

# Apuntes de Laboratorio 1

15 de noviembre de 2016

# Índice general

<b>1. Medidas y errores</b>	<b>2</b>
1.1. Introducción	2
1.1.1. Diferentes formas de expresar el error de una medida	2
1.2. Precisión y exactitud	3
1.3. Promedio y desviación estándar	5
1.4. Cifras significativas	5
1.5. Aproximación y Redondeo	6
1.6. Orden de magnitud	7
1.7. Precisión, sensibilidad, y cifras significativas	7
1.7.1. Caso de una única medición	7
1.7.2. Caso de mediciones repetidas	8
<b>2. Noción de Probabilidades</b>	<b>9</b>
2.1. Introducción	9
2.1.1. Experimento aleatorio	9
2.1.2. Población	10
2.1.3. Espacio muestral	10
2.1.4. Evento	10
2.2. Interpretación de la Probabilidad	10
2.3. Axiomas de probabilidad	11
2.3.1. Consecuencias directas de los axiomas	11
2.4. Reglas de adición	11
2.5. Probabilidad condicional	12
2.5.1. Reglas de multiplicación	13
2.5.2. Reglas de Probabilidad total	14
2.5.3. Teorema de Bayes	14
2.6. Variables aleatorias	14
2.7. Histogramas	15
2.8. Distribuciones de probabilidad	16
2.8.1. Distribuciones de Poisson	18
2.8.2. Distribución de probabilidad de una variable aleatoria continua: densidad de probabilidad	19
2.8.3. Distribución de Gauss o Normal	20
<b>3. Ajuste de curvas</b>	<b>23</b>
3.1. Modelos de regresión	23
3.2. Regresión Lineal Simple	25
3.3. Regresión Polinomial	27
3.3.1. Regresión lineal múltiple	28
3.3.2. Generalización de la idea de Regresión y mínimos Cuadrados	29
3.4. Métodos de Ajuste	31
3.4.1. mínimos Cuadrados	31
3.4.2. mínimos Cuadrados Ponderados	31
3.4.3. máxima Verosimilitud	32
3.5. análisis (Gráfico e histograma) de residuos	34
<b>4. Estimadores y método de máxima Verosimilitud</b>	<b>35</b>
4.1. Promedio, mediana, desviación estándar, varianza	35
4.2. Definición de máxima verosimilitud.	35

# Capítulo 1

## Medidas y errores

### 1.1. Introducción

En Física, y en general en Ciencias, no es posible determinar en forma única, con infinita precisión, el valor de una magnitud física por medio de un experimento. Todo experimento tiene asociado algún grado de error, variabilidad, o incertidumbre. Por esto, en la práctica no sólo es relevante conocer el valor de una cierta magnitud, sino también la precisión con la que se determina este valor.

Por ejemplo, el resultado de un experimento que permita medir la aceleración de gravedad en algún lugar particular de la Tierra podrá expresarse en la forma “ $x \pm \Delta x$ ” como

$$g = (9.70 \pm 0.15) \text{m/s}^2 \quad (1.1)$$

Más adelante discutiremos con mayor profundidad lo que el “error”  $\pm 0.15 \text{ m/s}^2$  significa, pero por ahora es suficiente enfatizar que este valor determina un rango de valores en el que consideramos que el experimento determina el valor de  $g$ : entre  $(9.70 - 0.15) \text{ m/s}^2 = 9.55 \text{ m/s}^2$  y  $(9.70 + 0.15) \text{ m/s}^2 = 9.85 \text{ m/s}^2$ , con algún grado de confianza (es decir, una cierta probabilidad. Por ahora, suponga que se tiene completa certeza que el valor de  $g$  está en ese intervalo, es decir, 100 % de probabilidad que el valor medido está dentro del rango). Entre más pequeño sea el valor de este error, más precisa será la medida reportada.

Una de las razones por las que debemos considerar el error asociado a las medidas, o más generalmente la incertidumbre de éstas, es que en muchas ocasiones nos interesa poner a prueba la predicción de un modelo o teoría. En otras ocasiones, nos interesará comparar nuestro resultado experimental con otro realizado en forma independiente (por otras personas, por ejemplo), y queremos saber si estos resultados pueden o no ser considerados como coincidentes. También es posible que nuestras mediciones puedan dar evidencia de la existencia de algún nuevo efecto o fenómeno físico. En todos estos casos, el valor del error de la incertidumbre de los datos reportados puede modificar drásticamente la conclusión a la que lleguemos.

Por lo tanto, el objetivo del “análisis de errores” es caracterizar las incertidumbres que toda medición tiene asociada. Esta preocupación por el análisis de errores no es de importancia sólo en experimentos de pregrado, sino que es (aún más) fundamental en la tarea del investigador.

Incluso las así llamadas “constantes universales” se determinan con cierta incertidumbre<sup>1</sup>. Por ejemplo el valor actualmente recomendado<sup>2</sup> para la constante gravitacional  $G$  es

$$G = (6.67384 \pm 0.00080) \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}, \quad (1.2)$$

correspondiendo a un error relativo de  $1.2 \times 10^{-4}$ .

#### 1.1.1. Diferentes formas de expresar el error de una medida

En la expresión

$$(\bar{x} \pm \Delta x) \text{ unidades} \quad (1.3)$$

el valor de  $\Delta x$  es llamado **error absoluto**. Otras formas equivalentes de expresar la incertidumbre en el resultado es usando:

---

<sup>1</sup>Excepto aquellas que tienen un cierto valor *por definición*. Por ejemplo, la velocidad de la luz en el vacío es hoy en día definida como  $c := 299792458 \text{ m/s}$  (exacto!).

<sup>2</sup>... por el CODATA (Committee on Data for Science and Technology). Para la recomendación 2010, ver [este link](#) y la ref. [1].

- El relativo:  $u_r = \Delta x / \bar{x}$ .
- El error porcentual:  $u_r \times 100 \% = (\Delta x / \bar{x}) \times 100 \%$ .

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

## 1.2. Precisión y exactitud

Cuando se analizan datos experimentales, existen dos conceptos que es útil distinguir: **precisión** y **exactitud**, puesto que se refieren a propiedades distintas de los datos y de las características del experimento con el que se obtuvieron. La idea cualitativa de estos conceptos es:

- **Precisión:** Se refiere a una medida de la *dispersión* de los datos asociados a una magnitud física. Una medición es “precisa” si la dispersión de los datos es pequeña, es decir si estos no se diferencian mucho entre sí. Como veremos más adelante, ésta noción está asociada al **número de cifras significativas** que representan una cantidad.
- **Exactitud:** Se refiere al grado en que los valores medidos se acercan al valor “verdadero” o al valor “aceptado” de la magnitud física en cuestión. Claramente, este concepto sólo puede aplicarse si se conoce el “valor real” de una variable (lo que en muchas ocasiones NO ocurre), o si se está comparando con algún valor aceptado por la comunidad científica (típicamente, luego de realizar experimentos independientes que intenten determinar la misma cantidad).

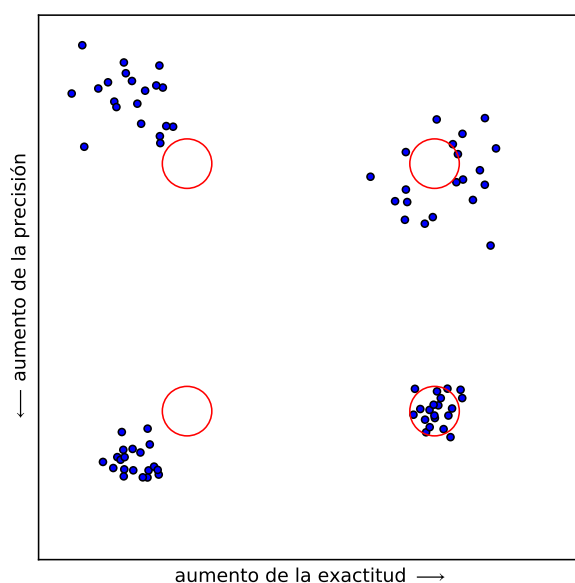


Figura 1.1: Exactitud versus precisión. Código Python [aquí](#).

Asociado a los conceptos de precisión y exactitud está la idea de clasificar las fuentes de error en (al menos) tres tipos:

- **Error aleatorio:** La propiedad definitoria de los errores aleatorios es que éstos producen resultados distintos al repetir una misma medición, incluso cuando se intentan dejar inalteradas todas las variables que determinan el resultado de éste. Al repetir una medida varias veces, se obtienen diferentes resultados con una cierta dispersión. Esta dispersión determina entonces la precisión de los datos obtenidos. Como veremos, una medida de la dispersión de los datos, y por consiguiente de la precisión de la medida, es la **desviación estándar**. Las causas que producen el error aleatorio en un experimento son

a menudo clasificadas como “ruidos técnicos” o “ruidos fundamentales”. Cada aparato tiene asociado un cierto límite de ruido fundamental determinado por la física asociada a su funcionamiento. Usualmente, sin embargo, el ruido técnico es el dominante, y puede deberse a múltiples causas (mecánicas, ambientales, variaciones en la forma en que se preparación del sistema para repetir la medición... y un largo etcétera!). Un(a) buen(a) físic@ experimental tiene experiencia y habilidad en identificar las posibles fuentes de ruido técnico e idear formas de minimizarlo.

- **Error sistemático:** Este tipo de error causa que los valores medidos se desplacen en alguna determinada dirección respecto al valor verdadero o aceptado. Por esto, el error sistemático está relacionado con la exactitud del resultado: a mayor exactitud menor es el error sistemático. Las fuentes de este tipo de error pueden ser una defectuosa calibración del instrumento de medición, o el uso de un método de medición que siempre sobre- o sub-estime la cantidad a determinar.
- **Equivocación:** Este tipo de error, en el que los métodos usados para estimar el error de una medida no son aplicables, se produce típicamente porque alguna persona se equivoca en alguna de sus tareas. Por ejemplo, escribe mal el valor marcado en el instrumento al traspassarlo a su vitácora (en papel, o en una planilla, o en un archivo con los datos, por ejemplo). También puede ocurrir este tipo de error cuando la persona no ha chequeado correctamente la puesta en marcha de su instrumento, o que éste esté midiendo en una escala distinta, etc. etc. Si bien las equivocaciones pueden ser difíciles de distinguir de los errores sistemáticos y aleatorios en la práctica, estos en principio no tienen relación directa con el sistema físico, los instrumentos o el método usado, y en principio pueden ser eliminados repitiendo cuidadosamente el experimento.

En los gráficos de la figura 1.2 se ilustra gráficamente las distintas posibilidades de error aleatorio y sistemático, por medio de **histogramas** (gráficos de barra de número de ocurrencias de las medidas repetidas). En particular, suponemos que el valor verdadero de una variable es 10, y que medimos el valor de esta variable en repetidas oportunidades, para el caso simple en que nuestro instrumento nos suministre sólo valores enteros.

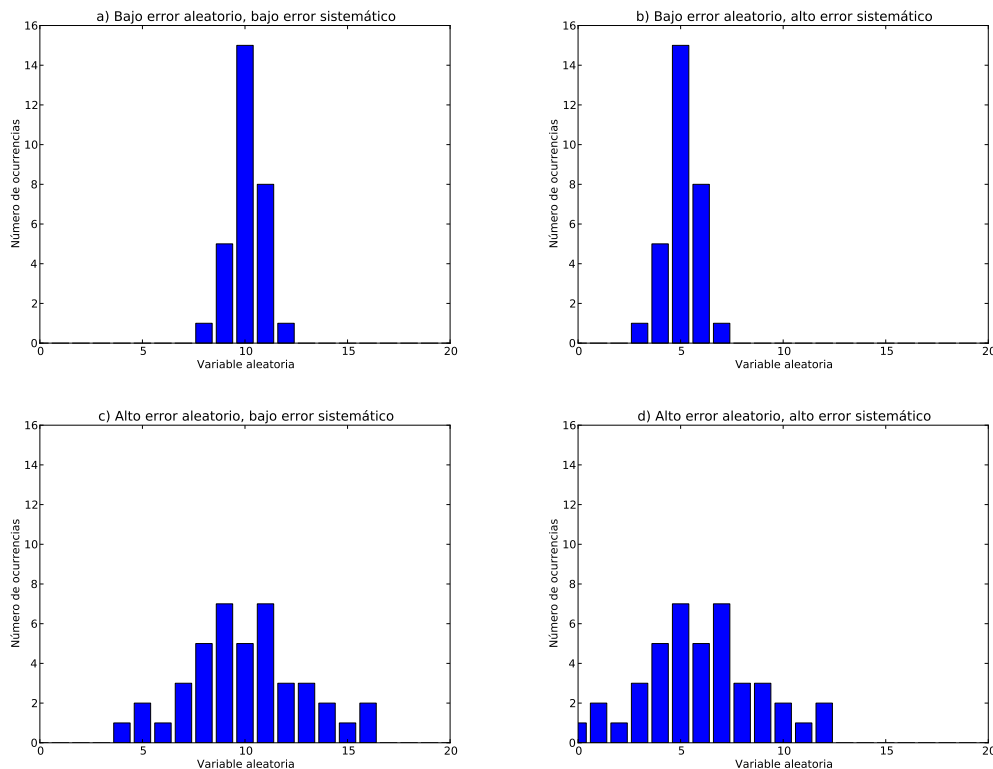


Figura 1.2: Distintas posibilidades para el tamaño del error aleatorio y sistemático. Código Python [aquí](#).

### 1.3. Promedio y desviación estándar

Cuando se realiza un experimento, generalmente se obtiene un conjunto de valores discretos  $\{x_i\}$ ,  $i = 1, \dots, N$ . En el análisis de estos datos, a menudo son de gran utilidad los conceptos de **media**, **desviación estándar**, **varianza**, **error estándar**, **suma residual de los cuadrados** y la distribución de los datos respecto del valor medio.

La medida estadística más común es la **media**, definida como

$$\bar{x} := \frac{1}{N} \sum_{i=1}^N x_i, \quad (1.4)$$

donde  $\{x_i\}$ ,  $i = 1, \dots, N$  representan los datos individuales y  $N$  el número total de éstos. Equivalentemente, podemos escribir

$$\bar{x} := \frac{1}{N} \sum_{a=1}^d n_a x_a, \quad N = \sum_{a=1}^d n_a, \quad (1.5)$$

donde  $n_a$  es el **número de veces que se repite el valor**  $x_a$  en la muestra y  $d$  es el **número de valores distintos**  $\{x_a\}$ ,  $a = 1, \dots, d$ .

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

$$s := \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}. \quad (1.6)$$

Otro símbolo usado para la misma cantidad es  $\sigma_{N-1}$ , ver por ejemplo [2].

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

$$\alpha := \frac{s}{\sqrt{N}}. \quad (1.7)$$

Si la dispersión de los datos respecto a la media aumenta,  $s$ , y  $\alpha$  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,

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2. \quad (1.8)$$

El valor del denominador,  $N-1$  en lugar de  $N$  se elige debido a que sólo existen  $N-1$  variaciones *independientes*  $(x_i - \bar{x})$ , ya que  $\sum (x_i - \bar{x}) \equiv 0$ . Formalmente decimos que se pierde un grado de libertad. Otra justificación es que la dispersión no está definida en caso de tener sólo un dato (en ese caso se obtiene 0/0).

En términos de los valores distintos  $\{x_a\}$ , y de su correspondiente número de repeticiones  $\{n_a\}$ , podemos escribir la varianza como

$$s^2 = \frac{1}{N-1} \sum_{a=1}^d n_a (x_a - \bar{x})^2. \quad (1.9)$$

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

$$c.v. := \frac{s}{\bar{x}} \times 100\%. \quad (1.10)$$

### 1.4. Cifras significativas

Al expresar una cierta magnitud física, decimos que su valor tiene un cierto número de **cifras significativas**, es decir, un cierto número de dígitos que, tomando en cuenta la incertidumbre en la determinación de su valor, entregan información relevante del valor de la magnitud en cuestión. Las cifras significativas pueden típicamente dividirse en cifras de las que se está completamente seguro, y en cifras que si bien no se tiene certeza absoluta de su valor, suministran información relevante para estimar el valor de la magnitud. Por ejemplo, la constante

gravitacional en (1.2) fue expresada usando 6 cifras significativas (6, 6, 7, 3, 8 y 4). Dada la incertidumbre en el valor de esta constante universal, sólo las tres primeras cifras son seguras, mientras que las últimas tres pueden cambiar. Sin embargo, en ese caso se ha considerado que las tres cifras no seguras también tienen información relevante de ser reportada. En general, no existe una regla absoluta y universalmente aceptada sobre cómo definir el número de cifras significativas que es necesario o útil de considerar al reportar un cierto resultado experimental. Por contraste, las siguientes reglas son univesalmente aceptadas para definir *cifras que no son significativas*:

- Todos los ceros a la izquierda de una cantidad.
- Ceros a la derecha de una cantidad, en el caso en que éstos indican solamente la escala del número (y/o las unidades de medida usadas).
- Cifras espurias que aparezcan producto de cálculos que van más allá de la precisión que los datos originales determinan y que por este motivo no suministren ninguna información físicamente relevante de la magnitud.

En el caso en que un cero a la derecha de una cantidad se considere como significativo (es decir, que suministra información relevante, ya sea porque es una cifra segura o porque si bien no es segura es la cifra que se considera más probable) se acostumbra a escribir esa cifra explícitamente. Por ejemplo, si decimos que una distancia es  $x_1 = 1,70$  km, estamos expresando que las cifras 1 y 7 y 0 son significativas, mientras que si escribimos  $x_2 = 1,7$  km estamos indicando que sólo 2 cifras son consideradas significativas. Si queremos expresar las mismas cantidades, pero en unidades de metros y no kilómetros, se acostumbra entonces escribir  $x_1 = 1,70 \times 10^3$  m (3 cifras significativas) y  $x_2 = 1,7 \times 10^3$  m (2 cifras significativas), respectivamente.

Ejemplos:

2	(1 cifra significativa)	
32	(2 cifras significativas)	
12,470	(5 cifras significativas)	
12,0010	(6 cifras significativas)	(1.11)
0,0023	(2 cifras significativas)	
000156,210	(6 cifras significativas)	
$2,3 \times 10^5$	(2 cifras significativas)	

## Ejercicio

Un(a) estudiante mide el largo de una mesa y como resultado nos entrega la siguiente cantidad  $(2,1 \pm 0,5)$  m. Este(a) estudiante, por conveniencia, decide expresar su resultado en milímetros. ¿Cómo tendría que expresar su resultado?, ¿ $(2100 \pm 50)$  mm?. **Resp.:** No, puesto que esto daría a entender que la cantidad fue expresada con 4 cifras significativas, cuando en realidad sólo tiene 2. La convención para expresar el resultado es:  $(2,1 \pm 0,5) \times 10^3$  mm.

## 1.5. 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. Además, se adoptan las siguientes reglas para *redondear* la  $n$ -ésima cifra:

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 la siguiente *convención*:
  - Si la  $n$ -ésima cifra es impar, se le agrega una unidad.
  - Si la  $n$ -ésima cifra es par la dejamos inalterable.

Esta convención intenta evitar la introducción del error sistemático que se agregaría si, por ejemplo, redondeáramos siempre “hacia arriba” la  $n$ -ésima cifra.

Note que estas convenciones de redondeo se aplican al valor principal de una magnitud física y no al error asociado, puesto que éste último siempre se redondea a la cifra superior.

## 1.6. 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).

## 1.7. Precisión, sensibilidad, y cifras significativas

En Física (y otras ciencias), estamos especialmente interesados en el último de los tres casos listados arriba, es decir, en distinguir las cifras que son relevantes a la hora de expresar el resultado de una cantidad física, de las que no lo son, *a partir de la precisión con la que medimos o calculamos la cantidad* en cuestión. Por ejemplo, si medimos una única vez el ancho de una hoja con una regla (o similar) que tiene una escala con una graduación mínima de un milímetro (decimos que la **sensibilidad del instrumento** es un 1 mm), y vemos que el valor buscado está en la zona interior entre las marcas de la regla correspondientes a 21.5 cm y 21.6 cm, podríamos expresar el resultado con 4 cifras significativas:  $(21.55 \pm 0.05)$  cm, o bien  $(215.5 \pm 0.5)$  mm. En este caso, la última cifra considerada como significativa es *estimada*, es decir, aproximada por el criterio de la persona que realiza la medida. En este ejemplo, siendo conservadores (o pesimistas) aseguraremos que (con 100 % de confianza), el ancho de la hoja está comprendido en el intervalo entre 21.5 cm y 21.6 cm de modo que (en el peor de los casos) tendremos 2 cifras seguras (2 y 1) y dos cifras que podrían variar, pero que nos entregan información física relevante a la medición. Por el contrario, en este tipo de medición no tiene ningún sentido físico reportar que el ancho de la hoja, medido con la misma regla, es 21.55275165359 cm puesto que las cifras 275165359 no son para nada confiables (la cifra anterior, 5, ya corresponde a una estimación que realizó la persona que realizó la medida). En general, toda cifra que está más a la derecha de la posición que la precisión de la medida determina es considerada como no significativa (a menos que existen muy buenos argumentos para considerar lo contrario!).

Para estimar la precisión que determina (entre otras cosas) el número de cifras significativas con las que se reporta un resultado o, equivalentemente el error o incertidumbre de la medida es útil considerar varios casos:

### 1.7.1. Caso de una única medición

En este caso se hace, a su vez, la distinción entre:

- **Medición realizada con un instrumento análogo:** Aquí decimos que la sensibilidad del instrumento es el valor de la mínima subdivisión en la escala que éste posee. Para la regla considerada en el ejemplo anterior, decimos que su sensibilidad es 1 mm. Es razonable considerar la incertidumbre o error de una medida realizada con un instrumento análogo como la mitad del valor de su sensibilidad: 0.5 mm en el caso de la regla. Esta es la “elección conservadora” para el error de la medida<sup>3</sup>
- **Medición realizada con un instrumento digital:** Suponga que mide el mismo ancho de la hoja, pero ahora usando un instrumento digital<sup>4</sup>, que suministra valores discretos en una pantalla, por ejemplo “21.57” cm. ¿Qué error asociamos a la medida?. Nuevamente, la

<sup>3</sup>En algunas ocasiones, sin embargo, esta elección puede ser claramente demasiado pesimista. Considere por ejemplo que se mide el mismo ancho de la hoja, pero con una regla que tenga una sensibilidad de 1 cm (la menor graduación marcada en ella). En este caso la medida quedará entre 21 cm y 22 cm, y el error “conservador” asociado sería de  $\pm 0.5$  cm, pudiendo expresar el resultado como  $(21.5 \pm 0.5)$  cm que, al ojo humano sano, puede parecer muy pesimista. Un(a) físic(a) experimental con experiencia y habilidad podría ser capaz de realizar una mejor estimación de la última cifra y acotar el error asociado asegurando, por ejemplo, que el ancho está entre 21.1 cm y 21.7 cm y por lo tanto informe que  $(21.4 \pm 0.3)$  cm. Dado que no existe una receta general y 100 % aceptada de cómo proceder en estos casos (que aseguren que las personas que reciben el resultado tengan confianza en él), adoptaremos aquí “pesimista”, es decir, que el error de una medición es igual a la mitad de la sensibilidad del instrumento.

<sup>4</sup>¿Cómo podría hacer esto?, ¿Con qué instrumento?



respuesta no es única, puesto que depende de cómo exactamente funciona el dispositivo, y de cómo realiza el redondeo y/o truncamiento de cifras que finalmente son mostradas en la pantalla. Algunos instrumentos digitales traen consigo especificaciones técnicas del fabricante que establecen que el error debe considerarse como “la mitad del último dígito”. En nuestro ejemplo reportaríamos entonces que el ancho es  $(21.570 \pm 0.005)\text{cm}$  (el último 0 sería una cifra significativa en este caso) o, equivalentemente, que el valor está en el intervalo entre 21.565 cm y 21.575 cm. Como podemos ver, este caso requiere que el instrumento (internamente) **redondee** (aproxime) el valor de la medida, y que no simplemente **trunque** (corte) el valor. En el caso que se desconozcan los detalles técnicos del instrumento y/o si éste aproxima o trunca las cifras es preferible, nuevamente, adoptar la postura “conservadora” y adotar como error de la medida a la unidad correspondiente al último dígito que el instrumento indica. En nuestro ejemplo,  $\pm 0.01\text{cm}$ , y reportar que  $(21.57 \pm 0.01)\text{cm}$ .

### 1.7.2. Caso de mediciones repetidas

En este caso, es necesario recurrir a herramientas estadísticas para caracterizar algunos aspectos de la precisión de los resultados, puesto que es esencial tomar en cuenta que se realizan varias (idealmente, muchas) medidas, y que éstas en general no arrojarán los mismos valores, por lo que los datos estudiados tendrán cierta dispersión. Existen distintas cantidades que son útiles para caracterizar el valor característico y la dispersión en los datos, pero que necesariamente reducen la riqueza y complejidad éstos. Teniendo esto en cuenta, es posible usar (por ejemplo) el promedio ( $\bar{x}$ ) y la desviación estándar ( $\sigma$ ) (es decir, sólo dos valores!) para expresar el valor característico y el error de una magnitud física que queremos determinar a partir de muchas medidas individuales. Hay, sin embargo, un precio que pagar: es necesario incluir además un valor para la **probabilidad de que el valor buscado esté dentro de cierto intervalo** (por ejemplo,  $(\bar{x} \pm \sigma)$ ).

## Capítulo 2

# Noción de Probabilidades

### 2.1. Introducción

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, en general la componente aleatoria sí está presente y su magnitud puede ser muy importante para obtener alguna conclusión. Por lo tanto, lo que buscamos es describir, cuantificar y modelar este tipo de variaciones.

En la figura 2.1 se muestra el esquema idealizado de un experimento, donde el resultado de un experimento depende sólo del sistema físico bajo estudio y de los parámetros iniciales del experimento. Sin embargo, en la práctica siempre existen variables no controladas que pueden

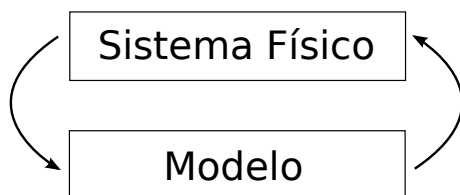


Figura 2.1: Sistema físico y modelo.

influir en el resultado (y/o puede ocurrir que el sistema bajo estudio presente propiedades intrínsecamente aleatorias, como ocurre con distintos sistemas con propiedades cuánticas), por lo que un esquema un poco más realista sería el que se muestra en la figura 2.1.



Figura 2.2: Entrada, sistema y salida.

A continuación introduciremos algunos conceptos básicos que resultan útiles en la definición y discusión de probabilidades.

#### 2.1.1. Experimento aleatorio

Un **experimento aleatorio** es aquel que proporciona diferentes resultados aún cuando se repita siempre de la “misma manera”. Esto puede ocurrir ya sea porque existen variables no controladas o desconocidas que influyen y modifican el resultado, o bien porque el fenómeno estudiado es intrínsecamente aleatorio.

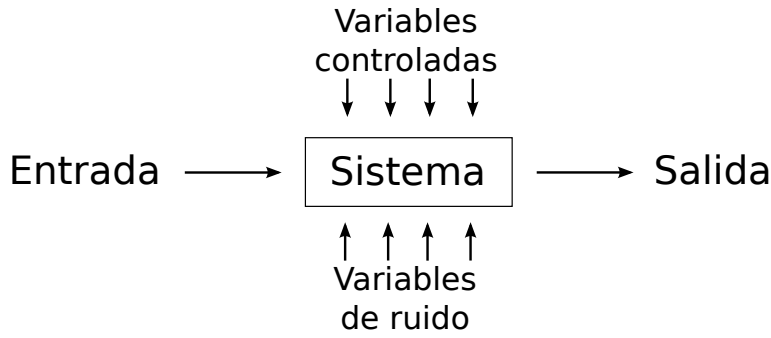


Figura 2.3: Entrada, sistema, salida y variables.

### 2.1.2. 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**.

### 2.1.3. 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.

### 2.1.4. 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**. Decimos que  $X_i$  ( $i = 1, 2, \dots$ ) son **eventos elementales** si son mutuamente excluyentes ( $X_i \cap X_j = \emptyset$ , para  $i \neq j$ ) y además  $\cup_i X_i = S$ . En otras palabras, los eventos  $X_i$  son tales que la ocurrencia de uno implica que ninguno de los otros eventos ocurre, y que cualquier otro evento no-elemental puede entenderse como la ocurrencia simultánea de ellos.

**Ejemplo 1:** Considere un experimento en el cual se miden dos variables y éstas pueden tomar dos valores,  $a$  o  $b$ . Los posibles resultados de dicho experimento serán:  $S = \{aa, ab, ba, bb\}$  y posibles eventos serán  $E_1 = \{aa\}$ ,  $E_2 = \{ab, ba\}$ ,  $E_3 = \{aa, ab, ba\}$ ,  $E_4 = \{ab, ba, bb\}$ ,  $E_5 = \{bb\}$ .

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

$$S_1 = \{12, 13, 21, 23, 31, 32\}, \quad (2.1)$$

$$S_2 = \{11, 12, 13, 21, 22, 23, 31, 32, 33\}, \quad (2.2)$$

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

## 2.2. Interpretación de la Probabilidad

Es útil cuantificar la posibilidad que se presente un resultado de un experimento aleatorio. La posibilidad de un resultado se cuantifica asignándole un número real en el 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<sup>1</sup> se basa en el modelo de la *repetición del experimento aleatorio*. Sea  $n(E)$  el número de veces que ocurre el evento  $E$  y  $N$  el número de

<sup>1</sup>Esta es la **interpretación frecuentista**. En Física también es usada la **interpretación Bayesiana**, que no discutiremos en mayor detalle en este curso.

veces que se realiza el experimento. Es intuitivamente razonable tomar como valor de  $P(E)$  a

$$P(E) := \lim_{N \rightarrow \infty} f(E) = \lim_{N \rightarrow \infty} \frac{n(E)}{N}, \quad (2.3)$$

es decir, como el valor límite de la fracción de veces que el resultado  $E$  aparece en  $N$  repeticiones del experimento aleatorio, a medida que  $N$  crece sin cota alguna.

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

## 2.3. 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,

$$P(S) = 1, \quad 0 \leq P(E) \leq 1. \quad (2.4)$$

- Si  $E_1$  y  $E_2$  son eventos excluyentes, es decir,  $E_1 \cap E_2 = \emptyset$ , entonces

$$P(E_1 \cup E_2) = P(E_1) + P(E_2). \quad (2.5)$$

### 2.3.1. Consecuencias directas de los axiomas

- La probabilidad de no obtener ningún resultado es nula:

$$P(\emptyset) = 0. \quad (2.6)$$

- La probabilidad de obtener el **evento complementario**  $E'$  a un evento  $E$ :

$$P(E') = 1 - P(E). \quad (2.7)$$

- La probabilidad de un evento  $E_1$  contenido en otro evento  $E_2$ , es decir tal que  $E_1 \subseteq E_2$ , es menor que la probabilidad de  $E_2$ :

$$P(E_1) \leq P(E_2). \quad (2.8)$$

**Ejemplo:** Considere que el espacio muestral de un experimento aleatorio es dado por  $S = \{a, b, c, d, e\}$ , donde  $a, \dots, e$  son eventos elementales con probabilidades 0.1, 0.1, 0.2, 0.4 y 0.2 respectivamente. Sean además los eventos  $A = \{a, b\}$  ( $a$  ó  $b$ ),  $B = \{c, d, e\}$  ( $c$  ó  $d$  ó  $e$ ). Determinar:

$$P(A) = 0.1 + 0.1, \quad (2.9)$$

$$P(B) = 0.2 + 0.4 + 0.2, \quad (2.10)$$

$$P(A \cup B) = 0.1 + 0.1 + 0.2 + 0.4 + 0.2, \quad (2.11)$$

$$P(A \cap B) = P(\emptyset) = 0. \quad (2.12)$$

## 2.4. Reglas de adición

Considere dos eventos  $A$  y  $B$  que no necesariamente son mutuamente excluyentes (decir, que  $A \cap B \neq \emptyset$  y por tanto en general  $P(A \cap B) \neq 0$ ). Entonces la probabilidad de obtener  $A$  o bien  $B$  es dada, ver figura 2.4, por

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (2.13)$$

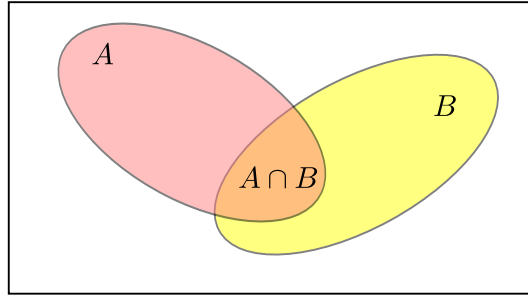


Figura 2.4: Dos conjuntos y su intersección.

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

Cuadro 2.1: Distribución de las 940 mediciones.

**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:

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

$$E_1 = \{\text{todos los resultados donde } X = B\}, \quad (2.14)$$

$$E_2 = \{\text{todos los resultados donde } Y = D\}. \quad (2.15)$$

Calcular la probabilidad de obtener como resultado los siguientes casos:

$$X = B, \quad Y = D, \quad X = B \wedge Y = D, \quad X = B \vee Y = D, \quad X \neq B \vee Y \neq D. \quad (2.16)$$

$$P(E_1) \approx \frac{358}{940}, \quad P(E_2) \approx \frac{314}{940}, \quad P(E_1 \cap E_2) \approx \frac{246}{940}, \quad (2.17)$$

$$P(E_1 \cup E_2) \approx \frac{426}{940}, \quad P((E_1 \cup E_2)') \approx \frac{514}{940}, \quad P((E_1 \cap E_2)') \approx \frac{694}{940}. \quad (2.18)$$

Otra manera de resolver el ejercicio:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) \approx \frac{358}{940} + \frac{314}{940} - \frac{246}{940} = \frac{426}{940}, \quad (2.19)$$

$$P((E_1 \cup E_2)') = 1 - P(E_1 \cup E_2) \approx 1 - \frac{426}{940} = \frac{514}{940}. \quad (2.20)$$

## 2.5. Probabilidad condicional

En muchas ocasiones la probabilidad de un evento depende de algunas condiciones, como por ejemplo si otro evento a ocurrido (antes o simultaneamente). En estos casos resulta útil el concepto de **probabilidad condicional**.

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

$$P(A|B) := \frac{P(A \cap B)}{P(B)}, \quad (2.21)$$

de modo que

$$P(A \cap B) = P(A|B) \cdot P(B). \quad (2.22)$$

En palabras: “la probabilidad de que ocurra  $A$  y  $B$  es igual a la probabilidad de que ocurra  $A$  *dado*  $B$ , multiplicada por la probabilidad de que ocurra  $B$ ”.

Decimos que  $A$  y  $B$  son **eventos independientes** si se tiene que

$$P(A|B) = P(A), \quad (2.23)$$

es decir, que el resultado de  $B$  no influye en la probabilidad de obtener  $A$ . En este caso (2.24) implica que

$$P(A \cap B) = P(A) \cdot P(B), \quad (2.24)$$

y además que

$$P(B|A) = P(B). \quad (2.25)$$

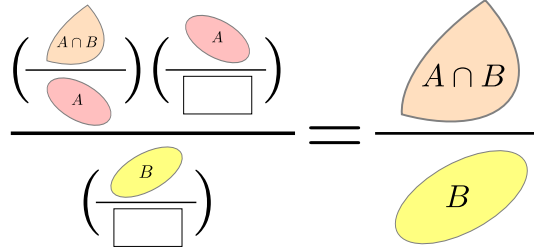


Figura 2.5: Representación diagramática de la Probabilidad condicional (2.21).

**Ejemplo:** 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:

	$A$	
$B$	Sí	No
Sí	12	18
No	24	212

Cuadro 2.2: Contaminantes en muestras de vino.

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \approx \frac{\frac{12}{266}}{\frac{36}{266}} = \frac{12}{36} = \frac{1}{3} \approx 0.33. \quad (2.26)$$

Calculemos  $P(A)$  y  $P(A|B)$  para construir un diagrama, llamado diagrama árbol, y así poder comprender mejor lo que está pasando.

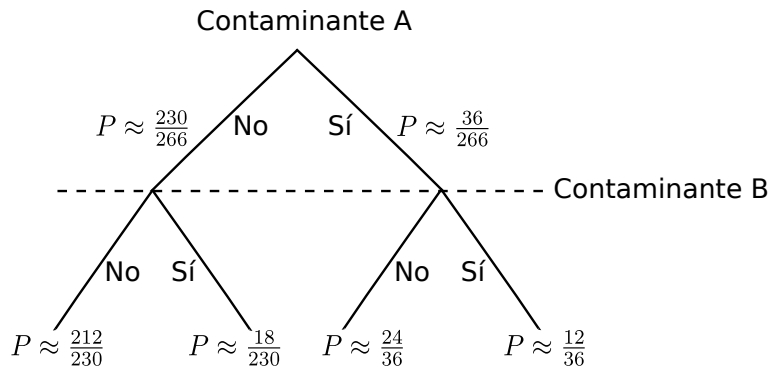


Figura 2.6: Diagrama árbol.

### 2.5.1. Reglas de multiplicación

Como vimos, la definición (2.21) de probabilidad condicional implica la relación (2.22). Esta regla de multiplicación puede ser extendida a un número finito de eventos  $A_i$ :

$$P\left(\bigcap_{i=1}^k A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|\bigcap_{i=1}^{k-1} A_i). \quad (2.27)$$

### 2.5.2. 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)

$$B = (A \cap B) \cup (A' \cap B). \quad (2.28)$$

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 obtenemos:

$$P(B) = P(A \cap B) + P(B \cap A') = P(B|A)P(A) + P(B|A')P(A'). \quad (2.29)$$

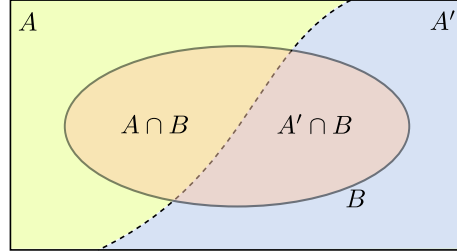


Figura 2.7: Diagrama de Venn para la probabilidad total.

### 2.5.3. Teorema de Bayes

Retomando la definición de probabilidad condicional (2.21) y las reglas de multiplicación, es directo probar que:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (2.30)$$

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. Calcular la probabilidad de encontrar Pb entre las muestras que presentaron As:  $P(\text{Pb}|\text{As})$ .

	Pb	No	Sí
As			
No		32	24
Sí		16	12

Cuadro 2.3: Pb y As en 84 muestras de vino.

Usando el teorema de Bayes tenemos que

$$P(\text{Pb}|\text{As}) = \frac{P(\text{As}|\text{Pb})P(\text{Pb})}{P(\text{As})}, \quad (2.31)$$

$$P(\text{Pb}|\text{As}) \approx \frac{\frac{12}{36} \frac{36}{84}}{\frac{28}{84}} = \frac{12}{28} = \frac{3}{7} \approx 0.43. \quad (2.32)$$

## 2.6. Variables aleatorias

**Definición:** Sea  $S = \{s_1, s_2, \dots, s_D\}$  un espacio muestral, cuyos elementos  $s_\alpha$  ( $\alpha = 1, \dots, D$ ) 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. Entonces  $S$  es el dominio de la variable aleatoria  $X$  y el rango de esta función es el conjunto de valores reales  $\{x_\alpha = X(s_\alpha)\}$ .

Decimos que la variable  $X$  es **discreta** si su rango es un conjunto discreto (finito o infinito numerable) de números reales, en caso contrario decimos que  $x$  es una variable **continua**.

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

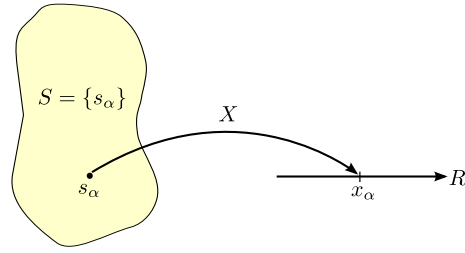


Figura 2.8: Espacio Muestral, función aleatoria y rango.

## 2.7. 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  $\{x_a\}$ ).

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.

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

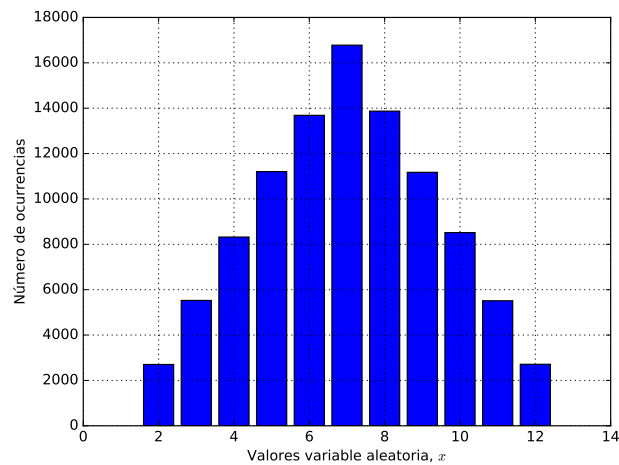


Figura 2.9: Histograma de valores de suma de dados, de acuerdo a los datos en la tabla 2.4. Código Python [aquí](#).



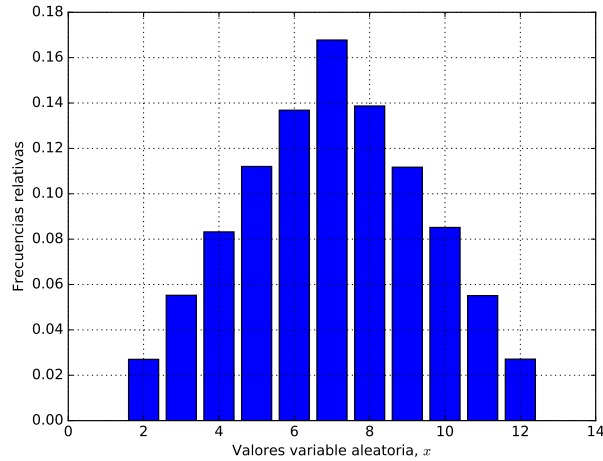


Figura 2.10: Frecuencias relativas de ocurrencia de los valores de suma de dados. Código Python [aquí](#).

## 2.8. 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  $D = 36$  posibles resultados, *que supondremos igualmente probables*, y para los cuales la variable aleatoria puede tomar  $d = 11$  valores distintos. Además, existe un cierto número de combinaciones distintas de los resultados de cada dado que generan el mismo valor de la variable aleatoria (la suma), como se muestra en la Tabla 2.5. 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 2.8, el 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 figura 2.8.

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

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. 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

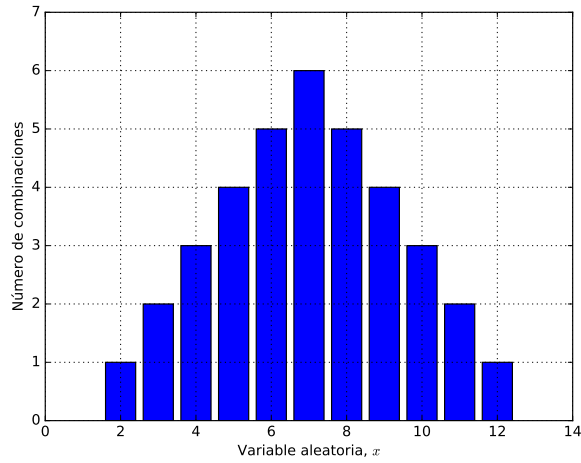


Figura 2.11: Gráfico de barras para los valores  $x$  de la variable aleatoria, de acuerdo a los datos en la tabla 2.5. Código Python [aquí](#).

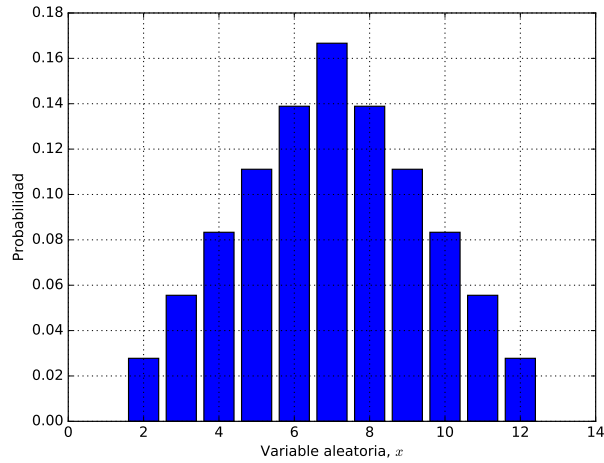


Figura 2.12: Probabilidad para las sumas de las dos caras de dos dados. Código Python [aquí](#).

a la **función de distribución acumulativa** de  $X$ , que definiremos más adelante.

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

$$p_a = P(X = x_a), \quad a = 1, \dots, d, \quad (2.33)$$

y debe satisfacer las siguientes condiciones:

$$p_a \geq 0, \quad a = 1, \dots, d, \quad (2.34)$$

$$\sum_{a=1}^d p_a = 1. \quad (2.35)$$

**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:

$$F(x) = P(X \leq x) = \sum_{x_a \leq x} p_X(x_a) = \sum_{x_a \leq x} p_a. \quad (2.36)$$

Por lo tanto, en el caso discreto, una variable aleatoria está caracterizada por la función de probabilidad  $p_a$ , la cual determina la probabilidad puntual que  $X = x_a$ , y por la función distribución acumulativa  $F(x)$ , la que representa la suma de las probabilidades puntuales hasta el valor  $x$  de  $X$ , inclusive.

### 2.8.1. Distribuciones de Poisson

**Ejemplo:** Se tiene una muestra de material radiactivo y se registra el número de partículas emitidas en los decaimientos (por ejemplo, usando un contador Geiger). El contador registra el número de decaimientos detectados y éstos tienen un comportamiento aleatorio, en el sentido que no puede ser predicho cuándo ocurrirá el siguiente decaimiento. Luego de detectadas las partículas emitidas, podemos realizar un conteo del número  $x$  de decaimientos registrados en cada intervalo de tiempo de largo  $\Delta t$  (por ejemplo, en cada intervalo de  $\Delta t = 3s$ ). Entonces  $x$  es una variable aleatoria discreta que puede asumir los valores  $x = 0, 1, 2, \dots$ , que tomarán valores distintos en cada intervalo  $\Delta t$ . Luego de muchas mediciones, podemos realizar el conteo de cuántas veces se detectaron  $x$  partículas en el intervalo  $\Delta t$ . Si denotamos como  $n(x)$  al número de veces que se detectaron  $x$  partículas (en intervalos de largo  $\Delta t$ ), entonces podemos construir un histograma  $n(x)$  versus  $x$ , y el gráfico de frecuencias relativas  $n(x)/N$  versus  $x$ , donde  $N = \sum_{x=1}^{\infty} n(x) = n(0) + n(1) + n(2) + \dots$  es el número total de mediciones realizadas (en este ejemplo, el número total de intervalos de largo  $\Delta t$  en los que se realizó el conteo). Finalmente, en el límite  $N \rightarrow \infty$  el gráfico  $n(x)/N$  versus  $x$  tiende al gráfico de distribución de probabilidad de medir  $x$  partículas en un intervalo (de largo  $\Delta t$ ). Bajo algunas hipótesis (que revisaremos a continuación) podemos modelar éste tipo de casos (y muchos otros) usando una **distribución probabilidad de Poisson**, que está caracterizada por un único parámetro  $\mu$ .

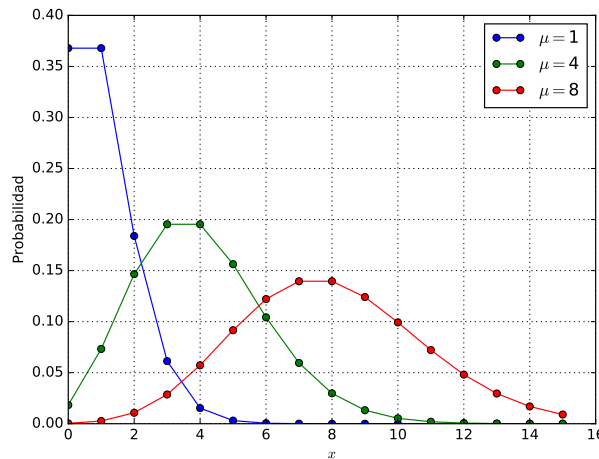


Figura 2.13: Función probabilidad de Poisson. Código Python en apéndice ??.

**Definición:** Sea  $X$  una variable aleatoria que representa el número de eventos aleatorios independientes que ocurren en un cierto intervalo (de tiempo, espacio, etc.). Se dice entonces que la variable aleatoria  $X$  tiene una **distribución de Poisson** si su función de probabilidad es dada por

$$P(x, \mu) = \frac{e^{-\mu} \mu^x}{x!}, \quad x = 0, 1, 2, 3, \dots, \quad \mu > 0. \quad (2.37)$$

El parámetro  $\mu$  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.). Esta distribución de probabilidad satisface las siguientes propiedades:

- Condición de normalización,

$$\sum_{x=0}^{\infty} P(x, \mu) = 1. \quad (2.38)$$

- Valor medio,

$$\mu_X = \sum_{x=0}^{\infty} xP(x, \mu) = \mu. \quad (2.39)$$

- Varianza,

$$\sigma^2 = \sum_{x=0}^{\infty} (x - \mu)^2 P(x, \mu) = \mu, \quad (2.40)$$

- Desviación estándar,

$$\sigma = \sqrt{\mu}. \quad (2.41)$$

En general, la distribución de probabilidad de Poisson es el modelo matemático que describe procesos en los que se satisfacen las siguientes propiedades: Se dice que el proceso es de Poisson si satisface que, al dividirse el conteo en sub-intervalos suficientemente pequeños:

- La probabilidad de *ocurrencia simultánea* de dos eventos es nula.
- Los eventos ocurren en forma *independiente*.
- La probabilidad de ocurrencia de un evento *por intervalo* (de tiempo, o distancia, etc.) es constante. Equivalentemente, la probabilidad de ocurrencia de un evento *es proporcional al intervalo* (de tiempo, o distancia, etc.) de observación.

En el ejemplo de los decaimientos radiactivos, las condiciones listadas arriba parecen hipótesis razonables, por lo que esperamos que este tipo de proceso pueda modelarse apropiadamente como un proceso de Poisson. En particular, si  $\nu$  es el **número medio de decaimientos por unidad de tiempo**<sup>2</sup> de que se produzca un decaimiento en el material, entonces  $\mu = \nu \Delta t$  es el número medio de decaimientos esperados en el intervalo de tiempo  $\Delta t$ .

**Ejemplo:** Se estima que cada año se producen aproximadamente 13000 sismos de magnitud mayor que 4 en el mundo. Suponiendo que éstos se distribuyen aleatoriamente en el año y que son independientes, deberíamos este proceso sea de Poisson. Estime la probabilidad de que se produzcan 5 sismos de estas características durante la misma hora.

### 2.8.2. Distribución de probabilidad de una variable aleatoria continua: densidad de probabilidad

Cuando la variable aleatoria puede tomar cualquier valor dentro de un cierto intervalo de números reales, diremos que la variable aleatoria es **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.  $P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f_X(u) du$ .

Esto permite interpretar  $f_X(x) dx$  como la probabilidad de que la variable  $X$  asuma un valor entre  $x$  y  $x + dx$ . Es importante recordar que  $f_X(x)$  no es directamente el valor de la probabilidad sino que es la “probabilidad por unidad de intervalo de  $x$ ”. Esto se refleja también en las unidades de medida de  $f_X(x)$ :

$$[f_X] = \frac{[P]}{[\Delta x]} = \frac{1}{[x]}. \quad (2.42)$$

Así, por ejemplo, si la variable aleatoria  $x$  representa una distancia, entonces  $[x] = L$  y  $[f_X] = L^{-1}$ .

<sup>2</sup>En Mecánica Cuántica es usual poder calcular la **probabilidad por unidad de tiempo**,  $\dot{P}_{\text{dec}}$ , que un núcleo radiactivo decaiga. En este caso,  $\nu = \mathcal{N} \dot{P}_{\text{dec}}$ , donde  $\mathcal{N}$  es el número total de núcleos presentes en la muestra analizada.

**Definición:** La **función distribución acumulada** (“cdf: cumulative distribution function”) de una variable aleatoria continua  $X$  es

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(u) du. \quad (2.43)$$

**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

$$\mu = \int_{-\infty}^{\infty} x f_X(x) dx. \quad (2.44)$$

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

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx. \quad (2.45)$$

La **desviación estándar** es entonces

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx}. \quad (2.46)$$

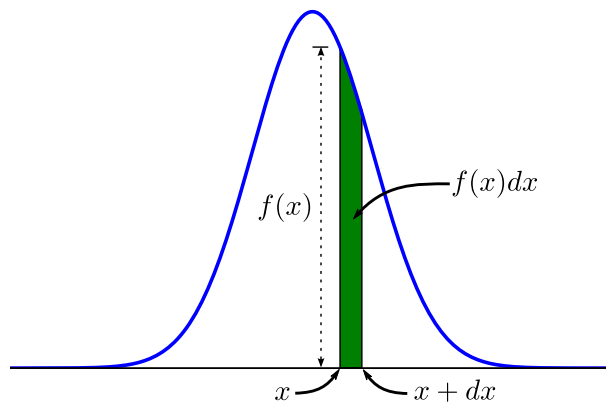


Figura 2.14: Densidad de probabilidad.

### 2.8.3. Distribución de Gauss o Normal

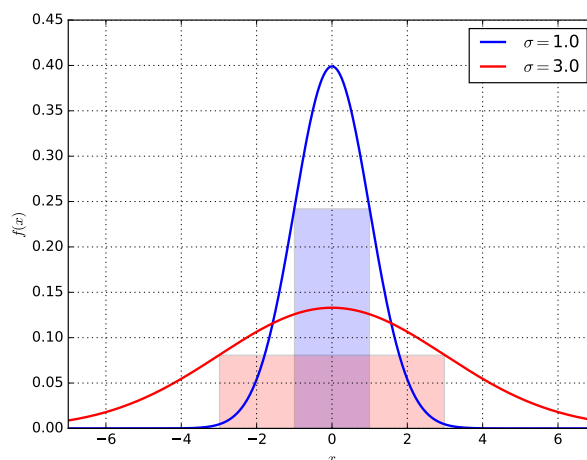


Figura 2.15: Distribución normal. Código Python [aquí](#).

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

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty, \quad (2.47)$$

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

### Propiedades

Si  $X$  es una variable aleatoria normal con valor medio  $\bar{x} = \mu$  y varianza  $\sigma_x^2 = \sigma^2$ , entonces la variable aleatoria

$$z = \frac{(x - \mu)}{\sigma} \quad (2.48)$$

es una **variable aleatoria normal** con  $\bar{z} = 0$  y varianza  $\sigma_z^2 = 1$ . 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 con distribución normal con media  $\mu$  y varianza  $\sigma^2$ . Entonces,

$$P(X \leq x) = P\left(\frac{(X - \mu)}{\sigma} \leq \frac{(x - \mu)}{\sigma}\right) = P(Z \leq z), \quad (2.49)$$

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

$$\Phi(z) = P(Z \leq z). \quad (2.50)$$

Los puntos de inflexión, la **media**, la **varianza** y la desviación estándar, usando la función distribución estandarizada, estarían dadas por

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}z^2}, \quad (2.51)$$

$$\frac{d^2 f(z)}{dz^2} \stackrel{!}{=} 0 \quad \Rightarrow \quad \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2} (z^2 - 1) = 0 \quad \Rightarrow \quad z = \pm 1 \quad \Rightarrow \quad x = \mu \pm \sigma \quad (2.52)$$

$$\bar{x} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (2.53)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma z) e^{-z^2/2} dz \quad (2.54)$$

$$= \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz + 0 \quad (2.55)$$

$$= \frac{\mu}{\sqrt{2\pi}} (\sqrt{2\pi}) \quad (2.56)$$

$$= \mu \quad (2.57)$$

$$\sigma_X^2 = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (2.58)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^2 e^{-z^2/2} dz \quad (2.59)$$

$$= -\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \frac{d}{dz} \left( e^{-z^2/2} \right) dz \quad (2.60)$$

$$= -\frac{\sigma^2}{\sqrt{2\pi}} \left[ z e^{-z^2/2} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} e^{-z^2/2} dz \right] \quad (2.61)$$

$$= -\frac{\sigma^2}{\sqrt{2\pi}} [0 - \sqrt{2\pi}] \quad (2.62)$$

$$= \sigma^2. \quad (2.63)$$

$$\sigma_x = \sqrt{\sigma_x^2} = \sigma, \tag{2.64}$$

respectivamente.

## Capítulo 3

# Ajuste de curvas

### 3.1. 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 modelar 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 3.1-3.1.

El primer intento de ajuste (ver fig. 3.1), 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. 3.1) 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



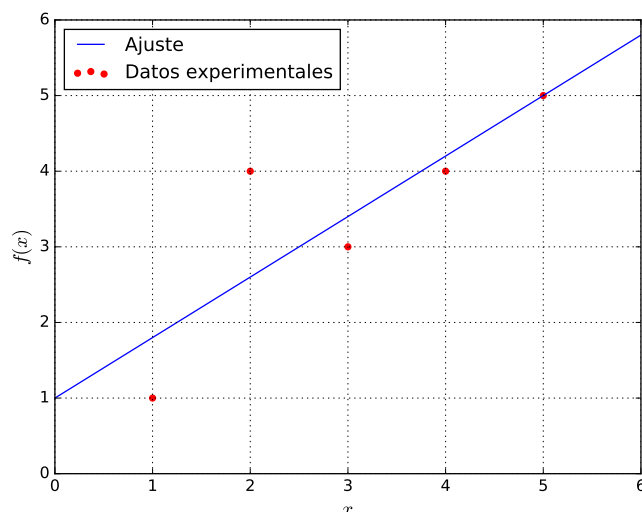


Figura 3.1: Ajuste lineal. Código Python [aquí](#).

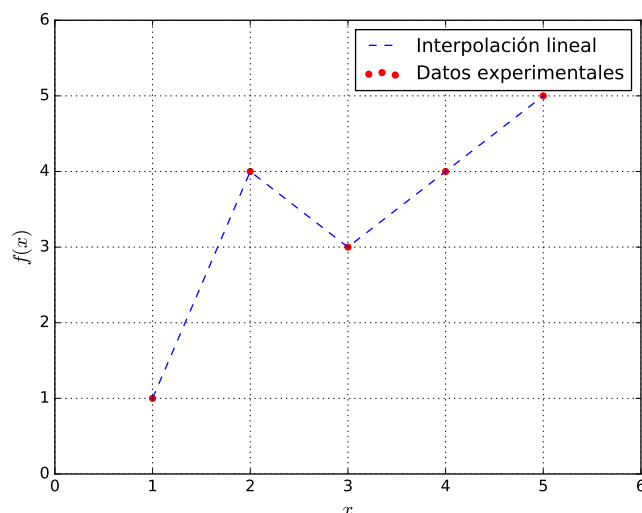


Figura 3.2: Interpolación lineal. Código Python [aquí](#).

lineal. En el tercer caso (ver fig. 3.1) 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.

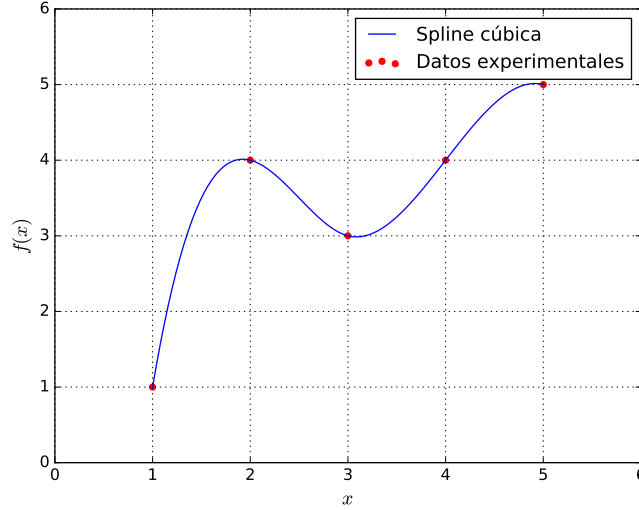


Figura 3.3: Interpolación spline cúbica. Código Python [aquí](#).

## 3.2. 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

$$E(Y) = \beta_0 + \beta_1 x, \quad (3.1)$$

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

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad (3.2)$$

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

$$y_i = b_0 + b_1 x_i + \epsilon_i, \quad i = 1, 2, \dots, N. \quad (3.3)$$

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

$$\chi^2 = \sum_{i=1}^N \epsilon_i^2 = \sum_{i=1}^N (y_i - b_0 - b_1 x_i)^2. \quad (3.4)$$

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

$$\frac{\partial \chi^2}{\partial b_0} = -2 \sum_{i=1}^N (y_i - b_0 - b_1 x_i) = 0, \quad (3.5)$$

$$\frac{\partial \chi^2}{\partial b_1} = -2 \sum_{i=1}^N (y_i - b_0 - b_1 x_i) x_i = 0. \quad (3.6)$$

Simplificando y reordenando términos, obtenemos:

$$Nb_0 + b_1 \sum_{i=1}^N x_i = \sum_{i=1}^N y_i, \quad (3.7)$$

$$b_0 \sum_{i=1}^N x_i + b_1 \sum_{i=1}^N x_i^2 = \sum_{i=1}^N y_i x_i, \quad (3.8)$$

$$\begin{pmatrix} N & \sum_{i=1}^N x_i \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N y_i x_i \end{pmatrix}. \quad (3.9)$$

Resolviendo el sistema se obtiene

$$b_0 = \bar{y} - b_1 \bar{x}, \quad (3.10)$$

$$b_1 = \frac{N \left( \sum_{i=1}^N y_i x_i \right) - \left( \sum_{i=1}^N y_i \right) \left( \sum_{i=1}^N x_i \right)}{N \left( \sum_{i=1}^N x_i^2 \right) - \left( \sum_{i=1}^N x_i \right)^2} = \frac{\overline{xy} - (\bar{x})(\bar{y})}{\overline{x^2} - (\bar{x})^2}, \quad (3.11)$$

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

$$S_{Y/x} = \sqrt{\frac{S_r}{N-2}}, \quad (3.12)$$

con

$$S_r = \sum_{i=1}^N (y_i - b_0 - b_1 x_i)^2. \quad (3.13)$$

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

$$r^2 = \frac{S_t - S_r}{S_t}, \quad (3.14)$$

donde

$$S_t = \sum_{i=1}^N (y_i - \bar{y})^2, \quad (3.15)$$

$$S_r = \sum_{i=1}^N (y_i - b_0 - b_1 x_i)^2. \quad (3.16)$$

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 3.1

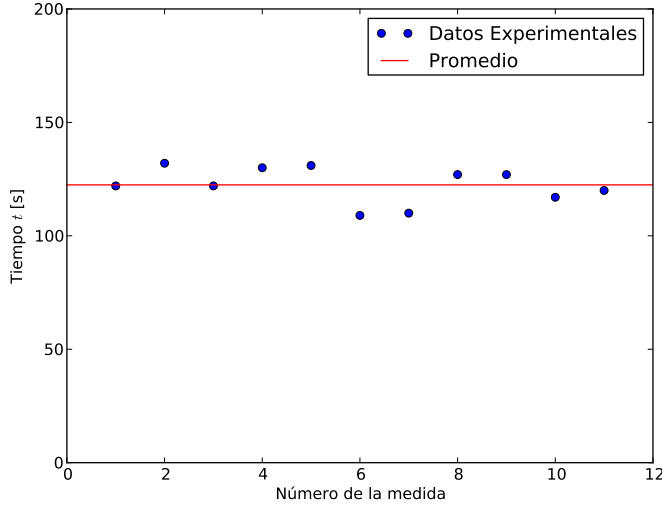


Figura 3.4: El ajuste por una constante es equivalente a calcular el promedio. Código Python en apéndice ??.

### 3.3. 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  $m$ :

$$y = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m. \quad (3.17)$$

En este caso

$$S_r = \sum_{i=1}^N (y_i - a_0 - a_1x_i - a_2x_i^2 - \cdots - a_mx_i^m)^2, \quad (3.18)$$

$$\frac{\partial \chi^2}{\partial a_0} = -2 \sum_{i=1}^N (y_i - a_0 - a_1x_i - a_2x_i^2 - \cdots - a_mx_i^m), \quad (3.19)$$

$$\frac{\partial \chi^2}{\partial a_1} = -2 \sum_{i=1}^N x_i (y_i - a_0 - a_1x_i - a_2x_i^2 - \cdots - a_mx_i^m), \quad (3.20)$$

$$\frac{\partial \chi^2}{\partial a_2} = -2 \sum_{i=1}^N x_i^2 (y_i - a_0 - a_1x_i - a_2x_i^2 - \cdots - a_mx_i^m), \quad (3.21)$$

$$\vdots = \vdots, \quad (3.22)$$

$$\frac{\partial \chi^2}{\partial a_m} = -2 \sum_{i=1}^N x_i^m (y_i - a_0 - a_1x_i - a_2x_i^2 - \cdots - a_mx_i^m). \quad (3.23)$$

Igualando a cero y reordenando términos, obtenemos

$$a_0 N + a_1 \sum_{i=1}^N x_i + a_2 \sum_{i=1}^N x_i^2 + \cdots + a_m \sum_{i=1}^N x_i^m = \sum_{i=1}^N y_i, \quad (3.24)$$

$$a_0 \sum_{i=1}^N x_i + a_1 \sum_{i=1}^N x_i^2 + a_2 \sum_{i=1}^N x_i^3 + \cdots + a_m \sum_{i=1}^N x_i^{m+1} = \sum_{i=1}^N x_i y_i, \quad (3.25)$$

$$a_0 \sum_{i=1}^N x_i^2 + a_1 \sum_{i=1}^N x_i^3 + a_2 \sum_{i=1}^N x_i^4 + \cdots + a_m \sum_{i=1}^N x_i^{m+2} = \sum_{i=1}^N x_i^2 y_i, \quad (3.26)$$

$$\vdots = \vdots, \quad (3.27)$$

$$a_0 \sum_{i=1}^N x_i^m + a_1 \sum_{i=1}^N x_i^{m+1} + a_2 \sum_{i=1}^N x_i^{m+2} + \cdots + a_m \sum_{i=1}^N x_i^{2m} = \sum_{i=1}^N x_i^m y_i. \quad (3.28)$$

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:

$$\begin{pmatrix} N & \sum x_i & \sum x_i^2 & \cdots & \sum x_i^m \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \cdots & \sum x_i^{m+1} \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \cdots & \sum x_i^{m+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_i^m & \sum x_i^{m+1} & \sum x_i^{m+2} & \cdots & \sum x_i^{2m} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \\ \vdots \\ \sum x_i^m y_i \end{pmatrix}. \quad (3.29)$$

**Ejemplo:** Ajustar un polinomio de grado 2, para la siguiente tabla de valores [3.1](#).

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 3.1: Posición de un movimiento unidimensional como función del tiempo.

### 3.3.1. Regresión lineal múltiple

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

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m. \quad (3.30)$$

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

$$S_r = \sum_{i=1}^N (y_i - a_0 - a_1 x_{i1} - a_2 x_{i2} - \cdots - a_m x_{im})^2. \quad (3.31)$$

En este caso las ecuaciones para los estimadores quedan determinadas por

$$\frac{\partial \chi^2}{\partial a_0} = -2 \sum_{i=1}^N (y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i} - \cdots - a_m x_{m,i}), \quad (3.32)$$

$$\frac{\partial \chi^2}{\partial a_1} = -2 \sum_{i=1}^N x_{1,i} (y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i} - \cdots - a_m x_{m,i}), \quad (3.33)$$

$$\frac{\partial \chi^2}{\partial a_2} = -2 \sum_{i=1}^N x_{2,i} ((y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i} - \cdots - a_m x_{m,i}), \quad (3.34)$$

$$\vdots = \vdots \quad (3.35)$$

$$\frac{\partial \chi^2}{\partial a_m} = -2 \sum_{i=1}^N x_{m,i} ((y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i} - \cdots - a_m x_{m,i}). \quad (3.36)$$

Igualando a cero y reordenando términos, obtenemos

$$a_0 N + a_1 \sum_{i=1}^N x_{1,i} + a_2 \sum_{i=1}^N x_{2,i} + \cdots + a_m \sum_{i=1}^N x_{m,i} = \sum_{i=1}^N y_i, \quad (3.37)$$

$$a_0 \sum_{i=1}^N x_{1,i} + a_1 \sum_{i=1}^N x_{1,i}^2 + a_2 \sum_{i=1}^N x_{1,i} x_{2,i} + \cdots + a_m \sum_{i=1}^N x_{1,i} x_{m,i} = \sum_{i=1}^N x_{1,i} y_i, \quad (3.38)$$

$$a_0 \sum_{i=1}^N x_{2,i} + a_1 \sum_{i=1}^N x_{2,i} x_{1,i} + a_2 \sum_{i=1}^N x_{2,i}^2 + \cdots + a_m \sum_{i=1}^N x_{2,i} x_{m,i} = \sum_{i=1}^N x_{2,i} y_i, \quad (3.39)$$

$$\vdots = \vdots \quad (3.40)$$

$$a_0 \sum_{i=1}^N x_{m,i} + a_1 \sum_{i=1}^N x_{m,i} x_{1,i} + a_2 \sum_{i=1}^N x_{m,i} x_{2,i} + \cdots + a_m \sum_{i=1}^N x_{m,i}^2 = \sum_{i=1}^N x_{m,i} y_i. \quad (3.41)$$

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

$$\begin{pmatrix} N & \sum x_{1,i} & \sum x_{2,i} & \cdots & \sum x_{m,i} \\ \sum x_{1,i} & \sum x_{1,i} x_{1,i} & \sum x_{1,i} x_{2,i} & \cdots & \sum x_{1,i} x_{m,i} \\ \sum x_{2,i} & \sum x_{2,i} x_{1,i} & \sum x_{2,i} x_{2,i} & \cdots & \sum x_{2,i} x_{m,i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{m,i} & \sum x_{m,i} x_{1,i} & \sum x_{m,i} x_{2,i} & \cdots & \sum x_{m,i} x_{m,i} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_{1,i} y_i \\ \sum x_{2,i} y_i \\ \vdots \\ \sum x_{m,i} y_i \end{pmatrix} \quad (3.42)$$

**Ejemplo:** Determine el “mejor” plano que ajusta los valores en la tabla 3.2 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 3.2: Datos para la Tarea: valores de  $x$ ,  $y$  y  $z$ .

### 3.3.2. 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  $y$ .

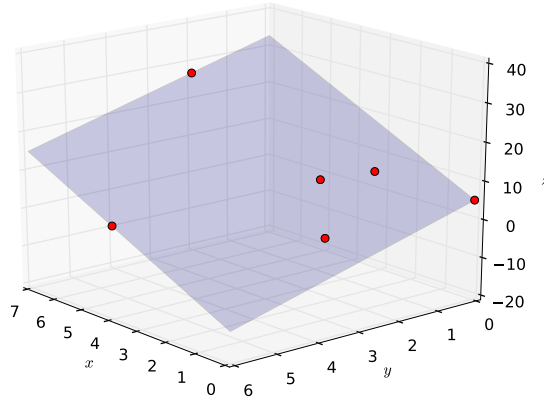


Figura 3.5: Ajuste por mínimos cuadrados de los datos por un plano. Código Python en apéndice ???. En este caso ajustamos una función  $z = a_0 + a_1x + a_2y$ . Los coeficientes resultantes son  $a_0 = 5.0$ ,  $a_1 = 4.0$  y  $a_2 = -3.0$ . El coeficiente de determinación resulta ser  $r^2 = 1$ .

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:

$$Y = f(x) + \epsilon, \quad (3.43)$$

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:

$$\chi^2 = \sum_{i=1}^N (y_i - f(x_i))^2. \quad (3.44)$$

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.

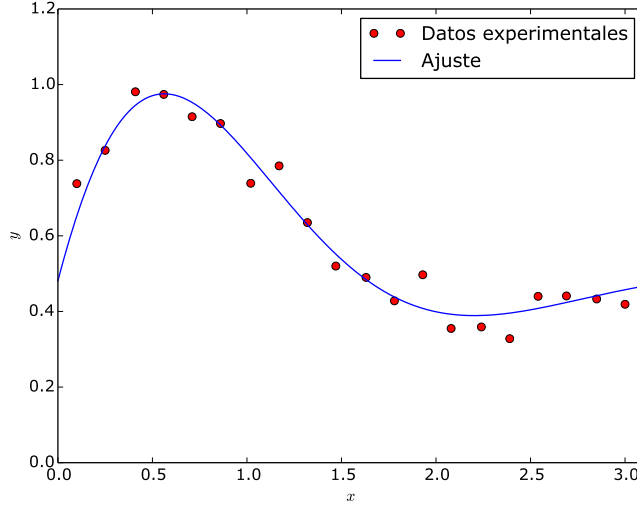


Figura 3.6: Ajuste no-polinomial de datos usando el método de mínimos cuadrados. Código Python en apéndice ??.

## 3.4. Métodos de Ajuste

### 3.4.1. mínimos Cuadrados

### 3.4.2. 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 3.5), 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

$$S_r = \sum_{i=1}^N w_i (y_i - f(x_i))^2. \quad (3.45)$$

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

$$b_0 = \frac{\left(\sum_{i=1}^N w_i y_i\right) \left(\sum_{i=1}^N w_i x_i^2\right) - \left(\sum_{i=1}^N w_i x_i y_i\right) \left(\sum_{i=1}^N w_i x_i\right)}{\left(\sum_{i=1}^N w_i\right) \left(\sum_{i=1}^N w_i x_i^2\right) - \left(\sum_{i=1}^N w_i x_i\right)^2}, \quad (3.46)$$



$$b_1 = \frac{\left(\sum_{i=1}^N w_i\right) \left(\sum_{i=1}^N w_i x_i y_i\right) - \left(\sum_{i=1}^N w_i x_i\right) \left(\sum_{i=1}^N w_i y_i\right)}{\left(\sum_{i=1}^N w_i\right) \left(\sum_{i=1}^N w_i x_i^2\right) - \left(\sum_{i=1}^N w_i x_i\right)^2}. \quad (3.47)$$

### Ajuste polinomial

$$\begin{pmatrix} \sum w_i & \sum w_i x_i & \sum w_i x_i^2 & \cdots & \sum w_i x_i^m \\ \sum w_i x_i & \sum w_i x_i^2 & \sum w_i x_i^3 & \cdots & \sum w_i x_i^{m+1} \\ \sum w_i x_i^2 & \sum w_i x_i^3 & \sum w_i x_i^4 & \cdots & \sum w_i x_i^{m+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum w_i x_i^m & \sum w_i x_i^{m+1} & \sum w_i x_i^{m+2} & \cdots & \sum w_i x_i^{2m} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \sum w_i y_i \\ \sum w_i x_i y_i \\ \sum w_i x_i^2 y_i \\ \vdots \\ \sum w_i x_i^m y_i \end{pmatrix}. \quad (3.48)$$

### Ajuste lineal multiple

$$\begin{pmatrix} \sum w_i & \sum w_i x_{1,i} & \sum w_i x_{2,i} & \cdots & \sum w_i x_{m,i} \\ \sum w_i x_{1,i} & \sum w_i x_{1,i} x_{1,i} & \sum w_i x_{1,i} x_{2,i} & \cdots & \sum w_i x_{1,i} x_{m,i} \\ \sum w_i x_{2,i} & \sum w_i x_{2,i} x_{1,i} & \sum w_i x_{2,i} x_{2,i} & \cdots & \sum w_i x_{2,i} x_{m,i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum w_i x_{m,i} & \sum w_i x_{m,i} x_{1,i} & \sum w_i x_{m,i} x_{2,i} & \cdots & \sum w_i x_{m,i} x_{m,i} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \sum w_i y_i \\ \sum w_i x_{1,i} y_i \\ \sum w_i x_{2,i} y_i \\ \vdots \\ \sum w_i x_{m,i} y_i \end{pmatrix}. \quad (3.49)$$

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 3.3: 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  $w_i = 1/(\Delta y_i)^2$ ).

### 3.4.3. 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,

$$Y = f(x) + \epsilon, \quad (3.50)$$

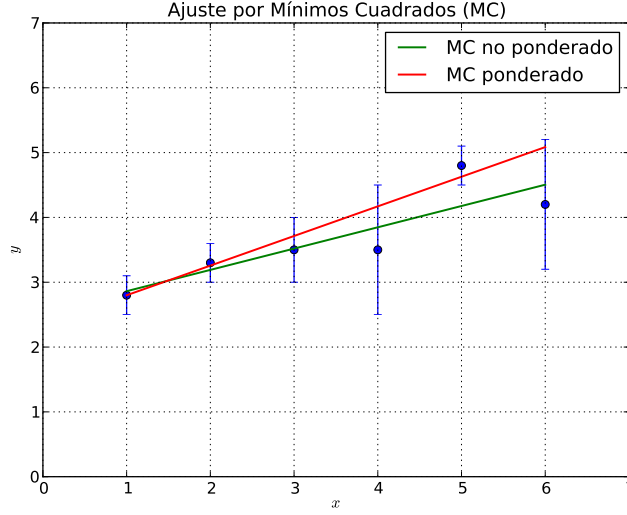


Figura 3.7: Ajuste por mínimos cuadrados con y sin ponderación. Código Python en apéndice ??.

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

$$L(b_0, b_1, \dots, \sigma^2) = \prod_{i=1}^N g(y_i - f(x_i)). \quad (3.51)$$

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

$$\frac{\partial \ln L(b_0, \dots, b_i, \dots, b_N, \sigma^2)}{\partial b_0} = 0, \quad (3.52)$$

$$\vdots = 0, \quad (3.53)$$

$$\frac{\partial \ln L(b_0, \dots, b_i, \dots, b_N, \sigma^2)}{\partial b_i} = 0, \quad (3.54)$$

$$\vdots = 0, \quad (3.55)$$

$$\frac{\partial \ln L(b_0, \dots, b_i, \dots, b_N, \sigma^2)}{\partial b_N} = 0. \quad (3.56)$$

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

$$\frac{\partial \ln L(b_0, \dots, b_i, \dots, b_N, \sigma^2)}{\partial \sigma^2} = 0. \quad (3.57)$$

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  $b_1, b_1, \dots, b_N$  que mínimos cuadrados.

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

$$f(\epsilon, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{\epsilon}{\sigma}\right)^2}. \quad (3.58)$$

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

$$L(b_0, b_1, \sigma^2) = \prod_{i=1}^N f(y_i - b_0 - b_1 x_i) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \left( \frac{y_i - b_0 - b_1 x_i}{\sigma} \right)^2}. \quad (3.59)$$

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

$$\frac{\partial \ln L(b_0, b_1, \sigma^2)}{\partial b_0} = 0, \quad (3.60)$$

$$\frac{\partial \ln L(b_0, b_1, \sigma^2)}{\partial b_1} = 0, \quad (3.61)$$

$$\frac{\partial \ln L(b_0, b_1, \sigma^2)}{\partial \sigma^2} = 0, \quad (3.62)$$

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 método de mínimos cuadrados.

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

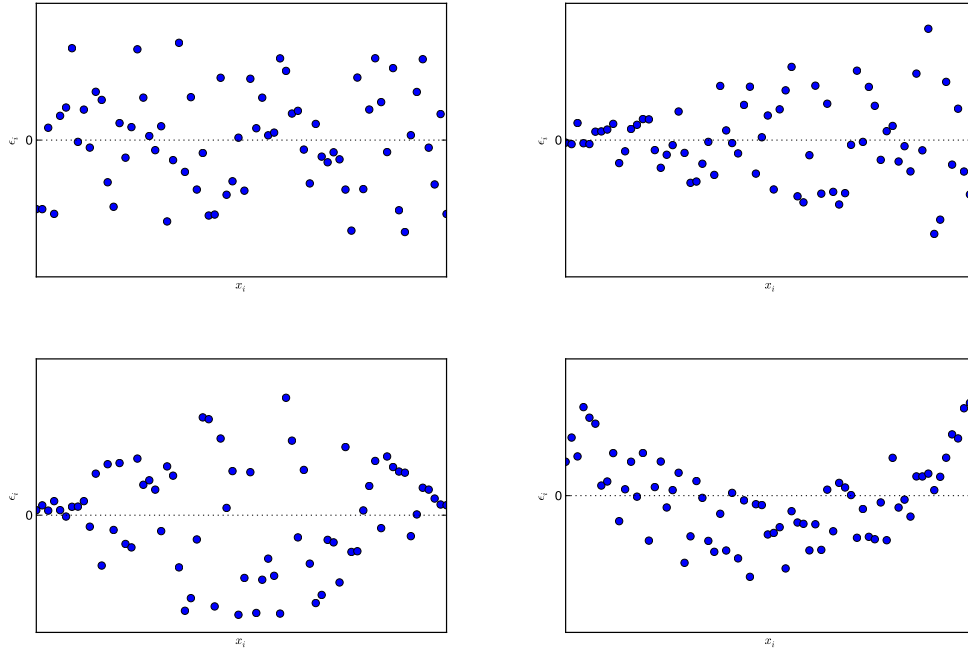


Figura 3.8: Ejemplos de distintas posibilidades para la dispersión.

## Capítulo 4

# 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).

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

### 4.2. 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

$$L(\theta) := f(x_1, \theta)f(x_2, \theta) \cdots f(x_N, \theta) \quad (4.1)$$

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

$$L(\theta) = P(X_1 = x_1, X_2 = x_2, \dots, X_N = x_N), \quad (4.2)$$

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

$$L(\mu, \sigma) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2} \quad (4.3)$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 \right]. \quad (4.4)$$

Calculando el logaritmo natural de ambos lados, encontramos

$$\ln L(\mu) = -\frac{N}{2} \ln(2\pi\sigma^2) - (2\sigma^2)^{-1} \sum_{i=1}^N (x_i - \mu)^2. \quad (4.5)$$

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

$$\frac{d \ln L(\mu)}{d\mu} = (\sigma^2)^{-1} \sum_{i=1}^N (x_i - \mu) \stackrel{!}{=} 0, \quad (4.6)$$

que tiene como solución para  $\mu$  a

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i = \bar{x}, \quad (4.7)$$

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 (4.3). Calculando ahora las derivadas parciales de (4.5) respecto a  $\mu$  y  $\sigma^2$ , e igualándolas a cero, encontramos

$$\frac{\partial \ln L(\mu, \sigma)}{\partial \mu} = (\sigma^2)^{-1} \sum_{i=1}^N (x_i - \mu) \stackrel{!}{=} 0, \quad (4.8)$$

$$\frac{\partial \ln L(\mu, \sigma)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^N (x_i - \mu)^2 \stackrel{!}{=} 0, \quad (4.9)$$

que tiene por soluciones a

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i = \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2. \quad (4.10)$$

\*\*\* 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.\*\*\*

# Bibliografía

- [1] Mohr *et al.*, *Rev. Mod. Phys.* **84**, 1527 (2012) <http://dx.doi.org/10.1103/RevModPhys.84.1527>.
- [2] I.G. Hughes and T.P.A. Hase, *Measurements and their Uncertainties: A practical guide to modern error analysis*, Oxford University Press (2010).