

# Apuntes de Laboratorio 1

5 de agosto de 2016

# Índice general

<b>1. Presentación de datos experimentales</b>	<b>3</b>
1.0.1. Gráficos . . . . .	3
<b>2. Medidas y errores</b>	<b>8</b>
2.1. precisión, exactitud y sensibilidad . . . . .	8
2.2. Cifras significativas . . . . .	9
2.2.1. Ejercicio . . . . .	9
2.3. número de cifras significativas para expresar resultados . . . . .	9
2.4. Aproximación y Redondeo . . . . .	10
2.5. Orden de magnitud . . . . .	10
2.6. Tipos de errores . . . . .	10
2.7. Representación grafica de error aleatorio y sistemático . . . . .	10
2.8. Diferentes formas de expresar el error aleatorio . . . . .	11
2.9. Propagación de incertezas en las operaciones . . . . .	11
<b>3. Noción de Probabilidades</b>	<b>16</b>
3.1. Introducción . . . . .	16
3.1.1. Experimento aleatorio . . . . .	16
3.1.2. Población . . . . .	17
3.1.3. Muestra . . . . .	17
3.1.4. Espacio muestral . . . . .	17
3.1.5. Evento . . . . .	17
3.2. Interpretación de la Probabilidad . . . . .	17
3.2.1. Axiomas de probabilidad . . . . .	18
3.2.2. Consecuencias directas de los axiomas . . . . .	18
3.3. Probabilidad condicional . . . . .	18
3.4. Reglas de adición . . . . .	19
3.5. Reglas de multiplicación . . . . .	20
3.6. Reglas de Probabilidad total . . . . .	21
3.7. Teorema de Bayes . . . . .	21
<b>4. Distribuciones de Probabilidad</b>	<b>22</b>
4.1. Fundamentos matemáticos . . . . .	22
4.2. Variables aleatorias, histogramas y distribuciones . . . . .	23
4.2.1. Variable aleatoria . . . . .	23
4.3. Histogramas . . . . .	23
4.4. Distribuciones de probabilidad . . . . .	24
4.4.1. Distribuciones de Poisson . . . . .	26
4.4.2. Distribución de Gauss o Normal . . . . .	27
4.4.3. Distribución normal . . . . .	27
<b>5. Estimadores y método de máxima Verosimilitud</b>	<b>32</b>
5.1. Promedio, mediana, desviación estándar, varianza . . . . .	32
5.2. Definición de máxima verosimilitud. . . . .	32

<b>6. Ajuste de curvas</b>	<b>34</b>
6.1. Modelos de regresión . . . . .	34
6.2. Regresión Lineal Simple . . . . .	36
6.3. Regresión Polinomial . . . . .	38
6.3.1. Regresión lineal múltiple . . . . .	39
6.3.2. Generalización de la idea de Regresión y mínimos Cuadrados . . . . .	40
6.4. Métodos de Ajuste . . . . .	42
6.4.1. mínimos Cuadrados . . . . .	42
6.4.2. mínimos Cuadrados Ponderados . . . . .	42
6.4.3. máxima Verosimilitud . . . . .	43
6.5. análisis (Gráfico e histograma) de residuos . . . . .	45
<b>7. Series de tiempo</b>	<b>46</b>
7.1. Introducción . . . . .	46
7.2. Componentes de una Serie Temporal . . . . .	46
7.3. Componentes de una serie: tendencia, estacionalidad y ruido . . . . .	48
7.4. análisis de la tendencia . . . . .	48
7.5. métodos de análisis de componentes . . . . .	50

# Capítulo 1

## Presentación de datos experimentales

Supongamos que medimos dos magnitudes físicas  $x$  e  $y$ , que están vinculadas de alguna manera y que se desea descubrir cómo se relacionan. La primera magnitud,  $x$ , se mide con un instrumento distinto a la segunda,  $y$ . De esta manera uno obtiene un conjunto de pares ordenados  $(x, y)$ , que son la base empírica para establecer la relación matemática entre ambas variables. Independiente de cómo sea la adquisición de datos, usted poseerá un conjunto de datos a analizar.

Una de las maneras de presentar este conjunto de datos es a través de una tabla.

- Número
- Título
- Definición de símbolos
- Unidades
- Error asociado a las medidas

Frecuencia $f$ , [Hertz] $\pm 1$	Voltaje $V$ , [Volt] $\pm 0,2$
5	2.6
20	5.0
40	7.8
60	11.5
80	13.9

Cuadro 1.1: Voltaje versus frecuencia en un circuito eléctrico.

Sin embargo de una tabla no se puede concluir la forma específica en que una de las variables depende de la otra. El descubrimiento de esa forma, se puede realizar a través de gráficos.

### 1.0.1. Gráficos

La idea básica al construir un gráfico es encontrar la relación matemática existente entre las dos variables en estudio. Al representar gráficamente los valores medidos se puede visualizar la distribución. Esa distribución determina la forma de la dependencia matemática entre ambas variables.

Un Gráfico debe contener la siguiente información (ver por ejemplo, la figura 1.0.1):

- Número
- Título
- Nombre de la magnitud física o símbolo de las magnitudes

- Unidades de medida en cada eje
- Si corresponde, se pueden incluir las barra de errores asociadas a cada punto.

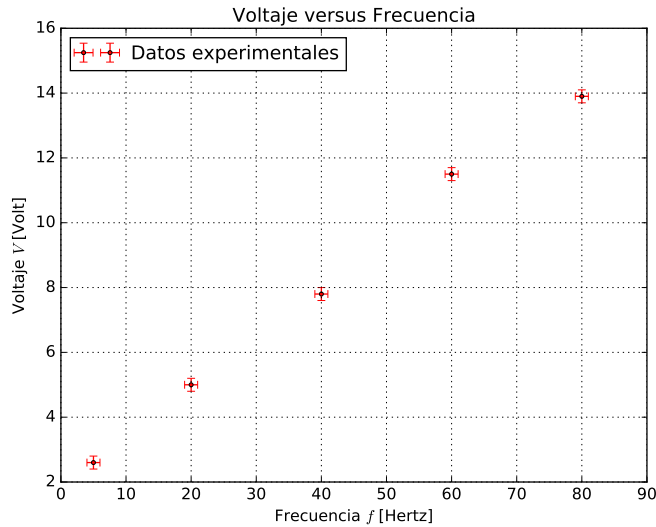


Figura 1.1: Datos experimentales de diferencia de potencial versus frecuencia. Código Python en apéndice ??.

## Lineal

La idea básica es graficar los datos obtenidos durante el experimento (ver fig. 1.0.1), recopilados en la tabla 1.1. En el caso particular mostrado en la fig. 1.0.1 usted puede ver que la dependencia es del tipo lineal. La mejor recta es aquella que está más cerca de todos los puntos a la vez, aunque no tiene necesariamente que pasar por todos los puntos, más aún, *puede no pasar por ninguno de ellos*.

La ecuación de la recta esta dada por  $y = mx + n$ , donde

$$m = \frac{\Delta V}{\Delta f} = 0,5, \quad (1.1)$$

$$n = V - mf = 2,0 - 0,15 \times 0 = 2,0, \quad (1.2)$$

$$V(f) = 0,15f + 2,0, \quad 5 < f < 80. \quad (1.3)$$

## Log-Log

En esta escala usted encuentra que los espacios están divididos en segmentos desiguales en ambas direcciones. En cada dirección las divisiones igualmente espaciadas son potencias de 10, correspondiendo a los ciclos que se repiten.

De manera análoga al caso anterior, se grafican los datos recopilados durante la experiencia (ver fig. 1.0.1), que se encuentran ordenados en la Tabla 1.2.

La ecuación de la recta para este caso, está dada por:

$$\log(L) = m \log(F) + \log(A), \quad (1.4)$$

donde  $m$  es la pendiente, definida como

$$m = \frac{\Delta \log(L)}{\Delta \log(F)} = \frac{\log(3) - \log(2)}{\log(40) - \log(18)} = 0,508. \quad (1.5)$$

Para determinar la constante  $A$ , evaluamos en un punto sobre la recta:

$$\log(L) = 0,508 \log(F) + \log(A). \quad (1.6)$$

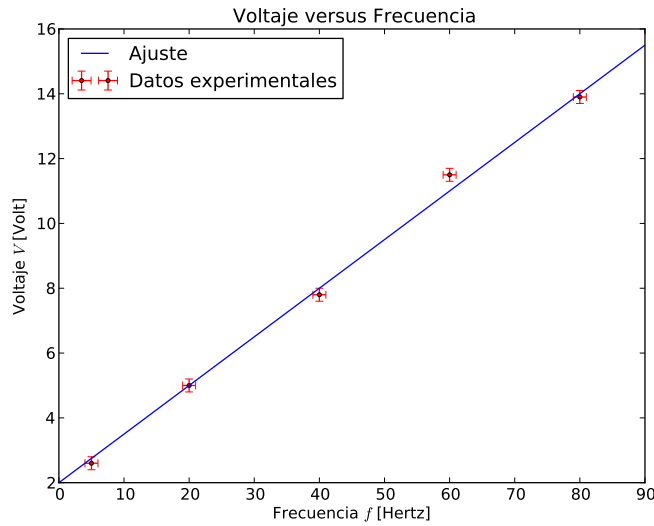


Figura 1.2: Voltaje versus frecuencia en un circuito eléctrico y ajuste lineal. Código Python en apéndice ??.

Fuerza $F$ , [N]	Deformación $L$ , [mm] $\pm 0,2$
$5,0 \pm 0,3$	1.0
$16 \pm 1$	2.0
$80 \pm 1$	4.0
$160 \pm 5$	6.0
$400 \pm 5$	9.8

Cuadro 1.2: Fuerza aplicada sobre una barra metálica.

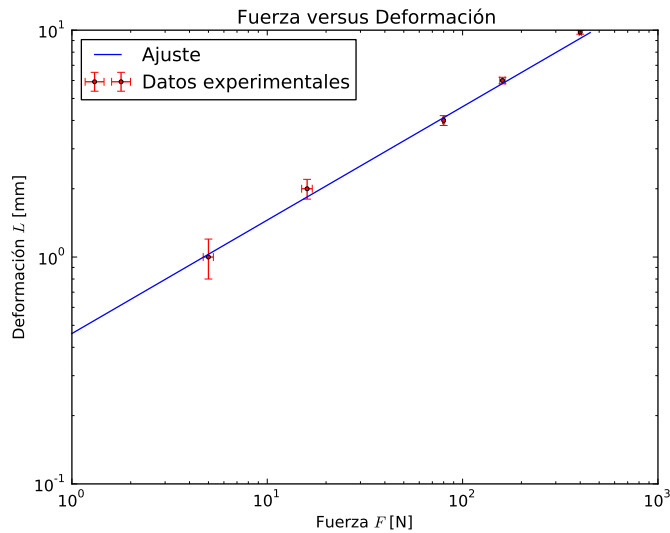


Figura 1.3: Fuerza aplicada sobre una barra metálica. Gráfico log-log y ajuste. Código Python en apéndice ??.

Evaluando en  $F = 1$ , encontramos

$$\log(L) = 0,508 \log(1) + \log(A) \quad (1.7)$$

$$= \log(A), \quad (1.8)$$

y por lo tanto  $L = A \approx 0,46$ . Por lo tanto, nuestro ajuste es en general de la forma (1.4), que es equivalente a

$$L = A \times F^m, \quad (1.9)$$

de modo que en este caso tendríamos

$$L \approx (0,46) \times F^{0,5} \text{ mm}, \quad 5 \leq F \leq 400. \quad (1.10)$$

### Semilog

En este caso usted encuentra que en una dirección las divisiones son desiguales como en la situación anterior y en la otra dirección las divisiones de los espacios están igualmente distribuidos.

Procediendo como en los casos anteriores, ordenamos los datos experimentales en la Tabla 1.3 y luego se grafican (ver fig. 1.0.1), donde se aprecia una dependencia lineal.

Tiempo $t$ , [ms] $\pm 0,1$	Carga eléctrica $Q$ , [mC] $\pm 0,002$
4,0	0,780
8,0	0,260
16,0	0,040
20,0	0,015
$24,0 \pm 5$	0,005

Cuadro 1.3: Carga eléctrica y tiempo en un condensador.

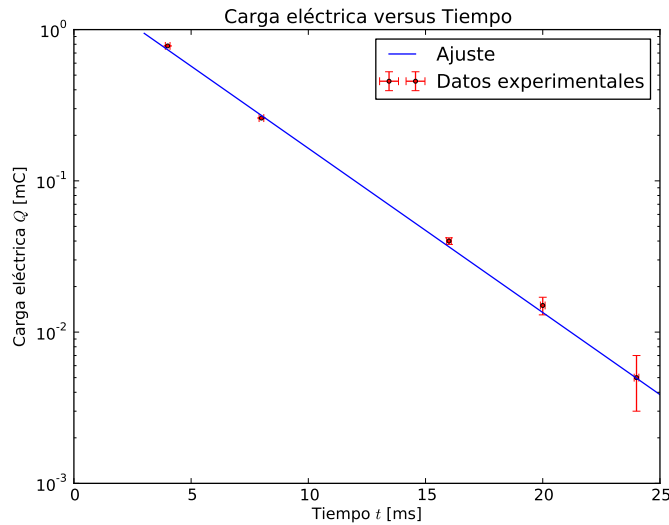


Figura 1.4: Carga eléctrica y tiempo en un condensador. Gráfico semilog y ajuste. Código Python en apéndice ??.

Para el ajuste de una recta, se tiene

$$\log(Q) = mt + \log(A), \quad \Leftrightarrow \quad Q = A \times 10^{mt} \quad (1.11)$$

$$m = \frac{\log(0,0022) - \log(0,046)}{(27 - 15) \times 10^{-3}} \approx -110 \quad (1.12)$$

En muchas ocasiones será conveniente expresar la función ajustada usando base  $e$ :

### Repaso:

$$y = e^x \Rightarrow \ln y = x, \quad \log(y) = x \log(e), \quad (1.13)$$

$$\frac{\log(y)}{\ln(y)} = \log(e) \Rightarrow \log(y) = \ln(y) \log(e). \quad (1.14)$$

Aplicando esta propiedad a la ecuación  $Q(t)$ , encontramos

$$m = \frac{\log(y_f) - \log(y_i)}{t_f - t_i} = \frac{(\ln(y_f) - \ln(y_i)) \log(e)}{t_f - t_i}, \quad (1.15)$$

$$|m| = \left| \frac{\ln(y_f) - \ln(y_i)}{t_f - t_i} \right| \log(e) =: \frac{\log(e)}{\tau}, \quad \Rightarrow \quad \tau \approx \frac{\log(e)}{110} \approx \frac{0,4343}{110} \approx 0,004 \quad (1.16)$$

Usando la definición de  $\tau$  podemos reescribir la expresión (1.11) como

$$Q(t) = A \times e^{-t/\tau}. \quad (1.17)$$

Si  $t = 0$  entonces  $Q(0) = A \approx 2$ . Por lo tanto, nuestro ajuste puede escribirse como

$$Q(t) \approx 2 \times e^{-t/0,004} \text{ mC}, \quad 4 \text{ s} \leq t \leq 24 \text{ s}. \quad (1.18)$$

Debemos hacer notar que en los tres casos mencionados para ajustar curvas, tanto para el cálculo de la pendiente como para el punto donde la recta corta el eje vertical, se usaron los valores de la “mejor” recta y no (directamente) los valores experimentales.



## Capítulo 2

# Medidas y errores

Si medimos una única vez, el resultado de nuestro experimento se expresa como:  $(x \pm \Delta x)$  unidades, donde  $x$  es nuestra medida y  $\Delta x$  es el error asociado a nuestra medida.

Para instrumentos digitales, el error estará dado por la cifra más pequeña que pueda medir nuestro instrumento (sensibilidad). Si el instrumento es análogo el error estará dado por la mitad de la sensibilidad del instrumento de medición.

Para el caso en que se mida varias veces la misma cantidad y suponiendo que los errores asociados a estas medidas se producen por causas aleatorias, un buen estimador del error o de la variabilidad del resultado, podría ser la desviación estándar; expresando el resultado de nuestro experimento como

$$\bar{x} \pm \Delta x, \quad (2.1)$$

donde  $\bar{x}$  es el **promedio** de nuestras medidas y  $\Delta x$  es el mayor error, entre el error asociado al instrumento y el error calculado:

$$\bar{x} := \frac{1}{N} \sum_{i=1}^N x_i \quad (\text{promedio}). \quad (2.2)$$

$$s := \sqrt{\frac{1}{(N-1)} \sum_{i=1}^N (x_i - \bar{x})^2}, \quad (\text{desviación estándar}). \quad (2.3)$$

En este punto es bueno tener presente que existen otros valores medios y otras estimaciones para el error, como:

$$x_g := \sqrt[N]{x_1 \cdot x_2 \cdots x_N}, \quad (\text{media geométrica}) \quad (2.4)$$

$$\frac{1}{x_a} := \frac{1}{N} \sum_{i=1}^N \frac{1}{x_i}, \quad (\text{media armónica}). \quad (2.5)$$

Otra medida de tendencia central es la **mediana**:

$$\bar{x} := \begin{cases} x_{[(N+1)/2]}, & N \text{ impar} \\ (x_{[N/2]} + x_{[N/2+1]})/2, & N \text{ par} \end{cases} \quad (2.6)$$

$$s := \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^N (x_i - \bar{x})^2}, \quad (\text{error estándar}). \quad (2.7)$$

### 2.1. precisión, exactitud y sensibilidad

- **Precisión:** Se refiere al número de cifras significativas que representan una cantidad o la extensión en lecturas repetidas de un instrumento.
- **Exactitud:** Se refiere a la aproximación de una medida al valor verdadero.
- **Sensibilidad:** Es la mínima magnitud que puede diferenciar el instrumento de medición.

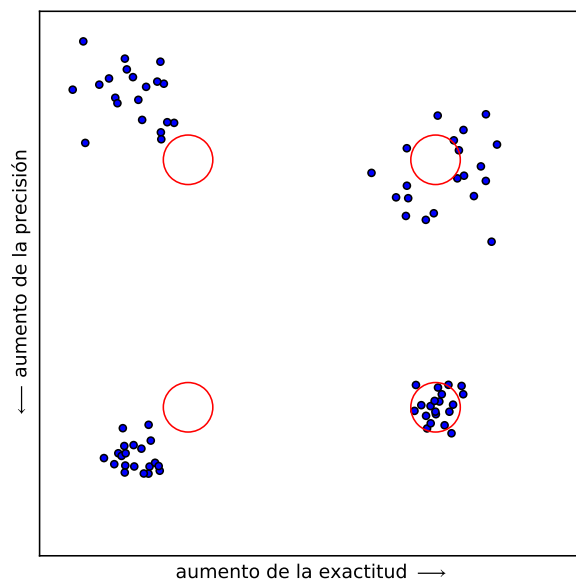


Figura 2.1: Exactitud versus precisión. Código Python en apéndice ??.

## 2.2. Cifras significativas

El número de cifras significativas de un número, lo obtenemos mirando los dígitos de dicho número, de izquierda a derecha y comenzando a contar desde el primer dígito distinto de cero hasta el último dígito, sea o no cero. La mencionada cantidad de dígitos nos entrega el número de cifras significativas.

Ejemplos:

2	(1 cifra significativa)	
32	(2 cifras significativas)	
12,470	(5 cifras significativas)	
12,0010	(6 cifras significativas)	(2.8)
0,0023	(2 cifras significativas)	
000156,210	(6 cifras significativas)	
$2,3 \times 10^5$	(2 cifras significativas)	

### 2.2.1. Ejercicio

Un(a) estudiante mide el largo de una mesa y como resultado nos entrega la siguiente cantidad  $(2,13 \pm 0,05)$  m. Este(a) estudiante, por conveniencia, decide expresar su resultado en milímetros. ¿Cómo tendría que expresar su resultado?, ¿ $(2130 \pm 50)$  mm?

**Respuesta:**

Si ésta fuese la respuesta, bastaría con cambiar de unidades para mejorar la sensibilidad de nuestro instrumento.

La forma correcta de expresar el resultado es:  $(2,13 \pm 0,05) \times 10^3$  mm.

## 2.3. número de cifras significativas para expresar resultados

1. Suma y sustracción: El resultado se expresa con tantos decimales como el número de menor cantidad de decimales que participa de la operación.

2. Multiplicación y división: El resultado se expresa con tantas cifras significativas como la magnitud que tenga menor número de cifras significativas.

## 2.4. Aproximación y Redondeo

Si queremos expresar el número  $\pi = 3,14159265358979 \dots$  con  $n$  cifras significativas, debemos desechar todas las cifras que se encuentren a la derecha del  $n$ -ésimo lugar.

### Reglas

1. Si la cifra que se encuentra en el lugar  $(n + 1)$  es mayor que 5, se le agrega una unidad a la cifra que se encuentra en el lugar  $n$ -ésimo.
2. Si la cifra que se encuentra en el lugar  $(n + 1)$  es menor que 5, dejamos la cifra  $n$ -ésima inalterable.
3. Si la cifra que se encuentra en el lugar  $(n + 1)$  es igual a 5, adoptaremos el siguiente criterio:
  - Si la  $n$ -ésima cifra es impar, se le agrega una unidad.
  - Si la  $n$ -ésima cifra es par la dejamos inalterable.

Nota 1: este criterio es totalmente arbitrario y para asegurarnos que realmente es al azar el que la cifra  $n$ -ésima aumente o quede inalterable, debemos cerciorarnos que la posibilidad que sea par es la misma que sea impar. Para estos efectos consideraremos al cero como número par.

Nota 2: este criterio de redondeo es aplicable a la magnitud que representa el resultado, ya que para el error siempre se redondea a la cifra superior.

## 2.5. Orden de magnitud

Es la potencia de diez más cercana. Ejemplos:

- 6,7 es de orden  $10^1$  (está más cerca de 10 que de 1).
- 5,3 es de orden  $10^0$  (está más cerca de 1 que de 10).
- 128,9 es de orden  $10^2$  (está más cerca de 100 que de 1000).

## 2.6. Tipos de errores

- **Error aleatorio:** Al repetir una medida varias veces, se obtienen diferentes resultados que están relacionados a la precisión de los datos obtenidos. Esta dispersión se debe a la presencia del error aleatorio.
- **Error sistemático:** Este tipo de error está relacionado con la exactitud del resultado, a mayor exactitud menor es el error sistemático. Las fuentes de este tipo de error podrían ser la mala calibración del instrumento de medición o un método que siempre sobre estime o sub estime un valor medido.
- **Equivocación:** Este tipo de error es cuando en el conjunto de medidas se encuentran algunos datos malos y no tienen relación con el experimento en sí. Un ejemplo de equivocación podría ser un error de escritura al transcribir los datos.

## 2.7. Representación grafica de error aleatorio y sistemático

Supongamos que el valor exacto de una variable es 10. Representemos un conjunto de medidas para los siguientes casos:

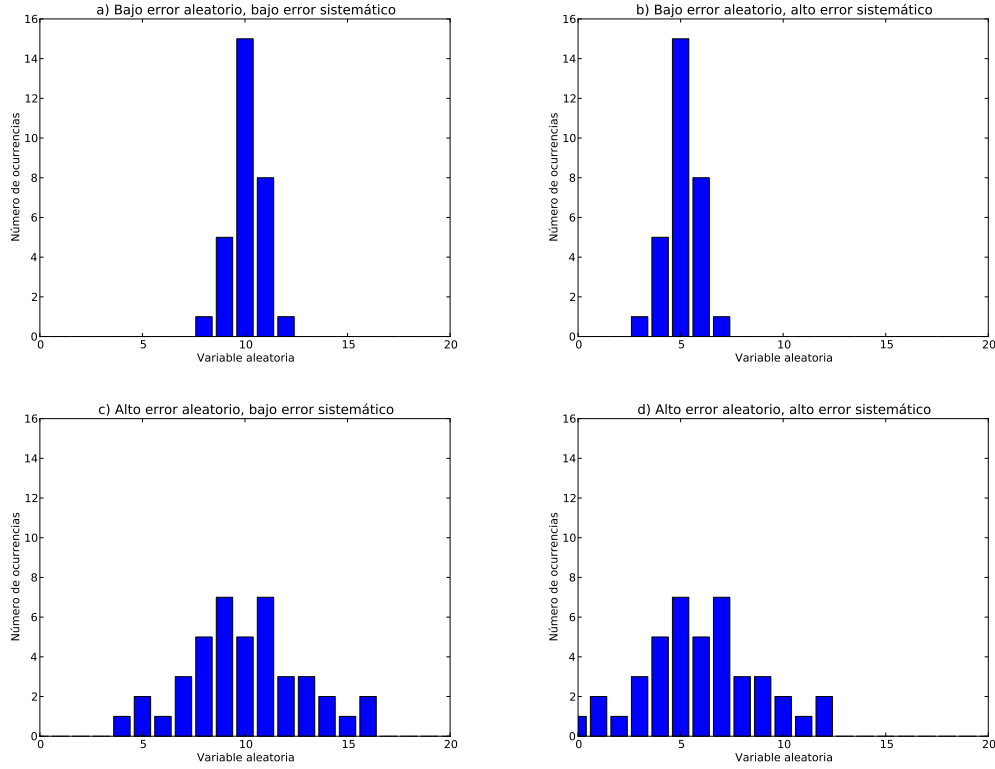


Figura 2.2: Distintas posibilidades para el tamaño del error aleatorio y sistemático. Código Python en apéndice ??.

## 2.8. Diferentes formas de expresar el error aleatorio

Si nuestro resultado está expresado como:

$$(\bar{x} \pm \Delta x) \text{ unidades} \quad (2.9)$$

Error absoluto:  $\Delta x$ .

Error relativo:  $\Delta x / \bar{x}$ .

Error porcentual:  $(\Delta x / \bar{x}) \times 100 \%$

Nota: El error absoluto habitualmente se expresa con *una cifra significativa* y tanto el error relativo como porcentual, generalmente se expresan con *2 cifras significativas*.

## 2.9. Propagación de incertezas en las operaciones

La propagación de incertezas consiste en conocer el error en la magnitud derivada a partir de los errores de las magnitudes medidas directamente.

Supongamos que la magnitud del valor que queremos hallar, depende de las variables  $x^1, x^2, \dots, x^m$  mediante la relación funcional  $y = f(x^\alpha)$ ,  $\alpha = 1, \dots, m$ . Luego de realizar  $N$  medidas para cada una de las  $m$  variables, tenemos

$$\begin{array}{cccccc} x_1^1 & x_2^1 & \cdots & x_N^1 & ; & \bar{x}^1 \\ x_1^2 & x_2^2 & \cdots & x_N^2 & ; & \bar{x}^2 \\ \vdots & \vdots & \vdots & \vdots & ; & \vdots \\ x_1^m & x_2^m & \cdots & x_N^m & ; & \bar{x}^m. \end{array}$$

Escribimos las desviaciones respecto al promedio como  $\Delta x_i^1, \Delta x_i^2, \dots, \Delta x_i^m$ , donde

$$\Delta x_i^1 := \bar{x}^1 - x_i^1, \quad \Delta x_i^2 := \bar{x}^2 - x_i^2, \quad (2.10)$$

etc.

Las magnitudes medidas indirectamente, estarían dadas por cada uno de los conjuntos de medidas mediante la relación funcional:

$$y_1 = f(x_1^1, x_1^2, \dots, x_1^m), \quad (2.11)$$

$$\vdots = \vdots \quad (2.12)$$

$$y_i = f(x_i^1, x_i^2, \dots, x_i^m), \quad (2.13)$$

$$\vdots = \vdots, \quad (2.14)$$

$$y_N = f(x_N^1, x_N^2, \dots, x_N^m). \quad (2.15)$$

Desarrollando en serie de Taylor a primer orden entorno de los valores medios, podemos escribir:

$$y_i = f(x_i^1, x_i^2, \dots, x_i^m) \quad (2.16)$$

$$\approx f(\bar{x}^1, \bar{x}^2, \dots, \bar{x}^m) + \frac{\partial f}{\partial x^1} \Big|_{\bar{x}} (x_i^1 - \bar{x}^1) + \dots + \frac{\partial f}{\partial x^m} \Big|_{\bar{x}} (x_i^m - \bar{x}^m). \quad (2.17)$$

Los errores asociados a cada conjunto de mediciones se pueden expresar como

$$\Delta y_i := y_i - f(\bar{x}^1, \dots, \bar{x}^m) \quad (2.18)$$

$$\approx \frac{\partial f}{\partial x^1} \Big|_{\bar{x}} (x_i^1 - \bar{x}^1) + \dots + \frac{\partial f}{\partial x^m} \Big|_{\bar{x}} (x_i^m - \bar{x}^m). \quad (2.19)$$

Calculando el error cuadrático medio

$$\Delta y := \sqrt{\sum_{i=1}^N \frac{(\Delta y_i)^2}{N}} \quad (2.20)$$

$$\begin{aligned} &\approx \frac{1}{\sqrt{N}} \left[ \sum_{i=1}^N \left[ \left( \frac{\partial f}{\partial x^1} \Big|_{\bar{x}} \right)^2 (x_i^1 - \bar{x}^1)^2 + \dots + \left( \frac{\partial f}{\partial x^m} \Big|_{\bar{x}} \right)^2 (x_i^m - \bar{x}^m)^2 \right. \right. \\ &\quad + 2 \left( \frac{\partial f}{\partial x^1} \Big|_{\bar{x}} \right) \left( \frac{\partial f}{\partial x^2} \Big|_{\bar{x}} \right) (x_i^1 - \bar{x}^1)(x_i^2 - \bar{x}^2) + \dots \\ &\quad \left. \left. + 2 \left( \frac{\partial f}{\partial x^{m-1}} \Big|_{\bar{x}} \right) \left( \frac{\partial f}{\partial x^m} \Big|_{\bar{x}} \right) (x_i^{m-1} - \bar{x}^{m-1})(x_i^m - \bar{x}^m) \right] \right]^{1/2}, \end{aligned} \quad (2.21)$$

donde  $\bar{x} = (\bar{x}^1, \dots, \bar{x}^m)$ .

Separamos esta última expresión en dos casos.

- Caso 1. Cuando las variables medidas están correlacionadas. Para este caso, el cálculo del error cuadrático medio está dado por la raíz cuadrada de la expresión que se obtuvo.
- Caso 2. Cuando las variables medidas no están correlacionadas. En este caso, la suma de todos los términos cruzados son cero (o despreciable!), por lo que la expresión se reduce a

$$\Delta y \approx \sqrt{\frac{1}{N} \sum_{i=1}^N \left[ \left( \frac{\partial f}{\partial x^1} \Big|_{\bar{x}} \right)^2 (x_i^1 - \bar{x}^1)^2 + \dots + \left( \frac{\partial f}{\partial x^m} \Big|_{\bar{x}} \right)^2 (x_i^m - \bar{x}^m)^2 \right]} \quad (2.22)$$

Una manera muy simple de visualizar que los términos cruzados no contribuyen al error cuadrático medio, consiste en observar la figura 2.9, donde en cada eje se graficó una variable distinta. Las dos imágenes superiores, de la misma figura, son casos sin correlación entre las variables medidas y las dos figuras inferiores tienen una clara correlación.

Luego, dividimos cada una de estas figuras en 4 cuadrantes. En todas las figuras el término cruzado  $\sum_i (x_i^\alpha - \bar{x}^\alpha)(x_i^\beta - \bar{x}^\beta)$  (con  $\alpha \neq \beta$ ) es positivo en los cuadrantes I y III y negativo en los cuadrantes II y IV.

Para las dos figuras superiores, el número de medidas en cada uno de los cuadrantes son iguales, por lo que la suma total de los cuatro cuadrantes será igual a cero. Para las dos figuras inferiores, se ve claramente una correlación, es decir, si una de las variables crece la otra tiene una tendencia definida, dejando en evidencia que si hay dependencia entre estas y al dividir en

4 cuadrantes el número de medidas en los cuadrantes I y III no son iguales a las medidas de los cuadrantes II y IV, tal que la suma total no es cero.

Supongamos que el valor exacto es 10 y representemos un conjunto de medidas para los siguientes casos:

El error de  $y$  cuando se conoce el error de cada  $x_i$  viene dado por:

$$\Delta y \approx \sqrt{\sum_{i=1}^N \left( \frac{\partial f}{\partial x_i} \Big|_{\bar{x}} \Delta x_i \right)^2} = \sqrt{\left( \frac{\partial f}{\partial x_1} \Big|_{\bar{x}} \Delta x_1 \right)^2 + \left( \frac{\partial f}{\partial x_2} \Big|_{\bar{x}} \Delta x_2 \right)^2 + \dots} \quad (2.23)$$

Donde  $\bar{x} = (\bar{x}_1, \dots, \bar{x}_N)$  son los valores medios y  $\Delta x_1, \dots, \Delta x_N$  son los errores asociados a cada una de las magnitudes medidas.

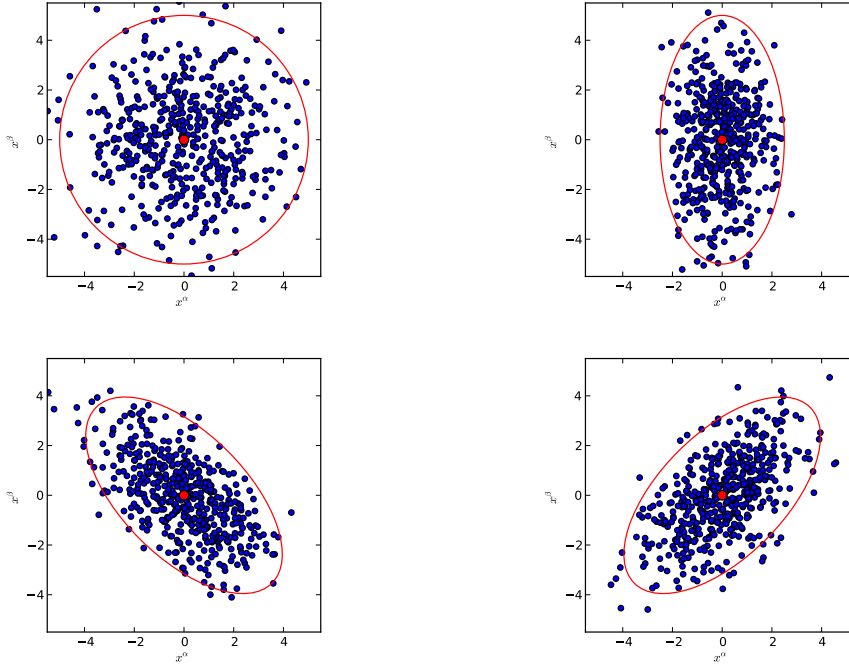


Figura 2.3: Distribuciones normales bivariadas. Código Python en apéndice ??.

Un ejemplo importante y frecuente en el laboratorio sobre las medidas indirectas es el siguiente:

Supongamos que queremos medir el periodo de un oscilador y, medimos el tiempo que tarda en hacer 10 oscilaciones, con un cronómetro de sensibilidad 0,1 s. Si el tiempo total es de 4,6 s, tendremos que el periodo medio es  $P = 0,46$  s.

$$P = \frac{t}{10} = \frac{4,6 \text{ s}}{10} = 0,46 \text{ s}, \quad (2.24)$$

$$\Delta P = \frac{0,1 \text{ s}}{10} = 0,01 \text{ s}, \quad (2.25)$$

$$P = (0,46 \pm 0,01) \text{ s}. \quad (2.26)$$

Para el ejemplo del péndulo:  $y = P(t) = t/N$

$$\Delta y = \sqrt{\left( \frac{\partial P}{\partial t} \Delta t \right)^2} = \sqrt{\left( \frac{1}{N} \Delta t \right)^2} = \frac{1}{N} \Delta t = \frac{1}{10} 0,1 = 0,01. \quad (2.27)$$

Otro ejemplo más ilustrativo es: Supóngase que determinamos el volumen de un gas a partir de la ecuación para un gas ideal,

$$V(p, T, n, R) = \frac{nRT}{p}, \quad (2.28)$$

donde cada magnitud física es medida con su respectivo error asociado:

$$p = (100 \pm 3) \times 10^3 \text{ Pa}, \quad T = (300,0 \pm 0,2) \text{ K}, \quad n = (0,10 \pm 0,01) \text{ mol}, \quad (2.29)$$

mientras que de una tabla de datos se obtiene

$$R = (8,31 \pm 0,01) \text{ J/mol K}. \quad (2.30)$$

El volumen medido es:

$$V(p, T, n, R) = 0,002493 \text{ m}^3. \quad (2.31)$$

Ahora evaluaremos el error asociado con el volumen,  $V$ , usando la ecuación (2.23):

$$\Delta V = \sqrt{\left(\left.\frac{\partial V}{\partial p}\right|_{\bar{x}} \Delta p\right)^2 + \left(\left.\frac{\partial V}{\partial T}\right|_{\bar{x}} \Delta T\right)^2 + \left(\left.\frac{\partial V}{\partial n}\right|_{\bar{x}} \Delta n\right)^2 + \left(\left.\frac{\partial V}{\partial R}\right|_{\bar{x}} \Delta R\right)^2}, \quad (2.32)$$

donde  $\bar{x} = (\bar{p}, \bar{T}, \bar{n}, \bar{R})$ .

$$\Delta V = \sqrt{\left(\frac{\bar{n}\bar{R}\bar{T}}{\bar{p}^2} \Delta p\right)^2 + \left(\frac{\bar{n}\bar{R}}{\bar{p}} \Delta T\right)^2 + \left(\frac{\bar{R}\bar{T}}{\bar{p}} \Delta n\right)^2 + \left(\frac{\bar{n}\bar{T}}{\bar{p}} \Delta R\right)^2} \quad (2.33)$$

$$\approx 25 \times 10^{-5} \text{ m}^3 \quad (2.34)$$

Por lo tanto,  $V \pm \Delta V \approx (0,002493 \pm 0,00025) \text{ m}^3 = (0,0025 \pm 0,0003) \text{ m}^3$ .

Para completar esta información daremos algunas de las expresiones que indican como debe entregarse el resultado de una medición cuando esta proviene de diferentes operaciones. La forma en que se propagan las incertezas cuando el resultado proviene de una suma, resta, multiplicación o división de mediciones.

$$(a \pm \Delta a) + (b \pm \Delta b) = (a + b) \pm (\Delta a + \Delta b), \quad (2.35)$$

$$(a \pm \Delta a) - (b \pm \Delta b) = (a - b) \pm (\Delta a + \Delta b), \quad (2.36)$$

$$(a \pm \Delta a)(b \pm \Delta b) \approx ab \pm (|a|\Delta b + |b|\Delta a), \quad (2.37)$$

$$\frac{(a \pm \Delta a)}{(b \pm \Delta b)} \approx \frac{a}{b} \pm \frac{a}{b} \left( \frac{\Delta a}{a} + \frac{\Delta b}{b} \right). \quad (2.38)$$

Note que las expresiones que hemos mostrado aquí son en verdad una primera aproximación a la forma de entregar un resultado experimental, es decir, son válidas sólo para el caso donde el error es mucho menor que la magnitud medida.

**Ejemplo:** Calculemos el área de un triángulo midiendo diferentes lados y diferentes alturas. Supongamos que disponemos de una regla graduada en milímetros y que los diferentes lados y sus correspondientes alturas están dadas por:

$$a = (16,09 \pm 0,02) \text{ cm}, \quad h_a = (10,66 \pm 0,03) \text{ cm}, \quad (2.39)$$

$$b = (10,81 \pm 0,02) \text{ cm}, \quad h_b = (15,84 \pm 0,03) \text{ cm}, \quad (2.40)$$

$$c = (17,78 \pm 0,02) \text{ cm}, \quad h_c = (9,66 \pm 0,03) \text{ cm}. \quad (2.41)$$

Para calcular el área del triángulo rectángulo debemos utilizar la expresión:

$$S = \frac{\text{base} \times \text{altura}}{2}. \quad (2.42)$$

En general la podemos escribir de la siguiente forma:

$$S = \frac{(b \pm \Delta b)(h \pm \Delta h)}{2} = \frac{bh \pm (h\Delta b + b\Delta h)}{2}, \quad (2.43)$$

$$S_a = (85,7597 \pm 0,3478) \text{ cm}^2 \Rightarrow S_a = (85,8 \pm 0,4) \text{ cm}^2, \quad (2.44)$$

$$S_b = (85,6152 \pm 0,345) \text{ cm}^2 \Rightarrow S_b = (85,6 \pm 0,4) \text{ cm}^2, \quad (2.45)$$

$$S_c = (85,8774 \pm 0,3227) \text{ cm}^2 \Rightarrow S_c = (85,9 \pm 0,4) \text{ cm}^2. \quad (2.46)$$

Los valores son iguales dentro de las imprecisiones experimentales. ¿Qué valor representa mejor el área del triángulo?

Utilizamos el promedio aritmético para encontrar el valor más probable

$$S = \bar{S} \pm \Delta S, \quad (2.47)$$

$$\bar{S} = \frac{1}{N} \sum_{i=1}^N S_i = \frac{1}{3} \sum_{i=1}^N S_i = 85,8, \quad (2.48)$$

$$\Delta S = \sqrt{\frac{1}{N} \sum_{i=1}^N (S_i - \bar{S})^2} = \sqrt{\frac{1}{3} (0^2 + 0,2^2 + 0,1^2)} = 0,2. \quad (2.49)$$

Finalmente el resultado queda expresado como:

$$\bar{S} = (85,8 \pm 0,4) \text{ cm}^2, \quad (2.50)$$

donde 0,4 es el mayor entre el error cuadrático medio y la propagación del error asociado a cada área.



## Capítulo 3

# Noción de Probabilidades

### 3.1. Introducción

Si se mide una variable física, lo que estamos realizando es un experimento. Al intentar repetir una misma medición, los resultados obtenidos no son exactamente iguales debido a pequeñas variaciones en las variables que no están totalmente controladas. Otra posible fuente de error, podría ser el instrumento de medición, ya que éste puede sufrir pequeñas variaciones para una misma medición. En consecuencia, se dice que dicho experimento tiene una componente aleatoria.

En algunos casos, las variaciones son tan pequeñas que podrían ignorarse. Sin embargo, la variación casi siempre está presente y su magnitud puede ser muy importante para obtener alguna conclusión.

Sin importar el cuidado al realizar un experimento, siempre se tendrán variaciones y la idea es comprender, cuantificar y modelar el tipo de variaciones. Cuando se incorporan estas variaciones en el análisis de los resultados, siempre se podrán obtener conclusiones de los resultados que no sean invalidados por las variaciones.

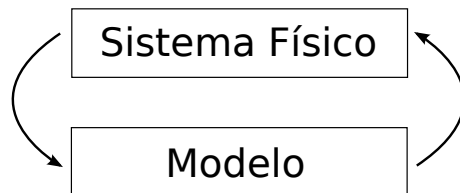


Figura 3.1: Sistema físico y modelo.

Un esquema idealizado de experimento podría ser



Figura 3.2: Entrada, sistema y salida.

Siendo la segunda alternativa la que describiría un experimento más real.

A menudo se requiere incorporar en el modelo las variaciones.

Por otro lado, es difícil especular sobre la magnitud de las variaciones sin mediciones empíricas. No obstante, con un número suficiente de éstas, es posible considerar el efecto que tienen sobre el desempeño de nuestro experimento (dispositivos como instrumentos de medición).

#### 3.1.1. Experimento aleatorio

Un experimento aleatorio es aquel que proporciona diferentes resultados aún cuando se repita siempre de la “misma manera”.

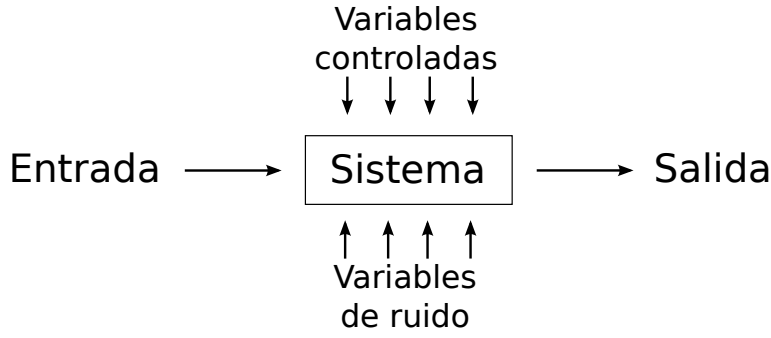


Figura 3.3: Entrada, sistema, salida y variables.

### 3.1.2. Población

La población es un conjunto de elementos (individuos, objetos, etc.) con alguna característica común observable. A los elementos que conforman la población se les llama **unidad observable** o **unidad de observación**. Cuando se posee información de todas las unidades observables de la población se está en presencia de un **censo**.

### 3.1.3. Muestra

Parte o porción extraída de una población por métodos que permitan considerarla como representativa de la misma.

### 3.1.4. Espacio muestral

El conjunto de los posibles resultados de un experimento aleatorio recibe el nombre de **espacio muestral** del experimento. El espacio muestral se denotará con la letra  $S$ .

Un espacio muestral es **discreto** si está formado por un conjunto de resultados contables.

### 3.1.5. Evento

Un evento es un subconjunto de un espacio muestral de un experimento aleatorio.

Dos eventos,  $E_1$  y  $E_2$ , tales que  $E_1 \cap E_2 = \emptyset$ , se dice que son mutuamente excluyentes.

**Ejemplo:** Considere un experimento en el cual se miden dos variables y éstas pueden tomar dos valores,  $X$  e  $Y$ . Las posibles soluciones a dicho experimento serán:  $S = \{XX, XY, YX, YY\}$  y posibles eventos serán  $E_1 = \{XX\}$ ,  $E_2 = \{XY, YX\}$ ,  $E_3 = \{XX, XY, YX\}$ ,  $E_4 = \{XY, YX, YY\}$ ,  $E_5 = \{YY\}$ .

**Ejemplo:** Supongamos que tenemos tres objetos  $(1, 2, 3)$  y sacamos un par de éstos. Todos los posibles resultados serán:

$$S_1 = \{12, 13, 21, 23, 31, 32\}, \quad (3.1)$$

$$S_2 = \{11, 12, 13, 21, 22, 23, 31, 32, 33\}, \quad (3.2)$$

dependiendo si una vez sacado cada objeto, éste es o no remplazado.

## 3.2. Interpretación de la Probabilidad

Con frecuencia es útil cuantificar la posibilidad que se presente un resultado de un experimento aleatorio. La posibilidad de un resultado se cuantifica asignándole un número del intervalo  $[0, 1]$  o un porcentaje entre 0 y 100 %. Mientras más grande sea el número, mayor será la probabilidad de obtener ese resultado.

Una posible interpretación de la probabilidad se basa en el modelo de la repetición del experimento aleatorio. Sea  $n(A)$  el número de veces que ocurre el suceso y  $n$  el número de veces que se realiza el experimento. Es intuitivamente razonable tomar como valor de  $P(A)$  a:

$$P(A) = f(A) = n(A)/n \quad (3.3)$$

como el valor límite de la proporción de veces que el resultado aparece en  $n$  repeticiones del experimento aleatorio, a medida que  $n$  crece sin cota alguna.

Cada vez que un espacio muestral está formado por  $N$  posibles resultados igualmente probables, la probabilidad de cada uno de ellos será  $1/N$ .

Para un espacio muestral discreto, la probabilidad de un evento  $E$ , denotada como  $P(E)$ , es igual a la suma de las probabilidades de los resultados en  $E$ .

### 3.2.1. Axiomas de probabilidad

La probabilidad es un número que se asigna a cada miembro de una colección de eventos de un experimento aleatorio y que satisface las siguientes propiedades.

Si  $S$  es el espacio muestral y  $E$  es cualquier evento del experimento aleatorio,

$$P(S) = 1, \quad 0 \leq P(E) \leq 1. \quad (3.4)$$

Si  $E_1$  y  $E_2$  son eventos excluyentes

$$E_1 \cap E_2 = \emptyset, \quad P(E_1 \cup E_2) = P(E_1) + P(E_2). \quad (3.5)$$

### 3.2.2. Consecuencias directas de los axiomas

$$P(\emptyset) = 0, \quad (3.6)$$

$$P(E') = 1 - P(E), \quad \text{si } E_1 \subseteq E_2 \Rightarrow P(E_1) \leq P(E_2). \quad (3.7)$$

**Ejemplo:** Si el espacio muestral de un experimento aleatorio es  $S = \{a, b, c, d, e\}$  con probabilidades 0,1, 0,1, 0,2, 0,4 y 0,2 respectivamente. Sean  $A = \{a, b\}$ ,  $B = \{c, d, e\}$ . Determinar:

$$P(A) = 0,1 + 0,1, \quad (3.8)$$

$$P(B) = 0,2 + 0,4 + 0,2, \quad (3.9)$$

$$P(A \cup B) = 0,1 + 0,1 + 0,2 + 0,4 + 0,2, \quad (3.10)$$

$$P(A \cap B)P(\emptyset) = 0. \quad (3.11)$$

## 3.3. Probabilidad condicional

Algunos fenómenos tienden a ocurrir en rachas, esto hace que la probabilidad que ocurra un primer evento es menor que la probabilidad que se repita el evento u ocurra inmediatamente otro asociado al primero.

**Ejemplo:** Un canal de comunicación digital tiene una tasa de error de un bit por cada mil transmitidos, es decir, la probabilidad que ocurra un error es de  $1/1000$ . Aunque los errores son muy escasos, tienden a ocurrir en rachas, por lo que una vez ocurrido un error, la probabilidad que ocurra un segundo error es mayor que  $1/1000$ .

**Definición:** La probabilidad condicional de un evento  $A$  dado un evento  $B$ , denotada por  $P(A|B)$ , es

$$P(A|B) := \frac{P(A \cap B)}{P(B)}. \quad (3.12)$$

Consideremos un caso en que todos los resultados de un experimentos aleatorio son igualmente probables. Si existe un total de  $n$  resultados

$$P(B) = \frac{(\# \text{ de resultados en } B)}{n}, \quad (3.13)$$

$$P(A \cap B) = \frac{(\# \text{ de resultados en } A \cap B)}{n}, \quad (3.14)$$

$$\frac{P(A \cap B)}{P(B)} = \frac{(\# \text{ de resultados en } A \cap B)}{(\# \text{ de resultados en } B)} = \text{frecuencias relativas}, \quad (3.15)$$

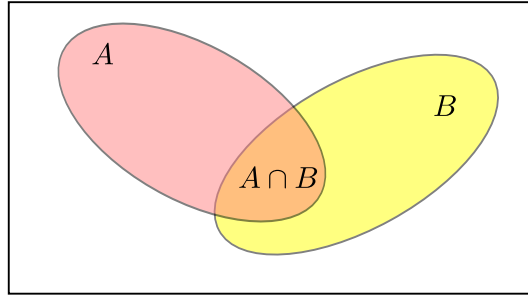


Figura 3.4: Dos conjuntos y su intersección.

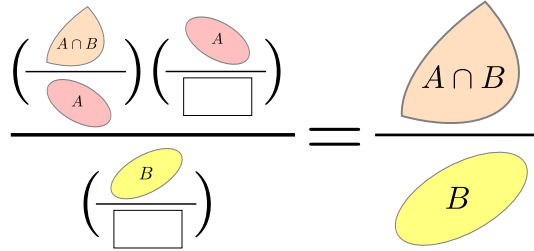


Figura 3.5: Representación diagramática de la Probabilidad condicional (3.12).

En general si son eventos independientes se tiene que

$$P(A|B) = P(A), \quad \text{o bien} \quad P(B|A) = P(B). \quad (3.16)$$

Suponga que medimos la presencia de los contaminantes  $A$  y  $B$  en muestras de vino. Medimos 266 muestras de vino, las cuales arrojan el siguiente resultado:

	$A$	Si	No
$B$			
Si	12	18	
No	24	212	

Cuadro 3.1: Contaminantes en muestras de vino.

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{12}{266}}{\frac{36}{266}} = \frac{12}{36} = \frac{1}{3}, \quad (3.17)$$

Calculemos  $P(A)$  y  $P(A|B)$  para construir un diagrama, llamado diagrama árbol, y así poder comprender mejor lo que está pasando. Para procesos independientes la probabilidad que ocurra un evento  $A$  cuando ya ha ocurrido otro evento  $B$  con anterioridad, es igual a si medimos la probabilidad que ocurra el evento  $A$  sin que hayamos medido previamente el evento  $B$ .

**Definición:** Dos eventos son independientes si, y sólo si, cualquiera de las proposiciones es verdadera.

- $P(A|B) = P(A)$ .
- $P(B|A) = P(B)$ .
- $P(A \cap B) = P(A)P(B)$ .

### 3.4. Reglas de adición

La probabilidad de un evento compuesto, como uniones, intersecciones y complementos, a menudo se pueden obtener a partir de las probabilidades de cada uno de los eventos individuales.

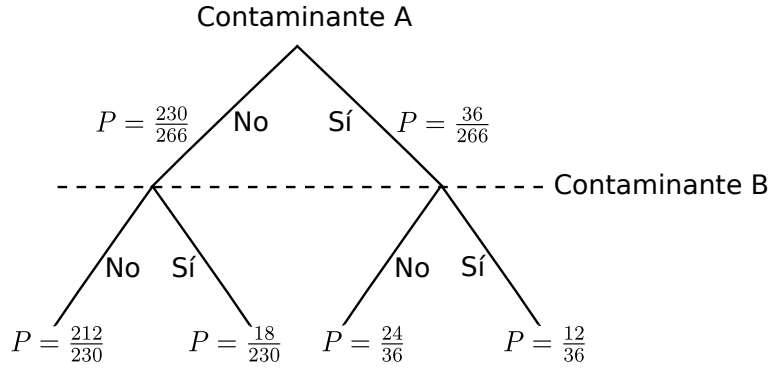


Figura 3.6: Diagrama árbol.

**Ejemplo:** Se realiza un experimento para determinar el valor de dos magnitudes físicas. El número de mediciones son 940, las que se distribuyen según la siguiente tabla:

	$Y$	$C$	$D$
$X$			
$A$		514	68
$B$		112	246

Cuadro 3.2: Distribución de las 940 mediciones.

Sean  $E_1$  y  $E_2$  los eventos definidos por:

$$E_1 = \{\text{todos los resultados donde } X = B\}, \quad (3.18)$$

$$E_2 = \{\text{todos los resultados donde } Y = D\}. \quad (3.19)$$

Calcular la probabilidad de obtener como resultado los siguientes casos:

$$X = B, \quad Y = D, \quad X = B \wedge Y = D, \quad X = B \vee Y = D, \quad X \neq B \vee Y \neq D. \quad (3.20)$$

$$P(E_1) = \frac{358}{940}, \quad P(E_2) = \frac{314}{940}, \quad P(E_1 \cap E_2) = \frac{246}{940}, \quad (3.21)$$

$$P(E_1 \cup E_2) = \frac{426}{940}, \quad P((E_1 \cup E_2)') = \frac{514}{940}, \quad P((E_1 \cap E_2)') = \frac{694}{940}. \quad (3.22)$$

Otra manera de resolver ejercicio:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) = \frac{358}{940} + \frac{314}{940} - \frac{246}{940} = \frac{426}{940}, \quad (3.23)$$

$$P((E_1 \cup E_2)') = 1 - P(E_1 \cup E_2) = 1 - \frac{426}{940} = \frac{514}{940}. \quad (3.24)$$

### 3.5. Reglas de multiplicación

Usando la definición de probabilidad condicional

$$P(A|B) := \frac{P(A \cap B)}{P(B)}, \quad P(B|A) := \frac{P(B \cap A)}{P(A)}, \quad (3.25)$$

se puede obtener

$$P(A \cap B) = P(A|B)P(B), \quad P(B \cup A) = P(B|A)P(A), \quad (3.26)$$

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) \quad (3.27)$$

Esta regla de multiplicación puede ser extendida a un número finito de sucesos  $A_i$

$$P\left(\bigcap_{i=1}^k A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\bigcap_{i=1}^{k-1} A_i). \quad (3.28)$$

### 3.6. Reglas de Probabilidad total

Cualquier evento puede escribirse como la unión de la parte de  $B$  que se encuentra en  $A$  más la parte de  $B$  que está en el complemento de  $A$ . (Ver figura)

$$B = (A \cap B) \cup (A' \cap B). \quad (3.29)$$

Note que  $(A \cap B)$  y  $(A' \cap B)$  son mutuamente excluyentes. Si calculamos la probabilidad del evento  $B$  usando la propiedad de unión y la regla de multiplicación tenemos:

$$P(B) = P(A \cap B) + P(B \cap A') = P(B|A)P(A) + P(B|A')P(A'). \quad (3.30)$$

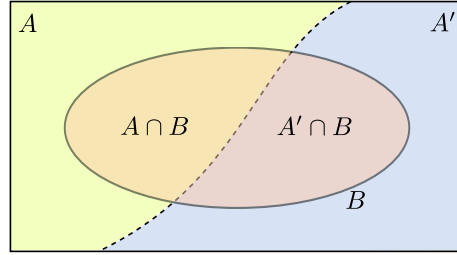


Figura 3.7: Diagrama de Venn para la probabilidad total.

### 3.7. Teorema de Bayes

Retomando la definición de probabilidad condicional y las reglas de multiplicación, tenemos:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (3.31)$$

Este resultado es de gran utilidad ya que permite determinar  $P(A|B)$  en función de  $P(B|A)$

Ejemplo: Supongamos que mediante un análisis de 84 muestras se detectan 36 muestras con Pb, 28 con As, 12 con ambos elementos pesados y 32 están libres de contaminación.

	Pb	
As	No	Sí
No	32	24
Sí	16	12

Cuadro 3.3: Pb y As en 84 muestras de vino.

Calcular  $P(\text{Pb}|\text{As})$

Usando el teorema de Bayes tenemos que

$$P(\text{Pb}|\text{As}) = \frac{P(\text{As}|\text{Pb})P(\text{Pb})}{P(\text{As})}, \quad (3.32)$$

$$P(\text{Pb}|\text{As}) = \frac{\frac{12}{36} \frac{36}{84}}{\frac{28}{84}} = \frac{12}{28} = \frac{3}{7}. \quad (3.33)$$

Si se conocen los siguientes datos  $P(\text{As}|\text{Pb}) = 0,33$ ,  $P(\text{Pb}) = 0,43$ ,  $P(\text{As}) = 0,33$ , calcular  $P(\text{Pb}|\text{As})$ . Usando el teorema de Bayes el cálculo es muy simple,

$$P(\text{Pb}|\text{As}) = \frac{0,33 \times 0,43}{0,33} = 0,43. \quad (3.34)$$

## Capítulo 4

# Distribuciones de Probabilidad

### 4.1. Fundamentos matemáticos

Cuando uno realiza un experimento aleatorio, generalmente obtiene un conjunto de puntos discretos (Muestra). Sin embargo, a veces lo que se busca es determinar una dependencia funcional entre dichos puntos, para lo cual son de gran utilidad los conceptos de **media**, **desviación estándar**, **varianza**, **error estándar**, **suma residual de los cuadrados** y la distribución de dichos puntos respecto del valor medio.

Éstos son conceptos útiles para definir de una forma cuantitativa la confiabilidad que se le puede atribuir a un cierto resultado al experimento aleatorio. La medida estadística más común es la **media**, definida como

$$\bar{x} := \frac{1}{N} \sum_{i=1}^N x_i, \quad (4.1)$$

donde  $x_i$  representa los datos individuales y  $N$  el número total de éstos. Equivalentemente, podemos escribir

$$\bar{x} := \frac{1}{N} \sum_{a=1}^{\hat{d}} n_a x_a, \quad (4.2)$$

donde  $n_a$  es el número de veces que se repiten el valor  $x_a$  en la muestra y  $\hat{d}$  es el número de valores distintos de  $x_a$  ( $\sum_{a=1}^{\hat{d}} n_a = N$ ).

La medida más común de la variabilidad de un conjunto de datos, es la **desviación estándar** respecto a la media:

$$s_x := \sqrt{\frac{S_t}{N-1}}, \quad (4.3)$$

donde  $S_t$  es la suma de los cuadrados de las diferencias entre cada dato y la media de éstos, definida por:

$$s_t := \sum_{i=1}^N (x_i - \bar{x})^2. \quad (4.4)$$

Otro estimador de la variabilidad, el **error estándar**, es definido por:

$$\alpha_x := \frac{s_x}{\sqrt{N}}. \quad (4.5)$$

Si la dispersión de los datos respecto a la media aumenta,  $s_t$ ,  $s_x$  y  $\alpha_x$  crecen. Por otro lado, si las medidas individuales están muy cerca del valor medio, estas cantidades serán menores. La dispersión también puede representarse por una magnitud llamada **varianza**, que por definición es el cuadrado de la desviación estándar,

$$s_x^2 = \frac{S_t}{N-1}. \quad (4.6)$$

Para ambos casos el denominador es  $N-1$ . Formalmente decimos que se pierde un grado de libertad. Esto quiere decir que se ha considerado el promedio, calculado previamente a partir de los datos de la muestra, para determinar el valor de  $s_t$ . Una justificación más intuitiva es que la dispersión no está definida en caso de tener sólo un dato (en ese caso se obtiene 0/0).

Otra medida estadística que es útil para cuantificar la dispersión es el **coeficiente de variación**, definido por:

$$c.v. := \frac{s_x}{\bar{x}} \times 100 \%. \quad (4.7)$$

## 4.2. Variables aleatorias, histogramas y distribuciones

### 4.2.1. Variable aleatoria

**Definición:** Sea  $S = \{s_1, s_2, \dots\}$  un espacio muestral, cuyos elementos  $s_a$  ( $a = 1, \dots, D$ ) representan los posibles resultados distintos de un experimento aleatorio. Sea  $X$  una función de valor real definida sobre  $S$ , de manera que asocie los resultados de  $S$  a valores reales. Se dice entonces que  $X$  es una **variable aleatoria** definida sobre el espacio muestral. ( $S$  es el dominio de la variable aleatoria  $X$  y el rango de esta función es el conjunto de valores reales  $\{X(s_a)\}$ ).

**Definición:** Se dice que la variable  $X$  es **discreta** si su rango es un conjunto discreto (finito o infinito numerable) de números reales.

**Definición:** Se dice que la variable aleatoria  $X$  es **continua** si sus valores consisten al menos de un intervalo de los números reales.

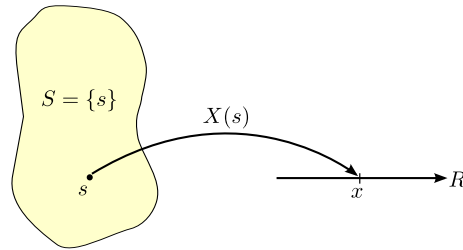


Figura 4.1: Espacio Muestral, función aleatoria y rango.

Denotaremos al conjunto de todos los *valores distintos* de  $x$  definidos por la variable aleatoria por  $\{x_1, x_2, \dots, x_d\}$  o bien  $\{x_\alpha\}$ , con  $\alpha = 1, \dots, d$ . Note que, si la función aleatoria asigna el mismo valor real a al menos dos elementos del espacio muestral, entonces  $d \leq D$ .

## 4.3. Histogramas

Luego de realizar un experimento aleatorio, se tiene un conjunto de resultados. Una de las mejores maneras de representar gráficamente la distribución de los resultados es construir un **histograma**, es decir, un gráfico de frecuencias (o número de ocurrencias) para cada valor distinto de la variable aleatoria (los  $\{x_\alpha\}$ ).

Como primer paso se divide el conjunto de valores de la variable aleatoria en un cierto número de intervalos (intervalos de clase o celdas). Luego se grafica en el eje de las ordenadas los intervalos de clase y en el eje de las abscisas el número de veces que los resultados están contenidos en cada uno de los respectivos intervalos. Otra forma de representar los resultados es mediante un gráfico de **frecuencias relativas**, que se diferencia del anterior en que en el eje de las abscisas ahora se grafica las frecuencias divididas por el número total de datos. De esta manera las frecuencias relativas son la probabilidad que la variable aleatoria tome cierto valor *en la muestra* (y que se espera que sea un buen estimador de la probabilidad de la población).

**Ejemplo:** Arrojemos dos dados y para cada lanzamiento sumemos las dos caras superiores. Luego de repetir este experimento 100.000 veces, construyamos un histograma y un gráfico de frecuencias relativas a partir de los datos experimentales. En este caso el gráfico de frecuencias relativas nos permite estimar cómo es la distribución de probabilidades de la variable aleatoria.



$x_\alpha$	2	3	4	5	6	7	8	9	10	11	12
$n_\alpha$	2703	5526	8320	11205	13685	16778	13870	11173	8517	5511	2712

Cuadro 4.1: Número de ocurrencias de cada valor de la suma de dos dados. Experimento (simulado) con  $10^5$  repeticiones.

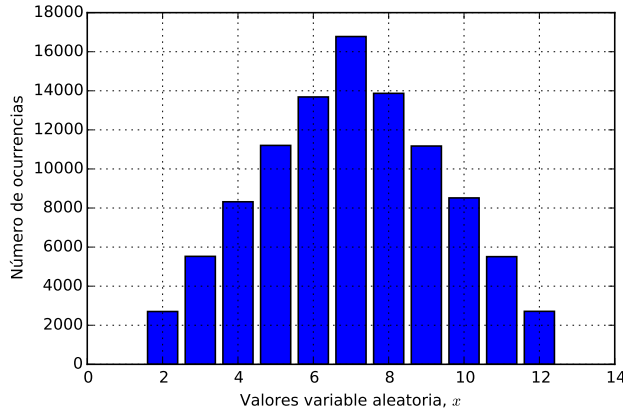


Figura 4.2: Histograma de valores de suma de dados, de acuerdo a los datos en la tabla 4.1.

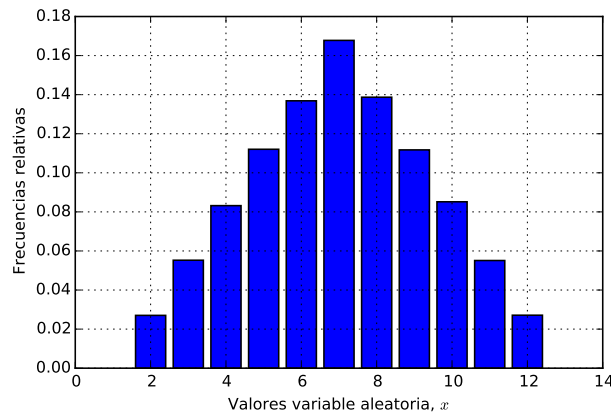


Figura 4.3: Frecuencias relativas de ocurrencia de los valores de suma de dados.

## 4.4. Distribuciones de probabilidad

Ahora consideraremos un modelo idealizado del experimento de los dados (de 6 caras, no cargados). Considere todos los posibles resultados al lanzar estos dos dados. Luego, definamos nuestra variable aleatoria como la suma de las caras de ambos dados, después de ser arrojados. Para este caso el espacio muestral está formado por los 36 posibles resultados, *que supondremos igualmente probables*, y para los cuales la variable aleatoria puede tomar 11 valores distintos. Además, existe un cierto número de combinaciones de los resultados de cada dado que generan el mismo valor de la variable aleatoria (la suma), como se muestra en la Tabla 4.2. Podemos construir un gráfico de barras del número de combinaciones en función de los valores que toma la variable aleatoria. En este ejemplo, ver figura 4.4, este gráfico muestra la distribución de combinaciones asociadas a la variable aleatoria. Más aún, normalizando este gráfico respecto al número total de combinaciones (36 en este ejemplo), encontramos un gráfico de la distribución de probabilidad asociada a cada uno de los posibles valores de la variable aleatoria. Ver fig. 4.4.

Como se desprende del ejemplo, si la variable aleatoria está claramente definida, es posible definir una función que asigne una probabilidad que dicha variable tome un determinado valor.

Espacio muestral, $S$ (resultados posibles)	Valor de la variable aleatoria, $X$	número de combinaciones
(1, 1)	2	1
(1, 2), (2, 1)	3	2
(1, 3), (2, 2), (3, 1)	4	3
(1, 4), (2, 3), (3, 2), (4, 1)	5	4
(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)	6	5
(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)	7	6
(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)	8	5
(3, 6), (4, 5), (5, 4), (6, 3)	9	4
(4, 6), (5, 5), (6, 4)	10	3
(5, 6), (6, 5)	11	2
(6, 6)	12	1

Cuadro 4.2: Número de combinaciones correspondientes a cada valor de la suma de dos dados. En este caso  $d = 11$ .

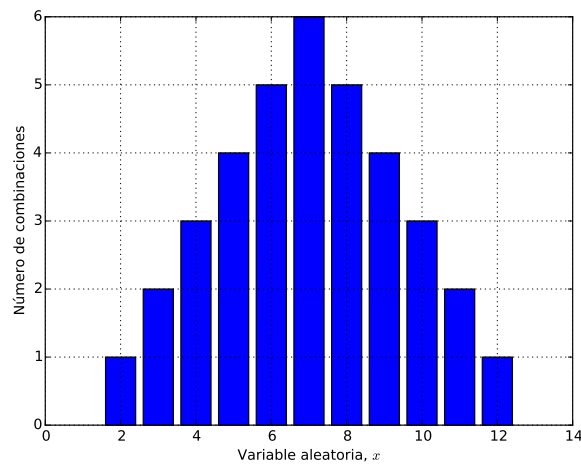


Figura 4.4: Gráfico de barras para los valores  $x$  de la variable aleatoria, de acuerdo a los datos en la tabla 4.2. Código Python en apéndice ??.

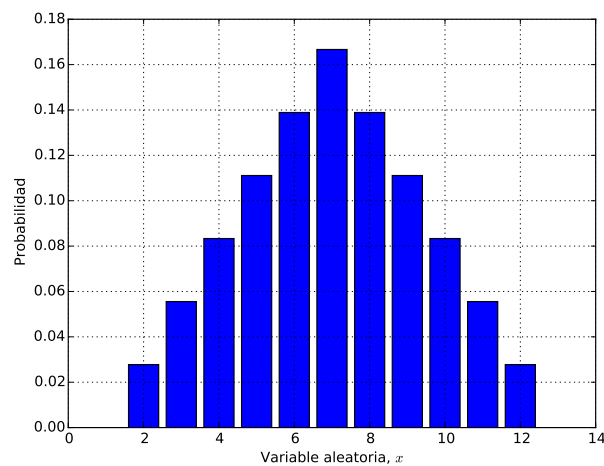


Figura 4.5: Probabilidad para las sumas de las dos caras de dos dados. Código Python en apéndice ??.

En este caso, esta probabilidad es proporcional al número de combinaciones del espacio muestral asociados a un mismo valor de la variable aleatoria. La mencionada función recibe el nombre de **función distribución de probabilidad** de la variable aleatoria  $X$ . El término más general, **distribución de probabilidad**, se refiere no sólo a la función de probabilidad, sino también a la **función de distribución acumulativa** de  $X$ , que definiremos más adelante.

**Definición:** Llamamos **evento** ( $E$ ,  $E \subseteq S$ ) a un conjunto de posibles resultados de un experimento aleatorio (Por ejemplo, el evento  $E$  puede ser el conjunto de todos los resultados tales que la variable aleatoria sea par). Denotamos por  $P(E)$  a la probabilidad de ocurrencia de este evento. (En nuestro ejemplo,  $P(E) = P(x = 2) + P(x = 4) + P(x = 6) \cdots P(x = 12)$ ).

**Definición:** La **función de probabilidades**  $p_X$  para una variable aleatoria discreta  $X$  está definida como el conjunto  $p_X = \{p_\alpha\}$  de todas las probabilidades de ocurrencia de cada valor posible  $x_\alpha$  de la variable aleatoria, con

$$p_\alpha = P(X = x_\alpha), \quad \alpha = 1, \dots, d, \quad (4.8)$$

y debe satisfacer las siguientes condiciones:

$$p_X(x_\alpha) \geq 0, \quad \alpha = 1, \dots, d, \quad (4.9)$$

$$\sum_{\alpha=1}^d p_X(x_\alpha) = 1. \quad (4.10)$$

**Definición:** La **función de distribución acumulativa** de una variable aleatoria discreta  $X$ , es la probabilidad que el valor de  $X$  sea menor o igual a un valor específico  $x$  y está dado por:

$$F(x) = P(X \leq x) = \sum_{x_\alpha \leq x} p_X(x_\alpha) = \sum_{x_\alpha \leq x} p_\alpha. \quad (4.11)$$

Por lo tanto, en el caso discreto, una variable aleatoria está caracterizada por la función de probabilidad  $p_\alpha$ , la cual determina la probabilidad puntual que  $X = x_\alpha$ , y por la función distribución acumulativa  $F_\alpha$ , la que representa la suma de las probabilidades puntuales hasta el valor  $x_\alpha$  de  $X$ , inclusive.

#### 4.4.1. Distribuciones de Poisson

**Ejemplo:** Las desintegraciones radiactivas son fenómenos físicos que puede ser modelados por una **distribución de Poisson**. Si construimos un gráfico, donde en el eje de las ordenadas representamos el número de veces que se repite un determinado valor para la variable aleatoria  $X$  ( $X$ : número de decaimientos radioactivos en un cierto intervalo de tiempo) y en el eje de las abscisas los posibles valores de la variable aleatoria, es decir, el número de decaimientos que se producen en los intervalos de clase considerados.

Dado un intervalo de números reales, supóngase que el conteo de ocurrencias es *aleatorio* en dicho intervalo. Si éste puede dividirse en sub-intervalos suficientemente pequeños, tales que:

- La probabilidad de más de una ocurrencia en el intervalo es cero.
- La probabilidad de ocurrencia en un sub-intervalo es la misma para todos los sub-intervalos, y es proporcional a la longitud de éstos.
- El conteo de ocurrencias en cada sub-intervalo es independiente del de los demás intervalos, entonces el experimento aleatorio recibe el nombre de proceso de Poisson.

**Definición:** Sea  $X$  una variable aleatoria que representa el número de eventos aleatorios independientes que ocurren a una rapidez constante sobre el tiempo o el espacio. Se dice entonces que la variable aleatoria  $X$  tiene una distribución de Poisson con función de probabilidad

$$P(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, 3, \dots, \quad \lambda > 0 \quad (4.12)$$

El parámetro  $\lambda$  de la distribución es el **número medio de sucesos esperados** (número promedio de ocurrencias del evento en el intervalo de observación, ya sea tiempo, distancia, etc.).

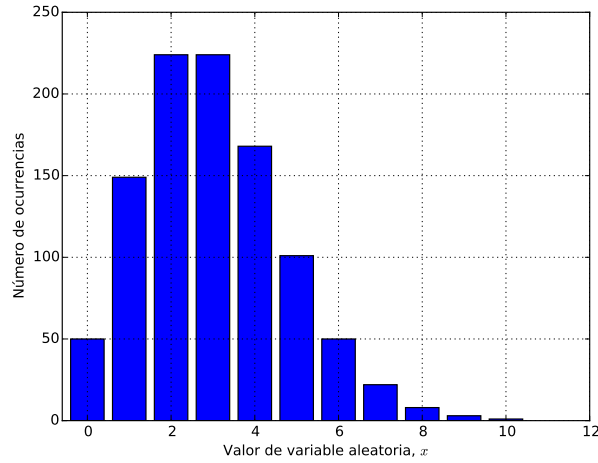


Figura 4.6: número de partículas  $\alpha$  emitidas en un intervalo de tiempo dado. Código Python en apéndice ??.

Para valores mayores que cero,  $\lambda > 0$  define una familia de distribuciones con una función de probabilidad determinada por la ecuación dada.

$$\sum_{x=0}^{\infty} P(x; \lambda) = 1, \quad (\text{condición de normalización}), \quad (4.13)$$

$$\mu_X = \sum_{x=0}^{\infty} xP(x; \lambda) = \lambda, \quad (\text{valor medio}), \quad (4.14)$$

$$\sigma^2 = \sum_{x=0}^{\infty} (x - \lambda)^2 P(x; \lambda) = \lambda, \quad (\text{varianza}), \quad (4.15)$$

$$\sigma = \sqrt{\lambda}, \quad (\text{desviación estándar}). \quad (4.16)$$

#### 4.4.2. Distribución de Gauss o Normal

**Ejemplo** Al hacer estudios del poder de frenado (Stopping Power) lo que se desea describir es la pérdida media de energía de un haz de partículas al viajar por el interior de un material. Supongamos que se hace incidir protones sobre una lámina delgada de oro. Si graficamos la energía de los protones dividida en intervalos de clase, luego que estos han atravesado la lámina de oro, en el eje de las abscisas y el número de protones correspondiente a cada intervalo en el eje de las ordenadas.

#### 4.4.3. Distribución normal

Cuando la variable aleatoria puede tomar cualquier valor dentro de un cierto intervalo de números reales, diremos que la variable aleatoria es continua.

**Definición:** Si el rango de una variable aleatoria  $X$  contiene un intervalo (finito o infinito) de números reales, entonces  $X$  es una **variable aleatoria continua**.

**Definición:** Una función  $f_X(x)$  es una **función de densidad de probabilidad** de la variable aleatoria continua  $X$  si para cualquier intervalo de números reales  $[x_1, x_2]$ :

1.  $f_X(x) \geq 0$ .
2.  $\int_{-\infty}^{\infty} f_X(x) dx = 1$ ,
3.  $P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f_X(u) du$

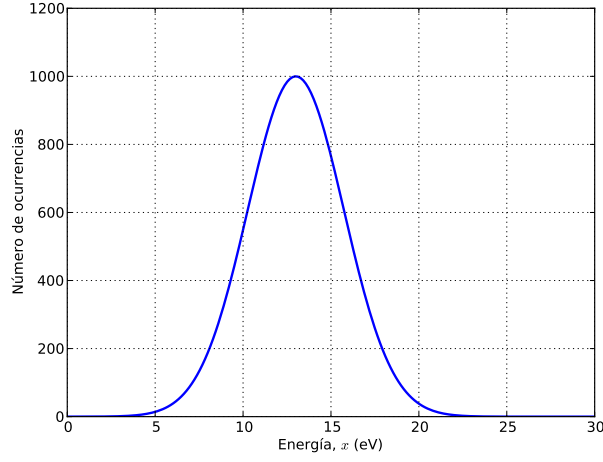


Figura 4.7: Pérdida de energía de protones luego de atravesar una lámina de oro. Código Python en apéndice ??.

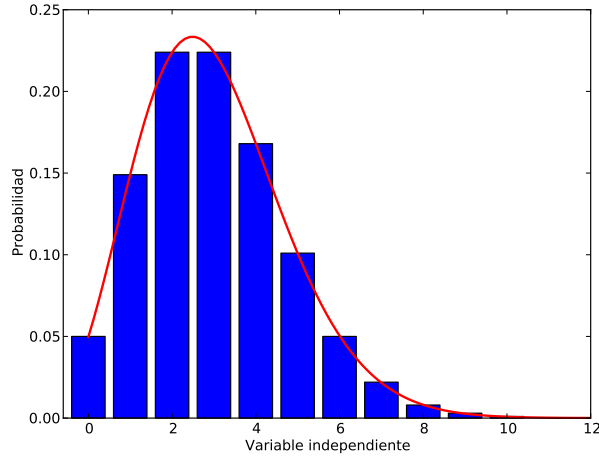


Figura 4.8: Función probabilidad de Poisson. Código Python en apéndice ??.

**Definición:** La **función distribución acumulada** de una variable aleatoria continua  $X$  es

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(u) du. \quad (4.17)$$

**Definición:** Se dice que una variable aleatoria  $X$ , se encuentra **normalmente distribuida** (Distribución de Gauss) si su función densidad de probabilidad está dada por:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty, \quad (4.18)$$

donde  $-\infty < \mu < \infty$  y  $\sigma > 0$  denotan la media y la desviación estándar de  $X$ , respectivamente.

En lo que resta de este capítulo, nos centraremos en la distribución de Gauss o distribución normal.

**Definición:** Si  $X$  es una variable aleatoria continua con función de densidad de probabilidad  $f_X(x)$ ,  $-\infty < x < \infty$ , entonces la **media** de  $X$ , está definida por

$$\mu = \int_{-\infty}^{\infty} x f_X(x) dx. \quad (4.19)$$

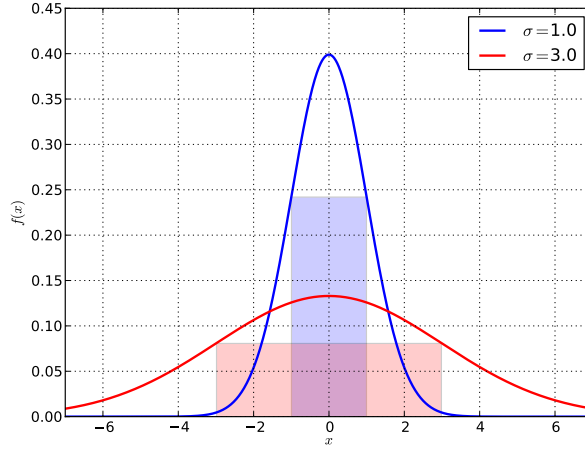


Figura 4.9: Distribución normal. Código Python en apéndice ??.

Además, la **varianza** de  $X$  es definida por

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx. \quad (4.20)$$

La **desviación estándar** es entonces

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx}. \quad (4.21)$$

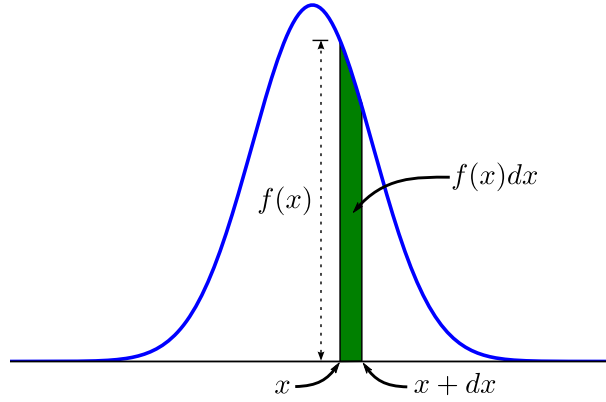


Figura 4.10: Densidad de probabilidad.

## Propiedades

La cantidad (infinitesimal)  $f(x)dx$  representa la **fracción de las medidas que se encuentran entre  $x$  y  $x + dx$** , es decir, **la probabilidad que una medida entregue como resultado un valor que se encuentre entre  $x$  y  $x + dx$** . Esta probabilidad tiene un máximo en  $x = \mu$  y es simétrica entorno al valor  $\mu$ .

Si  $X$  es una variable aleatoria normal con valor medio  $E(X) = \mu$  y varianza  $V(X) = \sigma_X^2$ , entonces la variable aleatoria

$$Z = \frac{(X - \mu)}{\sigma_X} \quad (4.22)$$

es una **variable aleatoria normal** con  $E(X) = 0$  y varianza  $V(X) = 1$ . Esto es,  $Z$  es una variable aleatoria normal estándar. La creación de una variable aleatoria con esta transformación se conoce como **estandarización**. La variable aleatoria  $Z$  representa la diferencia de  $X$  y su promedio, en unidades de desviaciones estándar.

**Utilidad de la estandarización:** Suponga que  $X$  es una variable aleatoria normal con media  $\mu$  y varianza  $V(X) = \sigma_X^2$ . Entonces,

$$P(X \leq x) = P\left(\frac{(X - \mu)}{\sigma} \leq \frac{(x - \mu)}{\sigma}\right) = P(Z \leq z), \quad (4.23)$$

donde  $Z$  es una variable aleatoria normal, y  $z := (x - \mu)/\sigma$  es el valor de  $z$  obtenido a través de la estandarización de  $X$ .

Análogamente, es posible definir la **función distribución acumulada** de una variable aleatoria normal estándar como

$$\Phi(z) = P(Z \leq z). \quad (4.24)$$

El **valor más probable**, los puntos de inflexión, la **media**, la **varianza** y la desviación estándar, usando la función distribución estandarizada  $(X - \mu)/\sigma$ , estarían dadas por

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (4.25)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}z^2} \quad (4.26)$$

$$(4.27)$$

$$\frac{df(z)}{dz} = \frac{-2z}{2\sqrt{2\pi\sigma^2}} e^{-z^2/2} = 0 \quad (4.28)$$

$$\Rightarrow z = \frac{(x - \mu)}{\sigma} = 0 \quad (4.29)$$

$$x = \mu \quad (4.30)$$

$$\frac{d^2f(z)}{dz^2} = \frac{d}{dz} \left( \frac{-z}{\sqrt{2\pi\sigma^2}} e^{-z^2/2} \right) = 0 \quad (4.31)$$

$$\Rightarrow \left( \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2} \right) (z^2 - 1) = 0 \quad (4.32)$$

$$\Rightarrow z = \pm 1 \quad (4.33)$$

$$\Rightarrow \frac{(x - \mu)}{\sigma} = \pm 1 \quad (4.34)$$

$$x = \mu \pm \sigma \quad (4.35)$$

$$E(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (4.36)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma z) e^{-z^2/2} dz \quad (4.37)$$

$$= \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz + 0 \quad (4.38)$$

$$= \frac{\mu}{\sqrt{2\pi}} (\sqrt{2\pi}) \quad (4.39)$$

$$= \mu \quad (4.40)$$

$$V(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (4.41)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^2 e^{-z^2/2} dz \quad (4.42)$$

$$= -\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \frac{d}{dz} \left( e^{-z^2/2} \right) dz \quad (4.43)$$

$$= -\frac{\sigma^2}{\sqrt{2\pi}} \left[ z e^{-z^2/2} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} e^{-z^2/2} dz \right] \quad (4.44)$$

$$= -\frac{\sigma^2}{\sqrt{2\pi}} \left[ 0 - \sqrt{2\pi} \right] \quad (4.45)$$

$$= \sigma^2. \quad (4.46)$$

$$s = \sqrt{V(X)} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (4.47)$$

$$= \sqrt{\sigma^2} = \sigma. \quad (4.48)$$

respectivamente.



## Capítulo 5

# Estimadores y método de máxima Verosimilitud

Uno de los mejores métodos para obtener un estimador puntual de un parámetro es el **método de máxima verosimilitud** y como lo dice su nombre, el estimador será el valor del parámetro que maximiza la **función de verosimilitud**, es decir, el valor que haga máxima la probabilidad de obtener la muestra observada. Para poder comprender este método comenzaremos por definir algunos conceptos.

**Definición:** **estadística o estadígrafos** es cualquier función de las observaciones contenidas en una muestra aleatoria (promedio, varianza, etc).

**Definición:** **Estimación puntual** es una estimación numérica de la estadística que se desee evaluar. Por ejemplo, el estimador de  $\mu$  (promedio de la población) de una muestra aleatoria es  $\bar{x}$  (promedio de la muestra).

### 5.1. Promedio, mediana, desviación estándar, varianza

### 5.2. Definición de máxima verosimilitud.

Supóngase que  $X$  es una variable aleatoria con distribución de probabilidad  $f(x, \theta)$ , donde  $\theta$  es un parámetro desconocido. Sean  $x_1, x_2, \dots, x_N$  los valores observados en una muestra aleatoria de tamaño  $N$ . La **función de verosimilitud** de la muestra es

$$L(\theta) := f(x_1, \theta)f(x_2, \theta) \cdots f(x_N, \theta) \quad (5.1)$$

Note que la función de verosimilitud es ahora una función del parámetro desconocido  $\theta$ . El **estimador de máxima verosimilitud** de  $\theta$  es el valor de  $\theta$  que *maximiza* la función de verosimilitud  $L(\theta)$ .

Note que para el caso de una variable aleatoria discreta, la interpretación de la función de verosimilitud  $L(\theta)$  es la probabilidad

$$L(\theta) = P(X_1 = x_1, X_2 = x_2, \dots, X_N = x_N), \quad (5.2)$$

es decir, la probabilidad de obtener los valores Muestrales  $x_1, \dots, x_N$ . Para que esta última afirmación sea válida, las probabilidades representadas por cada miembro del producto deben ser independientes. Así, el estimador de máxima verosimilitud es un estimador que maximiza la probabilidad de ocurrencia de los valores muestrales.

**Ejemplo:** Sea  $X$  una variable aleatoria con *distribución normal*, con media desconocida y varianza conocida. La función verosimilitud de una muestra de tamaño  $N$  es

$$L(\mu, \sigma) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2} \quad (5.3)$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 \right]. \quad (5.4)$$

Calculando el logaritmo natural de ambos lados, encontramos

$$\ln L(\mu) = -\frac{N}{2} \ln(2\pi\sigma^2) - (2\sigma^2)^{-1} \sum_{i=1}^N (x_i - \mu)^2. \quad (5.5)$$

Derivando respecto al parámetro desconocido  $\mu$  e igualando a cero, llegamos a

$$\frac{d \ln L(\mu)}{d\mu} = (\sigma^2)^{-1} \sum_{i=1}^N (x_i - \mu) \stackrel{!}{=} 0, \quad (5.6)$$

que tiene como solución para  $\mu$  a

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i = \bar{x}, \quad (5.7)$$

es decir, un estimador de la media  $\mu$  es el valor de la media muestral.

**Ejemplo:** Sea  $X$  una variable aleatoria con distribución normal, donde tanto la media como la varianza son desconocidas. Ya sabemos que la función verosimilitud está dada por (5.3). Calculando ahora las derivadas parciales de (5.5) respecto a  $\mu$  y  $\sigma^2$ , e igualándolas a cero, encontramos

$$\frac{\partial \ln L(\mu, \sigma)}{\partial \mu} = (\sigma^2)^{-1} \sum_{i=1}^N (x_i - \mu) \stackrel{!}{=} 0, \quad (5.8)$$

$$\frac{\partial \ln L(\mu, \sigma)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^N (x_i - \mu)^2 \stackrel{!}{=} 0, \quad (5.9)$$

que tiene por soluciones a

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i = \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2. \quad (5.10)$$

\*\*\* Ojo que éste no es exactamente  $\sigma$ , porque está dividido por  $N$ !, lo cual quiere decir que el método de Máxima verosimilitud no garantiza obtener estimadores insesgados.\*\*\*

# Capítulo 6

## Ajuste de curvas

### 6.1. Modelos de regresión

Generalmente, al realizar un experimento se obtiene como resultado un conjunto de puntos discretos y, como ya hemos mencionado, en muchas ocasiones se requiere determinar la *dependencia funcional* entre dichos puntos. En otras ocasiones se requieren puntos *entre* esos valores discretos. Para lograr algunos de estos objetivos se requiere de **técnicas de ajuste de curvas** tanto para obtener valores intermedios como para determinar la forma en que se relacionan dichas variables. Otras veces, lo que se busca es una expresión simplificada de una función muy complicada, que se ajuste bien en algún rango deseado. Para encontrar esta función simplificada, suele evaluarse la función más complicada para varios puntos y tratar estos puntos con el mismo criterio de ajuste que los datos obtenidos experimentalmente.

Las técnicas de ajuste las separaremos en dos grupos generales:

1. Cuando los datos obtenidos muestran imprecisión, es decir un grado significativo de error aleatorio (ruido) y lo que se busca es determinar la *tendencia* y no necesariamente un modelo que describa detalladamente cada variación sistemática de los datos experimentales.

La estrategia aquí es derivar una curva que represente el comportamiento general, es decir, un modelo. Dicho modelo no necesariamente interceptará cada uno de los puntos. No debemos olvidar que estamos estudiando el caso donde los datos presentan ruido y por este motivo es que no podemos considerar cada punto de manera individual, ya que éste podría ser incorrecto. Con este criterio de búsqueda nos queda claro que el modelo (la dependencia funcional entre las variables) que deseamos encontrar debe describir el patrón del conjunto de puntos o datos experimentales. Para este caso usaremos un **modelo de regresión**.

2. Cuando los datos obtenidos muestran una gran precisión o un grado mínimo de error aleatorio (poco ruido) y lo que se desea determinar son valores entre datos experimentales sin estar interesados en modelar el fenómeno, es decir, no se está buscando la dependencia entre las variables involucradas.

En este caso, tenemos como objetivo encontrar una curva o una serie de curvas que pasen exactamente por cada uno de los puntos. La diferencia fundamental con el caso anterior es que cada punto se considera como correcto.

A la estimación de valores entre puntos discretos se le conoce con el nombre de **interpolación**.

Suponga que se obtiene experimentalmente un conjunto de datos que son presentados en las figuras 6.1-6.1.

El primer intento de ajuste (ver fig. 6.1), no pretende conectar los puntos, sólo trata de caracterizar el crecimiento de los datos mediante una línea recta. Esta técnica ofrecería una estimación adecuada solamente para el caso lineal. Para el segundo caso (ver fig. 6.1) se utilizaron segmentos rectos entre cada par de puntos, es decir, una **interpolación lineal** que conecta dichos puntos. Esta técnica, ofrecería una estimación adecuada solamente para el caso donde los puntos están muy cercanos unos de otros y cada uno de ellos hubiese sido medido con un error aleatorio mínimo, tal que cada punto sea significativo.

Sin embargo, cuando la relación subyacente es altamente no-lineal o cuando los datos están muy separados entre sí, se pueden introducir errores importantes al realizar una interpolación

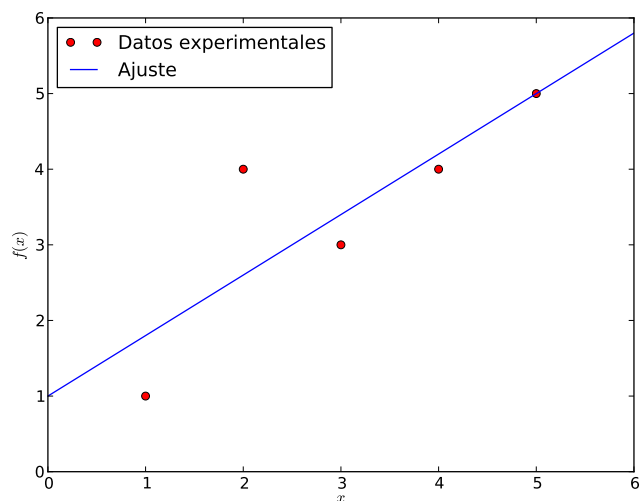


Figura 6.1: Ajuste lineal. Código Python en apéndice ??.

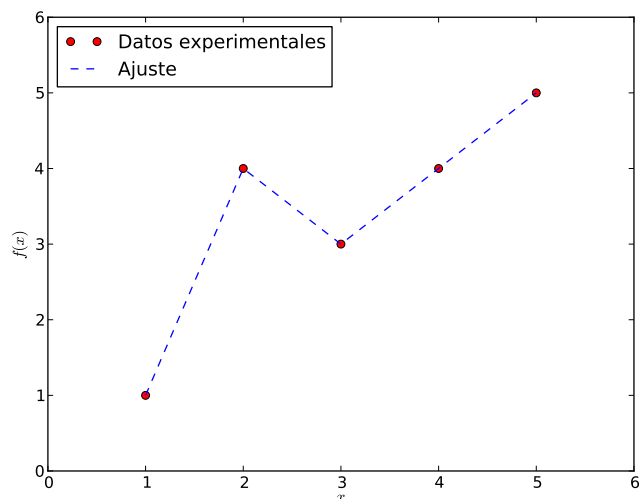


Figura 6.2: Interpolación lineal. Código Python en apéndice ??.

lineal. En el tercer caso (ver fig. 6.1) se usaron curvas que intentan capturar el comportamiento general de los datos. Este criterio para ajustar la curva será el adecuado si cada uno de los puntos del conjunto de datos está medido con un error suficientemente grande tal que el conjunto nos de información del comportamiento general, pero no cada uno de los puntos individualmente.

Lo comentado deja de manifiesto la necesidad de desarrollar métodos sistemáticos y objetivos con el propósito de determinar la curva más adecuada, ya sea para modelar o interpolar.

El **análisis de tendencias** representa el proceso de usar el patrón de los datos y hacer predicciones, pudiéndose usar polinomios de interpolación para el caso en que los datos fueron tomados con alta precisión. Este tipo de análisis se usa para predecir valores de la variable dependiente, **interpolaciones** (predecir dentro del rango de datos medidos). Otra de las aplicaciones del ajuste de curvas experimentales consiste en poner a prueba hipótesis. Esto consiste en comparar nuestro modelo teórico con los valores medidos. Pudiendo a veces ajustar los coeficientes desconocidos del modelo para que éste se ajuste mejor al experimento.

Finalmente, estos métodos de ajuste pueden usarse para derivar funciones simples que se aproximen, dentro de un rango, a funciones complicadas.

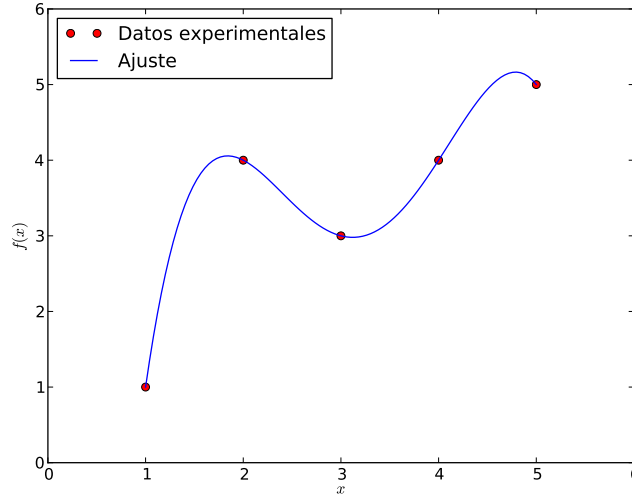


Figura 6.3: Interpolación curvilínea. Código Python en apéndice ??.

## 6.2. Regresión Lineal Simple

Este modelo considera sólo un **regresor, predictor** o variable independiente  $x$  y una variable dependiente  $Y$ . *Suponiendo* que en promedio la relación entre la variable independiente y dependiente es lineal, es decir, el valor esperado de  $Y$ ,  $E(Y)$  está dado por

$$E(Y) = \beta_0 + \beta_1 x, \quad (6.1)$$

donde el coeficiente de posición  $\beta_0$  y la pendiente  $\beta_1$  son los coeficientes desconocidos de la regresión. Por otro lado, supongamos que cada una de las observaciones de  $Y$  quede descrita por el siguiente modelo

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad (6.2)$$

donde  $\epsilon_i$  es un *error aleatorio* (valores no correlacionados) con valor medio igual a cero y varianza constante  $\sigma^2$ , es decir,  $V(\epsilon_i) = \sigma^2$ . Entenderemos que la varianza es constante cuando para un  $x_i$  dado, la variabilidad en  $y_i$  queda descrita por un mismo  $\sigma^2$  para todos los elementos de la muestra aleatoria.

Por otro lado, a partir de un conjunto de datos experimentales (una muestra) podemos hacer estimaciones de los coeficientes  $\beta_0$  y  $\beta_1$ . Un método comúnmente usado para estimar dichos parámetros es el **método de mínimos cuadrados**.

Si tenemos una muestra aleatoria formada por  $N$  pares de observaciones  $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$  y proponemos un modelo lineal  $y = b_0 + b_1 x$  para caracterizar el comportamiento de los datos experimentales, es posible expresar cada par de observaciones como

$$y_i = b_0 + b_1 x_i + \epsilon_i, \quad i = 1, 2, \dots, N. \quad (6.3)$$

La suma de los cuadrados de las diferencias entre los datos medidos y la predicción realizado por el modelo propuesto (o **residuos**) está dada por

$$\chi^2 = \sum_{i=1}^N \epsilon_i^2 = \sum_{i=1}^N (y_i - b_0 - b_1 x_i)^2. \quad (6.4)$$

Minimizando la suma de los cuadrados con respecto a los coeficientes desconocidos, obtenemos los estimadores  $b_0$  y  $b_1$  de los parámetros  $\beta_0$  y  $\beta_1$ . Las ecuaciones que determinan los estimadores son

$$\frac{\partial \chi^2}{\partial b_0} = -2 \sum_{i=1}^N (y_i - b_0 - b_1 x_i) = 0, \quad (6.5)$$

$$\frac{\partial \chi^2}{\partial b_1} = -2 \sum_{i=1}^N (y_i - b_0 - b_1 x_i) x_i = 0. \quad (6.6)$$

Simplificando y reordenando términos, obtenemos:

$$Nb_0 + b_1 \sum_{i=1}^N x_i = \sum_{i=1}^N y_i, \quad (6.7)$$

$$b_0 \sum_{i=1}^N x_i + b_1 \sum_{i=1}^N x_i^2 = \sum_{i=1}^N y_i x_i, \quad (6.8)$$

$$\begin{pmatrix} N & \sum_{i=1}^N x_i \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N y_i x_i \end{pmatrix}. \quad (6.9)$$

Resolviendo el sistema se obtiene

$$b_0 = \bar{y} - b_1 \bar{x}, \quad (6.10)$$

$$b_1 = \frac{N \left( \sum_{i=1}^N y_i x_i \right) - \left( \sum_{i=1}^N y_i \right) \left( \sum_{i=1}^N x_i \right)}{N \left( \sum_{i=1}^N x_i^2 \right) - \left( \sum_{i=1}^N x_i \right)^2} = \frac{\overline{xy} - (\bar{x})(\bar{y})}{\overline{x^2} - (\bar{x})^2}, \quad (6.11)$$

donde  $\bar{x}$ ,  $\bar{y}$ ,  $\overline{xy}$  y  $\overline{x^2}$  representan los promedios de  $\{x_i\}$ ,  $\{y_i\}$ ,  $\{x_i y_i\}$  y  $\{x_i^2\}$  respectivamente.

Una manera de cuantificar la dispersión de los datos en torno del modelo es calcular la desviación estándar

$$S_{Y/x} = \sqrt{\frac{S_r}{N-2}}, \quad (6.12)$$

con

$$S_r = \sum_{i=1}^N (y_i - b_0 - b_1 x_i)^2. \quad (6.13)$$

Note que para ajustar el modelo se introdujeron dos valores medios, es decir, se perdieron dos grados de libertad. Otra justificación del término  $N-2$  en el cálculo de la varianza es que si ajustamos una recta para sólo dos puntos no habría dispersión.

Además, el *coeficiente de determinación*  $r^2$  es definido por

$$r^2 = \frac{S_t - S_r}{S_t}, \quad (6.14)$$

donde

$$S_t = \sum_{i=1}^N (y_i - \bar{y})^2, \quad (6.15)$$

$$S_r = \sum_{i=1}^N (y_i - b_0 - b_1 x_i)^2. \quad (6.16)$$

Así  $r^2$  cuantifica la mejora del ajuste respecto del promedio y lo normaliza respecto a las desviaciones de la media  $S_t$ .

Si  $r^2 = 1$  el modelo toma en cuenta el 100 % de la variabilidad presente en los datos y la recta obtenida pasa exactamente por los puntos ajustados. Por otro lado, si  $r^2=0$  significa que el modelo no representa ninguna mejora respecto del promedio o que los datos no están correlacionados.

En este punto es conveniente mencionar que un coeficiente de determinación con valor cercano a 1, no significa necesariamente que el modelo ajustado es el más adecuado. Se recomienda, luego de graficar y evaluar el coeficiente de determinación, graficar los residuos con el propósito de intentar identificar algún patrón en ellos, o en su defecto que éstos son aleatorios. Si los residuos son aleatorios, será útil construir un **histograma de los residuos** y Así, conociendo la distribución de los residuos, poder agregar información sobre las bondades del modelo ajustado.

Ejemplo: Ver figura 6.1

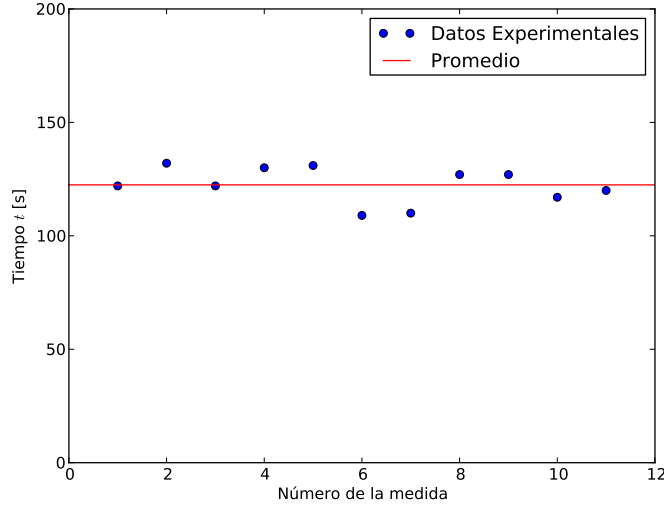


Figura 6.4: El ajuste por una constante es equivalente a calcular el promedio. Código Python en apéndice ??.

### 6.3. Regresión Polinomial

Algunos datos se representan pobremente mediante una línea recta. Para estos casos es mejor usar otro tipo de modelos. Por ejemplo, la llamada **regresión polinomial** se basa en ajustar el siguiente polinomio de grado  $m$ :

$$y = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m. \quad (6.17)$$

En este caso

$$S_r = \sum_{i=1}^N (y_i - a_0 - a_1x_i - a_2x_i^2 - \cdots - a_mx_i^m)^2, \quad (6.18)$$

$$\frac{\partial \chi^2}{\partial a_0} = -2 \sum_{i=1}^N (y_i - a_0 - a_1x_i - a_2x_i^2 - \cdots - a_mx_i^m), \quad (6.19)$$

$$\frac{\partial \chi^2}{\partial a_1} = -2 \sum_{i=1}^N x_i (y_i - a_0 - a_1x_i - a_2x_i^2 - \cdots - a_mx_i^m), \quad (6.20)$$

$$\frac{\partial \chi^2}{\partial a_2} = -2 \sum_{i=1}^N x_i^2 (y_i - a_0 - a_1x_i - a_2x_i^2 - \cdots - a_mx_i^m), \quad (6.21)$$

$$\vdots = \vdots, \quad (6.22)$$

$$\frac{\partial \chi^2}{\partial a_m} = -2 \sum_{i=1}^N x_i^m (y_i - a_0 - a_1x_i - a_2x_i^2 - \cdots - a_mx_i^m). \quad (6.23)$$

Igualando a cero y reordenando términos, obtenemos

$$a_0 N + a_1 \sum_{i=1}^N x_i + a_2 \sum_{i=1}^N x_i^2 + \cdots + a_m \sum_{i=1}^N x_i^m = \sum_{i=1}^N y_i, \quad (6.24)$$

$$a_0 \sum_{i=1}^N x_i + a_1 \sum_{i=1}^N x_i^2 + a_2 \sum_{i=1}^N x_i^3 + \cdots + a_m \sum_{i=1}^N x_i^{m+1} = \sum_{i=1}^N x_i y_i, \quad (6.25)$$

$$a_0 \sum_{i=1}^N x_i^2 + a_1 \sum_{i=1}^N x_i^3 + a_2 \sum_{i=1}^N x_i^4 + \cdots + a_m \sum_{i=1}^N x_i^{m+2} = \sum_{i=1}^N x_i^2 y_i, \quad (6.26)$$

$$\vdots = \vdots, \quad (6.27)$$

$$a_0 \sum_{i=1}^N x_i^m + a_1 \sum_{i=1}^N x_i^{m+1} + a_2 \sum_{i=1}^N x_i^{m+2} + \cdots + a_m \sum_{i=1}^N x_i^{2m} = \sum_{i=1}^N x_i^m y_i. \quad (6.28)$$

Note que el número de incógnitas  $(a_0, a_1, a_2, \dots, a_m)$  es igual al número de ecuaciones  $(m+1)$ . El sistema de ecuaciones anterior es lineal en las incógnitas, y puede escribirse en la forma estándar como sigue:

$$\begin{pmatrix} N & \sum x_i & \sum x_i^2 & \cdots & \sum x_i^m \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \cdots & \sum x_i^{m+1} \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \cdots & \sum x_i^{m+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_i^m & \sum x_i^{m+1} & \sum x_i^{m+2} & \cdots & \sum x_i^{2m} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \\ \vdots \\ \sum x_i^m y_i \end{pmatrix}. \quad (6.29)$$

El error en el ajuste polinomial se cuantifica mediante el **error estándar** de la aproximación,

$$S_{Y/x} = \sqrt{\frac{S_r}{N - (m+1)}}, \quad (6.30)$$

donde  $m$  es el orden del polinomio. Esta cantidad se divide por  $N - (m+1)$  ya que se usaron  $m+1$  coeficientes. Además del error estándar podemos calcular el coeficiente de determinación usando nuevamente la definición (6.14) y (6.15), pero donde ahora  $S_r$  es dado por (6.18).

**Ejemplo:** Ajustar un polinomio de grado 2, para la siguiente tabla de valores 6.1.

Posición $x, [\text{m}] \pm 0,1$	Tiempo $t, [\text{s}] \pm 0,01$
0.00	2.1
1.00	7.7
2.00	13.6
3.00	27.2
4.00	40.9
5.00	61.1

Cuadro 6.1: Posición de un movimiento unidimensional como función del tiempo.

### 6.3.1. Regresión lineal múltiple

Una extensión útil de la regresión lineal es cuando  $y$  es función de dos o más variables:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m. \quad (6.31)$$

Análogamente a los casos anteriores, los “mejores” coeficientes quedan determinados minimizando el cuadrado de los residuos:

$$S_r = \sum_{i=1}^N (y_i - a_0 - a_1 x_1 - a_2 x_2 - \cdots - a_m x_m)^2. \quad (6.32)$$



En este caso las ecuaciones para los estimadores quedan determinadas por

$$\frac{\partial \chi^2}{\partial a_0} = -2 \sum_{i=1}^N (y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i} - \cdots - a_m x_{m,i}), \quad (6.33)$$

$$\frac{\partial \chi^2}{\partial a_1} = -2 \sum_{i=1}^N x_{1,i} (y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i} - \cdots - a_m x_{m,i}), \quad (6.34)$$

$$\frac{\partial \chi^2}{\partial a_2} = -2 \sum_{i=1}^N x_{2,i} ((y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i} - \cdots - a_m x_{m,i}), \quad (6.35)$$

$$\vdots = \vdots \quad (6.36)$$

$$\frac{\partial \chi^2}{\partial a_m} = -2 \sum_{i=1}^N x_{m,i} ((y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i} - \cdots - a_m x_{m,i}). \quad (6.37)$$

Igualando a cero y reordenando términos, obtenemos

$$a_0 N + a_1 \sum_{i=1}^N x_{1,i} + a_2 \sum_{i=1}^N x_{2,i} + \cdots + a_m \sum_{i=1}^N x_{m,i} = \sum_{i=1}^N y_i, \quad (6.38)$$

$$a_0 \sum_{i=1}^N x_{1,i} + a_1 \sum_{i=1}^N x_{1,i}^2 + a_2 \sum_{i=1}^N x_{1,i} x_{2,i} + \cdots + a_m \sum_{i=1}^N x_{1,i} x_{m,i} = \sum_{i=1}^N x_{1,i} y_i, \quad (6.39)$$

$$a_0 \sum_{i=1}^N x_{2,i} + a_1 \sum_{i=1}^N x_{2,i} x_{1,i} + a_2 \sum_{i=1}^N x_{2,i}^2 + \cdots + a_m \sum_{i=1}^N x_{2,i} x_{m,i} = \sum_{i=1}^N x_{2,i} y_i, \quad (6.40)$$

$$\vdots = \vdots \quad (6.41)$$

$$a_0 \sum_{i=1}^N x_{m,i} + a_1 \sum_{i=1}^N x_{m,i} x_{1,i} + a_2 \sum_{i=1}^N x_{m,i} x_{2,i} + \cdots + a_m \sum_{i=1}^N x_{m,i}^2 = \sum_{i=1}^N x_{m,i} y_i. \quad (6.42)$$

Esto define el siguiente sistema lineal de ecuaciones para  $(a_0, a_1, a_2, \dots, a_m)$ :

$$\begin{pmatrix} N & \sum x_{1,i} & \sum x_{2,i} & \cdots & \sum x_{m,i} \\ \sum x_{1,i} & \sum x_{1,i} x_{1,i} & \sum x_{1,i} x_{2,i} & \cdots & \sum x_{1,i} x_{m,i} \\ \sum x_{2,i} & \sum x_{2,i} x_{1,i} & \sum x_{2,i} x_{2,i} & \cdots & \sum x_{2,i} x_{m,i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{m,i} & \sum x_{m,i} x_{1,i} & \sum x_{m,i} x_{2,i} & \cdots & \sum x_{m,i} x_{m,i} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_{1,i} y_i \\ \sum x_{2,i} y_i \\ \vdots \\ \sum x_{m,i} y_i \end{pmatrix} \quad (6.43)$$

**Ejemplo:** Determine el “mejor” plano que ajusta los valores en la tabla 6.2 y evalúe el error estándar.

$x$	$y$	$z$
0	0	5
2	1	10
2.5	2	9
1	3	0
4	6	3
7	2	27

Cuadro 6.2: Datos para la Tarea: valores de  $x$ ,  $y$  y  $z$ .

### 6.3.2. Generalización de la idea de Regresión y mínimos Cuadrados

Sean  $\{x_i, y_i\}$ ,  $i = 1, \dots, N$ , un conjunto de  $N$  pares de datos medidos experimentalmente. La manipulación de estos datos tiene normalmente como objetivo determinar una ley experimental, o bien comprobar alguna ley previamente supuesta. En cualquiera de estos casos la ley se expresa en forma de una relación  $y = f(x)$  que entregue de manera fidedigna la correlación entre  $x$  y  $y$ .

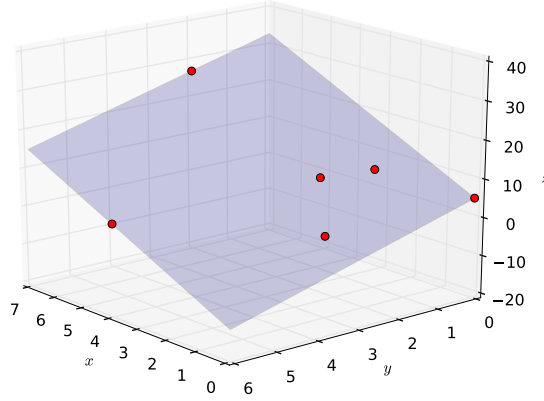


Figura 6.5: Ajuste por mínimos cuadrados de los datos por un plano. Código Python en apéndice ???. En este caso ajustamos una función  $z = a_0 + a_1x + a_2y$ . Los coeficientes resultantes son  $a_0 = 5,0$ ,  $a_1 = 4,0$  y  $a_2 = -3,0$ . El coeficiente de determinación resulta ser  $r^2 = 1$ .

La idea de regresión es siempre la misma, elegir un modelo bajo algunos supuestos y luego elegir algún método para estimar los parámetros de dicho modelo. Entonces, debemos tener claro que el método de mínimos cuadrados o cualquier otro método de ajuste, no selecciona modelos, solamente ajusta de manera eficiente los parámetros que contempla un modelo dado (por ejemplo, para un modelo lineal que depende de  $\beta_0$  y  $\beta_1$  un método muy usado para estimar ambos parámetros es el de mínimos cuadrados, donde los estimadores son  $b_0$  y  $b_1$ ).

Supongamos que el modelo propuesto está dado por:

$$Y = f(x) + \epsilon, \quad (6.44)$$

donde  $\epsilon$  es un error aleatorio con  $E(\epsilon) = 0$  y varianza  $V(\epsilon) = \sigma^2$  constante. Tal como se mencionó anteriormente, la idea de varianza constante consiste en que si el intervalo es dividido en sub-intervalos todos poseen la misma varianza. Esto mismo visto de manera intuitiva, consiste en que la dispersión de los datos (en  $y$ ) respecto a la curva ajustada (modelo) debe ser (aproximadamente) independiente de  $x$ .

### Ajuste de una función arbitraria por mínimos cuadrados

Esto consiste en definir la siguiente funcional:

$$\chi^2 = \sum_{i=1}^N (y_i - f(x_i))^2. \quad (6.45)$$

Por construcción  $S_r$  satisface:

1.  $\chi^2$  es una magnitud definida como positiva, es decir  $\chi^2 \geq 0$ .
2.  $\chi^2 = 0$  si y sólo si  $y_i = f(x_i)$ ,  $\forall i$ .

Además, se exige que los residuos satisfagan  $E(\epsilon) = 0$ . El método de mínimos cuadrados garantiza automáticamente que se satisface esta condición.

Normalmente no se cumple la condición  $y_i = f(x_i)$ , por lo que  $\chi^2$  nunca es cero. La forma de determinar los parámetros ajustables de la función  $f$  es hacer que  $\chi^2$  sea lo más próximo a cero, lo que se logra *minimizando* su valor con respecto a los parámetros ajustables. Al realizar esto, se obtiene una función que describe, sólo de manera aproximada, el comportamiento global de los puntos experimentales. Por lo tanto, el problema de encontrar  $f$  se traduce en minimizar la funcional  $S_r$  para una familia de funciones dada.

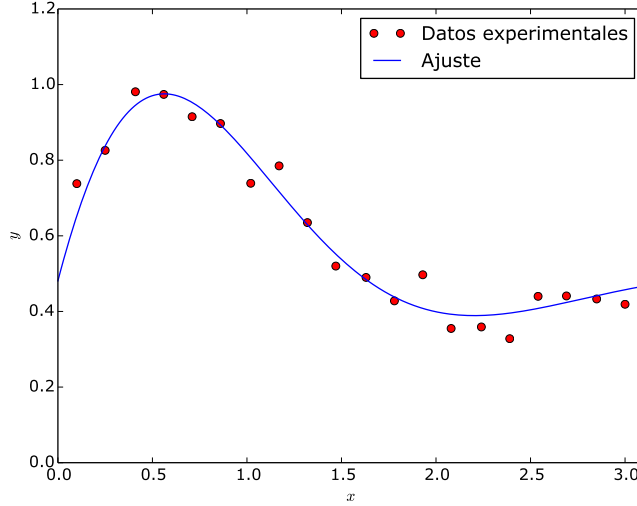


Figura 6.6: Ajuste no-polinomial de datos usando el método de mínimos cuadrados. Código Python en apéndice ??.

## 6.4. Métodos de Ajuste

### 6.4.1. mínimos Cuadrados

### 6.4.2. mínimos Cuadrados Ponderados

En todos los casos anteriores hemos exigido que se cumplan las siguientes condiciones para los residuos:

- Valor esperado de los residuos igual a cero  $E(\epsilon_i) = 0$ ,
- varianza constante  $V(\epsilon_i) = \sigma^2$ ,
- los valores de los residuos son aleatorios, es decir, no correlacionados, y
- las incertezas se consideran solamente para la variable dependiente.

Si la varianza de los residuos no es constante (ver figura 6.5), podemos introducir una mejora para estimar los parámetros de ajuste por mínimos cuadrados, introduciendo la idea de mínimos cuadrados ponderados. Al ajustar un modelo a este conjunto de puntos deberíamos incorporar el hecho que la varianza no es la misma, es decir, la dispersión no es la misma para cada uno de los  $y_i$ , lo que implicaría que no todos los  $y_i$  tienen la misma variabilidad. Una posibilidad para incluir este hecho en el método de ajuste, es considerar mínimos cuadrados ponderados, método que contempla diferentes pesos para cada uno de los valores medidos, según el grado de dispersión de cada uno de éstos. A cada medición se le asocia un peso  $w_i$ , tal que

$$S_r = \sum_{i=1}^N w_i (y_i - f(x_i))^2. \quad (6.46)$$

Luego repetimos el mismo procedimiento utilizado por el método de mínimos cuadrados ordinarios visto anteriormente. Si para cada una de las expresiones de los estimadores obtenidas usando mínimos cuadrados ordinarios cambiamos  $\sum_{i=1}^N$  por  $\sum_{i=1}^N w_i$ , obtenemos las expresiones para mínimos cuadrados ponderados.

#### Ajuste de un modelo lineal

$$b_0 = \frac{\left(\sum_{i=1}^N w_i y_i\right) \left(\sum_{i=1}^N w_i x_i^2\right) - \left(\sum_{i=1}^N w_i x_i y_i\right) \left(\sum_{i=1}^N w_i x_i\right)}{\left(\sum_{i=1}^N w_i\right) \left(\sum_{i=1}^N w_i x_i^2\right) - \left(\sum_{i=1}^N w_i x_i\right)^2}, \quad (6.47)$$

$$b_1 = \frac{\left(\sum_{i=1}^N w_i\right) \left(\sum_{i=1}^N w_i x_i y_i\right) - \left(\sum_{i=1}^N w_i x_i\right) \left(\sum_{i=1}^N w_i y_i\right)}{\left(\sum_{i=1}^N w_i\right) \left(\sum_{i=1}^N w_i x_i^2\right) - \left(\sum_{i=1}^N w_i x_i\right)^2}. \quad (6.48)$$

### Ajuste polinomial

$$\begin{pmatrix} \sum w_i & \sum w_i x_i & \sum w_i x_i^2 & \cdots & \sum w_i x_i^m \\ \sum w_i x_i & \sum w_i x_i^2 & \sum w_i x_i^3 & \cdots & \sum w_i x_i^{m+1} \\ \sum w_i x_i^2 & \sum w_i x_i^3 & \sum w_i x_i^4 & \cdots & \sum w_i x_i^{m+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum w_i x_i^m & \sum w_i x_i^{m+1} & \sum w_i x_i^{m+2} & \cdots & \sum w_i x_i^{2m} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \sum w_i y_i \\ \sum w_i x_i y_i \\ \sum w_i x_i^2 y_i \\ \vdots \\ \sum w_i x_i^m y_i \end{pmatrix}. \quad (6.49)$$

### Ajuste lineal multiple

$$\begin{pmatrix} \sum w_i & \sum w_i x_{1,i} & \sum w_i x_{2,i} & \cdots & \sum w_i x_{m,i} \\ \sum w_i x_{1,i} & \sum w_i x_{1,i} x_{1,i} & \sum w_i x_{1,i} x_{2,i} & \cdots & \sum w_i x_{1,i} x_{m,i} \\ \sum w_i x_{2,i} & \sum w_i x_{2,i} x_{1,i} & \sum w_i x_{2,i} x_{2,i} & \cdots & \sum w_i x_{2,i} x_{m,i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum w_i x_{m,i} & \sum w_i x_{m,i} x_{1,i} & \sum w_i x_{m,i} x_{2,i} & \cdots & \sum w_i x_{m,i} x_{m,i} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \sum w_i y_i \\ \sum w_i x_{1,i} y_i \\ \sum w_i x_{2,i} y_i \\ \vdots \\ \sum w_i x_{m,i} y_i \end{pmatrix}. \quad (6.50)$$

Notemos que no hemos dado una forma explícita para los pesos  $w_i$  y eso se debe a que no hay una única manera de definirlos. Por ejemplo, un buen criterio podría ser que los pesos sean inversamente proporcionales a la incertidumbre, es decir, le daríamos mayor credibilidad a los  $y_i$  con menos incertidumbre y menor credibilidad a los  $y_i$  con mayor incertidumbre. Para esto es común elegir  $w_i = 1/(\Delta y_i)^2$ . La idea de definir pesos consiste en no considerar todos los datos por igual, por lo que para cada experimento podrán usarse diferentes criterios para definir los pesos más adecuados que no necesariamente estarán basados en errores asociados si no también en argumentos físicos, como por ejemplo que la curva deba pasar por el origen o algún valor determinado.

**Ejemplo:** Considere los datos de la siguiente tabla: Podemos comparar el resultado de ajustar

$x$	$y \pm \Delta y$
1,0	$2,8 \pm 0,3$
2,0	$3,3 \pm 0,3$
3,0	$3,5 \pm 0,5$
4,0	$3,5 \pm 1,0$
5,0	$4,8 \pm 0,3$
6,0	$4,2 \pm 1,0$

Cuadro 6.3: Valores de  $x$  e  $y$ , con error en  $y$ .

una recta a estos datos usando el método de mínimos cuadrados tradicional (sin ponderar, es decir, sin tomar en cuenta los valores de  $\Delta y$ ) y el método de mínimos cuadrados ponderado ( eligiendo  $w_i = 1/(\Delta y_i)^2$ ).

#### 6.4.3. máxima Verosimilitud

Otro método para estimar los parámetros de los modelos es maximizando la **función de verosimilitud**.

Supongamos que disponemos del siguiente modelo a ser ajustado,

$$Y = f(x) + \epsilon, \quad (6.51)$$

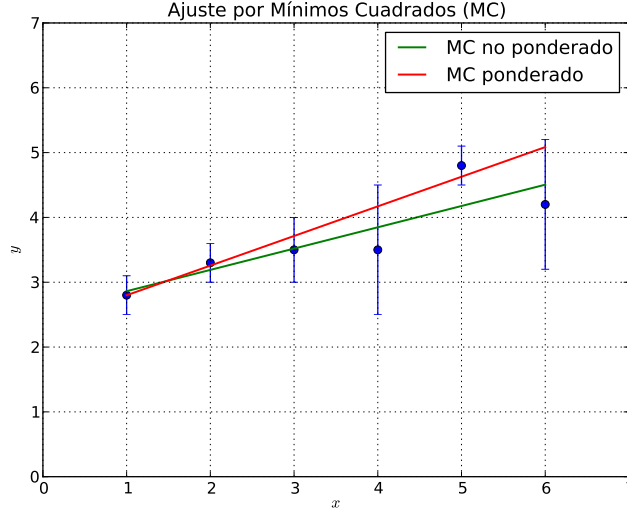


Figura 6.7: Ajuste por mínimos cuadrados con y sin ponderación. Código Python en apéndice ??.

y que conocemos la distribución de los residuos  $g(\epsilon)$ . En este caso podríamos construir la función verosimilitud, definida por

$$L(b_0, b_1, \dots, \sigma^2) = \prod_{i=1}^N g(y_i - f(x_i)). \quad (6.52)$$

Maximizando la función verosimilitud (o, equivalentemente, su logaritmo) con respecto a los  $N+1$  parámetros ajustables o estimadores de los parámetros del modelo propuesto, encontramos

$$\frac{\partial \ln L(b_0, \dots, b_i, \dots, b_N, \sigma^2)}{\partial b_0} = 0, \quad (6.53)$$

$$\vdots = 0, \quad (6.54)$$

$$\frac{\partial \ln L(b_0, \dots, b_i, \dots, b_N, \sigma^2)}{\partial b_i} = 0, \quad (6.55)$$

$$\vdots = 0, \quad (6.56)$$

$$\frac{\partial \ln L(b_0, \dots, b_i, \dots, b_N, \sigma^2)}{\partial b_N} = 0. \quad (6.57)$$

Agregamos otra condición que nos permite determinar el valor de la varianza:

$$\frac{\partial \ln L(b_0, \dots, b_i, \dots, b_N, \sigma^2)}{\partial \sigma^2} = 0. \quad (6.58)$$

Resolviendo este conjunto de  $N+2$  ecuaciones, quedan determinados los estimadores de los coeficientes y la varianza.

Note que si la distribución de los residuos es normal, el método de máxima verosimilitud entrega los mismos valores para los parámetros  $b_0, b_1, \dots, b_N$  que mínimos cuadrados.

**Ejemplo:** Supongamos que deseamos ajustar un modelo lineal y sabemos que los residuos tienen una distribución normal

$$f(\epsilon, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{\epsilon}{\sigma}\right)^2}. \quad (6.59)$$

Modelo:  $Y = \beta_0 + \beta_1 x + \epsilon$  y para los datos de la muestra de tamaño  $N$  se cumple que  $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$  con  $i = 1, \dots, N$ .

En este caso, la función verosimilitud queda definida como

$$L(b_0, b_1, \sigma^2) = \prod_{i=1}^N f(y_i - b_0 - b_1 x_i) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \left( \frac{y_i - b_0 - b_1 x_i}{\sigma} \right)^2}. \quad (6.60)$$

Maximizando  $\ln L(b_0, b_1, \sigma^2)$  respecto de los parámetros  $b_0, b_1$  y  $\sigma^2$  se obtienen tres ecuaciones,

$$\frac{\partial \ln L(b_0, b_1, \sigma^2)}{\partial b_0} = 0, \quad (6.61)$$

$$\frac{\partial \ln L(b_0, b_1, \sigma^2)}{\partial b_1} = 0, \quad (6.62)$$

$$\frac{\partial \ln L(b_0, b_1, \sigma^2)}{\partial \sigma^2} = 0, \quad (6.63)$$

donde las dos primeras ecuaciones nos conducen a las mismas soluciones de ambos estimadores  $b_0$  y  $b_1$ , para el caso de distribución normal de los residuos, que el metodo de mínimos cuadrados.

## 6.5. análisis (Gráfico e histograma) de residuos

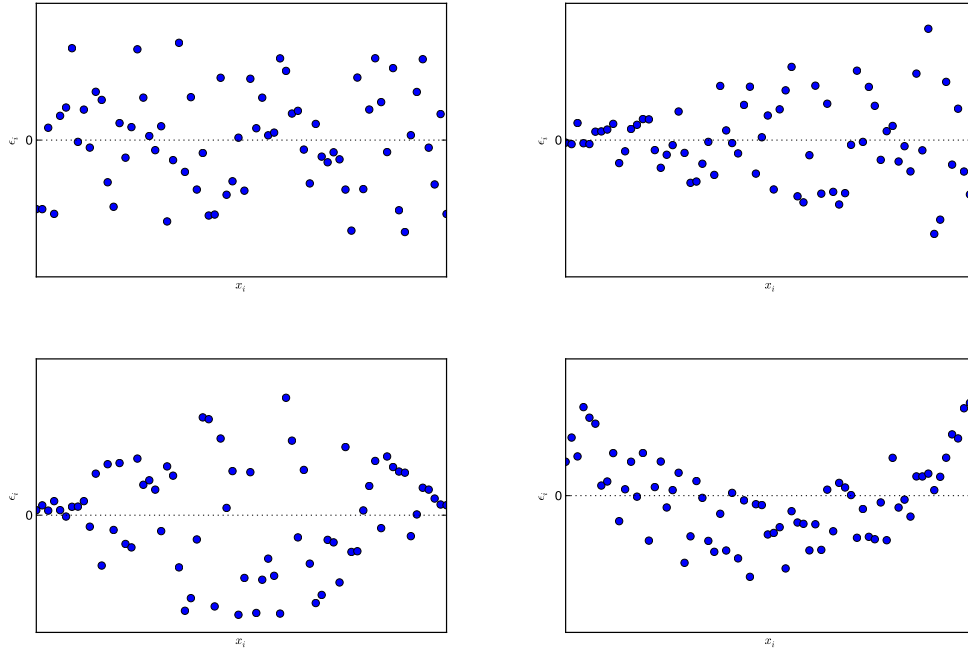


Figura 6.8: Ejemplos de distintas posibilidades para la dispersión.

# Capítulo 7

## Series de tiempo

### 7.1. Introducción

Una serie de tiempo es una secuencia de observaciones ordenadas cronológicamente y dependientes entre si. Ejemplos: Temperatura en Concepción, producto interno bruto etc. En los casos mencionados vemos que el valor esta basado en datos anteriores.

Las series de tiempo las podemos clasificar como: Discretas: Cuando el conjunto de observaciones es finito o infinito numerable  $y_t$ .

Continuas: Cuando el conjunto de observaciones es infinito no numerable.

Determinísticas: Cuando se puede usar un modelo para predecir exactamente los valores futuros de la serie.

Estocásticos: Cuando los valores futuros de la serie sólo pueden ser determinados en términos probabilísticos, pues el modelo tiene un factor aleatorio.

Una de las características especiales de las series de tiempo, es que observaciones sucesivas no son independientes, de modo que su análisis debe considerar el orden de dichas observaciones.

Como objetivo nos plantearemos:

1.- Encontrar un modelo (o familia de modelos) que describa estas series. 2.- Con el modelo filtrar la señal de ruido, predecir valores futuros y controlar valores futuros.

### 7.2. Componentes de una Serie Temporal

El análisis clásico de las series temporales se basa en la suposición de que los valores que toma la variable de observación es la consecuencia de tres componentes, cuya acción conjunta da como resultado los valores medidos.

Las tres componentes principales son: a.- Tendencia ( $T_t$ ) se puede definir como un cambio de largo plazo y esta relacionado con el cambio de la media. Esta tendencia podría ser modelada por regresión y los coeficientes ajustados por algunos de los métodos usados (Mínimos cuadrados, mínimos cuadrados ponderados o Máxima verosimilitud)

b.- Componente estacional ( $E_t$ ) Muchas series de tiempo presentan ciertas variaciones periódicas. Es importante hacer notar que podrían haber 2 o más periodos y si un periodo es mucho mayor que el otro, suelen llamar al periodo mayor ciclo. Como en muchos casos lo que se busca es determinar la tendencia, la estacionalidad debe ser eliminada del modelo predictor. Este proceso se llama desestacionalizar la serie. Ejemplo: El índice de precios al consumidor IPC, en este caso suelo buscarse la tendencia del IPC y no las variaciones estacionales, las cuales se repiten todos los años en determinados meses y no aportan a conocer la tendencia de largo plazo.

c.- Componente aleatoria ( $I_t$ ) esta componente no responde a ningún patrón de comportamiento, sino que es el resultado de factores aleatorios que inciden en la serie de tiempo.

Los dos casos más simples a ser estudiados en estos apuntes son:

1.- Si la serie estuviese formada por la suma de las mencionadas componentes.

$$Y_t = T_t + E_t + I_t \quad (7.1)$$

Incluir gráfico de la serie como suma Incluir gráficos de  $T_t$   $E_t$   $I_t$  por separado cada componente Filtro Media móvil Ejemplo suavizar borrando estacionalidad y ruido para determinar tendencia. luego hacer análisis de Fourier para encontrar estacionalidad análisis de residuos para ver

su es realmente ruido o tiene estructura. Mostrar como caracterizar el ruido unido de las componentes. Graficar ruido hacer historama del ruido

2.- Si la serie estuviese formada por el producto de las componentes.

$$Y_t = T_t * E_t * I_t \quad (7.2)$$

Mostrar el grafico de una serie para este caso. Graficar cada componente por separado. Usar media movil para suavizar la serie. Usar media movil para borrar estacionalidad

Cuando hablamos de una secuencia de valores observados a lo largo del tiempo, la denominamos, en un sentido amplio, **serie temporal**. Resulta difícil imaginar una rama de la ciencia en la que no aparezcan datos que puedan ser considerados como series temporales.

Si, conocidos los valores pasados de la serie, no fuera posible predecir con total certeza el próximo valor de la variable, decimos que la serie es **no determinista** o **aleatoria**, y lógicamente es de éstas de las que se ocupa el cuerpo de doctrina denominado “análisis de series temporales” y al que vamos a dedicar esta breve introducción.

El análisis estadístico de series temporales se usa hoy día en muchas áreas de la Ciencia, fundamentalmente en física, Ingeniería y en Economía.

Los objetivos del análisis de series temporales son diversos, pudiendo destacar la predicción, el control de un proceso, la simulación de procesos, y la generación de nuevas teorías físicas o biológicas.

Denominamos **predicción** a la estimación de valores futuros de la variable en función del comportamiento pasado de la serie. Este objetivo se emplea ampliamente en el campo de la Ingeniería y de la Economía, incluyendo en esta última rama también la sanidad pública y la vigilancia de la salud. Así por ejemplo, la predicción mediante modelos basados en la teoría de series temporales, puede servir para una buena planificación de recursos sanitarios, en función de la demanda que se espera en el futuro, prevista por el modelo. Otro de los campos en los que se aplica la predicción mediante series temporales es el de la meteorología o en la predicción de otros fenómenos naturales.

En la teoría de control de procesos, se trata de seguir la evolución de una variable determinada con el fin de regular su resultado. Esta teoría se utiliza en Medicina en los Centros de Control de Enfermedades. La **simulación** se emplea en investigación aplicada, cuando el proceso es muy complejo para ser estudiado de forma analítica.

Evidentemente, aunque el valor futuro de una serie temporal no sea predecible con total exactitud, para que tenga interés su estudio, el resultado tampoco puede ser completamente aleatorio, existiendo alguna regularidad en cuanto a su comportamiento en el tiempo, lo que hará posible su modelado y por ende, en su caso, la predicción. La búsqueda de regularidades y de patrones ha sido siempre una de las tareas básicas de la Ciencia, y muchas veces se descubren simetrías que sirven de fundamento para la predicción del comportamiento de los fenómenos, incluso antes de que se entienda la razón o causa que justifica esa regularidad. Esto ocurrió, por ejemplo, con el sistema periódico de los elementos, descrito por Mendeleiev (1834-1907), quien organizó de forma muy correcta los elementos químicos en base a las simetrías observadas entre ellos, antes de que se comprendiese la razón de esas simetrías o periodicidad, razones que luego se fundamentaron sobre todo en trabajos de Schrödinger (1887-1961) y Pauli (1900-1958).

Por lo tanto, si podemos encontrar patrones de regularidad en diferentes secciones de una serie temporal, podremos también describirlas mediante modelos basados en distribuciones de probabilidad. La secuencia ordenada de variables aleatorias  $x(t)$  y su distribución de probabilidad asociada, se denomina **proceso estocástico**. Un proceso estocástico es por tanto el modelo matemático para una serie temporal.

Un concepto importante que encontramos en este ámbito, es el de procesos estacionarios. Si examinamos por ejemplo la temperatura para un determinado mes a lo largo de los años en una determinada zona geográfica, y se está produciendo un cambio climático, aunque haya fluctuaciones, habrá una tendencia creciente. De una manera informal, diremos que una serie es **estacionaria** cuando se encuentra en equilibrio estadístico, en el sentido de que sus propiedades no varían a lo largo del tiempo, y por lo tanto no pueden existir tendencias. Un proceso es **no-estacionario** si sus propiedades varían con el tiempo, como el clima.

Vamos ahora a presentar tres enfoques diferentes, aunque relacionados, para el análisis de series temporales.

El primer paso obligatorio para analizar una serie temporal es presentar un Gráfico de la evolución de la variable a lo largo del tiempo, como puede ser el de la figura 7.1:

El siguiente paso consistirá en determinar si la secuencia de valores es completamente aleatoria o si, por el contrario, se puede encontrar algún patrón a lo largo del tiempo, pues sólo en este caso podremos seguir con el análisis.



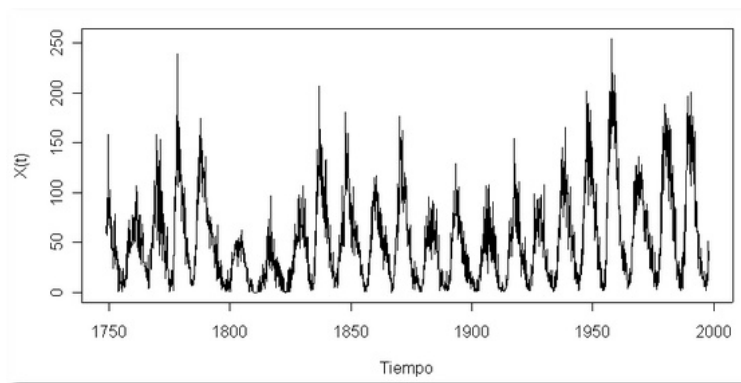


Figura 7.1: Serie temporal.

La metodologíaa tradicional para el estudio de series temporales es bastante sencilla de comprender, y fundamentalmente se basa en descomponer las series en varias partes: tendencia, variación estacional o periódica, y otras fluctuaciones irregulares.

**Tendencia.** Es la dirección general de la variable en el periodo de observación, es decir, el cambio a largo plazo de la media de la serie.

**Estacionalidad.** Corresponde a fluctuaciones periódicas de la variable, en periodos relativamente cortos de tiempo.

**Otras fluctuaciones irregulares.** Después de extraer de la serie la tendencia y variaciones cíclicas, nos quedará una serie de valores residuales, que pueden ser o no totalmente aleatorios. Volvemos a estar como en el punto de partida, pues ahora también nos interesa determinar si esa secuencia temporal de valores residuales puede o no ser considerada como aleatoria pura.

En la figura 7.2 vemos un ejemplo de una serie temporal en la que se aprecia la existencia de las distintas componentes comentadas

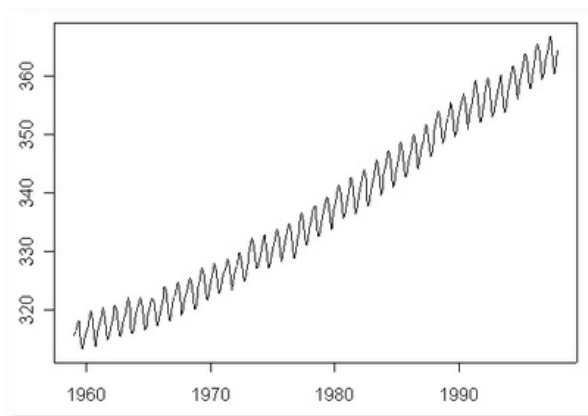


Figura 7.2: Tendencia, estacionalidad y fluctuaciones irregulares.

### 7.3. Componentes de una serie: tendencia, estacionalidad y ruido

### 7.4. análisis de la tendencia

Una primera idea sobre la presencia de tendencia en la serie la obtendremos en su representación gráfica. Pero no siempre estará tan clara como en la figura 7.2. Por ejemplo, en los datos de la figura 7.3 sigue habiendo tendencia, pero ya no es tan marcada.

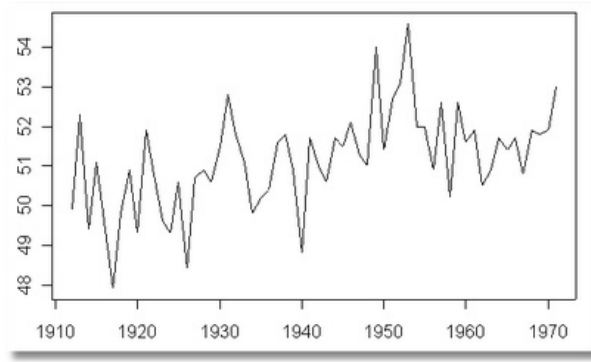


Figura 7.3: Tendencia, estacionalidad y fluctuaciones irregulares.

Los medios más utilizados para detectar y eliminar la **tendencia** de una serie se basan en la aplicación de **filtros** a los datos. Un filtro no es más que una función matemática que aplicada a los valores de la serie produce una nueva serie con unas características determinadas. Entre esos filtros encontramos las **medias móviles**.

Una media móvil se calcula, para cada punto, como un promedio del mismo número de valores a cada lado de ese punto. así una media móvil de tres puntos se calcula como:

$$m(x_t) = \frac{x_{t-1} + x_t + x_{t+1}}{3}. \quad (7.3)$$

Mientras que una media móvil de cuatro puntos viene dada por

$$m(x_t) = \frac{x_{t-2}/2 + x_{t-1} + x_t + x_{t+1} + x_{t+2}/2}{4}. \quad (7.4)$$

Cuando la cantidad de puntos de la media móvil es par, se toma la mitad de los valores extremos.

Existen otros procedimientos para extraer la tendencia, como **ajuste de polinomios**, entre otros.

Una clase de filtro, que es particularmente útil para eliminar la tendencia, se basa en aplicar **diferencias** a la serie hasta convertirla en estacionaria. Una diferencia de primer orden se obtiene restando dos valores contiguos:

$$\Delta x_{t+1} := x_{t+1} - x_t. \quad (7.5)$$

Si volvemos a diferenciar esa serie, restando los nuevos valores consecutivos obtenemos una nueva serie más suavizada.

$$\Delta^2 x_{t+2} := \Delta x_{t+2} - \Delta x_{t+1}. \quad (7.6)$$

Una vez que se aplica un proceso clásico de descomposición mediante un procedimiento de medias móviles a los datos de la figura 7.2, se obtiene las siguientes series:

Para analizar la **estacionalidad** de una serie introduciremos un concepto de gran interés en el análisis de series temporales: la **función de autocorrelación**.

La **función de autocorrelación** mide la correlación entre los valores de la serie distanciados un lapso de tiempo  $k$ .

La fórmula del coeficiente de correlación simple, dados  $N$  pares de observaciones  $y, x$ :

$$r := \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum (y_i - \bar{y})^2 \sum (x_i - \bar{x})^2}}. \quad (7.7)$$

De igual forma, dada una secuencia temporal de  $N$  observaciones  $x_1, \dots, x_N$ , podemos formar  $N - 1$  parejas de observaciones contiguas  $(x_1, x_2), (x_2, x_3), \dots, (x_{N-1}, x_N)$  y calcular el coeficiente de correlación de estas parejas.

A este coeficiente lo denominaremos **coeficiente de autocorrelación** de orden 1 y lo denotamos como  $r_1$ . Análogamente se pueden formar parejas con puntos separados por una distancia 2, es decir  $(x_1, x_3), (x_2, x_4)$ , etc. y calcular el nuevo coeficiente de autocorrelación de orden 2. De forma general, si preparamos parejas con puntos separados una distancia  $k$ , calcularemos el coeficiente de autocorrelación de orden  $k$ .

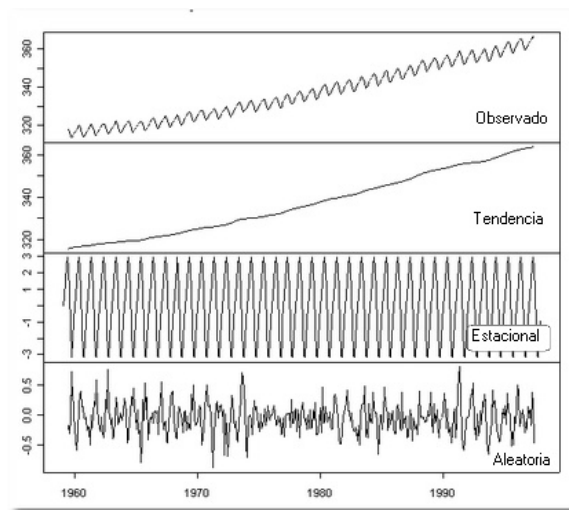


Figura 7.4: Observado, tendencia, estacional, aleatoria.

Al igual que para el coeficiente de correlación lineal simple, se puede calcular un error estándar y por tanto un intervalo de confianza para el coeficiente de autocorrelación.

La **función de autocorrelación** es el conjunto de coeficientes de autocorrelación  $r_k$  desde 1 hasta un máximo que no puede exceder la mitad de los valores observados, y es de gran importancia para estudiar la estacionalidad de la serie, ya que si ésta existe, los valores separados entre sí por intervalos iguales al periodo estacional deben estar correlacionados de alguna forma. Es decir que el coeficiente de autocorrelación para un retardo igual al periodo estacional debe ser significativamente diferente de 0.

## 7.5. métodos de análisis de componentes