tiempo se debe esperar.

A final de cuentas la naturaleza humana es muy extraña, y por más que se presuma nuestro avanzado intelecto y el estudio de nuestra conciencia, tenemos comportamientos que están profundamente enraizados en nosotros. Tendemos a actuar por miedo, miedo a perder nuestras esperanzas.

La gente educada financieramente contempla este tipo de actividades como una diversión, cuando es en realidad un gasto ya que no se tiene promesa alguna de que obtengamos beneficio alguno. Para observar numéricamente este resultado es útil utilizar el concepto de esperanza matemática definido en el punto anterior. Resulta pues, que existen dos posibles resultados: ganar el premio gordo o perder el costo de la participación en el sorteo. Más explícitamente:

$$\mathbb{E}[X] = wp(X = \textit{ganar}) + lp(X = \textit{perder})$$

Dado que las probabilidades de ganar una lotería son bajas, resulta que, al observar el término que acompaña el evento de ganar, resulta que el valor de la operación es pequeño, dependiendo de cada lotería, alrededor del orden de 10^{-5} hasta del orden de 10^{-7} por lo que en realidad aporta poco a la suma, lo que realmente lleva a considerar el gasto que se realiza.

Nuestro cerebro es más sensible a la posibilidad de perder algo que a la opción de ganar algo, usualmente llamado aversión a la pérdida. Nos negamos a perder cosas. Tenemos miedo de perder cosas. Nuestra naturaleza es evitar cualquier tipo de pérdida, por más pequeña que sea, y peor si nos comparamos con los demás, somos muy codiciosos.

Julian Baggini, filósofo y analista de la mente del University College de London dice que un pájaro en mano vale 2.48 pájaros volando. Para comprobar su idea, realizó un experimento: a dos grupos los separaron en dos habitaciones. Al primer grupo se le regaló una taza de barro, que era completamente suya. Al segundo grupo solo se les presentó la taza y se les permitió examinarla. Más tarde se pidió a cada grupo que valoraran económicamente el valor aproximado de la taza. El primer grupo le dio un valor más alto a la taza que al segundo. Al analizar los resultados numéricos obtenidos, se encontró que el primer grupo valoraba su taza 2.48 veces más que el segundo grupo.

Parece paradójico, pero el hecho de que la probabilidad de ganar sea tan pequeña es lo que más atrae a los jugadores. Por otro lado, no se pierde mucho, se tiene más que ganar. La estadística se vuelve irrelevante, nuestro cerebro sólo se fija en lo barato que resulta probar.

Nuestro cerebro a la hora de decidir si jugará o no intentará buscar información sensible, racional. Al momento de no encontrar *la* información racional, buscará cualquier razón a la que nosotros le demos la suficiente lógica, nos engañamos a nosotros considerando que aquel número que no podemos imaginar es en realidad algo alcanzable. A final de cuentas es suerte, azar; una probabilidad no deja de ser más que eso y frecuentemente olvidamos lo que significa.

8.5 Tensor covariante y contravariante

Los conceptos de covarianza, contravarianza e invarianza son una de las herramientas más importantes de la formulación matemática de las leyes físicas.

Supongamos que en un determinado espacio (euclídeo o no, las condiciones necesarias para su validez no nos son importantes) de dimensión N tenemos una base de N vectores linealmente independientes. Por ejemplo, en \mathbb{R}^2 la base de vectores ortonormales

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

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

Se define como base principal a la base de vectores a la cual referiremos las coordenadas de los demás elementos del espacio, y como elementos covariantes con ella a los que se transforman del mismo modo que ella bajo un cambio de coordenadas.

Los elementos covariantes se especifican escribiendo sus índices bajo los mismos. De este modo tenemos en el ejemplo:

$$e_1 = i$$

$$e_2 = j$$

Se define como base contravariante al conjunto de N vectores (con el índice arriba) que cumplen la siguiente condición con respecto a los productos escalares con los elementos de la base principal:

$$e_i e^j = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

es decir, para cada elemento i de la base principal existe un único elemento j con el cual posee un producto escalar igual a la unidad, y es ortogonal a todos los demás (en base a que si el producto escalar es nulo los vectores son ortogonales). Para expresarlo más cómodamente se define la delta de kronecker y se incluye en la ecuación:

$$\delta_i^j = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

$$e_i e^j = \delta_i^j$$

En el ejemplo, podemos encontrar la base contravariante:

$$e^1 = \begin{pmatrix} e^{11} \\ e^{22} \end{pmatrix}$$

$$e^2 = \begin{pmatrix} e^{21} \\ e^{22} \end{pmatrix}$$

$$\begin{cases} e_1 e^1 = 1 \\ e_2 e^1 = 0 \\ e_1 e^2 = 0 \\ e_2 e^2 = 1 \end{cases}$$

$$\begin{cases} e^{11} = 1 \\ e^{12} = 0 \\ e^{21} = 0 \\ e^{22} = 1 \end{cases}$$

Supongamos una transformación de la base principal que la lleva a otra base de N elementos asociados uno a uno con la anterior, pongamos por ejemplo:

$$e'_1 = e_1$$

$$e'_2 = e_1 + e_2$$

mantenemos el primer vector de la base idéntico, y hacemos que en esta ocasión el segundo vector base sea la suma de los dos anteriores. Definimos como matriz de transformación covariante a la matriz Λ que expresa este cambio de coordenadas:

$$= \begin{pmatrix} e'_1 \\ e'_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$$

Resulta de nuevo más cómodo introducir una notación más clara para esta expresión, y para ello la expresaremos con la notación de Einstein. Podemos expresar cada elemento de la matriz según su fila i y su columna i' , de modo que con la notación de Einstein el hecho de que dos elementos tengan la misma letra en un índice y uno de los índices esté abajo (sea covariante) y el otro esté arriba (sea contravariante) implica que se suman todos los posibles valores del índice. Para verlo, se escribe la transformación de nuevo y después se expone lo que representa en nuestro ejemplo:

$$\begin{aligned} e_{i'} &= \Lambda_{i'}^i e_i \\ e_{i'} &= \Lambda_{i'}^1 e_1 + \Lambda_{i'}^2 e_2 \\ e_1 &= \Lambda_1^1 e_1 + \Lambda_1^2 e_2 \\ e_2 &= \Lambda_2^1 e_1 + \Lambda_2^2 e_2 \end{aligned} \tag{8}$$

En este nuevo sistema coordenado, la base contravariante será distinta también, pues el conjunto de vectores que cumplen los requisitos adecuados ya no son los mismos. Podemos calcularlos de modo análogo a como hicimos antes:

$$e'^1 = \begin{pmatrix} e'^{11} \\ e'^{22} \end{pmatrix}$$

$$e^2 = \begin{pmatrix} e'^{21} \\ e'^{22} \end{pmatrix}$$

$$\begin{cases} e'_1 e'^1 = 1 \\ e'_2 e'^1 = 0 \\ e'_1 e'^2 = 0 \\ e'_2 e'^2 = 1 \end{cases}$$

$$\begin{cases} e'^1 = 1 \\ e'^1 + e'^2 = 0 \\ e'^2 = 0 \\ e'^1 + e'^2 = 0 \end{cases}$$

$$e'^1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$e'^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

9 Tarea 9

9.1 Ventaja

El término ventaja se puede definir como la cualidad de tener una posición superior o más favorable. Puede ser muy ambiguo el utilizar una definición tan general, sin embargo podemos intentar estructurar algo más concreto: las apuestas.

En el mundo de las apuestas, cuando se habla acerca de dos jugadores (equipos, personas, etc.) se informa quien tiene una ventaja sobre el otro. Por otro lado, en los juegos de azar hay algo conocido como apuesta con ventaja. El término hace referencia a métodos legales usados para poner en favor la balanza de un juego al apostar, usualmente usado en juegos en los que la casa tiene (casi) todas las de ganar, aunque existen juegos entre varios jugadores como el poker que también entran en éste campo. Juegos como el blackjack pueden ser vencidos mediante técnicas como el conteo de cartas, trazo al barajar, ordenar los cantos y demás métodos.

Mario Puzo se refería en su famosa novela *Fools Die* cuando su personaje ficticio, un jefe de un casino, Gronevelt, comentaba: "... Los porcentajes nunca mienten. Hemos construido todos estos hoteles en porcentajes. Nos hacemos ricos en el porcentaje. Usted puede perder la fé en todo, la religión y Dios, las mujeres y el amor, el bien y el mal, la guerra y la paz. Lo que sea. Pero el porcentaje siempre estará firme".

Puzo argumentaba lo correcto acerca del dinero en los casinos. Sin el "riesgo", no existirían los casinos. Con esta ventaja, ya causa de un resultado matemático famoso llamado la ley de los grandes números, un casino está garantizado a ganar en el largo plazo.

Las posibilidades del jugador de ganar en un juego de casino y la velocidad a la que gana o pierde dinero depende del juego, las reglas en vigor para ese juego, y para algunos juegos, de su nivel de habilidad. La cantidad de dinero que el jugador puede esperar ganar o perder en el largo plazo (si la apuesta se hace una y otra vez) se llama valor esperado apuesta del jugador (*EV*), o expectativa. Cuando la expectativa de la apuesta del jugador es negativa, va a perder dinero en el largo plazo.

Para una apuesta de \$5 en el color rojo en la ruleta, por ejemplo, la expectativa es -\$0.263. En promedio, el jugador perder poco ms de un cuarto para cada apuesta de \$5 en el color rojo.

En la ruleta existen los siguientes casos:

Cuando la expectativa de la apuesta se ve desde la perspectiva de los casinos (el negativo de las expectativas del jugador) y se expresa como un porcentaje, se tiene la ventaja de la casa. Para el ejemplo de la ruleta, la ventaja de la casa es de 5,26% (\$0.263 dividido por \$5). El cálculo formal es la siguiente:

$$EV = (+5)(18/38) + (-5)(20/38) = -0.263$$
$$(\text{Ventaja de la casa} = 0.263/5 = 5.26\%)$$

Cuando este cálculo EV se realiza para una cantidad de 1 unidad, el negativo del valor resultante es la ventaja de la casa. Estos son los cálculos para la apuesta a un solo número en doble cero y la ruleta de un solo cero:

Ruleta con doble cero (apuesta a un número sencillo):

$$EV = (+35)(1/38) + (-1)(37/38) = -0.053$$
$$(\text{Ventaja de la casa} = 5.3\%)$$

Ruleta con cero sencillo (cero sencillo):

$$EV = (+35)(1/37) + (-1)(36/37) = -0.027$$
$$(\text{Ventaja de la casa} = 2.7\%)$$

La ventaja de la casa representa el porcentaje de largo plazo del dinero apostado que será retenido por el casino. También se le llama el borde de la casa, las "probabilidades" (es decir, se debe evitar juegos con malas probabilidades), o simplemente el "porcentaje".

Algunos juegos de casino son pura casualidad: ninguna cantidad de habilidad o estrategia puede alterar las probabilidades. Estos juegos incluyen ruleta, dados, baccarat, keno, los seis grandes ruedas de la fortuna, y máquinas tragamonedas. De éstos, baccarat y craps ofrecen las mejores probabilidades, con las ventajas de la casa de 1.2% y menos del 1%, respectivamente. Ruleta y tragaperras cuestan el jugador más (ventajas de la casa de 5,3% para el doble cero ruleta y el 5% y el 10% para las ranuras) mientras que la rueda de la fortuna alimenta al casino cerca de 20% de las apuestas y keno es una mina de oro para el casino con una ventaja de la casa en promedio cercano al 30%.

La probabilidad representa la razón a largo plazo de (# de veces que el evento ocurre) entre (# de veces que el experimento es realizado). Las "probabilidades en contra" representan la razón a largo plazo de (# de veces que el evento no ocurre) entre (# de veces que el experimento es realizado). Si una carta es elegida al azar de un mazo estándar de 52 cartas, la probabilidad de que sea una espada es 1/4, por lo que la probabilidad en contra (*odds* son 3 a 1). El pago justo cuando se juega en contra de las probabilidades es aquel que hace la apuesta en el evento justo. En la ruleta, al apostar por un número el pago debe ser de 37 a 1, cuando es pagado 35 a 1.

Fichas	Apuesta	Se juega a	Premio
1	Rojo/Negro	Se apuesta al color de número ganador, si será rojo o negro. Con esta apuesta se está jugando a 18 números ya que en la ruleta hay 18 números rojos y 18 números negros.	1 x 1
2	Par/Impar	Se apuesta a si el número donde cae la bola será par o impar. Con esta apuesta se está jugando a 18 números, bien a los 18 números pares o los 18 números impares que están en la ruleta.	1 x 1
3	Pasa/Falta	Se trata de apostar si el número estará comprendido entre los números del 1 al 18 (falta) o entre los números del 19 al 36 (pasa). Por tanto, con esta apuesta se está jugando a 18 números.	1 x 1
4	Docena	Se trata de apostar en que docena estará el número ganador. El tapete se divide en 3 docenas, cada una de ellas abarca 12 números, por tanto al apostar por una docena se juega a 12 números.	2 x 1
5	Columna	Se trata de apostar en que de que columna será el número ganador. El tapete se divide en 3 columnas, cada una de ellas alberga 12 números. Por tanto, al apostar por una columna se juega a 12 números.	2 x 1
6	Dos Docena	Se trata de apostar con una sola apuesta a dos docenas, esta apuesta solo se puede hacer para dos docenas contiguas, es decir, se puede apostar a las docenas 1 y 2, o a las docenas 2 y 3. Al apostar a 2 docenas se juega a 24 números.	0,5 x 1
7	Dos Columnas	Se trata de apostar con una sola apuesta a dos columnas, esta apuesta solo se puede realizar para dos columnas contiguas. Por tanto, con la apuesta de dos columnas se podrá apostar a las columnas 1 y 2 o a las columnas 2 y 3. Se juega a 24 números.	0,5 x 1
8	Seisena	Se trata de apostar a 6 números con una sola apuesta. Los 6 números sobre los que se realiza este tipo de apuesta se encuentran en dos filas contiguas.	5 x 1
9	Cuadro	Se trata de apostar a 4 números con una sola apuesta. Esta apuesta se realiza sobre 4 números que forman un cuadrado en el tapete.	8 x 1
10, 11,12	Transversal	Se trata de apostar a 3 números con una sola apuesta, con la apuesta se apuesta a los 3 números de una fila. Existen dos variaciones de esta apuesta, la apuesta transversal para apostar a los 3 números 0,1 y 2 y la apuesta transversal para apostar a los 3 números 0, 2 y 3.	11 x 1
13,14	Caballo	Se trata de apostar a 2 números con una sola apuesta, los 2 números deberán estar contiguos en el tapete de manera horizontal o vertical.	17 x 1
15	Pleno	Se trata de apostar a un solo número.	35 x 1

Table 7: Reglas de la ruleta

Juego	Ventaja de la casa
Roulette (double-zero)	5.3%
Craps (pass/come)	1.4%
Craps (pass/come with double odds)	0.6%
Blackjack - average player	2.0%
Blackjack - 6 decks, basic strategy*	0.5%
Blackjack - single deck, basic strategy*	0.0%
Baccarat (no tie bets)	1.2%
Caribbean Stud*	5.2%
Let It Ride*	3.5%
Three Card Poker*	3.4%
Pai Gow Poker (ante/play)*	2.5%
Slots	5% - 10%
Video Poker*	0.5% - 3%
Keno (average)	27.0%

*Estrategia óptima

Table 8: Ventajas de la casa en juegos populares [10]

La teoría estadística se puede utilizar para predecir la magnitud de la diferencia entre el porcentaje de victorias real y el porcentaje de victorias teórico para un número dado de apuestas. Al observar el porcentaje real para ganar que un jugador (o casino) pueden experimentar, ¿cuál es el grado de variación de la victoria teórica que se puede esperar? ¿Qué es una fluctuación normal? La base para el análisis de tales preguntas de volatilidad es una medida estadística llamada la desviación estándar (esencialmente la desviación media de todos los posibles resultados esperados). Junto con el teorema del límite central (una forma de la ley de los grandes números), la desviación estándar (SD) se puede utilizar para determinar los límites de confianza con las siguientes directrices de volatilidad:

- Sólo el 5% de las veces se los resultados será de más de 2 desviaciones estándar de resultado esperado.
- Casi nunca (0,3%) se resultados será de más de 3 desviaciones estándar del resultado esperado.

9.2 El problema del caballero De Meré

Alrededor del año 1651, el caballero De Meré, gran aficionado a los juegos de azar, y los matemáticos Pascal y Fermat, hicieron un estudio exhaustivo sobre el cálculo de probabilidades.

De Meré estudió la frecuencia con la que aparecían ciertos sucesos relacionados con los juegos de azar, y estas observaciones llevaron a plantear a Pascal determinados problemas que, a su vez, dieron origen a una correspondencia entre Pascal y el matemático Pierre de Fermat, para tratar de darles solución.

Uno de estos problemas fue:

Deseo averiguar si es o no ventajoso jugar apostando cantidades iguales a que, por lo menos, aparece un 6 en cuatro tiradas de un dado.

La solución que dio Pascal fue la siguiente:

La probabilidad de que en una tirada no salga un 6 es igual a $5/6$; todas las tiradas son independientes entre sí, y el resultado de una no influye en la otra; por tanto, la probabilidad de que en las cuatro tiradas no salga ningún 6 será:

$$P(\text{no sacar ningún 6}) = \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \approx 0.518$$

Esa probabilidad es ligeramente mayor que 0.5, por lo que es una apuesta ventajosa, aunque no en exceso, ya que se deberían jugar bastantes partidas para que se apreciara esa ligera diferencia sobre 0.5. De cada 100 partidas se ganarían 51.8, es decir, unas 52, y se perderían 48.

10 Tarea 10

10.1 Congruencia de Zeller

La congruencia de Zeller es un algoritmo ideado por Julius Christian Johannes Zeller para calcular el día de la semana de cualquier fecha del calendario.

Para el calendario gregoriano la congruencia de Zeller es

$$h = \left(q + \left\lfloor \frac{(m+1)26}{10} \right\rfloor + K + \left\lfloor \frac{K}{4} \right\rfloor + \left\lfloor \frac{J}{4} \right\rfloor - 2J \right) \mod 7$$

para el calendario juliano es

$$h = \left(q + \left\lfloor \frac{(m+1)26}{10} \right\rfloor + K + \left\lfloor \frac{K}{4} \right\rfloor + 5 - J \right) \mod 7$$

donde

- h es el día de la semana (0 = sábado, 1 = domingo, 2 = lunes,...)
- q es el día del mes
- m es el mes
- J es la centuria (año / 100) y
- K el año de la centuria (año mod100).

Enero y febrero se cuentan como meses 13 y 14 del año anterior; el 2 de enero del 2013, es $m=13$; año=2012.

En las implementaciones informáticas en las que el módulo de un número negativo es negativo, la manera más sencilla de obtener un resultado entre 0 y 6 es reemplazar $-2J$ por $+5J$ y $-J$ por $+6J$.

Cada término de la fórmula se usa para calcular el desplazamiento necesario para obtener el día correcto de la semana. Para el caso del calendario Gregoriano, que es el que rige todo occidente: q representa la progresión del día de la semana basada en el día del mes, dado que cada día sucesivo resulta en un desplazamiento adicional de 1 en el día de la semana. K representa la progresión del día de la semana basada en el año. Suponiendo que cada año tiene 365 días, la misma fecha de cada año sucesivo será desplazada por un valor de $365 \bmod 7 = 1$.

Como hay 366 días en cada año bisiesto, esto se debe tener en cuenta añadiendo un día adicional al valor de desplazamiento del día de la semana. Esto se logra añadiendo $\lfloor \frac{K}{4} \rfloor$ al desplazamiento. Este término se calcula como un resultado entero. Cualquier resto que pueda haber es descartado.

Usando una lógica similar, se puede calcular la progresión del día de la semana para cada centuria observando que hay 36524 días en una centuria normal, y 36525 en cada centuria divisible por 400. Dado que $36525 \bmod 7 = 6$ y $36524 \bmod 7 = 5$, el término: $\lfloor \frac{J}{4} \rfloor - 2J$ refleja esto (de nuevo usando división entera y descartando cualquier resto fraccional). Para evitar los números negativos, este término se puede reemplazar por $5J + \lfloor \frac{J}{4} \rfloor$ con un resultado equivalente.

El término $\lfloor \frac{(m+1)26}{10} \rfloor$ se puede explicar de la siguiente manera. Zeller observó que, al iniciar cada año el 1 de marzo, el día de la semana de cada mes sucesivo progresaba multiplicando el mes por un valor constante y descartando el resto fraccional.

La función global, $\bmod 7$, normaliza el resultado para que se encuentre en el intervalo de 0 a 6, lo que da el índice del día de la semana correcto para la fecha analizada.

La implementación del algoritmo en C++

```
#include <iostream>

int Zeller(int dia, int mes, int ao){
    int h, K, J;

    if(mes<=2){
        mes += 12;
        ao -= 1;
    }
    else{
        mes-=2;
    }

    K = ao \% 100;
    J = ao / 100;

    h = ((700+((26*mes-2)/10)+dia+K+(K/4)+((J/4)+5*J))\%7);

    return h;
}
```

```

}

using namespace std;
int main(int argc, char *argv[]) {
int dia;
int ao;
int mes;

cout<<Zeller(dia, mes, ao);
}

```

11 Tarea 11

11.1 Momentos

El momento r de una variable aleatoria X alrededor de la media μ , también conocido como el momento central r , se define [11] como

$$\mu_r = E[(X - \mu)^r]$$

donde $r = 1, 2, \dots$. Se deduce que $\mu_0 = 1$, $\mu_1 = \mu$, $\mu_2 = \sigma^2$, es decir, el segundo momento central es la varianza. Se tiene que

$$\mu_r = \sum_{j=1}^n (x_j - \mu)^r f(x_j) = \sum (x - \mu)^r f(x) \text{ (variables discetas)}$$

$$\mu_r = \int_{-\infty}^{\infty} (x - \mu)^r f(x) dx \text{ (variables continuas)}$$

El momento de X alrededor del origen se define como

$$\mu'_r = E[X^r]$$

donde $r = 0, 1, 2, \dots$ y se pueden hallar fórmulas análogas a las anteriores donde $\mu = 0$.

La relación entre éstos momentos se expresa por

$$\mu_r = \mu'_r - \binom{r}{1} \mu'_r \mu + \dots (-1)^j \binom{r}{j} \mu'_{r-j} \mu^j + \dots + (-1)^r \mu'_0 \mu^r$$

11.2 Función generadora de momentos

La función generadora de momentos se define como

$$M_X(t) = E(e^{tX})$$

es decir

$$M_X(t) = \sum_{j=1}^n e^{tx_j} f(x_j) = \sum e^{tx} f(x)$$

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

Se calcula la expansión en serie de Taylor de la función generadora de momentos alrededor de 0:

$$\begin{aligned} M_X(t) &= E[e^{tX}] = E \left[1 + tX + \frac{t^2 X^2}{2} + \frac{t^3 X^3}{6} + \dots \right] \\ &= 1 + tE[X] + \frac{t^2}{2} E[X^2] + \frac{t^3}{6} E[X^3] + \dots \\ &= 1 + \mu'_1 t + \mu'_2 \frac{t^2}{2} + \mu'_3 \frac{t^3}{6} \end{aligned}$$

Se puede observar que los coeficientes de ésta expansión permite hallar los momentos (de ahí su nombre). Si una expansión de Taylor es de la forma

$$f(t) = \sum_{n=0} c_n (t-a)^n$$

entonces

$$c_n = \frac{1}{n!} \frac{d^n}{dt^n} f(t) \Big|_{t=a}$$

por lo que

$$\mu'_r = \frac{d^r}{dt^r} f(t) \Big|_{t=0}$$

es decir, μ_r es la r -ésima derivada de $M_X(t)$ evaluada en $t = 0$.

11.3 Sesgo

Con frecuencia una distribución no es simétrica con respecto a un máximo sino que tiene una cola más larga que la otra. Si la cola más larga se extiende a la derecha se dice que la distribución está sesgada a la derecha mientras que si la cola más larga se extiende a la izquierda se dice que está sesgada a la izquierda.

Las medidas que describen esta simetría se denominan coeficientes de sesgo o simplemente sesgo. Una de las medidas es:

$$\alpha_3 = \frac{E[x - \mu]^3}{\sigma^3}$$

la cual es una cantidad adimensional. La medida α_3 es positiva o negativa, si la distribución está sesgada a la derecha o a la izquierda, respectivamente, y si es igual a cero se dice que la distribución es simétrica. Otra fórmula empleada para calcular el sesgo es

$$g_1 = \frac{\sum_{i=1}^N \frac{(Y_i - \bar{Y})^3}{N}}{s^3}$$

donde \bar{Y} es la media, s es la desviación estándar y N es el número de datos de la muestra. Ésta fórmula se conoce como el coeficiente de Fisher-Pearson de sesgo.

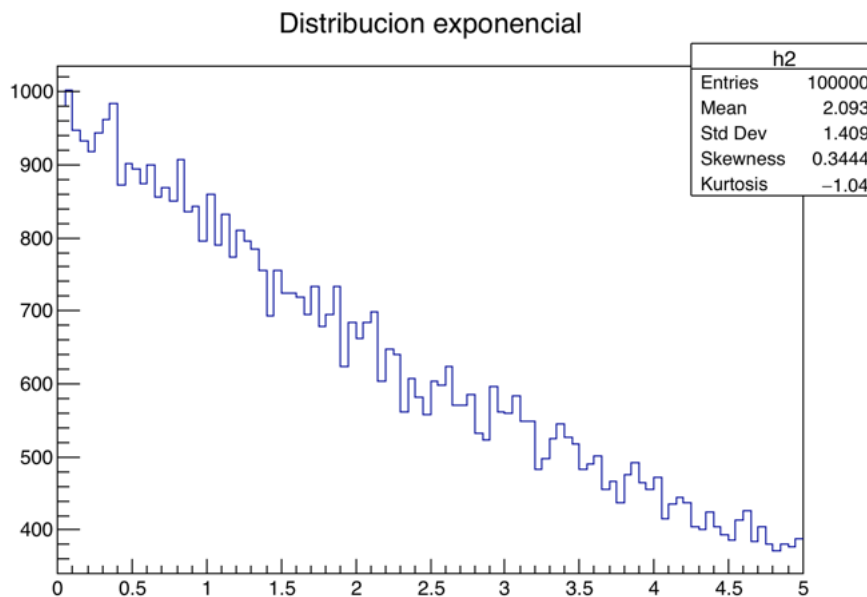


Figure 11: Curtosis y sesgo de un histograma de datos que siguen una distribución exponencial

11.4 Curtosis

En algunos casos las distribuciones pueden tener sus valores concentrados cerca de la media así que la distribución tiene un pico grande, como en la distribución normal. En otros casos, la distribución puede ser relativamente plana; en el caso anterior, el pico es menos alto por lo que los datos se mueven hacia las colas.

Medidas del grado de apuntamiento de una distribución se llaman coeficientes de curtosis o simplemente curtosis. Una medida empleada frecuentemente está dada por

$$\alpha_4 = \frac{E[(x - \mu)^4]}{\sigma^4} = \frac{\mu_4}{4}$$

que igualmente resulta en una cantidad adimensional. Otra fórmula para calcular el coeficiente de curtosis es

$$g_2 = \frac{\sum_{i=1}^N \frac{(Y_i - \bar{Y})^4}{N}}{s^4}$$

La curtosis de una distribución normal es 3, por lo que también es utilizada la siguiente fórmula:

$$g_2 = \frac{\sum_{i=1}^N \frac{(Y_i - \bar{Y})^4}{N}}{s^4} - 3$$

Esta definición se utiliza para que la distribución normal tenga curtosis cero. Además, un valor positivo indica una distribución de “cola pesada” mientras que un valor negativo indica una distribución de “cola ligera”.

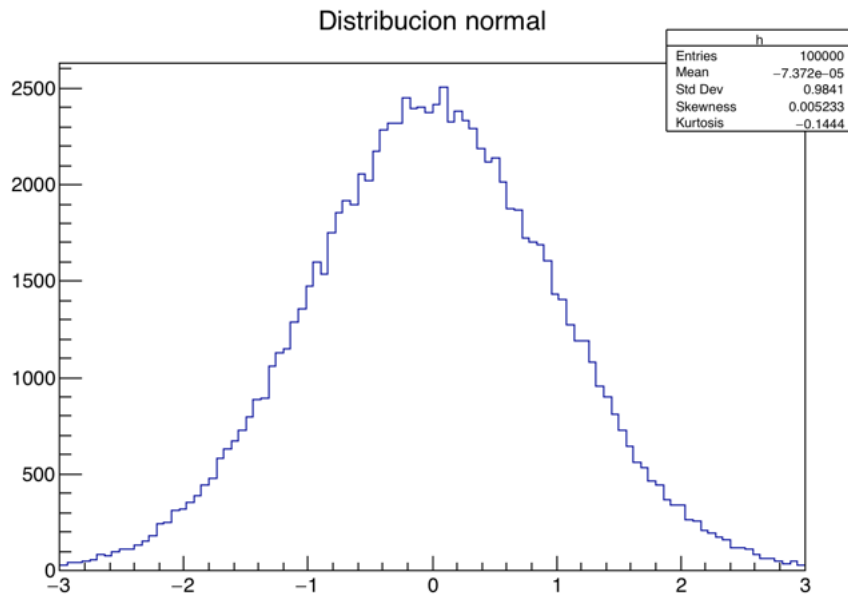


Figure 12: Curtosis y sesgo de un histograma de datos que siguen una distribución normal

12 Tarea 12

12.1 II

La belleza de los números irracionales puede ser observada desde distintos puntos de vista, ya sea en el carácter cotidiano que engloba, el marco finito en que se desarrollan a pesar de su carácter infinito, etc. El número irracional π demuestra ser un número con carácter especial para la comunidad intelectual. A continuación, se muestran algunas fórmulas que históricamente se ubicaron como las líderes en su tiempo; evidentemente, éstas fórmulas tienen una pesada parte matemática en relación a su demostración, las cuales omitimos pero hacemos referencia a [9].

El primer algoritmo como tal lo sugirió Arquímedes alrededor de 250 A.C. La idea consistía en considerar un círculo de diámetro unidad y polígonos circunscritos e inscritos de $3 \cdot 2^n$ lados. Denotando sus perímetros con a_n y b_n respectivamente, Arquímedes, con razonamientos puramente geométricos, demostró las relaciones

$$a_1 = 2\sqrt{3}, \quad b_1 = 3, \quad a_{n+1} = \frac{2a_nb_n}{a_n + b_n}, \quad b_{n+1} = \sqrt{a_{n+1}b_n}$$

de tal forma que $b_n < \pi < a_n$ y las sucesiones a_n y b_n convergen hacia π .

Desde éste desarrollo, se tuvo que esperar cerca de 1800 años para que en 1573 Viète obtuviera un desarrollo con radicales anidados infinitos:

$$\frac{\pi}{2} = \sqrt{\frac{1}{2}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}}} \cdots$$

El método de Viète se basa en considerar un círculo de diámetro 2 e ir aproximando su área sucesivamente por las áreas de un cuadrado inscrito, de un octágono inscrito, un polígono de 16 lados inscrito, etc. De esta manera Viète obtiene π con 9 cifras decimales: 3.141592653 [9].

Desde aquí, podemos surcar por las aguas del cálculo infinitesimal de Newton y Leibniz. Newton obtuvo la siguiente fórmula:

$$\pi = \frac{3\sqrt{3}}{4} + 2 - \frac{3}{4} \sum_{n=0}^{\infty} \frac{\binom{2n}{n}}{(n+1)(2n+5)16^n}$$

con la cual obtuvo [9] 15 decimales de π en 1665. Una serie similar a la de Newton pero más sencilla es

$$\frac{\pi}{3} = \sum_{n=0}^{\infty} \frac{\binom{2n}{n}}{(2n+1)16^n}$$

En 1706 John Machin demostró la fórmula que lleva su nombre

$$\frac{\pi}{4} = 4 \arctan \frac{1}{5} - \arctan \frac{1}{239}$$

Gauss también participó en la búsqueda de series para π de convergencia rápida y en 1863 demostró la siguiente identidad

$$\frac{\pi}{4} = 12 \arctan \frac{1}{18} + 8 \arctan \frac{1}{57} - 5 \arctan \frac{1}{239}$$

Gran parte de la obra del genio matemático hindú Ramanujan se relaciona con la teoría de las funciones modulares, que tiene su origen en el estudio de las integrales elípticas de primera y segunda especie.

En 1914 estas investigaciones le llevan a la obtención de extraordinarias series para $\frac{1}{\pi}$ totalmente diferentes de todas las conocidas hasta entonces y que son de la siguiente forma

$$\frac{1}{\pi} = \sum_{n=0}^{\infty} B_n z^n (bn + a)$$

siendo $b, a > 0$ y $-1 < z < 1$ números algebraicos, y que quedan clasificadas en 4 familias de acuerdo a las siguientes posibilidades de B_n

$$\frac{(2n)!^3}{n!^6} \quad \frac{(4n)!}{n!^4} \quad \frac{(2n)!(3n)!}{n!^5} \quad \frac{(6n)!}{(3n)!n!^3}$$

Actualmente, programas como Mathematica para calcular π utilizan el algoritmo de Chudnovsky. El algoritmo de Chudnovsky es un algoritmo creado por David Volfovich Chudnovsky y Gregory Volfovich Chudnovsky, hermanos y matemáticos ucranianos nacionalizados estadounidenses, mediante el cual se puede obtener muy buenas aproximaciones del π . Se basa en la siguiente expresión relacionada con el número π que encontró Ramanujan:

$$\frac{1}{\pi} = \frac{2\sqrt{2}}{9801} \sum_{k=0}^{\infty} \frac{(4k)!(1103 + 26390k)}{(k!)^4 396^{4k}}$$

La expresión del algoritmo de Chudnovsky es la siguiente:

$$\frac{1}{\pi} = 12 \sum_{k=0}^{\infty} \frac{(-1)^k (6k)!(13591409 + 545140134k)}{(3k)!(k!)^3 640320^{3k+3/2}}$$

y con ella se obtienen 14 decimales exactos más de π con cada término de la misma.

12.2 Algoritmo de Gauss-Legendre

Se parten de los siguientes dados iniciales:

$$a_0 = 1 \quad b_0 = \frac{1}{\sqrt{2}} \quad t_0 = \frac{1}{4} \quad p_0 = 1$$

A partir de aquí, se calculan lo siguiente:

$$\begin{aligned} A_{n+1} &= \frac{a_n + b_n}{2} \\ b_{n+1} &= \sqrt{a_n b_n} \\ t_{n+1} &= t_n - p_n (a_n - a_{n+1})^2 \\ p_{n+1} &= 2p_n \end{aligned}$$

Por lo que π se aproxima de la siguiente forma:

$$\pi \approx \frac{(a_n + b_n)^2}{4t_n}$$

El método tiene convergencia de segundo orden, es decir, en cada iteración se duplica el número de dígitos exactos obtenidos en la iteración anterior.

12.3 Algoritmo de Borwein

Se parte de los siguientes datos iniciales

$$a_0 = 6 - 4\sqrt{2}$$

$$y_0 = \sqrt{2} - 1$$

después se opera de la siguiente forma

$$y_{k+1} = \frac{1 - (1 - y_k^4)^{\frac{1}{4}}}{1 + (1 - y_k^4)^{\frac{1}{4}}}$$

$$a_{k+1} = a_k(1 + y_{k+1})^4 - 2^{2k+3}y_{k+1}(1 + y_{k+1} + y_{k+1}^2)$$

y se tiene lo siguiente

$$\lim_{k \rightarrow \infty} a_k = \frac{1}{\pi}$$

La convergencia de este método es cuártica. Es decir, en cada iteración se consiguen el cuádruple de dígitos exactos que en la iteración anterior.

12.4 Enteros largos

En C existen tipos de datos como char, short, int y long que pueden usarse para guardar números enteros de diferentes tamaos. Un problema de este tipo emerge al querer calcular números “grandes” como el factorial, por lo que las estructuras de datos que comprenden los lenguajes de manera estándar resultan insuficientes.

Una manera de verlos es por medio de un arreglo de chars que solo almacena dígitos. Extraído del libro Robert Lafore - Object-Oriented Programming in C++, 4a Edición:

```
// verylong.cpp
// implements very long integer type
#include "verylong.h" //header file for verylong
//-----
void verylong::putvl() const //display verylong
{
    char temp[SZ];
    strcpy(temp,vlstr); //make copy
    cout << strrev(temp); //reverse the copy
} //and display it
//-----
void verylong::getvl() //get verylong from user
{
    cin >> vlstr; //get string from user
    vlen = strlen(vlstr); //find its length
    strrev(vlstr); //reverse it
}
//-----
verylong verylong::operator + (const verylong v) //add verylongs
{
    char temp[SZ];
    int j;
```

```

        //find longest number
int maxlen = (vlen > v.vlen) ? vlen : v.vlen;
int carry = 0; //set to 1 if sum >= 10
for(j = 0; j<maxlen; j++) //for each position
{
    int d1 = (j > vlen-1) ? 0 : vlstr[j]-'0'; //get digit
    int d2 = (j > v.vlen-1) ? 0 : v.vlstr[j]-'0'; //get digit
    int digitsum = d1 + d2 + carry; //add digits
    if( digitsum >= 10 ) //if there's a carry,
        { digitsum -= 10; carry=1; } //decrease sum by 10,
    else //set carry to 1
        carry = 0; //otherwise carry is 0
    temp[j] = digitsum+'0'; //insert char in string
}
if(carry==1) //if carry at end,
    temp[j++] = '1'; //last digit is 1
temp[j] = '\0'; //terminate string
return verylong(temp); //return temp verylong
}

//-----
verylong verylong::operator * (const verylong v) //multiply
{ //verylongs
    verylong pprod; //product of one digit
    verylong tempsum; //running total
    for(int j=0; j<v.vlen; j++) //for each digit in arg
    {
        int digit = v.vlstr[j]-'0'; //get the digit
        pprod = multdigit(digit); //multiply this by digit
        for(int k=0; k<j; k++) //multiply result by
            pprod = mult10(pprod); // power of 10
        tempsum = tempsum + pprod; //add product to total
    }
    return tempsum; //return total of prods
}

//-----
verylong verylong::mult10(const verylong v) const //multiply
{ //arg by 10
    char temp[SZ];
    for(int j=v.vlen-1; j>=0; j--) //move digits one
        temp[j+1] = v.vlstr[j]; // position higher
    temp[0] = '0'; //put zero on low end
    temp[v.vlen+1] = '\0'; //terminate string
    return verylong(temp); //return result
}

//-----
verylong verylong::multdigit(const int d2) const

```

```

{
    char temp[SZ];
    int j, carry = 0;
    for(j = 0; j<vlen; j++)
    {
        int d1 = vlstr[j]-'0';
        int digitprod = d1 * d2;
        digitprod += carry;
        if( digitprod >= 10 )
        {
            carry = digitprod/10;
            digitprod -= carry*10;
        }
        else
            carry = 0;
        temp[j] = digitprod+'0';
    }
    if(carry != 0)
        temp[j++] = carry+'0';
    temp[j] = '\0';
    return verylong(temp);
}

//multiply this verylong
//by digit in argument

//for each position
//    in this verylong
//get digit from this
//multiply by that digit
//add old carry
//if there's a new carry,
//carry is high digit
//result is low digit

//otherwise carry is 0
//insert char in string

//if carry at end,
//it's last digit
//terminate string
//return verylong

/ verylong.h
// class specifier for very long integer type
#include <iostream>
#include <string.h>          //for strlen(), etc.
#include <stdlib.h>          //for ltoa()
using namespace std;

const int SZ = 1000;
    //maximum digits in verylongs

class verylong
{
private:
    char vlstr[SZ];          //verylong number, as a string
    int vlen;                //length of verylong string
    verylong multdigit(const int) const; //prototypes for
    verylong multi10(const verylong) const; //private functions
public:
    verylong() : vlen(0)     //no-arg constructor
    { vlstr[0]='\0'; }
    verylong(const char s[SZ]) //one-arg constructor
    { strcpy(vlstr, s); vlen=strlen(s); } //for string
    verylong(const unsigned long n) //one-arg constructor

```

```

    {
        ltoa(n, vlstr, 10);          //convert to string
        strrev(vlstr);               //reverse it
        vlen=strlen(vlstr);          //find length
    }
    void putvl() const;              //display verylong
    void getvl();                   //get verylong from user
    verylong operator + (const verylong); //add verylongs
    verylong operator * (const verylong); //multiply verylongs
};

```

12.5 Variable aleatoria

Una variable aleatoria X de un espacio muestral S es una función S en el conjunto \mathbb{R} de los números reales tal que la imagen inversa de cada intervalo de \mathbb{R} es un evento (o suceso de) de S .

Si S es un espacio discreto en el cual cada subconjunto es un suceso, entonces cada función de valores reales de S es una variable aleatoria. Por otra parte, se puede comprobar que si S es no contable, entonces ciertas funciones de valores reales de S no son variables aleatorias.

Si X y Y son variables aleatorias se sigue que:

$$(X + Y)(s) = X(s) + Y(s)$$

$$(X + k)(s) = X(s) + k$$

$$(kX)(s) = kX(s)$$

$$(XY)(s) = X(s)Y(s)$$

para todo $s \in S$.

Usamos la notación abreviada $P(X = a)$ y $P(a \leq X \leq b)$ para la probabilidad de los sucesos “ X toma al valor a ” y “ X toma valores en el intervalo $[a, b]$ ”. Significados análogos se dan a $P(X \leq a)$, $P(X = a, Y = b)$, etc.

12.5.1 Distribución y esperanza de una variable aleatoria finita

Sea X una variable aleatoria de un espacio muestral S con el conjunto imagen finito; a saber, $X(s) = |x_1, x_2, \dots, x_n|$. Convertimos $X(s)$ en un espacio de probabilidad definiendo la probabilidad de x_i como $P(X = x_i)$ que se escribe $f(x_i)$. Esta función f de $X(S)$, definida como $f(x_i) = P(X = x_i)$ se llama función de distribución o probabilidad de X y se expresa generalmente en forma de tabla:

x_1	x_2	\dots	x_n
$f(x_1)$	$f(x_2)$	\dots	$f(x_n)$

La dsitribución f satisface las siguientes condiciones:

- (i) $f(x_i) \geq 0$
- (ii) $\sum_{i=1}^n x_i f(x_i) = 1$

Ahora si X es una variable aleatoria con la distribución anterior, entonces la media o esperanza (o valor esperado) de X , denotado por $E[X]$ o μ_x se define como

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + \dots + x_n f(x_n) = \sum_{i=1}^n x_i f(x_i)$$

Esto es, $E(X)$ es el promedio ponderado de los valores posibles de X , cada valor ponderado por su probabilidad.

Los primeros teoremas en relación con la noción de valor esperado para operaciones de variables aleatorias son

Teorema 1. Sea X una variable aleatoria y k un número real. Entonces (i) $E(kX) = kE(X)$, y (ii) $E(X + k) = E(X) + k$.

Teorema 2. Sean X y Y variables aleatorias del mismo espacio muestral S . Entonces $E(X + Y) = E(X) + E(Y)$.

Un simple argumento de inducción conduce al siguiente resultado

Corolario 1. Sean X_1, X_2, \dots, X_n variables aleatorias de S . Luego

$$E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n)$$

12.5.2 Varianza y desviación estándar

La media de una variable aleatoria X mide, en cierto sentido, el valor “promedio” de X . El concepto de la varianza de X mide el “esparcimiento” o dispersión de X .

Sea X una variable aleatoria con la siguiente distribución: Entonces la varianza de X , denotada

x_1	x_2	\dots	x_n
$f(x_1)$	$f(x_2)$	\dots	$f(x_n)$

por $var(X)$ se define como

$$var(X) = \sum_{i=1}^n (x_i - \mu)^2 f(x_i) = E((X - \mu)^2)$$

donde μ es la media de X . La desviación estándar de X , denotada por σ_x es la raíz cuadrada (no negativa) de $var(X)$:

$$\sigma_x = \sqrt{var(X)}$$

El siguiente teorema muestra una alternativa a veces más útil para calcular la varianza de la variable aleatoria X .

Teorema 3.

$$var(X) = \sum_{i=1}^n x_i^2 f(x_i) - \mu^2 = E(X^2) - \mu^2$$

Ahora se establecen algunas propiedades de la varianza.

Teorema 4. Sea X una variable aleatoria y k un número rel. Entonces (i) $\text{var}(X + k) = \text{var}(X)$, y (ii) $\text{var}(kX) = k^2 \text{var}(X)$. Por lo tanto $\sigma_{X+k} = \sigma_X$ y $\sigma_{kX} = |k|\sigma_X$

12.5.3 Distribución conjunta

Sean X y Y variables aleatorias de un espacio muestral S con los respectivos conjuntos imagen

$$X(S) = \{x_1, x_2, \dots, x_n\} \quad Y(S) = \{y_1, y_2, \dots, y_m\}$$

Se calcula

$$X(S) \times Y(S) = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_m)\}$$

en un espacio de probabilidad definiendo la probabilidad de la perja ordenada (x_i, y_j) como $P(X = x_i, Y = y_j)$ que se escribe $h(x_i, y_j)$. Esta función h de $X(S) \times Y(S)$, se llama distribución conjunta o función de probabilidad conjunta de X y Y .

Las funciones

$$f(x_i) = \sum_{j=1}^m h(x_i, y_j) \quad g(y_j) = \sum_{i=1}^n h(x_i, y_j)$$

son llamadas distribuciones marginales y son, de hecho, distribuciones (individuales) de X y Y respectivamente. La distribución conjunta h satisface las siguientes condiciones:

- (i) $h(x_i, y_j) \geq 0$
- (ii) $\sum_{i=1}^n \sum_{j=1}^m h(x_i, y_j) = 1$

Ahora, si X y Y son variables aleatorias con la distribución conjunta anterior, con respectivas medias μ_X y μ_Y , entonces, la covarianza de X y Y denotada por $\text{cov}(X, Y)$ se define por

$$\text{cov}(X, Y) = \sum_{i,j} (x_i - \mu_X)(y_j - \mu_Y)h(x_i, y_j) = E[(X - \mu_X)(Y - \mu_Y)]$$

o equivalentemente como

$$\text{cov}(X, Y) = \sum_{i,j} x_i y_j h(x_i, y_j) - \mu_X \mu_Y = E(XY) - \mu_X \mu_Y$$

La correlación de X y Y , denotada por $\rho(X, Y)$ se define por

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

La correlación no tiene dimensión y tiene las siguientes propiedades

- (i) $\rho(X, Y) = \rho(Y, X)$
- (ii) $-1 \leq \rho \leq 1$
- (iii) $\rho(X, X) = 1, \rho(X, -X) = -1$

(iv) $\rho(aX + b, cY + d) = \rho(X, Y)$ si $a, c \neq 0$

La noción de una distribución conjunta de h se extiende a un número finito de variables aleatorias X, Y, \dots, Z de manera obvia, donde h es producto de $X(S) \times Y(S) \times \dots \times Z(S)$ definida por

$$h(x_i, y_j, \dots, z_k) = P(X = x_i, Y = y_j, \dots, Z = z_k)$$

12.5.4 Variables aleatorias independientes

Se dice que un número finito de variables aleatorias X, Y, \dots, Z de un espacio muestral S son independientes si

$$P(X = x_i, Y = y_j, \dots, Z = z_k) = P(X = x_i)P(Y = y_j) \cdots P(Z = z_k)$$

para valores x_i, y_j, \dots, z_k . En particular, X y Y son independientes si

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

Ahora, si X y Y tienen distribuciones f y g , respectivamente, y la distribución conjunta de h , entonces la ecuación anterior se puede escribir como

$$h(x_i, y_j) = f(x_i)g(y_j)$$

En otras palabras, X y Y son independientes si cada elemento $h(x_i, y_j)$ es el producto de sus elementos marginales.

Se establecen algunas propiedades importantes de variables aleatorias que no se cumplen en general

Teorema 5. Sean X y Y variables aleatorias independientes. Entonces

$$(i) E(XY) = E(X)E(Y)$$

$$(ii) \text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$$

$$(iii) \text{cov}(X, Y) = 0$$

De la parte (ii) del teorema anterior

Teorema 6. Sean X_1, X_2, \dots, X_n variables aleatorias independientes. Entonces

$$\text{var}(X_1 + \dots + X_n) = \text{var}(X_1) + \dots + \text{var}(X_n)$$

12.5.5 Funciones de una variable aleatoria

Sean X y Y variables aleatorias del mismo espacio muestral S . Entonces se dice que Y es una función de X si Y puede representarse por alguna función ϕ de valor real de una variable real $Y = \phi(X)$; esto es, si $Y(s) = \phi[X(s)]$ para todo $s \in S$.

Teorema 7. Sean X y Y variables aleatorias de un mismo espacio muestral S con $Y = \phi(X)$. Entonces

$$E(Y) = \sum_{i=1}^n \phi(x_i)f(x_i)$$

donde f es la función de distribución de X .

Similarmente, se dice que una variable aleatoria Z es una función de X y Y si Z se puede representar por $Z = \phi(X, Y)$ donde ϕ es una función de valor real de dos variables reales, esto es

$$Z(s) = \phi[X(s), Y(s)]$$

para todo $s \in S$. Conforme al teorema anterior

Teorema 8. Sean X, Y y Z variables aleatorias de un mismo espacio muestral S con $Z = \phi(X, Y)$. Entonces

$$E(Z) = \sum_{i,j} \phi(x_i, y_j) h(x_i, y_j)$$

donde h es la función de distribución conjunta de X y Y .

12.5.6 Variables aleatorias discretas en general

Ahora supongase que X es una variable aleatoria de S con un conjunto imagen infinito contable, o sea $X(S) = \{x_1, x_2, \dots\}$. Tales variables aleatorias son llamadas variables aleatorias discretas. Como en el caso finito, se construye $X(S)$ en un espacio de probabilidad definiendo la probabilidad de x_i como $f(x_i) = P(X = x_i)$ y llamamos f la distribución de x : El valor esperado $E(X)$ y la

x_1	x_2	x_3	\dots
$f(x_1)$	$f(x_2)$	$f(x_3)$	\dots

varianza $var(X)$ se definen como

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + \dots = \sum_{i=1}^{\infty} x_i f(x_i)$$

$$var(X) = (x_1 - \mu)^2 f(x_1) + (x_2 - \mu)^2 f(x_2) + \dots = \sum_{i=1}^{\infty} (x_i - \mu)^2 f(x_i)$$

cuando las series pertinentes convergen absolutamente. $var(X)$ existe si y sólo si $\mu = E(X)$ y $E(X^2)$ existeb ambos y la fórmula

$$var(X) = E(X^2) - \mu^2$$

es válida como en el caso finito. Cuando $var(X)$ existe,

$$\sigma_X = \sqrt{var(X)}$$

. Para la distribución conjunta, variables aleatorias independientes y funciones de variables aleatorias se extienden directamente del caso general. Si X y Y están definidas en el mismo espacio muestral S y si $var(X)$ y $var(Y)$ existen, entonces las series

$$cov(X, Y) = \sum_{i,j} (x_i - \mu_X)(y_j - \mu_Y) h(x_i, y_j)$$

convergen absolutamente y la relación

$$cov(X, Y) = E(XY) - \mu_X \mu_Y$$

se cumple igual que en el caso finito.

12.5.7 Variables aleatorias continuas

Supóngase que X es una variable aleatoria cuyo conjunto imagen $X(S)$ es un conjunto continuo de números tales como un intervalo. Recalcamos de la definición de variables aleatorias que el conjunto $\{a \leq X \leq b\}$ es un suceso de S y por consiguiente, la probabilidad $P(a \leq X \leq b)$ está bien definida. Suponemos que existe una función continua especial $f : \mathbb{R} \rightarrow \mathbb{R}$ tal que $P(a \leq X \leq b)$ es igual al área bajo la curva de f entre $x = a$ y $x = b$. De manera más formal,

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

En este caso se dice que X es una variable aleatoria continua. La función f se llama función de distribución o de probabilidad continua (o función de densidad) de X , que satisface las siguientes condiciones:

- (i) $f(x) \geq 0$
- (ii) $\int_{\mathbb{R}} f(x)dx = 1$

Es decir, f es no negativa y el área total bajo su curva es 1. El valor esperado $E(X)$ se define por

$$E(X) = \int_{\mathbb{R}} xf(x)dx$$

cuando existe. Las funciones de variable aleatorias, si $Y = \phi(X)$

$$E(Y) = \int_{\mathbb{R}} \phi(x)f(x)dx$$

cuando el miembro de la derecha existe. La varianza $var(x)$ se define por

$$var(X) = E((X - \mu)^2) = \int_{\mathbb{R}} (x - \mu)^2 f(x)dx$$

cuando existe. $var(X)$ existe si y sólo si μ , $E(X)$ y $E(X^2)$, y por tanto

$$var(X) = E(X^2) - \mu^2 = \int_{\mathbb{R}} x^2 f(x)dx - \mu^2$$

La desviación estándar σ_X se define por $\sigma_X = \sqrt{var(X)}$ cuando $var(X)$ existe.

Un número finito de variables aleatorias continuas, a saber X, Y, \dots, Z se dice que son independientes si para unos intervalos $[a, a']$, $[b, b']$, \dots , $[c, c']$.

$$P(a \leq x \leq a', b \leq Y \leq b', \dots, c \leq Z \leq c') = P(a \leq x \leq a')P(b \leq Y \leq b') \cdots P(c \leq Z \leq c')$$

12.5.8 Función de distribución acumulada

Sea X una variable aleatoria (discreta o continua). La función de distribución acumulada F de X es la función $F : \mathbb{R} \rightarrow \mathbb{R}$ definida por

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

Si X es una variable aleatoria discreta de distribución f , entonces F es la función escalonada definida por

$$F(x) = \sum_{x_i \leq x} f(x_i)$$

Por otra parte, si X es una variable aleatoria continua de distribución f , entonces

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

En ambos casos, F es monótona creciente, esto es $F(a) \leq F(b)$ siempre que $a \leq b$ y el límite de F a la izquierda es 0 y a la derecha es 1.

12.5.9 Desigualdad de Tchebycheff y Ley de los Grandes Números

Teorema 9. (*Desigualdad de Tchebycheff*): Sea X una variable aleatoria con promedio μ y desviación estándar σ . Entonces para cada $\epsilon > 0$

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

Teorema 10. (*Ley de los grandes números*) Sea X_1, X_2, \dots es una sucesión de variables aleatorias independientes con la misma distribución con promedio μ y varianza σ^2 . Sea

$$\bar{S}_n = (X_1 + X_2 + \dots + X_n)/n$$

Entonces para $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(|\bar{S}_n - \mu| \geq \epsilon) = 0$$

o bien

$$\lim_{n \rightarrow \infty} P(|\bar{S}_n - \mu| < \epsilon) = 1$$

13 Tarea 13

13.1 Is Economics the Next Physical Science?

En los últimos años, se ha observado una oleada de físicos estudiando temas que son del interés para los economistas, en una nueva área denominada econofísica. Pero, ¿en qué difiere el enfoque del economista al del físico?

Los físicos se han visto envueltos en multitud de áreas. Daniel Bernoulli introdujo en 1738 la idea de la utilidad de describir las preferencias de las personas.

El rango de temas que abordan (de la parte económica) va desde la parte financiera, la observación empírica de regularidades en los datos del mercado, la dinámica de la información de los precios, la comprensión de las burbujas y los pánicos, los métodos para la fijación de precios las opciones y otros derivados, y la construcción de carteras. Temas más amplios en la economía incluyen la

distribución del ingreso, la aparición del dinero, y las implicaciones de la simetría y la escala para el funcionamiento del mercado.

El desarrollo de las sociedades y las economías puede estar supeditada a los accidentes de la historia y gira en torno a los aspectos complejos de la conducta humana.

Mercados y otras instituciones económicas traen consigo los conceptos de eficiencia y optimización de los deseos humanos, algo que en sí resulta bastante complejo y subjetivo.

Por un lado, la física se dirige mediante la búsqueda de leyes universales. Como se mencionó anteriormente, las relaciones y el medio social tan complejos, suele verse ésta búsqueda como algo imposible o difícil de modelar.

Gran parte del trabajo de los físicos trata de leyes de potencias. Una ley de potencia es una relación asintótica de la forma

$$f(x) \rightarrow x^{-\alpha}$$

donde la variable $\alpha > 0$ es una constante. La existencia de leyes de potencia en los cambios de precios es interesante desde un punto de vista práctico, ya que tiene implicaciones para el riesgo de las inversiones financieras, y desde un punto de vista teórico, ya que sugiere una escala de independencia y posibles analogías con fenómenos críticos y el comportamiento de no equilibrio en los procesos que generan rendimientos financieros.

Una propiedad interesante y sorprendente de la caminata aleatoria en precios es que la velocidad de difusión no es constante. El tamaño de la variación de los precios en un momento t se correlaciona con que en $\tau + t$ a pesar de que las direcciones de los cambios en los precios no están correlacionadas: Este fenómeno se denomina volatilidad en clúster. La relación decae como una ley de potencia de la forma $\tau^{-\gamma}$.

Los físicos han descubierto que la volatilidad es sólo uno de varios procesos de memoria larga en los mercados. La oferta y demanda es uno de éstos fenómenos. Si uno asigna +1 a una orden de compra y -1 a una orden de venta, la serie resultante de números tiene una correlación positiva que se desintegra como una ley de potencia con un exponente de alrededor de 0.6. La correlación positiva persiste en niveles estadísticamente significativas con respecto a decenas de miles de pedidos, por semanas.

Una gran cantidad de trabajos empíricos utiliza métodos y analogías de física. Por ejemplo, la teoría de matrices aleatorias, desarrollada en física nuclear, y las correlaciones ultramétricas han demostrado ser útiles para la comprensión de la correlación entre los movimientos de los precios de diferentes compañías.

La diferencia fundamental entre un sistema físico y uno económico es que está habitado por personas, que tienen interacciones estratégicas. La gente piensa y planea para después tomar decisiones basadas en sus estrategias, lo que resulta más complicado que tratar con átomos. Como consecuencia de ello, las técnicas matemáticas y la filosofía de modelado en la economía se han distanciando de los de la física. A pesar de esta divergencia es claramente necesaria, muchos í podrían argumentar que la brecha es más amplia de lo que debería ser.

Como se puede observar, el área está lejos de darse por terminada. El trabajo cada vez se formaliza en programas de estudio y muchos artículos son presentados en nuevas revistas dedicadas al área.

La manera de aproximarse puede no ser del todo correcta, pero poco a poco se va cambiando el paradigma de investigación.

13.2 Función de distribución de probabilidad

En el caso discreto, que la probabilidad de que Y tome el valor y , $P(Y = y)$, se define como la suma de las probabilidades de todos los puntos muestrales en S a los que se asigna el valor y . A veces se denota $P(Y = y)$ por $p(y)$.

Como $p(y)$ es una función que asigna probabilidades a cada valor y de las variables aleatoria Y , a veces recibe el nombre de función de probabilidad para Y .

Nótese que $p(y) \geq 0$ para toda y , pero la distribución de probabilidad para una variable aleatoria discreta asigna probabilidades diferentes de cero a sólo un número contable de valores y distintos. Se entiende que cualquier valor y no explícitamente asignado a una posibilidad positiva es tal que $p(y) = 0$

Para cualquier distribución de probabilidad discreta, se cumple lo siguiente

1. $0 \leq p(y) \leq 1$ para toda y
2. $\sum_y p(y) = 1$ donde la sumatoria es para todos los valores de y con probabilidad diferente de cero.

Si la variable aleatoria es continua, hay infinitos valores posibles de la variable y entre cada dos de ellos se podrían definir infinitos valores. En estas condiciones no es posible deducir la probabilidad de un valor puntual de la variable como se puede hacer en el caso de las variables discretas. Pero sí es posible calcular la probabilidad acumulada hasta un cierto valor (función de distribución) y cómo cambia esa probabilidad acumulada en cada punto (densidad de probabilidad). Por tanto, cuando la variable aleatoria sea continua hablaremos de función de densidad.

Sea X una variable aleatoria continua, se llama función de densidad y se representa como $f(x)$ a una función no negativa definida sobre la recta real, tal que para cualquier intervalo que estudiemos se verifica:

$$P(X \in A) = \int f(x)dx$$

Si $F(y)$ es una función de distribución entonces

1. $F(-\infty) \equiv \lim_{y \rightarrow -\infty} F(y) = 0$
2. $F(\infty) \equiv \lim_{y \rightarrow \infty} F(y) = 1$
3. $F(y)$ es una función no decreciente de y .

14 Tarea 14

14.1 Análisis ROC

Una curva ROC (Receiver Operating Characteristic o Característica Receptiva Operativa) es una representación gráfica que ilustra el desempeño de un sistema de clasificador binario con cambios en su límite de discriminación. La curva es creada graficando la razón positiva verdadera contra la razón del falso positivo.

La sensibilidad y la especificidad son medidas estadísticas del desempeño de una prueba de clasificación binaria. La sensibilidad (o razón del verdadero positivo) mide la proporción de resultados positivos que son correctamente identificados como tales. La especificidad (o razón del verdadero negativo) mide la proporción de resultados negativos que son correctamente identificados como tales. De manera más concreta, la sensibilidad cuantifica los falsos negativos evitados mientras que la especificidad los falsos positivos.

En general, si se conocen las distribuciones de probabilidad para la detección y de falsas alarmas, la curva ROC se puede generar mediante el trazado de la función de distribución acumulativa (área bajo la distribución de probabilidad de $-\infty$ al límite de discriminación) de la probabilidad de detección en el eje y frente a la función de distribución acumulativa de la probabilidad de falsa alarma en el eje x .

La sensibilidad se define como [12]:

$$\frac{\text{Número de positivos verdaderos}}{\text{Número casos positivos}}$$

y la especificidad:

$$\frac{\text{Número de falsos verdaderos}}{\text{Número de casos falsos}}$$

La precisión en éste caso, se relaciona con la fracción de la población que fue catalogada de manera correcta, es decir

$$\text{Precisión} = [\text{Sensibilidad}] \cdot [\text{Fracción de la población del estudio que es positiva}] + [\text{Especificidad}] \cdot [\text{Fracción de la población del estudio que es negativa}]$$

Anteriormente mencionamos los conceptos de fracción del verdadero positivo (TPF) y la fracción del falso positivo (TNF). De manera alterna, se tiene el concepto inverso de éstas razones: la fracción del falso positivo (FPF) y la fracción del falso negativo (FNF), que miden la razón de los casos positivos y negativos, respectivamente de los casos positivos y negativos que son asignados de manera incorrecta. De manera sencilla se puede observar que:

$$\text{TPF} + \text{FNF} = 1$$

y

$$\text{TNF} + \text{FPF} = 1$$

Es común denotar las razones anteriores usando probabilidad condicional ya que cada asignación (la fracción) representa una estimación de la probabilidad de una asignación en particular, pero con la restricción (o condición) que un caso de manera individual tiene un cierto valor binario.

Si D representa la decisión, \bar{D} representa la decisión contraria. Sea T el resultado de la asignación. Se tiene los siguientes casos:

$$\text{FPF} = P(T|\bar{D})$$

$$\text{TPF} = P(T|D)$$

$$\text{FNF} = P(\bar{T}|D)$$

$$\text{TNF} = P(\bar{T}|\bar{D})$$

Observemos que éste caso puede llevar a la paradoja del falso positivo que se trató con anterioridad.

Para cualquier prueba binaria, la información obtenida de las pruebas no necesariamente caen, dentro de una las dos categorías definidas por el usuario. Por ejemplo, [12], ejemplifica el razonamiento con un ejemplo con resultados de diagnósticos como ensayos de suero en la sangre, donde la distribución de los valores de los resultados en positivo y negativo se traslapa. En pocas palabras,

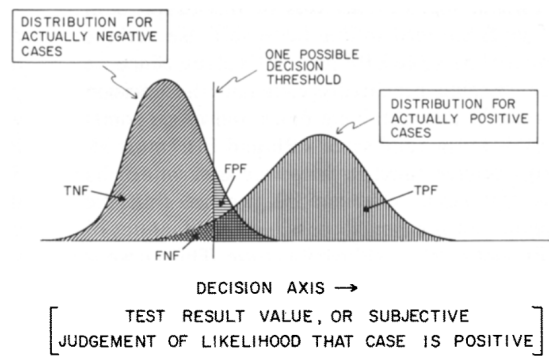


Figure 13: Límite para la prueba

elegir cual es el parámetro para separar un caso del otro se elige de manera arbitraria, pero trae consigo problemas al momento de aplicar las pruebas ya que arroja resultados diferentes. Para resolver este dilema, de manera forzada se varía el límite de aceptación, observando los cambios que ocasiona en las razones de decisión.

Si se cambia explícitamente el umbral de decisión varias veces como se describe anteriormente, obtendremos varios pares diferentes de TPF y FPF. Estos pares se pueden representar como la coordenada y y x de puntos en un gráfico. Los ejes de este gráfico van de cero a uno, porque son los límites de posibles valores para el FPF TPF.

Como se puede variar repetidamente el umbral de decisión y la obtención de más y más puntos en este gráfico, y como el TPF y el FPF deben cambiar de manera conjunta determinada por las distribuciones de los resultados de prueba, se observa que los puntos que representan todas las combinaciones posibles de TPF y FPF tienen que estar en una curva.

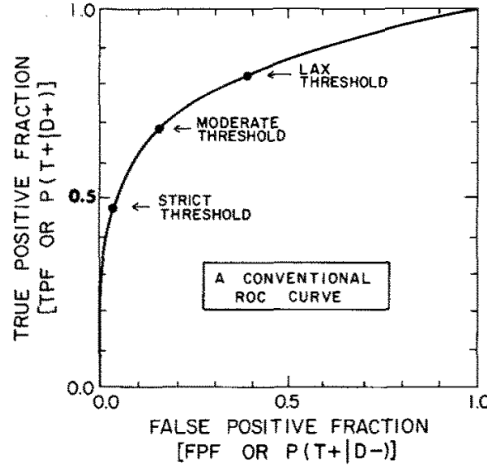


Figure 14: Una curva ROC convencional mostrando tres posibles puntos operativos

Los puntos anteriores muestran la flexibilidad que se tiene al catalogar los resultados: una actitud stricta (por ejemplo, positivo con 99.9% de certeza), moderado (90% de certeza) o laxo (se llama positivo bajo cualquier indicio que lleve a sospechar que es positivo).

15 Tarea 15

15.1 Algunas demostraciones para valores medios

Propiedad 1.

$$\overline{\Delta u} = \overline{u - \bar{u}} = \bar{u} - \bar{u} = 0$$

Demostración.

Por definición, $\Delta u = u - \bar{u}$ y

$$\bar{u} = P(u_i)u_i \quad (9)$$

Sustituyendo en Eq. (9) se obtiene

$$\begin{aligned} \sum_{y \in U} y P(u_i - \bar{u} = y) &= \sum (u_i - \bar{u}) P(X = u_i) \\ &= \sum u_i P(X = u_i) - \bar{u} \sum P(X = u_i) \\ &= \bar{u} - \bar{u} \\ &= 0 \end{aligned} \quad (10)$$

Propiedad 2.

$$\overline{(\Delta u)^2} \equiv \sum P(u_i)(u_i - \bar{u})^2 \geq 0$$

Demostración:

Por definición, $\Delta u = u - \bar{u}$. Elevando al cuadrado, se obtiene que

$$(\Delta u)^2 = (u - \bar{u})^2 \geq 0 \quad (11)$$

para cualquier valor. Por otro lado, si se calcula el valor medio de $(\Delta u)^2$ se obtiene que, junto con el resultado de la Eq.(11)

$$(\Delta u)^2 = \sum P(u_i)(u_i - \bar{u})^2 \geq 0 \quad (12)$$

ya que $0 \leq P(x) \leq 1$.

Propiedad 3.

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

Demostración: La idea consiste en desarrollar el cuadrado adentro de la sumatoria

$$\overline{(u - \bar{u})^2} = \sum P(u_i)(u_i - \bar{u})^2 \quad (13)$$

$$= \sum (u_i)(u_i^2 - 2\bar{u}u_i + \bar{u}^2) \quad (14)$$

$$= \sum P(u_i)u_i^2 - 2\bar{u} \sum P(u_i)u_i + \bar{u}^2 \sum P(u_i) \quad (15)$$

$$= \sum P(u_i)u_i^2 - \bar{u}^2 \quad (16)$$

$$= \overline{u^2} - \bar{u}^2 \quad (17)$$

Propiedad 4.

$$\overline{u^2} \geq \bar{u}^2$$

Demostración:

De la Eq.(11) y el resultado de la propiedad 3 se sigue que

$$\overline{u^2} \geq \bar{u}^2$$

15.2 Una demostración útil

Se tiene que

$$p^2 \left(\frac{\partial}{\partial p} \right)^2 (p^{n_1}) = p^2 n_1^2 p^{n_1-2} \quad (18)$$

$$= n_1^2 p^{n_1} \quad (19)$$

15.3 Aplicaciones de caminatas aleatorias

Uno de los ejemplos más comunes para la caminata aleatoria es en el área de finanzas, que pueden ser vistas como juegos de azar. El objetivo es calcular la posición S_n al tiempo n del precio de una acción. Sin embargo, se puede adoptar un enfoque distinto para la caminata aleatoria, debido

que la magnitud de los cambios se considera aproximadamente proporcional a la magnitud del cambio.

Para reflejar ésta visión, se considera que el precio en el periodo n es producto del precio en $n - 1$ y un multiplicador aleatorio Y_n , es decir,

$$Z_n = Z_{n-1} * Y_n$$

que implica que

$$Z_n = Z_0(Y_1 \cdot Y_2 \cdot \dots \cdot Y_n) \quad n \geq 1$$

Si se asume que los multiplicadores Y_n son independiente e idénticamente distribuidos se tiene que

$$\log(Z_n) = \log(Z_0) + S_n$$

donde S_n es una caminata aleatoria con pasos $X_n \equiv \log(Y_n)$.

15.4 Caminata aleatoria

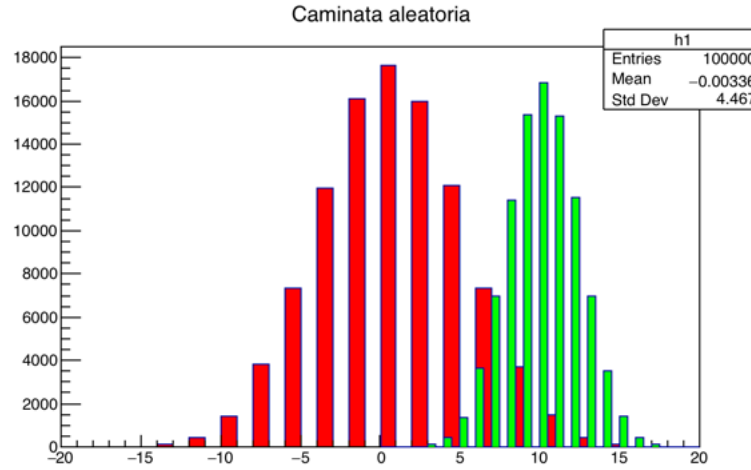


Figure 15: Caminata aleatoria en una dimensión para 20 pasos con $p = q = 0.5$, en rojo la posición final, en verde el número de pasos que se dan a la derecha

16 Tarea 16

16.1 Función Gamma

La función Gamma fue definida por Euler mediante

$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt \quad , \quad x > 0 \quad (20)$$

donde la condición $x > 0$ es exigida para la convergencia de la integral.

Se tiene que $e^{-t}t^{x-1} \rightarrow 0$ cuando $x \rightarrow \infty$, de manera que no se esperan inconvenientes en el límite superior de la integral. Cerca del límite inferior $t = 0$, el integrando se aproxima a t^{x-1} ya que $e^{-t} \simeq 1$. De esta manera

$$\begin{aligned}\Gamma(x) &= \int_0^c t^{x-1} dt + \int_c^\infty e^{-t} t^{x-1} dt \\ &= \left[\frac{t^x}{x} \right]_0^c + \int_c^\infty e^{-t} t^{x-1} dt\end{aligned}$$

y para que el primer término de la derecha permanezca finito, se debe tener $x > 0$.

De la Eq.(20) se puede observar que

$$\Gamma(1) = \int_0^\infty e^{-t} dt = 1$$

y

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty e^{-t} t^{-1/2} dt = 2 \int_0^\infty e^{-u^2} du = 2 \left(\frac{\sqrt{\pi}}{2} \right) = \sqrt{\pi}$$

Una relación básica de la función Gamma es

$$\Gamma(x+1) = x\Gamma(x) \tag{21}$$

la cual se deduce a partir de Eq.(20) e integrando por partes, como sigue:

$$\begin{aligned}\Gamma(x+1) &= \int_0^\infty e^{-t} t^{x-1} dt \\ &= [-e^{-t} t^x]_0^\infty + x \int_0^\infty e^{-t} t^{x-1} dt \\ &= x\Gamma(x)\end{aligned}$$

Si se toma $x = n$ (n entero positivo) y se usa Eq.(21) repetidamente, se tiene que

$$\begin{aligned}\Gamma(n+1) &= n\Gamma(n) \\ &= n(n-1)\Gamma(n-2) \\ &\vdots \\ &= n(n-1)(n-2) \dots 1 \cdot \Gamma(1)\end{aligned}$$

es decir

$$\Gamma(n+1) = n! \tag{22}$$

Esta última expresión puede usarse para definir $0!$, si se aplica para $n = 0$, obteniéndose

$$0! = \Gamma(1) = 1$$

16.2 Otras demostraciones para distribuciones de probabilidad para N grande

Ecuación 1. *Expandiendo $\ln W$ en una serie de Taylor se tiene que*

$$\ln W(n_1) = \ln W(\tilde{n}_1) + \beta_1 \eta + \frac{1}{2} \beta_2 \eta^2 + \frac{1}{6} \beta_3 \eta^3 + \dots$$

Como es de interés saber el comportamiento alrededor de $n_1 = \tilde{n}_1$, se usa

$$n_1 \equiv \tilde{n}_1 + \eta$$

Recordemos que una serie de Taylor se obtiene de la siguiente manera

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

Como se busca evaluar la función $\ln(a)$ alrededor de $a = n_1 \equiv \tilde{n}_1 + \eta$ obteniendo

$$\ln W(n_1) = \ln W(\tilde{n}_1) + \frac{d \ln W(\tilde{n})}{dn_1} \frac{(n_1 - \tilde{n}_1)}{1!} + \frac{d^2 \ln W(\tilde{n})}{dn_1^2} \frac{(n_1 - \tilde{n}_1)^2}{2!} + \dots$$

Si se sustituye $\eta = n_1 - \tilde{n}_1$ y sea $B = \frac{d^k \ln W(\tilde{n}_1)}{dn_1^k}$, donde B es la k -ésima derivada evaluada en \tilde{n} , donde $k \geq 1$ se obtiene:

$$\ln W(n_1) = \ln W(\tilde{n}_1) + \beta_1 \eta + \frac{1}{2} \beta_2 \eta^2 + \frac{1}{6} \beta_3 \eta^3 + \dots$$

Ecuación 2.

$$\ln W(n_1) = \ln N! - \ln n_1! \ln(N - n_1)! + n_1 \ln p + (N - n_1) \ln q$$

De la ecuación de la distribución binomial se tiene que

$$W(n_1) = \frac{N!}{n_1!(N - n_1)!} p^{n_1} q^{N - n_1}$$

Aplicando logaritmos en ambos lados de la ecuación

$$\ln(W(n_1)) = \ln\left(\frac{N!}{n_1!(N - n_1)!} p^{n_1} q^{N - n_1}\right)$$

desarrollando la expresión anterior

$$\begin{aligned} \ln(W(n_1)) &= \ln\left(\frac{N!}{n_1!(N - n_1)!}\right) + \ln(p^{n_1}) + \ln(q^{N - n_1}) \\ &= \ln(N!) - \ln(n_1!) - \ln(N - n_1)! + \ln(p^{n_1}) + \ln(q^{N - n_1}) \\ &= \ln(N!) - \ln(n_1!) - \ln(N - n_1)! + n_1 \ln(p) + (N - n_1) \ln(q) \end{aligned}$$

Ecuación 3.

$$\frac{d \ln n}{dn} \approx \ln n$$

Considerando $n \gg 1$, al aplicar la fórmula de Stirling

$$\ln n! \approx n \ln n - n$$

Derivando ambos lados de la ecuación

$$\begin{aligned} \frac{d \ln n}{dn} &\approx \frac{d(n \ln n - n)}{dn} \\ &\approx \frac{n}{n} + \ln n - 1 \\ &\approx \ln n \end{aligned}$$

Ecuación 4.

$$\frac{d \ln W}{dn_1} = -\ln n_1 + \ln(N - n_1) + \ln p - \ln q$$

Dada la ecuación

$$\ln W(n_1) = \ln(N!) - \ln(n_1!) - \ln(N - n_1)! + n_1 \ln(p) + (N - n_1) \ln(q)$$

derivando con respecto a n_1

$$\begin{aligned} \frac{\ln W(n_1)}{n_1} &= \frac{\ln(N!) - \ln(n_1!) - \ln(N - n_1)! + n_1 \ln(p) + (N - n_1) \ln(q)}{dn_1} \\ &= \frac{\ln(N!)}{dn_1} - \frac{\ln(n_1!)}{dn_1} - \frac{\ln(N - n_1)!}{dn_1} + \frac{n_1 \ln(p)}{dn_1} + \frac{(N - n_1) \ln(q)}{dn_1} \end{aligned}$$

Pero, se puede observar que

$$\begin{aligned} \frac{\ln(N!)}{dn_1} &= 0 \\ \frac{n_1 \ln(p)}{dn_1} &= \ln p \\ \frac{(N - n_1) \ln(q)}{dn_1} &= -\ln q \end{aligned}$$

además

$$\frac{d \ln n!}{dn} \approx \ln n!$$

por lo tanto

$$\begin{aligned} \frac{d \ln n!}{dn} &\approx \ln n! \\ \frac{d \ln(N - n_1)!}{dn} &\approx \ln(N - n_1)! \end{aligned}$$

por lo que finalmente se obtiene

$$\frac{d \ln W}{dn_1} = -\ln n_1 + \ln(N - n_1) + \ln p - \ln q$$

Ecuación 5.

$$\ln \left(\frac{N - \tilde{n}_1}{\tilde{n}_1} \cdot \frac{p}{q} \right) = 0$$

Igualando la expresión anterior a 0 donde $n_1 = \tilde{n}_1$, lo que implica que W es el máximo

$$\begin{aligned} \frac{d \ln W}{dn_1} &= -\ln \tilde{n}_1 + \ln(N - \tilde{n}_1) + \ln p - \ln q = 0 \\ \ln(N - \tilde{n}_1) - \ln(\tilde{n}_1) + \ln p - \ln q &= 0 \\ \ln \left(\frac{N - \tilde{n}_1}{\tilde{n}_1} \right) + \ln \left(\frac{p}{q} \right) &= 0 \\ \ln \left(\frac{N - \tilde{n}_1}{\tilde{n}_1} \cdot \frac{p}{q} \right) &= 0 \end{aligned}$$

Ecuación 6.

$$(N - \tilde{n}_1)p = \tilde{n}_1 q$$

De la ecuación anterior

$$\begin{aligned} \ln \left(\frac{N - \tilde{n}_1}{\tilde{n}_1} \cdot \frac{p}{q} \right) &= 0 \\ e^{\ln \left(\frac{N - \tilde{n}_1}{\tilde{n}_1} \cdot \frac{p}{q} \right)} &= e^0 \\ \frac{N - \tilde{n}_1}{\tilde{n}_1} \cdot \frac{p}{q} &= 1 \\ (N - \tilde{n}_1) \cdot p &= \tilde{n}_1 q \end{aligned}$$

Ecuación 7.

$$Np = \tilde{n}_1$$

De la ecuación anterior

$$\begin{aligned} (N - \tilde{n}_1) \cdot p &= \tilde{n}_1 q \\ Np - \tilde{n}_1 p &= \tilde{n}_1 q \\ Np &= \tilde{n}_1 q + \tilde{n}_1 p \\ Np &= \tilde{n}_1 (q + p) \\ Np &= \tilde{n}_1 \end{aligned}$$

donde $p + q = 1$.

Ecuación 8.

$$\frac{d^2 \ln W}{dn_1^2} = -\frac{1}{n_1} - \frac{1}{N - n_1}$$

De la primera derivada en la Eq. 4

$$\frac{d \ln W}{dn_1} = -\ln n_1 + \ln(N - n_1) + \ln p - \ln q$$

volvemos a derivar,

$$\begin{aligned}\frac{d}{dn_1} \left(\frac{d \ln W}{dn_1} \right) &= \frac{d}{dn_1} (-\ln n_1 + \ln(N - n_1) + \ln p - \ln q) \\ &= \frac{d(-\ln n_1)}{dn_1} + \frac{d \ln(N - n_1)}{dn_1} + \frac{d \ln p}{dn_1} - \frac{d \ln q}{dn_1}\end{aligned}$$

y se obtiene

$$\frac{d^2 \ln W}{dn_1^2} = -\frac{1}{n_1} - \frac{1}{N - n_1}$$

Ecuación 9.

$$B_2 = -\frac{1}{Npq}$$

De la ecuación 7 y 8,

$$\begin{aligned}B_2 &= -\frac{1}{Np} - \frac{1}{N - Np} \\ &= -\frac{1}{Np} - \frac{1}{N(1 - p)} \\ &= -\frac{1}{Np} - \frac{1}{Nq} \\ &= -\frac{1}{N} \left(\frac{1}{p} + \frac{1}{q} \right) \\ &= -\frac{1}{N} \frac{q + p}{pq} \\ &= -\frac{1}{Npq}\end{aligned}$$

16.3 Simulación de Movimiento Browniano

17 Tarea 17

17.1 Función delta de Dirac

La función delta de Dirac esta directamente inspirada en el delta de Kronecker que se define como

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

en donde i y j son números enteros. Lo que se quiere es tomar el concepto del delta de Kronecker haciéndolo extensivo para subíndices que representan cantidades continuas que en principio pueden tomar cualquier valor que forme parte del conjunto de los números reales. Con esto en mente, se define a la función delta de Dirac como una función límite.

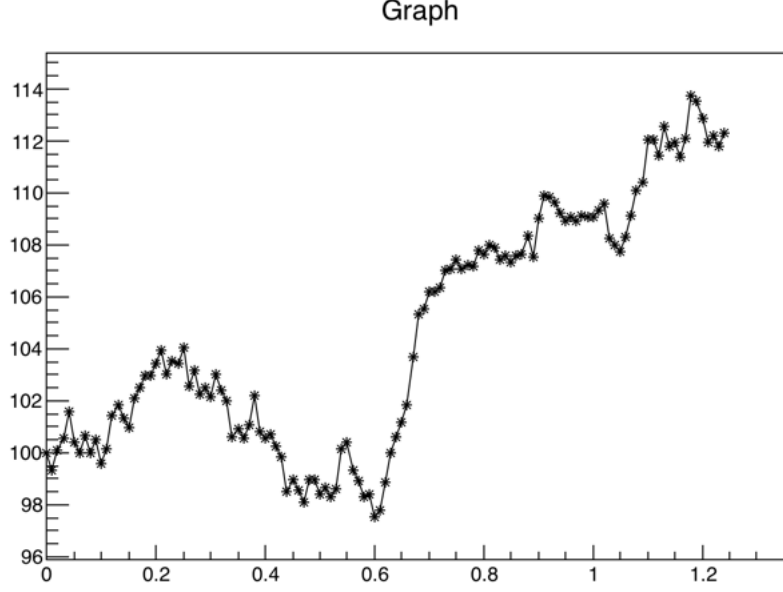


Figure 16: Simulación de movimiento browniano de la forma $dS_t = S_t \mu dt + S_t \sigma dz$ donde $dz = \epsilon \sqrt{\Delta t}$ y $\epsilon \sim N(0, 1)$. Los parámetros son $S_0 = 100$, $\sigma = 0.07$, $\mu = 0.1$ y $\Delta t = 0.01$ con $n = 200$ pasos de tiempo.

El área de la figura 17, colocada de modo tal que está centrada en la coordenada $x = x^*$, es igual a la unidad, o sea que $A = (a)(1/a) = 1$. Si se recorta cada vez más la base de la figura, de modo que se mantiene el valor de su área constante para cada iteración, se necesita que la altura se incremente bajo el mismo factor. Al considerar el límite de la anchura cuando tiende a cero, se puede hacer que la altura sea infinita y que tenga la misma área, como se puede observar en el proceso descrito, lo cual puede llegar a ser contradictorio.

En base al proceso extremo que se ha llevado a cabo, es tentador definir a esta función de la manera siguiente situando el “pico” en una posición $x = a$ en donde a puede ser igual a cero (situando el “pico” en el centro del sistema de coordenadas):

$$\begin{cases} \delta(x - a) = \infty & \text{si } x = a \\ \delta(x - a) = 0 & \text{si } x \neq a \end{cases}$$

Una forma de ampliar la definición de la función extrema sería basarse en el hecho de que el “área bajo la curva” es igual a la unidad, y en el cálculo infinitesimal ya hay una forma de definir el “área bajo la curva” para cualquier función dentro de un rango $a \leq x \leq b$ para la variable independiente:

$$\int_a^b f(x) dx$$

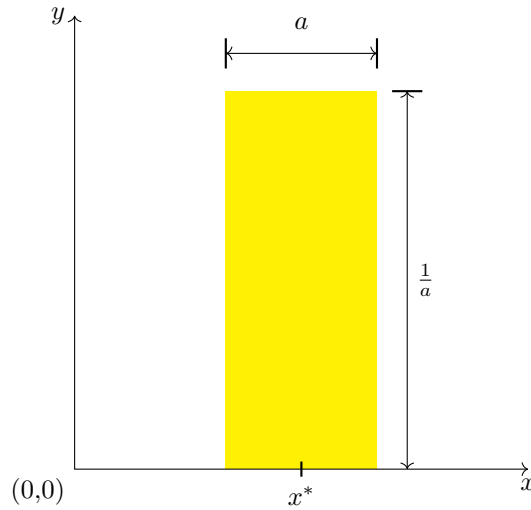


Figure 17: Idea inicial para la función delta de Dirac

En el proceso extremo, se lleva a cabo la subdivisión hasta el infinito, haciendo la anchura de cada rectángulo lo más cercano a cero que se pueda ($\Delta x \rightarrow 0$) y sumando las contribuciones de una cantidad prácticamente infinita de rectángulos, lo cual en el proceso límite de la definición de la integral puede ser simbolizado de la siguiente manera:

$$\int f(x)dx$$

Puesto que en el proceso de construcción de la función extrema se utilizó precisamente el área bajo la curva manteniéndose todo el tiempo igual a la unidad, y puesto que a los rectángulos, o mejor dicho, al rectángulo límite, se le fue disminuyendo su anchura Δx hasta hacerla igual a cero para fines prácticos, se puede definir a la función no por sí sola sino dentro de una integral, una integral que podemos especificar que sea llevada a cabo desde $-\infty$ hasta ∞ sabiendo que el “pico” tiene que estar situado necesariamente entre los dos extremos del infinito para la coordenada que representa a la variable independiente, entre el infinito negativo y el infinito positivo. Esto da la clave para tratar de darle una definición matemática completa a la función delta de Dirac $\delta(x)$ con la siguiente condición:

$$\int_{-\infty}^{\infty} \delta(x)dx = 1$$

Sin embargo, hay muchos teóricos que consideran que, siendo la función delta de Dirac una función extraordinariamente impropia, se le debe definir por el efecto que tiene sobre otra función cuando multiplica a dicha función bajo el signo de la integral. Si la función a la que multiplica es una función cualquiera $f(x)$, entonces la definición de la función delta de Dirac vendría siendo la siguiente:

$$f(a) = \int_{-\infty}^{\infty} f(x)\delta(x-a)dx$$

Se debe resaltar el hecho de que la función delta de Dirac en realidad sólo tiene sentido cuando se utiliza bajo el signo integral. Y este es el punto central que nunca hay que perder de vista: La función delta de Dirac por sí sola carece de sentido matemático y físico; dicha función solo tiene sentido cuando es usada bajo el signo de integración.

Se debe aclarar que, en rigor de verdad, la función delta de Dirac no es una función en el sentido usual que se le da a dicha palabra en las matemáticas, ya que cualquier función que sea igual a cero excepto en un solo punto debería de tener un valor de integración igual a cero, y aunque para fines formales se le puede definir como una distribución, la costumbre ha hecho que se le siga identificando como una función.

Tal vez la mejor manera de visualizar a la función delta de Dirac sea como lo que realmente es, como un proceso límite en el cual se tiene, por ejemplo, un rectángulo (que supondremos centrado en el origen $x = 0$) cuya área es igual a la unidad con su lado izquierdo situado en $x = -\epsilon$ y con su lado derecho situado en $x = +\epsilon$.

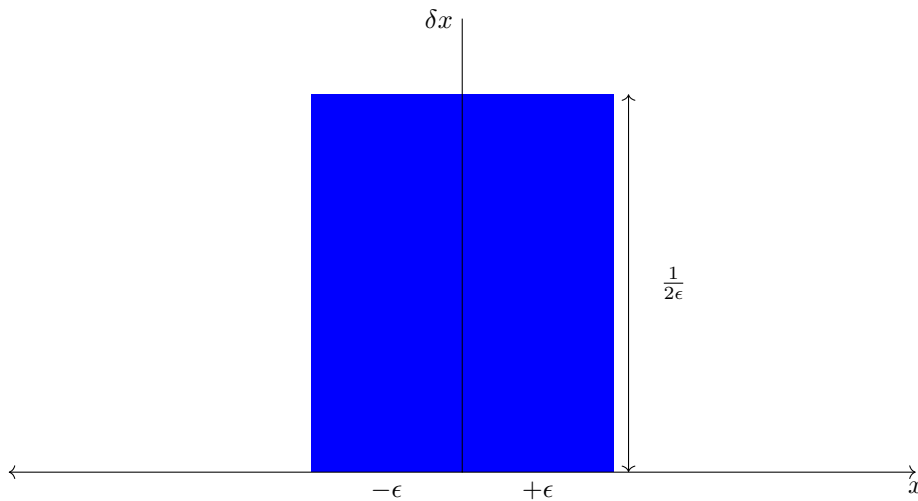


Figure 18: Función delta de Dirac como límite

De este modo, al llevar a cabo una integración desde $-\infty$ hasta $+\infty$, en realidad se está llevando a cabo una integración sobre tres regiones: la región que está a la izquierda de la función delta de Dirac (a la izquierda de $-\epsilon$), la región en la cual está situado el rectángulo hipotético que se está utilizando como función delta (entre $-\epsilon$ y $+\epsilon$), y la región que está situada a la derecha de la función delta (a la derecha de $+\epsilon$). De este modo, al llevar a cabo la integración desde $-\infty$ hasta $+\infty$ sobre una función $\delta(x)$, el proceso matemático viene siendo:

$$\int_{-\infty}^{\infty} \delta(x) dx = \int_{-\infty}^{-\epsilon} \delta(x) dx + \lim_{\epsilon \rightarrow 0} \int_{-\epsilon}^{+\epsilon} \delta(x) dx + \int_{+\epsilon}^{+\infty} \delta(x) dx$$

Obsérvese que a la segunda integral se le ha escrito como un proceso límite en el cual $\epsilon \rightarrow 0$, simbolizando con ello al rectángulo cuya anchura se va haciendo infinitamente pequeña. Puesto

que fuera de la región situada entre $-\epsilon$ y $+\epsilon$ la función $\delta(x)$ es igual a cero, la primera integral y la tercera integral son igual a cero, quedando únicamente la integral intermedia, de modo que:

$$\int_{-\infty}^{\infty} \delta(x) dx = \int_{-\infty}^{-\epsilon} 0 \cdot dx + \lim_{\epsilon \rightarrow 0} \int_{-\epsilon}^{+\epsilon} \delta(x) dx + \int_{+\epsilon}^{+\infty} 0 \cdot dx$$

De este modo, se llega a la raíz esencial de la definición ampliada de la función delta de Dirac:

$$\int_{-\infty}^{\infty} \delta(x) dx = \lim_{\epsilon \rightarrow 0} \int_{-\epsilon}^{+\epsilon} \delta(x) dx = 1$$

La notación intermedia dando como límites inferior y superior a $-\epsilon$ y $+\epsilon$ resulta particularmente útil para la resolución de problemas al momento de llevar a cabo una integración que involucre a la función $\delta(x)$. Específicamente, estamos hablando de una situación en la cual se lleva a cabo una integración de una función $\psi(x)$ cualesquiera tomando el producto de la misma con la función delta de Dirac. De acuerdo con el procedimiento que se acaba de delinear, se tiene lo siguiente:

$$\int_{-\infty}^{\infty} \psi(x) \delta(x) dx = \int_{-\infty}^{-\epsilon} \psi(x) \cdot 0 \cdot dx + \lim_{\epsilon \rightarrow 0} \int_{-\epsilon}^{+\epsilon} \psi(x) \delta(x) dx + \int_{+\epsilon}^{+\infty} \psi(x) \cdot 0 \cdot dx$$

Puesto que fuera de la región cubierta por la función delta de Dirac, el valor de $\delta(x)$ se toma como cero, esto vuelve cero la primera integral y la tercera integral dejándonos únicamente la integral intermedia. Al tomarse el límite $\epsilon \rightarrow 0$, se puede dar por hecho de que justo en el punto $x = 0$ la función $\psi(x)$ toma el valor $\psi(0)$, un valor fijo y constante, obteniéndose

$$\int_{-\infty}^{\infty} \psi(x) \delta(x) dx = \psi(0)$$

Del mismo modo, si esta centrada en $x = a$, se obtiene que

$$\begin{aligned} \int_{-\infty}^{\infty} \psi(x) \delta(x - a) dx &= \int_{-\infty}^{a-\epsilon} \psi(x) \cdot 0 \cdot dx + \lim_{\epsilon \rightarrow 0} \int_{a-\epsilon}^{a+\epsilon} \psi(x) \delta(x - a) dx + \int_{a+\epsilon}^{+\infty} \psi(x) \cdot 0 \cdot dx \\ &= \psi(a) \lim_{\epsilon \rightarrow 0} \int_{a-\epsilon}^{a+\epsilon} \delta(x - a) dx \\ &= \psi(a) \cdot 1 \\ &= \psi(a) \end{aligned}$$

Demostrar que para la función de paso unitario, también conocida como la función de Heaviside $H(x)$ la derivada de dicha función es igual a la función delta de Dirac.

$$H(x) = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x < 0 \end{cases}$$

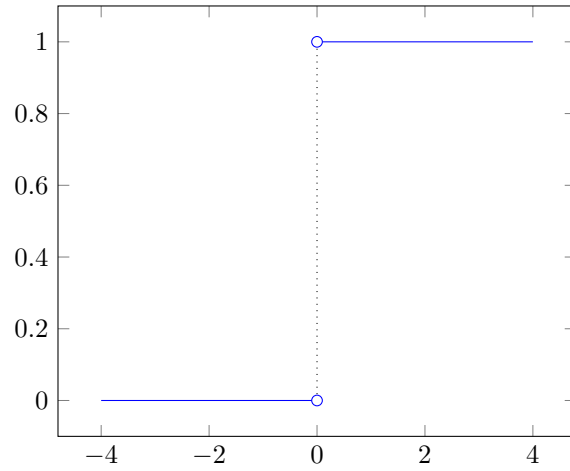


Figure 19: Función de Heaviside

La definición de la derivada para cualquier función incluyendo la función de paso unitario $H(x)$ está basada en el siguiente concepto esencial del límite:

$$\frac{dH}{dx} = \lim_{\Delta x \rightarrow 0} \frac{\Delta H}{\Delta x}$$

A continuación se superpondrán a la función de paso unitario los extremos para los cuales se considerará la diferencia Δx , uno en:

$$\Delta x = -\frac{1}{2a}$$

y el otro en

$$\Delta x = +\frac{1}{2a}$$

Gráficamente, para las regiones en las que x es menor que $x = 0$ y mayor que $x = 0$, la derivada $\frac{dH}{dx}$ es muy clara: $H(x)$ debe ser una línea plana en esas regiones y la pendiente (derivada) de una línea plana es igual a cero. En términos de la definición de derivada, puesto que H no cambia entonces $\Delta H = 0$ y consecuentemente $\frac{dH}{dx} = 0$ en dichas regiones. Pero si se toma dos puntos igualmente espaciados con respecto a $x = 0$, uno de ellos en:

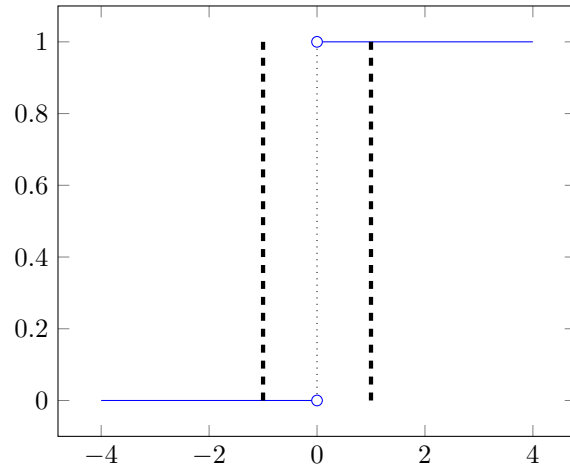
$$x_- = -\frac{a}{2}$$

y el otro en

$$x_+ = +\frac{a}{2}$$

entonces $\Delta H = 1$ y $\Delta x = a$. No importa que tan pequeño sea a , ΔH se mantendrá igual. Entonces la derivada de $H(x)$ para $x = 0$ será:

$$\frac{dH}{dx} = \lim_{a \rightarrow 0} \frac{1}{a}$$



Gráficamente, esto es también bastante claro. $H(x)$ salta de $H(x) = 0$ a $H(x) = 1$ en el punto $x = 0$, de modo tal que la pendiente es esencialmente vertical o infinita. Juntando los resultados para $\frac{dH}{dx}$ en uno solo tenemos lo siguiente:

$$\frac{dH}{dx} = \begin{cases} 0 & \text{si } x < 0 \\ \infty & \text{si } x = 0 \\ 0 & \text{si } x > 0 \end{cases}$$

que resulta en la definición de la delta de Dirac.

Se había mencionado previamente que para fines formales a la función delta de Dirac se le puede tomar como una distribución. No resulta difícil ver el origen de tal idea.

La curva Gaussiana, es una función normalizada a la unidad, esto es, el área bajo la curva es igual a 1, con lo cual se da a entender que la probabilidad de encontrar cualquier elemento de una población bajo cualquier parte de la curva Gaussiana (la cual se extiende desde menos infinito hasta más infinito) es igual a la certeza. Y resulta que “el área bajo la curva” de la función delta de Dirac también es igual a la unidad, definida también entre los límites de menos infinito y más infinito. La distribución Gaussiana, al igual que la función delta de Dirac, también tiene su “pico máximo”, aunque el “pico” de la distribución Gaussiana tiene una altura finita mientras que el “pico” de la función delta tiene una altura infinita.

Al variar los parámetros de la distribución Gaussiana, la anchura de la curva disminuye haciéndose a la vez más y más pronunciado el “pico máximo” de la curva Gaussiana de modo tal que el área bajo la curva Gaussiana siga manteniendo un valor igual a la unidad.

Llevando hacia el extremo el procedimiento de ir disminuyendo el valor de σ , haciéndolo casi igual a cero (aunque nunca a cero exactamente, ya que ello conduce a una indefinición matemática) se obtiene gráficamente lo que para cualquiera parecerá ser una función delta de Dirac, aunque no

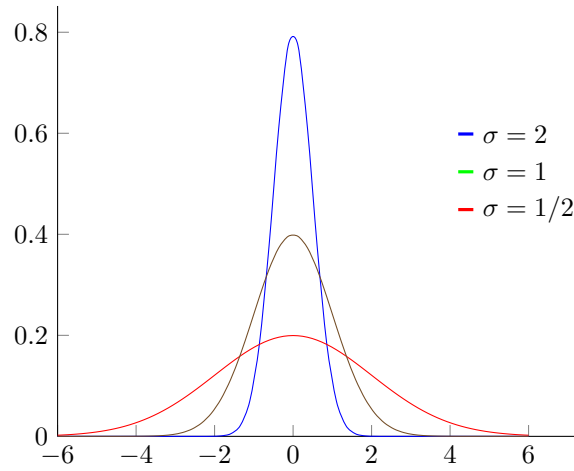


Figure 20: Distribución Gaussiana

pueda serlo. La función delta de Dirac $\delta(k)$ se define del siguiente modo:

$$\delta(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx} dx$$

Llevando a cabo una simple diferenciación bajo el signo de la integral con respecto a k de $\delta(k)$, se puede dar la siguiente definición formal a la derivada de la función $\delta(k)$:

$$\frac{d(\delta(k))}{dk} = \frac{1}{2\pi} \int_{-\infty}^{\infty} ike^{ikx} dx$$

17.2 Caminata aleatoria (parte 2)

17.3 Caminata aleatoria en dos dimensiones

18 Tarea 18

18.1 Relación entre la distribución normal y la Gaussiana

Teorema 11. La función generadora de momentos de una variable aleatoria con distribución $B(n, p)$ es

$$M_x(\theta) = (q + pe^{\theta})^n$$

donde $q = 1 - p$.

Demostración.

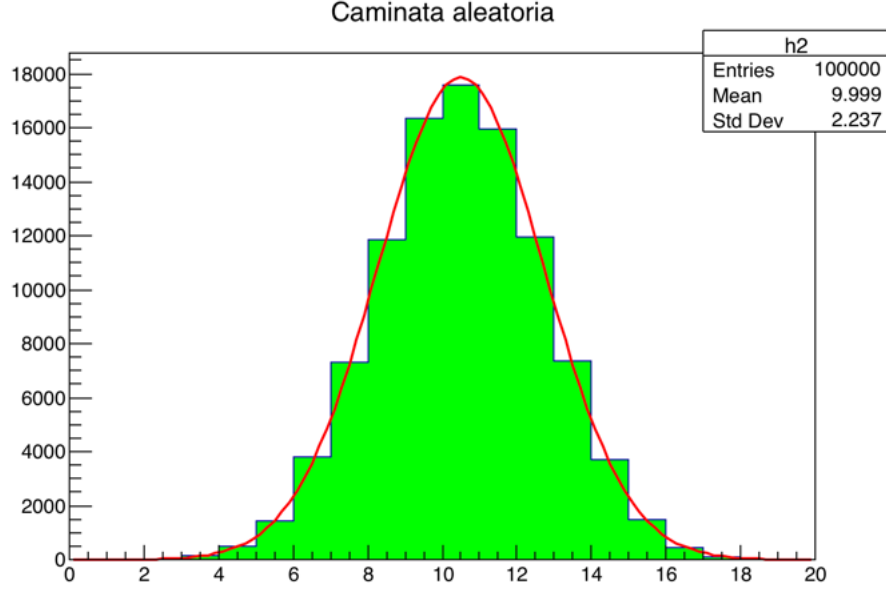


Figure 21: Histograma para la caminata aleatoria con respecto a n (pasos a la derecha)

Usando la definición de una distribución binomial y como se vió anteriormente, la definición de la función generadora de momentos, se tiene que

$$M_x(\theta) = \sum_{x=0}^{\infty} e^{\theta x} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{\infty} \binom{n}{x} p e^{\theta x} q^{n-x} = (q + p e^{\theta})^n$$

Teorema 12. Si x es una variable aleatoria con distribución $B(n, p)$, entonces para n lo suficientemente grande, la siguiente variable aleatoria tiene una distribución normal estándar:

$$z = \frac{(x - \mu)}{\sigma} \sim N(0, 1)$$

donde

$$\mu = np$$

$$\sigma^2 = np(1 - p)$$

Demostración.

Como la función generadora de momentos es una transformación lineal, se tiene que

$$M_z(\theta) = e^{-\frac{\mu\theta}{\sigma}} M_x\left(\frac{\theta}{\sigma}\right) = e^{-\frac{\mu\theta}{\sigma}} (q + p e^{\frac{\theta}{\sigma}})^n$$

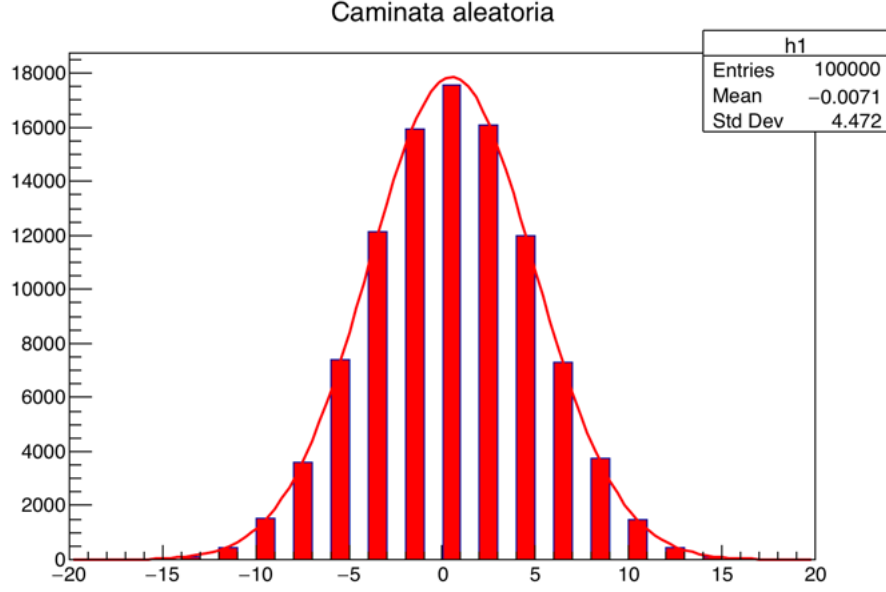


Figure 22: Histograma para la caminata aleatoria con respecto a m (posición final)

Tomando el logaritmo natural en ambos lados y expandiendo la serie de potencias de $e^{\theta/\sigma}$ se obtiene

$$\ln M_z(\theta) = -\frac{\mu\theta}{\sigma} + n \ln(q + pe^{\frac{\theta}{\sigma}}) = -\frac{\mu\theta}{\sigma} + n \ln \left(q + p \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\theta}{\sigma} \right)^k \right)$$

Como $p + q = 1$

$$\ln M_z(\theta) = -\frac{\mu\theta}{\sigma} + n \ln \left(1 + p \sum_{k=1}^{\infty} \frac{1}{k!} \left(\frac{\theta}{\sigma} \right)^k \right)$$

Si se hace n lo suficientemente grande, $\sigma = \sqrt{npq}$ puede hacerse lo suficientemente grande para que, para cualquier θ fija, el valor absoluto de la suma anterior sea menor a 1. Sea

$$z = p \sum_{k=1}^{\infty} \frac{1}{k!} \left(\frac{\theta}{\sigma} \right)^k$$

Entonces, para n lo suficientemente grande, $|z| < 1$. El término que involucra a \ln en la expresión anterior es $\ln(1 + z)$ donde $|z| < 1$, por lo que se expande de la siguiente manera:

$$\ln(1 + z) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{z^k}{k}$$

Lo que implica

$$\ln M_z(\theta) = -\frac{\mu\theta}{\sigma} + n \ln(1 + z) = -\frac{\mu\theta}{\sigma} + n \sum_{k=1}^{\infty} (-1)^{k-1} \frac{z^k}{k}$$

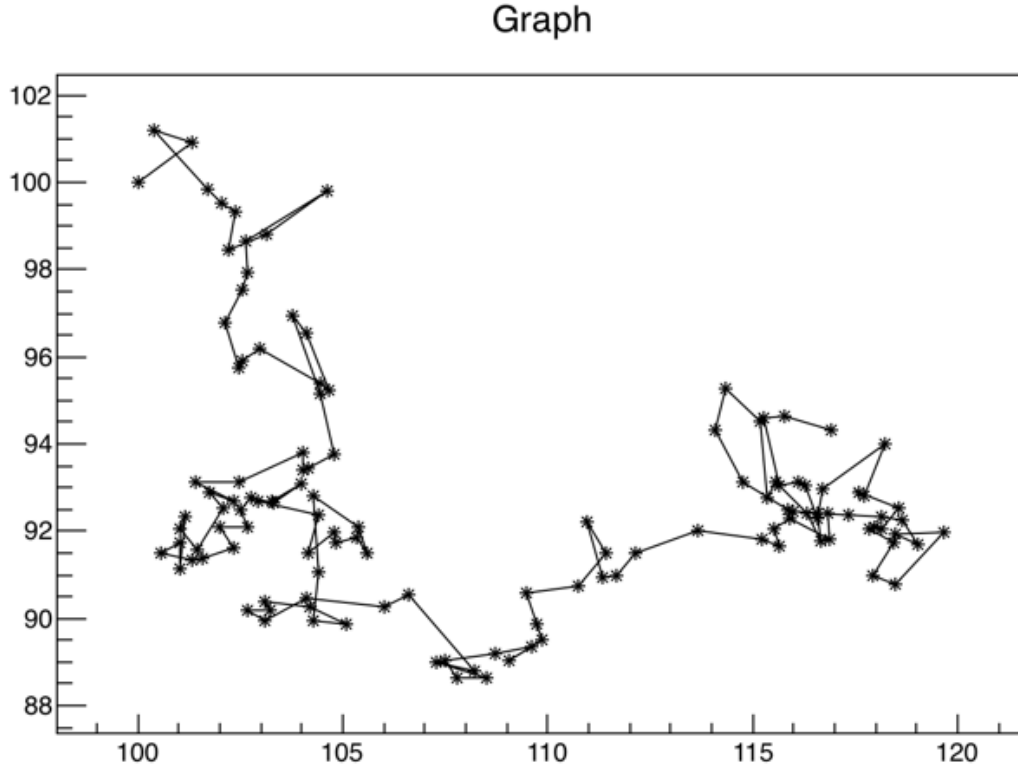


Figure 23: Caminata aleatoria en dos dimensiones

Acomodando los términos

$$\ln M_z(\theta) = -\frac{\mu\theta}{\sigma} + n \ln(1+z) = -\frac{\mu\theta}{\sigma} + n \sum_{k=1}^{\infty} (-1)^{k-1} \frac{z^k}{k} = \left(-\frac{\mu}{\sigma} + \frac{np}{\sigma}\right) \theta + n \left(\frac{p}{\sigma^2} - \frac{p^2}{\sigma^2}\right) \frac{\theta^2}{2!} + n \sum_{k=3}^{\infty} \frac{c_k}{\sigma^k} \theta^k$$

para valores de c_k que no implican a n, σ o θ . Como $\mu = np$ y $\sigma^2 = np(1-p)$, el coeficiente del término que implica a θ es cero y el término que implica a θ^2 es uno. Entonces,

$$\ln M_z(\theta) = \frac{\theta^2}{2} + \sum_{k=3}^{\infty} \frac{nc_k}{\sigma^k} \theta^k$$

Como el coeficiente de cada término en la suma es de la forma

$$\frac{nc_k}{\sigma^k} = \frac{nc_k}{(npq)^{k/2}} = \frac{nc_k}{(pq)^{k/2} n^{k/2-1}} \rightarrow p \text{ cuando } n \rightarrow \infty$$

Por lo tanto

$$\lim_{n \rightarrow \infty} \ln M_z(\theta) = \frac{\theta^2}{2}$$

y

$$\lim_{n \rightarrow \infty} M_z(\theta) = e^{\frac{\theta^2}{2}}$$

Pero la función generadora de momentos de una distribución normal con parámetros $\mu = 0$ y $\sigma^2 = 1$ esta dada por

$$M_z(\theta) = e^{\frac{\theta^2}{2}}$$

Si dos distribuciones tienen la misma función generadora de momentos, se considera que son la misma, por lo que, para la distribución binomial, a la larga se considera que converge a la distribución normal.

18.2 Graficar puntos de mathematica en Tikz

El comando usado es:

```
\tikz \draw plot coordinates {(0,0) (1,1) (2,0) (3,1) (2,1) (10:2cm)};
```

Leyendo de un archivo externo:

La segunda manera para especificar puntos es ponerlos en un archivo externo llamado ¡nombre!. Cada línea del archivo debe contener una línea que comienza con dos números separados por un espacio, lo que siga después de los dos números es ignorado. Además, las líneas que contengan % o # son ignorados, al igual que las líneas en blanco.

```
\tikz \draw plot[mark=x,smooth] file {plots/pgfmanual-sine.table};
```

```
#Curve 0, 20 points
#x y type
0.00000 0.00000 i
0.52632 0.50235 i
1.05263 0.86873 i
1.57895 0.99997 i
...
9.47368 -0.04889 i
10.00000 -0.54402 i
```

18.3 Delta de dirac en coordenadas polares

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

19 Tarea 19

19.1 Distribuciones de probabilidad

19.1.1 Disitribución de Bernoulli

Para comprender el proceso de Bernoulli se puede pensar, por ejemplo, en situaciones en las que sólo hay dos posibles resultados mutuamente excluyentes (verdadero/falso, en un test; defectuoso/no defectuoso, en los artículos que salen de una fábrica; aprobado/suspendido, en los resultados de un examen, etc.). Se dice que son mutuamente excluyentes porque no pueden darse simultáneamente (un examen no puede estar aprobado y suspendido al mismo tiempo; una respuesta no puede ser al mismo tiempo verdadera o falsa, etc.). Una manera común de designar estos dos resultados es como Éxito (E) o Fracaso (F). Una segunda característica de los fenómenos que siguen el denominado Proceso de Bernoulli es que las pruebas de las que se obtienen los éxitos o los fracasos son independientes. Así, el hecho de que un artículo salga defectuoso en una línea de producción no tiene que ver con el resultado obtenido en el siguiente artículo que se examina. Por último, una tercera característica de este proceso es que las probabilidades de éxito o fracaso son constantes.

Se denota p a la probabilidad de éxito y $1 - p$ a la probabilidad de fracaso. Se define una variable aleatoria X tal que $X = 1$ si el resultado es exitoso y $X = 0$ si el resultado es un fracaso, es decir,

$$\begin{aligned}P(X = 1) &= p \\P(X = 0) &= 1 - p\end{aligned}$$

Para calcular la media se utiliza la ecuación del primero momento alrededor de 0, es decir

$$\mu = E(X) = \sum_{j=1}^2 x_j f(x_j) = 1 \cdot p + 0 \cdot (1 - p) = p$$

y para calcular la varianza

$$\begin{aligned}\sigma^2 &= \sum_{j=1}^2 (x_j - \mu)^2 f(x_j) \\&= \sum_{j=1}^2 (x_j - p)^2 f(x_j) \\&= (1 - p)^2 p + (0 - p)^2 q = (1 + p^2 - 2p)p + p^2 q \\&= p + p^3 - 2p^2 + p^2 - p^3 = (1 - p)p = pq\end{aligned}$$

Es decir, la distribución de Bernoulli tiene una media

$$\mu = p$$

y varianza

$$\sigma^2 = pq$$

19.1.2 Distribución binomial

Sea un experimento aleatorio en el que pueden obtenerse dos resultados posibles, mutuamente excluyentes, con probabilidades constantes en el que p denota la probabilidad de éxito.

Supongamos que se realizan n pruebas independientes (es decir, se dan las condiciones de Bernoulli) y tenemos una sucesión de Bernoulli de tamaño n . Sea X la variable definida como el número de éxitos resultantes en la sucesión de Bernoulli. X se distribuye como una distribución binomial.

Pensemos, por ejemplo, en un agente de seguros que tiene como posibles resultados de su gestión hacer un seguro (éxito) o no hacerlo (fracaso). Todos los eventos, si son una sucesión de Bernoulli de tamaño n , se distribuyen como una distribución binomial de parámetros n y p , y se denota como

$$X \sim B(n, p)$$

En donde n es el tamaño de la sucesión de Bernoulli, el número de veces que se repite el experimento y p es la probabilidad de éxito.

La función de cuantía de la distribución binomial es

$$f(X = x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & \text{si } x = x_i \\ 0 & \text{si } x \neq x_i \end{cases}$$

La función de distribución esta dada por

$$F(x) = P(X \leq x) = \sum_{i \leq x} \binom{n}{i} p^i q^{n-i}$$

Para calcular la media de una distribución binomial, hay que recordar que ésta se define como una suma de procesos de Bernoulli, es decir $X = X_1 + \dots + X_n$ donde X_i es un proceso de Bernoulli y X es la variable con distribución binomial. Luego,

$$\begin{aligned} E[X] &= E[X_1 + X_2 + \dots + X_n] \\ &= E[X_1] + \dots + E[X_n] \\ &= p + \dots + p = np \end{aligned}$$

De igual modo para la varianza

$$\begin{aligned} V[X] &= V[X_1 + \dots + X_n] \\ &= V[X_1] + \dots + V[X_n] \\ &= np(1 - p) \end{aligned}$$

19.1.3 Distribución uniforme

Una variable aleatoria X se distribuye uniformemente en un intervalo $[a, b]$, representada como $X \sim U(a, b)$ cuando su función de densidad es

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{en otro caso} \end{cases}$$

y su función de distribución es

$$F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b \end{cases}$$

Para calcular la media de una variable que se distribuye de manera uniforme en el intervalo $[a, b]$

$$\begin{aligned} \int_a^b x \frac{1}{b-a} dx &= \frac{1}{b-a} \int_a^b x dx \\ &= \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b \\ &= \frac{1}{b-a} \left[\frac{b^2 - a^2}{2} \right] \\ &= \frac{1}{b-a} \cdot \frac{(b-a)(b+a)}{2} = \frac{b+a}{2} \end{aligned}$$

De igual modo para calcular la varianza

$$\begin{aligned} \int_a^b x^2 \frac{1}{b-a} dx &= \frac{1}{b-a} \int_a^b x^2 dx \\ &= \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b \\ &= \frac{1}{b-a} \left[\frac{b^3 - a^3}{3} \right] \\ &= \frac{b^2 + ab + a^2}{3} \end{aligned}$$

Finalmente

$$\sigma^2 = E(X^2) - E(X)^2 = \frac{(b-a)^2}{12} \quad (23)$$

19.1.4 Distribución exponencial

Se dice que una variable aleatoria X sigue una distribución exponencial con parámetro $\lambda > 0$ y denotada por $X \sim \exp(\lambda)$ cuya función de densidad es

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

y su función de distribución esta dada por

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Para calcular su media

$$\begin{aligned} \mu &= \int_0^{\infty} x f(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx \\ &= -x e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = \int_0^{\infty} e^{-\lambda x} dx \\ &= -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^{\infty} = \frac{1}{\lambda} \end{aligned}$$

Para calcular la varianza, se puede utilizar la siguiente fórmula

$$E[(x - \mu)^2] = E[x^2] - E[x]^2$$

Por lo que

$$\begin{aligned} E[x^2] &= \int_0^{\infty} x^2 f(x) dx = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx \\ &= -x^2 e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} 2x e^{-\lambda x} dx \\ &= \frac{2}{\lambda} \left[-\frac{1}{\lambda} e^{-\lambda x} \right]_0^{\infty} \\ &= \frac{2}{\lambda^2} \end{aligned}$$

Por lo tanto

$$\sigma^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda} \right)^2 = \frac{1}{\lambda^2}$$

19.1.5 Distribución normal

La distribución normal es la más común entre todas las distribuciones de probabilidad utilizadas en Estadística y tiene importantes aplicaciones en el modelado de variables estadísticas asociadas a los

elementos de una población. Por ejemplo, las medidas físicas del cuerpo humano en una población, las características medidas por una prueba de inteligencia o personalidad, las medidas de calidad en muchos procesos industriales, etc, siguen distribuciones normales.

Una justificación de la frecuente aparición de la distribución normal es el teorema central del límite que establece que cuando los resultados de un experimento sean debidos a un conjunto muy grande de causas independientes, que actúan sumando sus efectos, siendo cada efecto individual de poca importancia respecto al conjunto, se espera que los resultados sigan una distribución normal.

Una variable aleatoria que sigue una distribución normal con parámetros μ y σ^2 se representa como $X \sim N(\mu, \sigma^2)$ con una función de densidad

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

La familia de distribuciones normales depende de los parámetros μ y σ^2 que coinciden con la media y la varianza, respectivamente.

La función de distribución de una normal toma la forma

$$F(x) = \int_{-\infty}^x f(x)dx = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Para calcular la media

$$\bar{x} = \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

realizando un cambio de variable $y = x - \mu$ y $dy = dx$ se obtiene

$$\bar{x} = \frac{1}{\sigma\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} y e^{-\frac{y^2}{2\sigma^2}} dy + \mu \int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma^2}} dy \right] \quad (24)$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \left[0 + \mu\sqrt{2\pi\sigma^2} \right] = \mu \quad (25)$$

19.1.6 Distribución de Poisson

Suponga que se desea hallar la distribución de probabilidad del número de accidentes viales ocurridos en un cruce particular durante un periodo de una semana. Considere el periodo de una semana dividido en n subintervalos, cada uno de los cuales es tan pequeño que a lo sumo un accidente podría ocurrir en él con probabilidad diferente de cero. Denotando p la probabilidad de un accidente en cualquier subintervalo, se tiene

$$P(\text{no ocurren accidentes en el subintervalo}) = 1 - p$$

$$P(\text{ocurre un accidente en el subintervalo}) = p$$

$$P(\text{ocurre más de un accidente en el subintervalo} = 0)$$

El número total de accidentes se obtiene al contar el número de intervalos que contienen un accidente. Si los eventos son independientes de un intervalo de a otro, el número total de accidentes tiene distribución binomial.

Aun cuando no hay forma única de seleccionar los subintervalos y no se conoce ni n ni p , si el intervalo se divide en intervalos más pequeños (más que n subintervalos), disminuye la probabilidad p de un accidente en cada nuevo subintervalo [4]. Si se hace $\lambda = np$ y al tomar el límite de la probabilidad binomial

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

cuando $n \rightarrow \infty$ se tiene que

$$\begin{aligned} \lim_{n \rightarrow \infty} \binom{n}{x} p^x (1-p)^{n-x} &= \lim_{n \rightarrow \infty} \frac{n(n-1) \cdots (n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \lim_{n \rightarrow \infty} \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n(n-1) \cdots (n-x+1)}{n^x} \left(1 - \frac{\lambda}{n}\right)^{-x} \\ &= \frac{\lambda^x}{x!} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right) \end{aligned}$$

Por otro lado, por definición

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

junto con el hecho de que todos los términos a la derecha del límite tienden a 1 cuando n tiende a infinito, se obtiene

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Se dice que una variable aleatoria X tiene una distribución de probabilidad de Poisson si y sólo si

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad \lambda > 0$$

Si X es una variable aleatoria que sigue una distribución de Poisson con parámetro λ entonces

$$\mu = \lambda \quad \text{y} \quad \sigma^2 = \lambda$$

Demostración

Por definición

$$E(X) = \sum_x x p(x) = \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda}$$

El primer término de esta suma es igual a cero, por lo que

$$E(X) = \sum_x x p(x) = \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} = \sum_{x=1}^{\infty} x \frac{\lambda^x}{(x-1)!} e^{-\lambda}$$

Al hacer el cambio de variable $y = x - 1$, la expresión anterior es una función de probabilidad. Los límites de la sumatoria se convierten en $y = 0$ cuando $x = 1$ y $y = \infty$ cuando $x = \infty$. Entonces

$$E(X) = \lambda \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} e^{-\lambda}$$

El truco es que ésta expresión es una función de probabilidad, siendo más específicos, resulta ser la misma distribución de Poisson pero con diferente notación, y $\sum_{y=0}^{\infty} p(y) = 1$. Por lo tanto

$$\mu = E(X) = \lambda$$

Para calcular la varianza, se puede usar el siguiente truco: calcular $E[X(X - 1)]$. Usando la definición de esperanza

$$\begin{aligned} E[X(X - 1)] &= \sum_{x=0}^{\infty} x(x - 1) \frac{\lambda^x}{(x)!} e^{-\lambda} \\ &= \sum_{x=2}^{\infty} \frac{\lambda^x}{(x - 2)!} e^{-\lambda} \\ &= \lambda^2 \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} e^{-\lambda} = \lambda^2 \end{aligned}$$

Nótese que se utilizó el mismo recurso que el de la demostración de la media. Ahora, viene lo interesante. Como la esperanza es un operador lineal, se tiene que

$$E[X(X - 1)] = E[X^2] - E[X] = \lambda^2$$

o visto de otra forma

$$E[X^2] = \lambda^2 + E[X] = \lambda^2 + \lambda$$

Como se demostró anteriormente

$$\begin{aligned} \sigma^2 &= E[X^2] - E[X]^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

19.1.7 Distribución hipergeométrica

Se dice que una variable aleatoria X tiene una distribución de probabilidad hipergeométrica si y sólo si

$$p(x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}$$

donde $x = 0, 1, 2, \dots, n$ sujeto a las restricciones $x \leq r$ y $n - x \leq N - r$

La media de la distribución hipergeométrica

$$\begin{aligned}
E(X) &= \sum_{x=0}^n x \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}} \\
&= \sum_{x=0}^n \frac{x(r-1)!r}{(r-x)!(x-1)!x} \frac{\binom{N-r}{n-x}}{\frac{N(N-1)!}{(N-n)!(n-1)!n}} \\
&= \frac{nr}{N} \sum_{x=0}^n \frac{\binom{r-1}{x-1} \binom{N-r}{n-x}}{\binom{N-1}{n-1}} \\
&= \frac{nr}{N} \sum_{x=1}^n \frac{\binom{r-1}{x-1} \binom{N-r}{n-x}}{\binom{N-1}{n-1}}
\end{aligned}$$

Este último paso es válido dada la convención $\binom{a}{b} = 0$ si $b > a$. Haciendo el cambio de variable $y = x - 1$

$$\begin{aligned}
&= \frac{nr}{N} \sum_{x=1}^{n-1} \frac{\binom{r-1}{y} \binom{N-r-1+1}{n-y-1}}{\binom{N-1}{n-1}} \\
&= \frac{nr}{N}
\end{aligned}$$

donde

$$\sum_{i=0}^m \binom{a}{i} \binom{b}{m-i} = \binom{a+b}{m}$$

(identidad de Vandermonde).

Para calcular la varianza, primero se calcula $E(x^2)$

$$\begin{aligned}
E(x^2) &= \sum_{x=0}^n x^2 \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}} \\
&= r \sum_{x=0}^n x \frac{\binom{r-1}{x-1} \binom{N-r}{n-x}}{\binom{N}{n}} \\
&= r \sum_{x=1}^n x \frac{\binom{r-1}{x-1} \binom{N-r}{n-x}}{\binom{N}{n}}
\end{aligned}$$

Sea $y = x - 1$, entonces

$$\begin{aligned}
&= r \sum_{y=0}^{n-1} (y+1) \frac{\binom{r-1}{y} \binom{N-r}{n-y-1}}{\binom{N}{n}} \\
&= r \left[\sum_{y=0}^{n-1} y \frac{\binom{r-1}{y} \binom{N-r}{n-y-1}}{\binom{N}{n}} + \sum_{y=0}^{n-1} \frac{\binom{r-1}{y} \binom{N-r}{n-y-1}}{\binom{N}{n}} \right] \\
&= r \left[\sum_{y=0}^{n-1} y \frac{\binom{r-1}{y} \binom{N-r}{n-y-1}}{\frac{N}{n} \binom{N-1}{n-1}} + \sum_{y=0}^{n-1} \frac{\binom{r-1}{y} \binom{N-r}{n-y-1}}{\frac{N}{n} \binom{N-1}{n-1}} \right]
\end{aligned}$$

La primera sumatoria resulta ser el valor medio cuando se tiene $r-1$, $n-1$ y $N-1$, es decir

$$\begin{aligned}
&= r \left[\frac{n}{N} \frac{(n-1)(r-1)}{N-1} + \frac{n}{N} \right] \\
&= \frac{rn}{N} \left[\frac{(n-1)(r-1)}{N-1} + 1 \right]
\end{aligned}$$

Ahora, para calcular la varianza

$$\begin{aligned}
\sigma^2 &= E(x^2) - E(x)^2 \\
&= \frac{rn}{N} \left[\frac{(n-1)(r-1)}{N-1} + 1 \right] - \frac{r^2 n^2}{N^2} \\
&= \frac{rn}{N} \left[\frac{(n-1)(r-1)}{N-1} + 1 - \frac{nr}{N} \right]
\end{aligned}$$

19.1.8 Distribución de Cauchy

La distribución de Cauchy (a veces también distribucin de Lorentz) es una distribución de probabilidad continua cuya función de densidad es

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

donde x_0 es el parámetro de corrimiento que especifica la ubicación del pico de la distribución, y γ es el parámetro de escala que especifica el ancho medio al máximo medio (half-width at half-maximum, HWHM).

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

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

19.1.9 Distribución gamma

Se dice que una variable aleatoria X tiene una distribución gamma con parámetros $\alpha > 0$ y $\beta > 0$ si y sólo si la función de densidad de X es

$$f(x) = \begin{cases} \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} & \text{si } 0 \leq x < \infty \\ 0 & \text{en cualquier otro punto} \end{cases}$$

donde

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

Si X tiene una distribución gamma con parámetros α y β , entonces

$$\mu = E(X) = \alpha\beta \quad \text{y} \quad \sigma^2 = V(X) = \alpha\beta^2$$

Para demostrar éstos resultados, se procede de manera usual que con los resultados anteriores. Para calcular la media, por definición

$$E(X) = \int_{-\infty}^\infty x f(x) dx = \int_0^\infty x \left(\frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} \right) dx$$

Por definición, la función de densidad gamma es tal que

$$\int_0^\infty \left(\frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} \right) dx = 1$$

Por tanto,

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

y

$$\begin{aligned} E(X) &= \int_0^\infty \left(\frac{x^\alpha e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} \right) dx = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^\alpha e^{-x/\beta} dx \\ &= \frac{1}{\beta^\alpha \Gamma(\alpha)} [\beta^{\alpha+1} \Gamma(\alpha+1)] \\ &= \frac{\beta^\alpha \Gamma(\alpha)}{\Gamma(\alpha)} = \alpha\beta \end{aligned}$$

Para calcular la varianza

$$\begin{aligned}
 E(X^2) &= \int_0^\infty x^2 \left(\frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} \right) dx \\
 &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha+1} e^{-x/\beta} dx \\
 &= \frac{1}{\beta^\alpha \Gamma(\alpha)} [\beta^{\alpha+2} \Gamma(\alpha+2)] \\
 &= \frac{\beta^2 (\alpha+1) \alpha \Gamma(\alpha)}{\Gamma(\alpha)} = \alpha(\alpha+1) \beta^2
 \end{aligned}$$

Finalmente

$$\begin{aligned}
 V(Y) &= E(X^2) - E(X)^2 \\
 &= \alpha(\alpha+1) \beta^2 - (\alpha\beta)^2 \\
 &= \alpha^2 \beta^2 + \alpha\beta^2 - \alpha^2 \beta^2 = \alpha\beta^2
 \end{aligned}$$

19.1.10 Distribución beta

La distribución de probabilidad beta es una función de densidad de dos parámetros definida sobre el intervalo cerrado $0 \leq y \leq 1$. Frecuentemente se usa como modelo para proporciones, como la proporción de impurezas en un producto químico o la proporción de tiempo que una máquina esta en reparación.

Se dice que una variable aleatoria X tiene una distribución de probabilidad beta con parámetros $\alpha > 0$ y $\beta > 0$ si y sólo si la función de densidad de X es

$$f(x) = \begin{cases} \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)} & \text{si} & 0 \leq x < 1 \\ 0 & \text{en cualquier otro punto} \end{cases}$$

donde

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

Para calcular la media de una variable aleatoria con distribución beta con parámetros $\alpha > 0$ y

$\beta > 0$

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_0^1 x \left[\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \right] dx \\
&= \frac{1}{B(\alpha, \beta)} \int_0^1 x^{\alpha}(1-x)^{\beta-1} dx \\
&= \frac{B(\alpha+1, \beta)}{B(\alpha, \beta)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)} = \frac{\alpha}{\alpha+\beta}
\end{aligned}$$

Para calcular la varianza, primero se calcula $E(X^2)$

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_0^1 x^2 \left[\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \right] dx \\
&= \frac{1}{B(\alpha, \beta)} \int_0^1 x^{\alpha+1}(1-x)^{\beta-1} dx \\
&= \frac{B(\alpha+2, \beta)}{B(\alpha, \beta)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\alpha(\alpha+1)\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)(\alpha+\beta+1)\Gamma(\alpha+\beta)} = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}
\end{aligned}$$

Finalmente

$$\begin{aligned}
Var(X) &= E(X^2) - E(X)^2 \\
&= \frac{(\alpha-1)\alpha}{(\alpha+\beta-1)(\alpha+\beta)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\
&= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
\end{aligned}$$

Por lo tanto, el valor de la varianza es

$$Var(X) = \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

19.2 Procesos NP

“P versus NP - a gift to mathematics from computer science”

Steve Smale

La Teoría de la Complejidad Computacional es una rama de la teoría de la computación que se centra en la clasificación de los problemas computacionales de acuerdo a su dificultad inherente, y en la relación entre dichas clases de complejidad.

Un problema se cataloga como “inherentemente difícil” si su solución requiere de una cantidad significativa de recursos computacionales, sin importar el algoritmo utilizado. La teoría de la complejidad computacional formaliza dicha aseveración, introduciendo modelos de cómputo matemáticos para el estudio de estos problemas y la cuantificación de la cantidad de recursos necesarios para resolverlos, como tiempo y memoria.

Las clases de complejidad más sencillas se definen teniendo en cuenta factores como:

- El tipo de problema computacional: los problemas más utilizados son los problemas de decisión, pero las clases de complejidad se pueden definir para otros tipos de problemas.
- El modelo de cómputo: el modelo de cómputo más común es la Máquina de Turing determinista, pero muchas clases de complejidad se basan en Máquinas de Turing no deterministas, Máquinas de Turing cuánticas, etc.
- El recurso (o recursos) que está(n) siendo acotado(s) y la(s) cota(s): estas dos propiedades usualmente se utilizan juntas, por ejemplo, “tiempo polinomial”, “espacio logarítmico”, “profundidad constante”, etc.

¿Por qué tiempo polinomial?

Para empezar, los polinomios tipifican las funciones de “crecimiento lento”. La cerradura de los polinomios bajo la suma, multiplicación y composición preserva la noción de eficiencia (por ejemplo, subrutinas o llamadas recursivas). Al adoptar esta notación, se elimina la necesidad de expresar de manera precisa el modelo computacional [12]. De manera general, resulta más fácil comprobar una solución que encontrarla.

De igual manera, es importante este concepto por que permite tener una idea del panorama en que muchas soluciones a problemas computacionales se encuentran. Por ejemplo, es deseable que un algoritmo se ejecute en tiempo lineal, es decir, que dependa sólo de la entrada de un parámetro. Existen algunos otros problemas que se ejecutan en tiempo cuadrático n^2 . Sin embargo, existen otros problemas cuya solución se puede calcular a lo menos, en tiempo exponencial, por lo que reducir cualquiera de éstos complejidades a tiempo polinomial es más que deseable.

La clase de problemas no deterministas aceptables polinómicamente, NP , contiene todos los problemas cuya solución puede ser verificada de manera polinomial. En otras palabras, solo se puede aceptar o rechazar la solución pero no se puede encontrar la solución.

Una reducción polinómica es una manera de relacionar dos problemas de decisión, de manera que la existencia de un algoritmo que resuelve el primer problema, garantiza inmediatamente, y a través de un tiempo polinómico, la existencia de un algoritmo que resuelve el segundo.

Formalmente, sean L y M lenguajes formales sobre los alfabetos Σ y Γ , respectivamente, una transformación polinómica de L en M es una función computable:

$$f : \Sigma^* \rightarrow \Gamma^*$$

que puede ser calculada en tiempo polinómico en función del tamaño de la entrada, y que está definida por:

$$w \in L \Leftrightarrow f(w) \in M$$

para todo elemento w de Σ^* . Cuando esta función f existe, se dice que L es polinómicamente transformable en M .

Un problema de decisión C es NP -completo si:

- C es un problema NP , y
- Todo problema de NP se puede transformar polinómicamente en C

Algunos problemas de complejidad NP

- Cobertura de vértice. Dado un grafo G y un número entero k , existe una colección de k vértices de tal manera que cada arista está conectado a uno de los vértices de la colección?
- Número cromático. Dada un grafo G y un entero k , ¿existe una manera para colorear los vértices con k colores tal que los vértices adyacentes sean de un color diferente?
- Problema del agente viajero. Dadas n ciudades y un número entero k , ¿existe un recorrido, de longitud menor que k , de las ciudades comenzando y terminando en la misma ciudad?
- Problema de la mochila. Dados n elementos, cada uno con un peso y un valor, y dos números enteros k y m ¿existe una colección de elementos con peso total menor que k con un valor total superior a m ?

19.3 El hombre anumérico capítulo 2

Una de las principales características de las personas anuméricas es la tendencia a sobrestimar la frecuencia de las coincidencias. Generalmente dan mucha importancia a todo tipo de correspondencias, y, en cambio, dan muy poca a evidencias estadísticas menos relumbrantes, pero absolutamente concluyentes.

La moraleja vuelve a ser que mientras es probable que ocurra algún hecho improbable, lo es mucho menos que se dé un caso concreto. El divulgador matemático Martin Gardner ilustra la distinción entre acontecimientos genéricos y acontecimientos concretos por medio de una ruleta con las veintiséis letras del alfabeto. Si se la hace girar cien veces y se apunta la letra que sale cada vez, la probabilidad de que salga la palabra GATO o FRÍO es muy baja, pero la probabilidad de que salga alguna palabra es ciertamente alta.

Otro problema de probabilidad sirve para ilustrar lo corrientes que pueden llegar a ser las coincidencias en otro contexto. El problema se formula a menudo como sigue: un número grande de hombres dejan sus sombreros en el guardarropa de un restaurante y el encargado baraja inmediatamente los números de orden de los sombreros. ¿Cuál es la probabilidad de que, a la salida, por lo menos uno

de los hombres recupere su propio sombrero? Lo natural es pensar que, al tratarse de un número grande de hombres, la probabilidad ha de ser muy pequeña. Sorprendentemente, el 63% de las veces por lo menos uno de los clientes recuperará su sombrero.

El ejemplo del cartero que ha de distribuir 21 cartas entre 20 buzones nos permitirá ilustrar un principio numérico que a veces sirve para explicar la certeza de un determinado tipo de coincidencias. Como 21 es mayor que 20, puede estar seguro, sin necesidad de mirar previamente las direcciones, que por lo menos uno de los buzones tendrá más de una carta. Este principio de sentido común, que se conoce a veces como principio del casillero o de los cajones de Dirichlet, puede servir a veces para llegar a conclusiones que no son tan obvias.

Hay una tendencia general muy fuerte a olvidar los fracasos y concentrarse en los éxitos y los aciertos. Los casinos abonan esta tendencia haciendo que cada vez que alguien gana un cuarto de dólar en una máquina tragaperras, parpadeen las lucecitas y la moneda tintinee en la bandeja de metal. Con tanta lucecita y tanto tintineo, no es difícil llegar a creer que todo el mundo está ganando. Las pérdidas y los fracasos son silenciosos.

Aunque lo más llamativo sean los valores extremos y las coincidencias, lo que suele proporcionar más información son los valores medios o los valores esperados. El valor esperado de una cantidad es la media de los valores que toma, pesados según sus probabilidades respectivas.

Cuando le dan a uno una mano de bridge de trece cartas, la probabilidad de que le den precisamente esa mano concreta es menor que una seiscientos mil millonésima. Y a pesar de ello, será absurdo que, después de recoger las trece cartas, esa persona las examine detenidamente, calcule que la probabilidad de tener precisamente esas trece cartas es menor que una seiscientos mil millonésima y concluya que no puede ser que le hayan dado precisamente esa mano porque es muy improbable que esto ocurra.

La mayoría de la gente suele pensar como si las desviaciones de la media estuvieran atadas a una banda elástica, de modo que, cuanto mayor fuera la desviación, mayor sería la fuerza recuperadora que tendiese a restaurar la media. La creencia errónea de que el hecho de que hayan salido varias caras seguidas hace más probable que la próxima vez salga cruz se conoce como sofisma del jugador (las mismas ideas valen para la ruleta y los dados).

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 aumento de los beneficios de las sociedades, el aumento de los tipos de interés o lo que sea para explicar los giros alcistas. Un comentarista casi nunca dice que la actividad de la bolsa de ese día o de tal semana ha obedecido, por lo general, a fluctuaciones aleatorias.

A final de cuentas, una gran parte de la población (debido a que esta parte odia de alguna manera las matemáticas y todo lo que tenga que ver con ellas) se deja llevar por pensamientos que les genera una sorpresa o indignación en cuanto fenómenos que comúnmente llamamos casualidades. Si se estiman los parámetros de la distribución (se puede asumir es normal) uno creea que obtiene una campana “perfecta”. Sin embargo, al construir el histograma de la frecuencia de dicho objeto uno puede observar un fenómeno llamado colas anchas, es decir, aquellos sucesos que se encuentran más lejos de la media suelen ocurrir con mucha frecuencia. Éste fenómeno es más común de lo que uno cree (por ejemplo, Mandelbrot).

El humano mortal ahora goza de muchos accesos a la información, sin embargo sigue siendo muy obstinado y arrogante al analizarla (si es que realmente hace eso), creyendo o refutando principios que le suenan “absurdos o sorprendentes a primera vista”. No se debe caer en esa tentación de refutar algo que se desconoce y es por eso que se debe estimular una cultura de aprendizaje y razonamiento matemático, que puede no a todos les gusten. Siguen apostando, siguen controlando nuestra voluntad a partir de los más bajos instintos del ser humano (en las apuestas por ejemplo, muchos desconocen el pago esperado de cada una) y éste tipo de conductas junto con la ignorancia matemática es lo que conlleva al abuso de los débiles. ¿Quién será nuestro salvador?.

19.4 Fork

Una bifurcación o fork, cuando se aplica en el contexto de un lenguaje de programación o un sistema operativo, hace referencia a la creación de una copia de sí mismo por parte de un programa, que entonces actúa como un “proceso hijo” del proceso originario, ahora llamado “padre”. Los procesos resultantes son idénticos, salvo que tienen distinto número de proceso (PID).

Más generalmente, una bifurcación en un entorno multihilo significa que un hilo de ejecución se bifurca.

```
#include <stdio.h>
#include <unistd.h>
#include <sys/types.h>
int main(void) \{
    pid_t idHijo;
    pid_t idPropio;
    idPropio = getpid(); //Se obtiene el id del proceso actual
    idHijo = fork(); //Se crea un proceso 'hijo'
    if (idHijo == -1) { //Si hay un código menor que cero, hubo un error
        printf("Error al realizar la bifurcación"); //Se notifica al usuario
        return 1; //Se interrumpe la ejecución del proceso con una salida distinta a cero
    }
    if (idHijo == 0)
        printf("Soy el hijo con id %ld id proceso original %ld\n", (long)getpid(), (long)idPropio);
    else //la ejecución de la llamada al sistema fork devuelve el identificador al proceso 'padre'
        printf("Soy el padre con id %ld id proceso original %ld\n", (long)getpid(), (long)idPropio);
    return 0;
    \}
```

19.5 Problemas selectos

1. Cierta proceso tiene la propiedad de que, sin importar que haya transcurrido en un intervalo $[0, t]$, la probabilidad de que un evento tome lugar en el intervalo $[t, (t + h)]$ es λh . Asuma que la probabilidad de que más de un evento es de orden mayor en h . Determine la probabilidad de que al tiempo t , n eventos hayan sucedido, pasando al límite donde h tiende a cero. Evalúe el valor promedio de n y n^2 para la función de distribución.

Suponemos que han sucedido n eventos en el intervalo $[t, t + h]$, es decir, h se divide en N partes, de las cuales n se observa dicho suceso y en $N - n$ no, es decir, la probabilidad de que sucedan n eventos es

$$P(X = n) = (\lambda h)^n (1 - \lambda h)^{N-n}$$

donde $N = \frac{t}{h}$ es el número de intervalos. Sin embargo, se necesitan considerar todas las combinaciones posibles para elegir n intervalo de los N totales:

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

por lo que, finalmente, la probabilidad es

$$P(X = n) = \frac{N!}{n!(N-n)!} (\lambda h)^n (1 - \lambda h)^{N-n}$$

Cuando $N \rightarrow \infty$ se tiene que

$$\begin{aligned} \lim_{N \rightarrow \infty} \binom{N}{n} p^n (1-p)^{N-n} &= \lim_{N \rightarrow \infty} \frac{N(N-1) \cdots (N-n+1)}{n!} \left(\frac{\lambda t}{N}\right)^n \left(1 - \frac{\lambda t}{N}\right)^{N-n} \\ &= \lim_{N \rightarrow \infty} \frac{(\lambda t)^n}{n!} \left(1 - \frac{\lambda t}{N}\right)^N \frac{N(N-1) \cdots (N-n+1)}{n^x} \left(1 - \frac{\lambda t}{N}\right)^{-n} \\ &= \frac{(\lambda t)^n}{n!} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda t}{N}\right)^N \left(1 - \frac{\lambda t}{N}\right)^{-n} \left(1 - \frac{1}{N}\right) \cdot \left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{N-1}{N}\right) \end{aligned}$$

Por otro lado, por definición

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda t}{N}\right)^n = e^{-\lambda t}$$

junto con el hecho de que todos los términos a la derecha del límite tienden a 1 cuando n tiende a infinito, se obtiene

$$p(x) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

Por definición

$$E(X = n) = \sum_n n p(n) = \sum_{n=0}^{\infty} n \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

El primer término de esta suma es igual a cero, por lo que

$$E(X = n) = \sum_n n p(n) = \sum_{n=1}^{\infty} n \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \sum_{n=1}^{\infty} n \frac{(\lambda t)^n}{(n-1)!} e^{-\lambda t}$$

Al hacer el cambio de variable $y = n - 1$, la expresión anterior es una función de probabilidad. Los límites de la sumatoria se convierten en $y = 0$ cuando $n = 1$ y $y = \infty$ cuando $n = \infty$. Entonces

$$E(X = n) = \lambda t \sum_{y=0}^{\infty} \frac{(\lambda t)^y}{y!} e^{-\lambda t}$$

El truco es que ésta expresión es una función de probabilidad, siendo más específicos, resulta ser la misma distribución de Poisson pero con diferente notación, y $\sum_{y=0}^{\infty} p(y) = 1$. Por lo tanto

$$\mu = E(X) = \lambda t$$

Para calcular el valor medio de n^2 , se puede usar el siguiente truco: calcular $E[X(X-1)]$. Usando la definición de esperanza

$$\begin{aligned} E[X(X-1)] &= \sum_{n=0}^{\infty} n(n-1) \frac{(\lambda t)^n}{n!} e^{-\lambda t} \\ &= \sum_{n=2}^{\infty} \frac{(\lambda t)^n}{(n-2)!} e^{-\lambda t} \\ &= (\lambda t)^2 \sum_{y=0}^{\infty} \frac{(\lambda t)^y}{y!} e^{-\lambda t} = (\lambda t)^2 \end{aligned}$$

Nótese que se utilizó el mismo recurso que el de la demostración de la media. Ahora, viene lo interesante. Como la esperanza es un operador lineal, se tiene que

$$E[X(X-1)] = E[X^2] - E[X] = \lambda^2$$

o visto de otra forma

$$E[X^2] = \lambda^2 + E[X] = (\lambda t)^2 + \lambda t$$

2. Existen cerca de 6500 estrellas que se pueden observar a simple vista. De vez en cuando, dos estrellas aparecen muy juntas, aunque al examinarlas cuidadosamente no existe una conexión física entre ellas dos. Dicha pareja se le conoce como estrella óptica doble.

- a) Asumiendo que las estrellas se distribuyen al azar en el esfera celeste, compute el valor de expectación del número de estrellas ópticas dobles con una separación de no más de de $1'$ de arco.
- b) ¿Cuál es la probabilidad de que existan precisamente 2 estrellas ópticas dobles.
- c) Estime toscamente la probabilidad de una estrella óptica triple.

Dos estrellas cuya separación angular es $\theta = 1'$ yace dentro de un ángulo sólido $\omega \approx \pi\theta^2$. Por lo tanto, la probabilidad de las estrellas de una pareja en particular es $p = \omega/4\pi \approx 2.1 \times 10^{-8}$. Si la estrella doble se define como el que las estrellas estén aisladas, es decir, no forme parte de una tripleta, etc, entonces la probabilidad de que dos estrellas formen una pareja es

$$p(1-p)^{N-2}$$

donde $N = 6500$. El factor $(1-p)^{N-2}$ representa la probabilidad de las estrellas sobrantes no se combinen con la pareja para formar una tripleta, etc. Sin embargo, si se incluye como estrellas dobles a las que forman tripletas, etc., la probabilidad es simplemente p . En este problema $Np \ll 1$ por lo que ambas definiciones llevan a una probabilidad p .

El número de parejas independientes uno puede formar de las N estrellas es $N(N-1)/2$. Por lo tanto el valor esperado de estrellas dobles es

$$\frac{N(N-1)}{2}p \approx \frac{N^2}{2}p = \lambda = 0.45$$

(b) La probabilidad de que dos parejas de estrellas formen dos dobles estrellas distintas mientras que las estrellas restantes no formen múltiples estrellas es

$$P_{22} = p^2(1-p)(1-2p) \cdots [1 - (N-3)p]$$

Esto es porque, al añadir la n -ésima estrella a la esfera celestial donde $n \leq 4$, la fracción de la esfera que debe ocupar sin combinarse con cualquiera de las 4 estrellas elegidas o las otras estrellas es $1 - (n-3)p$. Se debe notar que, si $(N-3)p \geq 1$, la probabilidad es cero. Esto es por que en éste caso las estrellas están tan densas que es imposible mantenerlas separadas para que formen solo dos estrellas dobles. La probabilidad de ver de dos estrellas no es más que una distribución de Poisson

$$DP_{22} = \frac{1}{2} \left(\frac{N^2 p}{2} \right)^2 e^{-\lambda} = \frac{1}{2} \lambda^2 e^{-\lambda} = 0.063$$

(c) Considere tres estrellas etiquetadas como α, β, γ . La probabilidad de que las estrellas β y γ se encuentren a $1'$ de α es p^2 . Si la tripleta es definida solo cuando tres estrellas estén aisladas de otras estrellas, entonces la probabilidad de que las tres formen la tripleta es $p^2(1-p)^{N-3}$.

El número de tripletas independientes uno puede formar de las N estrellas es

$$\frac{N!}{3!(N-3)!} \approx N^3/3!$$

Por lo tanto la probabilidad de observar una tripleta es $N^3 p^2 / 3! = 2 \times 10^{-5}$

20 Tarea 20

20.1 Eigenvalores y Eigenvectores

Para todos los ejercicios, calcular los valores propios y vectores característicos de las siguientes matrices.

1.

$$A = \begin{pmatrix} 4 & 2 \\ 3 & 3 \end{pmatrix}$$

Primero se calcula $|A - \lambda I| = 0$, es decir

$$|A - \lambda I| = \begin{vmatrix} 4 - \lambda & 2 \\ 3 & 3 - \lambda \end{vmatrix} = \lambda^2 - 7\lambda + 6 = 0$$

cuyas raíces son $\lambda_1 = 6$ y $\lambda_2 = 1$. Para $\lambda_1 = 6$ se resuelve el sistema $(A - 6I)v = 0$ se tiene como solución $x_1 = x_2$ o bien $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Para el caso de $\lambda_2 = 1$, se resuelve el sistema $(A - I)v = 0$ que satisface la ecuación $3x_1 + 2x_2 = 0$ y un vector posible es $v_2 = \begin{pmatrix} 2 \\ -3 \end{pmatrix}$.

2.

$$A = \begin{pmatrix} 2 & -1 \\ -4 & 2 \end{pmatrix}$$

Se calcula $|A - \lambda I| = 0$, es decir

$$|A - \lambda I| = \begin{vmatrix} 2 - \lambda & -1 \\ -4 & 2 - \lambda \end{vmatrix} = \lambda^2 - 4\lambda = 0$$

cuyas soluciones son $\lambda_1 = 0$ y $\lambda_2 = 4$. El espacio característico de 0 es simplemente el kernel o núcleo de A . Se calcula

$$\begin{pmatrix} 2 & -1 \\ -4 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

o bien $2x_1 = x_2$ y un vector característico es $v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Para el segundo vector se tiene que

$$\begin{pmatrix} -2 & -1 \\ -4 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

por lo que $v_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$.

3.

$$A = \begin{pmatrix} 3 & -5 \\ 1 & -1 \end{pmatrix}$$

Se calcula de $|A - \lambda I| = 0$:

$$|A - \lambda I| = \begin{vmatrix} 3 - \lambda & -5 \\ 1 & -1 - \lambda \end{vmatrix} = \lambda^2 + 2\lambda + 2 = 0$$

cuyas soluciones son $\lambda = 1 \pm i$. De esta manera, si $\lambda_1 = 1 + i$ se tiene que

$$\begin{pmatrix} 2 - i & -5 \\ 1 & -2 - i \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

y se obtiene $-(2 + i)x_2 + x_1 = 0$. De esta manera $x_1 = (2 + i)x_2$, cuyo vector característico es $v_1 = \begin{pmatrix} 2 + i \\ 1 \end{pmatrix}$. De manera análoga para el caso donde $\lambda_2 = 1 - i$ se obtiene que

$$\begin{pmatrix} 2 + i & -5 \\ 1 & -2 + i \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

cuya solución del sistema esta dada por $x_1 + (-2 + i)x_2 = 0$, lo cual proporciona $x_1 = (2 - i)x_2$ y su vector característico es $v_2 = \begin{pmatrix} 2 - i \\ 1 \end{pmatrix}$.

4.

$$A = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$$

Se calcula $|A - \lambda I| = 0$, es decir

$$|A - \lambda I| = \begin{vmatrix} 4 - \lambda & 0 \\ 0 & 4 - \lambda \end{vmatrix} = (\lambda - 4)^2 = 0$$

Por lo tanto $\lambda = 4$ es un valor característico de multiplicidad algebraica 2. Es evidente que

$Av = 4v$ para todo vector $v \in \mathbb{R}^2$ por lo tanto los vectores característicos son $v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ y

$$v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

5.

$$A = \begin{pmatrix} 1 & -3 & -9 \\ 0 & 5 & 18 \\ 0 & -2 & -7 \end{pmatrix}$$

Entonces, $|A - \lambda I| = 0$:

$$|A - \lambda I| = \begin{vmatrix} 1 - \lambda & -3 & -9 \\ 0 & 5 - \lambda & 18 \\ 0 & -2 & -7 - \lambda \end{vmatrix} = -(\lambda + 1)^3 = 0$$

De esta manera, $\lambda = -1$ es un valor característico de multiplicidad algebraica 3. Para calcular los vectores característicos

$$(A - I)v = \begin{pmatrix} 0 & -3 & -9 \\ 0 & 6 & 18 \\ 0 & -2 & -6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Así, $-2x_2 - 6x_3 = 0$ o bien $x_2 = -3x_3$ y x_1 es arbitrario. Haciendo $x_1 = 0$ y $x_3 = 1$ se obtiene

$$v_1 = \begin{pmatrix} 0 \\ -3 \\ -1 \end{pmatrix} \text{ y haciendo } x_1 = 1 \text{ y } x_3 = 1, v_2 = \begin{pmatrix} 1 \\ -3 \\ -1 \end{pmatrix}.$$

20.2 Matrices similares

Sean $A, B \in M_n(F)$. Se dice que A y B son similares (o semejantes) si existe una matriz $P \in M_n(F)$ tal que P es invertible y

$$P^{-1}AP = B$$

En este caso escribimos $A \sim B$.

La relación de semejanza de matrices es reflexiva, simétrica y transitiva.

Demostración. 1. Demostremos que \sim es reflexiva. Sea $A \in M_n(F)$. Entonces la igualdad $P^{-1}AP = A$ se cumple, por ejemplo, con $P = I_n$. 2. Propiedad simétrica. Sean $A, B \in M_n(F)$

tales que $A \sim B$. Se tiene que demostrar que $B \sim A$. Sea $P \in M_n(F)$ tal que P es invertible y $P^{-1}AP = B$. Supongamos $Q = P^{-1}$. Entonces Q es invertible y $Q^{-1} = P$. Además de la igualdad $P^{-1}AP = B$ se obtiene que $A = PBP^{-1}$, esto es, $A = Q^{-1}BQ$. Acabamos de demostrar que $B \sim A$. 3. Propiedad transitiva. Sean $A, B, C \in M_n(F)$ tales que $A \sim B$ y $B \sim C$. Se va a demostrar que $A \sim C$. Usando la definición de matrices similares encontramos matrices $P, P \in M(F)$ tales que P y P son invertibles, $B = P_1^{-1}AP_1$ y $C = P_2^{-1}BP_2$. Sea $P_3 = P_1P_2$. Siendo un producto de matrices invertibles la matriz P_3 es invertible, su inversa es $P_3^{-1} = P_2^{-1}P_1^{-1}$, y

$$C = P_2^{-1}BP_2 = P_2^{-1}P_1^{-1}AP_1P_2 = P_3^{-1}AP_3$$

lo cual significa que $A \sim C$.

Sea V un espacio vectorial de dimensión finita sobre un campo F , sea $T \in L(V)$ y sean A y B algunas bases de V . Entonces las matrices asociadas TA y TB están relacionadas por la fórmula

$$T = P_{BA}^{-1}TA P_{PB}$$

por lo cual son semejantes: $TA \sim TB$.

Propiedades de matrices semejantes Proposición. Sean $A, B \in M_n(F)$ tales que $A \sim B$. Entonces:

1. $r(A) = r(B)$
2. $\det(A) = \det(B)$
3. A y B son ambas invertibles o ambas no invertibles.
4. Para cada $\lambda \in F$, $\lambda I_n - A \sim \lambda I_n - B$
5. $CA = CB$
6. $spA = spB$

21 Tarea 21

21.1 Matrices poco densas

Una matriz poco densa o dispersa, es una matriz cuyos elementos son mayormente cero. La fracción de elementos distintos de cero de la matriz poco densa sobre el número total de elementos igual a cero representa la densidad de dicha matriz.

Conceptualmente, las matrices poco densas representan sistemas que son debilmente conectados. El concepto de dispersión es especialmente útil en combinatoria y teoría de redes, ya que tiene una baja densidad de información útil o conexiones. Por otro lado, matrices dispersas grandes suelen aparecer al resolver ecuaciones diferenciales parciales.

Algunas matrices poco densas comunes son las matrices diagonales, las triangulares superiores e inferiores o las matrices en banda. Por ejemplo

$$\begin{pmatrix} 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 8 & 0 & 0 \\ 0 & 0 & 7 & 3 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

o bien

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

Como el concepto surge en el campo de álgebra lineal numérica, es de interés utilizar estructuras de datos eficientes para manejar de manera eficiente estas matrices. Para guardar una matriz dispersa se puede usar:

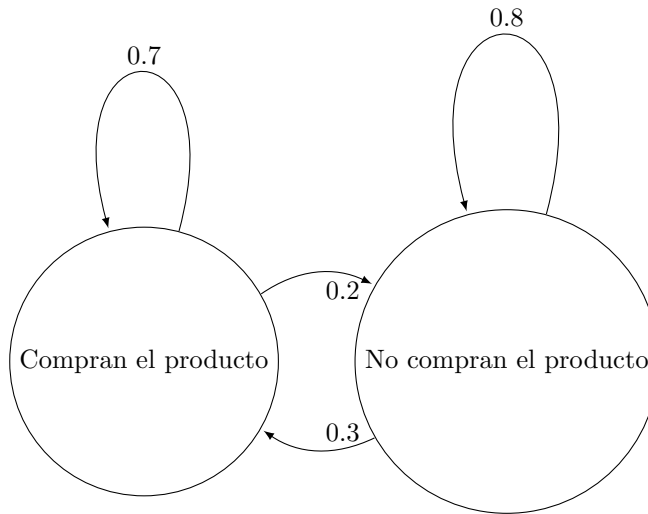
- Diccionario de llaves: consiste en un diccionario que mapea la columna y la llave al valor de los elementos. Los elementos faltantes en el diccionario se consideran cero.
- Lista de listas: almacenan una lista por fila, donde cada entrada contiene el índice de la columna y el valor.
- Lista de coordenadas: los valores de la matriz se guardan en una tripleta con un formato (fila, columna, valor)

Como se mencionó anteriormente, las matrices poco densas surgen al resolver ecuaciones diferenciales parciales, especialmente al utilizar el método de elemento finito.

El método de elementos finitos también tiene la flexibilidad necesaria para hacer frente a los problemas que varían rápidamente. Utilizando el método de elementos finitos para resolver EDPs implica los siguientes pasos:

1. Separar el dominio del problema en elementos discretos.
2. Formular la PDE en un problema variacional equivalente.
3. Construir un subespacio de dimensión finita y determinar su base en los elementos discretos.
4. Utilizar las funciones básicas en el problema variacional y formularlo en un sistema de ecuaciones lineales.
5. Atravesar todos los elementos discretos y montar la matriz de coeficientes.
6. Resolver el sistema de ecuaciones lineales para obtener la solución de la EDP originales.

La matriz de coeficientes es dispersa. Además, el tamaño de la matriz de coeficientes es grande con el fin de obtener una aproximación precisa a la solución de EDPs. Por lo tanto, las aplicaciones prácticas usando el método de elementos finitos siempre se basan en matrices dispersas y operaciones con matrices dispersas.



21.2 Cadenas de Markov

1. El departamento de estudios de mercado de una fábrica estima que el 20% de la gente que compra un producto un mes, no lo comprará el mes siguiente. Además, el 30% de quienes no lo compren un mes lo adquirirá al mes siguiente. En una población de 1000 individuos, 100 compraron el producto el primer mes. ¿Cuántos lo comprarán al mes próximo? ¿Y dentro de dos meses?

Solución

La cadena de Markov que modela el problema es:

De manera matricial

$$P^{(0)} = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}$$

Como se busca calcular el vector (x, y) donde x representa a las personas que comprarán el producto y y a las que no. Entonces, dentro de un mes comprarán

$$(x, y) = (100 \quad 900) \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix} = (350, 650)$$

Para calcular el número de personas que comprarán el producto dentro de dos meses, primero se calcula $P^{(2)}$:

$$\begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix} \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix} = \begin{pmatrix} 0.7 & 0.3 \\ 0.45 & 0.55 \end{pmatrix}$$

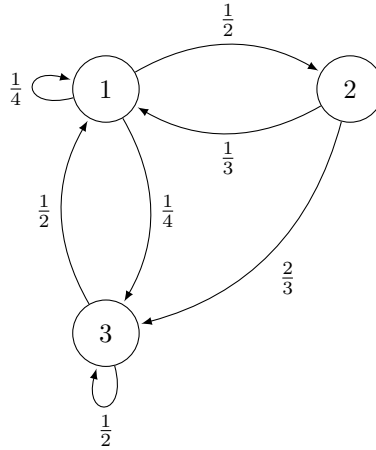
Finalmente

$$(x, y) = (100 \quad 900) \begin{pmatrix} 0.7 & 0.3 \\ 0.45 & 0.55 \end{pmatrix} = (475, 525)$$

2. Considere la cadena de Markov mostrada a continuación. Asuma que $X_0 = 1$, y sea R la primera vez que la cadena regresa al estado 1, es decir

$$R = \min\{n \geq 1 : X_n = 1\}.$$

Encuentre $E[R|X_0 = 1]$.



Solución

Sea r el valor esperado del tiempo para regresar al estado 1, entonces

$$r_1 = 1 + \sum_k t_k p_{1k},$$

donde t_k es el tiempo esperado hasta que la cadena llega al estado 1 dado que $X_0 = k$. De manera específica

$$\begin{aligned} t_1 &= 0, \\ t_k &= 1 + \sum_j t_j p_{kj}, \quad \text{for } k \neq 1. \end{aligned}$$

Se calculan los valores de t_k

$$\begin{aligned} t_2 &= 1 + \frac{1}{3}t_1 + \frac{2}{3}t_3 \\ &= 1 + \frac{2}{3}t_3, \end{aligned}$$

$$\begin{aligned} t_3 &= 1 + \frac{1}{2}t_3 + \frac{1}{2}t_1 \\ &= 1 + \frac{1}{2}t_3. \end{aligned}$$

Resolviendo las ecuaciones anteriores

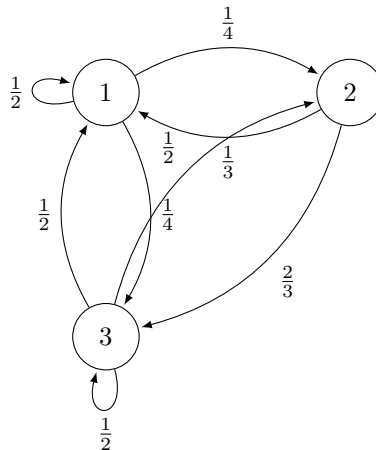
$$t_3 = 2, \quad t_2 = \frac{7}{3}.$$

Finalmente

$$\begin{aligned} r_1 &= 1 + \frac{1}{4}t_1 + \frac{1}{2}t_2 + \frac{1}{4}t_3 \\ &= 1 + \frac{1}{4} \cdot 0 + \frac{1}{2} \cdot \frac{7}{3} + \frac{1}{4} \cdot 2 \\ &= \frac{8}{3}. \end{aligned}$$

3. Considere la cadena de Markov de la siguiente figura

- ¿La cadena es irreducible?
- ¿La cadena es aperiódica?
- Encuentre la distribución estacionaria para esta distribución.
- ¿La distribución estacionaria es una probabilidad límite para la cadena?



Solución

- La cadena es irreducible ya que se puede ir de cualquier estado a cualquier estado en un número finito de pasos.
- La cadena es aperiódica ya que existe un loop en el estado 1 cuya probabilidad es mayor que cero.

c. Para encontrar la distribución estacionaria se resuelve el siguiente sistema

$$\begin{aligned}\pi_1 &= \frac{1}{2}\pi_1 + \frac{1}{3}\pi_2 + \frac{1}{2}\pi_3, \\ \pi_2 &= \frac{1}{4}\pi_1 + \frac{1}{2}\pi_3, \\ \pi_3 &= \frac{1}{4}\pi_1 + \frac{2}{3}\pi_2, \\ \pi_1 + \pi_2 + \pi_3 &= 1.\end{aligned}$$

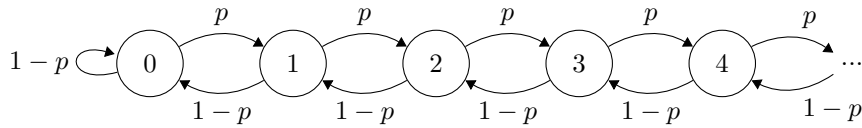
cuyas soluciones son

$$\pi_1 \approx 0.457, \pi_2 \approx 0.257, \pi_3 \approx 0.286$$

d. Como la cadena es irreducible y aperiódica se sigue que la distribución estacionaria es una distribución límite para la cadena.

4. Considere la cadena de Markov que se muestra en la siguiente figura. Asuma que $\frac{1}{2} < p < 1$. ¿La cadena siguiente tiene una distribución límite? Para todo $i, j \in \{0, 1, 2, \dots\}$ encuentre

$$\lim_{n \rightarrow \infty} P(X_n = j | X_0 = i).$$



Solución

La cadena es irreducible ya que todos los estados se comunican entre ellos y también es aperiódica ya que contiene un loop en el estado 0 cuya probabilidad es mayor que cero. Basta con escribir las ecuaciones para la distribución estacionaria. Para el estado 0 se tiene que

$$\pi_0 = (1-p)\pi_0 + (1-p)\pi_1,$$

que resulta

$$\pi_1 = \frac{p}{1-p}\pi_0.$$

Para el estado 1

$$\begin{aligned}\pi_1 &= p\pi_0 + (1-p)\pi_2 \\ &= (1-p)\pi_1 + (1-p)\pi_2,\end{aligned}$$

que resulta

$$\pi_2 = \frac{p}{1-p}\pi_1.$$

Como se puede observar, se encuentra un patrón de la forma

$$\pi_j = \alpha \pi_{j-1},$$

donde $\alpha = \frac{p}{1-p}$. Como $\frac{1}{2} < p < 1$, se concluye que $\alpha > 1$. Se obtiene

$$\pi_j = \alpha^j \pi_0, \quad \text{para } j = 1, 2, \dots$$

Finalmente, se debe obtener que

$$\begin{aligned} 1 &= \sum_{j=0}^{\infty} \pi_j \\ &= \sum_{j=0}^{\infty} \alpha^j \pi_0, & (\text{donde } \alpha > 1) \\ &= \infty \pi_0. \end{aligned}$$

Por lo tanto, la ecuación no se puede satisfacer si $\pi_0 > 0$. Si $\pi_0 = 0$, se sigue que todo $\pi_j = 0$ por lo que la suma no es igual a 1. Esto implica que, o todos los estados son transitivos o recurrentes nulos. En cualquier caso se tiene que

$$\lim_{n \rightarrow \infty} P(X_n = j | X_0 = i) = 0, \text{ para todo } i, j$$

5. Considera la cadena de Markov con tres estados $S = \{1, 2, 3\}$ que tiene la siguiente matriz de transición

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

Si se sabe que $P(X_1 = 1) = P(X_1 = 2) = \frac{1}{4}$ encuentre $P(X_1 = 3, X_2 = 2, X_3 = 1)$.

Solución

Primero se obtiene

$$\begin{aligned} P(X_1 = 3) &= 1 - P(X_1 = 1) - P(X_1 = 2) \\ &= 1 - \frac{1}{4} - \frac{1}{4} \\ &= \frac{1}{2} \end{aligned}$$

Finalmente

$$\begin{aligned} P(X_1 = 3, X_2 = 2, X_3 = 1) &= P(X_1 = 3) \cdot p_{32} \cdot p_{21} \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} \\ &= \frac{1}{12} \end{aligned}$$

21.3 Teorema del punto fijo

En matemáticas, un teorema del punto fijo es un teorema que especifica condiciones bajo las cuales se puede afirmar que una función f sobre un dominio dado (con rango en el mismo dominio) tiene, al menos, un punto fijo; es decir, que existe un punto x en dicho dominio para el cual: $f(x) = x$.

En diferentes ámbitos de estudio (espacios topológicos, espacios de Banach, espacios métricos, espacios euclídeos, etc) hay diferentes teoremas del punto fijo.

21.3.1 Teorema del punto fijo de Banach

En análisis matemático el teorema del punto fijo de Banach (también llamado teorema de la aplicación contractiva) es una de las herramientas más importantes para demostrar la existencia de soluciones de numerosos problemas matemáticos. El teorema garantiza la existencia y unicidad de puntos fijos de ciertas funciones definidas sobre espacios métricos y proporciona un método para encontrarlos.

Sea (X, d) un espacio métrico completo y f una aplicación. Se dice que f es contractiva si existe una constante K con $0 < K < 1$ tal que $d(f(x), f(y)) \leq Kd(x, y)$ para cualesquiera $x, y \in X$. Un punto fijo x_0 de f es un punto de X tal que $f(x_0) = x_0$. Entonces el teorema del punto fijo de Banach dice:

Sea (X, d) un espacio métrico completo y sea $f : X \rightarrow X$ una aplicación contractiva en X . Entonces existe un único punto fijo de f .

Además, el teorema establece que para todo punto x de X la sucesión $x, f(x), f(f(x)), \dots$ converge a dicho punto fijo.

21.3.2 Teorema del punto fijo de Brouwer

La forma más simple del teorema de Brouwer asume por hipótesis que la función f está definida sobre un intervalo cerrado y acotado, de extremos diferentes, J en sí mismo. De manera más general, la función está definida sobre un conjunto convexo y compacto K de un espacio euclídeo y a valores en K .

Existen varias versiones del teorema, según el contexto de utilización. La más simple toma a veces la forma siguiente:

En el plano: Toda aplicación continua f de un disco cerrado en sí mismo admite al menos un punto fijo.

Es posible generalizarlo a cualquier dimensión finita:

En un espacio euclídeo: Toda aplicación continua de una bola cerrada de un espacio euclídeo en sí misma admite un punto fijo.

Puede ser aún más general:

En un convexo compacto: Toda aplicación continua de un convexo compacto no vacío K de un espacio euclídeo a valores en K , admite un punto fijo.

Una formulación aún más general es conocida como:

Teorema del punto fijo de Schauder: Toda aplicación continua de un convexo compacto no vacío K de un espacio de Banach, en K , admite un punto fijo.

22 Tarea 22

22.1 Clasificación de la inteligencia artificial

La inteligencia artificial (IA) es una de las ramas de la informática, con fuertes raíces en otras áreas como la lógica y las ciencias cognitivas. Existen muchas definiciones de lo que es la inteligencia artificial. Sin embargo, todas ellas coinciden en la necesidad de validar el trabajo mediante programas.

H.A. Simon, uno de los padres de la IA, nos sirve de ejemplo, pues afirmó, en un artículo en 1995, que “el momento de la verdad es un programa en ejecución”. Las definiciones difieren en las características o propiedades que estos programas deben satisfacer.

La inteligencia artificial nace en una reunión celebrada en el verano de 1956 en Dartmouth (Estados Unidos) en la que participaron los que más tarde han sido los investigadores principales del área. Para la preparación de la reunión, J. McCarthy, M. Minsky, N. Rochester y C. E. Shannon redactaron una propuesta en la que aparece por primera vez el término “inteligencia artificial”. Parece ser que este nombre se dio a instancias de J. McCarthy.

La propuesta citada más arriba de la reunión organizada por J. McCarthy y sus colegas incluye la que puede considerarse como la primera definición de inteligencia artificial. El documento define el problema de la inteligencia artificial como aquel de construir una máquina que se comporte de manera que si el mismo comportamiento lo realizara un ser humano, este sería llamado inteligente.

Existen, sin embargo, otras definiciones que no se basan en el comportamiento humano. Son las cuatro siguientes.

1. Actuar como las personas. Esta es la definición de McCarthy, donde el modelo a seguir para la evaluación de los programas corresponde al comportamiento humano. El llamado Test de Turing (1950) también utiliza este punto de vista. El sistema Eliza, un bot (programa software) conversacional es un ejemplo de ello.
2. Razonar como las personas. Lo importante es cómo se realiza el razonamiento y no el resultado de este razonamiento. La propuesta aquí es desarrollar sistemas que razonen del mismo modo que las personas. La ciencia cognitiva utiliza este punto de vista.

3. Razonar racionalmente. En este caso, la definición también se focaliza en el razonamiento, pero aquí se parte de la premisa de que existe una forma racional de razonar. La lógica permite la formalización del razonamiento y se utiliza para este objetivo.
4. Actuar racionalmente. De nuevo el objetivo son los resultados, pero ahora evaluados de forma objetiva. Por ejemplo, el objetivo de un programa en un juego como el ajedrez será ganar. Para cumplir este objetivo es indiferente la forma de calcular el resultado.

Además de las definiciones mencionadas más arriba, hay aún otra clasificación de la inteligencia artificial según cuáles son los objetivos finales de la investigación en este campo. Son la inteligencia artificial fuerte y la débil.

La inteligencia artificial débil (también llamada estrecha) se define como la inteligencia artificial racional que se centra típicamente en una tarea estrecha. Siri es un buen ejemplo de la inteligencia estrecha. Siri opera dentro de un rango limitado previamente definido, no hay ninguna inteligencia genuina, sin conciencia, sin vida, a pesar de ser un ejemplo sofisticado de IA débil.

Se considera que las computadoras únicamente pueden simular que razonan, y únicamente pueden actuar de forma inteligente. Los partidarios de la inteligencia artificial débil consideran que no será nunca posible construir ordenadores conscientes, y que un programa es una simulación de un proceso cognitivo pero no un proceso cognitivo en sí mismo.

La Inteligencia Artificial Fuerte es aquella inteligencia artificial que iguala o excede la inteligencia humana promedio, la inteligencia de una máquina que exitosamente puede realizar cualquier tarea intelectual de cualquier ser humano. Es un objetivo importante para la investigación sobre IA y un tópico interesante para la ciencia ficción.

Muchas definiciones de inteligencia han sido propuestas, como el que se pase la prueba de Turing, pero hasta ahora no existe una definición que satisfaga todas. Sin embargo, existen ciertos convenios entre investigadores de inteligencia artificial en los cuales se especifica qué es lo que debe realizar cierta inteligencia:

- Razonar, resolver problemas y hacer juicios bajo incertidumbre (tomar decisiones)
- Representar el conocimiento, incluyendo el sentido común
- Planea
- Aprende
- Se comunica usando lenguaje natural
- Integrar las herramientas anteriores para lograr metas en común

Otras habilidades incluyen la habilidad de sentir y la habilidad de interactuar con objetos en un ambiente “intelectual”. Para confirmar AGI se han propuesto distintas pruebas:

- Prueba de Turing
- Prueba del café de Goertzel: una máquina debe hacer un café, desde encontrar la máquina, sacar el café, la taza, etc.
- Prueba del robot estudiante: un estudiante debe atravesar todo el proceso de matricularse en una universidad, aprobando todas las materias y obtener un título

- Prueba del empleado: a una máquina se le da un trabajo económicamente importante y debe resolverlo mejor que la contraparte humana

22.2 Problema de los 4 colores

A mediados del siglo XIX Francis Guthrie se dio cuenta mientras coloreaba un mapa de los condados de Inglaterra de que necesitaba al menos cuatro colores para que se cumpliera la condición de que dos regiones con frontera común tuvieran colores distintos (si dos regiones se tocan en un único punto se entiende que no tienen frontera común). Francis le comentó el tema a su hermano Frederick, que a su vez se lo planteó a Augustus de Morgan (profesor suyo en un curso de matemáticas en aquel momento), que aunque no supo responderle se encargó de difundir el asunto entre otros matemáticos. En 1878 Arthur Cayley lo presenta formalmente a la London Mathematical Society y así el problema queda abierto con un enunciado como éste:

Definición 1. *Todo mapa plano puede colorearse con, como máximo, cuatro colores con la condición de que regiones con frontera común tengan colores distintos.*

El año siguiente, 1879, es una fecha importante en relación con este problema. Ese año Alfred Kempe publica una demostración del mismo. En efecto parece ser que con cuatro colores era suficiente y el problema estaba resuelto. Y así fue hasta 1890, año en el que Percy Heawood encontró un error insalvable en la demostración de Kempe, por lo que el problema volvía a estar abierto. A partir de aquí muchos matemáticos (entre ellos el propio Heawood) atacaron el problema, pero ninguno de ellos consiguió dar con la respuesta.

Pero todos esos intentos fallidos no fueron en vano. Por el camino quedaron demostraciones de que para colorear un mapa dibujado en un toro hacen falta, como máximo, siete colores, y que para colorear un mapa en una banda de Möbius hacen falta, a lo sumo, seis colores. También se demostró que cinco colores eran suficientes para un mapa plano. Pero parecía que todos los mapas podían colorearse con cuatro, que no hacían falta esos cinco.

Hemos comentado al principio que este problema pertenece a la teoría de grafos, pero no hemos dicho cómo relacionar mapas con grafos. Lo que se hace a partir de cada mapa plano es calcular su grafo dual, que se construye asignando un vértice a cada región y uniendo dos vértices con una arista si en el mapa las dos regiones correspondientes a dichos vértices tenían frontera común.

De esta forma el problema queda planteado de la siguiente forma:

¿Es cierto que los vértices de todo grafo plano pueden colorearse con, a lo sumo, cuatro colores de forma que dos vértices unidos por una arista tengan colores distintos?

Con todo esto entramos en el siglo XX y sobre 1950 se comienza a pensar que las computadoras podrían ser de gran ayuda en este problema. El matemático alemán Heinrich Heesch fue uno de los pioneros en este sentido, y sus investigaciones acabaron siendo fundamentales para el desenlace del asunto. Pero los auténticos protagonistas de la demostración del teorema de los cuatro colores son Kenneth Appel y Wolfgang Haken. Ellos fueron los que, en 1976, anunciaron que “Cuatro colores son suficientes”. La demostración que presentaron no es “habitual”, ya que una buena parte de la misma

se realizó con ayuda de una computadora. Más adelante se presentaron mejoras a la demostración de Appel y Haken, y hasta hay alguna propuesta que no utiliza alguna computadora.

22.3 Experimentos de Mendel

Gregor Mendel desarrolló el modelo de herencia que ahora lleva su nombre por experimentos sobre diversas características de plantas de guisante: altura (largo vs. corto); color de la semilla (amarillo vs. verde); la capa de asiento (liso vs arrugada), etc. La siguiente explicación utiliza el rasgo de alto / bajo. Los otros rasgos Mendel estudió pueden ser sustituidos por alto y bajo.

Mendel comenzó con plantas que "germinan puramente". Es decir, cuando las plantas altas fueron autopolinizadas (o polinizadas cruzadamente con otras como ellas), las plantas en las generaciones siguientes eran todas altas; cuando las plantas bajas se auto-polinizan (o polinizadas con otras como ellas) las siguientes generaciones eran cortas.

Mendel descubrió que si se cruzan una línea genéticamente pura Tall [T] con plantas de cría pura cortas [t], toda la nueva generación de plantas, denominados $F1$, son todos altos. A continuación, se demostró que las plantas autógamas $F1$ (o cruzada con otras plantas polinizadas $F1$) producen una generación $F2$ con $3/4$ de las plantas altas y $1/4$ corto.

$1/4$ de la generación $F2$ son plantas cortas, que producen sólo las plantas cortas en la generación $F3$, si son autopolinizadas (o cruzas con otras plantas $F2$ cortas;) estas plantas $F2$ se reproducen puras.

$1/4$ de la generación $F2$ ($1/3$ de las plantas altas) son plantas altas que producen sólo plantas altas en la generación $F3$, si son autopolinizadas estas plantas $F2$ altas se reproducen puras.

$1/2$ de la generación $F2$ ($2/3$ de las plantas altas) son plantas altas que producen $1/4$ de plantas cortas y $3/4$ de plantas altas en la generación siguiente ($F3$), si son autopolinizadas. Esta es la misma proporción de plantas cortas y altas que las plantas $F1$ producen.

El modelo de Mendel para la generación $F1$ se resume en la siguiente tabla. El modelo establece que cada rasgo es controlado por un par de paquetes hereditarios que ahora llamamos genes. Un paquete viene de cada padre. Los alelos del gen para la altura son los mismos en plantas puras reproductoras (plantas padres TT y tt). Una cruce mixta TT con plantas tt produce plantas Tt en la primera generación o $F1$. Las plantas $F1$ reciben un alelo T del padre alto y un alelo t del padre corto. Las plantas $F1$ son altas debido a que el alelo T se expresa y cubre el alelo t . Por lo que el alelo T se llama dominante y el alelo t se llama recesivo.

	T	t
T	TT	Tt
t	tT	tt

El siguiente diagrama muestra como el modelo de Mendel explica la relación $3 : 1$ de alto para plantas bajas en la generación $F2$. En la generación $F1$ cada planta tenía un alelo T y un alelo t del gen que controla la altura. Las plantas en la generación $F2$ tuvieron la oportunidad del 50% de conseguir una T o una t de cada planta madre. El diagrama muestra que esto se traduce en 1 de

cada 4 plantas que consiguen solamente genes t y 3 plantas que consiguen al menos un gen T (que hace a la planta alta, porque T domina a t).

El diagrama muestra también que la generación F_2 en realidad tiene tres tipos de plantas: $1/4$ son plantas tT , que son cortas y producen plantas solo cortas en las siguientes generaciones de autopolinización. De los $3/4$ plantas altas restantes, $1/4$ son TT , que son altos y producen solo plantas altas en las siguientes generaciones si se autopolinizan. El restante $2/4$ obtiene una T de un padre y un t del otro. Cuando se autopolinizan, se producen un patrón exactamente igual al de la generación F_1 : 1 planta corta por cada 3 plantas altas. Estas plantas son exactamente igual que la generación F_1 .

	T	T
t	Tt	Tt
t	tT	tT

Un cierto rasgo es determinado por un par de genes, el cual puede ser de dos tipos, digamos G y g . Entonces un individuo puede tener

- GG (dominante)
- gG o Gg (híbrido)
- gg (recesivo)

Al momento de heredar el gen del padre, éste es seleccionado al azar, independientemente del otro (gen). Ésto determina la probabilidad de ocurrencia de cada tipo de cruce. La cruce de:

- dos padres GG debe ser GG
- dos padres gg debe ser gg
- un padre G y el otro g debe ser Gg
- algún otro caso debe ser examinado con detalle

Primer caso: los padres son Gg y Gg . Como en todos los casos, cada padre contribuye con un gen. Por lo tanto existen 4 posibles resultados: GG , gG , Gg y gg , es decir

- $1/4$ de ser GG
- $1/4$ de ser gg
- $1/2$ de ser Gg

Segundo caso: los padres son GG y Gg .

- $1/2$ de ser GG
- $1/2$ de ser Gg

Tercer caso: los padres son gg u Gg

- $1/2$ de ser gg
- $1/2$ de ser Gg

Considere ahora un proceso de cruas continuas

- Comienza con un individuo de genética conocida o desconocida y se cruza con un híbrido
- Asuma que existen a lo menos un brote: eliga uno de ellos al azar y cruzelos con un híbrido
- Repita este proceso a través de un número de generaciones

El tipo de genética del brote escogido en generaciones sucesivas puede ser representado por una cadena de Markov, con estados GG , Gg y gg . Por lo que existen tres posibles estados: $S_1 = GG$, $S_2 = Gg$ y $S_3 = gg$.

	GG	Gg	gg
GG	0.5	0.5	0
Gg	0.25	0.5	0.25
gg	0	0.5	0.5

De forma matricial

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Se procede a calcular la distribución de probabilidad estacionaria, es decir, se resuelve el sistema

$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix}$$

que resulta en el siguiente sistema de ecuaciones

$$\begin{aligned} w_1 &= \frac{1}{2}w_1 + \frac{1}{4}w_2 \\ w_2 &= \frac{1}{2}w_1 + \frac{1}{2}w_2 + \frac{1}{2}w_3 \\ w_3 &= \frac{1}{4}w_2 + \frac{1}{2}w_3 \end{aligned}$$

que tiene como solución el vector $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.

References

- [1] J. K. Blitzstein y J. Hwang. *Introduction to Probability*. Chapman & Hall/CRC Texts in Statistical Science, CRC Press, 2014.
- [2] J. M. Pliego and L. R. M. PEREZ. *Fundamentos de Probabilidad*. Ediciones Paraninfo S.A., 2006
- [3] J. M. Pliego, L. R. M. Perez and J. M. Lorenzo *Problemas de Probabilidad*. Ediciones Paraninfo S.A., 2009
- [4] Wackerly, Mendenhall y Scheaffer *Estadística Matemática con Aplicaciones*. Cengage Learning, 2008.
- [5] Ralph P. Grimaldi *Matemáticas discretas y combinatoria*. Pearson Education, 2009.
- [6] William Feller *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons, 1957
- [6] Sahil Mohnani *The trinomial theorem and Pascal's Tetrahedron*. <https://sahilmohnani.wordpress.com/2013/03/27/the-trinomial-theorem-and-pascals-tetrahedron/>
- [7] Francisco Bellot Rosado *Seminario de Problemas: Una primera lección de combinatoria*.
- [8] Resnick, Halliday, Krane *Física Volumen 1*. Pearson Education, 2000
- [9] Jesús Guillerá Goyanes *Historia de las fórmulas y algoritmos para π* . La Gaceta de la RSME (10), 159-178, Enero 2007
- [9] Elsa Vera and Maria A. Blasco *Beyond average: potential for measurement of short telomeres*. AGING 4 (6), 379-392, June 2012
- [10] Robert Hannum *Understanding Casino Math*. Disponible en: <http://gaming.unlv.edu/casinomath.html>
- [11] Murray R. Spiegel *Probabilidad y Estadística. Teora y 760 problemas resueltos*. McGraw-Hill, 1976
- [12] Charles E. Metz *Basic Principles of ROC Analysis*. Seminars in Nuclear Medicine 8, 283-298, October 1978
- [12] Avi Wigderson *P, NP and mathematics - a computational complexity perspective*. Disponible en : <http://www.math.ias.edu/avi/PUBLICATIONS/MYPAPERS/W06/w06.pdf>