

THE UNIVERSITY OF LOS ANDES  
BOOK

---

# Métodos Computacionales

---

*Manuel Alejandro Segura Delgado*

PhD



# Contents

<b>List of figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>ix</b>
<b>1 Derivación e integración numérica</b>	<b>1</b>
1.1 Taylor Series . . . . .	1
1.2 Error de redondeo . . . . .	2
1.3 Error de truncamiento . . . . .	2
1.4 Derivación . . . . .	4
1.4.1 Derivada Progresiva . . . . .	4
1.4.2 Derivada Regresiva . . . . .	5
1.4.3 Derivada Central . . . . .	5
1.4.4 Error Local . . . . .	6
1.5 Segunda Derivada . . . . .	6
1.6 Ejercicios: Derivación . . . . .	6
1.7 Método de Newton-Raphson . . . . .	11
1.7.1 Criterio de parada . . . . .	11
1.8 Método de bisección . . . . .	11
1.9 Ejercicios: Raíces de polinomios . . . . .	13
1.10 Interpolación de Lagrange . . . . .	14
1.10.1 Error . . . . .	15
1.11 Interpolación de Newton-Gregory . . . . .	15
1.12 Ejercicios: Interpolación de Lagrange . . . . .	17
1.13 Integración . . . . .	20
1.13.1 Método de trapecio simple . . . . .	20

1.13.2	Método de trapecio compuesto . . . . .	21
1.13.3	Método de Simpson simple 1/3 . . . . .	22
1.13.4	Método de Simpson compuesto . . . . .	24
1.13.5	Cuadratura Gaussiana . . . . .	25
1.13.6	Cuadratura de Gaus-Legendre . . . . .	27
1.13.7	Ejercicios: Integración . . . . .	30
<b>2</b>	<b>Método de MonteCarlo</b>	<b>37</b>
2.1	Generación pseudo-aleatoria de números . . . . .	37
2.2	Integración MonteCarlo . . . . .	38
2.3	Ley de grandes números . . . . .	39
2.4	Método de la transformada inversa . . . . .	40
2.5	Método de aceptación y rechazo . . . . .	41
2.6	Incertidumbre del método de MonteCarlo . . . . .	41
2.7	Ejercicios: Método de MonteCarlo . . . . .	42
<b>3</b>	<b>Álgebra lineal</b>	<b>47</b>
3.1	Método de Jacobi . . . . .	47
3.2	Norma matricial . . . . .	49
3.2.1	Radio espectral . . . . .	50
3.3	Método de Gauss-Seidel . . . . .	51
3.4	Método de relajación sucesiva . . . . .	52
3.5	Método generalizado de Newton-Raphson . . . . .	54
3.6	Método del descenso gradiente . . . . .	55
3.7	Propagacion del error en una red neuronal . . . . .	56
3.8	Ejercicios: Álgebra lineal . . . . .	58
<b>4</b>	<b>Probabilidad</b>	<b>69</b>
4.1	Definición de probabilidad . . . . .	69
4.1.1	Definición frecuentista de probabilidad . . . . .	71
4.2	Ejercicios: Axiomas de la probabilidad . . . . .	72
4.2.1	Independencia de eventos . . . . .	73
4.2.2	Probabilidad condicional . . . . .	73
4.2.3	Teorema de la probabilidad total . . . . .	74

4.3	Ejercicios: Probabilidad condicional y total . . . . .	76
4.4	Teorema de Bayes . . . . .	77
4.5	Ejercicios: Teorema de Bayes . . . . .	79
4.6	Técnicas de conteo . . . . .	80
4.7	Ejercicios: Técnicas de conteo . . . . .	82
4.8	Ejercicios: Generales de probabilidad . . . . .	83
4.9	Distribuciones discretas de probabilidad . . . . .	86
4.9.1	Covarianza . . . . .	88
4.9.2	Correlación . . . . .	89
4.9.3	Distribución de Bernoulli . . . . .	89
4.9.4	Distribución Binomial . . . . .	89
4.9.5	Distribución de Poisson . . . . .	90
4.10	Ejercicios: Distribuciones discretas de probabilidad . . . . .	90
4.11	Distribuciones continuas de probabilidad . . . . .	91
4.11.1	Distribución normal . . . . .	91
4.11.2	Distribución $\chi^2$ . . . . .	92
4.11.3	Distribución t-Student . . . . .	92
4.12	Ejercicios: Distribuciones continuas de probabilidad . . . . .	92
4.13	Procesos de Markov . . . . .	94
4.14	Ejercicios: Cadenas de Markov . . . . .	97
4.15	Metropolis-Hastings algorithm . . . . .	97
4.16	Ejercicios: Metropolis-Hastings . . . . .	99
4.17	Hidden Markov models . . . . .	103
4.18	Ejercicios: Hidden Markov models . . . . .	106
<b>5</b>	<b>Estadística</b> . . . . .	<b>109</b>
5.1	Muestreo (sampling) . . . . .	109
5.1.1	Aleatorio (probabilístico) . . . . .	110
5.1.2	Muestreo aleatorio simple . . . . .	111
5.1.3	Muestreo aleatorio estratificado . . . . .	111
5.1.4	Muestreo aleatorio sistemático . . . . .	112
5.1.5	Muestreo no aleatorio . . . . .	113
5.2	Distribución de frecuencias (Histogramas) . . . . .	113

5.3	Estadísticos . . . . .	115
5.4	Valor medio muestral . . . . .	115
5.5	Desviación estándar . . . . .	116
5.6	Teorema del límite central . . . . .	117
5.7	Ejercicios: Muestreo . . . . .	117
<b>6</b>	<b>Estimación de parámetros</b>	<b>121</b>
6.1	Mínimos cuadrados . . . . .	121
6.1.1	Espacio Columna . . . . .	121
6.2	Varianza de la estimación de parámetros . . . . .	124
6.3	Función de costo . . . . .	125
6.4	Ejercicios: Mínimos cuadrados . . . . .	125
6.5	Función de verosimilitud (Likelihood) . . . . .	129
6.6	Método de máxima verosimilitud . . . . .	130
6.6.1	Unbinned likelihood fit . . . . .	130
6.6.2	Binned likelihood fit . . . . .	131
6.6.3	Varianza de los estimadores . . . . .	132
6.7	Ejercicios: Parameter Estimation . . . . .	132
<b>7</b>	<b>Confidence Intervals</b>	<b>137</b>
7.1	Intervalos de confianza para la media . . . . .	137
7.2	Intervalo de confianza para la diferencia de medias . . . . .	138
7.3	Intervalo de confianza para una proporción . . . . .	138
7.4	Intervalo para la varianza . . . . .	138
7.5	Ejercicios: Intervalos de confianza . . . . .	139
<b>8</b>	<b>Pruebas de Hipótesis</b>	<b>143</b>
8.1	Ejercicios: Tablas de contingencia . . . . .	144
8.2	Prueba de hipótesis para generadores de números aleatorios . . . . .	146
8.3	Ejercicios: Hypotesis testing . . . . .	147
<b>References</b>		<b>155</b>

# List of Figures

1.1	Campo de velocidades cerca de un cilindro sólido. . . . .	8
1.2	Electric field of an electric charge close to grounded conducting plane. Note that the Gauss' Law is satisfied because of the electric field is normal in the surface. Nice! . . . . .	9
1.3	Interpolación usando el método de Newton-Gregory para el conjunto de datos descritos en el problema. . . . .	19
1.4	Domo esférico . . . . .	32
1.5	Accuracy vs n-points for the Gauss-Laguerre quadrature. . . . .	34
2.1	Correlaciones en los primeros k-vecinos para un mal generador (izquierda) y un buen generador (drand48) de números aleatorios. . . . .	39
2.2	Primeros $k = 10$ vecinos del generador Numpy como función del número de eventos generados. . . . .	43
2.3	Correlaciones de los primeros $k = 30$ vecinos del generador Numpy como función de $k$ -esimo vecino. Note que el valor debe fluctuar alrededor del valor teórico $C(k) = 1/4$ . . . . .	44
2.4	Generación de eventos cartesiana (izquierda) y esférica para $R = 1$ (derecha). . .	44
3.1	Esquema de colisión de dos discos rígidos. . . . .	60
3.2	Esquema para la conservación del momentum angular. . . . .	61
3.3	Conservación del momento lineal total antes y después del choque. . . . .	63
3.4	Conservación del momento angular total antes y después del choque. . . . .	63
3.5	Variables cinemáticas los dos discos después del choque. . . . .	64
3.6	Estrella naciente: Puntos originales (azul), los ejes principales del cuerpo (rojo) y puntos sobre los ejes principales (verde.) . . . . .	64

3.7	Esquema del espejo del telescopio espacial. Los vértices del cuadrado registran las temperaturas en grados kelvin. En la parte derecha, se ilustra una rotación para minimizar la temperatura en el punto $P = (0, 0.5)$ .	65
3.8	Mapa de temperatura al interior del espejo del telescopio.	66
4.1	Probabilidad $\mathbb{P}(n \leq 80)$ de que tengan una edad diferente.	85
4.2	Superficie de probabilidad del evento $A$ .	86
4.3	Probabilidades de transición de los estados para $n = 10$ pasos.	97
4.4	Relación entre la distancia cuadrática media y el número de pasos de un proceso Markoviano. Entender este proceso fue uno de los pasos definitivos para aceptar la existencia de las moléculas empezando el siglo XX.	100
4.5	Nube electrónica encontrada con el algoritmo de Metropolis asociada al electrón en cada núcleo.	102
4.6	Valor esperado de la energía potencial eléctrica de la molécula de Hidrógeno. Note que existe un estado ligado que conocemos como enlace covalente, la energía del pozo de potencial está alrededor de $U_{min} \approx 2.0 \text{ eV}$ y se da a una separación nuclear de $d_{equilibrium} \approx 2.1a_0$ .	102
4.7	Valor esperado de la energía potencial eléctrica de la molécula de Hidrógeno incluyendo el efecto de Heitler y London. Note que existe un estado ligado que conocemos como enlace covalente, la energía del pozo de potencial está alrededor de $U_{min} \approx 2.05 \text{ eV}$ y se da a una separación nuclear de $d_{equilibrium} \approx 2.3a_0$ .	103
4.8	Probabilidad de cada secuencia oculta para el estado $\Omega_O = [V, A, R]$ .	105
4.9	Probabilidad de cada secuencia oculta para el estado observado $\Omega_O = [S, C, C, C, S, C, S, C]$ .	107
5.1	Distribución de muestreo (izquierda) y distribución acumulada de probabilidad (derecha).	119
6.1	Distancia $\ \mathbb{A}\vec{x}^* - \vec{b}\ $ en la región rectangular definida en el problema. La distancia mínima debe corresponder con el resultado del método de mínimos cuadrados.	126
6.2	Gráfica de curvas de nivel de la función $\mathcal{L}(\mu, \epsilon)$ (izquierda) y superficie (derecha).	135
6.3	Medida del equipo de 3 canales.	135
8.1	Trayectoria de la partícula (Toy model). En este experimento la partícula le tomo 33 pasos salir de la caja.	150

8.2 Ajuste a los datos usando la distribución de Weibull exponencial.	150
8.3 Comparación entre la distribución observada y esperada para el número de pasos de escape de la partícula.	151
8.4 Distribución $\chi^2_{99}$ , el valor observado es menor que el valor crítico de la distribución. $p - value = 0.111$ , se acepta $H_0$ .	152
8.5 Distribución $\chi^2_{999}$ , el valor observado es menor que el valor crítico de la distribución. $p - value = 0.190$ , se acepta $H_0$ .	154



# List of Tables

3.1	Parámetros del cálculo. Recuerde que el momento de inercia de un sólido respecto a su centro de masas es $I = kmr^2$ .	62
4.1	Tabla de contingencia de genero y el uso de gafas.	77
4.2	Distribución de probabilidad	87
4.3	Distribución de probabilidad del número de microchips defectuosos.	90
5.1	Distribución de estudiantes según su carrera universitaria.	111
5.2	Distribución muestral según su carrera universitaria.	112
5.3	Distribución de pesos y distribución acumulada de pesos para un conjunto de mediciones de 100 estudiantes.	114
5.4	Distribución de salarios para 65 empleados de la empresa anónima.	118
8.1	Tabla de contingencia entre artículos vendidos y los días de la semana	144
8.2	Tabla de contingencia asociada a la encuesta realizada por la universidad.	145



# Chapter 1

## Derivación e integración numérica

### Section 1.1

#### Taylor Series

Let  $f(x)$  be a function of class  $C^\infty$ . The Taylor polynomial is defined as:

$$p_n(x) = \sum_{n=0}^{\infty} a_n(x - x_0)^n, \quad (1.1)$$

where  $x_0$  is a specific point, where one makes the expansion. Using  $x_0 = 0$  (Maclaurin serie), the representation of the  $f(x)$  is therefore given by:

$$f(x) = \sum_{n=0}^{\infty} a_n x^n. \quad (1.2)$$

One inside for obtaining the coefficients is driven by the value of the function and its derivatives; Hence:

$$\begin{aligned} f(0) &= a_0 + a_1 0 + a_2 0^2 + \dots \\ f'(0) &= a_1 + 2a_2 0 + \dots \\ f''(0) &= 2a_2 + 3 \times 2a_3 0 + \dots \end{aligned} \quad (1.3)$$

Note that,  $a_0 = f(0)$ ,  $a_1 = f'(0)/1!$ ,  $a_2 = f''(0)/2!$ , in general, we get:

$$a_n = \frac{f^{(n)}(x)}{n!} \quad (1.4)$$

This expression is known as the Maclaurin serie of  $f(x)$ .

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!} x^n. \quad (1.5)$$

For expanding in any arbitrary point (a), the expression is:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x-a)^n. \quad (1.6)$$

We are interesting in any point of the discretization ( $x'$ ) and its neighborhood  $x' + h$ , where  $h \ll 1$ . The expansion to around  $x'$  is given by:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x')}{n!} (x-x')^n. \quad (1.7)$$

Therefore, the value of the function in  $x' + h$  is:

$$f(x' + h) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x')}{n!} (x' + h - x')^n = \sum_{n=0}^{\infty} \frac{f^{(n)}(x')}{n!} (h)^n \quad (1.8)$$

Which is an useful formulae. the expansion up to first order is given by:

$$f(x + h) = f(x) + h f'(x) + \mathcal{O}(h^2) \quad (1.9)$$

### Section 1.2

## Error de redondeo

Es la diferencia entre el valor exacto de un número y la aproximación calculada debida al redondeo. Por ejemplo,  $\pi = 3.1415926535\dots$  si se aproxima a 3.1416 el error es  $7.3464 \times 10^{-6}$ , entonces la pregunta natural es: ¿cuál es el número más pequeño que podemos aproximar usando el computador? en otras palabras ¿cuál es el valor de  $\epsilon$  para que se cumpla  $1 + \epsilon = 1$ .

### Section 1.3

## Error de truncamiento

El error de truncamiento aparece cuando se usan aproximaciones en lugar de las expresiones exactas, en general, este tipo de error depende del tipo de aproximación que se realiza. Por

ejemplo, cuando expandimos una cierta función alrededor de un punto y despreciamos términos de orden superior, se introducen error de truncamiento a nuestras estimaciones.

$$\begin{aligned} \sin(x) &\cong x + \mathcal{O}(x^3) \\ \sin(x) &\cong x - \frac{x^3}{3!} + \mathcal{O}(x^5) \end{aligned} \tag{1.10}$$

tiene diferente error de truncamiento para la estimación de la función  $\sin(x)$ .

### Section 1.4

## Derivación

Para construir la derivada numérica se define la siguiente discretización de puntos nodales equi-espaciados.

$$x_j = x_0 + jh, \quad (1.11)$$

donde  $h$  se denomina paso, siendo una variación "pequeña" de la función. El conjunto de puntos nodales con sus respectivas imágenes se denomina conjunto soporte de la discretización:

$$\Omega = \{(x_0, f(x_0)), (x_1, f(x_1)), \dots, (x_n, f(x_n))\} \quad (1.12)$$

En el caso de más dimensiones, el conjunto soporte debe estar bien definido:

$$\Omega_2 = \{(x_0, y_0, f(x_0, y_0)), (x_1, y_1, f(x_1, y_1)), \dots, (x_n, y_n, f(x_n, y_n))\} \quad (1.13)$$

### 1.4.1 Derivada Progresiva

Dada esta condición podemos hacer una expansión en series de Taylor de  $f(x)$  (el dominio son los puntos nodales).

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \dots \quad \text{para} \quad h \ll 1 \quad (1.14)$$

despejando la primera derivada tenemos:

$$f'(x) = \frac{f(x+h) - f(x)}{h} - \underbrace{\frac{h}{2}f''(x)}_{\mathcal{O}(h)} \quad (1.15)$$

para algún punto de la partición:

$$f'(x_j) \cong \frac{f(x_{j+1}) - f(x_j)}{h} \quad (1.16)$$

La última expresión se denomina la derivada progresiva del punto  $x_j$ , la cuál tiene orden  $\mathcal{O}(h)$  en la estimación.

### 1.4.2 Derivada Regresiva

Para obtener la derivada regresiva se expande:

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2}f''(x) + \dots \quad \text{para } h \ll 1 \quad (1.17)$$

despejando la primera derivada tenemos:

$$f'(x) = \frac{f(x) - f(x-h)}{h} + \underbrace{\frac{h}{2}f''(x)}_{\mathcal{O}(h)} \quad (1.18)$$

para algún punto de la partición:

$$f'(x_j) \cong \frac{f(x_j) - f(x_{j-1})}{h} \quad (1.19)$$

Para la derivada regresiva se tiene un orden de aproximación de orden  $\mathcal{O}(h)$ .

### 1.4.3 Derivada Central

Para estimar la derivada central se compara las expresiones de ambos desarrollos de Taylor.

$$\begin{aligned} f(x+h) &= f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{3!}f'''(x) + \dots \\ f(x-h) &= f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{3!}f'''(x) + \dots \end{aligned} \quad (1.20)$$

Restamos las dos expresiones tenemos:

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} - \underbrace{\frac{h^2}{3}f'''(x)}_{\mathcal{O}(h^2)} \quad (1.21)$$

para algún punto de la partición:

$$f'(x_j) \cong \frac{f(x_{j+1}) - f(x_{j-1})}{2h}. \quad (1.22)$$

Notar que la estimación central tiene un orden  $\mathcal{O}(h^2)$  en la estimación.

### 1.4.4 Error Local

Una medida del error local es la distancia (euclíadiana) entre el valor estimado y el valor real.

$$\sigma(D^n f(x_j)) = |f^n(x_j) - D_{exact}^n f(x_j)| \quad (1.23)$$

#### Section 1.5

### Segunda Derivada

Para estimar la segunda derivada numérica, se suma los dos desarrollos de Taylor en la Ecuación (1.20).

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} + \mathcal{O}(h^2) \quad (1.24)$$

para algún punto de la partición:

$$f''(x_j) \cong \frac{f(x_{j+1}) - 2f(x_j) + f(x_{j-1})}{h^2} \quad (1.25)$$

Notar que la estimación tiene un orden  $\mathcal{O}(h^2)$  en la estimación.

#### Section 1.6

### Ejercicios: Derivación

1. (**Theoretical**) Demuestre la formula alternativa para la estimación de la segunda derivada discreta:

$$\frac{d^2 f(x_i)}{dx^2} = \frac{f(x_{i+2}) - 2f(x_i) + f(x_{i-2})}{4h^2} \quad (1.26)$$

2. Usando la definición de derivada central (con  $h = 0.05$ ) estime la derivada de la función:

$$f(x) = \frac{1}{\sqrt{1 + e^{-x^2}}}, \quad (1.27)$$

a) En el intervalo  $-10 \leq x \leq 10$ .

b) Para el intervalo anterior, estimar el error en cada punto nodal.

3. (**Machine Learning analogy**) El operador derivada central se puede definir a través de la operación de convolución usando un kernel muy específico. Sea  $M$  una máscara de convolución:

$$\mathbb{M} = [1, 0, -1] \quad (1.28)$$

El operador derivada central queda expresado como la convolución discreta entre la función  $y$  y la máscara  $M$ :

$$Df(x_n) = \frac{1}{2h} \sum_{m=-\infty}^{\infty} M[m+1]f(x_{n-m}). \quad (1.29)$$

Note que el kernel está centrado en  $m = 0$ , de modo que las sumas realmente van entre  $m = -1$  hasta  $m = 1$ . Implemente este algoritmo para calcular la derivada de la función del punto 2).

4. Diseñe un kernel de convolución para expresar el operador segunda derivada.
- Dar la expresión matemática:  $D^2f(x) = ?$ .
  - Implementar el cálculo de esta derivada a la función del punto 2).
5. (**Theoretical**) Show that the  $D^4f$  operator is given by:

$$D^4f(x_j) \cong \frac{f(x_{j+2}) - 4f(x_{j+1}) + 6f(x_j) - 4f(x_{j-1}) + f(x_{j-2})}{h^4} \quad (1.30)$$

For this operator, what is the order ( $\mathcal{O}(h^k)$ ) of the approximation?

6. Calcular el campo de velocidades cerca de la superficie de un cilindro de radio  $R = 2 \text{ cm}$ . Para esta tarea realizar los siguientes pasos:
- Definir una discretización en los ejes  $x$  e  $y$ , donde la región es:  $A \in [-4, 4]$  con 25 puntos en cada eje.
  - Definir la función potencial del flujo dada por:

$$\phi(x, y) = Vx \left( 1 - \frac{R^2}{x^2 + y^2} \right) \quad (1.31)$$

donde  $V = 2 \text{ cm/s}$

- c) Calcule y guarde adecuadamente el campo de velocidades usando la definición de derivada parcial central como:

$$\begin{aligned} v_x &= \frac{\partial \phi}{\partial x} \\ v_y &= -\frac{\partial \phi}{\partial y} \end{aligned} \quad (1.32)$$

use  $h = 0.001$ . Note que al interior del cilindro el campo de velocidades debe ser igual a cero.

- d) Dibuje el campo de velocidades usando el método: `ax.quiver(x[i],y[j],Vx[i,j],Vy[i,j])`. Debería obtener algo como 1.2:

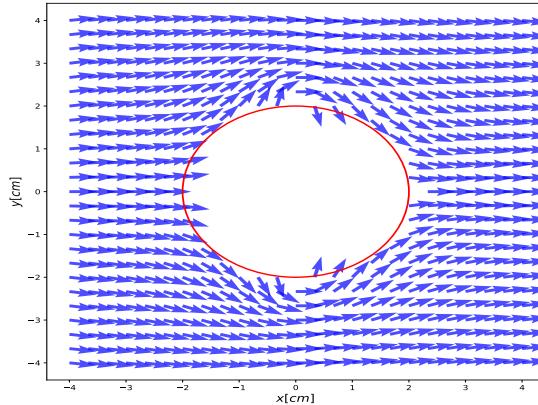


Figure 1.1: Campo de velocidades cerca de un cilindro sólido.

7. A classical electrodynamics problem is the following: suppose a point charge  $q$  held a distance  $d$  above an infinite grounded conducting plane. The electric potential is given by:

$$V(x, y) = \frac{q}{4\pi\epsilon_0} \left[ \frac{1}{\sqrt{(x - \vec{r}_q[0])^2 + (y - d)^2}} - \frac{1}{\sqrt{(x - \vec{r}_q[0])^2 + (y + d)^2}} \right] \quad (1.33)$$

For this numerical estimation the position of the charge is  $\vec{r}_q = (0.51, 0.21) \text{ m}$ . Note that  $d = \vec{r}_q[1]$ .

- a) Calculate the electric field in arbitrary units where  $\frac{q}{4\pi\epsilon_0} = 1$ .

$$\vec{E} = -\nabla V(x, y) \quad (1.34)$$

The derivation region is  $R \in [0., 1.] \times [0., 1.] \text{ m}^2$  and the step of discretization is  $h = 0.05$ .

- b) Plot the vectorial field using: `ax.quiver(x[i],y[j],Ex[i,j],Ey[i,j])`

The result looks like this 1.2:

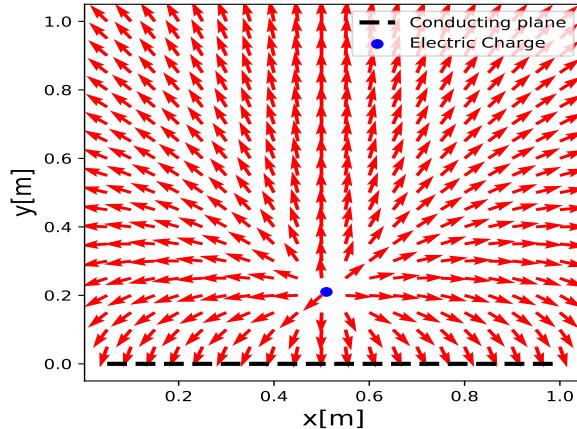


Figure 1.2: Electric field of an electric charge close to grounded conducting plane. Note that the Gauss' Law is satisfied because of the electric field is normal in the surface. Nice!

8. Es posible construir una aproximación de orden  $\mathcal{O}(h^2)$  para la derivada progresiva. Para tal propósito, se escribe el polinomio de interpolación de grado 2 para el conjunto soporte  $\Omega = \{(x_0, f(x_0)), (x_1, f(x_1)), (x_2, f(x_2))\}$ , y posteriormente se calcula la derivada de este polinomio.

- a) Calcular analíticamente el polinomio que interpola el conjunto soporte.  
 b) Derivar el polinomio interpolador para encontrar la derivada en el punto  $x_0$ :

$$f'(x_0) \approx p'(x_0) = \frac{1}{2h}(-3f(x_0) + 4f(x_1) - f(x_2)). \quad (1.35)$$

Si la discretización es equidistante, tenemos:

$$f'(x) \cong \frac{1}{2h}(-3f(x) + 4f(x+h) - f(x+2h)). \quad (1.36)$$

- c) (Python) Para  $f(x) = \sqrt{\tan(x)}$  estimar la derivada progresiva de orden  $\mathcal{O}(h^2)$  (expresión anterior) en el intervalo  $[0.1, 1.1]$  con  $h = 0.01$ .
- d) (Python) Para  $f(x) = \sqrt{\tan(x)}$  estimar la derivada central de orden  $\mathcal{O}(h^2)$  en el intervalo  $[0.1, 1.1]$  con  $h = 0.01$ .
- e) Calcule analíticamente la derivada de la función  $f(x)$ , y grafique con la estimación central y progresiva de orden  $\mathcal{O}(h^2)$ .
- f) Grafique el error nodal para ambas aproximaciones. ¿Tienen efectivamente el mismo orden de precisión ambos resultados?

**Section 1.7****Método de Newton-Raphson**

Es un método iterativo para encontrar las raíces reales polinomios usando conceptos de cálculo diferencial. Tomemos un punto cualquiera  $x_n$  y construimos la ecuación de la recta usando la derivada de  $f(x)$  en  $x_n$ .

$$Df(x_n) = \frac{f(x_{n+1}) - f(x_n)}{x_{n+1} - x_n} \quad (1.37)$$

Se pretende que el siguiente punto en la iteración sea un raíz de  $f(x_{n+1}) = 0$ , por tanto:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (1.38)$$

Este es conocida como la forma iterativa de Newton-Raphson. Otro camino para deducir esta formula consiste en expandir  $f(x)$  alrededor de  $x_n$ .

$$f(x) = f(x_n) + f'(x_n)(x - x_n) + \frac{(x - x_n)^2}{2!} f''(x_n) + \dots \quad (1.39)$$

Si se trunca la función hasta orden  $\mathcal{O}(x^2)$  y se evalúa en el siguiente punto  $x_{n+1}$ , el cuál se considera un raíz de  $f(x)$ ; se llega a la formula deseada.

**1.7.1 Criterio de parada**

Podemos usar el error relativo en cada iteración para detener el método.

$$\epsilon = \frac{|x_{k+1} - x_k|}{|x_{k+1}|} \quad (1.40)$$

el cual detiene el método cuando sea menor a una tolerancia, i.e.,  $\epsilon < 10^{-6}$ .

**Section 1.8****Método de bisección**

Este método consiste en obtener una estimación de la raíces de una función, la cuales se suponen existe en cierto intervalo  $[a,b]$ . La existencia de las raíces (el teorema no establece cuántas raíces existen en dicho intervalo) se garantizan por medio del teorema de Bolzano, que establece lo siguiente:

Sea  $f : [a, b] \rightarrow \mathbb{R}$  una función continua en  $[a, b]$  tal que  $f(a) < 0 < f(b)$ . Entonces, existe al menos un punto  $c \in [a, b]$  tal que  $f(c) = 0$ .

Si se cumple el teorema de Bolzano en nuestro intervalo de estudio, podemos acercarnos a la raíz de la solución dividiendo el intervalo en su punto medio iterativamente.

$$x_m = \frac{a + b}{2} \quad (1.41)$$

De los dos intervalos elegimos el intervalo que cumpla la condición de Bolzano:  $f(a)f(x_m) < 0$  o  $f(x_m)f(b) < 0$ . Repitiendo el proceso hasta que el intervalo sea lo suficientemente pequeño para acotar la raíz solución dentro de una vecindad arbitrariamente pequeña. La noción de distancia está dada por:

$$|x^k - x^{k-1}| < \epsilon^k \quad (1.42)$$

El algoritmo de bisección se puede describir como sigue:

- a) En el intervalo  $[a, b]$  se debe cumplir la condición de Boltzano:  $f(a)f(b) < 0$ . De lo contrario, no existe la raíz en dicho intervalo.
- b) De Divide el intervalo en su punto medio, donde tenemos dos intervalos:  $[a, x_m]$  y  $[x_m, b]$ .
- c) Si  $f(x_m) = 0$ , el punto medio es la raíz, con un cierto error y la iteración termina.
- d) Si  $f(a)f(x_m) < 0$ . la raíz está en  $[a, x_m]$ . Hacer  $b = x_m$  y repetir paso b). De lo contrario, la raíz está en  $[x_m, b]$ . Hacer  $a = x_m$  y repetir el paso b).

Dado que en cada paso la longitud del intervalo se divide por dos, la longitud del intervalo después de  $k$  pasos se reduce a  $\frac{b-a}{2^k}$ . La aproximación a la raíz será el punto medio de dicho intervalo, de este modo, el error en la aproximación está dada por:

$$|\epsilon^k| < \frac{1}{2} \frac{b-a}{2^k}. \quad (1.43)$$

$|\epsilon^k|$  es asignado arbitrariamente. Para cierta precisión se tiene la siguiente expresión del número de pasos que acotan la solución en esa vecindad ( $|x^k - x_{exacta}| < \epsilon^k$ ):

$$k > \frac{\ln((b-a)/\epsilon^k)}{\ln(2)} - 1 \quad (1.44)$$

**Section 1.9****Ejercicios: Raíces de polinomios**

1. ¿De qué tipo es el error asociado a la estimación de raíces usando el método de Newton-Raphson?
2. ¿Cómo ajustar la precisión para estimar raíces con el método de Newton-Raphson?
3. Calcular todas las raíces reales de:

$$f(x) = 3x^5 + 5x^4 - x^3 \quad (1.45)$$

4. (**Sympy**) Calcular todas las raíces reales de los primeros 20 polinomios de Legendre. La formula de rodriques es:

$$p_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n. \quad (1.46)$$

El intervalo donde existen las raíces es:  $-1 \leq x \leq 1$ .

5. (**Sympy**) Calcular todas las raíces reales de los primeros 20 polinomios de Laguerre. La formula de rodriques es:

$$L_n(x) = \frac{e^x}{n!} \frac{d^n}{dx^n} (e^{-x} x^n). \quad (1.47)$$

El intervalo donde existen las raíces es:  $0 \leq x \leq \infty$ .

**Section 1.10**

## Interpolación de Lagrange

**Teorema 1.** (*Teorema de aproximación de Weierstrass*). Sea  $f \in [a, b]$ , dado  $\epsilon > 0$  existe un polinomio  $p(x)$  definido en  $[a, b]$  tal que:

$$|f(x) - p(x)| < \epsilon \quad \forall x \in [a, b]. \quad (1.48)$$

Descubierto por Edwarg Waring en 1779 y redescubierto por Leonhard Euler en 1783, fue publicado por Lagrange en 1795. Sea el conjunto soporte  $\Omega = \{(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)\}$  de  $n + 1$  puntos diferentes, existe un polinomio interpolador de grado  $n$ :

$$p_n(x) = \sum_{i=0}^n f(x_i) \mathcal{L}_i(x), \quad (1.49)$$

donde  $\mathcal{L}_i(x)$  es la base de Lagrange (también conocidas como funciones cardinales).

$$\mathcal{L}_i(x) = \prod_{j=0, j \neq i}^n \frac{x - x_j}{x_i - x_j} \quad (1.50)$$

Este polinomio cumple que  $p(x_k) = y_k$  para todo  $k$  en  $\{0, \dots, n\}$ .

### Ejemplo:

Encontrar las funciones cardinales ( $\mathcal{L}_i(x)$ ) y el polinomio interpolador para el siguiente conjunto:  
 $\Omega = \{(5, 10), (10, 15)\}$ .

$$\mathcal{L}_0(x) = \frac{x - 10}{5 - 10} = -\frac{1}{5}(x - 10) \quad (1.51)$$

$$\mathcal{L}_1(x) = \frac{x - 5}{10 - 5} = \frac{1}{5}(x - 5) \quad (1.52)$$

Por tanto, el polinomio interpolador es:

$$\begin{aligned} p_1(x) &= 10\mathcal{L}_0(x) + 15\mathcal{L}_1(x) \\ p_1(x) &= x + 5 \end{aligned} \quad (1.53)$$

### 1.10.1 Error

Sea  $f : I \rightarrow \mathbb{R}$ ,  $\{x_i\}_{i=0}^n \subseteq I$ ,  $x_i \neq x_j$  para  $i \neq j$  y suponemos que  $f$  es derivable  $n+1$  veces. El error asociado a la interpolación está dado por:

$$E = f(x) - p(x) = \frac{f^{n+1}(\xi_x)}{(n+1)!}(x-x_0)(x-x_1)\dots(x-x_n) \quad (1.54)$$

donde  $p(x)$  es el polinomio interpolador en  $\{x_i\}_{i=0}^n$  y  $\xi_x \in$  al intervalo que contiene los puntos.

**proof:**

Si  $x$  es un punto  $x_k$  la identidad se satisface para cualquier  $\xi$ . De lo contrario, si  $x$  es fijo y diferente  $x_k$  se considera una función auxiliar:

$$F(t) = f(t) - p(t) - cL(t), \quad \text{donde} \quad c = \frac{f(x) - p(x)}{L(x)} \quad (1.55)$$

$L(x) = \prod_{i=0}^n (x-x_i)$ . Si evaluamos la función auxiliar en los puntos  $x_k$ ,  $F(x_k) = y_k - y_k - 0 = 0$  para todo  $k$ . Por tanto,  $F$  tiene  $n+1$  ceros. Adicionalmente  $F(x) = f(x) - p(x) - cL(x) = 0$ , dada la definición de  $c$ . entonces la función  $F$  tiene  $n+2$  ceros en el intervalo  $I$ . Ahora, por el teorema de Rolle,  $F'$  debe tener al menos  $n+1$  ceros en el intervalo que tiene a los puntos  $x_k$ ; entonces la  $(n+1)$ -ésima derivada debe tener al menos un cero. Sea  $\xi_x$  ese cero. Entonces derivamos  $(n+1)$  veces y evaluamos en  $\xi_x$ :

$$F^{n+1}(\xi_x) = f^{n+1}(\xi_x) - c(n+1)! = 0 \quad (1.56)$$

debido a que la  $(n+1)$ -ésima derivada de  $p(x)$  es cero. Entonces:

$$c = \frac{f^{n+1}(\xi_x)}{(n+1)!} \rightarrow cL(x) = f(x) - p(x) = \frac{1}{(n+1)!} f^{n+1}(\xi_x) L(x) \blacksquare \quad (1.57)$$

#### Section 1.11

### Interpolación de Newton-Gregory

Sean los  $n+1$  puntos de la discretización  $x_0, x_1, \dots, x_n$ , tal que  $x_i - x_{i-1} = h$  y sus respectivas imágenes:  $f(x_0), f(x_1), \dots, f(x_n)$ . Se propone el polinomio de interpolación:

$$p(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \dots + a_n(x - x_0)(x - x_1)\dots(x - x_{n-1}) \quad (1.58)$$

Se requiere el cálculo de los coeficientes del este polinomio. Se establece la notación  $f(x_0) = f_0$ , evaluando el polinomio en  $x_0$ :

$$p(x_0) = f_0 \quad (1.59)$$

Entonces  $a_0 = f_0$ : evaluando en  $x_1$ :

$$\begin{aligned} p(x_1) &= f_1 \\ f_1 &= f_0 + a_1(x_1 - x_0) \\ f_1 &= f_0 + a_1 h \end{aligned} \quad (1.60)$$

Note que  $a_1$  es la derivada derecha. Evaluando en  $x_2$ :

$$\begin{aligned} p(x_2) &= f_2 \\ f_2 &= f_0 + a_1(x_2 - x_0) + a_2(x_2 - x_0)(x_2 - x_1) \\ f_2 &= f_0 + \frac{f_1 - f_0}{h}(2h) + a_2(2h)(h) \end{aligned} \quad (1.61)$$

Entonces  $a_2$  se relaciona con la segunda derivada central:

$$a_2 = \frac{f_2 - 2f_1 + f_0}{2h^2} \quad (1.62)$$

Pensemos en un término general para los coeficientes:

$$a_n = \frac{\Delta^n f_0}{n!h^n} \quad (1.63)$$

donde  $\Delta^{n+1} f_i = \Delta^n f_{i+1} - \Delta^n f_i$ , que se conoce como las diferencias finitas de la función.

Por ejemplo: podemos ver los primeros términos:

$$\begin{aligned} \Delta^{0+1} f_i &= \Delta^0 f_{i+1} - \Delta^0 f_i \\ &= f_{i+1} - f_i \end{aligned} \quad (1.64)$$

$$\begin{aligned} \Delta^{1+1} f_i &= \Delta^1 f_{i+1} - \Delta^1 f_i \\ &= f_{i+2} - f_{i+1} - (f_{i+1} - f_i) \\ &= f_{i+2} - 2f_{i+1} + f_i \end{aligned} \quad (1.65)$$

$$\begin{aligned}
 \Delta^{2+1} f_i &= \Delta^2 f_{i+1} - \Delta^2 f_i \\
 &= f_{i+3} - 2f_{i+2} + f_{i+1} - (f_{i+2} - 2f_{i+1} + f_i) \\
 &= f_{i+3} - 3f_{i+2} + 3f_{i+1} - f_i
 \end{aligned} \tag{1.66}$$

$$\begin{aligned}
 \Delta^{3+1} f_i &= \Delta^3 f_{i+1} - \Delta^3 f_i \\
 &= f_{i+4} - 3f_{i+3} + 3f_{i+2} - f_{i+1} - (f_{i+3} - 3f_{i+2} + 3f_{i+1} - f_i) \\
 &= f_{i+4} - 4f_{i+3} + 6f_{i+2} - 4f_{i+1} + f_i
 \end{aligned} \tag{1.67}$$

El cálculo de estas diferencias resulta más evidente en el marco de matrices. Por ejemplo:

$$\Delta = \left( \begin{array}{ccccccc}
 \Delta^0 & \Delta^1 & \Delta^2 & \Delta^3 & \Delta^4 & \Delta^5 & \Delta^6 \\
 \hline
 39 & 0 & 0 & 0 & 0 & 0 & 0 \\
 19 & -20 & 0 & 0 & 0 & 0 & 0 \\
 -21 & -40 & -20 & 0 & 0 & 0 & 0 \\
 -57 & -36 & 4 & 24 & 0 & 0 & 0 \\
 -65 & -8 & 28 & 24 & 0 & 0 & 0 \\
 -21 & 44 & 52 & 24 & 0 & 0 & 0 \\
 99 & 120 & 76 & 24 & 0 & 0 & 0
 \end{array} \right) \tag{1.68}$$

En este caso, el polinomio interpolador será de orden tres.

### Section 1.12

## Ejercicios: Interpolación de Lagrange

1. (**Theoretical**) Demuestre que el polinomio interpolador es único.
2. (**Theoretical**) Compruebe que las funciones cardinales son base (i.e,  $L_i(x) = \delta_{ij}$  para cada  $j \in \{0, 1, \dots, n\}$ ).
3. ¿Con qué grado de exactitud podemos calcular  $\sqrt{114}$  mediante la interpolación de Lagrange para la función  $f(x) = \sqrt{x}$ , si elegimos los puntos  $x_0 = 100$ ,  $x_1 = 121$ ,  $x_2 = 144$ . Rpta:  $|E| \simeq 1.8 \times 10^{-3}$ .

4. En el lanzamiento de una bala, una cámara fotográfica registra las siguientes posiciones en metros respecto al arma homicida (tome  $\vec{g} = -9.8 \text{ m/s}^2 \hat{j}$ ):

<https://raw.githubusercontent.com/asegura4488/Database/main/MetodosComputacionalesReferenciales/Parabolico.csv>

Estime el vector velocidad inicial, que estaría definido por la magnitud y dirección. Rpta:  $V_0 = 10 \text{ m/s}$  y  $\theta = 20^\circ$ . Hint: Encuentre el termino lineal y cuadrático de la interpolación y compare con la ecuación de trayectoria de la bala.

5. *Interpolación Newton-Gregory:* Para el siguiente conjunto de puntos:

<https://raw.githubusercontent.com/asegura4488/Database/main/MetodosComputacionalesReferenciales/InterpolacionNewtonNoequi.csv>

Encuentre el polinomio interpolante de menor grado usando el método Newton-Gregory.

$$p(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \dots + a_n(x - x_0)(x - x_1)\dots(x - x_{n-1}) \quad (1.69)$$

Recuerde que las pendientes están definidas de forma recursiva:

$$\begin{aligned} f_0(x_i) &= \text{ términos de la secuencia} \\ f_1(x_0, x_1) &= \frac{f_0(x_1) - f_0(x_0)}{x_1 - x_0} \\ f_2(x_0, x_1, x_2) &= \frac{f_1(x_1, x_2) - f_1(x_0, x_1)}{x_2 - x_0} \\ f_i(x_0, x_1, \dots, x_{i-1}, x_i) &= \frac{f_{i-1}(x_1, \dots, x_{i-1}, x_i) - f_{i-1}(x_0, x_1, \dots, x_{i-1})}{x_i - x_0} \end{aligned} \quad (1.70)$$

Usando las pendientes, el polinomio de grado  $n$  queda definido por:

$$\begin{aligned} p_0(x) &= f_0(x_0) \\ p_1(x) &= p_0(x) + f_1(x_0, x_1)(x - x_0) \\ p_2(x) &= p_1(x) + f_2(x_0, x_1, x_2)(x - x_0)(x - x_1) \\ p_i(x) &= p_{i-1}(x) + f_i(x_0, x_1, \dots, x_{i-1}, x_i) \prod_{j=0}^{i-1} (x - x_j) \end{aligned} \quad (1.71)$$

El resultado se muestra en la Figura [1.3]:

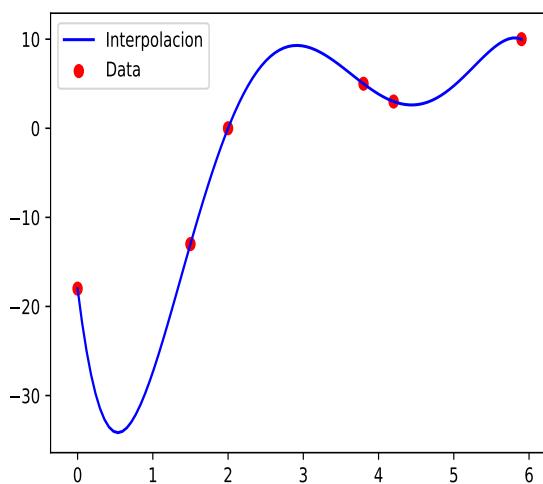


Figure 1.3: Interpolación usando el método de Newton-Gregory para el conjunto de datos descritos en el problema.

### Section 1.13

## Integración

Para el calculo de integrales definidas existe una familia de métodos denominados *Métodos de Newton-Côtes*, los cuales se basan en cambiar el integrando  $f(x)$  a un polinomio interpolador que aproxima a  $f(x)$  en el intervalo de integración. El grado del polinomio interpolador queda definido por el número de puntos a considerar, por ejemplo, en los casos más simples de interpolación líneal (Regla de trapecio) e interpolación cuadrática (Regla de simpson), se tiene expresiones sencillas que son fáciles de implementar computacionalmente.

### 1.13.1 Método de trapecio simple

Para la integral definida:

$$I = \int_a^b f(x)dx, \quad (1.72)$$

el método de trapecios simple cambia el integrando por un polinomio interpolador de grado uno. Este polinomio interpolador esta definido en el conjunto  $\Omega = \{(a, f(a)), (b, f(b))\}$  que finalmente conduce a la siguiente aproximación:

$$f(x) \approx p_1(x) = \frac{x - b}{a - b}f(a) + \frac{x - a}{b - a}f(b), \quad \forall x \in [a, b]. \quad (1.73)$$

De modo que la integral tiene un valor aproximado:

$$I = \int_a^b f(x)dx \cong \int_a^b p_1(x)dx = \frac{b - a}{2}(f(a) + f(b)) \quad (1.74)$$

El error en la estimación está asociado al procedimiento de interpolación. Suponiendo que  $f(x)$  es continua y derivable de clase  $C^2$  en el intervalo  $[a,b]$ :

$$f(x) = p_1(x) + \epsilon(x), \quad (1.75)$$

donde

$$\epsilon(x) = \frac{f''(\xi)}{2}(x - a)(x - b), \quad a \leq \xi \leq b. \quad (1.76)$$

Entonces el error asociado a la integración por le método del trapecio está dado por ( $h = b - a$ ):

$$E = \int_a^b \epsilon(x) dx = -\frac{h^3}{12} f''(\xi) \quad (1.77)$$

De esta forma el error alcanza un valor máximo para algún valor de  $\xi$ , donde la segunda derivada de  $f(x)$  se maximice; de modo que podemos acotar el error máximo en esta estimación.

$$|E| \leq \frac{h^3}{12} \max_{a \leq \xi \leq b} |f''(\xi)|. \quad (1.78)$$

Notar que el error es de orden  $\mathcal{O}(h^3)$ .

### 1.13.2 Método de trapecio compuesto

Para la integral definida:

$$I = \int_a^b f(x) dx, \quad (1.79)$$

el método de trapecios compuesto genera una partición *equi-espaciada* tal que:  $x_{i+1} - x_i = h$ ,  $\forall i = [1, \dots, n]$  sobre el conjunto  $\Omega = \{(x_0, f(x_0)), (x_1, f(x_1)), \dots, (x_n, f(x_n))\}$ . Las condiciones de borde corresponden con los límites de integración ( $x_0 = a$  y  $x_n = b$ ), definiendo el paso de integración  $h = \frac{b-a}{n}$ . La integral se puede escribir como:

$$\int_a^b f(x) dx = \int_{x_0}^{x_1} f(x) dx + \int_{x_1}^{x_2} f(x) dx + \dots + \int_{x_{n-1}}^{x_n} f(x) dx, \quad (1.80)$$

aplicando el método de trapecios simple se tiene:

$$\int_a^b f(x) dx \approx \frac{h}{2} (f(x_0) + f(x_1)) + \frac{h}{2} (f(x_1) + f(x_2)) + \dots + \frac{h}{2} (f(x_{n-1}) + f(x_n)), \quad (1.81)$$

tenemos la expresión para la regla de trapecio compuesta:

$$\int_a^b f(x) dx \approx \frac{h}{2} (f(a) + f(b)) + h \sum_{i=1}^{n-1} f(x_i). \quad (1.82)$$

La estimación del error se calcula sumando los errores en cada sub-intervalo.

$$E = \sum_{i=1}^n E_i = -\frac{h^2}{12} (f''(\xi_1) + f''(\xi_2) + \dots + f''(\xi_n)) \quad (1.83)$$

El error puede ser acotado por el valor máximo promedio que tome la segunda derivada en el intervalo  $[a, b]$ .

$$|E| \leq \frac{h^2(b-a)}{12} \underbrace{\max}_{a \leq \xi \leq b} |f''(\xi)|. \quad (1.84)$$

Notar que el error es de orden  $\mathcal{O}(h^2)$ .

### 1.13.3 Método de Simpson simple 1/3

Para la integral definida:

$$I = \int_a^b f(x) dx, \quad (1.85)$$

el método de Simpson simple cambia el integrando por un polinomio interpolador de grado dos. Este polinomio interpolador esta definido en el conjunto  $\Omega = \{(a, f(a)), (x_m, f(x_m)), (b, f(b))\}$ , donde  $x_m = \frac{a+b}{2}$  es el punto medio del intervalo. La interpolación conduce a la siguiente aproximación:

$$f(x) \approx p_2(x) = \frac{(x-b)(x-x_m)}{(a-b)(a-x_m)} f(a) + \frac{(x-a)(x-b)}{(x_m-a)(x_m-b)} f(x_m) + \frac{(x-a)(x-x_m)}{(b-a)(b-x_m)} f(b), \quad \forall x \in [a, b]. \quad (1.86)$$

De modo que la integral tiene un valor aproximado:

$$\int_a^b f(x) dx \cong \int_a^b p_2(x) dx = \frac{h}{3} (f(a) + 4f(x_m) + f(b)). \quad (1.87)$$

Adicionalmente, es importante mencionar que *la discretización debe ser par*. El error de la aproximación tiene una característica interesante, dado que el error de la interpolación genera un resultado idénticamente nulo. Suponiendo que  $f(x)$  es continua y derivable de clase  $C^3$  en el intervalo  $[a,b]$ :

$$f(x) = p_2(x) + \epsilon(x), \quad (1.88)$$

donde

$$E = \int_a^b \epsilon(x) dx = \int_a^b \frac{f'''(\xi)}{4!} (x-a)(x-b)(x-(a+b)/2) dx = 0, \quad a \leq \xi \leq b. \quad (1.89)$$

Esto significa que la regla de Simpson es exacta a orden  $\mathcal{O}(h^3)$ , entonces necesitamos otro camino para calcular el error de la estimación a orden  $\mathcal{O}(h^4)$ . Para tal propósito, se va a expandir en series de Taylor alrededor del punto medio  $x_m$  a  $f(x)$  y al polinomio interpolador  $p_2(x)$ . Suponiendo que  $f(x)$  es continua y derivable de clase  $C^4$  en el intervalo  $[a,b]$ , a orden  $\mathcal{O}(h^4)$  se tiene:

$$f(x) = f(x_m) + f'(x_m)(x - x_m) + \frac{f''(x_m)}{2!}(x - x_m)^2 + \frac{f'''(x_m)}{3!}(x - x_m)^3 + \epsilon(x), \quad (1.90)$$

donde el error es de orden  $\mathcal{O}(h^4)$ .

$$\epsilon(x) = \frac{f^{(4)}(\xi)}{4!}(x - x_m)^4. \quad (1.91)$$

Para reemplazar en la regla simple de Simpson, se debe encontrar  $f(a)$  y  $f(b)$ , entonces:

$$\begin{aligned} f(a) &= f(x_m - h) = f(x_m) + f'(x_m)(-h) + \frac{f''(x_m)}{2!}(-h)^2 + \frac{f'''(x_m)}{3!}(-h)^3 + \epsilon(x_m - h) \\ f(b) &= f(x_m + h) = f(x_m) + f'(x_m)(+h) + \frac{f''(x_m)}{2!}(+h)^2 + \frac{f'''(x_m)}{3!}(+h)^3 + \epsilon(x_m + h) \end{aligned} \quad (1.92)$$

Por tanto, la regla de Simpson queda expresada como (Notar como el término de orden  $\mathcal{O}(h^3)$  no contribuye):

$$\begin{aligned} \frac{h}{3}(f(a) + 4f(x_m) + f(b)) &= \frac{h}{3}(6f(x_m) + f''(x_m)h^2 + \epsilon(x_m + h) + \epsilon(x_m - h)) \\ &= \frac{h}{3}(6f(x_m) + f''(x_m)h^2 + \frac{1}{12}f^{(4)}(\xi)h^4) \end{aligned} \quad (1.93)$$

Ahora integrando el desarrollo de Taylor dado por la Ecuación (1.90):

$$\int_a^b f(x)dx = 2hf(x_m) + \frac{f''(x_n)}{3}h^3 + \frac{f^{(4)}(\xi)}{60}h^5. \quad (1.94)$$

De modo que es posible calcular el error como la diferencia en las dos integrales.

$$\begin{aligned} E &= \int_a^b f(x)dx - \int_a^b p_2(x)dx \\ &= \frac{f^{(4)}(\xi)}{60}h^5 - \frac{f^{(4)}(\xi)}{36}h^5 = -\frac{1}{90}f^{(4)}(\xi)h^5 \end{aligned} \quad (1.95)$$

El error puede ser acotado por el valor máximo promedio que tome la cuarta derivada en el intervalo  $[a, b]$ ; notar que el error es de orden  $\mathcal{O}(h^5)$ .

$$|E| \leq \frac{1}{90} h^5 \underbrace{\max}_{a \leq \xi \leq b} |f^{(4)}(\xi)|. \quad (1.96)$$

### 1.13.4 Método de Simpson compuesto

Para la integral definida:

$$I = \int_a^b f(x) dx, \quad (1.97)$$

el método de Simpson compuesto genera una partición *equi-espaciada* tal que:  $x_{i+1} - x_i = h$ ,  $\forall i = [1, \dots, n]$  sobre el conjunto  $\Omega = \{(x_0, f(x_0)), (x_1, f(x_1)), \dots, (x_n, f(x_n))\}$ . Las condiciones de borde corresponden con los límites de integración ( $x_0 = a$  y  $x_n = b$ ). Definiendo el paso de integración como  $h = \frac{b-a}{n}$ , la integral se puede escribir:

$$\int_a^b f(x) dx = \int_a^{x_2} f(x) dx + \int_{x_2}^{x_4} f(x) dx + \dots + \int_{x_{n-2}}^b f(x) dx \quad (1.98)$$

Notar que el número de puntos de la partición deber ser par y que los puntos medios serían el sub-conjunto  $\{x_{2n+1}\}$ . Aplicando el método simple a cada integral tenemos:

$$\int_a^b f(x) dx \approx \frac{h}{3} \left( f(a) + 4f(x_1) + f(x_2) + f(x_2) + 4f(x_3) + f(x_4) + \dots + f(x_{n-2}) + 4f(x_{n-1}) + f(b) \right) \quad (1.99)$$

En general, podemos escribir la expresión final para el método de Simpson compuesto:

$$\int_a^b f(x) dx \approx \frac{h}{3} \left( f(a) + 4 \sum_{i=1, \text{impares}}^{n-1} f(x_i) + 2 \sum_{i=2, \text{pares}}^{n-2} f(x_i) + f(b) \right) \quad (1.100)$$

Para la estimación del error total, se debe considerar el error cometido en cada una de las integrales. De este modo:

$$|E| = \frac{1}{90} h^5 \left( f^{(4)}(\xi_1) + f^{(4)}(\xi_2) + \dots + f^{(4)}(\xi_{n/2}) \right) \quad (1.101)$$

Finalmente, acotando superiormente la estimación tenemos el error de la regla de Simpson compuesta a orden  $\mathcal{O}(h^4)$ .

$$|E| \leq \frac{b-a}{180} h^4 \underbrace{\max}_{a \leq \xi \leq b} |f^{(4)}(\xi)|. \quad (1.102)$$

Cabe resaltar que la regla de Simpson compuesta que aproxima el integrando por un polinomio de orden tres, reduce en dos ordenes el error cometido por método de trapecio compuesto.

### Ejemplo:

Para la integral:

$$\int_{-1}^1 \sqrt{1 + e^{-x^2}} dx, \quad (1.103)$$

Usando la regla de Simpson compuesta, ¿cuál debe ser el número de puntos para aproximar la integral con un error menor a 0.001?

Usando el operador discreto  $D^4 f(x_j)$  se estima el máximo en  $M = 3.183$ . Entonces:

$$n \geq \sqrt[4]{\frac{(b-a)^5}{E180}} M \approx \sqrt[4]{\frac{(2)^5}{180}} 3.183 \approx 4.87 = 6 \quad \text{Intervalos} \quad (1.104)$$

debido que el número de intervalos debe ser par!

### 1.13.5 Cuadratura Gaussiana

La idea es aproximar la integral a una suma ponderada de la función  $f(x)$ , que use los valores más óptimos posibles para su determinación. En general, la formula de cuadratura está dada por:

$$G_M(f) = \sum_{i=1}^n w_i^{(n)} f(x_i^{(n)}) \cong \int_{-1}^1 f(x) dx. \quad (1.105)$$

Donde  $\{w_i^{(n)}\}$  son los pesos de ponderación y los  $\{x_i^{(n)}\}$  son los puntos de Gauss, ambos cantidades dependen del número de puntos que se elijan. A diferencia de los métodos anteriores, queremos encontrar una regla que nos permita maximizar el grado de precisión en nuestras estimaciones. En general, para calcular los coeficientes se exige que la regla sea exacta para el polinomio de mayor grado que se tenga. Tenemos entonces:

$$\int_{-1}^1 x^k dx = \sum_{i=1}^n w_i^{(n)} (x_i^{(n)})^k \quad k = 0, 1, \dots, N \quad (1.106)$$

donde  $N$  es tan grande como sea posible. Necesitamos  $2n$  condiciones para integrar polinomios de grado  $2n - 1$ , esto se denomina la regla de  $n$  puntos, para el grado polinomial  $2n - 1$ .

**Lema 1:** Si  $N = 2n$ , entonces no existe el conjunto  $\{w_i^{(n)}\}$  tal que se verifique la igualdad.

Supongamos que existe dicho polinomio, si definimos  $L(x) = \prod_{j=1}^n (x - x_j^n)^2$  entonces  $L(x) \geq 0$  y por tanto la integral  $\int_{-1}^1 L(x) dx > 0$ . Sin embargo,  $\sum w_i^{(n)} L(x_i^{(n)}) = 0$ , entonces hay una contradicción.

**Lema 2:** Sea  $\{w_i^{(n)}\}$  los pesos y  $\{x_i^{(n)}\}$  los puntos de Gaus tal que se cumpla la Ecuación (1.106) para  $k = 0, 1, 2, \dots, N = 2n - 1$ , los pesos deben satisfacer:

$$w_i^{(n)} = \int_{-1}^1 L_i^{(n)}(x) dx, \quad (1.107)$$

con

$$L_i^{(n)}(x) = \prod_{k=1, k \neq i}^n \frac{x - x_k^{(n)}}{x_i^{(n)} - x_k^{(n)}} \quad (1.108)$$

Las bases cardinales son de grado  $\leq 2n - 1$  y  $L_i^{(n)}(x_j) = \delta_{ij}$ , como la regla de cuadratura debe ser exacta para los polinomios de grado  $2n - 1$ , entonces:

$$\int_{-1}^1 L_i^{(n)}(x) dx = \sum_{j=1}^n w_j^{(n)} L_i^{(n)}(x_j) = \sum_{j=1}^n w_j^{(n)} \delta_{ij} = w_i^{(n)} \quad (1.109)$$

Ahora, tenemos un polinomio de grado  $n$ ,  $P_n(x) = \prod_{i=1}^n (x - x_i^{(n)})$  y supongamos que tenemos un polinomio cualquier  $Q(x)$  de grado  $\leq n - 1$ , de modo que la multiplicación tiene grado  $\leq 2n - 1$ . Por tanto la identidad de la cuadratura se debe satisfacer.

$$\int_{-1}^1 P_n(x) Q(x) dx = \sum_{i=1}^n w_i^{(n)} P_n(x_i^{(n)}) Q(x_i^{(n)}). \quad (1.110)$$

Por la definición de  $P_n(x)$  la integral es nula y los polinomios deben formar una base ortonormal. Adicionalmente, el conjunto  $\{x_i^n\}$  deben ser las raíces de los  $P_n(x)$ .

### Ejemplo:

Obtener la regla de cuadratura para de  $f(x)$  en el intervalo  $[-3, 3]$  usando la siguiente partición  $P = \{-1, 0, 1\}$ .

Calculando las funciones de Lagrange:

$$\begin{aligned} L_1(x) &= \frac{1}{2}x(x-1) \\ L_2(x) &= -(x+1)(x-1) \\ L_3(x) &= \frac{1}{2}x(x+1) \end{aligned} \quad (1.111)$$

Integrando dichas funciones en el intervalo  $[-3, 3]$  obtenemos el conjunto de pesos:

$$w_i^3 = \{9, -12, 9\} \quad (1.112)$$

Entonces la regla de cuadratura es:

$$\int_{-3}^3 f(x)dx \approx 3[3f(-1) - 4f(0) + 3f(1)]. \quad (1.113)$$

Note que es exacta para  $f(x) = x^2$ .

### 1.13.6 Cuadratura de Gaus-Legendre

De la necesidad del uso de polinomios ortonormales y el cálculo de sus raíces, surge la cuadratura de Gaus-Legendre. Este calculo usa los polinomios de Legendre como conjunto ortonormal.

$$p_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n, \quad (1.114)$$

Los primeros polinomios son:

$$\begin{aligned}
 p_0(x) &= 1 \\
 p_1(x) &= x \\
 p_2(x) &= \frac{1}{2}(3x^2 - 1) \\
 p_3(x) &= \frac{1}{2}(5x^3 - 3x)
 \end{aligned} \tag{1.115}$$

cumplen con la siguiente relación de completez:

$$\int_{-1}^1 p_n(x)p_m(x)dx = \begin{cases} 0 & m \neq n \\ \frac{2}{2n+1} & m = n \end{cases} \tag{1.116}$$

y las siguientes relaciones de recurrencia.

$$(n+1)p_{n+1}(x) = (2n+1)xp_n(x) - np_{n-1}(x). \tag{1.117}$$

$$(x^2 - 1) \frac{dp_n}{dx} = nxp_n - np_{n-1}. \tag{1.118}$$

Adicionalmente se tiene el siguiente coeficiente director.

$$a_n = \frac{(2n)!}{2^n(n!)^2} \tag{1.119}$$

Para calcular los pesos de la cuadratura de Gauss-Legendre, se requiere escribir las funciones cardinales en términos de los polinomios de Legendre. Si definimos las funciones base de Lagrange como:

$$L(x) = \frac{p_n(x)}{x - x_k} \tag{1.120}$$

Tenemos una forma indeterminada cuando  $x = x_k$ . Para conocer dicho límite, si existe, usemos la regla de L'Hôpital.

$$\lim_{x \rightarrow x_i} L(x) = \left. \frac{\frac{dp_n(x)}{dx}}{\frac{d(x-x_k)}{dx}} \right|_{x=x_k} = p'_n(x_k) \tag{1.121}$$

como las funciones de Lagrange deben ser base y valor 1 justo en cada raíz, se tiene la siguiente expresión.

$$L(x) = \frac{1}{p'_n(x_k)} \frac{p_n(x)}{x - x_k} \quad (1.122)$$

Por tanto, los pesos pueden ser evaluados por:

$$w_k = \frac{1}{p'_n(x_k)} \int_{-1}^1 \frac{p_n(x)}{x - x_k} dx \quad (1.123)$$

**Ejemplo:**

Calcular los pesos de ponderación para dos puntos ( $n = 2$ ). El polinomio de Legendre sería  $p_2(x) = \frac{1}{2}(3x^2 - 1)$ , cuyas raíces son:  $x_0 = \frac{1}{\sqrt{3}}$ ,  $x_1 = -\frac{1}{\sqrt{3}}$ . Los pesos están dados por:

$$w_0 = \frac{1}{3(\frac{1}{\sqrt{3}})} \int_{-1}^1 \frac{\frac{1}{2}(3x^2 - 1)}{x - \frac{1}{\sqrt{3}}} dx = 1. \quad (1.124)$$

$$w_1 = \frac{1}{3(-\frac{1}{\sqrt{3}})} \int_{-1}^1 \frac{\frac{1}{2}(3x^2 - 1)}{x + \frac{1}{\sqrt{3}}} dx = 1. \quad (1.125)$$

Por tanto, la regla de cuadratura para dos puntos es:

$$\int_{-1}^1 f(x) dx = f\left(\frac{1}{\sqrt{3}}\right) + f\left(-\frac{1}{\sqrt{3}}\right) \quad (1.126)$$

para integrales que tiene un límite de integración diferente al de los polinomios de Legendre, es posible hacer un cambio de variable de modo que -1 coincida con el límite inferior y 1 con el límite superior de la integral. Planteando la siguiente proporción:

$$\frac{x - a}{b - a} = \frac{t - (-1)}{1 - (-1)}, \quad (1.127)$$

de manera que:

$$\begin{aligned} x &= \frac{1}{2}[t(b - a) + a + b] \\ dx &= \frac{1}{2}(b - a) \end{aligned} \quad (1.128)$$

Usando este cambio de variable, podemos escribir la regla de cuadratura para cualquier intervalo de integración.

$$\int_a^b f(x) dx = \frac{1}{2}(b - a) \sum_{k=0}^n w_k f\left(\frac{1}{2}[t_k(b - a) + a + b]\right) \quad (1.129)$$

La Ecuación (1.123) aunque es correcta, resulta poco útil para calcular los pesos a alto orden computacionalmente. Es posible encontrar una versión integrada de dicha expresión usando el siguiente teorema:

**Teorema 2.** (*Identidad de Christoffel-Darboux*) *Sea  $\{p_{n(x)}\}_{n=0}^{\infty}$  el sistema ortonormal de polinomios respecto a un peso  $\omega(x)$ . Entonces, para todo  $x, y \in R$ ,  $x \neq y$  y  $n \geq 1$  se cumple.*

$$\sum_{k=0}^{n-1} p_k(x)p_k(y) = \frac{a_{n-1}}{a_n} \left[ \frac{p_n(x)p_{n-1}(y) - p_{n-1}(x)p_n(y)}{x-y} \right] \quad (1.130)$$

siendo  $p_n(x) = a_n x^n + \dots$  y  $a_n > 0$

Usando esta identidad y haciendo  $y = x_k$  las raíces de los polinomios tenemos.

$$\sum_{k=0}^{n-1} p_k(x)p_k(x_k) = \frac{a_{n-1}}{a_n} \frac{p_n(x)p_{n-1}(x_k)}{x-x_k} \quad (1.131)$$

Integrando, tenemos:

$$1 = \frac{a_{n-1}}{a_n} p_{n-1}(x_k) \int_{-1}^1 \frac{p_n(x)}{x-x_k} dx \quad (1.132)$$

Por tanto, podemos escribir la Ecuación (1.123) como:

$$w_k = \frac{a_n}{a_{n-1} p_{n-1}(x_k) p'_n(x_k)} \quad (1.133)$$

Un cálculo sencillo muestra que  $\frac{a_n}{a_{n-1}} = 2/n$  y usando la Ecuación (1.118).

$$w_k = \frac{2}{(1-x_k^2)[p'_n(x_k)]^2}, \quad k = 1, \dots, n \quad (1.134)$$

Esta última expresión permite calcular los pesos usando librerías como `Simpy`.

### 1.13.7 Ejercicios: Integración

1. (**Theoretical**) Hacer pasos intermedios para regla de trapecio simple, Ecuación (1.74).
2. (**Theoretical**) Encontrar el error para regla de trapecio simple, Ecuación (1.77).
3. (**Theoretical**) Hacer los pasos intermedios para encontrar la regla de Simpson simple, Ecuación (1.87).

4. (**Theoretical**) Verificar el resultado presentado en la Ecuación (1.89).

5. Resolver la siguiente integral con el método del trapecio:

$$\int_0^1 e^{-x^2} dx \quad (1.135)$$

usando la condición que el error debe ser menor a 0.005 en el cálculo (si el número de puntos ( $n$ ) es decimal tomar el mayor entero).

6. Para encontrar la inductancia exacta de una bobina toroidal es necesario calcular la integral:

$$\int_{-a}^a \frac{\sqrt{a^2 - x^2}}{R + x} dx = \pi(R - \sqrt{R^2 - a^2}), \quad (1.136)$$

donde  $R = 0.5\text{ cm}$  es el radio al centro del toroide y  $a = 0.01\text{ cm}$  es el radio de la sección transversal del toroide. Estimar la integral con el método del trapecio y la regla de Simpson 1/3, con un error menor al 0.5%. Realizar la integral es un verdadero reto analítico.

7. Una forma de generalizar el método de integración del trapecio para una integral doble de una función  $f(x, y)$  consiste en dividir el plano  $xy$  en un grilla de cuadrados iguales y calcular el promedio del valor de la función de cada uno de los 4 vértices de cada cuadrado pequeño de la grilla. Calcule numéricamente el volumen de una semiesfera [1.4] de radio  $R = 1$  como sigue:

- a) Cree una grilla entre  $-R$  y  $R$  en el plano  $xy$ , donde el número de cuadrados en cada lado de la grilla sea  $n$ . Es decir, la grilla tendría  $n + 1$  puntos en cada eje, y  $n^2$  cuadrados pequeños.
  - b) Para cada cuadrado pequeño calcule el promedio de la función en los cuatro vértices y multiplique por el área del cuadrado pequeño. Si el punto queda fuera de la esfera asuma que el valor de la función  $f(x, y)$  es cero.
8. (**Sympy**) La regla de Simpson 3/8 consiste en aproximar el integrando por un polinomio interpolador de orden 3.

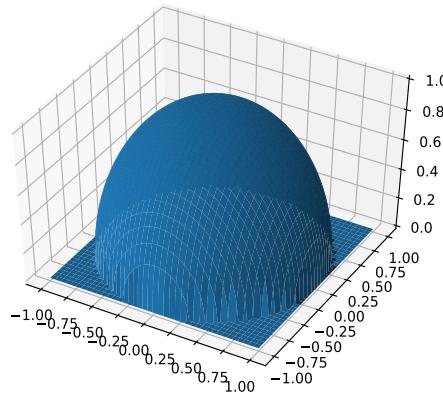


Figure 1.4: Domo esférico

- a) Encontrar las funciones cardinales de dicha interpolación e integrar para demostrar que:

$$\int_a^b f(x) \cong \frac{3h}{8} \left[ f(a) + 3f\left(\frac{2a+b}{3}\right) + 3f\left(\frac{a+2b}{3}\right) + f(b) \right]. \quad (1.137)$$

El número de puntos de la discretización debe ser múltiplo de tres.

*Hint:* Use la siguiente discretización  $\{0, h, 2h, 3h\}$ .

- b) Dado que  $h = (b - a)/3$ , muestre que los puntos intermedios son  $\frac{2a+b}{3}$  y  $\frac{a+2b}{3}$  respectivamente.

9. (**Sympy**) Muestre que el error asociado a la regla de Simpson 3/8 simple está dado por:

$$E = \frac{f^{(4)}(\xi)}{4!} \int_a^b (x - x_0)(x - x_1)(x - x_2)(x - x_3) dx = -\frac{3}{80} h^5 f^{(4)}(\xi). \quad (1.138)$$

*Hint:* Considere la siguiente integral:

$$\mathcal{I} = \int_0^{3h} (x)(x - h)(x - 2h)(x - 3h) dx \quad (1.139)$$

10. (**Theoretical**) Muestre que para la regla de Simpson 3/8 compuesta, el error puede ser acotado por:

$$|E| \leq \frac{(b-a)}{80} h^4 \underbrace{\max}_{a \leq \xi \leq b} |f^{(4)}(\xi)|. \quad (1.140)$$

11. Using the Simpson's 3/8 rule, calculate the following integral:

$$I = \int_{-1}^1 \sqrt{1 + e^{-x^2}} dx \approx 2.6388571169 \quad (1.141)$$

12. Evaluar:

$$\int_1^2 \frac{1}{x^2} dx, \quad (1.142)$$

por medio de la cuadratura de Gauss-Legendre con dos y tres puntos. Rpta:  $I_2 = 0.497041$ ,  $I_3 = 0.499874$ .

13. Escribir el polinomio  $p(x) = 3 + 5x + x^2$  en la base de Legendre. Rpta:  $p(x) = \frac{10}{3}p_0(x) + 5p_1(x) + \frac{2}{3}p_2(x)$

14. (**Sympy**) Dada la aproximación de cuadratura gausiana:

$$\int_{-1}^1 f(x) = \sum_{k=0}^n w_k f(x_k), \quad (1.143)$$

donde  $w_0, w_1, \dots, w_n$  son los coeficientes ponderados ó pesos.

- (a) Halle los ceros de los primeros 20 polinomios de Legendre.
  - (b) Halle los pesos de ponderación para los primeros 20 polinomios de Legendre.
15. Estime la siguiente integral usando el método de cuadratura de Gaus-Legendre:

$$\int_0^\infty \frac{1}{x^4 + 1} \approx 1.110721 \quad (1.144)$$

*Hint:* Dividir la integral para tener dos integrales con límites [-1,1] y [0,1].

16. In the black-body radiation problem the following integral appears:

$$\int_0^\infty \frac{x^3}{e^x - 1} dx = \frac{\pi^4}{15}. \quad (1.145)$$

- Compute this integral using the Gauss-Laguerre quadrature method for  $n=3$  evaluation points.
- For this estimation, plot the relative error ( $\epsilon_r(n) = I_{estimated}(n)/I_{exact}$ ) as a function of the evaluation points, with  $n = [2, 3, \dots, 10]$  [1.5].

*Hint:* For the Gauss-Laguerre method, the weights are given by:

$$w_k = \frac{x_k}{(n+1)^2[L_{n+1}(x_k)]^2} \quad (1.146)$$

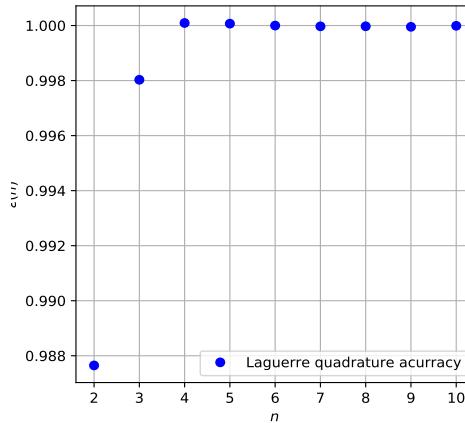


Figure 1.5: Accuracy vs n-points for the Gauss-Laguerre quadrature.

17. (Sympy) La cuadratura de Gaus-Hermite está definida para integrales de la forma:

$$I = \int_{-\infty}^{\infty} f(x)e^{-x^2} dx \quad (1.147)$$

que tiene la siguiente representación en cuadraturas:

$$I \approx \sum_{i=1}^N \omega_i f(x_i) \quad (1.148)$$

donde los puntos  $x_i$  son las raíces de los polinomios de Hermite ( $H_n(x)$ ) y la formula de pesos está dada por:

$$\omega_i = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 H_{n-1}(r_i)^2} \quad (1.149)$$

El estado de un oscilador armónico en mecánica cuántica está dado por la funciones de probabilidad:

$$\psi_n(\xi) = \frac{1}{\sqrt{2^n n!}} \left( \frac{m\omega}{\pi\hbar} \right)^{1/4} e^{-\xi^2/2} H_n(\xi) \quad (1.150)$$

donde  $\xi = \sqrt{\frac{m\omega}{\hbar}}x$ . Haga  $\sqrt{\frac{m\omega}{\hbar}} = 1$ , es decir,  $\xi = x$  para la aplicación numérica. Estime numéricamente el valor cuadrático medio de la posición de la partícula en el primer estado exitado ( $n = 1$ ). El valor exacto de la integral está dado por:

$$\langle x^2 \rangle = \int_{-\infty}^{\infty} |\psi_1(x)|^2 x^2 dx = 3/2 \quad (1.151)$$

El polinomio de Hermite de orden 1 está dado por:

$$H_1(x) = 2x \quad (1.152)$$

La formula de rodrigues que genera los polinomios de Hermite está dada por:

$$H_n(\xi) = (-1)^n e^{\xi^2} \frac{d^n}{d\xi^n} e^{-\xi^2} \quad (1.153)$$

Encontrar los ceros de los polinomios y los pesos de la cuadratura.



# Chapter 2

## Método de MonteCarlo

### Section 2.1

#### Generación pseudo-aleatoria de números

En general, no es posible generar auténticos números aleatorios por computador. Para obtener una secuencia aparentemente aleatoria, se utilizan métodos deterministas con altos períodos de repetición. Los métodos de congruencia lineal están definidos por:

$$x_{n+1} = (ax_n + c) \bmod m \quad n \geq 0, \quad (2.1)$$

donde  $0 \leq a < m$  se denomina multiplicador,  $0 \leq c < m$  es el incremento y  $m$  es el módulo del método. En la secuencia  $x_0$  se denomina la semilla del generador y debe ser ajustada *inteligentemente* para obtener secuencias diferentes en cada generación. En la expresión congruencial, el modulo representa el resto entre  $(ax_n + c)$  y  $m$ . Por ejemplo si  $x_0 = 1$ ,  $a = 16$ ,  $c = 0$  y  $m = 7$  tenemos la siguiente secuencia  $x_1 = 2$ ,  $x_2 = 4$ , .... Uno de los generadores más conocidos y usados tanto en C++ como en Python está definido por:

$$\begin{aligned} a &= 5DEECE66D \text{ (base 16)} \\ c &= B \text{ (base 16)} \\ m &= 2^{48} \end{aligned} \quad (2.2)$$

Para obtener un secuencia de números normalizada entre cero y uno ( $0 \leq r \leq 1$ ), se debe normalizar al modulo ( $m$ ) la ecuación congruencial.

$$r = \frac{x_{n+1}}{m} = \frac{(ax_n + c) \bmod m}{m} \quad (2.3)$$

Para obtener una distribución de número en otro intervalo cualquiera ( $A \leq x \leq B$ ) se tiene debe escalar la distribución  $\mathcal{U}(0, 1)$ .

$$x = A + (B - A)r. \quad (2.4)$$

Por otro lado, es importante garantizar la aleatoriedad de la secuencia, para tal propósito, se tienen algunas pruebas a la función de distribución de dicha secuencia [2], por ejemplo, podemos usar el momento de la distribución ( $P(x)$ ):

$$\frac{1}{N} \sum_{i=1}^N x_i^k \cong \int_0^1 x^k P(x) dx \cong \frac{1}{k+1} + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) \quad (2.5)$$

También se puede estudiar las correlación entre los  $k$ -vecinos cercanos, donde  $k$  es pequeño:

$$C(k) = \frac{1}{N} \sum_{i=1}^N x_i x_{i+k}, \quad k = 1, 2, \dots \quad (2.6)$$

Si la secuencia es aproximadamente uniforme podemos aproximar la suma como:

$$\frac{1}{N} \sum_{i=1}^N x_i x_{i+k} \cong \int_0^1 dx \int_0^1 dy xy P(x, y) = \frac{1}{4}. \quad (2.7)$$

De modo que, si la distribución de números pseudo-aleatorio sigue una función de distribución uniforme,  $C(k)$  debe tener como máximo  $1/4$ . La Figura [2.1] muestra las correlaciones de los primeros 30 vecinos para un generador de números aleatorios malo y uno bueno. Note que un buen generador de eventos debe mantener las correlaciones alrededor de  $1/4$ .

## Section 2.2

### Integración MonteCarlo

La integración Montecarlo es un método de integración de funciones generales que se sustenta en la ley de grandes números.

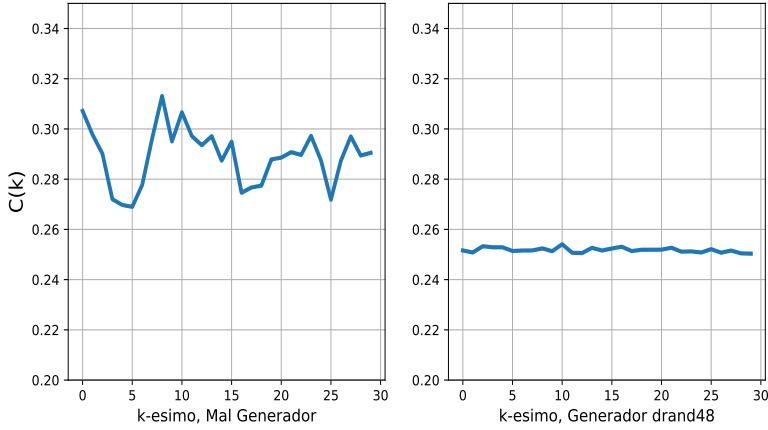


Figure 2.1: Correlaciones en los primeros  $k$ -vecinos para un mal generador (izquierda) y un buen generador (drand48) de números aleatorios.

### Section 2.3

## Ley de grandes números

Sea  $\{x_i\}$  un conjunto de puntos muestrales (observaciones o datos simulados por computador) y el promedio muestral de dicho conjunto.

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i. \quad (2.8)$$

Al aumentar el tamaño de la muestra, el promedio muestra converge al valor esperado con probabilidad  $i$ .

Muestral  $\rightarrow$  Poblacional

$$\bar{x} \rightarrow \mathbb{E}(x)$$

De esta manera, el operador integración  $\int_0^1 f(x)dx$  puede verse como el valor esperado de  $E(f(x))$  donde  $x \sim \mathcal{U}(0, 1)$  tiene una distribución uniforme. De este modo, tomemos una muestra  $x_1, x_2, \dots \sim \mathcal{U}(0, 1)$  y evaluamos en la función  $f(x_1), f(x_2), \dots$ , entonces:

$$\frac{1}{N} \sum_{i=1}^N f(x_i) \underset{N \rightarrow \infty}{\xrightarrow{\hspace{1cm}}} \mathbb{E}(f(x)) \quad (2.9)$$

El valor promedio de una función está dado por:

$$\bar{f} = \frac{1}{b-a} \int_a^b f(x)dx, \quad (2.10)$$

en consecuencia:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f(x_i) = \frac{1}{b-a} \int_a^b f(x)dx. \quad (2.11)$$

Otro resultado interesante es que si  $x$  es una variable aleatoria con densidad de probabilidad  $f$  y  $g$  es una función  $g : \mathbb{R} \rightarrow \mathbb{R}$ , entonces:

$$\mathbb{E}(g(x)) = \int_{-\infty}^{\infty} g(x)f(x)dx = \sum_{i=1}^N g(x_i)f(x_i). \quad (2.12)$$

#### Section 2.4

### Método de la transformada inversa

Sea  $x$  una variable aleatoria con densidad de probabilidad  $f$ , y sea  $F(x) = \int_{-\infty}^x f(s)ds$  la función acumulada de  $f$ . Dado que, la función acumulada de probabilidad es monótona creciente resulta ser inyectiva. De manera que, existe un único valor en el rango tal que:

$$F^{-1}(u) = x. \quad (2.13)$$

Por tanto, podemos generar una distribución uniforme y tomar la pre-imagen en la distribución acumulada. El resultado es el siguiente:

Sea  $U \sim \mathcal{U}(0, 1)$  y  $F$  una distribución acumulada, entonces  $F$  es inyectiva en algún intervalo pre-imagen de  $[0,1]$ . Si  $X$  se define como  $F^{-1}(U) = X$ , entonces  $X$  tiene distribución  $F$ .

**proof:**

Queremos ver si  $P(X \leq x) = F(x)$ , lo que significa que  $x$  tiene distribución  $F$ . Por definición de  $X$ :

$$P(X \leq x) = P(F^{-1}(U) \leq x) \quad (2.14)$$

Como  $F$  es monótona creciente en el intervalo  $[0,1]$  es invertible:

$$\begin{aligned} P(X \leq x) &= P(F(F^{-1}(U)) \leq F(x)) \\ &= P(U \leq F(x)) \\ &= F(x) \end{aligned} \tag{2.15}$$

El último paso se justifica dado que  $U$  tiene distribución uniforme  $U \sim \mathcal{U}(0, 1)$ . El algoritmo se resume como sigue:

1. Generar un número aleatorio que siga una distribución uniforme  $U \sim \mathcal{U}(0, 1)$ .
2. Tomar la pre-imagen de  $U$ ,  $X = F^{-1}(U)$ .

### Section 2.5

## Método de aceptación y rechazo

En la gran mayoría de los casos es imposible encontrar la función inversa para usar el método de transformada inversa. En estos casos, se utiliza el método de aceptación y rechazo, el cual mediante la distribución uniforme de números puede calcular el área bajo la curva de distribuciones complicadas. Este método consiste se base en los siguientes pasos:

1. Generamos un  $x_i$  que siga una distribución uniforme entre contenido en el intervalo de integración  $[a,b]$ .
2. Para  $x_i$ , generamos un  $y_i$  que siga una distribución uniforme entre 0 y el máximo de  $f(x)$ .
3. Si  $y_i < f(x_i)$  incluimos el valor  $x_i$  en la lista, de otro modo, no incluimos el valor  $x_i$ .

### Section 2.6

## Incertidumbre del método de MonteCarlo

La incertidumbre en el método de MonteCarlo se estima propagando el error cometido en cada uno de los puntos muestrales  $x_i$ . Recordemos que la media muestral está dada por:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \tag{2.16}$$

entonces:

$$\frac{\partial \bar{x}}{\partial x_i} = \frac{1}{N} \quad (2.17)$$

Teniendo en cuenta que cada medición es independiente la varianza asociada a la media muestra es:

$$Var(\bar{x}) = \sum_{i=1}^N \left( \frac{\partial \bar{x}}{\partial x_i} \right)^2 \delta_i^2 = \sum_{i=1}^N \frac{1}{N^2} \delta_i^2 \quad (2.18)$$

Por tanto, la incertidumbre está dada por:

$$\sqrt{Var(\bar{x})} \cong \frac{\delta}{\sqrt{N}} \quad (2.19)$$

$\delta$  es despreciable con el número de eventos simulados, en general se ajusta a 1 para tener una relación exacta con  $1/\sqrt{N}$ .

En esta ecuación es importante resaltar que la precisión del Método disminuye con  $\sqrt{N}$ , lo cuál debe tenerse en cuenta para cada calculo realizado con esta técnica. En el cálculo del número aureo usando la sucesión Fibonnaci aunque los puntos no provienen de un proceso aleatorio, la precisión del método tiene una dependencia similar.

### Section 2.7

## Ejercicios: Método de MonteCarlo

1. Demostrar la Ecuación (2.4).
2. Programar un mal generador de números aleatorios y usar drand48 en Python para reproducir la Figura [2.1].
3. Un test simple para probar la calidad de un generador de eventos es evaluar los momentos de la distribución de datos, la cual difiere del valor de la distribución uniforme en un orden de aproximación  $\mathcal{O}(\frac{1}{\sqrt{N}})$ ; esto se debe al error asociado al método de MonteCarlo.

$$\frac{1}{N} \sum_{i=1}^N x_i^k \cong \int_0^1 x^k P(x) dx \cong \frac{1}{k+1} + \mathcal{O}(\frac{1}{\sqrt{N}}) \quad (2.20)$$

- a) Implemente un código que estime los primeros  $k = 10$  momentos de la distribución de datos generados por Numpy. Genere `np.logspace(2,6,5)` eventos y grafique el eje x en escala logarítmica. Los momentos de la distribución de datos se muestran en la Figura [2.2]

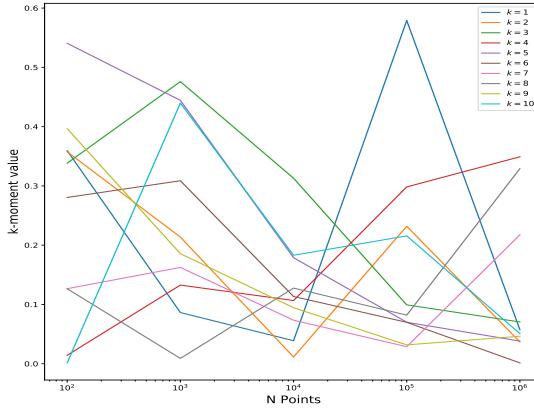


Figure 2.2: Primeros  $k = 10$  vecinos del generador Numpy como función del número de eventos generados.

4. Otro método para probar la calidad de un generador de eventos es evaluar las correlaciones con los  $k$ -vecinos más cercanos, donde  $k \sim 30$ .

$$C(k) = \frac{1}{N} \sum_{i=1}^N x_i x_{i+k}, \quad (k = 1, 2, 3\dots) \quad (2.21)$$

- a) Implemente un código que estime los coeficientes de correlación para los primeros  $k=30$  vecinos, con  $N = 10^4$  eventos de la distribución de datos generados por Numpy. Las correlaciones se muestran en la Figura [2.3].
5. Escriba un código que genere una secuencia uniforme de eventos, contenidas en una esfera de radio  $R = 2$  como se muestra en la Figura [2.4].
6. Usando la generación de puntos sobre una esfera estime la siguiente integral (en C++ y en Python), para  $\{x^2 + y^2 + z^2 \leq 1\}$ :

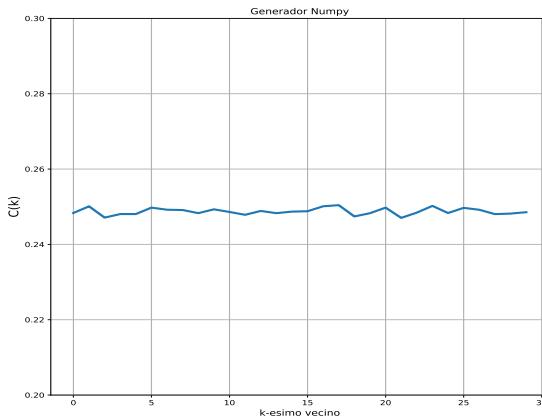


Figure 2.3: Correlaciones de los primeros  $k = 30$  vecinos del generador Numpy como función de  $k$ -esimo vecino. Note que el valor debe fluctuar alrededor del valor teórico  $C(k) = 1/4$ .

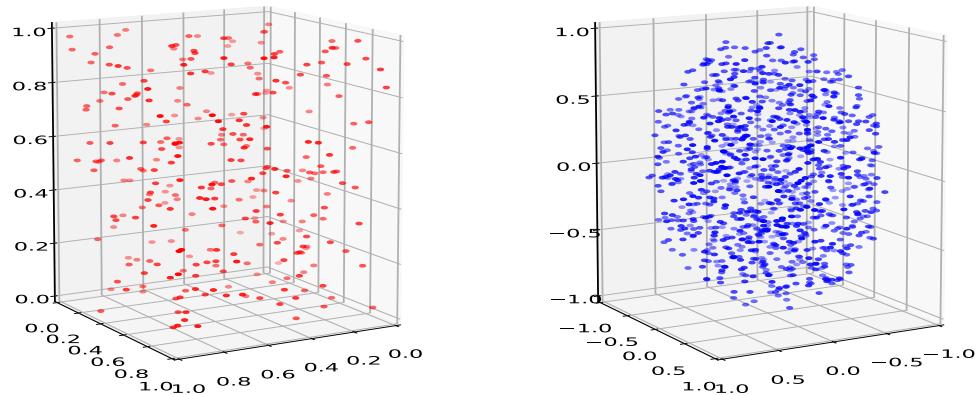


Figure 2.4: Generación de eventos cartesianas (izquierda) y esférica para  $R = 1$  (derecha).

$$\int \int \int e^{\sqrt{x^2+y^2+z^2}} dx dy dz = 4\pi(e-2) \quad (2.22)$$

7. Haga una gráfica entre la precisión  $\epsilon(I_{estimated}/I_{exact})$  y el número de puntos muestrales, el cuál debe tener un rango de valores adecuado para ver el comportamiento. ¿Qué dependencia se observa?
8. La distribución Beta está dada por:

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 \leq x \leq 1 \quad (2.23)$$

donde  $\Gamma(n) = (n - 1)!$ . Para  $f(x; 2,4)$ , halle el área bajo la curva usando el método de aceptación y rechazo con una incertidumbre del 1%.

9. La siguiente integral multidimensional:

$$\int_0^1 \cdots \int_0^1 2^{-7} \left( \sum_{i=1}^8 x_i \right)^2 dx_1 dx_2 \dots dx_8, \quad (2.24)$$

tiene el valor exacto  $\frac{25}{192}$ , usando el método de MonteCarlo estime esta integral con tres cifras de precisión.



# Chapter 3

## Álgebra lineal

### Section 3.1

#### Método de Jacobi

El método de Jacobi es usado para resolver sistemas lineales del tipo  $\mathbb{A}x = b$ , el cuál construye una sucesión descomponiendo la matriz  $\mathbb{A}$ .

$$\mathbb{A} = \mathbb{D} + \mathbb{R} \quad (3.1)$$

donde  $\mathbb{D}$  es la matriz diagonal y  $\mathbb{R}$  es la suma de la matriz triangular inferior y la matriz triangular superior.

$$\begin{aligned} \mathbb{D}x &= \mathbb{R}x + b \\ x &= \mathbb{D}^{-1}(b - \mathbb{R}x) \end{aligned} \quad (3.2)$$

si  $a_{ii} \neq 0$  tenemos un método iterativo.

$$x_i^{k+1} = \frac{1}{a_{ii}} \left( b_i - \sum_{j \neq i} a_{ij} x_j^k \right) \quad (3.3)$$

A nivel iterativo se necesitan todos los valores del vector  $x$  en el paso  $k$  para calcular el valor del vector en  $k + 1$ .

**Ejemplo:** Sea el sistema lineal:

$$\begin{aligned}
 3x - y - z &= 1 \\
 -x + 3y + z &= 3 \\
 2x + y + 4z &= 7
 \end{aligned} \tag{3.4}$$

encontrar la ecuación iterativa del método de Jacobi.

$$\mathbb{D}^{-1}b = \begin{pmatrix} 1/3 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/4 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 7 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1 \\ 7/4 \end{pmatrix} \tag{3.5}$$

$$\mathbb{D}^{-1}\mathbb{R} = \begin{pmatrix} 1/3 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/4 \end{pmatrix} \begin{pmatrix} 0 & -1 & -1 \\ -1 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1/3 & -1/3 \\ -1/3 & 0 & 1/3 \\ 1/2 & 1/4 & 0 \end{pmatrix} \tag{3.6}$$

Finalmente:

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \\ z^{k+1} \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1 \\ 7/4 \end{pmatrix} - \begin{pmatrix} 0 & -1/3 & -1/3 \\ -1/3 & 0 & 1/3 \\ 1/2 & 1/4 & 0 \end{pmatrix} \begin{pmatrix} x^k \\ y^k \\ z^k \end{pmatrix} \tag{3.7}$$

Este método se puede interpretar como una sucesión de vectores que ocurre en un punto arbitrario del espacio vectorial al vector solución, la cual es generada por un operador de traslación  $\mathbb{T}$  definido por el sistema de ecuaciones.

$$x^{k+1} = c + \mathbb{T}x^k \tag{3.8}$$

A nivel computacional para detener el método se calcula una métrica entre el vector solución y el vector calculado. Se define el residuo como:

$$Re = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{3.9}$$

para detener el método se usa la siguiente condición:

$$\|\mathbf{b} - \mathbb{A}\mathbf{x}\| < \delta \tag{3.10}$$

**Convergencia débil:** El método de Jacobi converge a la solución del problema lineal si la matriz asociada al sistema de ecuaciones es diagonal dominante.

**Convergencia Fuerte:** La sucesión  $x^{k+1} = \mathbb{T}x^k + c$ , para  $k \geq 0$  converge a la sucesión única  $x = \mathbb{T}x + c$  si y sólo si  $\rho(\mathbb{T}) < 1$ .

### Section 3.2

## Norma matricial

Una norma matricial sobre el conjunto de las matrices  $n \times n$  ( $M_n(\mathbb{R})$ ) es una aplicación:

$$\| \cdot \| : M_n(\mathbb{R}) \longrightarrow \mathbb{R}$$

que satisface las siguientes propiedades:

- a)  $\|\mathbb{A}\| \geq 0$ ,  $\forall \mathbb{A} \in M_n(\mathbb{R})$ .
- b)  $\|\mathbb{A}\| = 0$ , si y solo si  $a_{ij} = 0$ ,  $1 \leq i, j \leq n$ .
- c)  $\|\alpha \mathbb{A}\| = |\alpha| \|\mathbb{A}\|$   $\forall \alpha \in \mathbb{R}$  y  $\forall \mathbb{A} \in M_n(\mathbb{R})$ .
- d)  $\|\mathbb{A} + \mathbb{B}\| \leq \|\mathbb{A}\| + \|\mathbb{B}\|$ ,  $\forall \mathbb{A}, \mathbb{B} \in M_n(\mathbb{R})$ .
- e)  $\|\mathbb{A} \cdot \mathbb{B}\| \leq \|\mathbb{A}\| \cdot \|\mathbb{B}\|$ ,  $\forall \mathbb{A}, \mathbb{B} \in M_n(\mathbb{R})$ .

Tenemos un par de ejemplos útiles:

$$\|\mathbb{A}\|_1 = \underbrace{\max}_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}| \quad (3.11)$$

$$\|\mathbb{A}\|_\infty = \underbrace{\max}_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}| \quad (3.12)$$

Hagamos un ejemplo específico:

$$\mathbb{A} = \begin{pmatrix} -1 & 0 & 3 \\ 4 & -2 & 1 \\ 0 & 6 & 5 \end{pmatrix} \quad (3.13)$$

Hence,

$$\|\mathbb{A}\|_1 = \max[5, 8, 9] = 9 \quad (3.14)$$

$$\|\mathbb{A}\|_\infty = \max[4, 7, 11] = 11 \quad (3.15)$$

### 3.2.1 Radio espectral

El radio espectral de una matriz  $\mathbb{A} \in M_n(\mathbb{R})$ , con notación  $\rho(A)$ , se define como:

$$\rho(\mathbb{A}) = \max\{|\lambda| : \lambda \text{ valor propio de } \mathbb{A}\} \quad (3.16)$$

Recordar que los valores propios de una matriz  $\mathbb{A}$  son las raíces de la ecuación característica  $\det(\mathbb{A} - \lambda\mathbb{I}) = 0$ . Note que, si el valor propio es un número complejo, digamos  $\lambda = \alpha + i\beta$ , con  $\alpha, \beta \in \mathbb{R}$ , entonces  $|\lambda| = \sqrt{\alpha^2 + \beta^2}$ .

*El radio espectral no es una norma! puesto que  $\rho(\mathbb{A}) = 0$  con  $\mathbb{A} \neq 0$*

**Teorema:** Si  $\mathbb{A} \in M_n(\mathbb{R})$  entonces,

$$\rho(\mathbb{A}) \leq \|\mathbb{A}\|, \quad (3.17)$$

para cualquier norma matricial inducida definida por:

$$\|\mathbb{A}\| = \underbrace{\max}_{x \neq 0} \frac{\|\mathbb{A}x\|}{\|x\|} \quad (3.18)$$

Elijamos el auto-vector  $x$  asociado al radio espectral de  $\mathbb{A}$ .

$$\rho(\mathbb{A}) = \frac{\|\mathbb{A}x\|}{\|x\|} = \frac{\|\lambda x\|}{\|x\|} = \frac{|\lambda| \|x\|}{\|x\|} = |\lambda| \quad (3.19)$$

cuando elegimos un valor propio diferente al valor propio mas grande tenemos:

$$\rho(\mathbb{A}) \leq \|\mathbb{A}\|. \quad (3.20)$$

Necesitamos un último resultado:

**Teorema:** Sea  $\mathbb{A} \in M_n(\mathbb{R})$  entonces,

$$\lim_{k \rightarrow \infty} \mathbb{A}^k = 0 \text{ si y solo si } \rho(\mathbb{A}) < 1. \quad (3.21)$$

consideremos que  $\lim_{k \rightarrow \infty} \mathbb{A}^k = 0$  y tomemos algún valor de propio de  $\mathbb{A}$ , entonces  $\mathbb{A}^k x = \lambda^k x$  con  $x \neq 0$ . Entonces, necesariamente  $\lim_{k \rightarrow \infty} \lambda^k = 0$ ; lo cual se cumple solo si  $|\lambda| < 1$ . Como  $\lambda$  es un valor propio cualquiera y por la definición de radio espectral tenemos que  $\rho(\mathbb{A}) < 1$ .

Finalmente, para entender la convergencia fuerte definamos el error en la aproximación  $k$ -ésima:

$$e^k = x(k) - x \quad (3.22)$$

$x$  representa el vector solución. Por otro lado,  $x = Tx + c$  de manera que:

$$e^{k+1} = x^{k+1} - x = \mathbb{T}x^k + c - (\mathbb{T}x + c) = \mathbb{T}(x^k - x) = \mathbb{T}e^k \quad (3.23)$$

por lo que podemos ver el error en el paso  $k$  como la sucesión matricial por el error inicial  $e^k = \mathbb{T}^k e^0$ . Debido al teorema anterior,  $e^k = 0$  solo sí  $\lim_{k \rightarrow \infty} \mathbb{T}^k = 0$  y eso solo es posible si  $\rho(\mathbb{T}) < 1$ .

### Section 3.3

## Método de Gauss-Seidel

EL método de Gauss-Seidel se basa en el método de Jacobi con la única diferencia que cada variable se actualiza con cada iteración interna. Usando el ejemplo anterior se evidencia como funciona:

$$\begin{aligned} x &= \frac{1+y+z}{3} \\ y &= \frac{3+x-z}{3} \\ z &= \frac{7-2x-y}{4} \end{aligned} \quad (3.24)$$

Si la iteración empieza en el vector nulo tenemos que los valores en la siguiente iteración son:

$$\begin{aligned}
 x &= \frac{1+0+0}{3} = 1/3 \\
 y &= \frac{3+1/3+0}{3} = 10/9 \\
 z &= \frac{7-2/3-10/9}{4} = 47/36
 \end{aligned}
 \tag{3.25}$$

Teniendo la misma condición de convergencia del método de Jacobi, el método de Gauss-Seidel acelera el tiempo de convergencia a la solución. En general se tiene:

$$\begin{aligned}
 x^{k+1} &= \frac{1+y^k+z^k}{3} \\
 y^{k+1} &= \frac{3+x^{k+1}-z^k}{3} \\
 z^{k+1} &= \frac{7-2x^{k+1}-y^{k+1}}{4}
 \end{aligned}
 \tag{3.26}$$

#### Section 3.4

### Método de relajación sucesiva

Estos métodos realizan un ponderación de la solución de un sistema lineal en una iteración específica. La expresión general para un peso  $w$  dado es:

$$x_i^j = wx_i^j + (1-w)x_i^{j-1} \tag{3.27}$$

donde el índice  $i$  se relaciona al valor incógnita y  $j$  al valor de la iteración. Dependiendo del valor del peso, se tiene la siguiente calificación.

- Sub-relajación  $0 < w < 1$ . No convergente (para movimiento de fluidos puede converger).
- No modificado  $w = 1$ .
- Sobre-relajación  $1 < w < 2$ . Acelera convergencia.
- Divergente  $w > 2$ .

Encontrar el valor adecuado del peso dada una discretización es un asunto complicado del álgebra matricial numérica. Por ejemplo, para una discretización cuadrada de  $N$  puntos, el peso optimo está dado por:

$$w = \frac{2}{1 + \frac{\pi}{N}} \quad (3.28)$$

**Ejemplo:**

Encontrar  $x_3^1$  usando el método de sobre-relajación del siguiente sistema de ecuaciones lineales. Considere  $w = 1.25$ . y  $\vec{x}_0 = (0, 0, 0)$ .

$$\begin{aligned} 4x_1 - x_2 &= 0 \\ -x_1 + 4x_2 - x_3 &= 6 \\ -x_2 + 4x_3 &= 2 \end{aligned} \quad (3.29)$$

Usando el acercamiento por Gauss-Seidel tenemos:

$$\begin{aligned} x_1^1 &= \frac{2 + x_2^0}{4} = \frac{2 + 0}{4} = 0.5 \\ x_2^1 &= \frac{6 + x_1^1 + x_3^0}{4} = \frac{6 + 0.5 + 0}{4} = 1.625 \\ x_3^1 &= \frac{2 + x_2^1}{4} = \frac{2 + 1.625}{4} = 0.906 \end{aligned} \quad (3.30)$$

Ahora usemos sobre-relajación (para mostrar como funciona).

$$\begin{aligned} x_1^1 &= \frac{2 + x_2^0}{4} = \frac{2 + 0}{4} = 0.5 \\ x_1^1 &= (1.25)(0.5) + (1 - 1.25)(0) = 0.625 \\ x_2^1 &= \frac{6 + x_1^1 + x_3^0}{4} = \frac{6 + 0.625 + 0}{4} = 1.656 \\ x_2^1 &= (1.25)(1.656) + (1 - 1.25)(0) = 2.070 \\ x_3^1 &= \frac{2 + x_2^1}{4} = \frac{2 + 2.070}{4} = 1.017 \\ x_3^1 &= (1.25)(1.017) + (1 - 1.25)(0) = 1.271 \end{aligned} \quad (3.31)$$

Aunque la solución exacta para  $x_3 = 27/28$ , el método de sobre-relajación parece más distantes de la solución exacta. Sin embargo, se deja al lector la implementación de ambas

rutinas para verificar la potencial del método de relajación sucesiva. Adicionalmente, es posible imponer un criterio de convergencia del método:

$$\left| \frac{x_i^j - x_i^{j-1}}{x_i^j} \right| < \epsilon \quad (3.32)$$

donde  $\epsilon$  es arbitrariamente pequeño. Podemos pensar en la siguiente factorización para la implementación de la relajación sucesiva en ecuaciones diferenciales parciales:

$$u_{i,j}^l = u_{i,j}^{l-1} + w[u_{i,j}^l - u_{i,j}^{l-1}] \quad (3.33)$$

definiendo  $r_{i,j} := u_{i,j}^l - u_{i,j}^{l-1}$  se tiene:

$$u_{i,j}^l = u_{i,j}^{l-1} + wr_{i,j} \quad (3.34)$$

Generalmente  $r_{i,j} \leq 10^{-10}$ .

### Section 3.5

## Método generalizado de Newton-Raphson

Sea el sistema de ecuaciones lineales o no lineales:

$$\begin{aligned} f_1(x, y) &= 0 \\ f_2(x, y) &= 0 \end{aligned} \quad (3.35)$$

Suponemos que las funciones son diferenciables, de modo que podemos expandir en series de Taylor alrededor de  $(x_0, y_0)$  a primer orden. Tenemos:

$$\begin{aligned} f_1(x_0, y_0) + \frac{\partial f_1(x_0, y_0)}{\partial x} \Delta x + \frac{\partial f_1(x_0, y_0)}{\partial y} \Delta y &\approx 0 \\ f_2(x_0, y_0) + \frac{\partial f_2(x_0, y_0)}{\partial x} \Delta x + \frac{\partial f_2(x_0, y_0)}{\partial y} \Delta y &\approx 0 \end{aligned} \quad (3.36)$$

con  $x = x_0 + \Delta x$  y  $y = y_0 + \Delta y$ . Vectorizando estas ecuaciones  $\vec{x} = (x, y)$ , tenemos:

$$\vec{F}(\vec{x}_0) + \mathbb{J}(x_0)(\vec{x} - \vec{x}_0) \approx 0 \quad (3.37)$$

Donde definimos el Jacobiano del sistema:

$$\mathbb{J} = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix} \quad (3.38)$$

Si el Jacobiano del sistema es invertible, tenemos:

$$\vec{x} = \vec{x}_0 - \mathbb{J}^{-1}(\vec{x}_0) \vec{F}(\vec{x}_0) \quad (3.39)$$

La solución al sistema de ecuaciones significa encontrar la sucesión de Cauchy dada por:

$$\vec{x}_{n+1} = \vec{x}_n - \mathbb{J}^{-1}(\vec{x}_n) \vec{F}(\vec{x}_n). \quad (3.40)$$

### Section 3.6

## Método del descenso gradiente

Vamos a pensar el problema de sistemas de ecuaciones como un vector de funciones. Sea  $\vec{G}(\vec{x})$ :

$$\vec{G}(\vec{x}) = \begin{pmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ \vdots \\ f_n(x_1, x_2, \dots, x_n) \end{pmatrix} \quad (3.41)$$

Este vector tendrá una norma asociada. Se define una función a minimizar usando dicha norma.

$$\vec{F}(\vec{x}) = \frac{1}{2} \vec{G}^T(\vec{x}) \vec{G}(\vec{x}) \quad (3.42)$$

La solución al sistema de ecuaciones estará definida por:

$$\|\vec{F}(\vec{x})\| = \vec{0}. \quad (3.43)$$

Debemos movernos en contra del gradiente de la función vectorial.

$$\vec{x}^1 = \vec{x}^0 - \gamma \nabla \vec{F}(\vec{x}_0), \quad (3.44)$$

donde  $\gamma$  se denomina tasa de aprendizaje. Lo que brinda la información direccional del gradiente de una función vectorial es la matriz jacobiana.

$$J_{ij} = \frac{\partial f_i}{\partial x_j} \quad (3.45)$$

El jacobiano es la matriz de derivadas parciales de primer orden de las funciones del vector  $\vec{G}(\vec{x})$ . Suponemos  $\vec{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  y que es diferenciable, entonces  $x \in \mathbb{R}^n$  y  $\vec{F}(\vec{x}) \in \mathbb{R}^m$ . Finalmente, el descenso del gradiente sobre una función vectorial está dada por:

$$\vec{x}^{n+1} = \vec{x}^n - \gamma \mathbb{J}^T(\vec{x}^n) \cdot \vec{G}(\vec{x}^n). \quad (3.46)$$

Note que la función gradiente es el transpuesto del Jacobiano, debido a que el gradiente es un vector fila definido sobre el complemento ortogonal; mientras que el vector de soluciones es un vector columna.

### Section 3.7

## Propagación del error en una red neuronal

Comencemos por las definiciones. Tenemos que el vector  $\delta^l$  está definido como:

$$\delta_i^l = \frac{\partial C}{\partial z_i^l},$$

donde  $z^l$  se define más adelante. Sea  $a^l$  los valores de las funciones de activación en la  $l$ -ésima capa. Tenemos que la función de costo es  $C = C(a^L)$ , ya que esta solo depende de la función de activación en la última capa. Adicionalmente, sea  $z^l$  el input de la capa  $l$ . Siendo  $f$  la función de activación, tenemos que

$$a^l = f^l(z^l), \quad z^l = W^l f^{l-1}(z^{l-1}).$$

Utilizamos el vector de derivadas como la derivada con respecto al vector. Esto es:

$$\left( \frac{\partial C}{\partial z^l} \right)_i = \frac{\partial C}{\partial z_i^l}.$$

Usando la regla de la cadena, tenemos que

$$\frac{\partial C}{\partial z_i^L} = \frac{\partial C}{\partial a^L} \frac{\partial a^L}{\partial z_i^L}.$$

En notación de vectores, por ende, tenemos que

$$\delta^L = \frac{\partial C}{\partial a^L} \frac{\partial a^L}{\partial z^L}.$$

Para las demás capas, nótese que  $z^{l-1}$  solo afecta a  $C$  a través de  $z^l$ . Esto quiere decir que tenemos

$$\delta_i^{l-1} = \frac{\partial C}{\partial z_i^{l-1}} = \sum_j \frac{\partial C}{\partial z_j^l} \frac{\partial z_j^l}{\partial z_i^{l-1}}.$$

Ahora, tenemos que los  $a^l$  son, en general, vectores.  $a_k^l$  sería la función de activación de la  $k$ -ésima neurona de la capa  $l$ . Similarmente, tenemos que  $W_{jk}^l$  sería el peso que la  $j$ -ésima neurona de la capa  $l$  le asigna a la  $k$ -ésima salida de la capa  $l - 1$ . Por ende, tenemos que

$$z_j^l = \sum_k W_{jk}^l a_k^{l-1}.$$

Derivando con respecto a  $z_i^{l-1}$ :

$$\frac{\partial z_j^l}{\partial z_i^{l-1}} = \sum_k W_{jk}^l \frac{\partial a_k^{l-1}}{\partial z_i^{l-1}}.$$

Reemplazando en nuestra expresión para  $\delta_i^l$ :

$$\delta_i^{l-1} = \sum_j \sum_k \frac{\partial C}{\partial z_j^l} W_{jk}^l \frac{\partial a_k^{l-1}}{\partial z_i^{l-1}}.$$

Reconocemos  $\delta_j^l$  en la suma, tal que

$$\delta_i^{l-1} = \sum_j \sum_k \delta_j^l W_{jk}^l \frac{\partial a_k^{l-1}}{\partial z_i^{l-1}}.$$

Generalizando la notación de derivadas de vectores introduciendo antes, definimos la derivada de un vector con respecto a otro como

$$\frac{\partial a_k^{l-1}}{\partial z_i^{l-1}} = \left( \frac{\partial a^{l-1}}{\partial z^{l-1}} \right)_{ki}.$$

Representando  $\delta^l$  como una matriz y reemplazando con esta definición, tenemos que

$$\delta_i^{l-1} = \sum_j \sum_k (\delta^l)_{j1} W_{jk}^l \left( \frac{\partial a^{l-1}}{\partial z^{l-1}} \right)_{ki}.$$

Transponemos  $\delta^l$  para obtener un producto matricial:

$$\delta_i^{l-1} = \sum_j \sum_k ((\delta^l)^T)_{1j} W_{jk}^l \left( \frac{\partial a^{l-1}}{\partial z^{l-1}} \right)_{ki}.$$

Reconociendo el producto matricial, llegamos al resultado deseado:

$$\delta^{l-1} = (\delta^l)^T W^l \frac{\partial a^{l-1}}{\partial z^{l-1}}.$$

### Section 3.8

## Ejercicios: Álgebra lineal

1. Implemente el algoritmo de Gaus-Seidel para resolver el sistema de ecuaciones visto en clase.
2. En Python, implemente una clase para solucionar sistemas lineales. El constructor de clase debe recibir la matriz y el vector independiente. El primer método de clase resuelve el sistema usando el método de Jacobi, y el segundo método resuelve el sistema usando el método de Gaus-Seidel. En la pantalla debe aparecer el número de iteraciones necesarias para resolver el problema.
3. Implemente un algoritmo que realice la multiplicación de dos matrices. Use el algoritmo para calcular:

$$\mathbb{A}\mathbb{B} = \begin{pmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ -2 & 3 & 1 \end{pmatrix} \begin{pmatrix} 4 & -2 & 1 \\ 0 & 3 & 7 \\ 0 & 0 & 2 \end{pmatrix} \quad (3.47)$$

4. (Theoretical) Muestre con detalle que la sustitución hacia adelante se expresa como:

$$x_i = b_i - \sum_{j=0}^{i-1} A_{ij} x_j \quad (3.48)$$

5. (Theoretical) Muestre con detalle que la sustitución hacia atrás se expresa como:

$$x_i = \frac{b_i - \sum_{j=i+1}^n A_{ij} x_j}{A_{ii}} \quad (3.49)$$

donde  $i = n, n - 1, \dots, 0$ . Note que la diagonal de la matriz triangular superior puede tener cualquier valor.

6. Usando los métodos de Newton-Raphson y descenso del gradiente, encuentre la solución de los siguientes sistemas de ecuaciones no lineales:

$$\begin{aligned} \ln(x_1^2 + x_2^2) - \sin(x_1 x_2) &= \ln(2) + \ln(\pi), \\ e^{x_1 - x_2} + \cos(x_1 x_2) &= 0. \end{aligned} \quad (3.50)$$

Use  $\mathbf{x}^{(0)} = (2, 2)$

$$\begin{aligned} 6x_1 - 2\cos(x_2 x_3) - 1 &= 0, \\ 9x_2 + \sqrt{x_1^2 + \sin(x_3) + 1.06} + 0.9 &= 0, \\ 60x_3 + 3e^{-x_1 x_2} + 10\pi - 3 &= 0. \end{aligned} \quad (3.51)$$

Use  $\mathbf{x}^{(0)} = (0, 0, 0)$

Ejercicios tomados de Numerical Analysis - Burden and Faires, Ninth Edition, Chapter 10.

7. (**Machine Learning application**) La 3-esfera está definida por todos los puntos equidistantes de un punto dado en  $\mathbb{R}^4$ , generalmente definida por:

$$S^3 = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1\} \quad (3.52)$$

Dado que no podemos visualizar la 3-esfera, se usa el método de descenso del gradiente para encontrar sistemáticamente los puntos que pertenecen a  $S^3$ . Posteriormente, se hace una proyección sobre  $\mathbb{R}^3$ . Para hacer el mapeo de  $S^3$  se sugiere la siguiente estrategia:

- a) Modifique el código de la clase para que  $S^3$  sea el único elemento del vector de funciones.
- b) Calcule la métrica de minimización como la magnitud del vector de funciones.
- c) Modifique el Jacobiano a una dimensión ( $1 \times 4$ ).

- d) Genere un punto aleatorio uniformemente distribuido en  $\mathbb{R}^4$  (es decir, un vector aleatorio de 4 componentes) en el intervalo  $[-1,1]$ .
- e) Corra el algoritmo de aprendizaje para que el vector aleatorio se ubique en algún punto de  $S^3$ . El código *aprende* a mover el punto minimizando la métrica.
- f) Repita el proceso de aprendizaje para al menos  $10^3$  puntos.
- g) Verifique que los puntos obtenidos pertenezcan a la 3-esfera.
- h) Con los puntos obtenidos de  $S^3$ , dibuje la proyección espacial  $(X, Y, Z)$ , es decir, las primeras tres columnas. ¿Qué figura geométrica resulta?

*Hint:* Primero haga el caso de la 2-esfera para entender el problema. Note que la proyección de la 2-esfera en el plano es el círculo.

8. La Figura [3.1] representa la colisión de dos discos de masa  $m_1$  y  $m_2$ , y radios  $r_1$  y  $r_2$  respectivamente. El primer disco tiene una velocidad inicial  $\vec{u}_1$  antes del choque y el segundo disco se encuentra en reposo  $\vec{u}_2 = 0$ . El parámetro de impacto  $b$ , que es la distancia entre la dirección de la velocidad  $\vec{u}_1$  y el centro del segundo disco caracteriza la interacción.

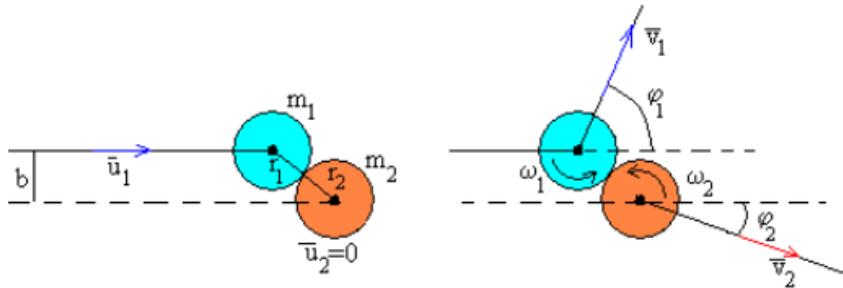


Figure 3.1: Esquema de colisión de dos discos rígidos.

Los siguientes son resultados teóricos que deben deducir y justificar físicamente.

- a) De la conservación del momento lineal demuestre que:

$$m_1 u_1 = m_1 v_{1x} + m_2 v_{2x} \quad (3.53)$$

$$0 = m_1 v_{1y} + m_2 v_{2y} \quad (3.54)$$

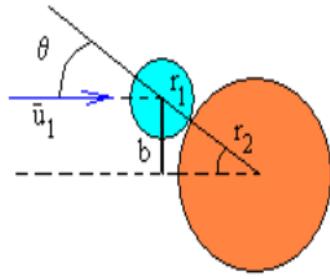


Figure 3.2: Esquema para la conservación del momentum angular.

- b) De la conservación del momento angular respecto del punto de contacto demuestre que (ver Figura [3.2]):

$$-m_1r_1u_1\sin(\theta) = I_1\omega_1 - m_1r_1(v_{1x}\sin(\theta) + v_{1y}\cos(\theta)) \quad (3.55)$$

$$0 = I_2\omega_2 + m_2r_2(v_{2x}\sin(\theta) + v_{2y}\cos(\theta)) \quad (3.56)$$

- c) De la definición del coeficiente de restitución demuestre que:

$$e = -\frac{(v_{1x}\cos(\theta) - v_{1y}\sin(\theta)) - (v_{2x}\cos(\theta) - v_{2y}\sin(\theta))}{u_1\cos(\theta)} \quad (3.57)$$

- d) Si no se considera deslizamiento de un disco respecto al otro en el punto de contacto  $P$ . Las velocidades de los discos en el punto  $P$  son iguales. Muestre que esta restricción conduce a:

$$r_1w_1 + v_{1x}\sin(\theta) + v_{1y}\cos(\theta) = -r_2w_2 + v_{2x}\sin(\theta) + v_{2y}\cos(\theta) \quad (3.58)$$

El sistema lineal queda descrito como:

$$\begin{pmatrix} \sin\theta & \cos\theta & -\sin\theta & -\cos\theta & 1 & 1 \\ 0 & M & 0 & 1 & 0 & 0 \\ M & 0 & 1 & 0 & 0 & 0 \\ -\cos\theta & \sin\theta & \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 & -k & 0 \\ 0 & 0 & \sin\theta & \cos\theta & 0 & k \end{pmatrix} \cdot \begin{pmatrix} v_{1x} \\ v_{1y} \\ v_{2x} \\ v_{2y} \\ r_1\omega_1 \\ r_2\omega_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ Mu_1 \\ eu_1\cos\theta \\ u_1\sin\theta \\ 0 \end{pmatrix}$$

Parámetros	
$m_1$	3 kg
$m_2$	1 kg
$M$	3
$r_1$	0.1 m
$r_2$	0.2 m
$k$	1/2
$e$	0.8
$\vec{u}_1$	$2\hat{i}$ m/s

Table 3.1: Parámetros del cálculo. Recuerde que el momento de inercia de un sólido respecto a su centro de masas es  $I = kmr^2$ .

Para la aplicación numérica definamos los parámetros del problema:

Variar el ángulo de incidencia ( $0 < \theta < \pi/2$ ) para describir algunos estados finales después del choque. Debe verificar que los principios de conservación se cumplan en este fenómeno, esto significa que el momento lineal y angular se mantengan constantes. La energía no se conserva porque hay restitución ( $e$ ) en la colisión, si todo va bien, graficar las seis variables cinemáticas de los discos para entender como se mueven los cuerpos después del proceso de "scattering". *Sería muy interesante realizar la animación de este fenómeno, sin embargo, las dificultades para animar la translación + rotación de los discos son enormes.*

Para que tengan una guía de los resultados, las graficas de este problema son:

9. En un instante de tiempo una estrella naciente está compuesta por un conjunto de masas unitarias ( $m = 1 \text{ kg}$ ). Las posiciones de cada una de las partículas están escritas en: <https://github.com/asegura4488/DataBase/blob/main/MetodosComputacionales/CuerposCelestes.dat>

Para un sistema de  $N$  partículas con masa  $m_k$  y posiciones  $r_k$ , el tensor de inercia se define como:

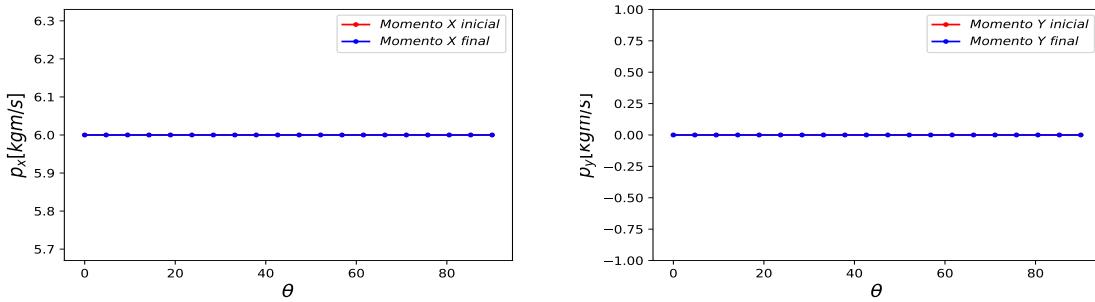


Figure 3.3: Conservación del momento lineal total antes y después del choque.

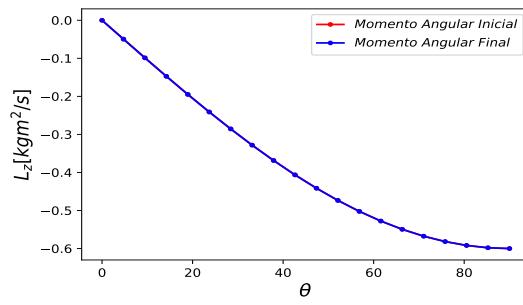


Figure 3.4: Conservación del momento angular total antes y después del choque.

$$\mathbf{I} = \sum_{k=1}^N m_k ((r_k \cdot r_k) \mathbf{E} - r_k \otimes r_k) \quad (3.59)$$

donde  $\mathbf{E}$  es el tensor unitario  $e_1 \otimes e_1 + e_2 \otimes e_2 + e_3 \otimes e_3$

- a) Muestre que el tensor de inercia para esta estrella está dado por:

$$\mathbf{I} = \begin{pmatrix} 1.1163 \times 10^4 & 2.0524 \times 10^0 & 7.4286 \times 10^{-1} \\ 2.0524 \times 10^0 & 8.9520 \times 10^3 & -2.3483 \times 10^3 \\ 7.4286 \times 10^{-1} & -2.3483 \times 10^3 & 4.2341 \times 10^3 \end{pmatrix}$$

Note que el tensor es simétrico  $I_{ij} = I_{ji}$  como es esperado.

- b) Encuentre los auto-valores y auto-vectores. Si dos valores propios son cercanos o iguales ¿Qué interpretación tienen los auto-valores respecto a las simetrías de este sistema?

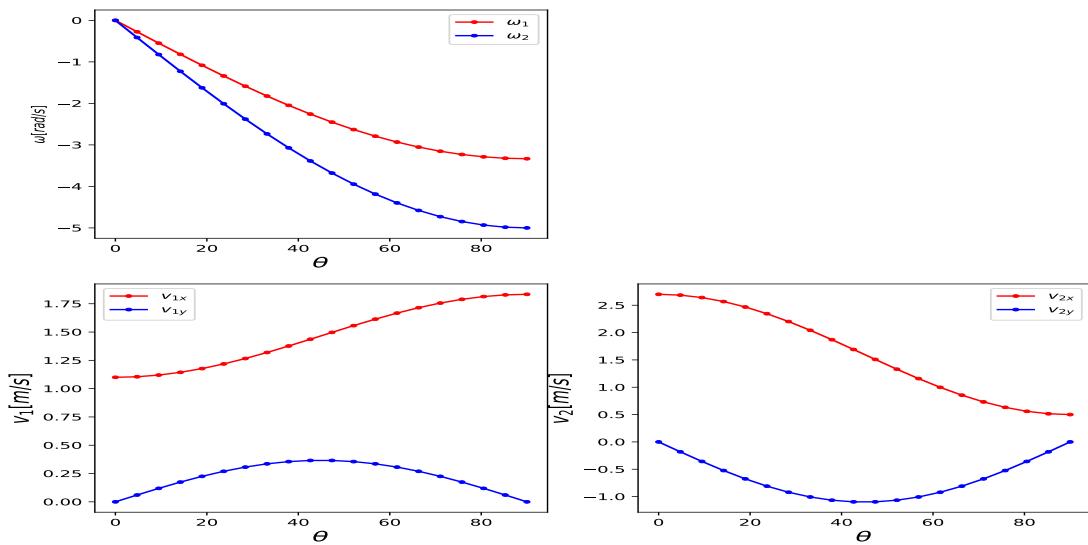


Figure 3.5: Variables cinemáticas los dos discos después del choque.

- c) Haga una gráfica en 3D de las estrella, los auto-vectores de la estrella y las estrella sobre sus ejes principales. (ver Figura [3.6])

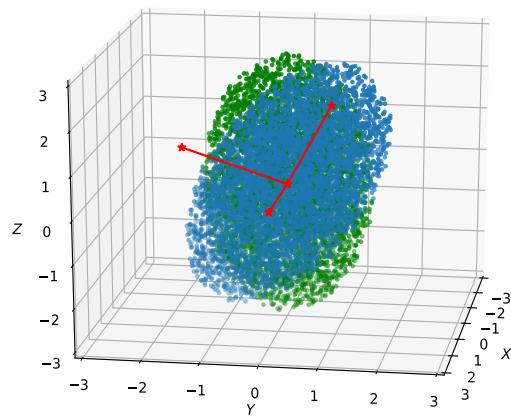


Figure 3.6: Estrella naciente: Puntos originales (azul), los ejes principales del cuerpo (rojo) y puntos sobre los ejes principales (verde.)

10. Imagine que uno de los espejos del telescopio espacial James Webb tiene una forma cuadrada como ilustra la Figura [3.7]. Los vértices están ubicados en las siguientes posiciones con sus respectivas temperaturas:

$$\begin{aligned}
 P_1 &= (+1, +1); & T(P_1) &= 1 \text{ K} \\
 P_2 &= (-1, +1); & T(P_2) &= 2 \text{ K} \\
 P_3 &= (-1, -1); & T(P_3) &= 0.5 \text{ K} \\
 P_4 &= (+1, -1); & T(P_4) &= 0.3 \text{ K}
 \end{aligned} \tag{3.60}$$

Lastimosamente en el punto  $P = (0, 0.5)$  se daño el sensor de la temperatura, lo que genera preocupación en los ingenieros de la NASA.

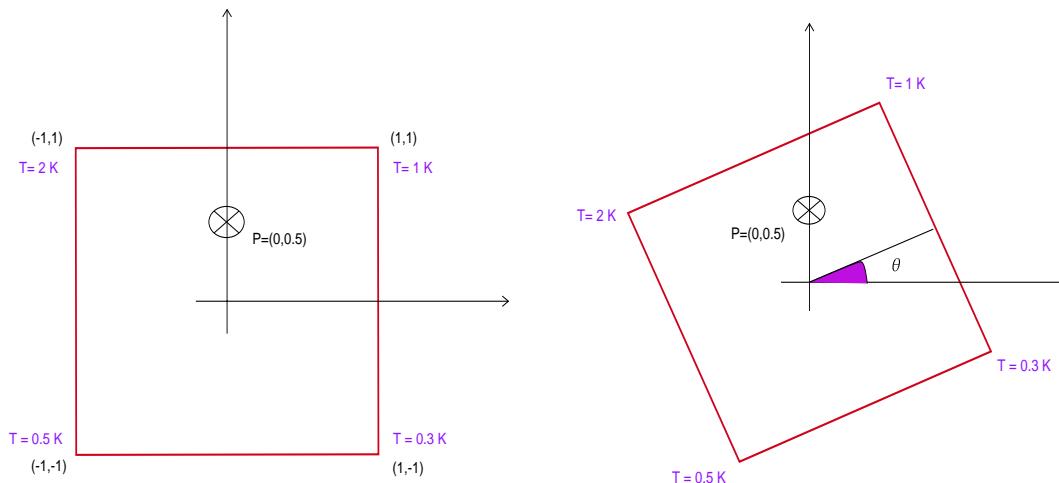


Figure 3.7: Esquema del espejo del telescopio espacial. Los vértices del cuadrado registran las temperaturas en grados kelvin. En la parte derecha, se ilustra una rotación para minimizar la temperatura en el punto  $P = (0, 0.5)$ .

La NASA quiere hacer una estimación de la temperatura en el punto donde se daño el sensor, usando la restricción de la temperatura en los vértices. Adicionalmente, se quiere rotar el espejo hasta que la temperatura tome el valor más bajo posible y garantizar la vida útil del espejo. Los estudiantes del curso de Métodos computacionales proponen calcular una interpolación espacial de la temperatura en el espejo para estimar la temperatura

en el punto  $P$ , y luego aplicar un operador de rotación para evaluar como cambia la temperatura en  $P$  a medida que se rota el espejo. Les propongo la siguiente estrategia:

- a) Proponga que la temperatura en la región del espejo está dada por una aproximación bilineal:

$$T(x, y) \cong \sum_{i=0}^1 \sum_{j=0}^1 a_{ij} x^i y^j \quad (3.61)$$

Programar esta función que define la interpolación de la temperatura en el espejo. Usar Sympy para verificar que la función es de la forma  $T(x, y) = a_{00} + a_{10}x + a_{01}y + a_{11}xy$ .

La función debe tener un parámetro  $p$  para llenar la matriz de coeficientes, tal que:

$$a_{00} = p_0, a_{10} = p_1, a_{01} = p_2, \text{ y } a_{11} = p_3.$$

- b) Defina el vector de posición de los vértices como: `position = np.zeros((4,2))`. Llene los vectores en el orden que aparecen en (3.60)
- c) Para calcular la matriz de coeficientes de la interpolación, debe resolver el problema matricial que aparece de exigir que la temperatura en los vértices tenga los valores presentados en (3.60). Por ejemplo, se debe cumplir  $1 = a_{00} + a_{10}(1) + a_{01}(1) + a_{11}(1)(1)$ .
- d) Verifique que la temperatura en los vértices cumpla con las condiciones de frontera.
- e) Grafique el mapa de temperatura en la región donde está el espejo. Debería obtener algo como: Figura [3.8].

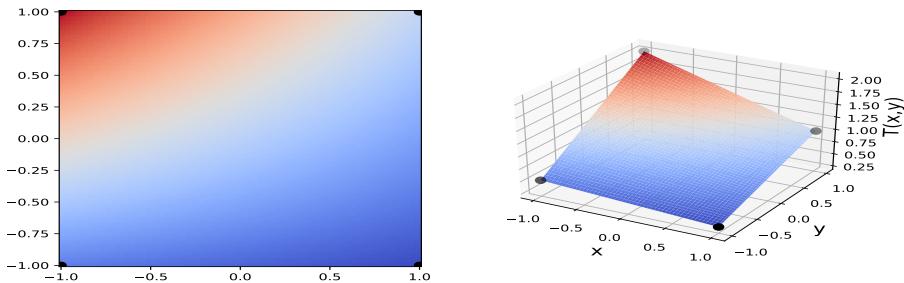


Figure 3.8: Mapa de temperatura al interior del espejo del telescopio.

- f) Estime la temperatura en el punto donde se daño el sensor:  $T(0, 0.5) \approx 1.225 K$ . (Si llegas hasta este punto tienes 3.0)

- g) Para el proceso de minimización, defina una función que rote el vector de los vértices del espejo.

$$\begin{pmatrix} P_x^i(\theta) \\ P_y^i(\theta) \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} P_x^i \\ P_y^i \end{pmatrix} \quad (3.62)$$

donde  $i = 0, 1, 2, 3$  es cada vértice del espejo. Dentro de la función debe realizar la rotación de los cuatro vértices.

- h) Una vez rote los cuatro vértices, realice nuevamente la interpolación usando los mismos valores de temperatura de los vértices y calcule la temperatura en el punto donde se daño el sensor.

- i) Defina una sucesión de ángulos, por ejemplo:

```
theta = np.linspace(0, 2*np.pi, 200),
```

¿Para cuál ángulo se minimiza la temperatura del espejo en el punto  $P$ . (Si encuentras el ángulo donde la temperatura en  $P$  es mínima tienes 5.0)

11. Descargue los datos del generador parabólico ubicados en: <https://github.com/asegura4488/DataBase/blob/main/MetodosComputacionales/knn.csv>.

- a) En cualquier fuente confiable de aprendizaje automático puede encontrar el algoritmo de k-vecinos más cercanos. Implemente una clase en Python que realice la clasificación de los puntos que caen fuera del círculo (0) y los puntos internos (1). Esta clase debe tener como mínimo un constructor, un método de entrenamiento y un predictor. Podría tener un método para calcular el *score* del método; pero es opcional.
- b) Mida el *score* del clasificador durante el entrenamiento (Training sample), que significa medir cuál es el porcentaje de eventos que se clasifican bien del total (My result is around 98% using  $k = 3$ , which is "quite good").
- c) Haga una optimización hiper-paramétrica del número de vecinos. Lo que significa variar el valor  $k$  hasta que el *score* sea lo más grande posible.
- d) Usando el código del generador parabólico genere una muestra nueva de 1000 eventos (Testing sample) y mida el *score* de la muestra de prueba. Esta sobre-entrenado (overfitting) su algoritmo? Justifique.

Este video podría ayudar: <https://www.youtube.com/watch?v=mQw2DCDsZVQ>

12. Modifique el código sobre redes neuronales para entrenar una red neuronal sencilla. La red neuronal debe comportarse como una compuerta XNOR. Dicha compuerta tiene la siguiente tabla de verdad:

A	B	Output
0	0	1
1	0	0
0	1	0
1	1	1

La red debe tener la topología más simple posible (un par de neuronas en un capa o dos neuronas en dos capas) y entrenada usando solo números aleatorios.

# Chapter 4

## Probabilidad

### Section 4.1

#### Definición de probabilidad

Para definir intuitivamente el concepto de probabilidad debemos definir el concepto de evento aleatorio. Un evento aleatorio es un suceso que puede ocurrir de diversas maneras, por ejemplo, ¿va a llover?, ¿cuánto va a tardar el próximo transporte?, ¿qué cara del lado va a salir?, etc. Para los antiguos griegos la ocurrencia de eventos se atribuye a la voluntad divina y por tanto no era posible determinar lo que iba a suceder. Por otro lado, los romanos consideraban que los eventos ocurren por cuestiones de suerte o azar. Esta naturaleza intrínseca de los eventos aleatorios llevan asociada un incertidumbre en el resultado. La probabilidad busca cuantificar el grado de plausibilidad de cada evento que puede suceder, en ese sentido, esta disciplina nos permite conocer aplicando algunas reglas, los eventos que pueden suceder con mayor frecuencia en un experimento.

Un evento en probabilidad está definido por el concepto de conjunto, el conjunto que contiene a todos los conjuntos se denomina espacio muestral  $\Omega$ . Tomemos un conjunto genérico  $A$  del espacio muestral, este conjunto debe pertenecer a una estructura denominada  $\sigma$ -álgebra; para ser considerado un evento válido. El concepto de  $\sigma$ -álgebra aparece en una rama de la matemática denominada teoría de la medida.

Sea  $\Omega$  un espacio muestral y definamos a  $\mathfrak{F}$  una colección de sub-conjuntos de  $\Omega$ . Esta colección se denomina  $\sigma$ -álgebra, si cumple las siguientes condiciones.

- $\emptyset \in \mathfrak{F}$ .
- $A \subset \mathfrak{F}$ , entonces  $A^c \subset \mathfrak{F}$ .
- $\{A_n\}_{n \in \mathbb{N}} \in \mathfrak{F}$ , entonces  $\cup_{i=1}^{\infty} A_i \in \mathfrak{F}$ .

De esta manera,  $(\Omega, \mathfrak{F})$  se denomina un espacio medible. Por ejemplo, tomemos el siguiente espacio medible:

$$\begin{aligned}\Omega &= \{0, 1, 2\} \\ \mathfrak{F} &= \{\emptyset, \Omega, \{0, 1\}, \{2\}\}\end{aligned}\tag{4.1}$$

El lector se puede convencer que se cumplen todas las condiciones de espacio medible. Dada la definición de espacio medible, podemos definir lo que es una medida de probabilidad. Sea  $\mathbb{P}$  una aplicación que mapea desde la  $\sigma$ -álgebra a los números reales  $\mathbb{R}$ .

$$\mathbb{P} : \mathfrak{F} \longrightarrow \mathbb{R}.\tag{4.2}$$

Esta aplicación asigna un número real a un conjunto de la  $\sigma$ -álgebra.

$$A \longrightarrow \mathbb{P}(A)\tag{4.3}$$

En 1933, Kolmogórov formula el sistema de axiomas para que los espacios medibles sean espacios de probabilidad. Los axiomas son los siguientes:

1.  $\mathbb{P}(\Omega) = 1$
2.  $\forall A \in \mathfrak{F}, \mathbb{P}(A) \geq 0$
3.  $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ , si  $A_i \cap A_j = \emptyset \forall i \neq j$ .

La estructura  $(\Omega, \mathfrak{F}, \mathbb{P})$  se denomina espacio de probabilidad. A nivel práctico, estamos interesados en contestar las siguientes preguntas:

1. ¿Cuál es la probabilidad de que ocurra  $A_1$  y  $A_2$ ?,  $\mathbb{P}(A_1 \cap A_2)$ .
2. ¿Cuál es la probabilidad de que ocurra  $A_1$  o  $A_2$ ?,  $\mathbb{P}(A_1 \cup A_2)$ .

3. ¿Cuál es la probabilidad de que ocurra  $A_1$  pero no ocurra  $A_2$ ,  $\mathbb{P}(A_1 - A_2)$ .
4. ¿Cuál es la probabilidad de que no ocurra  $A$ ,  $\mathbb{P}(A^c)$ .

Para contestar estas preguntas, se requiere establecer la regla de cálculo en esos espacios de probabilidad. Esencialmente, la cardinalidad de un conjunto es una medida del número de elementos de un conjunto y permite calcular la probabilidad de un evento  $A$  en un espacio muestral  $\Omega$ . En particular, cuando los resultados de un evento aleatorio son equiprobables se tiene la regla de Laplace. La regla de Laplace establece que la probabilidad del evento  $A$ , es la cardinalidad del evento dividido entre el cardinal del espacio muestral completo:

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}. \quad (4.4)$$

La regla de Laplace también se cumplirá en la composición de experimentos aleatorios independientes con resultados equiprobables. Por ejemplo, el lanzamiento de una moneda equilibrada, seguido del lanzamiento de un dado equilibrado.

**Ejemplo 1:** En una ciudad el 30% de sus ciudadanos leen sobre ciencia, el 60% leen sobre arte, y el 80% uno de los dos. ¿Cuál es la probabilidad de que una persona elegida al azar lea ambos?

Nombremos  $A = \text{Leer sobre ciencia}$   $B = \text{Leer sobre arte}$ . Entonces  $\mathbb{P}(A) = 0.3$  y  $\mathbb{P}(B) = 0.6$ . Leer uno o el otro es  $\mathbb{P}(A \cup B) = 0.8$ . Usamos:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) = 0.3 + 0.6 - 0.8 = 0.1 \quad (4.5)$$

### 4.1.1 Definición frecuentista de probabilidad

En términos prácticos, la probabilidad se mide en términos de la frecuencia de aparición de un evento específico  $A$ . Sea  $N$  el número de veces que se repite una experiencia aleatoria. Denominamos la frecuencia absoluta del evento  $A$ , al número de veces que ocurre  $A$  y se denomina como  $f(A)$ . De la misma forma, definimos la frecuencia relativa como la proporción de eventos que aparece  $A$ , la cual está dada por:

$$f_r(A) = \frac{f(A)}{N} \quad (4.6)$$

Cuando se realiza la experiencia aleatoria un número grande de veces y siguiendo la ley de los grandes números, la frecuencia relativa del evento  $A$  se aproxima asintóticamente a la probabilidad de ocurrencia.

$$\lim_{N \rightarrow +\infty} f_r(A) = \mathbb{P}(A). \quad (4.7)$$

Esto liga las experiencias aleatorias numéricas como: MonteCarlo, Bootstrapping, estimación de parámetros, etc; con las predicciones teóricas que provienen de la probabilidad.

#### Section 4.2

### Ejercicios: Axiomas de la probabilidad

1. Sean  $\mathbb{P}_1$  y  $\mathbb{P}_2$  dos medidas de probabilidad. Definamos  $\mathbb{P} = a_1\mathbb{P}_1 + a_2\mathbb{P}_2$ , donde  $a_1 + a_2 = 1$  y  $a_1, a_2 \in \mathbb{R}^+$ . ¿Es  $\mathbb{P}$  una medida de probabilidad?  
*Hint:* Verifique los axiomas de Kolmogorov para  $\mathbb{P}$ .
2. Sea  $\Omega = \{1, 2\}$ ,  $\mathfrak{F} = \sigma(\Omega)$  y  $\mathbb{P}$  una aplicación definida sobre  $\mathfrak{F}$  dada por:

$$\mathbb{P}(A) = \begin{cases} 0 & \text{si } A = \{\emptyset\} \\ 1/3 & \text{si } A = \{1\} \\ 2/3 & \text{si } A = \{2\} \\ 1 & \text{si } A = \{1, 2\} \end{cases} \quad (4.8)$$

Muestre detalladamente que  $\mathbb{P}$  es una medida de probabilidad. <sup>1</sup>

3. Sea  $(\Omega, \mathfrak{F}, \mathbb{P})$  un espacio de probabilidad. Demuestre las siguientes propiedades básicas de esta medida usando los axiomas de Kolmogorov y diagramas de Venn:
  - a)  $\mathbb{P}(\emptyset) = 0$ .
  - b)  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ .
  - c) Si dos eventos  $A$  y  $B$  son tales que  $A \subset B$ , entonces  $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B - A)$ .
  - d) Dado un evento  $A$ ,  $\mathbb{P}(A) \leq 1$ ,  $\forall A \in \mathfrak{F}$ . Use el teorema anterior.

<sup>1</sup>Tomado de Probabilidad, Liliana Blanco Castañeda UNAL.

- e)  $A \subseteq B$ , entonces  $\mathbb{P}(A) \leq \mathbb{P}(B)$ .
- f)  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ .
- g)  $\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)$ .
- h) Probabilidad de la diferencia: Use  $A - B = A \cap B^c$ , para mostrar:  $\mathbb{P}(A - B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$ .
- i) Probabilidad de la diferencia simétrica: Use  $(A - B) \cup (B - A)$ , para mostrar:  $\mathbb{P}((A \cap B^c) \cup (B \cap A^c)) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B)$ . Esto significa la probabilidad que ocurra  $A$  o  $B$  pero no ambos eventos al tiempo.

### 4.2.1 Independencia de eventos

Uno de los conceptos más importantes en probabilidad es la independencia de eventos. Se establece que dos eventos  $A$  y  $B$  son independientes, si la probabilidad de la intersección está dada por:

$$\mathbb{P}(A \cap B) := \mathbb{P}(A) \cdot \mathbb{P}(B). \quad (4.9)$$

Esto significa que el hecho de que ocurra  $A$  no influye en  $B$  y recíprocamente.

### 4.2.2 Probabilidad condicional

En otras ocasiones el resultado del evento  $B$  puede influir en el resultado del evento  $A$ . En este caso, se define la probabilidad de  $A$  condicionado al evento  $B$  como:

$$\mathbb{P}(A/B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (4.10)$$

Esto puede interpretarse siguiendo la noción de cardinalidad. Dado que ya ocurrió  $B$ , el espacio completo de redujo al conjunto  $B$  y el numerador indica el cardinalidad del conjunto donde ocurre  $A$  y  $B$  a la vez. Note que es posible, dada la probabilidad condicional, calcular la probabilidad de la intersección.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A/B)\mathbb{P}(B). \quad (4.11)$$

Esta formula es válida para eventos no independientes.

### 4.2.3 Teorema de la probabilidad total

Supongamos que tenemos una partición del espacio muestral  $\Omega$  tal que:

$$A_1, A_2, \dots, A_n \subseteq \Omega, \quad (4.12)$$

que cumple las siguientes condiciones:

- a)  $\Omega = A_1 \cup A_2 \cup \dots \cup A_n$
- b)  $A_i \cap A_j = \emptyset, i \neq j$

Si tenemos un suceso  $B$  de una partición de un espacio muestral. El teorema de la probabilidad total establece que la probabilidad  $\mathbb{P}(B)$  es:

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B/A_i) \mathbb{P}(A_i). \quad (4.13)$$

donde  $\mathbb{P}(B/A_i)$  es la probabilidad del suceso  $B$  restringido a cada uno de los elementos de la partición y  $\mathbb{P}(A_i)$  es la probabilidad de que ocurra ese elemento de la partición.

**Ejemplo 1:** En una sala de aeropuerto, 15 hombres y 20 mujeres. De los 15 hombres, 5 llevan gafas y de las 20 mujeres, 8 llevan gafas.

- a) Si se elige una persona al azar, ¿cuál es probabilidad de que tenga gafas?

Identificamos las cantidades  $\mathbb{P}(H) = 15/35$ ,  $\mathbb{P}(M) = 20/35$ . Note que la probabilidad condicional de que lleve gafas dado que es hombre es:  $\mathbb{P}(G/H) = 5/15$  y la probabilidad condicional de que lleve gafas dado que es mujer es:  $\mathbb{P}(G/M) = 8/20$ . La probabilidad total está dada por:

$$\begin{aligned} \mathbb{P}(G) &= \mathbb{P}(G/H)\mathbb{P}(H) + \mathbb{P}(G/M)\mathbb{P}(M) \\ &= (5/15)(15/35) + (8/20)(20/35) = 13/35 \approx 0.37 \end{aligned} \quad (4.14)$$

- b) Si se elige una persona al azar, ¿cuál es probabilidad de que no tenga gafas?

Identificamos las cantidades  $\mathbb{P}(H) = 15/35$ ,  $\mathbb{P}(M) = 20/35$ . Note que la probabilidad condicional de que no lleve gafas dado que es hombre es:  $\mathbb{P}(\bar{G}/H) = 10/15$  y la probabilidad condicional de que no lleve gafas dado que es mujer es:  $\mathbb{P}(\bar{G}/M) = 12/20$ . La probabilidad total está dada por:

$$\begin{aligned}\mathbb{P}(\bar{G}) &= \mathbb{P}(\bar{G}/H)\mathbb{P}(H) + \mathbb{P}(\bar{G}/M)\mathbb{P}(M) \\ &= (10/15)(15/35) + (12/20)(20/35) = 22/35 \approx 0.63\end{aligned}\quad (4.15)$$

**Ejemplo 2:** Se lanzan tres monedas al aire. Calcular la probabilidad de:

1. Obtener 3 caras.

Identificamos las probabilidades de obtener cara o cruz,  $\mathbb{P}(c) = \mathbb{P}(x) = 1/2$  respectivamente. En el caso,  $\mathbb{P}(ccc) = \mathbb{P}(c) \cdot \mathbb{P}(c) \cdot \mathbb{P}(c) = (1/2)(1/2)(1/2) = 1/8$ , dado que el lanzamiento de cada moneda es independiente.

2. Obtener 2 caras y una cruz.

Identificamos las probabilidades de obtener cara o cruz,  $\mathbb{P}(c) = \mathbb{P}(x) = 1/2$  respectivamente, y la probabilidad de obtener dos caras y una cruz con  $\mathbb{P}(A)$ . En este caso, hay tres ramas diferentes del diagrama de árbol que cumplen con la condición:  $\mathbb{P}(ccx)$ ,  $\mathbb{P}(cxc)$  y  $\mathbb{P}(xcc)$ . La probabilidad total será:

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(ccx) + \mathbb{P}(cxc) + \mathbb{P}(xcc) \\ &= \mathbb{P}(c) \cdot \mathbb{P}(c) \cdot \mathbb{P}(x) + \mathbb{P}(c) \cdot \mathbb{P}(x) \cdot \mathbb{P}(c) + \mathbb{P}(x) \cdot \mathbb{P}(c) \cdot \mathbb{P}(c) \\ &= 1/8 + 1/8 + 1/8 = 3/8\end{aligned}\quad (4.16)$$

Este problema también se puede solucionar usando combinaciones sin repetición dado que no importa el orden.

$$\mathbb{P}(A) = \binom{3}{2} \left(\frac{1}{2}\right)^2 \frac{1}{2} = \frac{3!}{(3-2)!2!} \left(\frac{1}{2}\right)^2 \frac{1}{2} = 3/8. \quad (4.17)$$

Formula que resulta más útil en los problemas donde el conteo no es sencillo de calcular usando el diagrama de árbol.

**Ejemplo 3:** Se tienen 2 urnas  $A$  y  $B$ , en la urna  $A$  se tienen 3 bolas negras y 4 blancas y en la urna  $B$  se tienen 5 bolas negras y 3 blancas. Se saca una bola de la urna  $A$ , y sin verla, se introduce en la urna  $B$ . Si se saca una segunda bola de la urna  $B$ , calcule la probabilidad de los siguientes sucesos:

- a) La segunda bola sea negra:  $\mathbb{P}(N_2)$ .
- b) La segunda bola sea blanca:  $\mathbb{P}(B_2)$ .

Calculemos las probabilidades de la primera bola: la probabilidad de que la primera bola sea negra es:  $\mathbb{P}(N_1) = 3/7$  y la probabilidad de que la primera sea blanca es:  $\mathbb{P}(B_1) = 4/7$ . Notar que si la primera bola es negra, en la urna  $B$  quedan 6 bolas negras y 3 blancas; de este modo la probabilidad condicional de obtener una bola negra y una bola blanca es  $\mathbb{P}(N_2/N_1) = 6/9$  y  $\mathbb{P}(B_2/N_1) = 3/9$ . Si por el contrario, la primera bola es blanca, en la urna  $B$  quedan 5 bolas negras y 4 blancas; de este modo la probabilidad condicional de obtener una bola negra y una bola blanca es  $\mathbb{P}(N_2/B_1) = 5/9$  y  $\mathbb{P}(B_2/B_1) = 4/9$ . Lo que conduce a la regla de probabilidad total:

$$\begin{aligned}\mathbb{P}(N_2) &= \mathbb{P}(N_2/N_1)\mathbb{P}(N_1) + \mathbb{P}(N_2/B_1)\mathbb{P}(B_1) \\ &= (6/9)(3/7) + (5/9)(4/7) = 38/63\end{aligned}\tag{4.18}$$

La probabilidad de que la segunda bola sea blanca es:

$$\begin{aligned}\mathbb{P}(B_2) &= \mathbb{P}(B_2/N_1)\mathbb{P}(N_1) + \mathbb{P}(B_2/B_1)\mathbb{P}(B_1) \\ &= (3/9)(3/7) + (4/9)(4/7) = 25/63\end{aligned}\tag{4.19}$$

### Section 4.3

#### Ejercicios: Probabilidad condicional y total

1. En una conferencia hay 1000 participantes distribuidos del siguiente modo: 185 hombres usan gafas, 415 hombres no usan gafas, 115 mujeres que usan gafas. Determine la probabilidad de:

- a) Ser hombre  $\mathbb{P}(H) = 3/5$ .

- b) Ser Mujer  $\mathbb{P}(M) = 2/5$ .
- c) Usar gafas  $\mathbb{P}(G) = 3/10$ .
- d) Que lleve gafas si sabemos que es mujer  $\mathbb{P}(G/M) = \frac{\mathbb{P}(G \cap M)}{\mathbb{P}(M)} = 23/80$ .

	Hombres	Mujeres	Totales
Usa gafas	185	115	300
No usa gafas	415	285	700
Totales	600	400	1000

Table 4.1: Tabla de contingencia de genero y el uso de gafas.

2. Lanzamos un dado de 6 caras. Si sale 1 o 2 extraemos una bola de la urna 1. Si sale 3,4,5,6 extraemos una bola de la urna 2. En la urna 1 tenemos 3 bolas rojas, 1 negra y 6 verdes, y en la urna 2 tenemos 6 bolas rojas, 2 negras y 2 verdes. Calcular la probabilidad de que la bola obtenida:
  - a) Sea roja:  $\mathbb{P}(R) = 1/2$ .
  - b) sea negra:  $\mathbb{P}(N) = 1/6$ .
  - c) Sea de la urna 1 si se ha obtenido una bola negra:  $\mathbb{P}(1/N) = 1/5$ .
  - d) Sea de la urna 2 si se ha obtenido una bola negra:  $\mathbb{P}(2/N) = 4/5$ .
3. Tengo una bolsa con dos dulces de limón y tres de fresa. Si me como un dulce y después otro, ¿cuál es la probabilidad de que ambos hayan sido de fresa.  $\mathbb{P}(F \cap F) = 3/10$

#### Section 4.4

### Teorema de Bayes

El teorema de Bayes responde la pregunta inversa de la probabilidad total. Dado que ya ocurrió el evento  $B$ , este teorema establece la regla para calcular la probabilidad de que ocurra el i-ésimo elemento de la partición.

$$\begin{aligned}
 \mathbb{P}(A_i/B) &= \frac{\mathbb{P}(B/A_i)\mathbb{P}(A_i)}{\mathbb{P}(B)} \\
 &= \frac{\mathbb{P}(B/A_i)\mathbb{P}(A_i)}{\sum_{i=1}^n \mathbb{P}(B/A_i)\mathbb{P}(A_i)}
 \end{aligned} \tag{4.20}$$

**Ejemplo 1:** En el curso de Métodos computacionales, el 20% de la población estudia Matemáticas, 30% estudia Geociencias y el resto estudia Física. El 30% de los estudiantes de matemáticas son mujeres, el 50% en Geociencias y el 45% en Física.

- a) Si un estudiante es elegido al azar y es mujer, ¿qué probabilidad habrá de que sea estudiante de matemáticas?

Identificamos las cantidades:  $\mathbb{P}(Mate) = 0.2$ ,  $\mathbb{P}(Geo) = 0.3$ ,  $\mathbb{P}(Fisi) = 0.5$ . Adicionalmente, tenemos las probabilidades condicionales de ser mujer.  $\mathbb{P}(M/Mate) = 0.3$ ,  $\mathbb{P}(M/Geo) = 0.5$ ,  $\mathbb{P}(M/Fisi) = 0.45$ .

Lo que se pide calcular es:

$$\mathbb{P}(Mate/M) = \frac{\mathbb{P}(M/Mate)\mathbb{P}(Mate)}{\mathbb{P}(M)} \tag{4.21}$$

Usemos el teorema de probabilidad total para calcular que sea mujer:

$$\begin{aligned}
 \mathbb{P}(M) &= \mathbb{P}(M/Mate)\mathbb{P}(Mate) + \mathbb{P}(M/Geo)\mathbb{P}(Geo) + \mathbb{P}(M/Fisi)\mathbb{P}(Fisi) \\
 &= (0.3)(0.2) + (0.5)(0.3) + (0.45)(0.5) = 0.435
 \end{aligned} \tag{4.22}$$

Entonces:

$$\mathbb{P}(Mate/M) = \frac{\mathbb{P}(M/Mate)\mathbb{P}(Mate)}{\mathbb{P}(M)} = \frac{(0.3)(0.2)}{0.435} = 4/29 \approx 0.14 \tag{4.23}$$

- b) Si un estudiante es elegido al azar y es hombre, ¿qué probabilidad habrá de que sea estudiante de Matemáticas?

Identificamos las cantidades:  $\mathbb{P}(Mate) = 0.2$ ,  $\mathbb{P}(Geo) = 0.3$ ,  $\mathbb{P}(Fisi) = 0.5$ . Adicionalmente, tenemos las probabilidades condicionales de ser hombre.  $\mathbb{P}(H/Mate) = 0.7$ ,  $\mathbb{P}(H/Geo) = 0.5$ ,  $\mathbb{P}(H/Fisi) = 0.55$ .

Lo que se pide calcular es:

$$\mathbb{P}(Mate/H) = \frac{\mathbb{P}(H/Mate)\mathbb{P}(Mate)}{\mathbb{P}(H)} \quad (4.24)$$

Usemos el teorema de probabilidad total para calcular que sea hombre:

$$\begin{aligned} \mathbb{P}(H) &= \mathbb{P}(H/Mate)\mathbb{P}(Mate) + \mathbb{P}(H/Geo)\mathbb{P}(Geo) + \mathbb{P}(H/Fisi)\mathbb{P}(Fisi) \\ &= (0.7)(0.2) + (0.5)(0.3) + (0.55)(0.5) = 0.565 \end{aligned} \quad (4.25)$$

Entonces:

$$\mathbb{P}(Mate/H) = \frac{\mathbb{P}(H/Mate)\mathbb{P}(Mate)}{\mathbb{P}(H)} = \frac{(0.7)(0.2)}{0.565} = 28/113 \approx 0.25 \quad (4.26)$$

**Ejemplo 2:** En el ejemplo 3 de la probabilidad total. Calcule la probabilidad de que la bola extraída en la urna  $A$  sea negra, dado que el resultado de la urna  $B$  es blanca  $\mathbb{P}(N_1/B_2)$ . Usando los valores calculados anteriormente:

$$\begin{aligned} \mathbb{P}(N_1/B_2) &= \frac{\mathbb{P}(B_2/N_1)\mathbb{P}(N_1)}{\mathbb{P}(B_2/N_1)\mathbb{P}(N_1) + \mathbb{P}(B_2/B_1)\mathbb{P}(B_1)} \\ &= \frac{(4/9)(4/7)}{(25/63)} = 16/25 \end{aligned} \quad (4.27)$$

#### Section 4.5

### Ejercicios: Teorema de Bayes

- Un médico ha observado que el 40% de sus pacientes fuma. De este 40% el 75% son hombres. Entre los pacientes que no fuman, el 60% son mujeres. Si se selecciona un paciente al azar:

- ¿Cuál es la probabilidad de que sea mujer?  $\mathbb{P}(M) = 23/50$ .

- b) ¿Cuál es la probabilidad de que sea hombre y fume?  $\mathbb{P}(F \cap H) = 3/10$ .
- c) Si el paciente es mujer, ¿cuál es la probabilidad de que fume?  $\mathbb{P}(F/M) = 5/23$ .
2. Suponga que se quiere estimar el número medio de estudiantes que se aproximan a la secretaría de física siguiendo una distribución de Poisson. El problema podría tener información inicial que puede ser utilizada como una función a priori de probabilidad. Esto asigna probabilidades a los posibles valores del parámetro  $\lambda$ . Supongamos que el número medio a priori de estudiantes es: 1,2,3,4 con las probabilidades  $\Pi(\lambda) = [0.4, 0.3, 0.2, 0.1]$  según las secretarías 1,2,3 y 4 (información subjetiva). El número de estudiantes observados durante cierto intervalo de tiempo es  $x = 4$ .
- a) Haga la estimación inicial del parámetro  $\hat{\lambda}$ .
- b) Calcule la función de verosimilitud de la observación para cada modelo  $\mathbb{P}(x/\lambda_i) = \mathcal{L}(x/\lambda_i)$  donde  $i = 1, 2, 3, 4$ .
- c) Aplicando el teorema de Bayes, calcule la función de distribución a posteriori para cada modelo:

$$\mathbb{P}(\lambda_i/x) = \frac{\mathcal{L}(x/\lambda_i)\Pi(\lambda_i)}{\sum_{\forall j} \mathcal{L}(x/\lambda_j)\Pi(\lambda_j)} \quad i, j = 1, 2, 3, 4 \quad (4.28)$$

- d) Verifique que la distribución posterior está normalizada.
- e) Indique el modelo que más probablemente explique los datos.
- f) Usando la información subjetiva de cada secretaria y la verosimilitud de la observación (i.e. la distribución a posteriori). Calcule el mejor parámetro  $\hat{\lambda}$ .

#### Section 4.6

### Técnicas de conteo

Dado que es necesario contar el número de elementos asociados a conjuntos finitos, necesitamos los axiomas fundamentales de conteo y algunas herramientas de combinatoria. Los axiomas fundamentales de conteo son los siguientes:

1. Principio multiplicativo: Si un evento  $A$  puede ocurrir de  $n$  maneras diferentes y otro evento  $B$  puede ocurrir de  $m$  diferentes, entonces ambos eventos pueden ocurrir de  $n \cdot m$  maneras distintas.

2. Principio sumativo: Si un evento  $A$  puede ocurrir de  $n$  maneras diferentes o  $m$  maneras diferentes, siendo incompatibles las unas con las otras, entonces existen  $n + m$  maneras de ocurrir  $A$ .

Adicionalmente, las posibles configuraciones depende si importa o no el orden en los sub-conjuntos. Si importa el orden se calculan las variaciones y permutaciones, y si no importa el orden, calculamos la combinaciones. Además, podrían o no repetirse elementos en cada sub-conjunto.

Variaciones sin repetición:

$$V_r^n = \frac{n!}{(n-r)!} \quad (4.29)$$

Variaciones con repeticion:

$$V_r^n = n^r \quad (4.30)$$

Permutaciones sin repetición:

$$P^n = n! \quad (4.31)$$

Permutaciones con repetición:

$$P_{n_1, n_2, \dots, n_r}^n = \frac{n!}{n_1! n_2! \dots n_r!} \quad (4.32)$$

Por otro lado, cuando no importa el orden de las configuraciones:

Combinaciones sin repetición:

$$C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!} \quad (4.33)$$

Combinaciones con repetición:

$$C_r^n = \binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!} \quad (4.34)$$

**Section 4.7****Ejercicios: Técnicas de conteo**

Realizar los siguientes cálculos estableciendo si es variación, permutación o combinación; con o sin repetición.

1. Carlos, Manuel, Sandra correrán los 100 metros planos. ¿De cuántas formas puede quedar el podio de primer y segundo lugar? Solo competirán ellos tres. R:6
2. ¿De cuántas formas se puede preparar una ensalada de frutas con solo 2 ingredientes, si se cuenta con plátano, manzana y uva? R:3
3. ¿De cuantas formas pueden hacer cola 5 amigos para entrar al cine? R:120
4. ¿De cuántas formas puede un juez otorgar el primero, segundo y tercer premio en un concurso que tiene 8 participantes R:336
5. El capitán de un barco solicita 2 marineros para realizar un trabajo, sin embargo, se presentan 10. ¿De cuántas formas podrá seleccionar a los 2 marineros? R:45
6. Eduardo tiene 7 Libros, ¿De cuántas maneras podrá acomodar cinco de ellos de un estante? R:2520
7. En un salón de 10 alumnos, ¿de cuántas maneras se puede formar un comité formado por 2 de ellos? R:45
8. ¿Cuántas palabras diferentes se puede formar con las letras de la palabra REMEMBER? R:1680
9. Un club de basketball tiene 12 jugadoras, una de ellas es la capitana María. ¿Cuántos equipos diferentes de 6 jugadoras se pueden formar, sabiendo que en todos ellos siempre debe estar la capitana María. R:462
10. Con 4 frutas diferentes, ¿cuántos jugos surtidos se pueden preparar?. Un jugo surtido se debe preparar con al menos 2 frutas. R:11
11. En un curso de 10 estudiantes se desea escoger presidente, vicepresidente y secretario. ¿De cuántas formas de pueden seleccionar los 3 estudiantes? R:720

12. En un campeonato compiten 8 equipos ¿de cuántas maneras diferentes se podrían ganar los premios de campeón y sub-campeón? R:56
13. ¿Cuántos números de 3 cifras distintas se pueden formar con los dígitos del 1 al 7? R:210
14. ¿Cuántos números de 3 cifras se pueden formar con los dígitos del 1 al 7? R:343
15. De un grupo de 10 estudiantes se quiere seleccionar un comité al azar de 3 estudiantes. ¿De cuántas maneras diferentes se puede seleccionar el comité? R:120
16. ¿Cuántas placas diferentes se pueden hacer con 3 letras y 3 dígitos? R:17576000
17. n Personas van a jugar cartas alrededor de una mesa, ¿de cuántas maneras diferentes se pueden sentar? R: $(n - 1)!$
18. En una heladería ofrecen 7 diferentes sabores, ¿cuántas combinaciones de helado de 3 sabores se pueden hacer? R:84
19. En un almacén venden 6 diferentes sabores de gaseosas, ¿de cuántas formas se pueden seleccionar 3 gaseosas?. ¿De cuántas formas diferentes se pueden seleccionar 3 gaseosas? R:56, R:20
20. Demostrar la formula de combinaciones con repetición.

$$C_r^n = \binom{n + r - 1}{r} \quad (4.35)$$

21. Calcular la probabilidad de ganar el Baloto comprando un solo billete. R: $7.1511 \times 10^{-8}$
22. Cuántas sumas de 3 enteros no negativos dan 10. R:66
23. Se tienen 9 llaves: 3 rojas, 3 azules y 3 verdes. Si elegimos 4, ¿de cuántas formas se pueden distribuir los colores? R:12

#### Section 4.8

### Ejercicios: Generales de probabilidad

Algunos tomados del libro de probabilidad y estadística de Walpole [4].

1. Se lanza un dado equi-probable dos veces. Dados los siguientes eventos:

- A) La suma de los resultados es menor o igual a 3.
- B) El resultado del primer lanzamiento es impar.

Calcular  $\mathbb{P}(A) = 1/12$ ,  $\mathbb{P}(B) = 1/2$ ,  $\mathbb{P}(A \cup B) = 19/36$  y  $\mathbb{P}(A^c) = 11/12$

2. Se prueban 5 celulares de un lote de 50 equipos donde existen 2 defectuosos. Si el muestreo se realiza sin reposición. Hallar la probabilidad que al menos un celular sea defectuoso.  $\mathbb{P}(A) = 47/245$ . *Piense en  $A^c$  y en el principio multiplicativo o en una estrategia combinatoria.*
3. En una cierta ciudad el 60% de los propietarios están suscritos al diario y el 80% al cable. Adicionalmente, el 50% están suscritos a ambos. Si un propietario es elegido al azar:
  - (a) ¿Cuál es la probabilidad que esté suscrito a uno de los dos servicios?  $\mathbb{P}(A \cup B) = 0.9$ .
  - (b) ¿Cuál es la probabilidad que esté suscrito al diario o al cable, pero no a ambos servicios?  $\mathbb{P}((A \cap B^c) \cup (B \cap A^c)) = 0.4$ .

4. Calcular la probabilidad que  $n$  personas ( $n \leq 365$ ) tengan fechas diferentes de cumpleaños, i.e, escribir la formula general de cálculo. Grafique la probabilidad  $\mathbb{P}(n \leq 80)$  como función de  $n$ . *Los números son demasiado grandes, pero Python puede manejar dichas cantidades.*

[https://www.youtube.com/watch?v=7uzx6D\\_0V7M&ab\\_channel=Derivando](https://www.youtube.com/watch?v=7uzx6D_0V7M&ab_channel=Derivando)

5. Se lanzan dos dados equi-probables y se observan los siguientes eventos:
  - a) La suma es 8:  $\mathbb{P}(A) = 5/36$ .
  - b) El segundo dado es impar:  $\mathbb{P}(B) = 1/2$ .

Calcule  $\mathbb{P}(A \cap B)$  y  $\mathbb{P}(A) \cdot \mathbb{P}(B)$ . ¿Qué podría concluir?
6. Se lanza simultáneamente 3 dados de 6 caras. ¿Cuál es la probabilidad de obtener 1 par?  $\mathbb{P}(A) = 5/12$ . Realice el cálculo de esta probabilidad usando un experimento virtual con  $N = 10^5$  eventos.
7. Se lanza simultáneamente 5 dados de 6 caras. ¿Cuál es la probabilidad de obtener?
  - a) 1 par:  $\mathbb{P}(A) = 25/54$ .

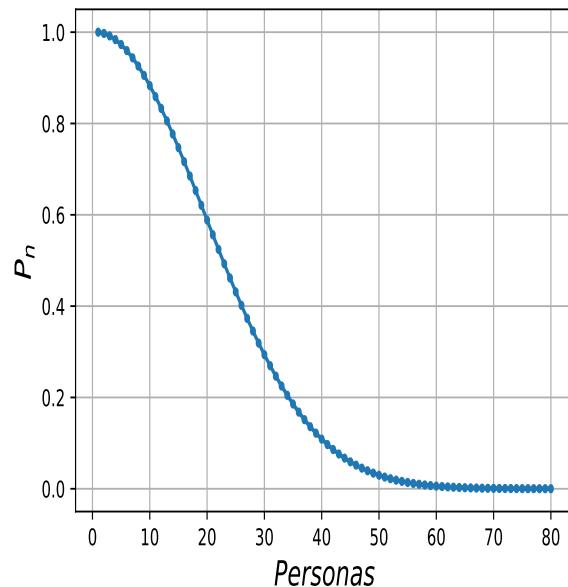


Figure 4.1: Probabilidad  $\mathbb{P}(n \leq 80)$  de que tengan una edad diferente.

- b) 2 pares distintos:  $\mathbb{P}(B) = 25/108$ .
- c) 4 de la misma cara:  $\mathbb{P}(C) = 25/1296$ .

Realice el cálculo de la probabilidad de los numerales a) y b) usando un experimento virtual con  $N = 10^5$  eventos.

8. Se lanzan simultáneamente 4 monedas. Determine la probabilidad de obtener dos caras y dos sellos  $\mathbb{P}(A) = 3/8$ . Realice el cálculo de esta probabilidad usando un experimento virtual con  $N = 10^5$  eventos, y etiquetando los resultados con +1 y -1 para cara y sello respectivamente.
9. En el ejercicio anterior, imagine que las monedas están truncadas de tal manera que la probabilidad de que la moneda 1 sea cara es  $p_1$  y que sea sello es  $1-p_1$ . Usando el árbol de probabilidad, ¿cuál es la expresión de la probabilidad de obtener dos caras y dos sellos de este evento? Si el truncamiento de las monedas 1 y 2 puede variar como:  $0.1 < p_1 < 0.9$  y  $0.1 < p_2 < 0.5$ , use el árbol de probabilidad para graficar la superficie de probabilidad del evento A. ¿En qué punto la probabilidad es mínima y máxima, y cuáles son sus valores?
10. En un juego de Poker que consta de 5 cartas, encuentre la probabilidad de tener:

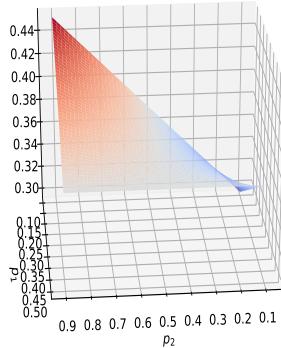


Figure 4.2: Superficie de probabilidad del evento  $A$ .

- a) 3 ases:  $\mathbb{P}(A) = 94/54145$
  - b) 4 cartas de corazones y 1 de bastos:  $\mathbb{P}(B) = 143/39984$
11. Si se seleccionan al azar 3 libros de un estante que contiene 5 novelas, 3 libros de poemas y un diccionario, ¿cuál es la probabilidad de que:
- a) se tome el diccionario:  $\mathbb{P}(A) = 1/3$
  - b) se escogen 2 novelas y un libro de poemas:  $\mathbb{P}(B) = 5/14$

#### Section 4.9

### Distribuciones discretas de probabilidad

Necesitamos una expresión que nos permita representar todas las probabilidades de una variable aleatoria  $X$ , esta función se representa por  $f(x)$  y tiene la siguiente connotación:

$$f(x) = \mathbb{P}(X = x) \quad (4.36)$$

el conjunto de pares ordenados  $(x, f(x))$  se denomina función de densidad de probabilidad.

**Definición 1.** *El conjunto de pares ordenados  $(x, f(x))$  se denomina función de densidad de probabilidad de la variable aleatoria  $X$  si, para cada resultado posible  $x$ :*

$$\begin{aligned}
 f(x) &\geq 0 \\
 \sum_x f(x) &= 1 \\
 \mathbb{P}(X = x) &= f(x)
 \end{aligned} \tag{4.37}$$

$x$	0	1	3
$\mathbb{P}(X = x)$	1/3	1/2	1/6

Table 4.2: Distribución de probabilidad

**Definición 2.** Sea  $X$  una variable aleatoria con distribución de probabilidad  $f(x)$ . La media o valor esperado de  $X$  es:

$$\hat{\mu} = \mathbb{E}(X) = \sum_x x f(x) \tag{4.38}$$

Del ejemplo anterior el valor esperado será:

$$\hat{\mu} = \sum_{i=0}^2 x_i f(x_i) = x_0 f(x_0) + x_1 f(x_1) + x_2 f(x_2) = 0(1/3) + 1(1/2) + 3(1/6) = 1 \tag{4.39}$$

Se tienen algunos resultados útiles para la esperanza de una variable aleatoria.

$$\begin{aligned}
 \mathbb{E}(c) &= c \\
 \mathbb{E}(cX) &= c\mathbb{E}(X) \\
 \mathbb{E}(c + X) &= c + \mathbb{E}(X) \\
 \mathbb{E}(X + Y) &= \mathbb{E}(X) + \mathbb{E}(Y)
 \end{aligned} \tag{4.40}$$

donde  $c$  es constante.

**Definición 3.** Sea  $X$  una variable aleatoria con distribución de probabilidad  $f(x)$ . La varianza de  $X$  es:

$$\hat{\sigma}^2 = Var(X) = \sum_x (x - \mathbb{E}(x))^2 f(x). \tag{4.41}$$

Esto es una medida de la dispersión en los datos. Se tienen algunos resultados útiles para la varianza de una variable aleatoria.

$$\begin{aligned}
 Var(c) &= 0 \\
 Var(cX) &= c^2Var(X) \\
 Var(X + c) &= Var(X) \\
 Var(X + Y) &= Var(X) + Var(Y), \quad X, Y \text{ independientes}
 \end{aligned} \tag{4.42}$$

donde  $c$  es constante. En el caso donde las variables  $X$  y  $Y$  no sean independientes, tenemos la definición de covarianza.

### 4.9.1 Covarianza

**Definición 4.** Sea  $X$  e  $Y$  dos variables aleatoria con distribución de probabilidad conjunta  $f(x, y)$ . La covarianza de  $X$  y  $Y$  está dada por:

$$\hat{\sigma}_{xy} = Cov(X, Y) = \sum_{x,y} (x - \mathbb{E}(x))(y - \mathbb{E}(y))f(x, y). \tag{4.43}$$

**Proof:** Definamos  $\mu_X = \mathbb{E}(X)$  y  $\mu_Y = \mathbb{E}(Y)$  y usando la propiedad de la esperanza  $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = \mu_X + \mu_Y$ . Entonces:

$$\begin{aligned}
 Var(X + Y) &= \mathbb{E}((X + Y - (\mu_X + \mu_Y))^2) \\
 &= \mathbb{E}(((X - \mu_X) + (Y - \mu_Y))^2) \\
 &= \mathbb{E}((X - \mu_X)^2) + \mathbb{E}((Y - \mu_Y)^2) + 2\mathbb{E}((X - \mu_X)(Y - \mu_Y))
 \end{aligned} \tag{4.44}$$

En la última línea se uso la linealidad de la esperanza, donde el último término representa la covarianza entre variables. Por tanto:

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y). \tag{4.45}$$

La covarianza satisface algunas propiedades:

$$\begin{aligned}
 Cov(aX, bY) &= abCov(X, Y) \\
 Cov(X, X) &= Var(X) \\
 Cov(X, Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)
 \end{aligned} \tag{4.46}$$

### 4.9.2 Correlación

La covarianza nos indica el nivel de asociación lineal que existe entre las variables, sin embargo, los valores dependen de las unidades en que se midan  $X$  y  $Y$ . En ese sentido, podemos definir un estimador independiente de unidades denominado coeficiente de correlación. El coeficiente de correlación de Pearson está dado por:

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}. \tag{4.47}$$

### 4.9.3 Distribución de Bernoulli

Sea  $X$  una variable aleatoria discreta  $X \sim Bernoulli(p)$ , descrita por un único experimento con dos posibles resultados denominados éxito y fracaso, y con parámetro  $0 < p < 1$ , donde  $p$  es la probabilidad de éxito. Esta función de probabilidad está dada por:

$$P(X = x) = p^x(1 - p)^{1-x}, \quad x = 0, 1 \tag{4.48}$$

Se puede demostrar que el valor medio está dado por  $\mathbb{E}(X) = p$ .

### 4.9.4 Distribución Binomial

Sea  $X$  una variable aleatoria discreta  $X \sim Bin(n, p)$  con parámetros  $n \in \mathbb{N}$  y  $0 < p < 1$ . Esta función de probabilidad está dada por:

$$P(X = x) = \binom{n}{x} p^x(1 - p)^{1-x}, \quad x = 0, 1, \dots, n \tag{4.49}$$

Esta distribución está conectada con la distribución de Bernoulli. Si  $X_1, X_2, \dots, X_n$  son  $n$  variables aleatorias con distribución de Bernoulli  $X_i \sim Bernoulli(p)$ , entonces:

$$\sum_i^n X_i \sim Bin(n, p). \tag{4.50}$$

Se puede demostrar que el valor medio está dado por  $\mathbb{E}(X) = np$ .

### 4.9.5 Distribución de Poisson

Sea  $X$  una variable aleatoria discreta  $X \sim Pois(\lambda)$  con parámetro  $\lambda \in \mathbb{R}^+$ . Esta función de probabilidad está dada por:

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, \dots \quad (4.51)$$

donde  $k$  es el número de ocurrencias del suceso.

Se puede demostrar que el valor medio está dado por  $\mathbb{E}(X) = \lambda$ .

#### Section 4.10

### Ejercicios: Distribuciones discretas de probabilidad

1. ¿Cuál es valor esperado del experimento de lanzar un dado? **Ans:**  $\hat{\mu} = 3.5$
2. ¿Cuál es valor esperado del experimento de lanzar dos dados y sumar sus resultados?  
**Ans:**  $\hat{\mu} = 7.0$
3. Un embarque de 10 microchips similares que se envía a distribución tiene 3 aparatos defectuosos. Si una empresa realiza un compra aleatoria de 2 de estos microchips.
  - a) Muestre que la distribución de probabilidad del número de microchips defectuosos.

$$f(x) = \frac{\binom{7}{2-x} \binom{3}{x}}{\binom{10}{2}} \quad (4.52)$$

$x$	0	1	2
$\mathbb{P}(X = x)$	7/15	7/15	1/15

Table 4.3: Distribución de probabilidad del número de microchips defectuosos.

- b) ¿Cuál es valor esperado de microchips defectuosos? **Ans:**  $\hat{\mu} = 3/5 = 0.6$
4. Una caja cuántica tiene 3 electrones, 2 protones y 3 neutrones. Se selecciona una muestra aleatoria de 4 partículas. Si  $x$  es el número de electrones e  $y$  es el número de protones.

a) Muestre que la distribución de probabilidad conjunta  $f(x, y)$  es:

$$f(x, y) = \frac{\binom{3}{x} \binom{2}{y} \binom{3}{4-x-y}}{\binom{8}{4}} \quad (4.53)$$

- b) Hallar las distribuciones marginales  $g(x)$  y  $h(y)$ .
- c) Halle el valor esperado de electrones:  $\mathbb{E}(x) = 105/70$ .
- d) Halle el valor esperado de protones:  $\mathbb{E}(y) = 1$ .
- e) Calcular la covarianza usando:  $\sigma_{xy} = \mathbb{E}(xy) - \mathbb{E}(x)\mathbb{E}(y) = -3/14$
- f) Calcular la covarianza usando:  $\sigma_{xy} = \mathbb{E}((x - \hat{\mu}_x)(y - \hat{\mu}_y)) = -3/14$
- g) Son las variables  $x$  e  $y$  independientes?

#### Section 4.11

### Distribuciones continuas de probabilidad

**Definición 5.** La función  $f(x)$  es una función de densidad de probabilidad para la variable aleatoria continua  $X$ , definida en el conjunto de los números reales, si:

$$\begin{aligned} f(x) &\geq 0 \text{ para toda } x \in \mathbb{R} \\ \int_{-\infty}^{\infty} f(x)dx &= 1 \\ P(a < X < b) &= \int_a^b f(x)dx \end{aligned} \quad (4.54)$$

**Definición 6.** Sea  $X$  una variable aleatoria con distribución de probabilidad  $f(x)$ . La media o valor esperado de  $X$  es:

$$\hat{\mu} = \mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx \quad (4.55)$$

#### 4.11.1 Distribución normal

La función de densidad de probabilidad normal está dada por:

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad (4.56)$$

donde  $\mu \in \mathbb{R}$  y  $\sigma^2 \in \mathbb{R}^+$  son los parámetros que caracterizan la distribución. Se puede demostrar que el valor medio está dado por  $\mathbb{E}(X) = \mu$ .

### 4.11.2 Distribución $\chi^2$

Si tenemos  $Z_1, Z_2, \dots, Z_k$  variables aleatorias con  $Z_i \sim N(0, 1)$  para  $i = 1, 2, \dots, k$ , entonces la variable aleatoria:

$$\sum_{i=1}^k Z_i^2 \sim \chi^2(k), \quad (4.57)$$

donde  $k > 0$  es el número de grados de libertad de la distribución. La función de densidad de probabilidad está dada por:

$$\chi^2(x, k) = \frac{\left(\frac{1}{2}\right)^k}{\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-x/2}. \quad (4.58)$$

Con  $x > 0$  y  $\Gamma$  es la función gamma.

### 4.11.3 Distribución t-Student

Si tenemos una variable aleatoria  $X_1 \sim N(0, 1)$  y otra variable aleatoria  $X_2 \sim \chi^2(n)$ . Se define la variable aleatoria:

$$X_n = \frac{X_1}{\sqrt{\frac{X_2}{n}}} \quad (4.59)$$

tiene una función de densidad t-Student dada por:

$$f(x, \nu) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\Gamma(\nu/2)} (1 + x^2/\nu)^{-(\nu+1)/2}, \quad (4.60)$$

donde  $\nu \in \mathbb{R}^+$  es el número de grados de libertad y  $\Gamma$  es la función gamma.

#### Section 4.12

### Ejercicios: Distribuciones continuas de probabilidad

1. Dada la función de probabilidad conjunta:

$$f(x, y) = \begin{cases} \frac{2}{3}(x + 2y) & 0 \leq x \leq 1, 0 \leq y \leq 1. \\ 0 & \text{otro caso} \end{cases} \quad (4.61)$$

encuentre analíticamente y a través del paquete SymPy los siguientes valores:

- a) Verifique que sea una función de densidad conjunta válida.
  - b) Hallar las distribuciones marginales  $g(x)$  y  $h(y)$ .
  - c)  $\mathbb{E}(x) = \frac{10}{18}$
  - d)  $\mathbb{E}(y) = \frac{11}{18}$
  - e) Calcular la covarianza usando:  $\sigma_{xy} = \mathbb{E}(xy) - \mathbb{E}(x)\mathbb{E}(y) = -0.00617$
  - f) Calcular la covarianza usando:  $\sigma_{xy} = \mathbb{E}((x - \hat{\mu}_x)(y - \hat{\mu}_y)) = -0.00617$
  - g) Son las variables  $x$  e  $y$  independientes?
2. Suponga que el error en la temperatura en un experimento controlado de laboratorio es una variable aleatoria continua  $X$ , que tiene la siguiente función de densidad de probabilidad:
- $$f(x) = \begin{cases} \frac{x^2}{3} & \text{si } -1 \leq x \leq 2 \\ 0 & \text{en otro caso} \end{cases} \quad (4.62)$$
- Encuentre las siguientes probabilidades:
- (a)  $\mathbb{P}(0 < X \leq 1) = 1/9$
  - (b)  $\mathbb{P}(1 < X \leq 2) = 7/9$
3. Una variable aleatoria continua  $X$  que puede asumir valores entre  $x = 2$  y  $x = 3$  tiene una función de densidad  $f(x) = 1/2$ .

- a) Demuestre que el área bajo la curva es igual a 1.
- b) Encuentre  $\mathbb{P}(2 < X < 2.5) = 1/4$
- c) Encuentre  $\mathbb{P}(X \leq 2.6) = 0.3$

4. Una variable aleatoria continua  $X$  tiene una función densidad:

$$f(x) = \begin{cases} e^{-x} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases} \quad (4.63)$$

Encuentre el valor esperado de  $g(X) = e^{2X/3} = 3$ . Recuerde que el valor esperado de la variable aleatoria  $g(X)$  está dada por:

$$\hat{\mu}_{g(X)} = \int_{-\infty}^{\infty} g(x)f(x)dx \quad (4.64)$$

### Section 4.13

## Procesos de Markov

**Definición 7.** *Un proceso de Markov es un proceso estocástico discreto  $\{x_n : n = 0, 1, \dots\}$ , con espacio de estados discreto  $S \subseteq \{0, 1, \dots\}$  que satisfacen la propiedad de Markov.*

$$\mathbb{P}(X_{n+1} = x_{n+1} / X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} = x_{n+1} / X_n = x_n) \quad (4.65)$$

Esto significa que la probabilidad de que la variable aleatoria tome el valor  $x_{n+1}$ , dado que el proceso ha pasado por los pasos anteriores  $x_0, x_1, \dots, x_n$ ; solo depende del estado anterior  $x_n$ . Equivalentemente, se puede escribir la probabilidad de que la variable aleatoria tome el valor  $x_{n+1}$  como:

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_{n+1} = x_{n+1}) = \mathbb{P}(X_0 = x_0)\mathbb{P}(X_1 = x_1 / X_0 = x_0)\mathbb{P}(X_{n+1} = x_{n+1} / X_n = x_n)$$

Note que el proceso completo, requiere un conjunto de probabilidades para el estado inicial.

**Definición 8.** *La distribución de la variable aleatoria  $X_0$  se denomina distribución inicial, es decir,  $\{\mathbb{P}(X_0 = 0), \mathbb{P}(X_0 = 1), \dots\}$ . Se suponen que la distribución inicial es conocida.*

Por otro lado, el conjunto de probabilidades condicionales que describen al proceso se denominan probabilidades de transición.

**Definición 9.** La probabilidad  $\mathbb{P}(X_{n+1} = i/X_n = j) = \mathbb{P}_{ij}(n, n+1)$  se denomina probabilidad de transición del estado  $j$  en el tiempo  $n$  al estado  $i$  en el tiempo  $n+1$ .

Se supone que los valores de la probabilidades de transición son independientes del valor específico de  $n$ , en ese caso, la cadena es estacionaria.

$$\mathbb{P}_{ij}(0, 1) = \mathbb{P}_{ij}(1, 2) = \dots = \mathbb{P}_{ij}(n, n+1). \quad (4.66)$$

**Ejemplo 1:**

Imaginemos que una partícula puede estar en cuatro diferentes estados 1,2,3 y 4. Esta partícula puede mantener o cambiar su estado con ciertas probabilidades. Medimos el estado de la partícula en intervalos de tiempo iguales  $t_0, t_1, \dots, t_n$ . Se tienen las siguientes reglas de transición:

$$\begin{aligned}\mathbb{P}(1/1) &= 0.30, \mathbb{P}(1/2) = 0.25, \mathbb{P}(1/3) = 0.15, \mathbb{P}(1/4) = 0.10 \\ \mathbb{P}(2/1) &= 0.30, \mathbb{P}(2/2) = 0.25, \mathbb{P}(2/3) = 0.55, \mathbb{P}(2/4) = 0.20 \\ \mathbb{P}(3/1) &= 0.30, \mathbb{P}(3/2) = 0.50, \mathbb{P}(3/3) = 0.10, \mathbb{P}(3/4) = 0.30 \\ \mathbb{P}(4/1) &= 0.10, \mathbb{P}(4/2) = 0.00, \mathbb{P}(4/3) = 0.20, \mathbb{P}(4/4) = 0.40\end{aligned} \quad (4.67)$$

Por ejemplo:  $\mathbb{P}(4/1)$  es la probabilidad de que pase al estado 4, dado que está en el estado 1.

- a) Si en  $t_0$  el estado de la partícula es 1, ¿cuál es la probabilidad de que esté en el estado 4 en  $t_2$ ?

Para pasar del estado 1 al estado 4 en dos pasos, la regla de probabilidad total define cuatro caminos independientes:  $E_1 \rightarrow E_1 \rightarrow E_4$ ,  $E_1 \rightarrow E_2 \rightarrow E_4$ ,  $E_1 \rightarrow E_3 \rightarrow E_4$ , y  $E_1 \rightarrow E_4 \rightarrow E_4$ . Entonces, la probabilidad es:

$$\begin{aligned}\mathbb{P}(1 \rightarrow 4) &= \mathbb{P}(1/1)\mathbb{P}(4/1) + \mathbb{P}(2/1)\mathbb{P}(4/2) + \mathbb{P}(3/1)\mathbb{P}(4/3) + \mathbb{P}(4/1)\mathbb{P}(4/4) \\ &= (0.3)(0.1) + (0.3)(0.0) + (0.3)(0.2) + (0.1)(0.4) = 0.13\end{aligned} \quad (4.68)$$

- b) Si en  $t_0$  el estado de la partícula es 1, ¿En cuál estado la partícula está con mayor probabilidad?

De las reglas de transición, calculamos las probabilidades restantes:

$$\begin{aligned}\mathbb{P}(1 \rightarrow 1) &= 0.22 \\ \mathbb{P}(1 \rightarrow 2) &= 0.35 \\ \mathbb{P}(1 \rightarrow 3) &= 0.30\end{aligned}\tag{4.69}$$

Lo que significa que la transición más probable es  $\mathbb{P}(1 \rightarrow 2)$ .

Note que en el ejemplo anterior hay una estructura intrínseca matricial. Por ejemplo, si pensamos en  $\mathbb{P}(1 \rightarrow 4)$  como la multiplicación de elementos de una matriz tenemos:

$$\mathbb{P}(1 \rightarrow 4) = p_{41}p_{11} + p_{42}p_{21} + p_{43}p_{31} + p_{44}p_{41} = \sum_{k=4} p_{4k}p_{k1} = \mathbb{P}_{41}.\tag{4.70}$$

Que corresponde con la multiplicación del cuarto renglón por la primera columna de la matriz de probabilidades de transición.

**Definición 10.** Para un proceso de Markov, la matriz de probabilidad de transición (estacionaria) de un paso es:

$$\mathbb{P} = \begin{pmatrix} P_{00} & P_{01} & \cdots & P_{0n} \\ P_{10} & P_{11} & \cdots & P_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n0} & P_{n1} & \cdots & P_{nn} \end{pmatrix}\tag{4.71}$$

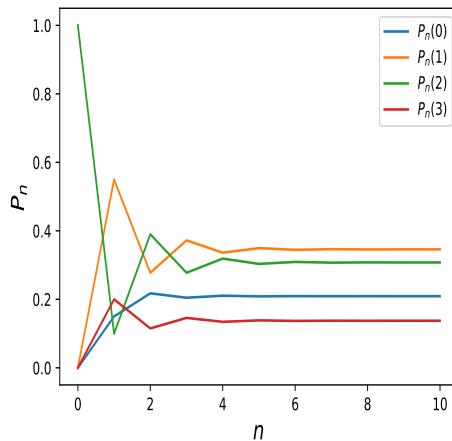
donde la información de la probabilidad de transición  $P_{ij}$  debe entenderse como la probabilidad de ir de la columna  $j$  al renglón  $i$ .

Esta matriz debe cumplir la propiedad básica de que la suma de probabilidades de cada columna debe ser igual a uno:

$$\sum_i P_{ij} = 1, \quad \forall j \leq n\tag{4.72}$$

Del ejemplo 1 de una cadena de dos pasos, la probabilidad de transición del estado 1 a todos los estados, puede calcularse como el producto matricial  $\mathbb{P}^2 \vec{E}$ :

$$\mathbb{P}(1 \rightarrow 1, 2, 3, 4) = \mathbb{P}^2 \vec{E} = \begin{pmatrix} 0.3 & 0.25 & 0.15 & 0.10 \\ 0.3 & 0.25 & 0.55 & 0.20 \\ 0.3 & 0.50 & 0.10 & 0.30 \\ 0.1 & 0.00 & 0.20 & 0.40 \end{pmatrix}^2 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.22 \\ 0.35 \\ 0.30 \\ 0.13 \end{pmatrix} \quad (4.73)$$

Figure 4.3: Probabilidades de transición de los estados para  $n = 10$  pasos.**Section 4.14****Ejercicios: Cadenas de Markov**

1. Calcule las probabilidades de transición estacionarias del ejemplo 1.

**Section 4.15****Metropolis-Hastings algorithm**

Suppose you wish to sample the random variable  $x \sim f(x)$ , but you can not use the direct simulation, the inverse CDF or accept-reject methods. But you can evaluate  $f(x)$  at least up to a proportionality constant; then you can use the Metropolis-Hastings algorithm.

Let  $f(x)$  be the (possible unnormalized) target density,  $x^j$  be a current value and  $q(x/x^j)$  be a proposal distribution, which might depend on the current value  $x^j$ . The algorithm is

summarized as follows:

- a) Sample  $x^* \sim q(x/x^j)$ .
- b) Calculate the acceptance probability:

$$\rho(x^j, x^*) = \min\left[1, \frac{f(x^*)q(x^j/x^*)}{f(x^j)q(x^*/x^j)}\right], \quad (4.74)$$

if  $\rho(x^j, x^*) > r$ ,  $x^*$  is accepted. Where  $r \sim \mathcal{U}(0, 1)$ .

- c) Set  $x^{j+1} = x^*$  with probability  $\rho(x^j, x^*)$ , otherwise set  $x^{j+1} = x^j$ .

In general, this sequence is not independent, moreover, we concentrate on symmetry chains (random walks) where  $q(x/x^j) = q(x^j/x)$ . The acceptance probability is, therefore:

$$\rho(x^j, x^*) = \min\left[1, \frac{f(x^*)}{f(x^j)}\right] \quad (4.75)$$

### How this algorithm works?

Let  $D_n(x)$  be the points density  $x_n$  around  $x$  and  $D_n(y)$  the points density  $y_n$  around  $y$ .  $x$  will be accepted or rejected. We can quantify the flux of points on the neighborhood  $x$  as:

$$\delta D_n(x) = \delta D_{n+1}(x) - \delta D_n(x) \quad (4.76)$$

$$\delta D_n(x) = \underbrace{\sum_y D_n(y) P(y \rightarrow x)}_{Gain} - \underbrace{\sum_y D_n(x) P(x \rightarrow y)}_{Lost} \quad (4.77)$$

$$\sum_y D_n(y) P(x \rightarrow y) \left[ \frac{P(y \rightarrow x)}{P(x \rightarrow y)} - \frac{D_n(x)}{D_n(y)} \right] \quad (4.78)$$

We have two properties:

1. The equilibrium is achieved asymptotically.
2. There is flux toward  $y$ , which tends  $D_n(x)$  to the equilibrium.

$$\frac{P(y \rightarrow x)}{P(x \rightarrow y)} < \frac{D_n(x)}{D_n(y)} \quad (4.79)$$

As a reminder  $\rho(x, y) > r$  is the acceptance condition. We can calculate  $P(x \rightarrow y)$  as  $p_{xy}A_{xy}$ , where  $p_{xy}$  is the probability of generation of  $y$  given  $x$ , and  $A_{xy}$  is the acceptance probability. Since the Metropolis strategy is symmetric ( $p_{xy} = p_{yx}$ ), we get:

$$\frac{P(y \rightarrow x)}{P(x \rightarrow y)} = \frac{A_{yx}}{A_{xy}} = \rho(x, y) = \frac{f(x)}{f(y)} \quad (4.80)$$

Therefore, the equilibrium distribution is given by:

$$D_{eq} = D(x) = f(x) \quad (4.81)$$

#### Section 4.16

### Ejercicios: Metropolis-Hastings

1. Se lanza una moneda  $n = 10$  veces y se encuentra que  $r = 7$  veces cae cara. Usando el algoritmo de Metrópolis:
  - a) Encuentre el parámetro asociado a la probabilidad de éxito  $\hat{p}$  de la distribución binomial, es decir, encuentre el máximo de la distribución posterior.
  - b) Usando la varianza binomial y los cuantiles de la distribución posterior, encuentre los errores asociados al parámetro  $\hat{p}$  a un nivel de confianza del 68% (i.e.,  $\sigma_-^+$ ).
  - c) ¿Podemos decir que la moneda está truncada?

Utilice la siguiente distribución a priori para el espacio de parámetros:

$$\pi(p) = \begin{cases} 1 & \text{si } 0 < p < 1 \\ 0 & \text{si } \text{Otro caso} \end{cases} \quad (4.82)$$

2. Resuelva el problema anterior usando el paquete Emcee.

<https://emcee.readthedocs.io/en/stable/>

3. Usando el algoritmo de Metrópolis, realice el muestreo de  $N = 1000$  eventos de una distribución normal:  $\mathcal{A} \sim \mathcal{N}(x; \mu = 2, \sigma = 0.5)$ .
4. Usando el algoritmo de Metrópolis, realice el muestreo de  $N = 1000$  eventos de una distribución estándar de Cauchy:  $\mathcal{A} \sim f(x; 0, 1)$ .

$$f(x; 0, 1) = \frac{1}{\pi(1 + x^2)} \quad (4.83)$$

5. Genere un ensamble de al menos 30 caminatas aleatorias como las generadas en clase. Calcule la distancia cuadrática desde el inicio del movimiento hasta el final, es decir  $R^2 = x^2 + y^2$  dado que se inicia siempre en  $[0,0]$  [4.4]. Calcule la distancia cuadrática media para cada uno de los siguientes pasos:

$$N = \{10, 100, 500, 1000, 5000, 10000, 20000, 100000\} \quad (4.84)$$

Debería obtener algo como:

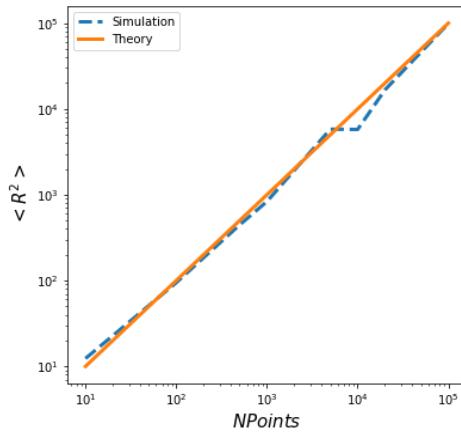


Figure 4.4: Relación entre la distancia cuadrática media y el número de pasos de un proceso Markoviano. Entender este proceso fue uno de los pasos definitivos para aceptar la existencia de las moléculas empezando el siglo XX.

6. Sea el volumen  $d$ -dimensional ( $V_d$ ):

$$V_d = \int_{-\infty}^{\infty} \cos(||r||) \exp(-||r||^2) dr^d, \quad (4.85)$$

donde  $||r|| = \sqrt{\sum_{i=1}^d x_i^2}$  y  $dr^d = dx^1 \cdots dx^d$ , con el factor de normalización:

$$f_n = (2\pi\sigma^2)^{\frac{d}{2}} \quad (4.86)$$

- a) Usando el método de Metrópolis-Hasting para  $N = 10^5$  eventos y  $\sigma = \frac{1}{\sqrt{2}}$ , muestre que los volúmenes  $d = 2, 3$  son respectivamente  $V_2 \approx 1.817671646$  y  $V_3 \approx 2.167233695$ .
7. La molécula de Hidrógeno está compuesta por dos núcleos separados una distancia  $L$ . En torno a cada núcleo existe un electrón que supondremos que está en el nivel más bajo de energía, en un órbital  $s$ . La función de onda que describe a cada electrón está dada por:

$$\psi(\vec{r}, \vec{R}) = \frac{1}{\sqrt{\pi a_0^3}} e^{-|\vec{r} - \vec{R}|/a_0} \quad (4.87)$$

donde  $a_0 = 0.529 \text{ \AA}$  es el radio de Bohr y  $\vec{R}$  es la posición del núcleo respecto al sistema de referencia. Adicionalmente, el operador de energía electrostática de la molécula (en electronvoltios), que incluye la interacción entre electrones y núcleos está dado por:

$$U(\vec{r}_1, \vec{r}_2, \vec{R}_1, \vec{R}_2) = \frac{1}{|\vec{r}_1 - \vec{r}_2|} + \frac{1}{|\vec{R}_1 - \vec{R}_2|} - \frac{1}{|\vec{r}_1 - \vec{R}_1|} - \frac{1}{|\vec{r}_2 - \vec{R}_1|} - \frac{1}{|\vec{r}_1 - \vec{R}_2|} - \frac{1}{|\vec{r}_2 - \vec{R}_2|} \quad (4.88)$$

Usar un sistema de unidades donde  $a_0 = 1$  y poner los núcleos en las posiciones:  $\vec{R}_1 = [0, 0, L/2]$  y  $\vec{R}_2 = [0, 0, -L/2]$ . La idea es calcular el valor esperado de la energía potencial como función de la separación de los núcleos de la molécula. Este promedio está dado (sin considerar el espín electrónico) por:

$$\langle U \rangle = \int d^3 r_1 d^3 r_2 |\psi(\vec{r}_1, \vec{R}_1)|^2 |\psi(\vec{r}_2, \vec{R}_2)|^2 U(\vec{r}_1, \vec{r}_2, \vec{R}_1, \vec{R}_2) \quad (4.89)$$

- a) Note que esta integral se puede ver como el valor promedio de la energía potencial evaluada en el muestreo de la densidad de probabilidad de cada electrón. Usando el algoritmo de Metropolis en coordenadas cartesianas, hacer el muestreo del cuadrado de la función de onda fijando  $L = 2$  y  $10^5$  pasos en la cadena. La distribución de puntos de la nube electrónica asociada a cada electrón está dada por [4.5]:
- b) Calcule el valor esperado de la energía potencial para el siguientes separaciones nucleares.
- `1 = np.linspace(1.0, 3.0, 10)`. Debería obtener algo similar a [4.7]:
- c) Incluya el efecto de Heitler y London (buscar el sentido físico):

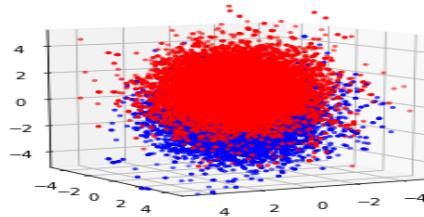


Figure 4.5: Nube electrónica encontrada con el algoritmo de Metropolis asociada al electrón en cada núcleo.

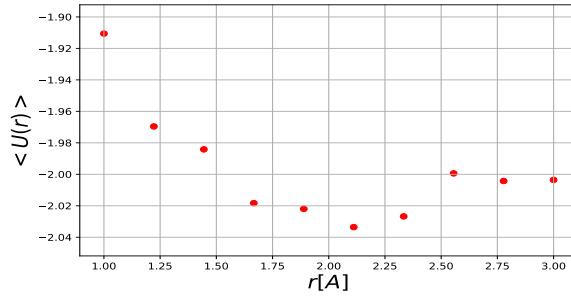


Figure 4.6: Valor esperado de la energía potencial eléctrica de la molécula de Hidrógeno. Note que existe un estado ligado que conocemos como enlace covalente, la energía del pozo de potencial está alrededor de  $U_{min} \approx 2.0 \text{ eV}$  y se da a una separación nuclear de  $d_{equilibrium} \approx 2.1a_0$ .

$$\langle U \rangle = \int d^3r_1 d^3r_2 |\psi(\vec{r}_1, \vec{R}_1)\psi(\vec{r}_2, \vec{R}_2) + \psi(\vec{r}_2, \vec{R}_1)\psi(\vec{r}_1, \vec{R}_2)|^2 U(\vec{r}_1, \vec{r}_2, \vec{R}_1, \vec{R}_2) \quad (4.90)$$

- d) Note que los valores están de acuerdo a los observados en la naturaleza: La distancia de separación de los protones es  $d_{equilibrium} = 2.4a_0$  y la energía de ligadura es 2.8 eV. Nice isn't it!

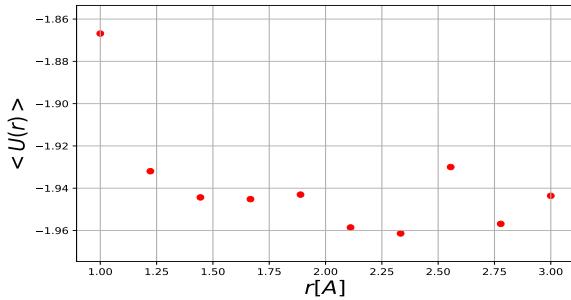


Figure 4.7: Valor esperado de la energía potencial eléctrica de la molécula de Hidrógeno incluyendo el efecto de Heitler y London. Note que existe un estado ligado que conocemos como enlace covalente, la energía del pozo de potencial está alrededor de  $U_{min} \approx 2.05 \text{ eV}$  y se da a una separación nuclear de  $d_{equilibrium} \approx 2.3a_0$ .

### Section 4.17

## Hidden Markov models

Un proceso de Markov oculto es un proceso estocástico definido por dos variables aleatorias, la variable oculta  $x(t)$  toma valores en el paso  $t$  y la variable observada  $y(t)$  toma valores en el paso  $t$ . La variable  $x(t)$  toma valores siguiendo la matriz de transición  $\mathbb{T}$  y la variable  $y(t)$  toma valores siguiendo la matriz de emisión  $\mathbb{E}$ . Por ejemplo, definimos las siguientes matrices de transición y emisión:

$$\mathbb{T} = \begin{pmatrix} & F & T \\ F & 0.7 & 0.5 \\ T & 0.3 & 0.5 \end{pmatrix} \quad (4.91)$$

$$\mathbb{E} = \begin{pmatrix} & F & T \\ R & 0.8 & 0.2 \\ V & 0.1 & 0.3 \\ A & 0.1 & 0.5 \end{pmatrix} \quad (4.92)$$

Las columnas de estas matrices representan las probabilidades de los estados en el tiempo  $t - 1$  y las filas las probabilidades asociadas al tiempo  $t$ . La suma de todos los valores en una columna debe ser igual a 1. Definamos los estados de transición como el estados de ánimo de una persona; Feliz y Triste como:  $S = [F, T]$  con las siguientes probabilidades a priori  $\pi = [0.4, 0.6]$ .

De esta forma, la probabilidad de estar feliz al día ( $t - 1$ ) y luego estar triste al siguiente día ( $t$ ) es:  $\mathbb{T}_{10} = 0.3$ . Los estados de transición que siguen este proceso aleatorio no son observables por lo que se denominan ocultos. Sin embargo, esta probabilidad de transición se propaga influyendo en la manera de vestir de la persona. Definamos los estados de emisión como el color de la ropa que usa la persona al tiempo  $t - 1$ ; Rojo, Verde, Azul como  $E = [R, V, A]$ . En este contexto, la probabilidad de que la persona use ropa azul dado que está feliz es  $\mathbb{E}_{20} = 0.1$ .

La secuencia de  $n$  eventos observados es:

$$\Omega_O = \{y(t_1), y(t_2), \dots, y(t_n)\} \quad (4.93)$$

Usando la propiedad de los procesos de Markov, se busca encontrar la secuencia de eventos ocultos más probable.

$$\Omega_{hidden} = \{x(t_1), x(t_2), \dots, x(t_n)\} \quad (4.94)$$

Suponga la siguiente secuencia de ropa usada por la persona:  $\Omega_O = [V, A, R]$ .

- a) Encuentre la probabilidad de cada secuencia de estados (ocultos) de ánimo.
- b) Encuentre el estado de ánimo más probable que tuvo la persona durante esos tres días.
- c) Calcule las probabilidades de cada estado observable ( $i$ ) como la suma de las probabilidades de todos los estados ocultos. Verifique que sumando todos los estados observables la medida es 1.

$$\sum \mathbb{P}_i = 1 \quad (4.95)$$

### Remarks:

$n_h$  es el número de estados ocultos y  $n_o$  es el número de estados observados. El número de secuencias posibles ( $n_s$ ) está dada por las variaciones con repetición de estas cantidades:

$$n_s = n_h^{n_o} = 8 \quad (4.96)$$

De las ocho secuencias ocultas posibles, supongamos que la secuencia oculta es  $S_1 = [F, F, T]$ . Entonces:

$$S_1 = \begin{pmatrix} F & F & T \\ V & A & R \end{pmatrix} \quad (4.97)$$

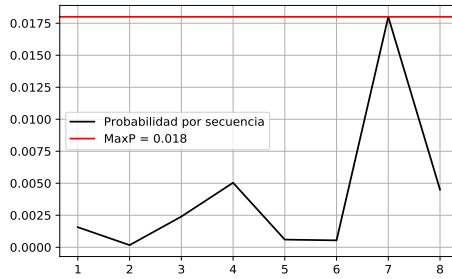


Figure 4.8: Probabilidad de cada secuencia oculta para el estado  $\Omega_O = [V, A, R]$ .

Recordemos la probabilidad condicional de dos sucesos:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A/B)\mathbb{P}(B) \quad (4.98)$$

Recordemos las matrices involucradas:

$$\mathbb{T} = \begin{pmatrix} & F & T \\ F & 0.7 & 0.5 \\ T & 0.3 & 0.5 \end{pmatrix} \quad (4.99)$$

$$\mathbb{E} = \begin{pmatrix} & F & T \\ R & 0.8 & 0.2 \\ V & 0.1 & 0.3 \\ A & 0.1 & 0.5 \end{pmatrix} \quad (4.100)$$

Por tanto, la probabilidad de la secuencia es:

$$\begin{aligned} \mathbb{P}(V, A, R \cap F, F, T) &= \underbrace{\mathbb{P}(V/F)\mathbb{P}(A/F)\mathbb{P}(R/T)}_{\text{Emission Matrix}} \times \underbrace{\mathbb{P}(F)}_{\text{Prior}} \underbrace{\mathbb{P}(F/F)\mathbb{P}(T/F)}_{\text{Transition Matrix}} \\ &= \mathbb{E}_{10}\mathbb{E}_{20}\mathbb{E}_{01}\mathbb{P}_0\mathbb{T}_{00}\mathbb{T}_{10} \\ &= 0.1 \cdot 0.1 \cdot 0.2 \times 0.4 \cdot 0.7 \cdot 0.3 = 1.68 \times 10^{-4} \end{aligned} \quad (4.101)$$

En general:

$$\mathbb{P}(O_i, q_i) = \prod_{i=1}^n \mathbb{P}(O_i/q_i) \prod_{i=1}^n \mathbb{P}(q_i/q_{i-1}) \quad (4.102)$$

**Section 4.18**
**Ejercicios: Hidden Markov models**

1. Un casino tiene dos tipos monedas una justa ( $J$ ) y una sesgada ( $B$ ), las probabilidades de obtener cara y sello son:  $\mathbb{P}_J = [0.5, 0.5]$  y  $\mathbb{P}_B = [0.9, 0.1]$  respectivamente. El tipo de moneda se escoge siguiendo esta ley de transición:

$$\mathbb{T} = \begin{pmatrix} & J & B \\ J & 0.8 & 0.2 \\ B & 0.2 & 0.8 \end{pmatrix} \quad (4.103)$$

Por otro lado, la matriz de emisión está dada por:

$$\mathbb{E} = \begin{pmatrix} & J & B \\ C & 0.5 & 0.9 \\ S & 0.5 & 0.1 \end{pmatrix} \quad (4.104)$$

Se realiza un experimento de 8 lanzamientos y se obtiene la siguiente secuencia:

$$\Omega_O = [S, C, C, C, S, C, S, C] \quad (4.105)$$

- a) Use la siguiente distribución de probabilidad a-priori  $\pi = [0.2, 0.8]$  para la moneda justa y sesgada.
- b) Encuentre la secuencia oculta más probable del tipo de moneda que se eligió en cada lanzamiento y su respectiva probabilidad  $\mathbb{P}_i$ .
- c) Calcule las probabilidades de cada estado observable ( $o$ ) como la suma de las probabilidades de todos los estados ocultos,  $\mathbb{P}_o = \sum_i \mathbb{P}_i$ .
- d) Verifique que la suma de todos los estados observables es 1.

$$\sum \mathbb{P}_o = 1 \quad (4.106)$$

- e) ¿Depende el resultado de la probabilidad a-priori?

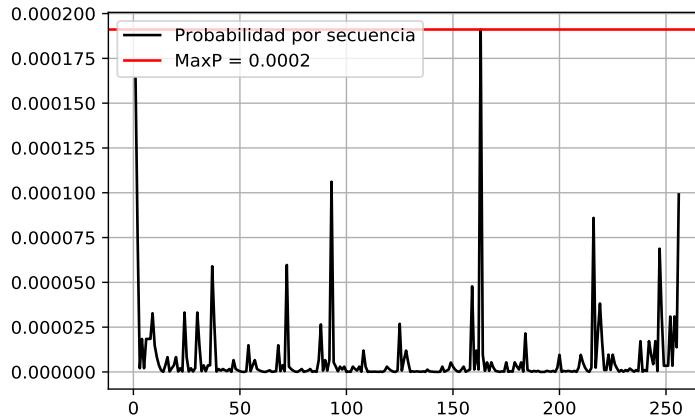


Figure 4.9: Probabilidad de cada secuencia oculta para el estado observado  $\Omega_O = [S, C, C, C, S, C, S, C]$ .

2. Las bases nitrogenadas fundamentales que componen el ADN son: Adenina (A), Citosina (C), Guanina (G) y Timina (T). Un gen se puede representar a través de una secuencia ordenada de dichas bases. Suponga la siguiente matriz de transición entre bases:

$$\mathbb{T} = \begin{pmatrix} & A & C & G & T \\ A & 0.4 & 0.2 & 0.2 & 0.2 \\ C & 0.25 & 0.25 & 0.25 & 0.25 \\ G & 0.3 & 0.3 & 0.1 & 0.3 \\ T & 0.1 & 0.1 & 0.1 & 0.7 \end{pmatrix} \quad (4.107)$$

La probabilidad a-priori está dada por:

$$\pi = [0.25, 0, 0.5, 0.25] \quad (4.108)$$

Encuentre la probabilidad de obtener el siguiente gen:

$$g = [T, G, C, T, C, A, A, A] \quad (4.109)$$

$$P_g = 7.5 \times 10^{-6}$$



# Chapter 5

## Estadística

La estadística estudia la organización y análisis de datos para dar conclusiones y tomar decisiones válidas. Es imposible estudiar el grupo completo de datos denominado población, en su lugar, se toma un sub-conjunto llamado muestra (que debe ser representativo). En general, la población puede ser finita o infinita. La descripción y análisis de únicamente la muestra, se denomina estadística descriptiva. Si por el contrario se desea inferir conclusiones sobre la población, con base a la muestra, se denomina estadística inferencial.

### Section 5.1

#### Muestreo (sampling)

Debido a que en general es imposible tener la población completa, se requiere una metodología sistemática y que sea imparcial para elegir elementos de una población. La técnica para elegir una muestra de una población de denomina muestreo. Suponiendo un tamaño de población  $N$  y de muestra ( $n$ ) determinado, el muestreo debe permitir hacer el estudio y estimaciones de las características de la población, por ejemplo, promedios, desviaciones típicas, etc. La muestra debe ser representativa para generalizar los resultados a la población.

#### Ventajas:

1. Permite tener un sub-conjunto de toda la población.
2. Si la muestra es representativa, el muestreo permite inferir los resultados de la muestra a la población.

3. Cuesta menos dinero.
4. Se pueden obtener resultados más rápido.

**Desventajas:**

1. Existe posibilidad de sesgo, i.e, preguntar sobre la preferencia de un candidato presidencial a un solo tipo de población.
2. Margen de error.
3. Dificultad para obtener los datos.

*Se tienen las siguientes técnicas de muestreo:*

### **5.1.1 Aleatorio (probabilístico)**

Es un método adecuado para seleccionar una muestra lo suficientemente representativa de una población, esta información se puede inferir al conjunto completo con cierto nivel de confianza. Este tipo de muestreo tiene las siguientes características:

- a) Se basa en que todos los elementos de la población tienen la misma probabilidad de ser elegidos.
- b) Este tipo de muestreo garantiza la representatividad de la muestra. (procesos de inferencia válidos).
- c) Permite la estimación de características poblacionales (promedios, total, proporciones, etc).
- d) Las estimaciones se realizan a través de funciones matemáticas denominadas **estimadores estadísticos**.
- e) Los errores se cuantifican mediante varianzas y desviaciones de los estimadores estadísticos.

*En particular, se tienen las siguientes técnicas aleatorias:*

### 5.1.2 Muestreo aleatorio simple

Es método consiste en seleccionar una muestra de tamaño  $n$ , cuyos elementos tienen igual probabilidad dentro de una población  $N$ . Esta muestra se obtiene elemento por elemento de forma aleatoria, adicionalmente, la construcción de la muestra no dependen del orden en que se obtiene la secuencia. Los elementos pueden ser elegidos con y sin reposición, cuando hace el muestreo con reposición, se puede obtener el mismo elemento más de una vez, en cuyo caso todas las muestras son equiprobables. Cuando realiza el muestreo sin reposición, cada elemento puede aparecer una sola vez en la muestra y cada muestra es equiprobable. La probabilidad de que un elemento de la población pertenezca a una muestra es:

$$\mathbb{P}_i = \frac{\text{Número de muestras que contienen el elemento } i}{\text{Número de muestras totales}} \quad (5.1)$$

$$\mathbb{P}_i = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{\frac{(N-1)!}{(n-1)!(N-n)!}}{\frac{N!}{n!(N-n)!}} = \frac{n}{N} \quad (5.2)$$

### 5.1.3 Muestreo aleatorio estratificado

Es posible construir la muestra a través de estratos o categorías ( $L$ ) que pueden ser etiquetadas como:  $n_1, n_2, \dots, n_i, \dots, n_L$ . Este esquema de muestreo está influenciado principalmente por tres factores: el número total de elementos en cada estrato, la variabilidad de las observaciones y en casos reales, por el costo de adquirir la muestra. El muestreo estratificado es óptimo y homogéneo, cuando se minimiza la desviación estándar de cada estrato, sin embargo, esta minimización resulta difícil de alcanzar en los casos reales. Para el muestreo estratificado, tenemos asignación igual o proporcional. Tomemos el siguiente ejemplo: Para cierta universidad tenemos el siguiente número de estudiantes por carrera:

Carrera	Estrato	Población
Medicina	1	1000
Ing. Industrial	2	500
Mercadeo	3	450
Física	4	50

Table 5.1: Distribución de estudiantes según su carrera universitaria.

Se requiere formar una muestra aleatoria estratificada de  $n = 400$  usando asignación igual y proporcional. Para asignación igual, se tiene el siguiente tamaño de estrato:

$$n_i = \frac{400}{4} = 100 \quad (5.3)$$

Para asignación proporcional tenemos los siguientes tamaños de estrato.

Carrera	Estrato	Población ( $N_i$ )	Muestra $n_i = n \times \frac{N_i}{N}$
Medicina	1	1000	200
Ing. Industrial	2	500	100
Mercadeo	3	450	90
Física	4	50	10
		$N = 2000$	$n = 400$

Table 5.2: Distribución muestral según su carrera universitaria.

#### 5.1.4 Muestreo aleatorio sistemático

Este método consiste en obtener los elementos de la muestra de una manera ordenada, donde los elementos se seleccionan en intervalos regulares. En particular, este método arroja buenos resultados cuando la variable de interés está ordenada de mayor a menor. Si la disposición de los elementos es aleatoria, el muestreo sistemático equivale al método aleatorio simple. El intervalo de selección de los elementos está dado por:

$$k = \frac{N}{n} \quad (5.4)$$

El primer valor ( $S_0$ ) se elige aleatoriamente de la lista entre el primer valor y  $k$ . Para seleccionar los siguientes  $k - 1$  cantidades se realiza una traslación sistemática de  $k$  unidades desde el primer valor:

$$\begin{aligned} i_0 &= np.random.randint(k) \\ i_{n+1} &= i_n + nk, \text{ con } n = 1, 2, \dots, k - 1 \end{aligned} \quad (5.5)$$

Tomemos la siguiente población:

$$\Omega = [0, -1, -3, -5, 5, 7, 4, 10, 12, 15, 20, 19, 50, 55, 60, 45] \quad (5.6)$$

Tomando  $k = 16/4 = 4$  y empezando con la lista ordenada en  $\Omega_1^{sorted} = 55$ , tenemos la muestra de tamaño  $k = 4$ :

$$S = [55, 19, 7, -1] \quad (5.7)$$

### 5.1.5 Muestreo no aleatorio

Son técnicas de muestreo donde la muestra se obtienen en procesos donde cada individuo no tienen la misma probabilidades de ser seleccionado.

1. Investigaciones exploratorias (por cuotas, intensional, bola de nieve).
2. La forma de selección, hace que la muestra no sea representativa. (procesos de inferencia no aplicables).

#### Section 5.2

## Distribución de frecuencias (Histogramas)

Una de las operaciones más sencillas de realizar en un conjunto de datos es la ordenación en términos de la frecuencia de repetición de los datos. Para tal fin, vamos a definir algunos parámetros de la muestra que nos permite ordenar la información.

1. Rango de la población: Es la diferencia entre el mayor y el menor valor encontrado en la muestra. Por ejemplo, se mide el peso de 300 estudiantes donde el valor mayor es 100 kg y el menor es 65.4 kg. El rango de la muestra es: 34.6 kg.
2. Frecuencia de clase: Es apropiado definir categorías o clases ( $C_i$ ) en los datos para separar la muestra en  $n$  clases. El número de datos pertenecientes a la clase  $C_i$  se denomina frecuencia de clase. Las clases tienen las siguientes propiedades:
  - (a)  $C_i \neq \emptyset$
  - (b)  $C_i \cap C_j = \emptyset$
  - (c)  $\bigcup_{i=1}^n C_i = \Omega$

3. Límite de clase: En el ejemplo del peso de los estudiantes, si tomáramos la clase  $C_1$  como todos los estudiantes que pesan entre 65 y 75 kg, entonces los límites de clase sería precisamente dichos valores.
4. Tamaño de clase: Es la diferencia entre los límites de clase, por ejemplo:  $75 - 65 \text{ kg} = 10 \text{ kg}$  es el tamaño de clase de  $C_1$ .
5. Marca de clase: Es el punto medio de cada clase, el cuál se obtiene promediando el valor de los límites de clase.

Con estas definiciones podemos definir el concepto de histograma. Un histograma es el conjunto de clases, que pueden ser representadas gráficamente como los rectángulos, cuya base esta centrada en las marcas de clase y su altura es la frecuencia de clase. Note que, el área total histograma es proporcional al tamaño total de la muestra. *En ciertos casos resulta muy útil hacer que el área del histograma sea igual a 1 para comparar la forma de la distribución con otras distribuciones.*

Así mismo, podemos definir una distribución de frecuencias acumulada, que nos da información de como se acumula la información en los datos para valores menores a una marca de clase específica. La grafica de la distribución acumulada de frecuencia se denomina ojiva. Imaginemos la siguiente situación:

Peso [kg]	Frecuencia	Acumulada
60 - 70	6	Menor a 60
70 - 80	35	Menor a 70
80 - 90	40	Menor a 80
90 - 100	19	Menor a 90
		Menor a 100
		100

Table 5.3: Distribución de pesos y distribución acumulada de pesos para un conjunto de mediciones de 100 estudiantes.

Una descripción sencilla de la distribución acumulada es que 81 estudiantes tienen un peso menor a 90 kg.

**Section 5.3****Estadísticos**

Sea  $X$  una variable aleatoria y sean  $\Omega = \{X_1, X_2, \dots, X_n\}$  variables aleatorias con la misma distribución que  $X$ . Diremos que el conjunto  $\Omega$  es una muestra aleatoria de tamaño  $n$  de  $X$ . Dada una muestra de una población, un estadístico es una función real de la muestra:

$$T = f(X_1, X_2, \dots, X_n) \quad (5.8)$$

como el estadístico es una variable aleatoria, entonces será en sí mismo una variable aleatoria. Esto significa que tendrá una distribución, una media, etc. En general, la distribución que siga el estadística se denomina distribución muestral.

**Section 5.4****Valor medio muestral**

En el caso de la media aritmética si los números  $x_1, x_2, \dots, x_n$  ocurren  $f_1, f_2, \dots, f_n$  veces, respectivamente. La media muestral está dada por:

$$\bar{X} = \frac{f_1x_1 + f_2x_2 + \dots + f_nx_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum f_i x_i}{\sum f_i}. \quad (5.9)$$

Del mismo modo, es posible definir pesos estadísticos que ponderen cada punto muestral.  $\{w_1, w_2, \dots, w_n\}$  tienen las siguiente características:

$$\begin{aligned} \sum_{i=1}^n w_i &= 1 \\ w_i &\geq 0 \quad \forall i, i = 1, 2, \dots, n \end{aligned} \quad (5.10)$$

La media ponderada está dada por:

$$\bar{X} = \frac{w_1x_1 + w_2x_2 + \dots + w_nx_n}{w_1 + w_2 + \dots + w_n} = \sum_{i=1}^n w_i x_i \quad (5.11)$$

Esta definición resulta muy útil en el caso de la media móvil, dado que, para todo punto  $i$  de la secuencia de observaciones y  $n$  puntos anteriores, la media móvil del siguiente punto  $i + 1$  queda definida por:

$$\bar{X}_{i+1}^m(n) = \sum_{j=1}^n w_j x_{i-j+1} \quad (5.12)$$

donde  $i \geq n$ . Esto significa que no es posible calcular la media móvil  $\bar{X}_2^m(3)$ , dado que en el punto muestral  $x_2$  solo se cuentan con dos observaciones anteriores  $\{x_1, x_2\}$ . Entonces, para los puntos  $i < n$  se mantienen los valores observados. Adicionalmente, es posible optimizar el valor de los pesos estadísticos para tener la mejor estimación posible. Para realizar la optimización se define un función de costo que tenga en cuenta la distancia entre las observaciones y las predicciones de la media móvil en cada punto muestral. La función de costo está dada por:

$$C(n) = \frac{1}{N} \sum_{i \geq n}^N |\bar{X}_i^m(n) - x_i| \quad (5.13)$$

La optimización se realiza bajo las restricciones dadas por la ecuación (5.10); usando algún paquete de minimización (i.e, `Scipy.optimize`).

La media armónica es útil en algunos casos particulares:

$$\frac{1}{H} = \frac{1}{N} \sum \frac{1}{x} \quad (5.14)$$

### Section 5.5

## Desviación estándar

1. La desviación estándar nos da una medida de la dispersión de los datos alrededor de la media. Para el caso poblacional está dada por:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})^2} = \sqrt{\mathbb{E}(x^2) - (\mathbb{E}(x))^2} \quad (5.15)$$

En el caso muestral, la desviación estándar está dada por:

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (5.16)$$

**Section 5.6****Teorema del límite central**

Sean  $X_1, X_2, \dots, X_n$  variables aleatorias independientes e idénticamente distribuidas con  $\mathbb{E}(X) = \mu$  y  $\mathbb{V}ar(X) = \sigma^2 < \infty$ . Se define:

$$Z_n := \frac{\frac{1}{n} \sum_{i=1}^n X_i - \mu}{\sigma / \sqrt{n}}, \quad (5.17)$$

La función de distribución de  $Z_n$  converge hacia la distribución normal estándar cuando  $n \rightarrow \infty$ . Entonces:

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n < z) = \Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx. \quad (5.18)$$

El teorema exige la existencia de la media y la varianza, sin embargo, no especifica la distribución de la variable  $X$ ; es un resultado general [3].

**Section 5.7****Ejercicios: Muestreo**

1. Use los datos almacenados en: [https://github.com/asegura4488/Estadistica\\_Herramientas/blob/main/Data/Sesion2/DatosEstratificados.csv](https://github.com/asegura4488/Estadistica_Herramientas/blob/main/Data/Sesion2/DatosEstratificados.csv) para realizar un muestreo estratificado con  $n = 200$  y  $n = 300$ . Verifique que la suma total de personas encuestadas en cada salón sea  $n$ .
2. Guarde la información final en el formato .xls para su distribución.
3. Use los datos almacenados en: [https://github.com/asegura4488/Estadistica\\_Herramientas/blob/main/Data/Sesion2/Datos\\_Sistematico.csv](https://github.com/asegura4488/Estadistica_Herramientas/blob/main/Data/Sesion2/Datos_Sistematico.csv) para realizar un muestreo sistemático con  $n = 200$ .
4. Guarde la información final en el formato .xls para su distribución.
5. La calificación de matemáticas de 80 estudiantes están en el siguiente archivo: <https://github.com/asegura4488/DataBase/blob/main/MetodosComputacionalesReforma/Matematicas.txt>. Hallar la siguiente información:
  - a) La calificación más alta.

- b) La más baja.
- c) El rango.
- d) La cinco más bajas.
- e) Las cinco más altas.
- f) La décima de mayor y menor.
- g) El número de estudiantes con calificación de 75 o más.
- h) El porcentaje de estudiantes con calificaciones mayores que 65 pero no superiores a 85.
- i) Las calificaciones de 0 a 100 que no aparecen en las notas.
6. La Tabla [5.4] muestra la distribución de frecuencia de los salarios de los empleados de una empresa anónima. Calcule la distribución acumulada de porcentajes y su respectiva gráfica.

Salarios	Número de empleados
250 - 259	8
260 - 269	10
270 - 279	16
280 - 299	14
290 - 300	10
300 - 309	5
310 - 319	2

Table 5.4: Distribución de salarios para 65 empleados de la empresa anónima.

7. Usando la distribución (empírica) de notas de matemáticas vista en clase, hacer las siguientes estimaciones:
- a) Calcular la media de la nota con `np.mean(data)`.
- b) Calcular la mediana de la nota con `np.percentile(data,50)`
- c) Calcular la distribución de frecuencia acumulada y la distribución acumulada de probabilidad  $\mathbb{P}(X \leq x)$ . En este caso, la mediana es el valor que toma la variable aleatoria que divide en 2 partes iguales la función de masa de probabilidad  $(X = x)$ .

- d) Usando la técnica de re-muestreo (Bootstrapping), obtenga la función de distribución de la media con `random.choices`, donde el tamaño de muestra sea igual al poblacional. En este caso, se analiza la variabilidad de la población respecto a la media.
- e) Hacer el fit gaussiano para obtener el valor de la media y su desviación estándar (Figura [5.1]).
- f) Calcular la mediana de la distribución bootstrapped usando la función acumulada de probabilidad (Figura [5.1]).
- g) Comparar todas las estimaciones.

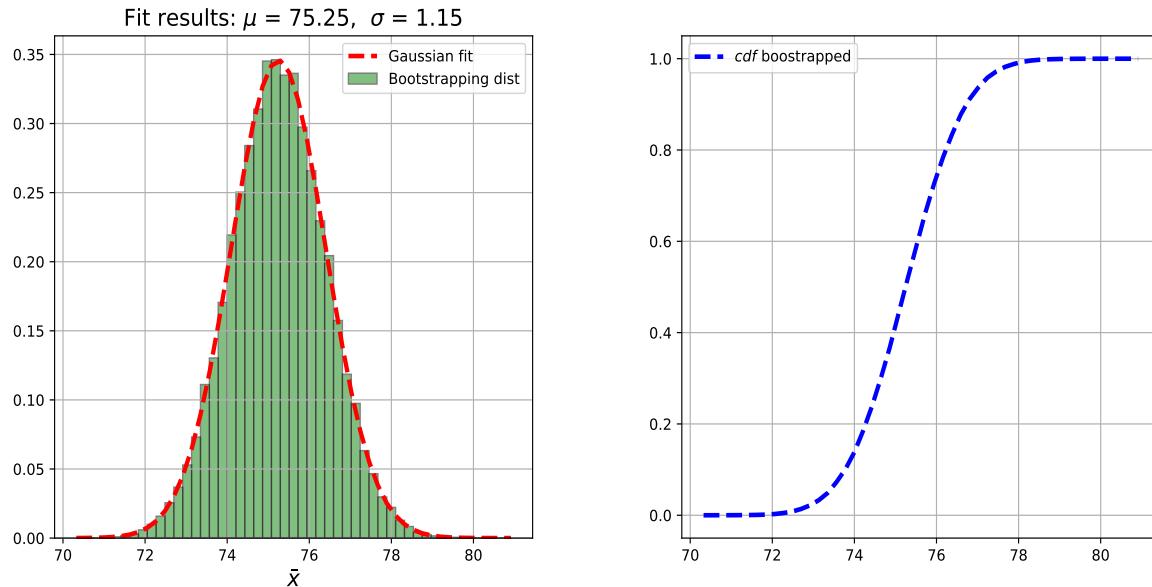


Figure 5.1: Distribución de muestreo (izquierda) y distribución acumulada de probabilidad (derecha).

8. Encuentre los valores óptimos de los pesos asociados a la predicción de las ventas. La pregunta puede ser muy complicada de responder!
9. Una persona viaja de  $A$  a  $B$  con una velocidad media de 30 millas por hora y regresa de  $B$  a  $A$  a una velocidad de media de 60 millas por hora. Hallar la velocidad media en el viaje completo. **Ans:** 40 mi/h

10. Un avión vuela  $d_1, d_2, d_3$  millas a velocidades  $v_1, v_2, v_3$  mi/h, respectivamente. Demuestre que su velocidad media es  $V$ , dada por:

$$\frac{d_1 + d_2 + d_3}{V} = \frac{d_1}{v_1} + \frac{d_2}{v_2} + \frac{d_3}{v_3} \quad (5.19)$$

es una media armónica ponderada.

11. Considere la variable aleatoria definida por una combinación lineal de otras variables aleatorias:  $X = a_1X_1 + a_2X_2 + \dots + a_nX_n$ , donde las variables aleatorias ( $X_i$ ) no son necesariamente idénticamente distribuidas, y las componentes  $[a_1, a_2, \dots, a_n] \in \mathbb{R}$ . La variable aleatoria  $X$  puede ser escrita como:  $X = a^T \vec{X}$  con la definición usual de producto interno entre vectores. El primer y segundo momento de la variable  $X$  quedan entonces expresados por:

$$\begin{aligned} E(X) &= E(a^T \vec{X}) = a^T E(\vec{X}) \\ Var(X) &= Var(a^T \vec{X}) = a^T Cov(\vec{X})a \end{aligned} \quad (5.20)$$

Donde  $Cov(\vec{X})$  es la matriz de covarianza de  $\vec{X}$ . Sea  $X_1 \sim \Gamma(2, 3)$ ,  $X_2 \sim N(5, 2)$  y  $X_3 \sim U(0, 10)$ , Genere  $N = 10^4$  eventos (que estabilice el valor de los momentos) para obtener la distribución de  $X = X_1 + 2X_2 - X_3$ . Calcule el primer y segundo momento de  $X$  a través de dos estrategias:

- a) Usando directamente el array de X.
- b) Usando las definiciones generales dadas en la Ecuación (5.20).
- c) Calcule el coeficiente de correlación de Pearson para las tres variables.
- d) Demuestre que la formula de la varianza de la media se puede escribir escalarmente como:

$$Var\left(\frac{1}{N} \sum_{i=1}^N X_i\right) = \frac{1}{N^2} \sum_{i=1}^N Var(X_i) + \frac{2}{N^2} \sum_{i=1}^N \sum_{j=i+1}^N Cov(X_i, X_j) \quad (5.21)$$

*Hint:* Encuentre el resultados para  $N = 2$  y use la intuición adquirida para el caso general.

# Chapter 6

## Estimación de parámetros

### Section 6.1

#### Mínimos cuadrados

El método de mínimos cuadrados surge como necesidad de plantear una solución aproximada al problema lineal  $\mathbb{A}\vec{x} = \vec{b}$ , donde la solución exacta no existe. Para entender el planteamiento desde el álgebra lineal se requiere el concepto de espacio columna de  $\mathbb{A}$ .

##### 6.1.1 Espacio Columna

Sea  $\mathbb{A}$  un matriz genérica de dimensiones  $m \times n$ , que se puede escribir como un arreglo de vectores columna  $\vec{v}_{m \times 1}$ :

$$A_{m \times n} = (\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n) \quad (6.1)$$

Cada uno de esos vectores tiene  $m$  componentes, por tanto  $\vec{v}_i \in \mathbb{R}^m$ , donde  $i = 1, 2, \dots, n$ . Entonces se puede definir al espacio columna de la matriz  $C(\mathbb{A})$ , como todas las combinaciones lineales de los vectores columnas de  $\mathbb{A}$ . Tomemos un vector genérico de ese subespacio:

$$\vec{a} = c_1\vec{v}_1 + c_2\vec{v}_2 + \dots + c_n\vec{v}_n \quad (6.2)$$

Notemos que el vector cero pertenece al espacio columna, dado que es una combinación lineal válida con  $c_i = 0$ , con  $i = 1, 2, \dots, n$ . Adicionalmente, vemos que es cerrado bajo la multiplicación por escalar.

$$k\vec{a} = kc_1\vec{v}_1 + kc_2\vec{v}_2 + \dots + kc_n\vec{v}_n \quad (6.3)$$

$$\vec{b} = d_1\vec{v}_1 + d_2\vec{v}_2 + \dots + d_n\vec{v}_n \quad (6.4)$$

Entonces  $\vec{b} \in C(\mathbb{A})$ . Finalmente, se puede demostrar que el espacio columna es cerrado frente a la suma. Dado que,  $\vec{a} \in C(\mathbb{A})$  y  $\vec{b} \in C(\mathbb{A})$  entonces  $\vec{a} + \vec{b} \in C(\mathbb{A})$ .

Por otro lado, para que el producto  $\mathbb{A}\vec{x}$  esté bien definido,  $\vec{x} \in \mathbb{R}^n$ . Usando la representación de la matriz  $\mathbb{A}$  se puede escribir:

$$\mathbb{A}\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \dots + x_n\vec{v}_n, \quad (6.5)$$

dado que estamos tomando todas las combinaciones lineales posibles por que  $\vec{x}$  es un vector genérico, la expresión anterior debe ser el espacio vectorial generado por los vectores columnas de la matriz  $\mathbb{A}$ , es decir,  $C(\mathbb{A})$ . En el contexto de los mínimos cuadrados, el sistema lineal  $\mathbb{A}\vec{x} = \vec{b}$  no tiene solución por que  $\vec{b} \notin C(\mathbb{A})$ .

Una vez demostrado que el vector  $\vec{b} \notin C(\mathbb{A})$ , se quiere encontrar un método que permita encontrar una solución aproximada lo más exacta posible. Entonces definimos  $\mathbb{A}\vec{x}^*$  como la proyección del vector  $\vec{b}$  en el espacio columna:

$$Proy_{C(\mathbb{A})}(\vec{b}) := \mathbb{A}\vec{x}^*. \quad (6.6)$$

Por construcción la distancia entre la proyección de  $\vec{b}$  en el espacio columna y el vector  $\vec{b}$ , está dada por:

$$\vec{d} = \mathbb{A}\vec{x}^* - \vec{b} \quad (6.7)$$

El método de mínimos cuadrados encuentra el vector  $\vec{x}^*$  que minimiza esta distancia.

$$\min(||\mathbb{A}\vec{x}^* - \vec{b}||) \quad (6.8)$$

Note que explícitamente se tiene:

$$|\vec{d}|^2 = \begin{pmatrix} P_1 - b_1 \\ P_2 - b_2 \\ \vdots \\ \vdots \\ P_n - b_n \end{pmatrix}^2 \quad (6.9)$$

Adicionalmente, note que el vector  $\vec{d}$  es ortogonal a  $C(\mathbb{A})$  y por tanto,  $\vec{d} \in C^\perp(\mathbb{A})$  que es el complemento ortogonal de  $\mathbb{A}$ . Se tiene:

$$C^\perp = \{\vec{x} \in \mathbb{R}^n \mid \vec{x} \cdot \vec{v} = 0 \ \forall \vec{v} \in C\} \quad (6.10)$$

El complemento ortogonal es el espacio nulo izquierdo de la matriz transpuesta de  $\mathbb{A}$ :

$$C^\perp(\mathbb{A}) = N(\mathbb{A}^T) \quad (6.11)$$

Recordemos que el espacio nulo de una matriz está dado por:

$$N(\mathbb{A}) = \{\vec{x} \in \mathbb{R}^n \mid \mathbb{A}\vec{x} = \vec{0}\}. \quad (6.12)$$

**proof:** Formemos una matriz cuyas filas son los  $n$  vectores de la base de  $C(\mathbb{A})$  (es decir,  $\mathbb{A}^T$ ):

$$\mathbb{A}^T = \begin{pmatrix} -\vec{v}_1 - \\ -\vec{v}_2 - \\ \vdots \\ -\vec{v}_n - \end{pmatrix} \quad (6.13)$$

de la definición de  $C^\perp(\mathbb{A})$ , si  $\vec{x} \in C^\perp(\mathbb{A})$  entonces debe ser perpendicular a todos los elementos de la base de  $C(\mathbb{A})$ :

$$\vec{x} \cdot \vec{v}_1 = \vec{x} \cdot \vec{v}_2 = \cdots = \vec{x} \cdot \vec{v}_n = \vec{0}. \quad (6.14)$$

esta expresión es la multiplicación matricial  $\mathbb{A}^T \vec{x} = 0$ , que es la definición de espacio nulo de  $\mathbb{A}^T$ :

$$N(\mathbb{A}^T) = \{\vec{x} \in \mathbb{R}^n \mid \mathbb{A}^T \vec{x} = \vec{0}\}. \quad (6.15)$$

Por tanto,  $C^\perp(\mathbb{A})$  es el espacio nulo de la matriz cuyas filas son una base de  $C(\mathbb{A})$ ; es decir,  $N(\mathbb{A}^T)$ . ■

Usando el último resultado se tiene:

$$\mathbb{A}^T(\mathbb{A}\vec{x}^* - \vec{b}) = 0 \quad (6.16)$$

$$\mathbb{A}^T\mathbb{A}\vec{x}^* = \mathbb{A}^T\vec{b} \quad (6.17)$$

### Section 6.2

## Varianza de la estimación de parámetros

En el caso de mínimos cuadrados, es posible encontrar expresiones cerradas y útiles computacionalmente para estimar la varianza de los parámetros encontrados usando este método. Se parte de la definición general de varianza:

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-p} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \frac{1}{n-p} \sum_{i=1}^n (y_i - A_i \hat{\beta})^2 \\ &= \frac{1}{n-p} (\mathbf{Y} - \mathbf{A}\hat{\beta})^T (\mathbf{Y} - \mathbf{A}\hat{\beta}) \end{aligned} \quad (6.18)$$

donde  $p$  es el número de parámetros que se ajustan en el modelo y  $\hat{\beta}$  son los parámetros encontrados. De esta forma, la desviación típica asociada a los parámetros está dada por:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}. \quad (6.19)$$

Más generalmente, la matriz de covarianza de los parámetros está dada por:

$$Var(\hat{\beta}) = \hat{\sigma}^2 (\mathbb{A}^T \mathbb{A})^{-1}. \quad (6.20)$$

**Section 6.3****Función de costo**

En el contexto de la estimación de parámetros es necesario introducir funciones de costo. En un marco conceptual Bayesiano, las funciones de costo miden una distancia entre los datos (que se consideran fijos) y el mejor modelo ajustado, por tal razón, la función de costo es una función de los parámetros a estimar. La función más sencilla y directamente relacionada con el método de mínimos cuadrados es  $\chi^2(\theta)$ , la cual está definida por:

$$\chi^2(\theta) = \sum_{i=1}^N \left( \frac{y_i - M(x_i, \theta)}{\sigma_i} \right)^2. \quad (6.21)$$

donde  $M(x_i, \theta)$  es el modelo que se propone para ajustar los datos. Los parámetros que minimizan la función de costo, son los valores óptimos que describen el mejor modelo que se ajusta a los datos. En general, están dados por:

$$\hat{\theta} = \arg \min_{\theta} \chi^2(\theta) = \arg \min_{\theta} \sum_{i=1}^N \left( \frac{y_i - M(x_i, \theta)}{\sigma_i} \right)^2. \quad (6.22)$$

**Section 6.4****Ejercicios: Mínimos cuadrados**

1. Se tienen tres líneas en  $\mathbb{R}^2$ :

$$\begin{aligned} 2x - y &= 2 \\ x + 2y &= 1 \\ x + y &= 4 \end{aligned} \quad (6.23)$$

- (a) Con el método de mínimos cuadrados encuentre el punto común a las tres líneas. Grafique las tres líneas y el punto solución, ¿qué interpretación puede dar?
- (b) Realice una búsqueda iterativa entre  $-5 \leq x \leq 5$  y  $-5 \leq y \leq 5$  con un paso de  $h = 0.01$  para encontrar la menor distancia del problema. Grafique la distancia y compare con el resultado obtenido con mínimos cuadrados.

$$\min(||\mathbb{A}\vec{x}^* - \vec{b}||) \quad (6.24)$$

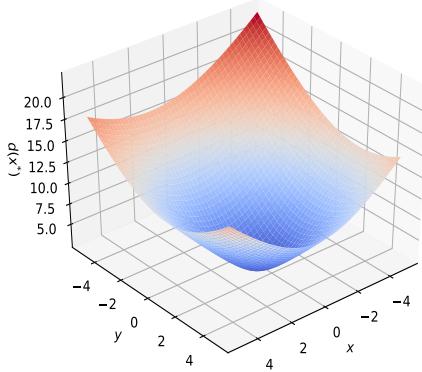


Figure 6.1: Distancia  $||\mathbb{A}\vec{x}^* - \vec{b}||$  en la región rectangular definida en el problema. La distancia mínima debe corresponder con el resultado del método de mínimos cuadrados.

2. Descargue los datos: <https://github.com/asegura4488/Database/blob/main/MetodosComputacionales/MinimosLineal.txt>. Realice el ajuste usando el método de mínimos cuadrados para encontrar los parámetros de:

$$f(x) = a_0 + a_1x \quad (6.25)$$

Grafique los datos y el ajuste mostrando el valor de las constantes en la etiqueta de la grafica.

3. Descargue los datos: <https://github.com/asegura4488/Database/blob/main/MetodosComputacionales/MinimosCuadratico.txt>. Realice el ajuste usando el método de mínimos cuadrados para encontrar los parámetros de:

$$f(x) = a_0 + a_1x + a_1x^2 \quad (6.26)$$

Grafique los datos y el ajuste mostrando el valor de las constantes en la etiqueta de la grafica.

4. Utilice el método `curve_fit` de Python para obtener los dos ajustes. Compare con los resultados anteriores.
5. Encontrar el intervalo de confianza a  $1\sigma = 68\%$  de los parámetros del ejercicio de mínimos cuadrados de *Bootstrapping*.
  - a) Haga un remuestreo con reemplazo con un tamaño de muestra igual al tamaño de los datos. `resample = np.random.choice(sample, size=len(sample), replace=True)`
  - b) Realice el ajuste del remuestreo usando los pasos del problema anterior.
  - c) Guarde el valor de cada parámetro en una lista diferente. Realice la experiencia aleatoria  $N = 10^5$  veces.
  - d) Dibuje el histograma asociado a la distribución de muestreo de cada parámetro.
  - e) De la distribución de muestreo estime el intervalo de confianza al 68% de cada parámetro del ajuste. ¿Cuál es más preciso?
6. En el caso de ajustes, es posible definir funciones de costo que midan la distancia entre los puntos muestrales y el modelo de ajuste. En el caso de mínimos cuadrados la función es  $\chi^2$ . Para  $n$  puntos y un modelo lineal la función de costo es:

$$\chi^2(a_0, a_1) = \sum_{i=1}^n (y_i - (a_0 + a_1 x_i))^2 \quad (6.27)$$

Si hablamos en términos Bayesianos,  $\Omega = \{a_0, a_1\}$  define el conjunto de modelos lineales que pueden explicar los  $n$  puntos muestrales. Al minimizar  $\chi^2(a_0, a_1)$  muestre (analíticamente) que los parámetros están dados por:

$$\begin{aligned} a_0 &= \bar{y} - a_1 \bar{x} \\ a_1 &= \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} \end{aligned} \quad (6.28)$$

donde  $\bar{x}$  y  $\bar{y}$  son los valores medios de puntos y sus imágenes. Para  $n$  puntos y un modelo cuadrático la función de costo es:

$$\chi^2(a_0, a_1, a_2) = \sum_{i=1}^n (y_i - (a_0 + a_1 x_i + a_2 x_i^2))^2 \quad (6.29)$$

Minimize  $\chi^2(a_0, a_1, a_2)$  para encontrar el siguiente sistema de ecuaciones. *Nota alguna regularidad?*

$$\begin{aligned} \sum_{i=1}^n [a_0 + x_i a_1 + x_i^2 a_2] &= y_i \\ \sum_{i=1}^n [a_0 x_i + a_1 x_i^2 + a_2 x_i^3] &= x_i y_i \\ \sum_{i=1}^n [a_0 x_i^2 + a_1 x_i^3 + a_2 x_i^4] &= x_i^2 y_i \end{aligned} \quad (6.30)$$

**Opcionalmente:** Encuentre los parámetros usando estas expresiones para los problemas 2) y 3)

7. (**Machine Learning: Logistic Regresion**) Descargue los datos de: <https://raw.githubusercontent.com/asegura4488/Database/main/MetodosComputacionalesReforma/Sigmoid.csv>

a) Defina el modelo de ajuste como:

$$M(x; \vec{\theta}) = \frac{\theta_0}{\theta_1 + e^{-\theta_2 x}} \quad (6.31)$$

donde el  $\vec{\theta}$  es el vector de parámetros del ajuste.

b) Defina la métrica (función de costo) a minimizar como:

$$\chi^2(\vec{\theta}) = \sum_{i=1}^N \left( \frac{y_i - M(x_i, \vec{\theta})}{\sigma_i} \right)^2. \quad (6.32)$$

donde  $\sigma_i = 1 \forall i$ , es decir, no se consideran los errores de  $y_i$ .

c) Muestre que las derivadas parciales de la métrica están dadas por:

$$\frac{\partial \chi^2(\vec{\theta})}{\partial \theta_i} = -2 \sum_{i=1}^N (y_i - M(x_i, \vec{\theta})) \frac{\partial M(x_i, \vec{\theta})}{\partial \theta_i} \quad (6.33)$$

d) Muestre que vectorialmente, el descenso del gradiente queda definido por:

$$\vec{\theta}_{j+1} = \vec{\theta}_j - \gamma \left( -2 \sum_{i=1}^N (y_i - M(x_i, \vec{\theta}_j)) \nabla_{\theta} M(x_i, \vec{\theta}_j) \right) \quad (6.34)$$

donde el gradiente es respecto a los parámetros:

$$\nabla_{\theta} M(x_i, \vec{\theta}) = \left[ \frac{\partial M(x_i, \vec{\theta})}{\partial \theta_0}, \quad \frac{\partial M(x_i, \vec{\theta})}{\partial \theta_1}, \quad \frac{\partial M(x_i, \vec{\theta})}{\partial \theta_2} \right] \quad (6.35)$$

- e) Use una taza de aprendizaje  $\gamma = 1 \times 10^{-3}$  o  $\gamma = 5 \times 10^{-4}$ ,  $\vec{\theta}_0 = [1, 1, 1]$ , un error de parada  $\epsilon = 0.01$  y un máximo de iteraciones de  $1 \times 10^4$ .
  - f) Grafique los datos y the best fit model con sus parámetros.
8. Calcule la proyección ortogonal del vector  $\vec{b} = (-3, -3, 8, 9)$  sobre el sub-espacio  $W$  generado por los vectores:

$$\begin{aligned} \vec{u}_1 &= (3, 1, 0, 1) \\ \vec{u}_2 &= (1, 2, 1, 1) \\ \vec{u}_3 &= (-1, 0, 2, -1) \end{aligned} \quad (6.36)$$

- a) Usando mínimos cuadrados matriciales. La proyección ortogonal es  $p_W(b) = Ax$ , donde las columnas de  $A$  son los vectores base y  $x$  es la solución de mínimos cuadrados.
- b) Con el proceso de Grand-Schmidt obtener una base ortonormal  $(v_1, v_2, v_3)$  y luego calcular la proyección sobre dicha base:  $p_W(b) = c_1v_1 + c_2v_2 + c_3v_3$ , donde  $c_i = \langle b, v_i \rangle$  para  $i = 1, 2, 3$ . Respuesta:  $p_W(b) = (-2, 3, 4, 0)$

Recuerde que el procedimiento de Grand-Schmidt es:

$$u_k = v_k - \sum_{j=1}^{k-1} \frac{\langle v_k, u_j \rangle}{\langle u_j, u_j \rangle} u_j. \quad (6.37)$$

### Section 6.5

## Función de verosimilitud (Likelihood)

Sea  $\vec{x}(x_1, x_2, \dots, x_n)$  un vector aleatorio, donde cada componente está definida como un canal de observación. Cada canal está asociado a una función de densidad de probabilidad que depende de  $m$  parámetros  $\theta_1, \theta_2, \dots, \theta_m$  dado un conjunto de datos, la función de verosimilitud está definida como:

$$\mathcal{L}(\vec{x}; \theta_1, \theta_2, \dots, \theta_m) = f_{\theta_1, \theta_2, \dots, \theta_m}(\vec{x}). \quad (6.38)$$

la cual es la función de densidad de probabilidad evaluada en el vector aleatorio  $\vec{x}$ . Si la muestra de datos es independiente e idénticamente distribuida (*iid*), la función de verosimilitud puede ser escrita como:

$$\begin{aligned} \mathcal{L}(\vec{x}; \theta_1, \theta_2, \dots, \theta_m) &= f(x_1; \theta_1, \theta_2, \dots, \theta_m) f(x_2; \theta_1, \theta_2, \dots, \theta_m) \dots f(x_n; \theta_1, \theta_2, \dots, \theta_m), \\ &= \prod_{i=1}^n f(x_i; \theta_1, \theta_2, \dots, \theta_m) \end{aligned} \quad (6.39)$$

### Section 6.6

## Método de máxima verosimilitud

Este método de estimación de parámetros consiste en maximizar la función de verosimilitud  $\mathcal{L}(\vec{x}; \theta_1, \theta_2, \dots, \theta_m)$ . El valor del parámetro que maximiza la función se denota por  $\hat{\theta}$ , y se conoce como estimador máximo verosímil. En la práctica, el cálculo de la función conjunta de verosimilitud conduce a cantidades muy grandes debido a la productoria, en ese sentido, se opta por calcular el logaritmo de la función de verosimilitud que transforma las productorias en sumatorios; y realizar un proceso de minimización cambiando de signo a la función. Adicionalmente, se tienen dos casos de estimación: la estimación de parámetros para datos no agrupados (unbinned likelihood fit) y estimación para datos agrupados (binned likelihood fit).

### 6.6.1 Unbinned likelihood fit

En este caso tenemos que minimizar:

$$\hat{\theta} = \arg \min_{\theta} - \sum_{x_i} \ln(\mathcal{L}(x_i; \theta)) \quad (6.40)$$

donde el índice  $i$  de la suma corre sobre todo el conjunto de datos.

**Ejemplo 1:**

Sea  $\mathcal{A}(x_1, x_2, \dots, x_n)$ , donde  $\mathcal{A} \sim Pois(\lambda)$  tiene distribución de Poisson con parámetro  $\lambda$ .

$$\mathcal{L}(\vec{x}; \lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \quad (6.41)$$

Tomando logaritmo natural tenemos:

$$Ln(\mathcal{L}(\vec{x}; \lambda)) = -n\lambda + \sum_{i=1}^n x_i Ln(\lambda) - Ln(\prod_{i=1}^n x_i!) \quad (6.42)$$

El parámetro máximo verosímil es la media muestral.

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}. \quad (6.43)$$

### Ejemplo 2:

Sea  $\mathcal{A}(x_1, x_2, \dots, x_n)$ , donde  $\mathcal{A} \sim Exp(\alpha)$  tiene distribución exponencial con parámetro  $\alpha$ .

$$\mathcal{L}(\vec{x}; \alpha) = \prod_{i=1}^n \frac{1}{\alpha} e^{-x_i/\alpha} = \left(\frac{1}{\alpha}\right)^n e^{-\sum_{i=1}^n x_i/\alpha} \quad (6.44)$$

Tomando el logaritmo natural de la función de verosimilitud.

$$Ln(\mathcal{L}(\vec{x}; \alpha)) = nLn(1/\alpha) - \sum_{i=1}^n \frac{x_i}{\alpha}. \quad (6.45)$$

Al maximizar  $Ln(\mathcal{L}(\vec{x}; \alpha))$ , el parámetro máximo verosímil es la media muestral.

$$\hat{\alpha} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X} \quad (6.46)$$

## 6.6.2 Binned likelihood fit

En el caso de datos agrupados la función de verosimilitud debe ser evaluada en las marcas de clase de la distribución de frecuencia. Para tal propósito, se debe construir el histograma que agrupa a los datos y se usará el índice  $j$  para aclarar que la suma no es sobre el conjunto de datos ( $\vec{x}$ ) si no sobre cada clase ( $\bar{x}_j$ ) del histograma. Si tenemos  $m$  clases:

$$\bar{x} = \{\bar{x}_0, \bar{x}_1, \dots, \bar{x}_m\}. \quad (6.47)$$

El histograma por definición está dado por:

$$\vec{h} = \{h_0, h_1, \dots, h_m\}, \quad (6.48)$$

donde  $h_j$  es el número de eventos contenidos en la  $j$ -ésima clase. El método de estimación de parámetros está dado por:

$$\hat{\theta} = \arg \min_{\theta} - \sum_j h_j \ln(\mathcal{L}(\bar{x}_j; \theta)). \quad (6.49)$$

### 6.6.3 Varianza de los estimadores

Vamos a expandir la función de verosimilitud alrededor del estimador, lo que significa que estamos en el punto máximo.

$$\ln(\mathcal{L}(\theta)) \cong \ln(\mathcal{L}(\hat{\theta})) + \frac{\partial \ln(\mathcal{L}(\theta))}{\partial \theta} \Big|_{\theta=\hat{\theta}} (\theta - \hat{\theta}) + \frac{1}{2} \frac{\partial^2 \ln(\mathcal{L}(\theta))}{\partial \theta^2} \Big|_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2 \quad (6.50)$$

Dado que estamos en el máximo:

$$\ln(\mathcal{L}(\theta)) \cong \ln(\mathcal{L}(\hat{\theta})) - \frac{(\theta - \hat{\theta})^2}{2\hat{\sigma}^2} \quad (6.51)$$

donde hemos definido:

$$\hat{\sigma}^2 = - \left( \frac{\partial^2 \ln(\mathcal{L}(\theta))}{\partial \theta^2} \right)^{-1} \quad (6.52)$$

Note que si evaluamos la función de verosimilitud en  $\theta = \hat{\theta} \pm \hat{\sigma}$  obtenemos

$$\ln(\mathcal{L}(\hat{\theta} \pm \hat{\sigma})) = \ln(\mathcal{L}(\hat{\theta})) - 1/2 \quad (6.53)$$

Esto significa que a  $1\sigma$  de desviación estándar el logaritmo de la función de verosimilitud evaluada desde el máximo debe cambiar en  $1/2$ .

#### Section 6.7

### Ejercicios: Parameter Estimation

1. **(Theoretical)** Sea  $\mathcal{A}(x_1, x_2, \dots, x_n)$ , donde  $\mathcal{A} \sim \mathcal{N}(\mu, \sigma)$  con parámetros  $\mu$  y  $\sigma$ . Muestre que los estimadores máximo verosímiles son:

$$\begin{aligned}\hat{\mu} &= \frac{1}{n} \sum_{i=1}^n x_i \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\end{aligned}\quad (6.54)$$

2. Para el ejercicio anterior, muestre que la matriz Hessiana evaluada en los estimadores es:

$$H(\hat{\mu}, \hat{\sigma}^2) = \begin{pmatrix} -n(\hat{\sigma}^2)^{-1} & 0 \\ 0 & -\frac{n}{2}(\hat{\sigma}^2)^2 \end{pmatrix}$$

¿Es esta matriz definida positiva o negativa?

3. Implemente el algoritmo de Metrópolis Hastings para hacer el ajuste de un histograma cuyas frecuencias relativas siguen una distribución normal.

a) Descargue los datos de: [Gaussian Data](#)

b) Utilice la siguiente distribución a priori uniforme:

$$\pi(\mu, \sigma) = \begin{cases} 1 & si \quad 3 \leq \mu \leq 5, \quad 0.5 \leq \sigma \leq 3.5 \\ 0 & si \quad \text{Otro caso} \end{cases} \quad (6.55)$$

c) Escriba la función de Likelihood Gaussiana.

$$\mathcal{L}(\mathbf{x}/\mu, \sigma) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\mu-x_i)^2/2\sigma^2} \quad (6.56)$$

d) Calcule el logaritmo de la distribución posterior.

$$\ln(\mathbb{P}(\mu, \sigma/\mathbf{x})) \sim \ln(\mathcal{L}(\mathbf{x}/\mu, \sigma)\pi(\mu, \sigma)) \quad (6.57)$$

Dado que la distribución encontrada no está normalizada, la relación que existe es de proporcionalidad.

- e) Use el algoritmo de Metrópolis para realizar el muestreo de  $\ln(\mathbb{P}(\mu, \sigma/\mathbf{x}))$  con  $N = 2 \times 10^4$  eventos.
- f) Estime el mejor valor de los parámetros del modelo  $(\hat{\mu}, \hat{\sigma})$ .

- g) Encuentre los errores  $\sigma_{\pm}^+$  de los parámetros en un intervalo de confianza del  $CL = 68\%$ .
4. Resuelva el problema anterior usando el paquete `Emcee`.  
<https://emcee.readthedocs.io/en/stable/>
  5. En general la varianza de estimadores es no calculable:

$$V(\hat{\theta}) = E(\hat{\theta}^2) - E(\hat{\theta})^2 \quad (6.58)$$

En el caso de la distribución exponencial tenemos un valor analítico dado por:

$$\begin{aligned} V(\hat{\theta}) &= \int_0^\infty \dots \int_0^\infty \left( \frac{1}{n} \sum_{i=1}^n x_i \right)^2 \frac{1}{\theta} e^{-x_1/\theta} \dots \frac{1}{\theta} e^{-x_n/\theta} dx_1 \dots dx_n \\ &- \left[ \int_0^\infty \dots \int_0^\infty \left( \frac{1}{n} \sum_{i=1}^n x_i \right) \frac{1}{\theta} e^{-x_1/\theta} \dots \frac{1}{\theta} e^{-x_n/\theta} dx_1 \dots dx_n \right]^2 \\ &= \frac{\theta^2}{n} \end{aligned} \quad (6.59)$$

- a) Intente encontrar este resultado analíticamente.
- b) Con el método de MonteCarlo compruebe este resultado para un conjunto de  $n=20$  variables aleatorias  $x_i \dots x_n \sim Exp(\theta = 2)$ . Generar varias muestras de distribuciones exponenciales para tener un buen promedio en el ensamble, por ejemplo:  $N = 10^6$  (Se obtiene algo como  $Var(\hat{\theta}) = 0.199$ ).
6. En física de altas energías el siguiente *Toy Model* es relevante. Se tiene un número observado de eventos totales ( $n = 10$ ), un número de eventos de física conocida  $b = 9$  y un número de eventos esperados de nueva física  $s = 4$ . La distribución está caracterizada por la siguiente función de Likelihood:

$$\mathcal{L}(\mu, \epsilon) = Poisson(n; \mu s + \epsilon b)Gauss(\epsilon; 1., 0.1) \quad (6.60)$$

$$\mathcal{L}(\mu, \epsilon) = \frac{1}{n!} e^{-(\mu s + \epsilon b)} (\mu s + \epsilon b)^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\epsilon-1)^2}{2\sigma^2}} \quad (6.61)$$

Donde  $\epsilon$  es la eficiencia de reconstrucción de los eventos de física conocida tiene valor medio  $\bar{\epsilon} = 1.$  y desviación estándar  $\sigma = 0.1$ . La contribución Gaussiana modela el error

sistemático asociado a la estimación de  $b$ .  $\mu$  se conoce como *signal strength*, el cuál es muy importante para descartar o no, nuevas teorías es física. Con esta información:

- Reproduzca las gráficas de la Figura [6.2]. Los parámetros pueden variar como  $0. < \mu < 2$ . y  $0. < \epsilon < 2$ .
- Usando el algoritmo de Metropolis-Hastings encuentre  $\hat{\mu}$  y  $\hat{\epsilon}$  y los errores asociados a cada parámetro.
- Elimine el error sistemático dado por la parte gaussiana. ¿Cuál sería el valor de  $\hat{\mu}$ ? A qué conclusiones llega?

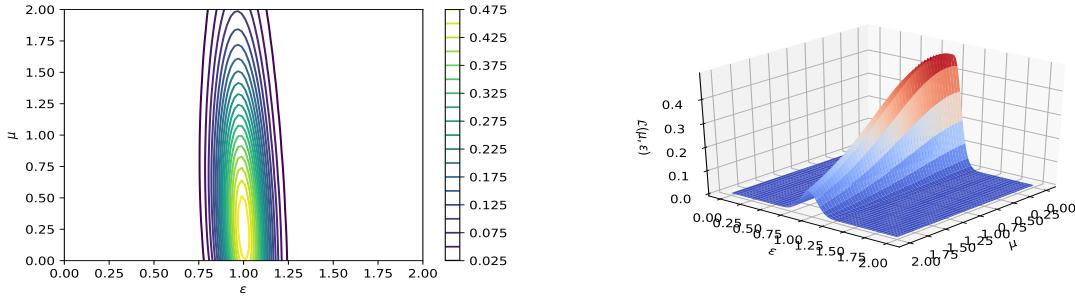


Figure 6.2: Gráfica de curvas de nivel de la función  $\mathcal{L}(\mu, \epsilon)$  (izquierda) y superficie (derecha).

- Un equipo de medición de 3 canales detecta los pulsos de una posible señal extraterrestre. En un momento determinado del tiempo se tiene la estadística descrita por la siguiente Figura [6.3]:

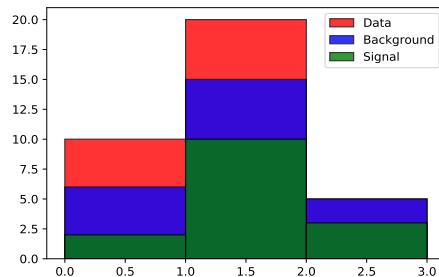


Figure 6.3: Medida del equipo de 3 canales.

El conjunto de datos es: Data=[10,20,5], el background o la señal esperada sin existencia extraterrestre bkg=[6,15,10] y un modelo de señal extraterrestre Signal=[2,10,3]. La función de Likelihood que describe esta observación para un canal está dada por:

$$\mathcal{L}_i(\mu, \epsilon) = Poisson(n; \mu s + \epsilon b)Gauss(\epsilon; 1., 0.1), \quad (6.62)$$

$$\mathcal{L}_i(\mu, \epsilon) = \frac{1}{n!} e^{-(\mu s + \epsilon b)} (\mu s + \epsilon b)^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\epsilon-1)^2}{2\sigma^2}}. \quad (6.63)$$

Donde  $\epsilon$  es la eficiencia de reconstrucción de los eventos sin existencia extraterrestre y  $\sigma = 0.1$  su desviación estándar.  $\mu$  es la fortaleza de la señal y es una medida para comprobar o descartar, en este caso, la señal extraterrestre. La función de Likelihood (the joint probability function) de todos los canales está dada por :

$$\mathcal{L}(\mu, \epsilon) = \prod_{i=1}^{NC_{channels}} \mathcal{L}_i(\mu, \epsilon). \quad (6.64)$$

- a) Use el algoritmo de Metropolis-Hastings para encontrar los mejores estimadores que describen esta observación  $\hat{\mu}$  y  $\hat{\epsilon}$  con sus respectivos errores (asuma que todos los canales tiene la misma resolución experimental ( $\sigma$ )). Los parámetros pueden variar para  $0. \leq \mu \leq 3.$  y para  $0.5 \leq \epsilon \leq 1.5.$  **Tenga cuidado cuando la caminata aleatoria tome valores negativos para  $\mu$**  ⊕.
- b) Si  $\mu$  es cercano a cero indica que los datos están bien descritos por la hipótesis de no existencia extraterrestre. Por el contrario, si  $\mu$  se acerca a uno *podría indicar* que existe vida extraterrestre. ¿Qué conclusiones reportaría usted sobre la observación realizada?. Nota: En realidad se debe hacer una prueba de hipótesis pero se pierde la alineación entre lo visto en clase y la evaluación.

# Chapter 7

## Confidence Intervals

### Section 7.1

#### Intervalos de confianza para la media

Sean  $x_1, x_2, \dots, x_n$  variables aleatorias independiente con  $\mathbb{E}(x_i) = \mu$  y  $Var(x_i) < \infty$ , entonces:

$$z_n := \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad (7.1)$$

la función de distribución de  $z_n$ , converge a una función de distribución normal estándar  $z \sim N(0, 1)$ .

El intervalo de confianza para la media con varianza conocida está dada por:

$$IC_{1-\alpha} = \left[ \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right] \quad (7.2)$$

Este intervalo es válido para población aproximada normal, u otra distribución poblacional con  $n \geq 30$ .

Cuando no se conoce la varianza poblacional y la muestra es pequeña, se necesita hacer una estimación del intervalo usando la distribución T-Student. El intervalo de confianza para la media está dado por:

$$IC_{1-\alpha} = \left[ \bar{x} - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{1-\alpha/2, n-1} \frac{S}{\sqrt{n}} \right] \quad (7.3)$$

donde  $S$  es la varianza muestral.

### Section 7.2

## Intervalo de confianza para la diferencia de medias

Para encontrar un intervalo de confianza para la diferencia de medias ( $n \leq 30$ ), se tiene la siguiente expresión:

$$IC_{1-\alpha} = \left[ (\bar{x}_1 - \bar{x}_2) - t_{\alpha/2, \nu} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{1-\alpha/2, \nu} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right] \quad (7.4)$$

Los grados de libertad se obtienen usando:

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}} \quad (7.5)$$

Para encontrar un intervalo de confianza para la diferencia de medias ( $n > 30$ ), se tiene la siguiente expresión:

$$IC_{1-\alpha} = \left[ (\bar{x}_1 - \bar{x}_2) - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right] \quad (7.6)$$

### Section 7.3

## Intervalo de confianza para una proporción

Para estimar el intervalo de confianza de un proporción, se tiene:

$$IC_{1-\alpha} = \left[ \hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < P < \hat{p} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] \quad (7.7)$$

donde se considera la aproximación binomial a la normal si  $n\hat{p} \geq 10$  y  $n(1-\hat{p}) \geq 10$ .

### Section 7.4

## Intervalo para la varianza

El estimador de la varianza muestral está dada por:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (7.8)$$

Este estadístico es un estimador insesgado de la varianza poblacional puesto que:

$$\mathbb{E}(s^2) = \sigma^2 \quad (7.9)$$

Con base a los criterios de la ley de los grandes números y suponiendo que las muestras siguen una distribución normal,  $X \sim N(\mu, \sigma^2)$ . La varianza muestral como variable aleatoria sigue una función de distribución  $\chi_{n-1}^2(s)$  de  $n - 1$  grados de libertad.

$$(n - 1) \frac{s^2}{\sigma^2} \sim \chi_{n-1}^2 \quad (7.10)$$

En problemas donde la varianza poblacional no es conocida, la distribución permite encontrar intervalos de confianza para la varianza poblacional; dada la varianza muestral.

### Section 7.5

## Ejercicios: Intervalos de confianza

- El contenido de botellas de un litro tienen una distribución aproximadamente normal con varianza de 0.15 *litros*<sup>2</sup>. Se toma una muestra aleatoria de cajas y se mide el contenido, obteniendo el siguiente resultado.

$$S = [0.974, 0.950, 0.932, 1.104, 1.038, 0.920, 0.935, 0.907, 0.810, 0.915] \quad (7.11)$$

- Construya el intervalo de confianza del 95% para la media población.

$$IC_{95\%} = \left[ 0.708 < \mu < 1.188 \right] \quad (7.12)$$

- Construya el intervalo de confianza del 95% usando el método de Bootstrapping ( $N = 50000$  toy experiments).

$$IC_{95\%} = \left[ 0.903 < \mu < 1.00 \right] \quad (7.13)$$

Notar que los IC no son consistentes, esto puede indicar que la población que describe a los datos no es normal. Usted puede calcular el IC en el caso de varianza desconocida y comprobar que se obtiene un mejor resultado.

2. Construir un IC de 90% de nivel de confianza para la estatura media de un grupo de estudiantes. Se sabe que la estatura (en  $cm$ ) es aproximadamente normal y está dada por:

$$S = [175, 176, 180, 164, 170, 170, 181, 169, 163, 190, 170, 171] \quad (7.14)$$

$$IC_{90\%} = [169.28 < \mu < 177.21] \quad (7.15)$$

3. Se tienen dos muestras de la medida de la resistencia eléctrica de un lote de dispositivo electrónicos, las cuales están dadas por:

$$\begin{aligned} S_1 &= [6.0, 5.0, 6.5, 5.0, 4.0, 5.0, 5.0, 5.0, 7.0, 5.5, 4.5] \\ S_2 &= [7.0, 8.0, 8.5, 7.4, 8.9, 6.7, 9.0, 8.4, 7.8, 5.3, 8.1] \end{aligned} \quad (7.16)$$

Determine el intervalo al 95% de nivel de confianza. ¿Pertenecen las muestras a la misma población?

$$IC_{95\%} = [-3.16 < \mu_1 - \mu_2 < -1.60] \quad (7.17)$$

Las muestras no pertenecen a la misma población, presenta una desviación a  $\sim 4\sigma$ .

4. Una moneda equilibrada es lanzada 500 veces. Sea  $X$  la variable aleatoria asociada al número de soles que aparecen en los  $n$  lanzamientos. Usando la distribución Binomial y la aproximación a la normal, estime:
- $\mathbb{P}(X \leq 260.) = 0.8261$  Binomial,  $0.8144$  Normal.
  - $\mathbb{P}(230 \leq X \leq 260) = 0.926$  Binomial,  $0.926$  Normal.
5. Se estudia la intención de voto de un candidato  $A$  a la presidencia de un país. Para ello, se toma una muestra aleatoria de 800 habitantes y 325 de ellos responden a la intención de votar por  $A$ . Encontrar el intervalo de confianza a 98%, para la proporción de votantes con intención de voto por el candidato  $A$ .

$$IC_{98\%} = [0.365 < P < 0.446] \quad (7.18)$$

6. Una empresa de equipo médico quiere saber la proporción de clientes que prefieren su marca. El gerente afirma que el 30% de los clientes del mercado prefieren sus productos. Una muestra de 100 clientes indicó que 24 usan dicha marca. Calcule el intervalo de confianza al 90% e indicar si el gerente tiene la razón.

$$IC_{90\%} = \left[ 0.169 < P < 0.310 \right] \quad (7.19)$$

El gerente está en lo correcto dentro de la confianza de la hipótesis nula.

7. Demuestre que  $\mathbb{E}(s^2) = \sigma^2$ . *Hint:* Recuerde que  $Var(x) = \sigma^2$ .
8. Genere una muestra de números  $x \sim N(0, 9)$  de tamaño 100.
- Escriba las funciones para calcular la varianza muestral y  $\chi^2_{n-1}$ .
  - Con el método de re-muestreo (Bootstrapping) genere ( $N = 50000$  toy experiments) muestras con reemplazo para encontrar la distribución de la varianza.
  - Dibuje el histograma y la función de distribución de probabilidad  $\chi^2_{n-1}$ , usando `chi2.pdf(x_, df=n-1)`.
  - Calcule el valor medio de la distribución usando el histograma y usando `chi2.mean(df=df)`.
  - Calcule el valor medio de la varianza poblacional (debería ser cercano a 9).
  - Calcule el intervalo de confianza de este estimador a un nivel de confianza del 95% (significancia  $\alpha = 0.05$ ).

$$IC_{95\%} = \left[ \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2}} \right] \quad (7.20)$$

9. A un grupo de individuos en un hospital se sometió a un medicamento experimental, al final se midió el nivel de glóbulos blancos en sangre. En una escala específica los resultados fueron los siguientes:

$$S = [6.0, 6.4, 7.0, 5.8, 6.0, 5.8, 5.9, 6.7, 6.1, 6.5, 6.3, 5.8] \quad (7.21)$$

Suponiendo que la población es normalmente distribuida, determine un intervalo de confianza al 95% para la varianza poblacional del nivel de colesterol.

$$IC_{95\%} = \left[ 0.0770 < \sigma^2 < 0.4428 \right] \quad (7.22)$$

10. Un fabricante de parte de automóvil garantiza que sus baterías duran en promedio 3 años con una desviación estándar de 1 año, si 5 baterías presentan una varianza de 0.815. Realice su estudio al 99% de nivel de confianza.
- a) ¿Está la observación de acuerdo con la afirmación del fabricante de que la desviación estándar es de 1 año?
  - b) ¿Cuál es el p-value de la observación?  $p = 0.620$ .
  - c) ¿Cuál es valor límite (crítico) que puede tener la varianza de las baterías para mantener la afirmación del fabricante?  $S_{up}^2 = 3.715$ .
11. Encuentre la probabilidad de que una muestra aleatoria de 30 observaciones de una población normal con varianza  $\sigma^2 = 10$ . Tenga una varianza muestral:
- (a)  $\mathbb{P}(S^2 \geq 9.) = 0.379$
  - (b)  $\mathbb{P}(3.5 \leq S^2 \leq 11.) = 0.675$
  - (c) Calcule el p-value de observar  $S^2 = 15$ ,  $p = 0.04$

# Chapter 8

## Pruebas de Hipótesis

El procedimiento estadístico para determinar si los datos de una muestra son compatibles con las características de una población. Para contextualizar el problema se realizan las siguientes definiciones:

- Una hipótesis estadística es una proposición sobre los parámetros de una población.
- Se denomina hipótesis nula a la hipótesis que se desea contrastar, la denominamos  $H_0$ . La hipótesis alternativa la denominamos con  $H_1$ .
- Un estadístico de prueba es un estadístico utilizado para determinar si se rechaza o no la hipótesis nula  $H_0$ .
- Región de no rechazo es el conjunto de valores del estadístico de prueba, para los cuales no rechazamos la hipótesis nula. Su región complementaria se llama la región de rechazo de  $H_0$  en favor de  $H_1$ .
- Dado que hay una probabilidad de que el estadístico genere un valor en la región de rechazo, se tienen dos tipos de errores: Error tipo-1 es rechazar la hipótesis nula cuando es verdadera y el Error tipo-2 es no rechazar la hipótesis nula cuando es falsa.
- El nivel de significación del test determina la probabilidad de que el estadístico genere un valor en la región de rechazo de  $H_0$ . Se denomina  $\alpha$  y está restringido al intervalo  $0 \leq \alpha \leq 1$ .

**Section 8.1****Ejercicios: Tablas de contingencia**

1. En cierta empresa, se calcula que la cantidad promedio de artículos vendidos por semana es 2000. Se quiere averiguar si hay una relación entre el día de la semana y la cantidad de artículos que se venden. La Tabla [8.1] muestra las observaciones.

	< 1800	1800-2200	> 2200
Lunes-Miércoles	38	55	52
Jueves-Viernes	39	65	55
Sábado-Domingo	43	60	63

Table 8.1: Tabla de contingencia entre artículos vendidos y los días de la semana

- a) Calcule el número total de observaciones.
- b) Calcule los vectores de frecuencias marginales.
- c) Calcule los vectores de probabilidad marginales.
- d) Calcule la matriz de probabilidad.
- e) Calcule la matriz de frecuencias esperadas.
- f) Calcule la función de costo:  $\chi^2$  dada por:

$$\chi^2 = \sum_{i=1}^n \frac{(f_e - f_o)^2}{f_e} \quad (8.1)$$

donde  $f_e$  es la frecuencia esperada y  $f_o$  es la frecuencia observada de cada punto de la tabla.

- g) Calcule el número de grados de libertad ( $\nu$ ).
- h) Dibuje la distribución  $\chi^2_\nu$  con el número correcto de grados de libertad.
- i) Describa la hipótesis nula y la alternativa.
- j) Calcule el p-value observado.
- k) ¿Cuál es valor crítico de la distribución para un nivel de confianza del 95%?
- l) Indique si se rechaza o no, la hipótesis nula.

2. La universidad de los Andes realiza una encuesta a  $N = 100$  estudiantes sobre la intención de voto de tres candidatos a la presidencia, los candidatos son:  $A$ ,  $B$  y  $C$ . Adicionalmente, se tuvo en cuenta el género de cada estudiante. Los resultados se pueden resumir en la siguiente Tabla:

	A	B	C
Mujeres	11	15	34
Hombre	23	6	11

Table 8.2: Tabla de contingencia asociada a la encuesta realizada por la universidad.

Los resultados mostrados en la Tabla [8.2] representan la frecuencias observadas ( $f_o$ ) de la intención de voto de cada candidato. La universidad desea saber si las variables género y candidato son o no independientes con un nivel de confianza del 95%. Para tal propósito, planteamos las siguientes hipótesis:

- a)  $H_0$  : Las variables género y candidato son independientes. (Hipótesis nula)
- b)  $H_1$  : Las variables género y candidato no son independientes. (Hipótesis alternativa)

Note que la hipótesis alternativa es la negación de la hipótesis nula  $H_1 = \neg H_0$ . **Suponemos que la hipótesis nula es verdadera** para calcular las frecuencias esperadas ( $f_e$ ). Entonces:

- a) Si las variables son independientes entonces  $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ . Entonces, deberá calcular las probabilidades marginales y luego calcular el valor de la probabilidad de cada entrada en la tabla. Usando el producto de Kronecker sería muy eficiente y claro.
- b) La frecuencia esperada en cada posición es:  $f_e^i = \mathbb{P}(A \cap B)^i \cdot N$ , i.e, la probabilidad de cada evento multiplicado por el tamaño de muestra.

La matriz de probabilidad  $\mathbb{P}(x, y) = \mathbb{G}(x) \otimes \mathbb{H}(y)$ , puede ser escrita como el producto de Kronecker de los vectores de probabilidad marginal (sería interesante que implementaran esta operación en su código). Finalmente, para medir la discrepancia entre las frecuencias esperadas y observadas se tiene el estadístico de prueba (muy conocido)  $\chi^2$  dado por:

$$\chi^2 = \sum_{i=1}^k \frac{(f_o^i - f_e^i)^2}{f_e^i} \quad (8.2)$$

Este valor se compará con la distribución  $\chi^2_{\alpha, (r-1)(c-1)}$ . En esta definición,  $\alpha = 0.05$  se refiere al nivel de significación de la prueba,  $r$  es el número de filas y  $c$  el número de columnas.  $(r - 1)(c - 1)$  son los grados de libertad (conceptos por definir) asociados a nuestra prueba estadística; entonces para nuestro ejemplo,  $\chi^2_{0.05, 2} \approx 5.991$  nos definirá el valor crítico del estadístico a un nivel de confianza del 95%.

En Python:

```
from scipy.stats import chi2
chi2.ppf(0.95, df=2)
```

`ppf` devuelve el percentil asociado a un (p-value)  $p = 0.05$ , para una distribución  $\chi^2$  de dos grados de libertad. Calcule el  $\chi^2 \approx 16.50$ , que es un valor más extremo que el valor crítico, por lo cuál, debemos rechazar la hipótesis nula con un nivel de confianza del 95%. Las variables son dependientes estadísticamente.

Remark:

Efectivamente, como muestra la tabla, las mujeres prefieren el candidato  $C$  y los hombres el candidato  $A$ .

### Section 8.2

## Prueba de hipótesis para generadores de números aleatorios

¿Sabías que el generador de números de Numpy (Mersenne - Twister 1998), tiene uniformidad en  $d=623$  dimensiones al 95% de nivel de confianza? Para probar la uniformidad de los números aleatorios generados por pseudo-generadores es posible realizar una prueba de hipótesis. Por lo cuál, se tienen las siguientes hipótesis:

- a)  $H_0$  : La secuencia de números sigue una distribución uniforme en  $k$ -dimensiones.

- b)  $H_1$  : La secuencia de números no sigue una distribución uniforme en  $k$ -dimensiones.

### Section 8.3

## Ejercicios: Hypotesis testing

1. (**Binomial**) Se realiza un experimento aleatorio del lanzamiento de 5 monedas  $N = 1000$ , cuyo espacio muestral está definido por  $\Omega = \{C, S\}$ . Se mide el número  $x$  de caras que se observan durante el experimento. El resultado del experimento se resume en: <https://raw.githubusercontent.com/asegura4488/Database/main/MetodosComputacionalesReforma/BinomialCoins.csv>. Note que es un archivo para cargar en **Pandas**. Se desea saber si el número de caras tiene una distribución binomial  $x \sim Bin(n = 5, p)$ . Se tienen las siguientes hipótesis:

- a)  $H_0$  : El número de caras tiene distribución binomial  $x \sim Bin(n = 5, p = \hat{p})$ .
- b)  $H_1$  : El número de caras tiene distribución binomial.

Suponemos que la hipótesis nula es verdadera, de este modo, necesitamos realizar los siguientes pasos para realizar la bondad del ajuste:

- a) Estimar el parámetro de la distribución  $\hat{p} = 0.4946$
- b) Suponiendo que  $H_0$  es verdadera, estimar la probabilidad de que se presenten el número de caras encontrados en el experimento. Ejemplo: la probabilidad de obtener 0 caras en un experimento es:

$$\mathbb{P}(X = 0 | \hat{p} = 0.4946) = \binom{5}{0} 0.4946^0 (1 - 0.4946)^{5-0} = 0.0329 \quad (8.3)$$

- c) Estimar las frecuencias esperadas asociadas a el número de errores  $f_e^i(x) = \mathbb{P}^i(x) \times N$ .
- d) Dibuje los histogramas correspondientes a las distribución esperada y observada.
- e) Medir la discrepancia que existe entre las frecuencias esperadas y observadas a través del estadístico de prueba  $\chi^2$ .

$$\chi^2 = \sum_{i=1}^k \frac{(f_o^i - f_e^i)^2}{f_e^i} \quad (8.4)$$

- f) Cuál es el p-value de esta observación.

- g) Calcular el valor crítico de la distribución  $\chi^2_{\alpha, k-t-1}$ , donde la significancia es  $\alpha = 0.05$ .  $k = 10$  es el número de categorías,  $t$  es el número de parámetros que se estimaron en el proceso, en este caso, se estima  $\hat{\lambda}$ . Note que los grados de libertad son  $gl = 4$ .
- h) Si el estadístico de prueba es menor que el valor crítico no rechazamos la hipótesis nula.
- i) Dar la conclusión.
2. (**Poisson**) Una empresa de motos realiza la construcción de  $N = 440$  motos. Durante las pruebas mecánicas y eléctricas, la empresa reportó el número de fallas que presenta cada artículo. El archivo que resume esta información es: <https://raw.githubusercontent.com/asegura4488/Database/main/MetodosComputacionalesReforma/PoissonCars.csv>. Note que es un archivo para cargar en **Pandas**. La empresa quiere saber si la distribución de errores es una distribución de Poisson y cuál es el estimador máximo verosímil que describe dicha distribución. De esta manera, se tienen las siguientes hipótesis.
- $H_0$  : El número de errores tiene distribución Poisson  $n \sim Poiss(\lambda = \hat{\lambda})$ .
  - $H_1$  : El número de errores no tiene distribución Poisson.

Suponemos que la hipótesis nula es verdadera, de este modo, necesitamos realizar los siguientes pasos para realizar la bondad del ajuste:

- Estimar el parámetro de la distribución  $\hat{\lambda} = 3.047$
- Suponiendo que  $H_0$  es verdadera, estimar la probabilidad de que se presenten el número de errores encontrados en las motos. Ejemplo: la probabilidad de obtener 0 errores en una moto es:

$$\mathbb{P}(x = 0 | \hat{\lambda} = 3.047) = \frac{e^{-3.047} 3.047^0}{0!} = 0.04746 \quad (8.5)$$

- Estimar las frecuencias esperadas asociadas a el número de errores  $f_e^i(x) = \mathbb{P}^i(x) \times N$ .
- Dibuje los histogramas correspondientes a las distribución esperada y observada.
- Medir la discrepancia que existe entre las frecuencias esperadas y observadas a través del estadístico de prueba  $\chi^2$ .

$$\chi^2 = \sum_{i=1}^k \frac{(f_o^i - f_e^i)^2}{f_e^i} \quad (8.6)$$

- f) Cuál es el p-value de esta observación.
- g) Calcular el valor crítico de la distribución  $\chi^2_{\alpha,k-t-1}$ , donde la significancia es  $\alpha = 0.05$ .  $k = 10$  es el número de categorías,  $t$  es el número de parámetros que se estimaron en el proceso, en este caso, se estimo  $\hat{\lambda}$ . Note que los grados de libertad son  $gl = 8$ .
- h) Si el estadístico de prueba es menor que el valor crítico no rechazamos la hipótesis nula.
- i) Dar la conclusión.
3. Una partícula se mueve aleatoriamente desde el centro de una caja cuadrada de longitud  $l = 5$ . El ángulo para realizar el paso se genera uniformemente entre 0 y  $2\pi$  y luego se actualiza la posición actual:

```
x += lstep * np.cos(theta)
y += lstep * np.sin(theta)
```

$lstep$  es el tamaño del paso, use  $lstep = 0.4$  para este experimento. Se quiere medir el número de pasos promedio que le toma a una partícula salir de la caja. Genere  $N = 200$  experimentos guardando el número de pasos que le toma a la partícula salir de la caja. Una posible trayectoria se describe en la Figura [8.1]

- a) Eliga una distribución de clases que tenga su valor mínimo y máximo en los valores mínimo y máximo de la lista y un ancho de paso  $h = 15$  steps:
- ```
bins = np.arange(min_,max_+h,h)
```
- b) Calcule el histograma de la distribución.
- c) Normalice el histograma a la unidad.
- ```
H1Norm = H1 / np.sum(H1*w)
```
- d) Realice el ajuste de la distribución usando la densidad de Weibull exponencial:

```
from scipy.stats import exponweib
a,c,d,e = exponweib.fit(lista)
```

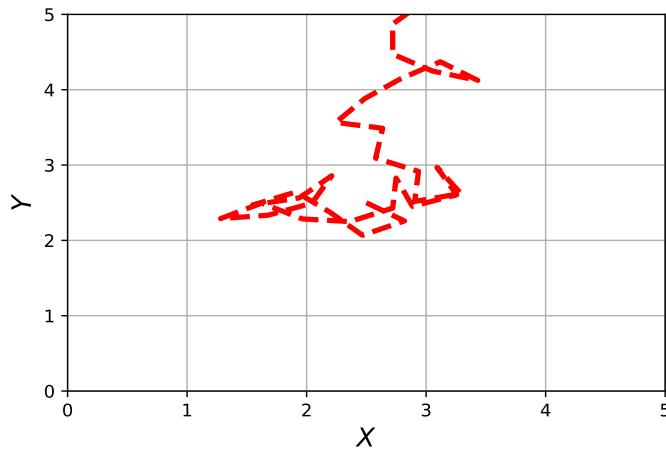


Figure 8.1: Trayectoria de la partícula (Toy model). En este experimento la partícula le tomo 33 pasos salir de la caja.

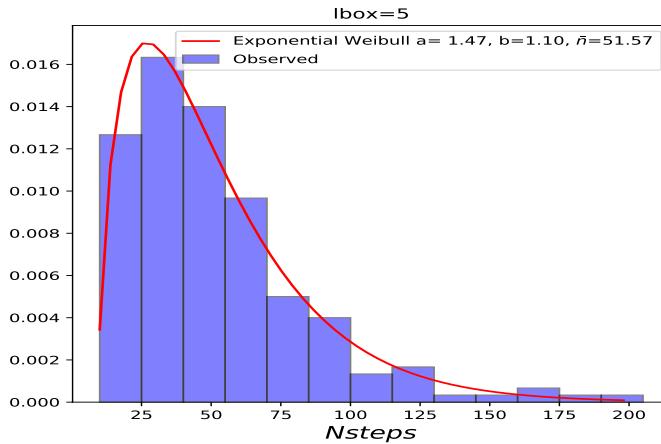


Figure 8.2: Ajuste a los datos usando la distribución de Weibull exponencial.

debería obtener un resultado similar a Figura [8.2]:

- e) Calcule el valor medio de pasos usando los métodos de `scipy`: `mean = exponweib.mean(a,c,d,e)`
- f) Ahora se quiere realizar la bondad en el ajuste. Genere la distribución esperada usando el método de generación de `Scipy`, con mismo número de eventos observados  $N = 200$ .  
`ExpFreq.append(exponweib.rvs(a,c,d,e))`

Debería obtener distribuciones similares como se muestra en la Figura [8.3].

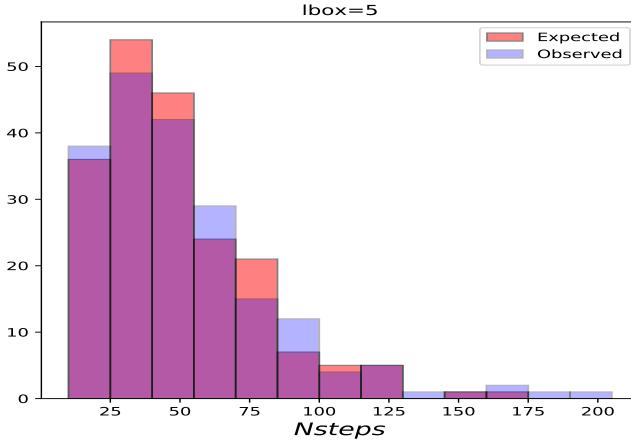


Figure 8.3: Comparación entre la distribución observada y esperada para el número de pasos de escape de la partícula.

- g) Calcule el estadístico de prueba  $\chi^2$ : (Ignore las clases que tengan menos de 5 entradas, es decir las colas de la distribución).

$$\chi^2 = \sum_{i=1}^k \frac{(f_o^i - f_e^i)^2}{f_e^i} \quad (8.7)$$

- h) Calcular el valor crítico de la distribución  $\chi_{\alpha, k-t-1}^2$ , donde la significancia es  $\alpha = 0.05$ .  $k$  es el número de categorías,  $t$  es el número de parámetros que se estimaron en el proceso, en este caso, se estima  $\hat{n}$ . Note que los grados de libertad son  $gl = \#Clases - 1 - 1$ .
- i) Si el estadístico de prueba es menor que el valor crítico no rechazamos la hipótesis nula.
- j) Dar la conclusión.

4. En el caso 2D, se divide el espacio entre  $[0, 1]$  en  $k^d$  celdas de igual área, luego se genera una muestra aleatoria de tamaño  $2n$  (note que el número de eventos es par), podemos construir  $n$  pares no solapados  $(x_0, x_1), (x_2, x_3), \dots, (x_{2n-1}, x_{2n})$  y estimar la frecuencia observada de puntos en cada celda. Si suponemos que la hipótesis nula es verdadera, la frecuencia esperada de puntos en cada celda es:

$$f_e^i = \frac{n}{k^d} \quad (8.8)$$

El estadístico de prueba que mide la discrepancia entre la frecuencia esperada y observada está dado por el estadístico  $\chi^2$ :

$$\chi^2 = \sum_{i=1}^k \frac{(f_o^i - f_e^i)^2}{f_e^i} \quad (8.9)$$

Calcule el valor crítico de la distribución  $\chi_{\alpha, k^d-1}^2$ , donde la significancia es  $\alpha = 0.05$ .  $k$  es el número de celdas en las que se divide el dominio. Note que los grados de libertad son  $gl = k^d - 1$ .

- (a) Genere una muestra de  $2 * (1000)$  números aleatorios con el generador de Numpy.
- (b) Use una partición de  $k = 10$  celdas.
- (c) Calcule las frecuencias observadas y estadístico de prueba. Guardar los conteos en objetos `M = np.array((k,k))`. *En este caso, es posible usar histogramas 2D para contar los puntos en cada celda.*
- (d) Calcule el p-value de la observación.
- (e) ¿Valida o rechaza la hipótesis al 95% de nivel de confianza de que los números tienen una distribución uniforme en  $d = 2$  dimensiones?

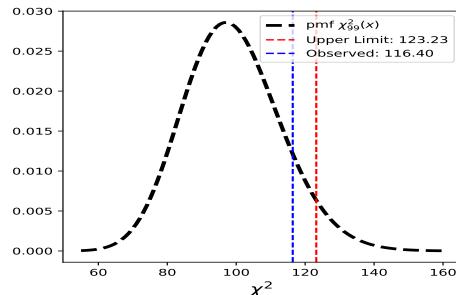


Figure 8.4: Distribución  $\chi_{99}^2$ , el valor observado es menor que el valor crítico de la distribución.  $p - value = 0.111$ , se acepta  $H_0$ .

5. En el caso 3D, se divide el espacio entre  $[0, 1]$  en  $k^d$  celdas de igual volumen, luego se genera una muestra aleatoria de tamaño  $3n$  (note que el número de eventos es multiplo de 3), podemos construir  $n$  pares no solapados  $(x_0, x_1, x_2), (x_3, x_4, x_5), \dots, (x_{3n-2}, x_{3n-1}, x_{3n})$  y estimar la frecuencia observada de puntos en cada celda. Si suponemos que la hipótesis nula es verdadera, la frecuencia esperada de puntos en cada celda es:

$$f_e^i = \frac{n}{k^d} \quad (8.10)$$

El estadístico de prueba que mide la discrepancia entre la frecuencia esperada y observada está dado por el estadístico  $\chi^2$ :

$$\chi^2 = \sum_{i=1}^k \frac{(f_o^i - f_e^i)^2}{f_e^i} \quad (8.11)$$

Calcule el valor crítico de la distribución  $\chi^2_{\alpha, k^d-1}$ , donde la significancia es  $\alpha = 0.05$ .  $k$  es el número de celdas en las que se divide el dominio. Note que los grados de libertad son  $gl = k^d - 1$ .

- (a) Genere una muestra de  $3 * (1000)$  números aleatorios con el generador de Numpy.
- (b) Use una partición de  $k = 10$  celdas.
- (c) Calcule las frecuencias observadas y estadístico de prueba. Guardar los conteos en objetos  $M = np.array((k,k,k))$ .
- (d) Calcule el p-value de la observación.
- (e) ¿Valida o rechaza la hipótesis al 95% de nivel de confianza de que los números tienen una distribución uniforme en  $d = 3$  dimensiones?

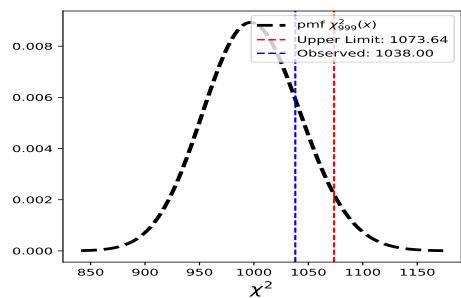


Figure 8.5: Distribución  $\chi^2_{999}$ , el valor observado es menor que el valor crítico de la distribución.  $p - value = 0.190$ , se acepta  $H_0$ .

# References

- [1] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927, <https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1927.0118>.
- [2] Rubin H Landau, José Páez, Manuel José Páez Mejía, and Cristian C Bordeianu. *A survey of computational physics: introductory computational science*. Princeton University Press, 2008, <https://psrc.aapt.org/items/detail.cfm?ID=11578>.
- [3] Sheldon M Ross. *Introduction to probability and statistics for engineers and scientists*. Academic press, 2020.
- [4] Ronald E Walpole, Raymond H Myers, Sharon L Myers, and Keying Ye. *Probabilidad y estadística para ingeniería y ciencias*. 4 Edición, 2007.