INFERENCIA ESTADÍSTICA

Julio Muñoz

Departamento de Matemáticas

ETSI Industriales. Universidad de Castilla-La Mancha

18 de Septiembre de 2008

Índice general

Capítulo 1

Estadísticos y Probabilidades.

1.1. Introducción.

La importancia de la Estadística es clara, no hay duda sobre su arraigo en la vida cotidiana, en las noticias, en los sondeos de opinión, en controles de calidad, en pruebas médicas y tecnológicas, en medioambiente, etc. Trataremos a lo largo de esta memoria proporcionar los rudimentos indispensables para poder iniciarse en esta ciencia.

Nuestro punto de partida es el de la investigación. El pilar fundamental en el que se asientan los principios de investigación es el de la observación y experimentación. Así, el hombre estudia todo aquello que le rodea y lo analiza de muchas formas (observa en la naturaleza o experimenta en un laboratorio). Sin embargo, es la *Estadística* la herramienta que le ayuda a conocer su entorno, le ayuda a entender cada *fenómeno* de su entorno a través de su observación o lo que es lo mismo, a través de sus concreciones.
Básicamente, el estudio en términos estadísticos de un fenómeno, consiste en clasificar y resumir tales concreciones o datos experimentales, con el fin de poner de manifiesto alguna característica o ley del propio fenómeno que se está estudiando.

Así pues, asociado a un fenómeno están sus posibles *concreciones o sucesos*, lo que constituye el *espacio muestral*. La información relativa al fenómeno se recoge a través de un número finito de concreciones (realizaciones u observaciones). Esta información es generalmente aleatoria, y por esta razón todo estudio estadístico del fenómeno tiene por objetivo primero el precisar o medir tal aleatoriedad. Para ello sirve la *Teoría de la Probabilidad*: la fase deductiva que consiste en que a partir de unos hechos conocidos, se trate de deducir la probabilidad de que ocurran ciertos eventos de interés particular, se llama *Probabilidad*. Seguidamente hay una fase de tipo inductivo en la que con ayuda de la Teoría de la Probabilidad se elabora la información para extraer conclusiones sobre el fenómeno, y que se realiza a través de Estimadores, Intervalos de Confianza, Contrastes, etc., se denomina *Inferencia Estadística*.

Conviene resaltar que los datos o concreciones del fenómeno son nuestra base de datos y a partir de ésta y mediante la inferencia, trataremos de cuantificar la incertidumbre de ciertos sucesos relacionados directa o indierectamente con el fenómeno que estamos estudiando.

En una primera fase y con el objeto de realizar un estudio posterior eficaz, habremos de recopilar, organizar y rersumir de manera conveniente todos los datos observados. De hecho, este proceso de análisis de datos sirve para hacer una primera aproximación cuantitativa del fenómeno en el que estamos interesados. Esto es lo que se llama *Estadística Descriptiva*. Un análisis más general, basado esencialmente en métodos matemáticos y en un de la Teoría de Probabilidad, es el objetivo de la Inferencia Estadística.

Previo al estudio resulta crucial analizar la naturaleza de los fenómenos que son objeto de estudio: podemos decir que se clasifican en dos tipos diferentes. Los fenómenos *Deterministas*, aquéllos en los que se sabe a priori cuáles van a ser sus realizaciones o concreciones, y los fenómenos *Aleatorios*, que son los fenómenos en los que sus concreciones tienen carácter aleatorio (no se sabe con certeza qué es lo que va a ocurrir en la siguiente observación). Un ejemplo de fenómeno aleatorio es el lanzamiento de un dado. También lo es el de una moneda o la elección de una carta en una baraja de naipes. El fenómeno que consiste en medir el área de cualquier círculo es de tipo determinista. También lo es analizar qué sucede con una cantidad de agua cuando ésta permanece en unas condiciones ambientales de temperatura a 100 grados Centígrados y de presión a 760 mm. El suceso seguro será la evaporación.

1.1.1. Elementos básicos

En el estudio estadístico de un fenómeno aleatorio es imprescindible tener en cuenta los siguientes elementos:

i) El conjunto de todos los posibles resultados, el denominado *Espacio Muestral, Universo o Población.*

ii) El cómo asignar de manera 'correcta' una probabilidad a cada uno de los sucesos que componen el Espacio Muestral.

Para manejar la información en términos aleatorios también es preciso establecer la noción de muestra:

iii) Una *Muestra* será un subconjunto cualquiera del espacio muestral. Está dada a través de un número finito de observaciones del fenómeno. Es la base de datos indispensable para todo tipo de conclusión estadística.

Hemos indicado que el estudio de un fenómeno está basado en la observación y que para su análisis es necesario el cálculo de probabilidades. La manera clásica de definir la probabilidad de un suceso se realiza mediante el uso de la *frecuencia relativa.* Asigna, a partir de una muestra, la probabilidad de un suceso empleando para ello la conocida fórmula de número de casos favorables dividido entre casos posibles (*Fórmula de Laplace*). El número de casos favorables es el número de veces que ha ocurrido el suceso y el de los casos posibles no es sino el número de elementos que forman la muestra , lo que llamaremos frecuencia absoluta. Esta forma de asignar la probabilidad es la que explícitamente se usa en Estadística Descriptiva.

Usando sólo Estadística Descriptiva se puede establecer conclusiones interesantes. Sin embargo, y aunque esta herramienta es heurísticamente buena, sucede que en muchos casos no es capaz de cumplir o adecuarse a los requisitos i) o ii). Es más, en algunas situaciones aparentemente sencillas los cálculos son sumamente complicados, sino imposibles de realizar. Tales complicaciones son frecuentes en fenómenos cuyos espacios muestrales constan de un gran número de elementos.

Un ejemplo en el que resulta imposible aplicar los puntos anteriores es aquél en el que se pretende calcular la probabilidad de que un dardo haga impacto sobre cierto subconjunto de una diana (se supone que los dardos caen sobre la diana con igual probabilidad sobre cada uno de los puntos de la misma). Obsérvese que en este caso el espacio muestral es infinito, y que por ende el uso de frecuencias relativas, tal y como las hemos descrito antes, no tiene sentido.

Resumiendo: en un problema de Estadística el investigador ha de observar lo ocurrido a través del tiempo (por ejemplo, el precio de un determinado artículo) o a través de las distintas observaciones (pensemos en un laboratorio experimental). Su objetivo es tratar de prever qué sucederá en el futuro, es decir, intentar establecer una especie de ley o tendencia por la que se regirá el fenómeno que estamos tratando. Utilizando muestras y midiendo probabilidades el investigador tratará de dar información más o menos buena sobre cómo es la ley que describe al fenómeno o cuáles son los aspectos más interesantes que describen al mismo. Como ya hemos señalado, existen dos maneras de llevar a cabo esto, una mediante la **Estadística Descriptiva**, que es aquella que mediante el recuento, ordenación y clasificación de los datos calcula parámetros estadísticos para caracterizar a la población y asignar probabilidades a aquéllos eventos de interés; y la otra manera de realizar el estudio es mediante la Estadística Inductiva o **Inferencia Estadística**, que establece conclusiones generales (infiere conclusiones) sobre la ley a partir de los datos concretos de la muestra y del uso riguroso de los métodos matemáticos propios de la Teoría de Probabilidades.

La organización de este capítulo es como sigue: en la Sección 1.2 damos una serie de ejemplos para motivar el uso de la estadística en campo medioambiental. La Sección 1.3 está dedicada a una introducción a la Estadística Descriptiva, mientras que la sección última concierne a un estudio elemental de la Teoría de la Probabilidad.

1.2. Ejemplos

EJEMPLO 1.1 *Planteamos el siguiente problema de verosimilitud: en tres grandes poblaciones de reptiles, A, B y C, que ocupan un 40 %, 45 % y 15 % de un determinado sistema, la proporción de individuos infectados por un virus es del 30, 60 y 10 % respectivamente. Se toma al azar una de las 3 poblaciones, y de ella elegimos a 10 reptiles al azar, resultando que 2 de ellos están infectados. ¿A qué población es más probable que pertenezcan?*

El principio en el que está basado el método estadístico apropiado para contestar a esta cuestión es el de la máxima verosimilitud. Básicamente consiste en calcular la probabilidad que se da para cada una de las alternativas y elegir de éstas la que dé lugar a una probabilidad mayor. Del enunuciado del problema se desprende lo siguiente: si Inf denota al suceso estar infectado y

A, B y C haber elegido a dichas poblaciones, entonces

$$P(\,Inf|\,A) = 3/10,\;\; P(\,Inf|\,B) = 6/10,\;\; P(\,Inf|\,C) = 1/10,$$

y

$$P(A) = 0.\,4,\;\; P(B) = 0.\,45,\;\; P(C) = 0.\,15.$$

Se sabe que S el suceso consistente en la realización de 10 obsevaciones con resultado de 2 infecciones ha tenido lugar. La pregunta es saber cuál de las siguientes cantidades es mayor:

$$P(\,A|\,S),\;\; P(B|\,S),\;\; P(\,C|\,S).$$

Calculemos: como veremos más tarde la probabilidad condicionada $P(\,A|\,S)$ se calcula mediante

$$P(\,A|\,S) = \frac{P(A \cap S)}{P(S)}$$

donde ahora $P(\,S|\,A)$ es la probabilidad de que hayan tenido lugar 2 infecciones de 10 individuos sabiendo que la probabilidad de infección en A es $3/10$*. Hacemos lo mismo para determinar $P(\,B|\,S)$ y $P(\,C|\,S)$. Puesto que el denominador que aparece en el cálculo de $P(\,A|\,S)$, $P(\,B|\,S)$ y $P(\,C|\,S)$ es el mismo, basta con comparar los numeradores. Éstos son*

$$\left(\frac{3}{10}\right)^2\left(\frac{7}{10}\right)^8(0{,}4)\,,\left(\frac{6}{10}\right)^2\left(\frac{4}{10}\right)^8(0{,}45)\,,\left(\frac{1}{10}\right)^2\left(\frac{9}{10}\right)^8(0{,}15)\,.$$

Por tanto, el valor más grande es el correspondiente a $P(\,A|\,S)$ y en base al Principio de Máxima Verosimilitud, la población con más probabilidad de haber sido elegida es A. Los detalles sobre cómo se cálcular por ejemplo $P(A \cap S)$ se darán en la Seccion 1.4.

EJEMPLO 1.2 *Se afirma que la altura media μ de una planta adulta en un campo de cultivo determinado es de 18 centímetros. Estamos interesados en contrastar esta afirmación con un nivel de confianza del 99 %. Esto significa que hallaremos dos valores próximos entre sí (μ_1 y μ_2, $\mu_1 < \mu_2$) y alrededor de 18 ($18 \in (\mu_1, \mu_2)$) para los que se cumpliría la siguiente propiedad: de 100 muestreos que hiciéramos, al menos en 99 de ellos el computo de la altura media (media aritmética) estaría comprendida entre estos dos valores. La idea es pues establecer el valor de la media con una gran probabilidad y esto lo hemos de hacer mediante con ayuda de un solo muestreo.*

Por ejemplo, supongamos que el tamaño de la muestra que realizamos es $n = 100$ *(de 100 mediciones) y que los datos (medidas)* $h_1, h_2, ..., h_{100}$ *arrojan una media aritmética*

$$\overline{h} = \frac{h_1 + h_2 + ... + h_{100}}{100} = 1.78,$$

Supongamos además que se tiene lugar el cómputo del índice siguiente:

$$\overline{s} = \sqrt[2]{\frac{\left(h_1 - \overline{h}\right)^2 + \left(h_2 - \overline{h}\right)^2 + ... + \left(h_{100} - \overline{h}\right)^2}{99}} = 0{,}1.$$

En tal situación los valores μ_i *que determinamos mediante métodos estadísticos son* $\mu_1 = 1.754$ *y* $\mu_2 = 1.806$*, y por tanto, el intervalo en el que aseguramos va a estar la media con una confianza del 99 % es*

$$I = [1.754, 1.806]$$

EJEMPLO 1.3 *El nivel medio de azúcar en sangre de una población normal es conocido y resulta ser de* $\mu_0 = 20$*. Los resultados para esta muestra fueron de una media muestral* 18.5 *y una desviación* $4 : \overline{x} = 18.5$ *y* $s = 4$*. ¿Se puede decir que la muestra es comparable con la pobación a un determinado nivel de significación, digamos de* $\alpha = 0.05$*? O de otro modo, ¿Confirma este resultado empírico la hipótesis de que el nivel medio de azucar es 20?*

Para esta situación se analiza si el valor empírico de la media está muy lejos del valor teórico que se asume de antemano. El cómo de alejado puede estar se lleva a cabo así: se estudia si $|\overline{x} - \mu_0| \geq \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$*, y si tiene lugar esta desigualdad rechazaremos la hipótesis de* $\mu_0 = 20$*, o lo que es equivalente, diremos que la muestra extraida no es comparable con la población. De hecho así es, ya que* $|\overline{x} - \mu_0| = 1.5$ *y* $\frac{\sigma}{\sqrt{n}} z_{\alpha/2} = \frac{4}{\sqrt{40}}(1.96) = 1.2396$*. Hemos contrastado negativamente la hipótesis que sostiene que el nivel medio de azucar es 20.*

EJEMPLO 1.4 *Un investigador está preocupado por la contaminación por plomo de un determinado bosque en el que se supone hay 4 tipos de suelo distintos. Lleva a cabo un muestreo sobre los suelos y realiza mediciones obteniendo una serie de medicones de contenido de plomo en función del tipo de suelo:*

suelo 1	0.59	0.61	0.63	0.61	0.6
suelo 2	0.63	0.66	0.64	0.64	0.65
suelo 3	0.48	0.52	0.5	0.47	0.51
suelo 4	0.6	0.62	0.61	0.58	

Con estos nos preguntamos si existen diferencias significativas en el nivel de contaminación según sea el tipo de suelo. La respuesta, en base a los datos y a un análisis estadístico riguroso establece que sí hay distintos niveles de contaminación según sea el suelo que tomemos.

EJEMPLO 1.5 *Se desea contrastar la hipótesis de rendimientos medios para tres tipos de terrenos diferentes simultáneamente con la hipótesis de igualdad de rendimientos medios para cinco variedades de maíz. Se toman al azar 5 áreas de cada uno de los terrenos y se siembra cada área con un tipo de maíz. Después de recoger el maíz los rendimientos, en miles de kgs, vienen dados en la tabla*

tipo terrreno\variedad maíz	*a*	*a*	*c*	*d*	*e*
terreno 1	280	300	310	270	330
terreno 2	250	240	260	270	240
terreno 3	310	304	290	280	300

Como hemos indicado, las hipótesis sobre las que vamos a emitir juicio son:

H_1 : *igualdad de rendimientos para los tres tipos de terreno*

H_2 : *igualdad de rendimientos para las cinco variedades de maíz.*

Los métodos estadísticos van a establecer que en base a esto datos hemos de rechazar H_1 *y aceptar* H_2.

1.3. Estadística descriptiva

Tratamos de analizar los datos o concreciones recogidos en una muestra con el objeto de establecer ciertos parámetros asociados a la ley del fenómeno. En este tipo de análisis, de manera implícita, estamos suponiendo que lo que ocurre en toda la población está fielmente representado por la muestra con la que estamos trabajando.
El manejo, la síntesis y la representación gráfica de un número finito de datos - los que componen la muestra - es el tipo de análisis empleado para emitir algún tipo de información relativa al fenómeno. Para esto es preciso introducir el concepto de *variable estadística o aleatoria.*

DEFINICIÓN 1.1 *Una variable estadística se entiende como el símbolo que representa a todas las posibles concreciones del fenómeno.*

EJEMPLO 1.6 *X = edad de una persona que vive en Ciudad Real.*

Si la variable expresa un solo carácter de cada elemento del espacio muestral diremos que es unidimensional, si expresa dos diremos que es bidimensional, etc.

EJEMPLO 1.7 *Un ejemplo de variable bidimensional es: X = (edad, peso) de un alumno de primero de Ingeniería Química.*

DEFINICIÓN 1.2 *Sea X una variable aleatoria (v.a.) y sean los datos de ésta*

$$x_1, x_2, ..., x_N \tag{1.1}$$

(i.e., N sucesos o concreciones de la v.a. X). Se define:

$$\boxed{\begin{aligned} \textit{Frecuencia del dato } x_i &= n^0 \textit{ de veces que aparece } x_i \textit{ en (1.1)} \\ N &= \textit{frecuencia total} \\ \textit{Frecuencia relativa del dato } x_i &= \frac{\textit{frecuencia del dato } x_i}{N} \end{aligned}}$$

La Tabla de Frecuencias asociada a los resultados (1.1) se construye del siguiente modo

Valores o datos de X	Frecuencias	Frecuencia relativas
x_1	n_1	$\frac{n_1}{N}$
$\vdots$	$\vdots$	$\vdots$
x_r	n_r	$\frac{n_r}{N}$

NOTA 1.1 $\sum_{i=1}^{r} n_i = N$, *o lo que es lo mismo,* $\sum_{i=1}^{r} \frac{n_i}{N} = 1$.

Suelen utilizarse representaciones gráficas de las frecuencias relativas a conjunto de datos $[x_1, x_2, ..., x_i, ...x_N]$ que forman la muestra. Según el caso se usan las siguientes:

1. *Distribución no agrupada.*
   Se toma cada dato x_i y su frecuencia n_i. Se representan los datos en pares (x_i, n_i) . Veámoslo:

   EJEMPLO 1.8 *Sea la muestra z de tamaño 21 dada por*

   $$z = [1,\ 1,\ 2,\ 2,\ 2,\ 2,\ 3,\ 3,\ 3,\ 3,\ 4,\ 4,\ 4,\ 4,\ 4,\ 5,\ 5,\ 5,\ 6,\ 6,\ 7]$$

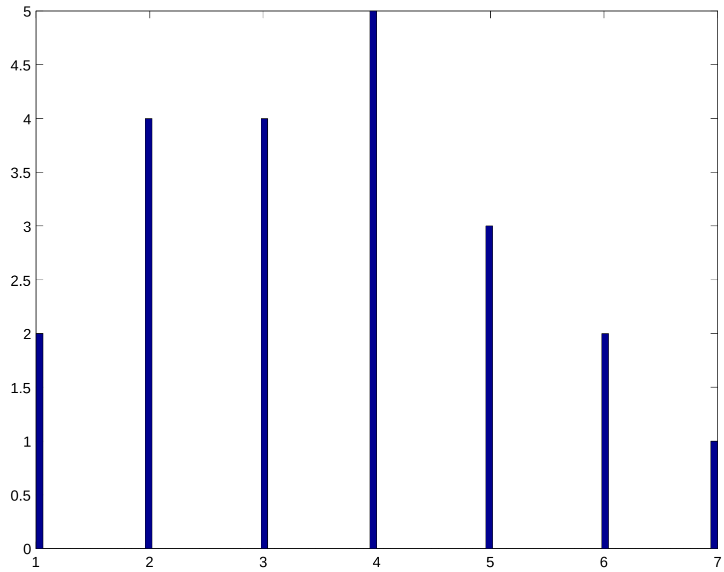
   

   Figura 1: Representación de los datos no agrupados de la muestra z
   La frecuencia de cada uno de los datos de esta muestra puede representarse como en la Figura 1.

2. *Distribución agrupada.*
   Se divide el intervalo $[\text{mín}(x_i), \text{máx}(x_i)]$ en una cantidad determinada de subintervalos (que a veces se llaman marcas o clases) y se cuentan las frecuencias de los datos en tales intervalos. La marca o clase que generalmente tomaremos será de la forma $(a, b]$ (con la excepción de la primera de todas, que la tomaremos del tipo $[\text{mín}(x_i), c]$).
   El número de subintervalos que suele tomarse es en torno a $2\sqrt{N}$. Nunca se agrupan los datos si se conoce la distribución original no agrupada, se pierde información.

   Para la misma muestra del ejemplo anterior, agrupando los datos en intervalos centrados en x_i y longitud 1, se tiene la siguiente representación

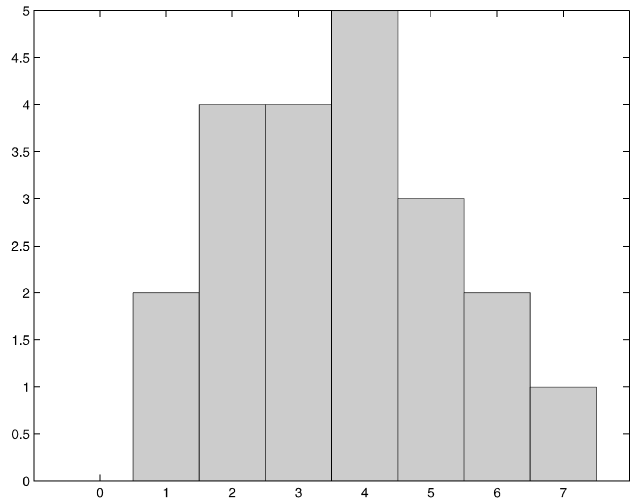


Figura 1.1: Figura 2: Distribución agrupada en torno a cada punto de la muestra z

de las frecuencias:

Obsérvese qué diferente es el aspecto del gráfico asociado a la muestra anterior si en ella llevamos a cabo una agrupación distinta:

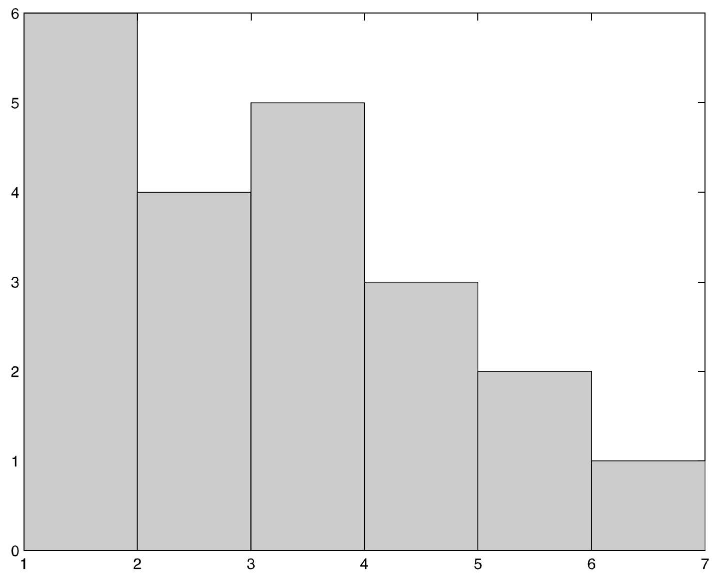


Figura 3

1.3.1. Reducción de datos.

Según se ha señalado, presentaremos distribuciones de frecuencias con las que analizar fenómenos. Esta presentación muestra en general dificultades de cálculo. Se precisa entonces el manipular estos datos mediante expresiones sintéticas de los mismos, expresiones sencillas que permitan trabajar sobre el conjunto de los sucesos, reduciendo o resumiendo éstos en una serie de medidas que llamaremos de *tendencia central.* La tendencia indica la propensión del fenómeno estudiado a una cierta conducta o característica. No obstante las medidas de tendencia central no son suficientes para caracterizar al fenómeno, veamos un ejemplo en el que quede reflejado esto:

Sea una comunidad cuya renta per cápita ha sido evaluada llegándose a la conclusión de que la media es de 600 dólares. De acuerdo con esta cifra entenderemos que todos y cada uno de los ciudadanos poseerán una renta así, despreciando la posibilidad de que un porcentaje significativo de ciudadanos ganen realmente más, o menos, que esa cifra.
Si en concreto suponemos que esa comunidad consta de dos individuos, uno ganando 200 y el otro 1.000 dólares, entonces está clara que la media -en este caso la medida de tendencia elegida- es de 600 dólares. Ahora bien, si consideramos otra comunidad de 2 individuos pero ganando 580 y 620, la tendencia sería la misma que en el caso anterior, pero la manera en que representa ésta al fenómeno es bastante mejor en el segundo caso que en el primero. -

Por tanto, la tendencia no representa en general de manera precisa al fenómeno, presenta una síntesis que debemos interpretar correctamente. Para esta interpretación introducimos lo que se llama *medidas de dispersión.* Con estas medidas pretendemos controlar cómo de alejados están los datos individuales de la tendencia central. Esta dispersión será la encargada de validar los resultados que se obtienen a partir de las medidas de tendencia.

1.3.2. Medidas de tendencia central o de posición

Estas son las más importantes:

Media Aritmética

Sean los datos

Datos	Frec.	Frec. relativas
x_1	n_1	n_1/N
$\vdots$	$\vdots$	$\vdots$
x_r	n_r	$n_r/N,$

relativos a una variable aleatoria X. Se define la **media aritmética** como

$$\begin{aligned} a &= \sum\nolimits_{i=1}^{r} x_i \frac{n_i}{N} = \\ &= x_1 \frac{n_1}{N} + x_2 \frac{n_2}{N} + \cdots + x_r \frac{n_r}{N}. \end{aligned}$$

Cuando se trate de una distribución agrupada el cálculo se realizará como sigue:

$$a = \sum_{i=1}^{r} y_i \frac{n_i}{N}$$

siendo y_i el punto medio de cada intervalo y n_i el número de datos que caen en dicho intervalo.

Un cálculo análogo tendrá lugar para otras medidas de posición cuya muestra viene dada en forma agrupada.

Media Geométrica

Se define como

$$\begin{aligned} g &= \left[x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}\right]^{\frac{1}{N}} = \\ &= \left[\Pi_{i=1}^{r} x_i^{n_i}\right]^{\frac{1}{N}}. \end{aligned}$$

Observamos que entonces la variable estadística $\log g$ es la media aritmética de $\log X$, i.e.

$$\log g = \frac{1}{N} \sum_{i=1}^{r} n_i \log x_i.$$

Media Armónica

Se define como

$$H = \frac{1}{\frac{1}{x_1}\frac{n_1}{N} + \frac{1}{x_2}\frac{n_2}{N} + \cdots + \frac{1}{x_r}\frac{n_r}{N}} = \frac{1}{\sum_{i=1}^{r} \frac{1}{x_i}\frac{n_i}{N}}.$$

Mediana

La mediana o mediana muestral es el valor de la muestra que deja a ambos lados el mismo número de observaciones realizadas. Si r es impar entonces es el centro geomético de la muestra. Cuando r sea par se toma el punto medio entre $x_{\frac{r}{2}}$ y $x_{\frac{r}{2}+1}$.

Valor Mediana

La valor mediana es aquel valor de la variable estadística que divide en 2 partes iguales a la distribución de frecuencias, es decir, es un valor tal que la suma de las frecuencias que quedan a su izquierda coincide con la suma de las frecuencias de las quedan a su derecha.

Si nos referiremos al caso de una distribución agrupada y consideramos

Datos	Frec.	Frec. acumulada
$x_1 - x_2$	n_1	N_1
$\vdots$	$\vdots$	$\vdots$
$x_{m-1} - x_m$	n_{m-1}	N_{m-1}

donde $N_1 = n_1$, $N_2 = n_1 + n_2, \cdots, N_j = \sum_{k=1}^{j} n_k, \cdots, N_{m-1} = \sum_{k=1}^{m-1} n_k = N$ son las frecuencias acumuladas. Distingamos dos casos:
i) Si el valor $N/2$ coincide con uno de los datos recogidos en la tabla, por ejemplo N_j, entonces la mediana es el extremo superior que corresponde a dicho intervalo, i.e., es x_{j+1}.
ii) si no ocurre i) hemos de proceder como sigue: habrá un valor en la columna de la derecha inferior a $N/2$ y otro inmediatamente posterior que será mayor. Supongamos que N_j es el valor de la frecuencia acumulada mayor que $N/2$, siendo N_{j-1} menor que $N/2$; para esta situación no hay forma de determinar la mediana y hablaremos en todo caso del llamado *intervalo mediana,* intervalo que según los datos dados es $[x_j, x_{j+1})$.

Se suele construir un método de interpolación con el que asignar un valor mediana para este último caso. Sin embargo esto conlleva hacer la suposición de que los datos se distribuyen de cierta forma dentro de cada intervalo, lo cual, en general, no es verdad.

NOTA 1.2 *El valor mediana admite una definición alternativa cuando la variable aleatoria es discreta (como es el caso de todo análisis descriptivo): se considera valor mediana aquel valor que deja tanto a su izquierda como* *derecha una probabilidad mayor o igual que* 0,5*. Con esto siempre se asegura su existencia, y cuando hay más de uno usualmente se opta por dar el intervalo de puntos (intervalo mediana) que verifican la definición. En ocasiones también se da el punto medio de dicho intervalo; además el valor mediana es un valor en el que se alcanza el mínimo de la función*

$$f(M) = \sum_{i=1}^{n} |x_i - M|$$

esto es, si M_e es tal valor, entonces

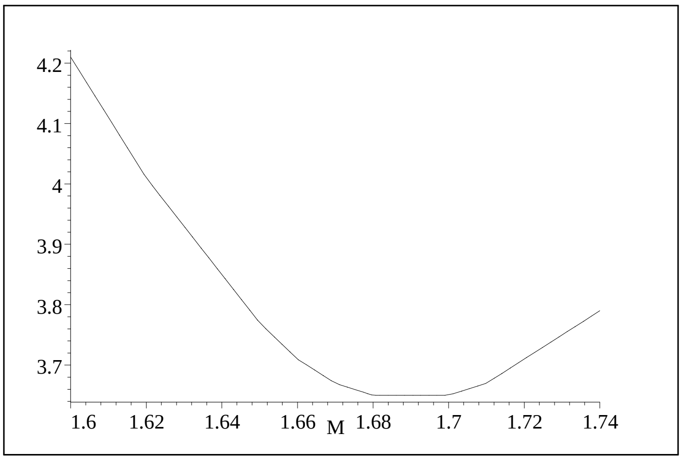
$$\min \left\{ \sum_{i=1}^{n} |x_i - M| \right\} = \sum_{i=1}^{n} |x_i - M_e|$$

Obsérvese que una definición aternativa de la media aritmética es aquel valor que minimiza a la función

$$g(M) = \left| \sum_{i=1}^{n} (x_i - M) \right|$$

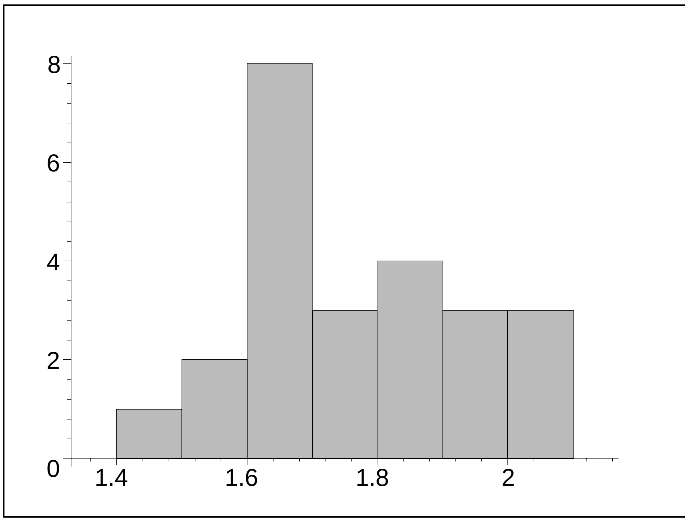
EJEMPLO 1.9 *Sea la muestra relativa a las alturas de los individuos de una población:*

$$z = \left[\begin{array}{c} 1{,}60, 1{,}75, 1{,}60, 1{,}41, 1{,}93, 2{,}00, 1{,}71, 1{,}68, 1{,}60, 1{,}67, 1{,}85, 1{,}83, \\ 1{,}57, 1{,}54, 1{,}62, 1{,}93, 1{,}84, 2{,}01, 1{,}70, 1{,}85, 1{,}05, 1{,}66, 1{,}90, 1{,}65 \end{array} \right].$$



Representación de $\sum_{j=1}^{24} |z_j - M|$

Los valores centrales son 1,70 *y* 1,71*, con lo que la mediana es* 1,705*. El* *intervalo mediana es* $M_e = [1.68, 1.705]$ *(véase el gráfico de arriba). Una representación de los datos agrupados en intervalos de longitud* 0,1 *queda como sigue a continuación:*



Representación agrupada de los datos en clases de longitud 0,1

Ejemplo 1.10 *Si por ejemplo el caso es el de la siguiente situación*

Datos	*Frec.*	*Frec. acumulada*
$2-4$	4	4
$4-6$	10	14
$6-8$	40	54
$8-10$	20	74
$10-12$	1	75

,

entonces $c = 2$, $N = 75$, $\frac{N}{2} = 37'5$, $N_j = 54$, $N_{j-1} = 14$, $x_j = 6$ *y* $n_j = 40$. *Por tanto el intervalo mediana es* $[6, 8)$.

Moda

Es el valor de la muestra que más se repite.

Valor modal

Se trata de aquél valor de la variable estadística tal que su frecuencia es superior a la del siguiente y a la del anterior. Podemos tener varios valores así, de entre éstos, el de mayor frecuencia se denomina *moda absoluta o valor*

modal, es el valor o intervalo que más se repite en la muestra.
Sea el caso de una distribución agrupada y sean $n_j > n_{j-1}$ y n_{j+1}, donde $x_k \leftrightarrow n_k$; entonces se de define el *intervalo modal* como $[x_j, x_{j+1}]$.

Debemos decir que la moda no tiene mucho sentido para distribuciones de tipo continuo (ya precisaremos esto más adelante). Es frecuente que el valor modal no sea un valor central. No obstante puede ser útil a la hora de predecir falta homogeneidad para la muestra (éste sería el caso si se existen varias modas).

EJEMPLO 1.11 *Sea el ejemplo cuya tabla es*

Datos	*Frec.*	*Frec. acumulada*
$2-4$	4	4
$4-6$	10	14
$6-8$	40	54
$8-10$	20	74
$10-12$	1	75

El intervalo modal es $[6,8)$.

EJEMPLO 1.12 *Para el caso del Ejemplo 1.8 se tiene después de un sencillo cálculo que* $a = 3.619$, *el intervalo mediana es*$[3,4)$ *y* $M_o = 4$.

Otras medidas de posición.

DEFINICIÓN 1.3 *Un* cuantil *de orden* k *es aquel valor de la variable para el cual los valores de la variable que (ordenados en sentido creciente) se conservan no superiores a él tienen una frecuencia* k. *Se denotará por* x_k, *y así* x_k *está caracterizado por tener una frecuencia acumulada* $N_k = k$.

Entre los cuantiles más importantes distinguimos:
Los *centiles o percentiles* son los valores de la variable estadística que agrupan, entre cada dos consecutivos al $1\,\%$ del conjunto de la distribución. De esta forma llamaremos *el primer percentil*, P_1, al valor de x tal que no mayores que él existe el $1\,\%$ de la distribución. El *segundo percentil*, P_2, el valor de x tal que no mayores que él existe el $2\,\%$ de la distribución. De manera análoga podemos definir hasta P_{99}.

Deciles: Son aquellos que agrupan al $10\,\%$ de la distribución entre dos consecutivos, así el primer decil, D_1, está caracterizado por ser el valor de la variable tal que el 10 % no es superior a él. De forma análoga se definen los *cuartiles*, éstos agrupan al 25 % de la distribución entre dos consecutivos. La notación empleada para el primer, segundo, ... cuartil es Q_1, Q_2, ...

1.3.3. Momentos. Medidas de Dispersión.

Los momentos son expresiones que aparecen al tratar de caracterizar un fenómeno a través del análisis de la tendencia. También sirven para el estudio de la dispersión.

Momentos.

DEFINICIÓN 1.4 *Definimos* el momento de orden h con respecto al origen *de la variable* X *a*

$$a_h = \sum_{i=1}^{r} x_i^h \frac{n_i}{N}.$$

Definimos el momento de orden h de X con respecto a a *como*

$$m_h = \sum_{i=1}^{r} (x_i - a)^h \frac{n_i}{N},$$

donde $a = \sum_{i=1}^{r} x_i \frac{n_i}{N}$.

Veamos cuál es la relación entre los momentos centrados y no centrados. Para ello basta desarrollar el binomio $(x_i - a)^h$. Al hacerlo obtenemos

$$\begin{aligned} m_h &= \binom{h}{0} a_h - \binom{h}{1} a\ a_{h-1} + ... + \\ &\quad + (-1)^j \binom{h}{j} a^j\ a_{h-j} + ... + (-1)^h a^h \\ &= \sum_{j=1}^{r} (-1)^j \binom{h}{j} a^j\ a_{h-j}. \end{aligned}$$

NOTA 1.3 $m_1 = 0$. *¿Por qué?.*

Medidas de dispersión.

Sirven para criticar la representatividad de los medidas de posición o tendencia. Definimos para ello distintas medidas de dispersión:

Por comparación entre valores de X. i) *Recorrido* o rango de variación: Es la diferencia entre el mayor y el menor valor de la variable estadística, así si $M = \max x_i$ y $m = \min x_i$, el recorrido es

$$R = M - m.$$

ii) *Intervalos intercuantílicos.* Éstos se definen como la diferencia entre dos cuantiles preestablecidos. Distinguimos a) Intervalos intercuantílico: Es la diferencia entre el cuantil tercero y el primero, i.e.

$$I_Q = Q_3 - Q_1.$$

b) *Intervalos interpercentiles.* Se obtienen mediante la diferencia entre dos percentiles simétricos, por ejemplo el intervalo 10-90 % es

$$I_p = P_{90} - P_{10},$$

el 7-93 es

$$I_p = P_{93} - P_7.$$

Por comparación entre X y medidas centrales. Si M_c es una medida de tendencia central, la medida de dispersión de una variable X con respecto a esta medida es

$$D = X - M_c.$$

La desviación absoluta con respecto a un valor x de la variable X se obtiene calculando

$$|D| = |x - M_c|\,;$$

suelen emplearse $M_c = M_e$ (valor mediana) o $M_c = a$ (media aritmética).

Desviación mediana. Definimos una nueva medida de dispersión, *la desviación mediana*: es el valor mediana de las desviaciones absolutas de la variable con respecto a M_e, es decir, el valor mediana de la variable aleatoria

$$Z = |X - M_e|\,.$$

La *desviación absoluta media* de X, $d\,(X)$, se define como la media aritmética de las desviaciones absolutas respecto de la medida de tendencia central elegida M_c, es decir

$$d(X) = \sum_{i=1}^{r} \frac{|x_i - M_c|\, n_i}{N}.$$

Se comprueba que esta función en la variable M_c alcanza su mínimo cuando $M_c = M_e$.

Dispersión respecto a la media. Al tomar como medida de tendencia central a la media aritmética, son de especial interés los siguientes criterios de dispersión:

1. *Desviación absoluta media respecto a la media*,
$$d_a(X) = \sum_{i=1}^{r} \frac{|x_i - a|\, n_i}{N}.$$

2. *Varianza muestral*: es la media aritmética del cuadrado de las desviaciones de X con respecto a la media aritmética,
$$v = \sum_{i=1}^{r} \frac{(x_i - a)^2\, n_i}{N}.$$
Observamos que
$$v = m_2$$
y
$$v = a_2 - a^2.$$
PROPOSICIÓN 1.1
(a) $0 \leq v \leq +\infty$.
(b) Si definimos la desviación cuadrática media *con respecto a un valor* W *como*
$$D^2(t) = \sum_{i=1}^{r} \frac{(x_i - t)^2\, n_i}{N},$$
entonces esta función en la variable W *alcanza el mínimo en* $t = a$, *y en tal caso* $D^2(t) = v$.

3. Se define *la desviación típica o estándar* de la variable X como
$$\sigma = +\sqrt{v}.$$

4. *Coeficiente de variación de Pearson*. Se define como
$$d_X = \frac{\sigma}{a}.$$
Dadas dos distribuciones diremos que posee menor dispersión aquélla cuyo coeficiente de Pearson sea menor. Este tipo de medición es bastante útil para comparar dos variables aleatorias cuando la escala de medición de una y otra difieren de manera ostensible.

Asimetría y Curtosis

Para analizar de manera completa el fenómeno, aparte de determinar su tendencia y dispersión, conviene hablar del coeficiente de asimetría y del grado de curtosis.

El *coeficiente de asimetría* indica el grado de inclinación de la distribución con respecto al media; analizando el histograma de frecuencias podremos hacernos una idea de este coeficiente de asimetría, el cual se define como

$$g_1 = \frac{m_3}{\sigma^3}.$$

Si éste es cero se dice que es simétrica, y hablaremos de asimetría positiva (negativa) cuando dicho coeficiente sea positivo (negativo respectivamente).

El *coeficiente de curtosis o de apuntamiento* medirá la agrupación de los valores de la variable en torno a la media comparados con los valores de una normal; cuanto mayor sea la diferencia mayor será el apuntamiento. Se define como

$$g_2 = \frac{m_4}{\sigma^4} - 3.$$

EJEMPLO 1.13 *Realizar un estudio descriptivo de la muestra 0.9501, 0.2311, 0.6068, 0.4860, 0.8913, 0.7621, 0.4565, 0.0185, 0.8214, 0.4447, 0.6154, 0.7919, 0.9218, 0.7382, 0.1763, 0.4057, 0.9355, 0.9169, 0.4103, 0.8936, 0.0579, 0.3529, 0.8132, 0.0099, 0.1389, 0.2028, 0.1987, 0.6038, 0.2722, 0.1988, 0.0153, 0.7468, 0.4451.*
Agrupemos los datos en torno a los puntos $0.j$ *con* $j = 0, ..., 9$ *y el dato* 1. *La representación de las frecuencias es como sigue:*

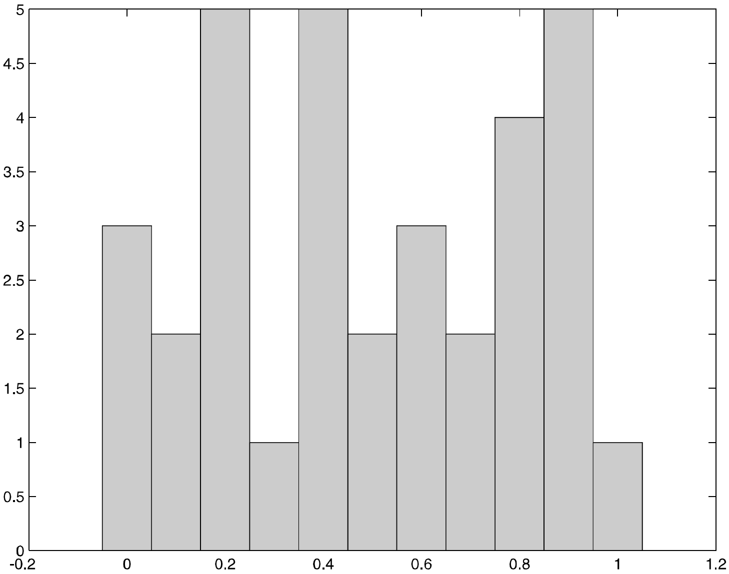


Figura 6

La media es 0,50092*, la mediana* 0,4565*, la varianza* 9.8087×10^{-2} *y la desviación típica* $\sigma = 0{,}31319$*. Nótese que con los datos agrupados se dan varias modas. Cerca de 0.2, 0.4 y 0.9. haciendo un promedio de éstas se obtiene 0.5, valor próximo a la media aritmética obtenida.*

Calculemos los coeficientes de asimetría y curtosis: $g_1 = \frac{m_3}{\sigma^3}$, $g_2 = \frac{m_4}{\sigma^4} - 3$. *Tenemos que* $m_3 = -0{,}001675$, $m_4 = 0{,}0149$ *luego*

$$g_1 = \frac{-0{,}001675}{(0{,}31319)^3} = -5.4524 \times 10^{-2}$$

y

$$g_2 = \frac{0{,}0149}{(0{,}31319)^4} - 3 = -1.4513.$$

El primero de los coeficientes informa sobre la agrupación de los datos en torno a la media. En este caso hay más frecuencia de datos a la izquierda de la media (se da asimetría negativa). El segundo informa que el grado de concentración en torno a la media es pequeño.

EJEMPLO 1.14 *Analizar descriptivamente la siguiente serie de datos sobre* *el tamaño de los huevos de pavo africano:*

Anchura	*número de huevos*
[13.75,14.75]	*2*
(14.75,15.75]	*14*
(15.75,16.75]	*54*
(16.75,17.75]	*95*
(17.75,18.75]	*22*
(18.75,19.75]	*19*
(19.75,20.75]	*5*

Se trata de una distribución agrupada. Calculemos las medidas de tendencia más interesantes: la media es

$$a = \sum_{i=1}^{7} x_i \frac{n_i}{211} = 17{,}189,$$

donde $x_1 = 14{,}25$ *y* $x_i = 14{,}25 + i$ *para* $i = 0, ..., 6$. *La desviación típica es* $\sigma = 1{,}1317$ *y (16.75,17.75) es tanto el intervalo mediana como el modal.*

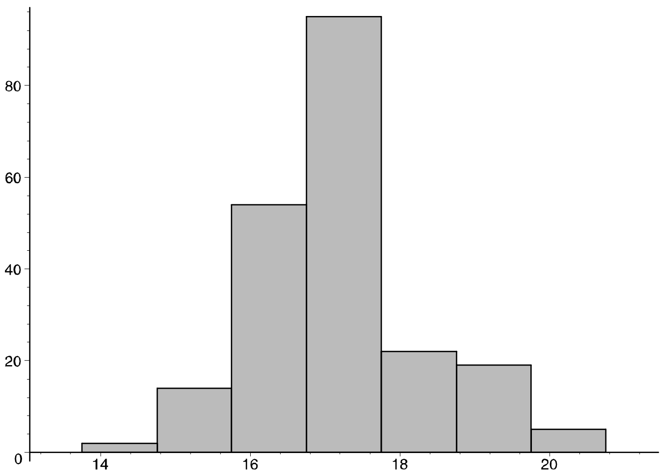


Figura 7

Los coeficientes de asimetría y curtosis son

$$g_1 = \frac{0{,}6587}{(1{,}1317)^3} = 0{,}454\,46$$

y

$$g_2 = \frac{5{,}6829}{(1{,}1317)^4} - 3 = 0{,}464\,53$$

respectivamente. Hay cierta asimetría a la derecha de la media y en cuanto a la concentración en torno a ésta podríamos decir que no es demasiado grande.

1.4. Teoría de la Probabilidad

1.4.1. Nociones generales

Conocemos ya fenómenos determinístas y aleatorios. Nos planteamos medir la incertidumbre o aleatoriedad mostrada por éstos. El elemento a usar para medir será la probabilidad ayudada por métodos matemáticos. Estos métodos darán rigor y firmeza así como una forma relativamente sencilla y sistemática de calcular probabilidades. La probabilidad tiene sus cimientos en una serie de axiomas que en todo momento han de ser respetadosy que son enunciados más adelante.
El primer paso en el estudio de un fenómeno es considerar todos las posibles concreciones como sucesos elementales o conjuntos que, todos ellos, junto con una serie de operaciones constituirán cierta estructura sumamente útil. Así tenemos:

1. **Conjuntos**,

$$\text{FENÓMENO} \rightarrow \Omega = \{\text{Espacio de sucesos}\} = \{\omega \in \Omega\}$$

y

$$\mathcal{A} = \{ \text{ sucesos } S \text{ susceptibles de ser medidos}\} = \{S \in \mathcal{A}\},$$

2. **Operaciones**. Al tratar de calcular probabilidades sería deseable que fuesen ciertas las siguientes implicaciones:

    a) Contraposición: $S \in \mathcal{A} \Rightarrow \overline{S} \in \mathcal{A}$,

    b) Unión: $S_1, S_2 \in \mathcal{A} \Rightarrow S_1 \cup S_2 \in \mathcal{A}$,

    c) Intersección: $S_1, S_2 \in \mathcal{A} \Rightarrow S_1 \cap S_2 \in \mathcal{A}$,

    d) Condicionamiento: $\exists S \in \mathcal{A}$ cuando $T \in \mathcal{A}$, i.e. $S|_T \in \mathcal{A}$.

    Nótese que $\mathcal{A}$ será el conjunto de todos los suceso que podemos medir.

Formalicemos esto:

DEFINICIÓN 1.5 *Diremos que el conjunto de sucesos $\mathcal{A}$ es un álgebra sobre Ω si*

1. $\Omega \in \mathcal{A}$,

2. $S \in \mathcal{A} \Rightarrow \overline{S} \in \mathcal{A}$,

3. $\cup_{finita} S_j \in \mathcal{A}$ *si* $S_j \in \mathcal{A}$ *para todo* j.

4. *Si además* $\cup_{j=1}^{\infty} S_j \in \mathcal{A}$, *con* $S_j \in \mathcal{A}$ *para todo* $j = 1, 2, 3, ...$, *entonces diremos que* $\mathcal{A}$ *es una* σ*-álgebra.*

EJEMPLO 1.15 *Supongamos que lanzamos un dado dos veces, entonces*

$$\Omega = \{\omega_1 = (1,1), \omega_2 = (1,2), ..., \omega_6 = (1,6), ..., \omega_{36} = (6,6)\}.$$

Parece natural elegir $\mathcal{A} = \mathcal{P}(\Omega)$ *(conjunto partes de* Ω*). Por ejemplo* $\{(1,1), (1,3)\} \in \mathcal{A}$, *y medir la probabilidad de que salga (1,1) ó (1,3) es calcular la probabilidad de* $\omega_1 \cup \omega_3$*).*

Al considerar espacios muestrales finitos se suele tomar como σ-álgebra a $\mathcal{A} = \mathcal{P}(\Omega)$. La elección dependerá en general de cómo sea la naturaleza fenómeno que vayamos a realizar.

Finalmente introducimos la herramienta encargada de medir aleatoriedad:

3. **Funciones de medida de probabilidad** (o simplemente probabilidad): Sea φ una función sobre una σ-álgebra $\mathcal{A}$.

$$\varphi : \mathcal{A} \to \mathbf{R}$$

tal que

$$\omega \in \mathcal{A} \to \varphi(\omega) \in \mathbf{R}.$$

φ se dice que es *completamente aditiva* sobre $\mathcal{A}$ si

$$\varphi(\emptyset) = 0 \quad \text{y} \quad \varphi\left(\cup_{j=1}^{\infty}\right) = \sum_{j=1}^{\infty} \varphi(S_j),$$

donde $S_j \in \mathcal{A}$ y son disjuntos.
Así, si φ es completamente aditiva sobre $\mathcal{A}$ y $\varphi(S) \geq 0$ $\forall S \in \mathcal{A}$ entonces decimos que φ es una *medida*. Si además $\varphi(\Omega) = 1$ se dirá entonces que φ es una *medida de probabilidad.*

El esquema es más o menos

$$\left.\begin{array}{l}\text{fenómeno}\\ \text{posibles concreciones}\end{array}\right\} \rightarrow \ \Omega,\ \text{medir incertidumbre} \rightarrow$$
$$\rightarrow \ \Omega \text{ y } \mathcal{A},\ (\Omega, \mathcal{A});$$

y si para medir la incertidumbre disponemos de una medida de probabilidad apropiada P, entonces constituiremos lo que se llama un **espacio de probabilidad**

$$(\Omega, \mathcal{A}, P)\,.$$

1.4.2. Axiomas de probabilidad.

Dado Ω espacio de muestras y $\mathcal{A}$ (σ-álgebra) definimos una (medida de) probabilidad como una función P sobre $\mathcal{A}$ que verifica los axiomas
(i) $P(\Omega) = 1$,
(ii) $P(S) \geq 0 \ \forall S \in \mathcal{A}$,
(iii) $P\left(\cup_{j=1}^{\infty} S_j\right) = \sum_{j=1}^{\infty} P(S_j)$, S_j disjuntos.

1.4.3. Operaciones con probabilidades.

Como consecuencia de los postulados enunciados anteriormente tenemos

TEOREMA 1.2 *Supongamos que P es una medida de probabilidad sobre $\mathcal{A}$ (σ-álgebra), entonces*

1. $P(\emptyset) = 0$,
2. $P(\cup_{finita} S_j) = \sum P(S_j)$, $S_j \in \mathcal{A}$ *disjuntos,*
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$,
4. $P(A \cup B) \leq P(A) + P(B)$,
5. $P(A) + P(B) \geq P(A \cap B)$,
6. $P(B) \leq P(A)$ *si* $B \subseteq A$,
7. $P(S) \in [0, 1]$, *y*
8. $P(A) + P(\overline{A}) = 1$.

La demostración de este teorema se deja como ejercicio.

1.4.4. Probabilidad condicionada.

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad asociado a cierto fenómeno y sea $B \in \mathcal{A}$.
Tomemos $A \in \mathcal{A}$ y calculamos lo que sería su probabilidad a priori, es decir, $P(A)$, supuesto que no disponemos de ninguna información extra sobre el fenómeno. Supongamos que el evento B ha ocurrido y nos preguntamos de nuevo por la probabilidad de A. En este caso la probabilidad podría cambiar con respecto a la probabilidad a priori, ahora se trata de evaluar $P(A|_B)$, la probabilidad a posteiori de A, la probabilidad de A sabiendo que B ha ocurrido. Esta probabilidad se denomina probabilidad de A condicionada por B, y se define, supuesto que $P(B) > 0$, como

$$P(A|_B) = \frac{P(A \cap B)}{P(B)}.$$

Ejemplo 1.16 *Se lanza un dado, sabemos que el resultado obtenido es par. Calcular las probabilidades de: (a) que el resultado sea 3, y (b) que sea 4.*

(a) $P(3|_{par}) = 0$ *ya que* $P(3 \cap par) = P(\emptyset) = 0$.

(b) $P(4|_{par}) = \dfrac{P(4 \cap par)}{P(par)} = \dfrac{1/6}{1/2} = \dfrac{1}{3}$.

También podríamos haber procedido restringiendo el espacio de muestras, en lugar de tomar $\Omega = \{1, ..., 6\}$, *trabajar con el espacio muestral* Ω_{par} $=\{2, 4, 6\}$. *En ese caso se ve claramente que*

$$P(4|_{par}) = P(4) = \frac{casos\ fav.}{casos\ pos.} = \frac{1}{3}.$$

Ejemplo 1.17 *Se tiran un par de dados no cargados y se sabe que los dos números que salen no son iguales. Evaluar la probabilidad de que la suma sea (i) 7 (ii) 4 (iii) 12.*

Denotemos

$$\begin{aligned} A &= \textit{los dos resultados distintos,} \\ B &= \textit{la suma da 7,} \\ C &= \textit{la suma da 4,} \\ D &= \textit{la suma da 12.} \end{aligned}$$

Las probabilidades a priori son

$$
\begin{aligned}
P(A) &= c.f./c.p. = 30/36 = 5/6. \\
P(B) &= 1/6. \\
P(C) &= 3/36 = 1/12. \\
P(D) &= 1/36.
\end{aligned}
$$

Entonces

$$
\begin{aligned}
P\left(B|_A\right) &= \frac{1/6}{5/6} = 1/5, \\
P\left(C|_A\right) &= \frac{2/36}{5/6} = 2/5, \\
P\left(D|_A\right) &= \frac{0}{5/6} = 0.
\end{aligned}
$$

DEFINICIÓN 1.6 *Se dice que dos eventos A y B de $\mathcal{A}$ son independientes si*

$$P\left(A|_B\right) = P(A) \quad o \quad P\left(B|_A\right) = P(B).$$

Independencia (estocástica) entre dos sucesos significa que el grado de incertidumbre que supone la realización de uno, no influye en la del otro. Nótese que los sucesos A y B son independientes sii $P(A \cap B) = P(A)P(B)$.

EJEMPLO 1.18 *Sea la tabla de probabilidad referente a una determinada población:*

	Desvío de columna	*no desvío de columna*
conductor	*0.5*	*0.2*
no conductor	*0.1*	*0.2*

Al seleccionar un individuo analizamos la posibilidad de que ocurran los eventos $A =$ tiene desvío de columna y $B =$ es conductor. ¿Son A y B sucesos independientes?.

$$
\begin{aligned}
P\left(A|_B\right) &= \frac{P(A \cap B)}{P(B)} = \frac{P(\textit{conduce y desvío})}{P(\textit{conduce})} = \frac{0{,}5}{0{,}7} = 0{,}714. \\
P(B) &= P(B \cap \left(A \cup \overline{A}\right)) = P(B \cap A) + P(B \cap \overline{A}) = 0{,}7. \\
P(A) &= 0{,}6. \\
P\left(B|_A\right) &= \frac{0{,}5}{0{,}6} = 0{,}83.
\end{aligned}
$$

Vemos pues que no hay indepenencia.

Ejercicio. Supongamos que A y B son conjuntos de $\mathcal{A}$ tales que $A \cap B = \emptyset$. ¿Son A y B independientes?

Definición 1.7 *Decimos que* $\{A_1, A_2, ..., A_n\}$ *son independientes si*

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)P(A_2)...P(A_n).$$

La axiomática definida permite establecer cierta similitud entre la geometría y la probabilidad. Esta similitud queda patente para los casos en los que un suceso S pueda dividirse en sucesos más simples S_j. En tal situación el cálculo de la probabilidad de S es más sencillo y se efectúa realizando el cálculo sobre cada uno de los S_j. Esto, no sólo establece un paralelismo entre la probabilidad y la geometría, sino que además reduce considerablemente los cálculos de la probabilidad. La discusión de un modo riguroso sería como sigue:
Sea $S \in \mathcal{A}$ y supongamos que $S \subset \cup_{j=1}^{N} S_j$, con los S_j disjuntos, esto es $S \subset \{S_1, ..., S_N\}$ y $S_i \cap S_j = \emptyset$ para $i \neq j$. Entonces observamos que

$$\begin{aligned} S &= S \cap \Omega = S \cap \left\{\cup_{j=1}^{N} S_j\right\} \\ &= \cup_{j=1}^{N} \left(S_j \cap S\right), \end{aligned}$$

luego

$$\begin{aligned} P(S) &= P(\cup_{j=1}^{N} \left(S_j \cap S\right)) = \\ &= \sum_{j=1}^{N} P\left(S_j \cap S\right) = \\ &= \sum_{j=1}^{N} P\left(S|_{S_j}\right) P(S_j), \end{aligned}$$

es decir:

$$P(S) = \sum_{j=1}^{N} P\left(S|_{S_j}\right) P(S_j),$$

igualdad esta última conocida con el nombre *fórmula de la probabilidad total.*

Una aplicación de la fórmula de la probabilidad total es el siguiente resultado:

TEOREMA 1.3 (BAYES) *Sean* S *y* T *de* $\mathcal{A}$ *tal que* $P(S), P(T) > 0$ *y* $S \subset \cup_{j=1}^{N} S_j$*, con los* S_j *disjuntos, entonces*

$$P\left(T|_S\right) = \frac{P(T \cap S)}{P(S)} = \frac{P\left(S|_T\right)P(T)}{\sum_{j=1}^{N} P\left(S|_{S_j}\right)P(S_j)}.$$

EJEMPLO 1.19 *Elegido un individuo de una determinada especie arborea, de un determinado bosque, y observado en laboratorio, se observó que estaba afectado por el hongo saza. La probabilidad de que en la población de la que se eligió el individuo, uno de ellos esté afectado por saza es de 0.01. Se sabe además que la probabilidad de que el aparato de laboratorio encargado de examinar la infección, detecte que un individuo está afectado estándolo es 0.97, y no estándolo es 0.001. ¿Qué podemos decir acerca de la probabilidad de acierto del laboratorio para el individuo que hemos elegido?*

La probabilidad de infección es 0.01, i.e. $P(\mathit{infec.}) = 0.01$ *y el resto de la información se escribe*

$$P\left(DI|\,\mathit{infec.}\right) = 0.97,\ P\left(DI|\,\overline{\mathit{infec.}}\right) = 0.001.$$

Para analizar el éxito del laboratorio en el caso dado hemos de calcular

$$\begin{aligned} P\left(\mathit{infec.}|\,DI\right) &= \frac{P\left(DI|\,\mathit{infec.}\right)P\left(\mathit{infec.}\right)}{P\left(DI\right)} \\ &= \frac{P\left(DI|\,\mathit{infec.}\right)P\left(\mathit{infec.}\right)}{P\left(DI|\,\mathit{infec.}\right)P\left(\mathit{infec.}\right) + P\left(DI|\,\overline{\mathit{infec.}}\right)P\left(\overline{\mathit{infec.}}\right)} \\ &= \frac{(0.97)\,(0.01)}{(0.97)\,(0.01) + (0.001)\,(1-0.01)} \approx 0.907 \end{aligned}$$

lo que se traduce en alta fiabiliada para la situación descrita.

EJEMPLO 1.20 *En un sistema de comunicación se transmite* 0 *ó* 1*. Debido al ruido del medio sucede que en ocasiones la señal no es recibida de manera correcta. Se dan los sucesos*

$$V_i = \text{se transmite } i,\ W_i = \text{se recibe } i,$$

para $i = 0$ *ó* $i = 1$*. Se supone que la probabilidades de que* V_0 *y* V_1 *se transmitan con éxito son* 0.7 *y* 0.8 *repectivamente. Se sabe además que la probabilidad de transmitir* 0 *es del* $60\,\%$*. Se pide*

1. Determinar la probabilidad de recibir 0 cuando se ha emitido 1.

2. Calcular la probabilidad de que una transmisión sea errónea.

En virutd de los datos lo que sabemos es lo siguiente:

$$P(V_0) = 0.6$$

y

$$\begin{aligned} P(W_0 | V_0) &= 0.7, \\ P(W_1 | V_1) &= 0.8. \end{aligned}$$

En la primera pregunta nos están pidiendo $P(V_0 | W_1)$ y para su cálculo empleamos el Toerema de Bayes.

$$\begin{aligned} P(V_0 | W_1) &= \frac{P(V_0 \cap W_1)}{P(W_1)} \\ &= \frac{P(W_1 | V_0) P(V_0)}{P(W_1 \cap V_0) + P(W_1 \cap V_1)} \\ &= \frac{(1 - P(W_0 | V_0)) P(V_0)}{P(W_1 | V_0) P(V_0) + P(W_1 | V_1) P(V_1)} \\ &= \frac{(1 - 0.7)(0.6)}{(1 - 0.7)(0.6) + (0.8)(0.4)} = 0{,}36 \end{aligned}$$

El cálculo de la probabilidad de que una transmisión sea errónea se computa con ayuda de la Fórmula de la Probabilidad Total:

$$\begin{aligned} P(\textit{errónea}) &= P((V_0 \cap W_1) \cup (V_1 \cap W_0)) \\ &= P(V_0 \cap W_1) + P(V_1 \cap W_0) \\ &= P(W_1 | V_0) P(V_0) + P(W_0 | V_1) P(V_1) \\ &= (1 - 0.7)(0.6) + (1 - 0.8)(0.4) = 0{,}26. \end{aligned}$$

EJEMPLO 1.21 *En tres grandes poblaciones de vegetales que crecen en un bosque A, B y C la proporción de individuos infectados por un determinado hongo es del 30, 60 y 10 % respectivamente. Se toma al azar una de las 3 poblaciones (las tres son igual de probables), y de ella elegimos a 10 individuos al azar, resultando que 2 de ellos están infectados. ¿A qué población es más probable que pertenezcan?*

Del enunciado del problema se desprende lo siguiente: si Inf denota estar infectado y A, B y C haber elegido a dichas poblaciones, entonces

$$P(Inf|\, A) = 3/10,\ \ P(Inf|\, B) = 6/10,\ \ P(Inf|\, C) = 1/10,$$

y

$$P(A) = P(B) = P(C) = \frac{1}{3}.$$

Se sabe que S el suceso consistente en la realización de 10 obsevaciones con resultado de 2 infecciones ha tenido lugar. La pregunta es saber cuál de las siguientes cantidades es mayor:

$$P(A|\, S),\ \ P(B|\, S),\ \ P(C|\, S).$$

Calculemos: se sabe por el Teorema de Bayes que $P(A|\, S)$ puede escribirse como

$$P(A|\, S) = \frac{P(A \cap S)}{P(S)} = \frac{P(S|\, A)P(A)}{P(S|\, A)P(A) + P(S|\, B)P(B) + P(S|\, C)P(C)},$$

ahora bien, $P(S|\, A)$ es la probabilidad de que hayan tenido lugar 2 infecciones de 10 individuos sabiendo que la probabilidad de infección en A es $3/10$. Esta probabilidad no es otra cosa que el valor de la masa de un binomial de parámetros $p = 3/10$ y $n = 10$; análogamente para las otras probabilidades condicionadas. Por tanto

$$P(S|\, A) = \binom{10}{2} \left(\frac{3}{10}\right)^2 \left(\frac{7}{10}\right)^8$$

y así

$$P(A|\, S) = \frac{\binom{10}{2} \left(\frac{3}{10}\right)^2 \left(\frac{7}{10}\right)^8}{\binom{10}{2} \left(\frac{3}{10}\right)^2 \left(\frac{7}{10}\right)^8 + \binom{10}{2} \left(\frac{6}{10}\right)^2 \left(\frac{4}{10}\right)^8 + \binom{10}{2} \left(\frac{1}{10}\right)^2 \left(\frac{9}{10}\right)^8}.$$

Puesto que el denominador que aparece en el cálculo de $P(A|\, S)$, $P(B|\, S)$ y $P(C|\, S)$ es el mismo, basta con comparar

$$\left(\frac{3}{10}\right)^2 \left(\frac{7}{10}\right)^8, \left(\frac{6}{10}\right)^2 \left(\frac{4}{10}\right)^8, \left(\frac{1}{10}\right)^2 \left(\frac{9}{10}\right)^8.$$

El valor más grande es el correspondiente a $P(A|\, S)$ [1]*. La población con más probabilidad de haber sido elegida es A.*

1.4.5. Espacios muestrales discretos y continuos.

Queremos ver como puede ser Ω en lo que se refiere al número de elementos que lo constituyen. Básicamente las distinciones a hacer son:

- Ω discreto, si el número de elementos que lo componen es finito o hay tantos como números naturales. En este último caso se dirá que es numerable.
- Ω continuo, si el número de elementos es infinito no numerable, es decir, hay tantos elementos como números reales.

EJEMPLO 1.22 *Lanzar un dado da lugar al espacio muestral* $\Omega = \{1, ..., 6\}$ *conjunto discreto.*

EJEMPLO 1.23 *La altura de una población: en este caso supone considerar un espacio muestral* Ω *de tipo continuo, por ejemplo* $\Omega = \{x : x \in (0, 3]\}$, *con* x *expresado en metros.*

Vemos como esencialmente hay 3 tipos de espacios muestrales (discreto finito, discreto numerable y continuo). La pregunta interesante que hemos de plantear en este momento es, ¿cómo debemos elegir la medida de probabilidad apropiada a nuestro fenómeno?

(i) si Ω finito$\rightarrow$ frecuencia relativa $\rightarrow$ espacio de probabilidad (ver ejemplos).

(ii) si Ω infinito numerable $\rightarrow$ ¿P?; si $\Omega = \{\omega_1, ..., \omega_n, ...\}$ (espacio muestral infinito numerable) la cuestión es ser capaces de asignar a cada evento ω_i una probabilidad p_i, es decir

$$P(\omega_i) = p_i \in [0, 1],$$

y como queremos que $P(\Omega) = 1$ entonces los p_i han de ser tales que $\sum_{i=1}^{\infty} p_i = 1$. Para estos caso la σ-álgebra que se suele eligir es $\mathcal{A} = \mathcal{P}(\Omega)$.

[1] Para cerciorarse de ello basta con esbozar la gráfica de la función $\psi(p) = p^2(1-p)^8$ y notar que el máximo se da para $p = 0{,}2$. Los valores en 0.6 y 0.1 están más lejos de este máximo que 0,3.

Veremos diferentes ejemplos más adelante.

(iii) si Ω infinito no numerable $\rightarrow$¿ P ?; una posibilidad para definir la probabilidad consiste en definir frecuencias relativas en un sentido más general; veamos un ejemplo bastante ilustrativo. Supongamos que Ω es un subconjunto de $\mathbb{R}^2$, digamos que es un cuadrado de área M. Si $\mathcal{A} = \mathcal{P}(\Omega)$, definimos

$$P(S) = \frac{área\,(S)}{M},\ \forall S \in \mathcal{A}.$$

Se puede demostrar que $(\Omega, \mathcal{P}(\Omega), P)$ es un espacio de probabilidad.

En el próximo capítulo veremos como para $\Omega = \mathbb{R}$ el conjunto de conjuntos

$$\mathcal{B} = \{(a, b] : a, b \in \mathbb{R} \cup \{\pm\infty\}\}$$

es una σ-álgebra y además dispondremos de una probabilidad capaz de actuar en cada uno de los conjuntos $(a, b]$, se definirá para ello

$$P((a, b]) = F(b) - F(a)$$

siendo F una función de distribución de probabilidad.

1.5. Problemas propuestos

1. La distribución por pesos de 70 alumnos de Ingeniería es la siguiente:

kgs.	frecuencia
$[50, 60]$	8
$[60, 70]$	15
$[70, 80]$	21
$[80, 90]$	18
$[90, 100]$	8

   Realizar un análisis descriptivo.

2. Se ha realizado la determinación del contenido de Potasio en la bioquímica automatizada de la sangre de 25 pacientes embarazadas. Los resultados han sido 2.7, 3.3,4.1, 3.2, 3.1, 3.3, 3.4, 2.7, 2.8, 2.7, 3.2, 2.3, 4.2, 3.5, 3.6, 3.7, 3.2, 3.3, 2.8, 3.5, .3,8 3.1, 3.2, 3.6, 2.4. Llevar a cabo un análisis descriptivo de esta muestra.

3. Demostrar las siguientes relaciones entre los momentos:

   a) $m_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\left(\alpha_1\right)^3.$

   b) $m_4 = \alpha_4 - 4\alpha_3\alpha_1 + 6\left(\alpha_1\right)^2\alpha_2 - 3\left(\alpha_1\right)^4.$

4. Sean A, B y C sucesos. Se pide

   a) Calcular $P(A \cup B \cup C)$

   b) Demostrar que

   $$P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$$

   supuesto que $P(C) > 0$.

   c) Demostrar las desigualdades

   $$P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B).$$

5. Demostrar que si A y B son sucesos independientes entonces también lo son A y $\overline{B}$. ¿$\overline{A}$ y $\overline{B}$?. ¿Es verdad que $P(A \cup B) = 1 - P(\overline{A})P(\overline{B})$?

6. Cuando la concentración en sangre de una determinada sustancia en un individuo sobrepasa cierto valor se dice que el individuo pertenece a cierto grupo de riesgo. Supongamos que el 5 % de los habitantes de una población pertenecen al grupo de riesgo. Determinar la probabilidad de que de tres individuos elegidos al azar, uno esté en el grupo de riesgo.

7. Determinar la probabilidad del suceso consistente en extraer una bola blanca de una urna que contiene 5 bolas de dos colores, blanco y negro. Se supone que las distintas composiciones de la urna son igualmente probables.

8. Un dispositivo de control consta de dos aparatos. La probabilidad de que falle el k-ésimo es $1 - \alpha_k$. Estimar la probabilidad de que ambos funcionen bien si:

   a) si el funcionamiento de dichos aparatos es independiente,

   b) si no se sabe si hay o no independencia entre el funcionamiento de los dos aparatos.

9. En una universidad el 4 % de los chicos y el 1 % de las chicas tienen una altura superior a 180 cm. El 60 % de los alumnos son chicas. Se toma un alumno al azar y se comprueba que mide más de 180 cm. Hallar la probabilidad de que tal alumno sea chico.

10. El volumen de ventas en un concesionario de coches es de 500 unidades (al año) para el deportivo, 1000 para el familiar y 2000 para el utilitario. Se sabe que el porcentaje de coches defectuosos es de un 2 % para el primero, un 1 % para el segundo y de un 3 % para el tercero. Se pide calcular la probabilidad del siguiente suceso:

    a) Elegido uno al azar, que éste no sea defectuoso.

    b) Habiendo elegido un automóvil defectuoso, que dicho automóvil pertnezca al grupo de los familiares.

11. Se disponen dos urnas U_1 y U_2. La urna U_1 contiene el 70 % de bolas blancas y el 30 % de negras, y en la urna U_2 hay un 30 % de blancas y un 70 % de negras. Se selecciona una urna al azar (se supone que ambas tienen la misma probabilidad de ser elegidas) y se toman 10 bolas una tras otra con reemplazamiento. El resultado fue el suceso $S = bnbbbbnbbb$, siend b bola blanca y n bola negra. ¿ Cuál es la probabilidad de que el suceso S proceda de la urna U_1?

12. Una empresa proyecta introducir un nuevo artículo y conjetura que con una probabilidad del 50 % será adquirido por un segmento del 5 % de sus clientes habituales, con una probabilidad de 25 % por el 10 % de sus clientes habituales, con un 20 % de probabilidad por un 15 % de sus clientes y con probabilidad del 5 % por el 20 %. Para asegurarse de ello recogen información preguntando a 20 de esos clientes, siendo 3 de éstos los que estarían dispuestos a comprar el nuevo artículo. ¿Cuál sería la probabilidad de que el 15 % comprara el mencionado artículo?

13. Sean 10 urnas. Nueve de ellas contienen 2 bolas blancas y dos negras y una contiene cinco blancas y una negra. Se elige una urna de forma aleatoria y sea extrae una bola resultando ser una blanca. ¿Cuál es la probabilidad de que la urna elegida sea la que posee 4 blancas y una negra?

14. La probabilidad de que tres jugadores de dardos hagan diana es $4/5$, $3/4$ y $2/3$. Supongamos que tiran los tres y se producen dos aciertos sobre la diana. ¿Cuál es la probabilidad de que haya fallado el tercero?

15. Sea $\Omega = (0, +\infty)$. Demostar que si definimos P sobre todos los intervalos $(a, b) \subset \Omega$ con ayuda de la fórmula

$$P((a, b)) = \int_a^b e^{-x} dx,$$

entonces P cumple todos los axiomas de la probabilidad.