

# Apuntes de Estadística

Daniel Vázquez Lago

February 2, 2023



# Contents

<b>1</b>	<b>Estadística descriptiva</b>	<b>7</b>
1.1	Introducción . . . . .	8
1.2	Medidas características de una distribución de frecuencias . . . . .	10
1.2.1	Momentos de una distribución de frecuencias . . . . .	11
1.2.2	Medidas de centralización . . . . .	11
1.2.3	Medidas de dispersión . . . . .	14
1.2.4	Medidas de asimetría . . . . .	16
1.2.5	Medidas de apuntamiento o curtosis . . . . .	17
1.3	Trasformaciones de una variable aleatoria . . . . .	18
1.4	Trasformaciones de una variable aleatoria . . . . .	19
1.4.1	Distribuciones de frecuencia multivariable . . . . .	19
<b>2</b>	<b>Teoría de probabilidades</b>	<b>23</b>
2.1	Fundamentos de probabilidad . . . . .	23
2.1.1	Función de probabilidad, densidad de probabilidad y distribución de probabilidad . . . . .	25
2.1.2	Normalización de una distribución continua: . . . . .	26
2.1.3	Esperanza: . . . . .	26
2.1.4	Medidas de centralización: . . . . .	26
2.1.5	Medidas de dispersión: . . . . .	27
2.1.6	Tipificación de la variable aleatoria: . . . . .	27
2.1.7	Momentos de la distribución de probabilidad . . . . .	27
2.1.8	Función generatriz de momentos . . . . .	27
2.1.9	Transformación de la variable aleatoria: . . . . .	29
2.1.10	Distribución binomial: . . . . .	30
2.1.11	Distribución de Pascal: . . . . .	31
2.1.12	Distribución de Poisson: . . . . .	31
2.1.13	Distribución Gaussiana: . . . . .	32
2.1.14	Distribuciones multivariadas . . . . .	33
2.1.15	Sumando variables aleatorias y Teorema del límite central: . . . . .	34
<b>3</b>	<b>Métodos estadísticos</b>	<b>35</b>
3.0.1	Estimación puntual de parámetros . . . . .	35
3.0.2	Intervalos de confianza . . . . .	39
3.0.3	Test de hipótesis estadísticas . . . . .	41
3.0.4	Método de máxima verosimilitud . . . . .	43

<b>4</b>	<b>Incertidumbre en la medida</b>	<b>49</b>
4.0.1	Introducción: error e incertidumbre . . . . .	49
4.0.2	Incertidumbre tipo A y tipo B . . . . .	50
4.0.3	Incertidumbre combinada . . . . .	51
4.0.4	Sistema Internacional de Unidades . . . . .	53
4.0.5	Cifras significativas . . . . .	53

# Introducción

En este texto trataremos temas como la estadística y probabilidad (partiendo desde lo más básico a conceptos mas complejos) de una manera teórica para luego poder avanzar y entender conceptos mas complicados, que quizá hayáis visto pero su mecanismo teórico es mucho mas profundo del que parece.

También analizaremos como hacer memorias, ejemplos de como estructurarlas, que hay que hacer con las incertidumbres, como se usan y combinan, ya que hay una gran falta de información de como proceder en las memorias para expresar correctamente los datos.

En general este texto será una introducción relativamente avanzada de como hacer una memoria, que factores hay que tener en cuenta, y cual es la razón teórica para hacer lo que hacemos.



# Chapter 1

## Estadística descriptiva

En las ciencias experimentales se basan en la contrastación de hipótesis con las medidas empíricas. Cuando se trata de determinar el valor de una magnitud física es inevitable cierto nivel de variabilidad o imprecisión, que puede venir dada por diversos motivos: imperfección de nuestros instrumentos, la propia fluctuación del mensurando o en la existencia de perturbaciones que no podemos controlar. Es necesario comprender que en cualquier conjunto de medidas de una propiedad física realizadas con instrumentos habrá una inevitable variación entre los valores de las medidas.

Supongamos que queremos calcular la masa de un objeto. Para determinarla necesitamos o bien una balanza con una masa patrón o una báscula. Supongamos que usamos una báscula con la precisión suficiente, donde obtenemos:

$$m \text{ (kg)} = \{1.00043; 1.00045; 1.00048; 1.00039; 1.00043\}$$

Ahora nos podemos hacer las siguientes preguntas: ¿Cuál es el valor que realmente podemos asignar a la masa? ¿Coincide con alguno de esos valores? ¿Realmente podemos calcular el valor real? ¿Con que nivel de certidumbre podemos concluir que conocemos el valor de la masa del cuerpo a partir del experimento anterior?

El origen de esta variabilidad en la determinación experimental de las magnitudes físicas es inherente a la naturaleza de los objetos, imputable a las limitaciones de nuestra capacidad de observación (precisión de los instrumentos...). El objeto de este capítulo es la introducción de algunas de las herramientas matemáticas desarrolladas en el ámbito de la Estadística matemática para la descripción de la información contenida en los datos y su posterior utilización para la evaluación de la incertidumbre asociada al proceso experimental. En concreto nos ocuparemos de los conjuntos de datos obtenidos del colectivo objeto de estudio mediante algún experimento (aleatorio) y que posteriormente pretende analizarse.

## 1.1 Introducción

En el análisis de una determinada propiedad física debemos especificar en primer lugar el conjunto de elementos o unidades estadísticas que manifiestan dicha cualidad. A dicho conjunto lo llamaremos *población*. Estas poblaciones pueden ser finitas o infinitas, aunque casi nunca disponemos de una población entera, normalmente estudiamos subconjuntos de las poblaciones, las denominadas *muestras*.

### Ejemplo 1.1

---

Si nuestra magnitud de interés es la vida media de un isótopo radioactivo, nos interesaría conocer todas las desintegraciones radioactivas de este isótopo a lo largo del tiempo. Eso sería nuestra población. Sin embargo estudiamos solo una cantidad limitada de este isótopo durante un intervalo de tiempo finito, lo que es la muestra

---

En general es en las muestras donde se realizan los experimentos aleatorios de edición de las diferentes cantidades de interés. A partir de la información suministrada por los experimentos sobre las muestras (supondremos que son representativas de la población) podremos obtener propiedades de las poblaciones.

Para el análisis asignamos a cada magnitud o característica de la población un determinado valor de tipo *cualitativo* (valores no numéricos) o de tipo *cuantitativo* (valor numérico). A su vez los valores cuantitativos pueden ser *discretos*, cuando pueden tomar un número finito o infinito pero numerable (por ejemplo cantidad de veces que ocurre un suceso, que va desde el 1 hasta el infinito), o *continuo*, cuando es posible que adopten cualquier valor dentro de un intervalo de la recta real.

Una vez que escogemos nuestra muestra y cual es su característica cuantitativa, debemos realizar un experimento que debe cumplir los siguientes requisitos:

1. Todos los posibles resultados de los ensayos del experimento son conocidos con anterioridad (sabemos que resultados numéricos pueden haber, el intervalo...)
2. No se puede predecir el resultado de los ensayos concretos
3. Los diferentes ensayos pueden repetirse en condiciones idénticas

Un ejemplo de un experimento aleatorio es el lanzamiento de una moneda, y dado que sabemos la variabilidad intrínseca de los experimentos al medir magnitudes podemos poner como ejemplo de un experimento aleatorio la determinación del número de desintegraciones en intervalos de tiempo consecutivos con un contador Geiger.

Consideremos un experimento donde realizamos  $N$  ensayos de la medición de una determinada característica que se le asocia una variable  $X$ . Tras la realización del experimento se habrá obtenido un conjunto de datos que representamos como:



$$\{x_i\}_{i=1}^N = \{x_1, x_2, \dots, x_N\}$$

### Ejemplo 1.2

---

En un experimento de determinación de la ecuación de estado  $P = P(V, T)$  de un gas donde se miden simultáneamente volumen y temperatura, suponemos que la variable estadística analizada  $\vec{X} = (V, T)$  es bidimensional.

---

Una vez obtenidos los resultados del experimento aleatorio llegamos a una fase crucial del proceso de descripción del fenómeno en cuestión, puesto que debemos proceder a organizar la información obtenida.

A partir de la muestra de  $N$  datos lo único que podemos afirmar es cuántas veces hemos obtenido cada uno de los valores posibles de la variable aleatoria analizada. Este número nos conduce al concepto estadístico fundamental que es la *distribución de frecuencia*.

Se denomina **frecuencia absoluta**,  $(n_i)$ , asociada al valor  $i$ -ésimo de los resultados de la variable aleatoria, al número de veces que se observa el valor en el total de las medidas realizadas. La suma de las frecuencias absolutas sobre todos los resultados es  $N$ :

$$\sum_{i=1}^l n_i = N$$

Se denomina **frecuencia relativa**,  $(f_i)$ , asociada al valor  $i$  de los resultados de la variable aleatoria, a la fracción de veces que se observa el valor en el total de las medidas realizadas  $f_i = \frac{n_i}{N}$ . La suma de las frecuencias relativas sobre todos los resultados es 1:

$$\sum_{i=1}^k f_i = 1$$

Además tenemos la **frecuencia acumulada** (absoluta o relativa) como el número de veces o fracción de las veces que se observan valores inferiores a uno dado en el total de las medidas realizadas. De esta forma si decimos la frecuencia acumulada relativa de un valor  $i$ -ésimo tendremos en cuenta no solo la frecuencia relativa de ese valor, si no todas las anteriores frecuencias más la de ese valor. En general (frecuencia absoluta y relativa acumulada):

$$N_i = \sum_{j=1}^i n_j \quad (1.1)$$

$$F_i = \sum_{j=1}^i f_j = \frac{N_i}{N} \quad (1.2)$$

Para una variable discreta con un número pequeño de variables las distribuciones de frecuencia pueden obtenerse en función de todos los valores (por ejemplo al tirar un dado de 6 caras). Sin embargo en el caso en los que el número de posibles valores observables sea comparable al número de observaciones, las distribuciones de frecuencia pierden toda su utilidad ya que  $f_i \rightarrow 0$ .

En estos supuestos se hace necesario agrupar los datos en clases. Mediante este procedimiento se divide el recorrido de la variable estadística analizada en una serie de subintervalos de tal forma que la unión de ellos cubra la totalidad del recorrido de la variable. Entonces asociamos a todos los valores dentro de cada *clase* a un único valor concreto, la *marca de clase*.

La frecuencia absoluta de observación de cada marca de clase será el número de datos originales que pertenezcan a su clase asociada. Aunque el agrupamiento de clases no sigue una norma particular, existen reglas para optimizarla. Las operaciones son:

1. Redondear los datos a dos o a lo sumo tres cifras significativas
2. El número de clases no debe ser demasiado grande ni demasiado pequeño. Se suele recomendar un número de clases  $r = \sqrt{N}$ .
3. Tomar las clases de la misma amplitud y sin ambigüedad. Para esto utilizamos intervalos semiabiertos (así no hay equivocación en los valores límites de los intervalos). En general, recorrido elegido aparecerá dividido en un conjunto de clases de la forma:

$$\{[a_0, a_1), [a_1, a_2), \dots, [a_{k-1}, a_k]\}$$

A cada clase se le asocia unas determinadas marcas de clase. Suelen tomarse como marcas de clase los centros de los diferentes subintervalos. Entonces la marca de clase asociada a los datos pertenecientes a la clase  $i$ -ésima será:

$$x_i = \frac{a_i + a_{i+1}}{2}$$

4. Finalmente se procede al recuento del número de datos crudos originales que pertenecen a las diferentes clases. El procedimiento supone establecer la equivalencia entre el experimento aleatorio original y un experimento con una variable discreta  $x_1, \dots, x_k$  en el que se hubiese obtenido esta última distribución de frecuencias.

## 1.2 Medidas características de una distribución de frecuencias

Las medidas características de una distribución de frecuencias son una serie de parámetros que resumen las propiedades generales de ésta. Hay que mencionar que las medidas que introduciremos a continuación proporcionan información relevante en el caso de muestras homogéneas de datos, mientras que pueden ser engañosas en el caso de muestras heterogéneas procedentes de distintas poblaciones. En general existen cuatro tipos de medidas características fundamentales:

1. **Medidas de posición central:** nos indican cuál es la tendencia central de los datos, entorno a qué valor concreto de la variable estadística aparecen distribuidos los datos de la muestra.
2. **Medidas de dispersión:** dan cuenta de la variabilidad inherente a los datos muestrales. Es decir, en que medida se desvían de su valor central.
3. **Medidas de asimetría:** proporcionan información acerca de la forma concreta de la distribución de datos, si predominan los datos menores o mayores de los valores centrales, o si la distribución de los datos es simétrica respecto al valor medio.
4. **Medidas de concentración:** nos informan acerca del grado de concentración de los datos, esto es, de cómo se reparte la frecuencia relativa entre los valores próximos al a media y los valores más extremos o alejados de esta.

### 1.2.1 Momentos de una distribución de frecuencias

Se define como momento de orden  $j$  de una distribución de frecuencias  $f_i$  con respecto al punto  $c$  como:

$$m_j(c) = \sum_{i=1}^k f_i(x_i - c)^j = \overline{(x_i - c)^j} \quad (1.3)$$

Si el punto  $c=0$  se denominan **momentos respecto al origen** y cuando  $c = \bar{x}$  **momentos centrales** o momentos respecto a la media.

La media será entonces el momento de primer orden respecto al origen y a la varianza el momento centradas de segundo orden.

### 1.2.2 Medidas de centralización

Existen diferentes medidas que proporcionan información acerca de la tendencia de los datos muestrales a distribuirse en torno a un valor o valores determinados. Tenemos:

- **Moda ( $Md$ ).** La moda de la muestra de datos corresponde al valor más frecuente de la distribución:

$$Md = \{x_i | f_i = \max_{j=1}^k f_j\} \quad (1.4)$$

Una distribución de datos puede tener una única moda o varias (unimodal o multimodal) lo que limita su utilidad. Además para esta medida de distribución es irrelevante el valor que tomen los demás datos.

- **Percentiles.** Se define el percentil  $q$ -ésimo de la muestra de datos ( $0 \leq q \leq 1$ ) como el valor de la variable aleatoria  $P_q$  tal que la frecuencia de todos los valores menores o iguales (frecuencia relativa acumulada) que  $P_q$  es igual a  $q$ .

$$f(X \leq P_q) = q \quad (1.5)$$

Los percentiles permiten describir la distribución de los datos muestrales de manera detallada. Hay algunos con nombre propio como la **mediana** ( $Me = P_{1/2}$ ), los **cuartiles** primero ( $P_{1/4}$ ) o tercero ( $P_{3/4}$ ), o los **deciles** (primero, segundo...) ( $q=0.1, 0.2, \dots, 0.9$ ).

En función de si la variable aleatoria es continua o discreta los valores de estos se pueden calcular de forma distinta:

- **Discreta:** si el valor coincide con alguna de las marcas de clase el valor del percentil será el valor medio de la marca de clase, el punto medio entre los extremos  $x_i$  y  $x_{i+1}$ , i.e.  $P_q = (x_i + x_{i+1})/2$ . En el caso de no coincidir en ningún valor  $N_i$  entonces el valor  $P_q$  será el mayor de los dos  $N_i$  entre los que esté comprendido, i.e.  $P_q = x_{i+1}$ .
- **Continua:** los datos se agruparan por clases, cuyas marcas de clase serán  $x_1, x_2, \dots, x_k$ . Ahora procedemos como en el caso discreto. Si el valor  $q \cdot N$  (que es al final  $P_q$  en una distribución continua) pertenece a una de las marcas de clase, tomaremos como el percentil  $q$ -ésimo el extremo superior de la clase correspondiente. Si no coincide se procederá por interpolación lineal entre las dos marcas de clase entre las que se encuentre.

$$\frac{N_{i+1} - N_i}{x_{i+1} - x_i} = \frac{q \cdot N - N_i}{P_q - x_i} \quad (1.6)$$

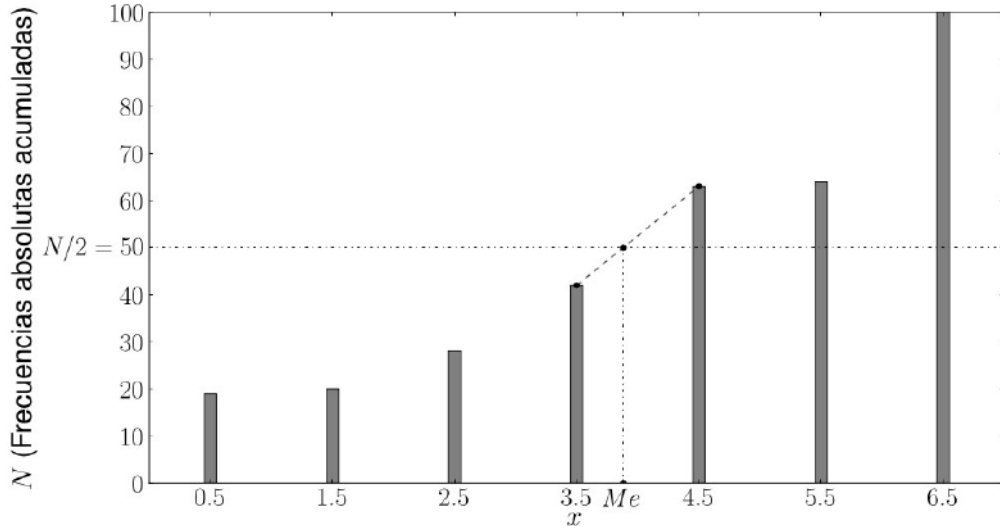


Figure 1.1: Ejemplo de interpolación lineal para una distribución

- **Mediana** ( $Me$ ). Es aquel valor de la variable para el cual los datos de valor inferior son tan frecuentes como los de valor superior. No es mas que un caso particular del percentil  $q = 1/2$ .

- **Media aritmética** ( $\bar{x}$ ). Se define como media el momento de primer orden de la distribución de frecuencias respecto al origen. Entonces:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (1.7)$$

En el caso de datos discretos agrupados en diferentes clases con un número  $k$  de valores diferentes tenemos:

$$\bar{x} = \sum_{i=1}^k f_i x_i \quad (1.8)$$

- **Media geométrica** ( $\bar{x}_g$ ). La media geométrica es la raíz N-ésima del productorio de todos los valores  $x_i$  de la medida:

$$\bar{x}_g = \sqrt[N]{\prod_{i=1}^N x_i} \quad (1.9)$$

Y en el caso de agrupar  $k$  marcas de clase de frecuencia  $f_i$  tenemos que:

$$\bar{x}_g = \sqrt[N]{\prod_{i=1}^k x_i^{n_i}} \quad (1.10)$$

Siendo  $n_i$  el número que aparece dicha marca de clase, que es  $n_i = f_i N$

- **Media cuadrática** ( $\bar{x}_q$ ). Corresponde a la raíz cuadrada del momento de orden dos respecto al origen:

$$\bar{x}_q = \sqrt{\sum_{i=1}^k f_i x_i^2} \quad (1.11)$$

- **Media armónica** ( $\bar{x}_a$ ). Corresponde al inverso de la suma de los inversos de los valores que constituyen la medida:

$$\frac{1}{\bar{x}_a} = \sum_{i=1}^n \left( \frac{f_i}{x_i} \right) \quad (1.12)$$

Aunque todas las medidas de centralización introducidas anteriormente nos dan una idea de en torno a qué valor o valores se distribuyen los datos de las muestras, son diferentes en lo referente a contenido y robustez frente a posibles cambios de la muestra. Así mientras la existencia de un solo dato extremo no modifica sensiblemente el valor de la mediana, variará considerablemente el valor de la media aritmética. Por ello las medias son mas sensibles a variaciones en la muestra o la existencia de datos atípicos (baja robustez), mientras que la mediana y percentiles son mucho más robustos frente a variaciones en la muestra. En general, es conveniente calcular todos los valores y compararlos, puesto que las divergencias entre moda y mediana y media suelen indicar asimetrías en la distribución o heterogeneidades en la muestra de datos.

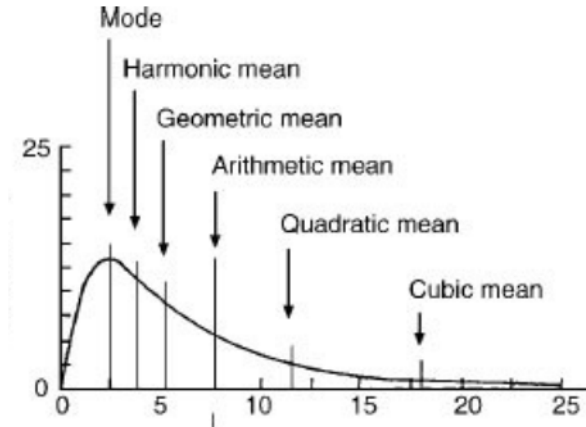


Figure 1.2: Ubicación de las medias para una distribución

### 1.2.3 Mediadas de dispersión

Las medidas de dispersión reflejan las fluctuaciones de los valores de la variable aleatoria. Mientras que las medidas de centralización nos indican los valores en torno a los que se distribuye la variable, estas nuevas medidas características nos dan una idea de la anchura de la distribución de la variable aleatoria en torno a los valores centrales.

Para construir una forma de cuantificar la dispersión, primero debemos preguntarnos en qué consiste la dispersión de una muestra de datos. La respuesta es que la dispersión de los datos se manifiesta en la separación de las diferentes observaciones respecto a un valor concreto que tomamos como referencia y que es siempre alguna de las medidas de la centralización de la muestra. Dependiendo del punto de referencia elegido y de la manera concreta de ponderar la desviación de los datos muestrales, obtendremos una u otra medida característica de dispersión:

- **Varianza** ( $s^2$ ). Se define la varianza de la muestra de datos como:

$$s^2 = \sum_{i=1}^k f_i (x_i - \bar{x})^2 \quad (1.13)$$

Es decir es la suma de las desviaciones cuadráticas de cada uno de los datos de la muestra respecto a la media aritmética. Es, sin duda, la medida de dispersión más importante. Desarrollando el cuadrado del sumatorio obtenemos que:

$$\begin{aligned} s^2 &= \sum_{i=1}^k f_i (x_i - \bar{x})^2 \\ &= \sum_{i=1}^k f_i x_i^2 - 2\bar{x} \sum_{i=1}^k f_i x_i + \bar{x}^2 \sum_{i=1}^k f_i \\ &= \bar{x}^2 - 2\bar{x}^2 + \bar{x}^2 = \bar{x}^2 - \bar{x}^2 \end{aligned} \quad (1.14)$$

- **Desviación típica** ( $s$ ). Se define como la raíz cuadrada de la varianza:

$$s = \sqrt{\sum_{i=1}^k f_i (x_i - \bar{x})^2} \quad (1.15)$$

La **desigualdad de Tchebychev** nos permite establecer, para una distribución genérica, cotas para la frecuencia de los datos que están mas alejados de la media que un cierto número de veces la desviación típica, lo que da una buena idea de la significación de este momento de la distribución.

**Desigualdad de Tchevychev:** establece que la frecuencia de los datos que distan de la media muestral mas de  $\alpha > 0$  la desviación típica está acotada de tal forma que:

$$f(x_i / |x_i - \bar{x}| > \alpha \cdot s) < \frac{1}{\alpha^2}$$

Es importante darse cuenta de que es aplicable a cualquier distribución de datos de varianza finita aunque desconozcamos la distribución de los mismos. Por ejemplo a partir de esta desigualdad sabemos que el 75% de los datos están en el intervalo  $(\bar{x} - 2s, \bar{x} + 2s)$ , es decir.

- **Desviación media respecto la media.** Se define como:

$$DM_{\bar{x}} = \sum_{i=1}^k f_i |x_i - \bar{x}| \quad (1.16)$$

Es conceptualmente semejante a la varianza ya que promedia desviaciones respecto a un punto medio, aunque las propiedades de la función valor absoluto hacen que sea menos empleada.

- **Desviación media respecto la mediana.** Se define como:

$$DM_{Me} = \sum_{i=1}^k f_i |x_i - Me| \quad (1.17)$$

Conceptualmente igual que la varianza o la desviación respecto la media, tiene el mismo problema que esta: los problemas analíticos intrínsecos de la función valor absoluto.

- **Coeficiente de variación de Pearson (CV).** Se define como:

$$CV = \frac{s}{\bar{x}} \quad (1.18)$$

Esta es la medida relativa (adimensional) de variabilidad de la muestra. Es útil cuando queremos estudiar no la incertidumbre absoluta, ya que una incertidumbre de un metro puede ser una gran incertidumbre si medimos centímetros pero muy pequeña si medimos con años luz.

- **Coefficientes de variación media.** De la misma manera que el coeficiente de variación de Pearson mide la desviación típica relativa, podemos hacer lo mismo con las desviaciones medias:

$$CV_{M_{\bar{x}}} = \frac{DM_{\bar{x}}}{\bar{x}} \quad (1.19)$$

$$CV_{M_{Me}} = \frac{DM_{Me}}{\bar{x}} \quad (1.20)$$

- **Recorrido.** Este parámetro de dispersión se define como la diferencia entre el valor máximo y mínimo de la variable estadística obtenidas en la muestra:

$$R = \max(x) - \min(x) \quad (1.21)$$

- **Recorrido intercuartílico.** Este parámetro de dispersión se define como la diferencia entre eltercer y primer cuartil de la variable estadística:

$$R_I = P_{3/4} - P_{1/4} \quad (1.22)$$

### 1.2.4 Medidas de asimetría

Otra información interesante de una distribución de datos es si esta es simétrica en puntos equidistantes respecto al valor central. Es decir, si los valores inferiores a la media que distan una cierta cantidad de esta son más, menos o igual de frecuentes que los valores superiores que se sitúan a la misma distancia. Para medir la asimetría de una distribución debemos introducir nuevos coeficientes. Los mas usados son:

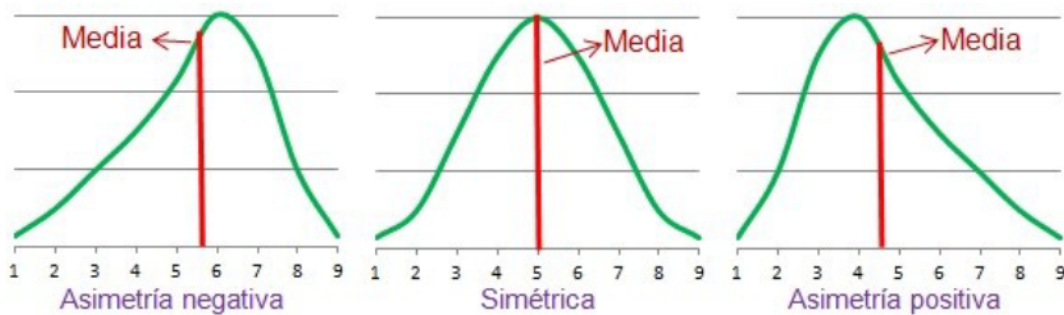


Figure 1.3: Figura que representa las diferentes asimetrías.

- **Coefficiente de asimetría de Pearson.** Este coeficiente adimensional se define como:

$$A_P = \frac{\bar{x} - Md}{s} \quad (1.23)$$

En el caso de que  $A_P > 0$  estamos ante una distribución asimétrica positiva y si  $A_P < 0$  estamos ante una distribución asimétrica negativa. Si es cero ( $A_P \approx 0$ ) estamos ante



una distribución aproximadamente simétrica. Nos da una información acerca de la distribución muy baja, además presenta problemas en una distribución multimodal.

- **Coefficiente de Bowley-Yule.** Se define a partir de los cuartiles de la distribución:

$$A_{BY} = \frac{P_{3/4} + P_{1/4} - Me}{P_{3/4} - P_{1/4}} \quad (1.24)$$

Cuando la distribución es simétrica la distancia entre el primer cuartil y la mediana será igual a la existente entre ésta y el tercer cuartil, por lo que el coeficiente será nulo. Si existe una asimetría dejará de ser nulo.

- **Coefficiente de asimetría de Fischer:** se obtiene a partir del momento central de tercer orden de la distribución:

$$A_F = \frac{m_3(\bar{x})}{s^3} = \frac{1}{s^3} \sum_{i=1}^k f_i (x_i - \bar{x})^3 \quad (1.25)$$

Su información es notablemente superior al de los otros coeficientes, ya que tiene en cuenta la contribución del conjunto de los datos para el cálculo de la distribución. Si  $A_F > 0$  estamos ante una distribución asimétrica positiva (los valores mas alejados son positivos y por eso su contribución al cubo es mayor) y si  $A_F < 0$  corresponde a una distribución asimétrica negativa.

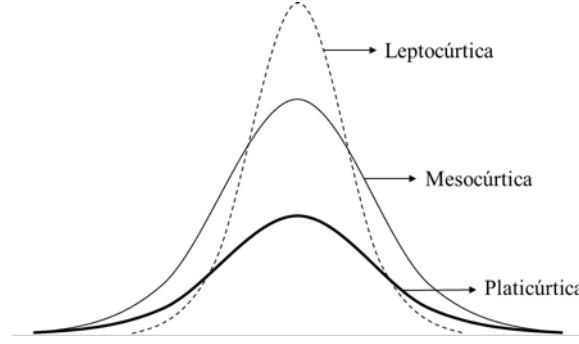
### 1.2.5 Medidas de apuntamiento o curtosis

Dos distribuciones de datos pueden tener idénticas medias y desviaciones típicas y tener formas distintas debido a la diferente concentración de los datos.

- **Coefficiente de apuntamiento o de curtosis:** se define como el cociente entre el momento central de orden 4 y el cuadrado de la varianza (o la desviación típica elevada a 4):

$$g = \frac{m_4(\bar{x})}{s^4} = \frac{1}{s^4} \sum_{i=1}^k f_i (x_i - \bar{x})^4 \quad (1.26)$$

La curtosis de una distribución normal o gaussiana es exactamente igual a 3, valor que se toma como referencia para el estudio del apuntamiento de las distribuciones estadísticas. Por eso las distribuciones con  $g > 3$  se llaman **leptocúrticas**, aquellas con  $g < 3$  las **platicúrticas** y aquellas con  $g = 3$  las **normocúrticas o mesocúrticas**.



### 1.3 Transformaciones de una variable aleatoria

En muchas ocasiones estamos interesados en realizar operaciones de los datos procedentes de los experimentos aleatorios, bien porque es mas conveniente para su manipulación o porque es necesario para acceder a una nueva variable de interés. En general distinguimos dos tipos de transformaciones: las lineales y las no lineales.

**Transformaciones lineales.** Consideremos una muestra  $\{x_i, i = 1^N$  de realizaciones de la variable estadística  $X$  con los posibles valores  $\{x_1, \dots, x_k\}$  y media aritmética  $\bar{x}$ . Ahora si transformamos esta variable a una nueva variable  $Y = a + bX$  tenemos los valores posibles  $\{a + bx_1, \dots, a + bx_k\}$ . Analicemos ahora las principales medidas características de la nueva variable. La variable:

$$\begin{aligned}\bar{y} &= \sum_{i=1}^k f_i y_i = \sum_{i=1}^k f_i (a + bx_i) \\ &= a + b \sum_{i=1}^k f_i x_i = a + b\bar{x}\end{aligned}\tag{1.27}$$

Como vemos la media aritmética se transforma de la misma manera que la variable. Por otra parte, la varianza de la nueva aleatoria será:

$$\begin{aligned}s_y^2 &= \sum_{i=1}^k f_i (y_i - \bar{y})^2 \\ &= \sum_{i=1}^k f_i (a + bx_i - a - b\bar{x})^2 \\ &= b^2 \sum_{i=1}^k f_i (x_i - \bar{x})^2 = b^2 s^2\end{aligned}\tag{1.28}$$

**Transformaciones no lineales.** En muchas veces usamos transformaciones no lineales, por ejemplo:

$$Y = X^P, Y = \ln(X), Y = e^X \dots$$

Las transformaciones no lineales, suelen variar la forma de las distribuciones originales de frecuencias, lo que nos obliga a ser cuidadosos en el análisis de las implicaciones de la transformación en las diferentes medidas características. Veamos los efectos de este tipo de transformaciones en algunas medidas características.

## 1.4 Distribuciones de frecuencia multivariable

Al estudiar varias variables aleatorias obtenidas simultáneamente en un experimento podemos comprobar y hallar relaciones entre estas variables, usando distribuciones de frecuencia multivariante.

Si consideramos dos variables X e Y con valores  $\{x_1, \dots, x_k\}$  y  $\{y_1, \dots, y_l\}$  tenemos que la frecuencia del par  $(x_i, y_j)$  será  $f_{ij}$ .

Llamamos **distribución de frecuencias marginal** para una variable X dentro de una distribución multivariante a la frecuencia asociada a los valores de X, sin importar los demás. Es decir es la frecuencia de que salga un  $x_i$  determinado respecto a N sin importar si el par es  $y_4$  o  $y_7$ . Es decir, si hablamos de frecuencias relativas:

$$f_{x_i} = \sum_j f_{ij}$$

### Independencia estadística y distribuciones de frecuencia condicionadas

Cuando hablamos de frecuencia de  $x$  condicionada a  $y_j$  nos referimos a las diferentes probabilidades de que salga un  $x_i$  *sabiendo* que ha salido  $y_j$ . Es decir es la probabilidad de que salgan los diferentes  $x_i$  sabiendo que la otra variable es una  $y_j$  previamente definida. Entonces si  $n_{y_j}$  es la frecuencia marginal de  $y_j$  tenemos que las diferentes frecuencias de  $x_i$  condicionadas a  $y_j$  son:

$$f(x_i|y_j) = \frac{f_{ij}}{n_{y_j}} \quad (1.29)$$

Hablamos de **independencia estadística** entre x e y cuando las distintas distribuciones de  $x_i$  condicionadas a cualquier  $y_j$  son exactamente iguales que las distribuciones marginales de  $x_i$ . Entonces

$$f(x_i|y_j) = n_{x_i} = f_{x_i}$$

Es decir cuando la frecuencia de un  $x_i$  no depende del  $y_j$  que haya salido es cuando hablamos de *independencia estadística*.

### Medidas características muestrales multivariantes

Al igual que con las medidas normales aquí tenemos **momentos de la distribución**. Llamamos momento de ordenes r y s respecto al punto (c,d) como:

$$m_{r,s}(c, d) = \sum_i \sum_j f_{ij} (x_i - c)^r (y_j - d)^s \quad (1.30)$$

Cuando hablamos momento respecto al origen hablamos momento respecto a  $(c,d)=(0,0)$  y momento central a  $(c,d)=(\bar{x}, \bar{y})$

Para definir media de una de las variables estadísticas hablamos de:

$$\bar{x} = m_{1,0}(0, d) = \sum_i \sum_j f_{ij}(x_i - 0)^1(y_j - d)^0 \quad (1.31)$$

Y la varianza como:

$$s_x^2 = m_{2,0}(\bar{x}, d) = \sum_i \sum_j f_{ij}(x_i - \bar{x})^2(y_j - d)^0 \quad (1.32)$$

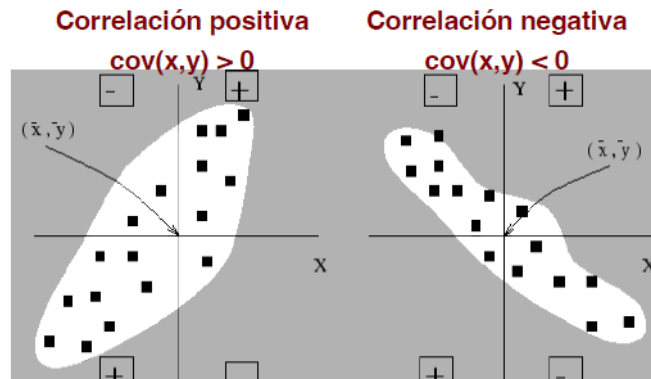
La **covarianza** es el momento central de primer orden en las dos variables estadísticas (donde  $\overline{xy} = \sum \sum f_{i,j} x_i y_j$ ):

$$\text{cov}(x, y) = m_{1,1}(\bar{x}, \bar{y}) = \overline{xy} - \bar{x}\bar{y} \quad (1.33)$$

Veamos ahora los cálculos para llegar a dicho resultado:

$$\begin{aligned} \text{cov}(x, y) &= m_{1,1}(\bar{x}, \bar{y}) = \sum_i \sum_j f_{ij}(x - \bar{x})(y - \bar{y}) = \\ &= \sum_i \sum_j f_{ij} x_i y_j - \sum_i \sum_j f_{ij} x_i \bar{y} - \sum_i \sum_j f_{ij} y_j \bar{x} + \sum_i \sum_j f_{ij} \bar{x} \bar{y} = \overline{xy} - \bar{x}\bar{y} - \bar{x}\bar{y} + \bar{x}\bar{y} \end{aligned}$$

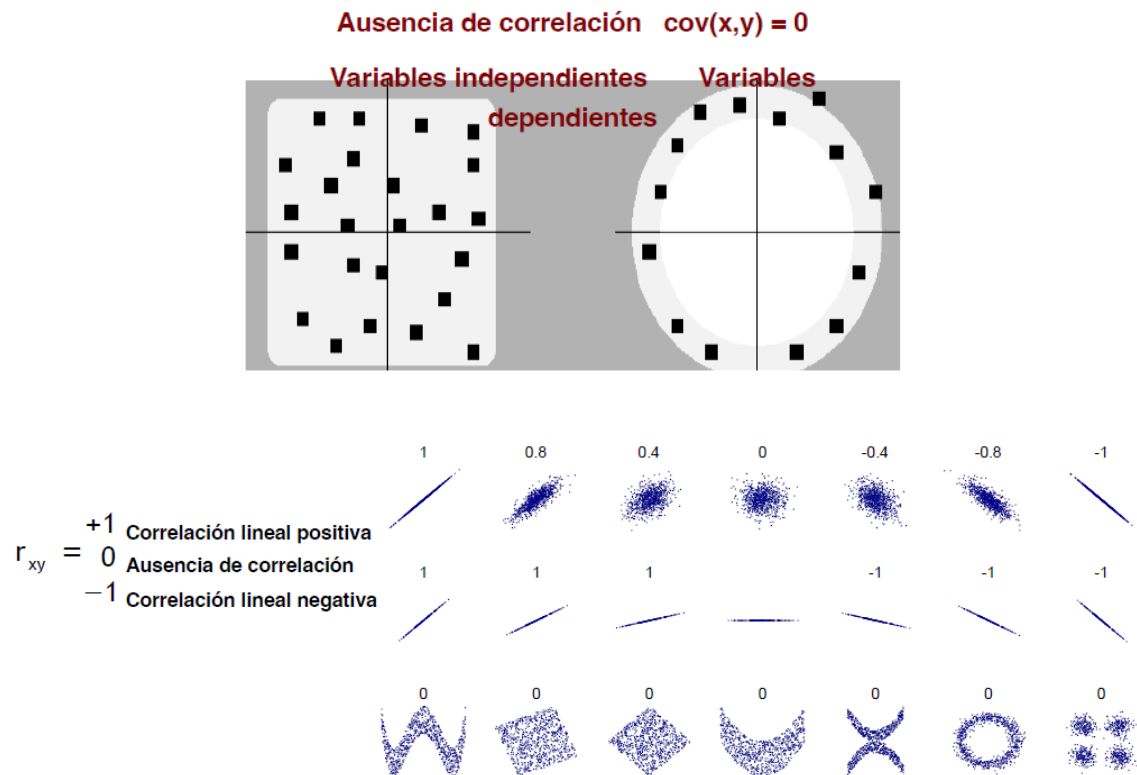
La covarianza tiene la característica que si  $x$  e  $y$  son independientes tiene valor de 0. Además informa sobre la correlación entre las variables estadísticas: si la covarianza es positiva hay una correlación positiva (como  $y=x$ ) y si es negativa hay una correlación negativa (como  $y=-x$ ). Véase mejor en la siguiente figura:



Llamamos **coeficiente de correlación lineal entre variables estadísticas (X,Y)** como:

$$r_{x,y} = \frac{\text{cov}(x, y)}{s_x + s_y} \quad (1.34)$$

Lo bueno de este coeficiente es que es adimensional ante cambios de escala (transformaciones lineales), y su dominio está acotado entre -1 y 1 siempre.





# Chapter 2

## Teoría de probabilidades

### 2.1 Fundamentos de probabilidad

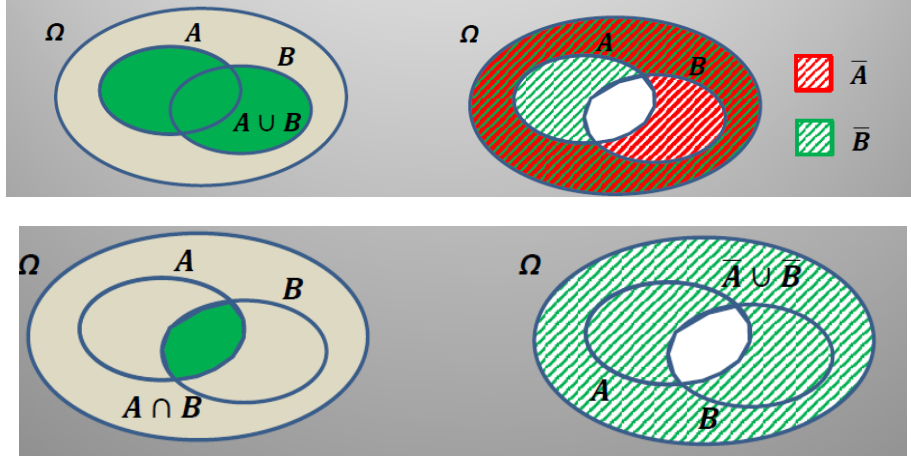
Llamamos **experimento aleatorio** a aquel que repetido sucesivas veces en condiciones idénticas produce diferentes resultados de manera no previsible. Un experimento determinista es aquel que repetido veces en condiciones idénticas siempre produce el mismo resultado. Para ello introduciremos los siguientes conceptos básicos:

- **Suceso elemental:** cada uno de los posibles resultados del experimento.
- **Espacio muestral  $\Omega$ :** el conjunto de todos los resultados posibles. Un espacio muestral puede ser:
  - **Discreto:** si cada uno de los sucesos se puede corresponder a un subconjunto de los números naturales
  - **Continuo:** si cada uno de los sucesos se puede corresponder a uno o mas intervalos de la recta real
- **Suceso:** cualquier subconjunto del espacio muestral (por ejemplo que al tirar un dado sea par).
- $\mathcal{A}$ : es el conjunto de todos los sucesos que es posible definir asociados a un cierto experimento aleatorio.

Supondremos que el lector ya estará familiarizado con los conceptos de unión, intersección, conjunto vacío... así como distinguir la incompatibilidad, diferencia de varios sucesos.

Considérense los sucesos A y B. Las **leyes de Morgan**, que son bastante intuitivas, nos dicen que:

- $\overline{A \cup B} = \bar{A} \cap \bar{B}$ :
- $\overline{A \cap B} = \bar{A} \cup \bar{B}$ :



Se define como **probabilidad de un suceso** (definición de probabilidad) como:

$$p(A) = \frac{\text{casos favorables}}{\text{casos posibles}} = \frac{n}{N} \quad (2.1)$$

Cuando hablamos de probabilidad hay muchas cosas que, al igual que antes, daremos por conocidas, como que la probabilidad de la muestra es 1, la probabilidad del conjunto vacío es nula, que la probabilidad del complementario de A viene dada por 1 menos la probabilidad de A... Consecuentemente pasaremos a ver las probabilidades condicionadas, ya que tienen mucho mas interés.

Nosotros conocemos la probabilidad condicionada del suceso A respecto B como la probabilidad de que ocurra A sabiendo que ha ocurrido el proceso B:

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \quad (2.2)$$

Por consecuencia siempre que  $p(B) > 0$  podremos escribir que:

$$p(A|B)p(B) = p(A \cap B)$$

Decimos que dos procesos son **estadísticamente independientes** cuando:

$$p(A|B) = p(A) \iff p(A \cap B) = p(A) p(B) \iff p(B|A) = p(B)$$

Cuando los sucesos A y B son independientes los siguientes pares de sucesos también lo son:

$$A, \bar{B}; \quad \bar{A}, B; \quad \bar{A}, \bar{B}$$

Llamamos a un conjunto de sucesos  $\{A_1, A_2, \dots, A_n | A_i \in \mathcal{A}\}$  es un **conjunto exhaustivo** si su unión cubre el espacio muestral:

$$\bigcup_{i=1}^n A_i = \Omega$$



El mismo conjunto de sucesos se le llama **conjunto completo** si además los sucesos entre sí tienen intersecciones vacías.

$$A_i \cap A_j = \emptyset; p(A_i \cap A_j) = 0 \quad \forall i \neq j$$

- **Teorema de Bayes:** dado un suceso B el Teorema de Bayes nos indica que:

$$p(A_i|B) = \frac{p(B|A_i)p(A_i)}{p(B)} = \frac{p(B|A_i)p(A_i)}{\sum_k p(B|A_k)p(A_k)}$$

### 2.1.1 Función de probabilidad, densidad de probabilidad y distribución de probabilidad

La función de **densidad de probabilidad** se define como:

$$\lim_{\Delta \rightarrow 0} \frac{p(x_0 - \Delta \leq x \leq x_0 + \Delta)}{\Delta} = f(x_0)$$

Para esto la función de probabilidad  $p(x)$  tiene que ser suave. Además tiene las siguientes propiedades:

- $f(x) \geq 0 \forall x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $\int_a^b f(x)dx = p(a \leq x \leq b)$

La **función de distribución de probabilidad**  $F(x)$  se define como:

$$F(x_0) = p(x \leq x_0) = \int_{-\infty}^{x_0} f(x)dx$$

Por lo tanto se cumplen las siguientes propiedades:

- $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$
- $F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1$
- $\frac{dF(x)}{dx} = f(x)$

### 2.1.2 Normalización de una distribución continua:

Supongamos que tenemos una función de distribución de probabilidad definida como:

$$y = \begin{cases} kx^2 & \text{si } 0 \leq x \leq 2 \\ 0 & \text{en otro caos} \end{cases}$$

Entonces como la distribución de probabilidad tiene que cumplir que  $\int_{-\infty}^{\infty} f(x)dx = 1$  entonces:

$$1 = \int_0^2 kx^2 dx = k \frac{8}{3} \implies k = \frac{3}{8}$$

De esta forma tendríamos normalizada la función.

### 2.1.3 Esperanza:

Dada una variable aleatoria definimos la esperanza o valor esperado de la variable como:

$$\mathbb{E}\{X\} = \begin{cases} \sum_i p(x_i)x_i & \text{si variable discreta} \\ \int_{-\infty}^{\infty} f(x)x dx & \text{si variable continua} \end{cases}$$

La esperanza de una variable aleatoria es su valor esperado, pero ¿Por qué? Cuando nosotros queremos calcular la media de cualquier cosa lo que hacemos es multiplicar la probabilidad de un que ocurra un suceso elemental por el propio suceso, eso en todos los sucesos posibles, sumarlos y dividirlos entre la cantidad de sucesos. En un caso integral lo que hacemos al poner  $x$  es poner estos sucesos elementales por su probabilidad ( $f(x)$ ) en todo su espacio muestral (que es el que tenemos que recorrer). Por eso el valor esperado de una variable es igual a la esperanza del suceso. Podemos poner en vez de  $x$   $H(x)$ , siendo este un cambio de variable cualquiera  $y=H(x)$ . Es el caso general.

En general solemos decir que la esperanza actúa como un operador lineal sobre la variable aleatoria, ya que:

$$\mathbb{E}\{\alpha H(X) + \beta G(X)\} = \alpha \mathbb{E}\{H(X)\} + \beta \mathbb{E}\{G(X)\}$$

### 2.1.4 Medidas de centralización:

Como ya dijimos la esperanza de la variable aleatoria  $X$  es igual al valor esperado (como la media). En general se escribe como:

$$\mu = \mathbb{E}\{X\} \tag{2.3}$$

La **mediana**  $Me$  es el valor tal que  $F(Me)=0.5$ . La **moda**  $Md$  es el valor mas probable. Por consiguiente:

$$\left. \frac{df}{dx} \right|_{Md} = 0$$

Por ser máximo.

### 2.1.5 Medidas de dispersión:

La **varianza**  $\sigma^2$  de una variable aleatoria  $X$  se define como:

$$\sigma^2 = \mathbb{E}\{(X - \mathbb{E}\{X\})^2\} = \mathbb{E}\{(X - \mu)^2\} \quad (2.4)$$

Y la **desviación típica**  $\sigma$  es la raíz cuadrada de la varianza. La desviación típica cualitativamente nos trata de dar una medida de la dispersión de los datos.

Las funciones de probabilidad pueden normalizarse (para que cuadren con las frecuencias relativas) y de esta forma su integral evaluada entre  $(-\infty, \infty)$ .

### 2.1.6 Tipificación de la variable aleatoria:

Lo que hacemos al tipificar una variable aleatoria es mover la media al punto 0, de tal forma que no se cambia la forma, ni la desviación típica; y luego crear una nueva variable donde "normalizamos" la varianza (de tal forma que  $\sigma^2 = 1$ ). Para esto construimos una nueva variable tal que:

$$Y = \frac{X - \mu_x}{\sigma_x} \quad (2.5)$$

$$\mathbb{E}\{Y\} = \mathbb{E}\left\{\frac{X - \mu}{\sigma}\right\} = \frac{1}{\sigma}\mathbb{E}\{X - \mu\} = \frac{1}{\sigma}\mathbb{E}\{X\} = 0$$

$$\sigma^2(Y) = \mathbb{E}\{[Y^2 - \mathbb{E}\{Y\}]^2\} = \mathbb{E}\{Y^2\} = \mathbb{E}\left\{\left[\frac{X - \mu}{\sigma}\right]^2\right\} = \frac{1}{\sigma^2}\mathbb{E}\{(X - \mu)^2\} = 1$$

Dada cualquier variable aleatoria podemos tipificarla (haciendo ese cambio de variable). De esta forma podremos transformar cualquier variable aleatoria para poder así comparar variables aleatorias con diferente media y varianza, pudiendo caracterizar la dispersión, asimetría...

### 2.1.7 Momentos de la distribución de probabilidad

Definimos momento de orden  $j$  de una distribución de probabilidad de la misma manera que en estadística:

$$M_j(c) = \mathbb{E}\{(x - c)^j\} = \int_{-\infty}^{+\infty} f(x)(x - c)^j \quad (2.6)$$

De esta modo tenemos la media (momento respecto el origen de orden uno), la varianza (momento central de orden dos)...

### 2.1.8 Función generatriz de momentos

Muchas veces es mas fácil analizar las propiedades de una función de distribución de probabilidad usando las técnicas del análisis matemático mediante la **función generatriz de momentos**:

$$M_{x-c}(t) = \mathbb{E}\{e^{[t(x-c)]}\} = \int_{-\infty}^{\infty} f(x)e^{t(x-c)}dx \quad (2.7)$$

Entonces  $M_{x-c}$  sea analítica (continua y diferenciable) en un intervalo:

$$\frac{d}{dt} M_{x-c}(t) = \int_{-\infty}^{+\infty} f(x)(x-c)e^{[t(x-c)]} dx \quad (2.8)$$

Haciendo el límite cuando  $t$  tiende a cero tenemos que la primera derivada de la función generatriz de momentos es igual al momento de orden 1 respecto  $c$ .

$$\lim_{t \rightarrow 0} \frac{d}{dt} M_{x-c}(t) = \int_{-\infty}^{\infty} f(x)(x-c) dx = M_1(c) \quad (2.9)$$

Si nos damos cuenta tenemos que:

$$\lim_{t \rightarrow 0} \left( \frac{d^j M_{x-c}(t)}{dt^j} \right) = \int_{-\infty}^{\infty} f(x)(x-c)^j dx = M_j(c) \quad (2.10)$$

Esto nos sirve para calcular momentos bastante altos de manera muy sencilla, incluso varianzas, curtosis, asimetrías (fischer), medias... o el momento que queramos. Ahora veremos el ejemplo para la función de probabilidad exponencial:

$$f(x) = \begin{cases} 0 & x \in (-\infty, 0) \\ \beta e^{-\beta x} & x \in [0, \infty) \end{cases}$$

Entonces tenemos que ( $c=0$ ):

$$M_x(t) = \int_{-\infty}^{\infty} f(x)e^{t(x-c)} dx = \int_0^{\infty} \beta e^{(t-\beta)x} dx = \frac{\beta}{\beta - t}$$

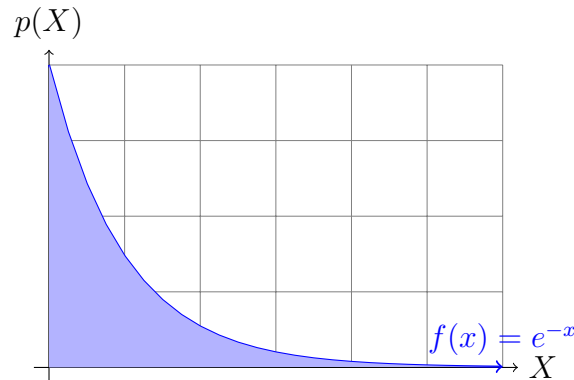
Entonces si derivamos:

$$\mu = M_1(x) = \lim_{t \rightarrow 0} \frac{dM_x(t)}{dt} = \lim_{t \rightarrow 0} \frac{\beta}{(\beta - t)^2} = \frac{1}{\beta}$$

$$\sigma^2 = \mathbb{E}\{x^2\} - \mu^2 = \lim_{t \rightarrow 0} \frac{d^2 M_x(t)}{dt^2} - \mu^2 = \lim_{t \rightarrow 0} \frac{2\beta}{(\beta - t)^3} - \frac{1}{\beta^2} = \frac{1}{\beta^2}$$

Entonces podremos calcular todos los momentos de distribución:

$$\frac{d^r M_x(t)}{dt^r} = \frac{r! \beta}{(\beta - t)^{r+1}} \implies M_0(x) = \frac{r!}{\beta^r}$$



### 2.1.9 Transformación de la variable aleatoria:

En el caso de una variable aleatoria continua (para variables aleatorias discretas se redefine de la misma manera)  $X$  podemos redefinirla mediante una aplicación  $g(x)$  tal que:

$$y = g(x)$$

Siendo  $f_X(x)$  y  $f_Y(y)$  las funciones de densidad, tenemos que en un intervalo muy pequeño  $[a,b]$  de  $x$  la probabilidad de que ocurra esos mismos valores en  $y$  ( $[g(a),g(b)]$ ) es exactamente la misma. Entonces:

$$|f_Y(y)dy| = |f_X(x)dx|$$

Es por el simple hecho de que si sale  $z$  veces los valores de  $X$  entre  $[a,b]$  tienen que salir  $z$  veces los valores de  $Y$  entre  $[g(a),g(b)]$ . Por lo tanto tenemos que:

$$f_Y(y) = \frac{f_X(x)}{\left|\frac{dy}{dx}\right|} \quad (2.11)$$

Muchos nos podríamos preguntar para que sirve está transformación, para que queríamos transformar una variable aleatoria y crear una nueva función de densidad de probabilidad. Pues uno de sus mayores usos son los simuladores en los ordenadores. En la mayoría de ordenadores encontramos generadores aleatorios de distribución pero de distribución de probabilidad uniforme, sin embargo, si queremos simular la interacción radiación materia, la mayoría de los procesos tiene una probabilidad de interacción exponencial. Entonces esta transformación nos permitiría simular en el ordenador estas interacciones:

$$\begin{cases} f_X(x) = 1 & 0 \leq x \leq 1 \\ f_Y(y) = \beta e^{-\beta x} & 0 \leq 0 < \infty \end{cases}$$

$$\left|\frac{dy}{dx}\right| = \frac{f_X(x)}{f_Y(y)} = \frac{1}{\beta e^{-\beta y}} \implies (1 - e^{-\beta y}) = x \implies y = \frac{-1}{\beta} \log(1 - x)$$

Y ese sería el cambio de variable que hay que hacer para hacer la simulación. También nos interesa estudiar como se transforma la media y la varianza de las variables aleatorias en un cambio de variable ( $y=g(x)$ ). Podemos aproximar  $g(x)$  por la serie de Taylor centrada en  $\mu_x$ :

$$g(x) = g(\mu_x) + \left.\frac{dg}{dx}\right|_{\mu_x} (x - \mu_x) + \frac{1}{2} \left.\frac{d^2g}{dx^2}\right|_{\mu_x} (x - \mu_x)^2 \dots$$

Entonces:

$$\mathbb{E}\{y\} = \mathbb{E}\{g(x)\} = g(\mu_x) + \left.\frac{dg}{dx}\right|_{\mu_x} \underbrace{\mathbb{E}\{(x - \mu_x)\}}_0 + \frac{d^2g}{dx^2}\bigg|_{\mu_x} \underbrace{\mathbb{E}\{(x - \mu_x)^2\}}_{\sigma^2(x)} + \dots$$

Entonces si podemos despreciar los términos de orden superior (a partir de dos) podemos despreciarlos y por lo tanto:

$$\mathbb{E}\{Y\} \approx g(\mu_x)$$

Podemos aplicar lo mismo para que:

$$\sigma_Y^2 \approx \left. \frac{dg}{dx} \right|_{\mu_x} \sigma_X^2$$

La demostración es análoga a la demostración de 1.32

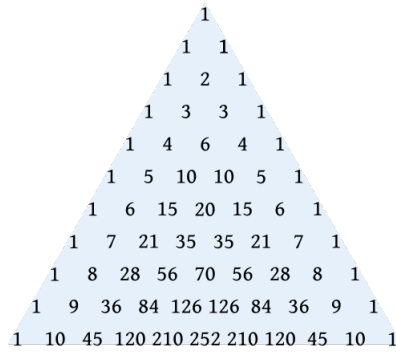
### 2.1.10 Distribución binomial:

Para entender la distribución binomial primero debemos entender que es un **suceso de Bernoulli**: supongamos un suceso aleatorio con dos posibles resultados que asignaremos 0 y 1. Sea la probabilidad de 1  $p$  ( $p(1)=p$ ), y la probabilidad de 0 entonces será  $1-p$  ( $p(0)=1-p=q$ ). Para esta variable aleatoria tenemos que.

$$\mu = \mathbb{E}\{x\} = \sum_i p_i x_i = p + 0 = p$$

$$\sigma^2 = \mathbb{E}\{(x - \mu)^2\} = \sum_i p_i (x - i - p)^2 = p(1 - p)^2 + q(0 - p)^2 = pq(q + p) = pq$$

Ahora imaginemos que tenemos  $n$  sucesos de bernouilli estadísticamente independientes, como si tiráramos  $n$  veces una moneda. ¿Cuál es la probabilidad de que aparecieran, por ejemplo,  $x$  caras? Si  $p$  es la probabilidad de sacar una cara pues dependerá de los caminos posibles hasta llegar allí y la probabilidad de que ocurra uno de esos caminos. Ahora bien, ¿Cuál es la cantidad de caminos posibles? Si nos fijamos en este caso y vamos realizando el típico esquema:



Tenemos que es el triángulo de pascal. Entonces la cantidad de caminos viene dada por la fórmula de la combinatoria:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \quad (2.12)$$

Siendo  $n$  el número de tiradas y  $r$  el numero de veces que quieres que salga tu suceso. Y ahora: ¿Cuál es la probabilidad de que ocurra un camino? Pues está claro: la probabilidad de que ocurra el suceso que queremos  $r$  veces por la probabilidad de que no ocurra  $n-r$  veces. Entonces ya tendríamos formulado la probabilidad de que ocurra un proceso  $x$  veces, que sería:

$$p(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x} \quad (2.13)$$

Entonces el valor esperado y la varianza:

$$\mathbb{E}\{X\} = \sum_{x=0}^n xp(x) = \sum_{x=0}^n x \binom{n}{x} p^x q^{n-x} = \sum_{x=1}^n x \binom{n}{x} p^x q^{n-x} = np \quad (2.14)$$

$$\sigma^2 = \sum_{x=0}^n xp(x) = npq \quad (2.15)$$

**Nota:** escribir demostración

### 2.1.11 Distribución de Pascal:

Se trata de evaluar el número de veces que se produce el experimento aleatorio hasta que el suceso A con probabilidad p no ocurra. La probabilidad de que siempre salga cara (probabilidad de que salga cara p) hasta que sale una cruz en n tiradas será:

$$p(x = n) = p^{n-1}q$$

Entonces el valor medio de está será:

$$\mu = \sum_{x=0}^{\infty} xp^{x-1} = \frac{1}{q} \quad (2.16)$$

$$\sigma^2 = \frac{p}{q^2} \quad (2.17)$$

**Nota:** escribir demostración

En muchos libros nos podemos encontrar la definición contraria, cambiando n por q. Es decir que la distribución de pascal es la probabilidad de que un suceso A no ocurra hasta la n, de tal forma que:

$$p(x = n) = q^{n-1}p$$

### 2.1.12 Distribución de Poisson:

En la distribución de Poisson es el límite de la distribución binomial si suponemos que el número de elementos (n) es muy grande pero la probabilidad de que ocurra un suceso (p) en un intervalo de elementos pequeño es realmente despreciable. Por ejemplo si cogemos la cantidad de accidentes de coche que hay en una carretera durante dos minutos, lo mas probable es que no ocurra ningún accidente. Sin embargo si cogemos la probabilidad a lo largo de 1 mes es probable encontrar un resultado de tal manera que la probabilidad de que p haya ocurrido sea una constante ( $\lambda$ ) que se repita todos los meses que lo hagamos, incluso que si lo hagamos 10 años la probabilidad de accidentes sea la misma (no la cantidad). Entonces tenemos que para sucesos con n grande el valor esperado se mantiene constante:

$$\mathbb{E}\{x\} = \lambda = np; \quad p = \lambda/n \quad (2.18)$$

Entonces tenemos que:

$$p(x = r) = \binom{n}{r} \left(\frac{\lambda}{n}\right)^r \left(1 - \frac{\lambda}{n}\right)^{n-r}$$

Del cual podemos tomar los límites  $n \rightarrow \infty$  de tal manera que (siendo  $x$  el número de intentos):

$$p(x = r) = \frac{\lambda^r}{r!} e^{-\lambda} \quad (2.19)$$

$$\mathbb{E}\{x\} = \lambda \quad (2.20)$$

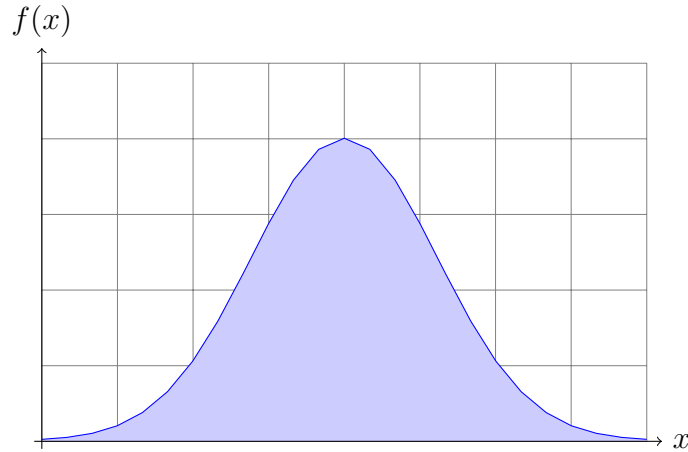
$$\sigma^2 = \lambda \quad (2.21)$$

**Nota:** escribir demostraciones

### 2.1.13 Distribución Gaussiana:

Una variable aleatoria continua se dice que tiene una distribución de probabilidad gaussiana cuando su densidad de probabilidad se puede escribir como:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \quad \sigma > 0, \quad x \in (-\infty, \infty) \quad (2.22)$$



Y su función de distribución de probabilidad:

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt; \quad \sigma > 0, \quad x \in (-\infty, \infty) \quad (2.23)$$

Entonces podríamos probar que:

$$\mathbb{E}\{x\} = \mu \quad (2.24)$$

$$\sigma^2 = \sigma^2 \quad (2.25)$$

Normalmente escribimos la gaussiana de media  $\mu$  y desviación típica  $\sigma$ :



$$x \in N(\mu, \sigma)$$

Si calculamos la función generatriz de momentos para la **función normal estándar** (gaussiana  $\mu = 0$ ,  $\sigma = 1$ ) tenemos que:

$$\begin{aligned} M_X(t) &= \mathbb{E}\{e^{tx}\} = \int_{-\infty}^{\infty} f(x)e^{tx} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{tx - \frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tx - \frac{x^2}{2}} dx \Rightarrow \\ M_X(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-x)^2}{2}} e^{\frac{t^2}{2}} dx = e^{t^2/2} (2\pi)^{-1/2} \underbrace{\int_{-\infty}^{\infty} e^{-(t-x)^2/2} dx}_{(2\pi)^{1/2}} \Rightarrow \\ M_X(t) &= e^{\frac{t^2}{2}} \end{aligned} \quad (2.26)$$

En general podemos convertir cualquier gaussiana a una normal estándar gracias a la tipificación que vimos en 2.1.6. Si calculamos la **curtuosis** de la distribución normal tenemos que ( $\sigma = 1$ ;  $\mu = 0$ ; como la media es 0 podemos calcular la curtuosis directamente como la derivada cuarta de la función generatriz de momentos):

$$g = \frac{M_4(\mu)}{\sigma^4} = \left. \frac{d^4 e^{t^2/2}}{dt^4} \right|_0 = \lim_{t \rightarrow 0} 3e^{t^2/2} + 6t^2 e^{t^2/2} + t^4 e^{t^2/2} = 3$$

### 2.1.14 Distribuciones multivariantes

Un suceso muchas veces implica calcular determinadas magnitudes a la vez. En este caso a la variable aleatoria se le llama **variable aleatorias multidimensionales o vectoriales**. Por ejemplo, al medir la estatura el peso y la edad en la población española resulta una variable aleatoria tridimensional, además que las 3 estarán relacionadas mediante una correlación (la estudiaremos mas adelante con la covarianza). En este caso el valor de la variable aleatoria es un conjunto de n valores numéricos.

Entonces para las distribuciones multivariantes definiremos una función de probabilidad conjunta  $f(x,y)$ . Llamamos distribuciones marginales para una de las componentes  $x_i$  de la distribución a la probabilidad de que salgan os diferentes valores de  $x_i$  ignorando las otras variables. Es decir consideramos la distribución univariante de dicha componente, su distribución en la población considerada aisladamente. De este modo sea la distribución bidimensional  $f(x,y)$ :

- $f(x) = \int_{-\infty}^{\infty} f(x,y) dy$
- $f(y) = \int_{-\infty}^{\infty} f(x,y) dx$

Además, como en el Tema ?? podemos hablar de independencia estadística y de probabilidades condicionadas. En el caso bidimensional (de ahora en adelante trataremos siempre el caso bidimensional, a menos que especifiquemos que no) tenemos que x e y son independientes estadísticamente si se cumple que:

$$f(x|y) = f(x); f(y|x) = f(y) / f(x|y) = \frac{f(x \cap y)}{f(y)}$$

También tenemos el **teorema de bayes** para distribuciones multivariantes, medias, varianzas, momentos... Quizás lo mas relevante sería la **covarianza**, ya que nos da una medida de la correlación entre las variables. Por ejemplo si dos variables son estadísticamente independiente tenemos que la covarianza es 0. Sin embargo el recíproco no es cierto. Definimos covarianza como:

$$\text{cov}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy \quad (2.27)$$

**Nota:** demostrar que si dos variables estadísticamente independientes su covarianza es 0

### 2.1.15 Sumando variables aleatorias y Teorema del límite central:

Consideremos dos variables aleatorias  $X$  e  $Y$ , tal que construimos una nueva variable  $Z = X + Y$ . Entonces la media y la varianza de  $Z$  vendrán dadas por:

$$\mathbb{E}\{z\} = \mathbb{E}\{x\} + \mathbb{E}\{y\} \quad (2.28)$$

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 + 2\text{cov}(x, y) \quad (2.29)$$

Y si  $X$  e  $Y$  son independientes estadísticamente  $\text{cov}(x, y) = 0$  y por lo tanto:

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2$$

Entonces si consideramos que sumamos  $n$  variables aleatorias independientes  $(X_1, X_2, \dots, X_N)$  con la misma distribución de probabilidad de tal manera que la media de  $X_i$  es  $\mu$  y la varianza  $\sigma^2$ . Entonces si definimos una variable aleatoria  $\bar{X}$  tal que:

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + X_3 + \dots + X_N)$$

La media y varianza de  $\bar{X}$  vendrá dada por:

$$\mathbb{E}\{\bar{X}\} = \mu; \quad \sigma^2(\bar{X}) = \frac{\sigma^2}{n}$$

Si tipificamos la variable aleatoria  $\bar{X}$  tenemos que:

$$\bar{Y} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (2.30)$$

Tal que  $\mu_Y = 0$  y  $\sigma_Y = 1$ . Tenemos que si  $n \rightarrow \infty$  la probabilidad de  $\bar{Y}$  tiene una distribución de probabilidad igual que una normal estándar  $N(0,1)$ .

**Nota:** demostrar esto

# Chapter 3

## Métodos estadísticos

En los métodos estadísticos estudiamos los métodos mediante los cuales podemos hacer afirmaciones o predicciones acerca del comportamiento de los elementos de una población a partir de la observación de subconjuntos de la población llamadas muestras. A este estudio también se le llama **inferencia estadística**, y en esta sección se introduzcan los siguientes temas:

1. **Estimación puntual y por intervalos:** en esta parte estudiamos las técnicas de estimación de parámetros de distribuciones de probabilidad a partir de los obtenidos en las distribuciones de frecuencia experimentales o empíricas. El caso mas importante es aquel en el que consideramos una población en la que tenemos definida una función de distribución de probabilidad. La información  $(\mu, \sigma \dots)$  de las distribuciones serán desconocidas, y tendremos que calcularlas a partir de nuestros datos experimentales mediante una serie de procedimientos. En particular nos centraremos en el tratamiento detallado de la estimación puntual de la media y varianza, así como la construcción de intervalos de confianza para estos.
2. **Test de hipótesis estadísticas:** gracias a los test de hipótesis estadísticas podremos estimar el nivel de aceptabilidad estadística de una afirmación acerca del comportamiento de una población con la información conocida de una muestra.
3. **Método de máxima verosimilitud:** proporciona un procedimiento de evaluación de los parámetros que maximiza la probabilidad de obtención de los datos determinada a una distribución. De esta forma podremos maximizar los parámetros de un ajuste (lineal, exponencial...) para maximizar la verosimilitud de la muestra (una vez suponemos que la muestra sigue una determinada distribución de probabilidad).

### 3.0.1 Estimación puntual de parámetros

Si consideramos una población de todas las posibles medidas de una magnitud de interés, siendo  $X$  la variable aleatoria y  $f(x)$  la función densidad de probabilidad, existen ciertos parámetros que describen esta distribución de probabilidad, tales como la media, la desviación típica, la curtosis...

Para resolver el problema primero debemos tener los datos que hemos obtenido. Considerando que tomamos  $l$  muestras de tamaño  $n$ ; y consideramos que cada una de las muestras corresponden a una variable  $n$  dimensional:

$$\vec{x}^i = (x_1^i, x_2^i, \dots, x_n^i)$$

Que siguen una distribución de probabilidad multivariante  $g(\vec{x})$ . Para poder garantizar que nuestro estudio se realiza en las mejores condiciones (y las mas sencillas de calcular) supondremos que:

- Todas las medidas son independientes entre sí: no hay correlación entre las medidas de una muestra.
- Las distribuciones de cada medida son idénticas entre sí, e iguales a la distribución de probabilidad de la población:

$$g_1(x_1) = g_2(x_2) = \dots = g_n(x_n) = f(x)$$

En general estas condiciones se dan en muchas medidas físicas aunque no son absolutas: podemos desconocer la distribución de probabilidad, o que no describa el comportamiento de la población; incluso no saber si realmente las medidas son estadísticamente independientes. Aquellas muestras que cumplan estas condiciones se le llaman **muestras aleatorias simples**.

Llamamos **estadístico** a una función de las variables estadísticas. Un estadístico es una variable aleatoria por el simple hecho de que son funciones de variables aleatorias. Por lo tanto un estadístico sigue una función de probabilidad. Lo veremos mas adelante.

Un **estimador S** de un parámetro  $\lambda$  cualquiera a un estadístico o función que nos permita estimar el parámetro que deseamos conocer:

$$S = S(x_1, x_2, \dots, x_n)$$

Para que un estimador este correctamente definido tiene que cumplir una serie de características, que son:

1. **Fidelidad:** se dice que un estimador lineal es fiel si para cualquier tamaño n de la muestra se cumple que:

$$\mathbb{E}\{S(x_1, x_2, \dots, x_n)\} = \lambda \quad (3.1)$$

2. **Consistencia:** un estimador se dice consistente si aumenta en precisión con el tamaño de la muestra:

$$\lim_{n \rightarrow \infty} \sigma(S) = 0 \quad (3.2)$$

3. **Eficiencia:** entre dos estimadores fieles y consistentes la eficiencia relativa será:

$$\eta = \frac{\mathbb{E}\{(S_1 - \lambda)^2\}}{\mathbb{E}\{(S_2 - \lambda)^2\}} = \frac{\sigma^2(S_1)}{\sigma^2(S_2)} \quad (3.3)$$

Siempre es preferido aquel estimador con la varianza mas pequeña, y la eficiencia nos puede ayudar a escoger entre varios estimadores, en el caso de duda.

## Estimación puntual de la media poblacional

La media poblacional  $\mu$  es fundamental en la estadística para la construcción de probabilidad de un fenómeno. Si consideramos el caso de un muestreo aleatorio simple con una distribución de probabilidad  $f(x)$  del que hemos tomado los datos  $\{x_i\}_{i=1}^n$ . El mejor estimador de la media poblacional será la media muestral  $\bar{x}$ , que se define como:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.4)$$

Para realmente comprobar que es el mejor (más eficiente, fiel y consistente) estimador de la media poblacional deberíamos comprobar punto por punto los anteriores requisitos anteriormente seleccionados. Sin embargo esto escapa de la intención de estos apuntes, muy básicos. En el caso de querer saber mas recomendamos la lectura de la bibliografía.

Como hemos dicho un estadístico es una variable aleatoria (ya que procede de variables aleatorias). Entonces el propio estadístico también sigue una distribución aleatoria. Ahora bien, ¿Cuál es la distribución que afecta a nuestro estadístico media muestral? Si la distribución de población madre (la de nuestras variables aleatorias iniciales)  $f(x)$  es normal entonces la suma de  $n$  variables normales seguirá una distribución normal. Entonces  $\bar{x}$  será una variable aleatoria de media  $\mu$  y varianza  $\sigma^2/n$ , es decir  $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ . Y si tipificamos y encontramos el cambio de variable:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad (3.5)$$

Tenemos que  $z$  seguirá una normal  $N(0, 1)$ . Aun en el caso de que  $f(x)$  no siga una densidad de probabilidad gaussiana, si tiene momentos finitos sabemos que sumando las suficientes variables aleatorias (es decir con una muestra lo suficientemente grande), la distribución de la media tenderá a una gaussiana.

## Estimación puntual de la varianza poblacional

Otro de los parámetros fundamentales de una distribución cualquiera es la anchura de la distribución, es decir, la varianza poblacional  $\sigma^2$ . Podríamos suponer que la varianza poblacional viene dado por el siguiente estimador:

$$\hat{s}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Pero a la hora de valorar su fidelidad nos aparece que:

$$\mathbb{E}\{\hat{s}^2\} = \frac{n-1}{n} \sigma^2$$

Por lo que este estimador no es fiel. Debido a la linealidad de la esperanza podríamos obtener que un estimador fiel sería:

$$\frac{n}{n-1} \mathbb{E}\{\hat{s}^2\} = \sigma^2 \implies \mathbb{E}\left\{\underbrace{\frac{n}{n-1} \hat{s}^2}_{s^2}\right\} = \sigma^2$$

Entonces el estimador lineal  $s^2$  si sería fiel, y lo podríamos construir de la siguiente forma:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.6)$$

La aparición de un factor  $n-1$  no es casualidad, ya que al introducir en la definición del estadístico un estadístico estamos imponiendo una ligadura entre los valores, ya que  $\bar{x}$  y  $\{x_i\}_{i=1}^n$  no son linealmente independientes. Es decir  $n-1$  son los grados de libertad del estimador  $s^2$ . Si  $n=1$  (solo tenemos un dato) esta forma de estimar la varianza sería bastante errónea.

En muchos casos la distribución de la varianza tiene un comportamiento asimétrico. Su forma depende tanto de  $f(x)$  como de  $n$ . De forma general se puede demostrar que el siguiente estimador lineal depende de una función llamada la chi cuadrada de person  $\chi^2$  con  $n-1$  grados de libertad:

$$\frac{(n-1)}{\sigma^2} s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{\sigma^2} = \chi_{n-1}^2 \quad (3.7)$$

La  $\chi^2$  tiene propiedades muy interesantes como:

1. La distribución de probabilidad de suma de dos variables que se comporten como  $\chi^2$  con  $n_1$  y  $n_2$  grados de libertad respectivamente es una variable que se comporta como una  $\chi^2$  de  $n_1 + n_2$  grados de libertad.
2. El valor esperado de una distribución  $\chi^2$  es  $n$ .
3. La varianza de una distribución  $\chi^2$  es  $2n$ .

Definiéndose la distribución de  $\chi^2$  para  $n$  grados de libertad con  $\lambda = \frac{n}{2}$  como:

$$f(\chi^2) = \frac{1}{\Gamma(\lambda)2^\lambda} (\chi^2)^{\lambda-1} e^{-\chi^2/2} \quad (3.8)$$

### Distribuciones de $\bar{x}$ para poblaciones madre normales

Como hemos mencionado si la variable aleatoria  $X$  tiene una distribución normal entonces  $\bar{x}$  tiene una distribución  $N(\mu, \frac{\sigma}{\sqrt{n}})$ . Si la función madre no tiene una distribución normal pero la muestra tiene un numero suficiente de datos ( $n \geq 30$ ) podemos crear un cambio de variables tal que  $z = \frac{x-\mu}{\sigma}$  de tal modo que la distribución se aproxime a una normal  $N(0, 1)$ . Sin embargo, muchas veces (o mas bien casi nunca) conocemos el valor de la varianza poblacional, y tenemos que usar nuestro estimador de la varianza poblacional, la varianza muestral  $s^2$ . En ese caso el cambio de variable será:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad (3.9)$$

Como podemos ver el cambio que se ha hecho es:

$$\sigma(\bar{x}) \simeq \frac{s}{\sqrt{n}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}} \quad (3.10)$$

Entonces podemos reescribir esto de tal forma que:

$$t = \frac{\frac{\bar{x}-\mu}{\sigma\sqrt{n}}}{s/\sigma} = \frac{\frac{\bar{x}-\mu}{\sigma\sqrt{n}}}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}}$$

Sea  $z$  una variable normal distribuida según  $N(0, 1)$  y sea una variable  $\chi_n^2$  distribuida según la chi cuadrado de Pearson con  $n$  grados de libertad tenemos que la variable

$$t_n = \frac{z}{\sqrt{\frac{\chi_n^2}{n}}} \quad (3.11)$$

sigue una distribución  $t$  de Student con  $n$  grados de libertad cuya densidad de probabilidad es:

$$f(t) = \frac{\Gamma(\frac{1}{2}(n+1))}{\Gamma(\frac{1}{2}n)\sqrt{\pi}\sqrt{n}} \left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)} \quad (3.12)$$

De todo esto podemos deducir que la distribución de la media muestral  $\bar{x}$  en el caso de no conocer la varianza poblacional  $\sigma^2$  sigue una distribución  $t$  de student con  $n-1$  grados de libertad (tras hacer el cambio de variables...). Cuando  $n$  es pequeña esta distribución de probabilidad es mas ancha que la normal, sin embargo si  $n$  es lo suficientemente grande la  $t$  de student tiende a una distribución gaussiana normal.

### 3.0.2 Intervalos de confianza

Normalmente, cuando tratamos de estimar un parámetro no solo nos interesa obtener un valor puntual, si no el rango donde debería encontrarse el verdadero valor del parámetro, con un determinado nivel de probabilidad o confianza. Recordemos que los propios estimadores son, a fin de cuentas, variables estadísticas que tienen incertidumbre. Aunque los intervalos de confianza se pueden dar a cualquier estimador (incluso el de la incertidumbre), en estos apuntes solo analizaremos los intervalos de confianza de la media poblacional estimada a partir de la media muestral y los intervalos de confianza para la estimación de la varianza a partir de la muestra en poblaciones madre normales (solo este caso).

#### Intervalos de confianza para la media poblacional

El objetivo ahora es obtener un rango, un intervalo en el cual con una probabilidad  $1 - \alpha$  se encuentre la media poblacional  $\mu$ . Entonces la probabilidad de que  $\mu$  este en el intervalo  $(a, b)$  será:

$$P(a < \mu < b) = 1 - \alpha$$

A  $1 - \alpha$  lo conocemos como el nivel de confianza y definirá el tamaño del intervalo. Para construir este intervalo de confianza nos apoyaremos en  $\bar{x}$ . Normalmente nos encontramos que nos dan un nivel de confianza en tanto por cierto, pero otras muchas nos dan la probabilidad de error. La probabilidad de error no es mas que el valor de  $\alpha$ , expresado también en tanto por ciento muchas veces. Estudiaremos ahora diferentes intervalos de confianza para la media poblacional, para poblaciones madre normales.

**A) Intervalo de confianza para la media  $\mu$  de una población madre con  $\sigma$  conocida**

Es obvio que si consideramos un intervalo  $(a, b)$  donde  $\mu$  se encuentre con una probabilidad  $1 - \alpha$ , que la media se encuentre por la izquierda ( $x < a$ ) como por la derecha del ( $b < x$ ) del intervalo tendrá una probabilidad de  $\alpha/2$ .

Dado que estamos considerando una población madre de la muestra de datos normal, de  $\sigma$  conocida. Entonces podemos crear una normal  $N(0, 1)$  a partir de la variable aleatoria:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Esta nueva variable nos permitirá estimar la media poblacional. El intervalo que buscamos lo podremos obtener a partir de los percentiles de la distribución normal estándar que dejan a su izquierda y derecha un área igual a  $\alpha/2$ . El valor en el que justamente toda el área hasta ese momento ocupa dicha área lo llamaremos  $-z_{\alpha/2}$ . Para el caso de área superior  $z_{\alpha/2}$ . Entonces:

$$-z_{\alpha/2} \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2} \quad (3.13)$$

Lo cual implica que el intervalo de confianza será:

$$\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \quad (3.14)$$

Ahora imaginemos que tenemos una población, y queremos obtener el tamaño de la muestra para que con una probabilidad  $1 - \alpha$  la diferencia entre la media muestral  $\bar{x}$  y la media poblacional  $\mu$  sea menor que  $D$ . Es decir, trataremos de obtener el tamaño de la muestra para el cual la media poblacional este a una distancia máxima  $D$  del valor verdadero con una probabilidad determinada. Entonces con estas consideraciones:

$$-\frac{\sigma}{\sqrt{n}} z_{\alpha/2} \leq \mu \leq \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \quad (3.15)$$

Por lo que la diferencia máxima  $D$ :

$$D = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (3.16)$$

Y si despejamos  $n$ :

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{D^2} \quad (3.17)$$

**B) Intervalo de confianza para la media  $\mu$  de una población madre con  $\sigma$  desconocida**

En este caso la diferencia es que en vez de obtener los percentiles y valores a partir de una normal  $N(0, 1)$  usando el cambio de variable  $z$ , usaremos la  $t$  de student para evaluar el intervalo de confianza. Entonces si consideramos exactamente lo mismo para el parámetro  $t$ :

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$



Que seguirá la distribución  $t$  de student con  $n - 1$  grados de libertad. Entonces para este caso:

$$-t_{\alpha/2} \leq \frac{\bar{x} - \mu}{s/\sqrt{n}} \leq t_{\alpha/2} \quad (3.18)$$

Lo cual implica que el intervalo de confianza será:

$$\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha/2} \leq \mu \leq \bar{x} + \frac{s}{\sqrt{n}} t_{\alpha/2} \quad (3.19)$$

Recordemos que la  $t$  de student cambia en función del número de datos a considerar.

### Intervalos de confianza para la varianza $\sigma^2$ de una distribución normal

Como hemos visto la variable aleatoria  $(n - 1)s^2/\sigma^2$  sigue una distribución  $\chi^2$  de Pearson con  $n-1$  grados de libertad. Ahora introduciremos un pequeño cambio, muy sutil respecto a lo anterior. Denotaremos como  $\chi_{\alpha/2,n}^2$  el valor que deja a su derecha un área  $\alpha/2$ . Entonces:

$$P \left[ \chi_{1-\alpha/2,n-1}^2 \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi_{\alpha/2,n-1}^2 \right] = 1 - \alpha \quad (3.20)$$

De lo que podemos deducir que:

$$P \left[ \frac{(n-1)s^2}{\chi_{\alpha/2,n-1}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{1-\alpha/2,n-1}^2} \right] = 1 - \alpha \quad (3.21)$$

Muchas veces en realidad no nos importa que la sigma se encuentre por debajo de el intervalo. De hecho, hasta nos conviene: cuanto mas se acerque a cero mejor, ya que significa que el experimento está mejor realizado. Entonces en ese caso podemos reescribir todo esto, teniendo en cuenta que ahora  $\chi_{1-\alpha,n}^2$  es aquel valor que deja a la derecha a la variable aleatoria  $\chi_{\alpha/2,n}^2$  con un nivel de confianza  $\alpha$ . Entonces como antes:

$$P \left[ \chi_{1-\alpha,n-1}^2 \leq \frac{(n-1)s^2}{\sigma^2} \right] = 1 - \alpha \quad (3.22)$$

De lo que podemos deducir que:

$$P \left[ \sigma^2 \leq \frac{(n-1)s^2}{\chi_{1-\alpha,n-1}^2} \right] = 1 - \alpha \quad (3.23)$$

### 3.0.3 Test de hipótesis estadísticas

Una hipótesis estadística es una afirmación que se realiza sobre una o mas características de la población. Una hipótesis de este tipo, puede ser, por ejemplo, la afirmación de que la fuerza que se ejerce sobre un muelle es proporcional al estiramiento del mismo con una constante  $k$ . Hipótesis de este tipo deben ser verificadas o contrastadas mediante la observación empírica o experimental de un fenómeno. Una hipótesis estadística estaría completamente contrastada si pudiéramos verificarlo con todos los elementos de la población, pero dado que la población tiene infinitos elementos (o no es accesible conseguirlos todos) en general la verificación estadística de una hipótesis se realiza mediante muestras de la población.

Consideramos dos tipos de hipótesis. La hipótesis simple y la hipótesis compuesta se diferencian principalmente en que la hipótesis simple establece un valor concreto para el parámetro de estudio (o varios) y la compuesta afirma un intervalo o recorrido para ellos.

El test de hipótesis alternativas se presenta mediante la contraposición de dos hipótesis, una que afirma lo que queremos que afirme (por ejemplo que el valor de la constante es  $k$ ) y otra que considera todos los demás casos. Se llaman, respectivamente, hipótesis nula  $H_0$  e hipótesis alternativa  $H_a$ :

$$\begin{aligned} H_0 ; \mu &= \mu_0 \rightarrow \text{hipótesis nula} \\ H_a ; \mu &\neq \mu_0 \rightarrow \text{hipótesis alternativa} \end{aligned} \quad (3.24)$$

El proceso de test de una hipótesis estadística sigue los siguientes pasos:

1. Definición de una hipótesis nula  $H_0$  y una hipótesis alternativa  $H_a$
2. Definición de una medida de discrepancia entre los datos muestrales y la hipótesis  $H_0$ . Para ello una función de los datos de la muestra.
3. Definición de la región de aceptación de  $H_0$ , constituida por el conjunto de valores que consideramos aceptable si  $H_0$  es cierta. Para ello debemos fijar un nivel de significación  $\alpha$  que es la probabilidad de rechazar  $H_0$  siendo correcta. En general es la probabilidad de fallar en la decisión de descartar o aceptar la hipótesis. La probabilidad de aceptar  $H_0$  siendo falsa se denota por  $\beta$ . A estos parámetros se les suele llamar error de tipo I y tipo II respectivamente.
4. Cálculo de la función de decisión llamado estadístico de contraste para aceptar o rechazar  $H_0$ .

Ahora analizaremos diferentes test de hipótesis estadísticas.

### Test de la media de una población normal

Consideremos una población gaussiana de media  $\mu$ . Entonces:

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_a &: \mu \neq \mu_0 \end{aligned} \quad (3.25)$$

Escogemos como estadístico de contraste:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Evidentemente mide la distancia (normalizada, ya que esta dividido entre la incertidumbre) entre los datos muestrales y la hipótesis estadística. Este estadístico sigue una  $t$  de Student con  $n - 1$  grados de libertad. La región crítica  $c$  es el valor de  $t$  para el cual rechazamos la hipótesis a un nivel de confianza  $1 - \alpha$ .

$$c = \{t : |t| > t_{\alpha/2; n-1}\} \quad (3.26)$$

Y la región de aceptación:

$$\Delta = \{t : |t| \leq t_{\alpha/2; n-1}\} \quad (3.27)$$

Aceptando la hipótesis si:

$$\frac{|\bar{x} - \mu_0|}{s/\sqrt{n}} \leq t_{\alpha/2; n-1} \quad (3.28)$$

Y rechazándola si:

$$\frac{|\bar{x} - \mu_0|}{s/\sqrt{n}} > t_{\alpha/2; n-1} \quad (3.29)$$

Entonces podremos aceptar la hipótesis siempre que el estadístico se encuentre en la región  $\Delta$ , o rechazarlo si está en la región  $c$ . Este contraste es de tipo bilateral, ya que lo rechazamos o aceptamos la hipótesis si el estadístico se encuentra por debajo o por encima de unos valores. Si solo quisiéramos rechazar datos en que el estadístico estuviera por encima o por debajo de unos valores tendríamos que hacer lo mismo pero cambiando un poco las regiones:

$$c = \{t : t > t_{\alpha; n-1}\} \quad (3.30)$$

Y la región de aceptación:

$$\Delta = \{t : t \leq t_{\alpha; n-1}\} \quad (3.31)$$

Esto está muy relacionado con los intervalos de confianza. Podemos visualizar un intervalo de confianza como el conjunto de hipótesis aceptables para ese nivel de confianza.

### 3.0.4 Método de máxima verosimilitud

Para completar el tratamiento de los métodos de determinación de parámetros vamos a introducir en la presente sección el método de máxima verosimilitud, que nos permitirá contemplar el problema de la estimación de parámetros y los ajustes de datos desde una perspectiva diferente. Supongamos que tenemos información para considerar que un determinado fenómeno se encuentra gobernado por una cierta distribución de probabilidad que depende de una serie de parámetros en principio desconocidos  $a_1, a_2, \dots, a_n$ . Para determinar estos parámetros debemos realizar un experimento para obtener los datos y calcular los parámetros. Pues bien, el método de máxima verosimilitud establece que a posteriori, una vez conocidos los datos, los parámetros óptimos para la distribución de probabilidad son aquellos que maximizan la probabilidad de obtención de ese conjunto de resultados experimentales. Este método es especialmente útil y sencillo en el caso de distribuciones normales.

Supongamos que esperamos que nuestros datos  $\{x_i\}_{i=1}^n$  siguen una cierta distribución de probabilidad que está parametrizada por una serie de parámetros  $(a_1, a_2, \dots, a_n)$ . Si denotamos por  $p_i$  la probabilidad de haber obtenido un cierto resultado  $x_i$ :

$$p \equiv p(x_i; a_1, a_2, \dots, a_n) \quad (3.32)$$

Si se verifica la independencia estadística de los resultados sucesivos del experimento aleatorio la probabilidad de haber obtenido el conjunto  $\{x_i\}_{i=1}^n$  será:

$$P \equiv p_1 \cdot p_2 \cdot p_3 \cdots p_n = \prod_{i=1}^n p_i \quad (3.33)$$

Esta probabilidad, una vez obtenidos los datos, se denomina función de verosimilitud:

$$L(a_1, a_2, \dots, a_n) = \prod_{i=1}^n p_i \quad (3.34)$$

Evidentemente la función de verosimilitud no puede tener un valor mayor que el producto de la probabilidad de cada uno de los elementos sin parametrizar (individualmente, tal que  $p_i = p(x_i)$ ). Está claro que, va a haber parámetros para los que este valor sea menor que el de los elementos individuales, y por lo tanto los valores que busquemos en la parametrización, siguiendo el método de máxima verosimilitud, serán aquellos que mas se acerquen a esta máxima probabilidad; i.e. trataremos de maximizar la función verosimilitud.

### Estimador de la media poblacional

Vamos a estudiar desde este nuevo punto de vista los estimadores de parámetros de una población a través de su muestra, tratando de calcular el parámetro  $\mu$ . Supongamos que tomamos  $n$  datos  $x_i$  que sabemos que siguen una distribución gaussiana de varianza  $\sigma^2$ , la probabilidad de obtener un dato será:

$$p_i(\mu) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2} \right] \quad (3.35)$$

Considerando la función de verosimilitud, que dependerá en este caso únicamente de la media poblacional:

$$L(\mu) = \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left[ -\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2} \right] \quad (3.36)$$

Como podemos ver maximizar la probabilidad de la función verosimilitud es minimizar la exponencial, y eso conlleva a minimizar una  $\chi^2$  de Pearson:

$$\chi^2(\mu) = -\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2} \quad (3.37)$$

Si tratamos de minimizar ahora la  $\chi^2$  de Pearson podemos llegar directamente al resultado obvio: que el parámetro óptimo de la media poblacional es la media muestral:

$$\mu' = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.38)$$

Y que su incertidumbre:

$$\sigma_{\mu'}^2 = \frac{\sigma^2}{n} \quad (3.39)$$

Ahora supongamos que cada uno de estos datos tiene una incertidumbre  $\sigma_i$  única asociada. Entonces si repetimos el mismo proceso tal que en este caso la función  $\chi^2$  de Pearson a minimizar es:

$$\chi^2(\mu) = -\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_i^2} \quad (3.40)$$

Entonces ahora el parámetro óptimo de la media poblacional será la llamada media ponderada, que es:

$$\mu' = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \quad (3.41)$$

La incertidumbre de la media ponderada en este caso será:

$$\sigma_{\mu'} = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} \quad (3.42)$$

### Ajuste por el método de mínimos cuadrados

En muchos casos de interés existe una correlación entre dos magnitudes medibles,  $x$  e  $y$ . La curva  $y = f(x)$  marque esta relación no es otra cosa que una hipótesis estadística de la dependencia de  $y$  respecto la variable independiente  $x$  (podríamos considerar el caso contrario, pero no cambiaría nada). A partir de este supuesto podemos obtener los parámetros óptimos de la función e ajuste que serán aquellos que maximicen la probabilidad de haber obtenido la muestra experimental de datos.

Antes de estudiar el tipo de ajustes por mínimos cuadrados que podemos hacer vamos a considerar un par de premisas, que usaremos de manera general. La primera que supondremos será que los datos  $x$  se verán afectados por una incertidumbre mucho menor que la de los datos  $y$ . De tal modo que  $\sigma_{y_i} \gg \sigma_x$ . Otra de las consideraciones que haremos (aunque será un caso particular de la anterior) es suponer que todas las incertidumbres de los datos  $y$  son iguales entre sí. Es decir:

$$\sigma_{y_1} = \sigma_{y_2} = \dots = \sigma_{y_n} = \sigma_y$$

Una vez supuesto esto estudiaremos cada uno de los casos. Primero veremos el método general de aproximación (a una función general con  $k$  parámetros) y luego miraremos casos mas concretos, como un ajuste lineal. También supondremos que la función madre de las  $y_i$  son gaussianas.

#### A) Caso general función arbitraria

En general no podemos hacer un ajuste a cualquier función arbitraria analítica, o cualquier función en general. Sin embargo el problema si admite una solución analítica siempre que sea la función lineal. Lineal nos referimos a que sea del tipo:

$$f(x, a_1, a_2, \dots, a_k) = \sum_{i=1}^k a_i f_i(x) \quad (3.43)$$

Pudiendo ser cualquier función analítica la  $f_i$ . La única condición es que los parámetros  $a_i$  deben multiplicar a la función y no encontrarse dentro de ella. En cualquier otro caso no podemos usar el ajuste de mínimos cuadrados para minimizar el ajuste.

$$y = f(x, a_1, a_2, \dots, a_k) \quad (3.44)$$

Para maximizar la función de verosimilitud tendremos que minimizar la función  $\chi^2$  de Pearson que viene dada por:

$$\chi^2 = \sum_{i=1}^n \frac{[y_i - f(x_i, a_1, a_2, \dots, a_k)]^2}{\sigma_i^2} = \sum_{i=1}^n \frac{[y_i - \sum_j a_j f_j(x_i)]^2}{\sigma_i^2} \quad (3.45)$$

Ahora solo tendríamos que derivar parcialmente respecto cada uno de estos parámetros e igualar a cero. De esto obtenemos un problema matricial tal que:

$$Y = HA \quad (3.46)$$

Siendo Y una matriz  $[k \times 1]$ , H una matriz  $[k \times k]$  y A la matriz que contiene cada uno de los parámetros que queremos calcular,  $[k \times 1]$ . Entonces el problema a resolver sería:

$$A = H^{-1}Y \quad (3.47)$$

Siendo, mas en concreto:

$$H \equiv \begin{pmatrix} \sum \frac{f_1^2(x_i)}{\sigma_i^2} & \dots & \sum \frac{f_1(x_i)f_k(x_i)}{\sigma_i^2} \\ \vdots & & \vdots \\ \sum \frac{f_1(x_i)f_k(x_i)}{\sigma_i^2} & \dots & \sum \frac{f_k^2(x_i)}{\sigma_i^2} \end{pmatrix} \quad (3.48)$$

$$Y \equiv \begin{pmatrix} \sum \frac{y_i f_1(x_i)}{\sigma_i^2} \\ \vdots \\ \sum \frac{y_i f_k(x_i)}{\sigma_i^2} \end{pmatrix} \quad (3.49)$$

Además podemos obtener las incertidumbres y las covarianzas entre los parámetros usando esta matriz  $H^{-1}$ . Sea  $(H^{-1})_{jk}$  el elemento fila  $j$  y columna  $k$  de la matriz tenemos que:

$$\sigma_{a_j}^2 = (H^{-1})_{jj} \quad (3.50)$$

$$\text{cov}(a_j, a_k) = (H^{-1})_{jk} \quad (3.51)$$

Como la matriz H es simétrica tenemos que las covarianzas son iguales. Una vez visto el caso general vamos a estudiar la regresión lineal.

## B) Regresión lineal con incertidumbres no constantes

Antes hemos dicho que la función a minimizar era la  $\chi^2$  de Pearson de ese tipo, pero, ¿Como llegamos ahí? Para verlo mejor (ya que en el caso anterior sería una embrollo de operaciones poco ilustrativo) usaremos este caso. Consideremos que la probabilidad de obtener un cierto valor  $y_i$  tiene una distribución normal gaussiana de media  $\bar{y}$ . Entonces:

$$p_i(\mu) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[ -\frac{1}{2} \frac{(y_i - \bar{y}_i)^2}{\sigma_i^2} \right] \quad (3.52)$$

Es razonable suponer que las medias de las distribuciones de los diferentes puntos experimentales se sitúan sobre la recta de regresión (ya que estamos suponiendo que esa regresión se verifica, solo tenemos que calcular los parámetros para los que esa función se verifica “mejor”). Por lo tanto:

$$\bar{y}_i = a + bx_i \quad (3.53)$$

Entonces la función de verosimilitud:

$$L(\mu) = \left( \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} \right) \exp \left[ -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - a - bx_i)^2}{\sigma_i^2} \right] \quad (3.54)$$

Y por lo tanto maximizarla equivale a minimizar la  $\chi^2$  de Pearson tal que:

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - a - bx_i)^2}{\sigma_i^2} \quad (3.55)$$

Ahora obtenemos los resultados directamente (los cálculos, una vez llegado hasta aquí, es lo mas sencillo):

$$a = \frac{\sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \sum_{i=1}^n \frac{y_i}{\sigma_i^2} - \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \sum_{i=1}^n \frac{x_i y_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \left( \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2} \quad (3.56)$$

$$b = \frac{\sum_{i=1}^n \frac{1}{\sigma_i^2} \sum_{i=1}^n \frac{x_i y_i}{\sigma_i^2} - \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \sum_{i=1}^n \frac{y_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \left( \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2} \quad (3.57)$$

Y ahora para calcular las incertidumbres de  $a$  y  $b$  consideraremos que a partir de la fórmula de propagación de incertidumbres:

$$\sigma_a^2 = \sum_{j=1}^n \sigma_j^2 \left( \frac{\partial a}{\partial y_j} \right)^2 \quad (3.58)$$

$$\sigma_b^2 = \sum_{j=1}^n \sigma_j^2 \left( \frac{\partial b}{\partial y_j} \right)^2 \quad (3.59)$$

De lo que podemos obtener directamente las incertidumbres:

$$\sigma_a^2 = \frac{1}{\Delta} \sum_{j=1}^n \frac{x_j^2}{\sigma_j^2} \quad (3.60)$$

$$\sigma_b^2 = \frac{1}{\Delta} \sum_{j=1}^n \frac{1}{\sigma_j^2} \quad (3.61)$$

Siendo  $\Delta$ :

$$\Delta = \sum_{j=1}^n \frac{1}{\sigma_j^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \left( \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \right)^2 \quad (3.62)$$

**C) Regresión lineal con incertidumbre constante** Para el caso en el que:

$$\sigma_{y_1} = \sigma_{y_2} = \dots = \sigma_{y_n} = \sigma_y$$

Podemos obtener los valores de  $a$  y  $b$  como:

$$a = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (3.63)$$

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (3.64)$$

Y sus incertidumbres:

$$\sigma_a^2 \simeq \frac{1}{\Delta} \sum x_i^2 \left[ \frac{1}{n-2} \sum (y_i - a - bx_i)^2 \right] \quad (3.65)$$

$$\sigma_b^2 \simeq \frac{1}{\Delta} \left[ \frac{1}{n-2} \sum (y_i - a - bx_i)^2 \right] \quad (3.66)$$

Teniendo en cuenta que  $\Delta$  es:

$$\Delta = n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \quad (3.67)$$



# Chapter 4

## Incertidumbre en la medida

### 4.0.1 Introducción: error e incertidumbre

Normalmente asociamos indistintamente error e incertidumbre, cuando son totalmente diferentes. La **incertidumbre** es un parámetro asociado con el resultado de una medida, inherente a cualquier dato empírico (sea físico, químico...), que caracteriza la dispersión de los valores que se pueden atribuir razonablemente a un mensurando. El **error** es la diferencia entre una medida y el resultado del valor verdadero del mensurando.

Es importante diferenciar esto, ya que cuando hablamos de error en la medida es algo que no podemos conocer, ya que no conocemos el valor verdadero de un mensurando, por eso durante muchos años conllevó a muchos problemas. Antes se distinguían varios tipos de errores (errores relativos, aleatorios...) que más tarde llevaron a la creación de la incertidumbre de tipo A y de tipo B. Vamos a ver estos diferentes errores más que nada porque en muchos casos aun en el presente se siguen usando, y conocerlos no está de más. Estos son:

- **Error relativo:** error de una medida dividido por el valor real del mensurando.
- **Error aleatorio:** resultado de la medida menos el valor medio que resultaría de infinitas medidas del mismo mensurando bajo condiciones de repetitividad. Proviene de variaciones estocásticas en las variables que influyen en el mensurando; estos errores se reducen incrementando el número de observaciones, de forma que su valor esperado tiende a cero.
- **Error sistemático:** media que resultaría de un conjunto infinito de medidas de un mensurando bajo condiciones de repetitividad, menos el valor real del mensurando. Son producidos por efectos reconocidos que provocan desviaciones en la medida, y que deben cuantificarse y corregirse si son significativos respecto a la exactitud requerida.

Como podemos ver estas definiciones son un poco ambiguas e imposibles de calcular. En primer lugar es imposible conocer el valor real del mensurando, por lo que estos errores ya serían imposibles de conocer. Lo de obtener medidas infinitas, en la práctica, también resultaría imposible. En general se usaba el error aleatorio como el error estadístico de la medida y en el sistemático todos aquellos factores que pudieran alterar la medida.

### 4.0.2 Incertidumbre tipo A y tipo B

Como ya hemos dicho, la incertidumbre es un parámetro asociado con el resultado de una medida que caracteriza la dispersión de los valores que se pueden atribuir razonablemente a un mensurando. Las fuentes de la incertidumbre en la medida pueden ser, entre otras:

- Definición incompleta del mensurando o realización experimental imperfecta de la definición del mensurando.
- Muestra no significativa o que no represente el mensurando adecuado.
- Conocimiento inadecuado de los efectos del entorno en el proceso de medida o medida imperfecta de las condiciones ambientales que afectan al proceso.
- Sesgo personal en la medida de los instrumentos analógicos.
- Resolución limitada en la instrumentación o umbrales de discriminación.
- Valores inexactos en los patrones, estándares de medida o de los materiales de referencia.
- Aproximaciones o suposiciones erróneas incorporadas en el proceso de medida.
- Valores inexactos de constantes y otros parámetros obtenidos de fuentes externas y utilizados en el algoritmo de tratamiento de datos.
- Variaciones en las observaciones repetidas del mensurando bajo condiciones aparentemente idénticas de medida.

En el laboratorio es muy importante saber de donde viene la incertidumbre, y conocer al máximo nuestro experimento para plasmar correctamente todas aquellas fuentes de incertidumbre. Cuantas mas conozcamos, y mejor las conozcamos, pues mas precisa será la incertidumbre. Además podemos distinguir dos tipos de incertidumbres en la medida: la incertidumbre de tipo A y la incertidumbre de tipo B.

La **incertidumbre de tipo A** se evalúa a partir de la serie de observaciones repetidas utilizando la varianza o desviación estándar de las funciones de densidad de probabilidad, derivadas de las distribuciones de frecuencia de resultados.

La **incertidumbre de tipo B** se evalúa utilizando el conocimiento previo obtenido, tipificado como varianza o desviaciones estándar obtenidas de funciones de densidad de probabilidad asumidas sobre la probabilidad (subjetiva) de que ocurra un suceso.

Veamos como se evalúan de manera mas concreta en los distintos casos

#### A) Incertidumbre de tipo A

La incertidumbre de tipo A se calcula usando estimadores estadísticos. Supongamos que tenemos  $n$  datos calculados. Entonces un buen estimador de la varianza de la media muestral es:

$$s^2(\bar{x}) = \frac{s^2(x_i)}{n} \quad (4.1)$$

Y por lo tanto tenemos que la desviación típica estadística es decir la incertidumbre típica de tipo A será:

$$u_A(\bar{x}) = \frac{s}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (4.2)$$

## B) Incertidumbre de tipo B

Para un estimador de una variable que no se ha obtenido se e valua a partir del juicio científico basado en las fuentes de información disponibles sobre la probabilidad de x. Para explicar correctamente la incertidumbre de tipo B debe explicitar la distribución de probabilidad que se ha usado, aunque si no se hace asumiremos que son distribuciones normales. La incertidumbre de tipo B se puede calcular en base a:

- Datos tomados previamente a los que no se tiene un acceso individualizado, resultados previos de sistemas equivalentes.
- Conocimiento general o experiencia con el comportamiento o propiedades de materiales relevantes o instrumentos de medida.
- Especificaciones del fabricante
- Datos obtenidos en calibraciones o en certificaciones de los aparatos
- Incertidumbres asignadas a datos de referencia o tomadas de bibliografía.

La incertidumbre de tipo B se adecua en cada caso: a veces es necesario aplicar una distribución gaussiana, otras triangular, cuadrada...

### 4.0.3 Incertidumbre combinada

Asumiendo que las incertidumbres de tipo A y tipo B son estadísticamente independientes obtendremos una incertidumbre combinada para la media  $u_C(x)$  definida como:

$$u_C^2(x) = u_A^2(x) + u_B^2(x) \quad (4.3)$$

Esto no es mas que un caso de propagación (combinación de incertidumbres estadísticamente independientes), que veremos a continuación de forma general para una magnitud de salida  $y = f(x_1, x_2, \dots, x_n)$  a partir de las magnitudes de entrada  $x_i$ . Cuando el resultado de una medición se obtiene a partir de lso valores de otras magnitudes la incertidumbre típica de este resultado se denomina incertidumbre típica combinada y se denota por  $u_c$ .

Supongamos un mensurando  $y$  procedente de  $n$  variables (no tienen porque ser independientes). En ese caso realizaremos  $p$  medidas de tal forma que:

$$y^i = f(x_1^i, x_2^i, x_3^i, \dots, x_n^i)$$

Si hacemos el desarrollo en serie de Taylor de la función centrada en la media de las variables tenemos que:

$$y^i = f(x_1^i, x_2^i, x_3^i, \dots, x_n^i) \simeq f(\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_n) + \sum_{j=1}^n \left. \frac{\partial f}{\partial x_j} \right|_{\bar{x}_j} (x_j^i - \bar{x}_j) \quad (4.4)$$

Si ahora calculamos la media de  $y$  tenemos que:

$$\bar{y} = \frac{1}{p} \sum_{i=1}^p y_i \simeq \sum_{i=1}^p \left( f(\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_n) + \sum_{j=1}^n \left. \frac{\partial f}{\partial x_j} \right|_{\bar{x}_j} (x_j^i - \bar{x}_j) \right) \quad (4.5)$$

Como sabemos que  $\sum_{i=1}^p (x_j^i - \bar{x}_j) = 0$  tenemos que:

$$y \simeq f(\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_n) \quad (4.6)$$

Siendo esta una aproximación de primer orden, que, en general, siempre nos funcionará. Como sabemos la incertidumbre de tipo A viene dado por el estimador:

$$u^2(\bar{y}) = \frac{1}{p(p-1)} \sum_{i=1}^p (y^i - \bar{y})^2 \simeq \frac{1}{p(p-1)} \sum_{i=1}^p \left( \sum_{j=1}^n \left. \frac{\partial f}{\partial x_j} \right|_{\bar{x}_j} (x_j^i - \bar{x}_j) \right)^2 \quad (4.7)$$

Teniendo en cuenta que:

$$\text{cov}(\bar{x}_j, \bar{x}_k) = \frac{1}{p(p-1)} \sum_{i=1}^p (x_j^i - \bar{x}_j)(x_k^i - \bar{x}_k); \quad u^2(\bar{x}_j) = \frac{1}{p(p-1)} \sum_{i=1}^p (x_j^i - \bar{x}_j)^2$$

Podemos extraer fácilmente que la incertidumbre de tipo A de la magnitud  $y$  derivada de la serie de  $p$  medidas y  $n$  variables:

$$u_A^2(\bar{y}) = \sum \left( \left. \frac{\partial f}{\partial x_j} \right|_{\bar{x}} \right)^2 u^2(\bar{x}_j) + 2 \sum_{j=1}^{n-1} \sum_{k=j+1}^n \left( \left. \frac{\partial f}{\partial x_j} \right|_{\bar{x}} \right) \left( \left. \frac{\partial f}{\partial x_k} \right|_{\bar{x}} \right) \text{cov}(\bar{x}_k, \bar{x}_j) \quad (4.8)$$

Esta será la expresión general de la incertidumbre de tipo A propagada. Sin embargo si a la hora de escribir  $u$  (y la covarianza) incluimos ya las variables de tipo A y de tipo B tendremos la incertidumbre combinada.

Cuando las variables  $x_i$  son independientes entre sí la covarianza entre ellas se anula. Es un caso muy significativo que simplifica el cálculo de la incertidumbre:

$$u^2(\bar{y}) = \sum \left( \left. \frac{\partial f}{\partial x_j} \right|_{\bar{x}} \right)^2 u^2(\bar{x}_j) \quad (4.9)$$

### Factor de cobertura

Cuando se conocen los niveles de confianza asociados o probabilidades asociadas a que el resultado se encuentre dentro de un número determinado de incertidumbres alrededor de la media, se puede utilizar un factor de cobertura  $k_c$  aplicado a la incertidumbre combinada, para así expresar la incertidumbre expandida  $U$ :

$$U = k_c \cdot u_c \quad (4.10)$$

Por ejemplo para una magnitud que cuya probabilidad sea normal los factores 1,2,3 se corresponden a niveles de confianza del 68,27%, 95,45% y 99,73%.

En muchos casos se suele considerar que una mejor aproximación a la incertidumbre expandida se obtiene considerando que la magnitud de salida sigue aproximadamente una  $t$  de Student con un número de grados de libertad  $v$ .

### Como expresar las incertidumbres

Utilizaremos la incertidumbre estándar combinada  $u_c(y)$  como el indicador de la incertidumbre de nuestras medidas. Una expresión correcta de la incertidumbre incluye:

- Describir claramente los métodos usados para calcular los resultados y sus incertidumbres.
- Hacer una lista de los componentes de la incertidumbre y documentar como se evaluaron.
- Presentar el análisis de datos de forma que cada paso importante se pueda entender y seguir y reproducir su cálculo de manera independiente.
- Dar todas las correcciones y constantes usadas en el análisis.

Para expresar la incertidumbre y el dato final podemos aplicar las siguientes expresiones (con ejemplos):

- $Z = 100,02123 \text{ g}$  con  $u_c = 0.35 \text{ mg}$ .
- $Z = 100,02123(35) \text{ g}$
- $Z = 100,02123(0,00035) \text{ g}$

## 4.0.4 Sistema Internacional de Unidades

### 4.0.5 Cifras significativas

La precisión de una medida no queda determinada por el número de dígitos con el que se exprese, ni por el número de éstos a la derecha de la coma decimal.

La precisión de un resultado experimental está implícita en el número de dígitos con los que se expresa la medida cuando esta se expresa correctamente, aunque hemos de añadir también el valor de la incertidumbre de la medida para identificar las cifras realmente significativas.

Para expresar de forma correcta una medida utilizamos las cifras significativas, que son las cifras que aportan información no ambigua ni superflua sobre la medida experimental. Una cantidad mayor de cifras significativas tiene una mayor precisión relativa, ya que se supone que la incertidumbre actúa sobre la última cifra significativa. Las reglas básicas en el conteo de cifras significativas son:

- La primera cifra no nula empezando por la izquierda es la cifra mas significativas.

- Si no hay coma decimal, la primera cifra no nula empezando por la derecha es la cifra menos significativa.
- Si, y solo si, hay una coma decimal, la primera cifra por la derecha es la cifra menos significativa, incluso si es un 0.
- Todas las cifras entre la cifra más y menos significativas son cifras significativas.
- Se recomienda el uso de la notación científica para evitar la inclusión de ceros adicionales a la derecha de las cifras significativas.

Un punto muy importante es que las incertidumbres siempre se deben expresar con dos cifras significativas. Por tanto las medidas no podrán exceder el número de posiciones de cifras significativas dada por la incertidumbre, siendo incorrecto un valor  $x = 1,23452$  cuando la incertidumbre es de  $u(x) = 0,73$ .

Para truncar un número tenemos en cuenta las cifras a eliminar como una fracción decimal. Si por ejemplo queremos quedarnos con cuatro cifras del número  $x = 123,6575$  la fracción a eliminar será 0,575. Entonces consideramos los casos:

- Si la fracción a remanente es mayor que 0,5 aumentamos en una unidad la cifra menos significativa.
- Si la fracción remanente es menor que 0,5 dejamos la cifra menos significativa tal y como está.
- Si la fracción es igual a 0,5; consideramos dos casos. Si la cifra menos significativa es impar la subimos una unidad. Si la cifra es par, la dejamos tal y como está.

Es importante tener en cuenta que el redondeo debe realizarse únicamente sobre el resultado final de un cálculo para evitar la propagación y magnificación de diferencias en operaciones matemáticas.