Miguel Ángel Gómez Villegas
Catedrático de Estadística e Investigación Operativa
Departamento de Estadística e Investigación Operativa
Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid

INFERENCIA ESTADÍSTICA

DIAZ DE SANTOS

Ediciones Díaz de Santos

Internet: http://www.diazdesantos.es
E-Mail: ediciones@diazdesantos.es

ISBN: 84-0000-000-X
Depósito legal: M. 49.000-2004

Diseño de cubierta:
Fotocomposición: Fer, S. A.
Impresión: Edigrafos, S. A.
Impreso en España

PRÓLOGO

Este libro contiene el material necesario para un curso de Inferencia Estadística de un año de duración. Aunque todavía en algunos planes de estudio, tanto de licenciaturas como de ingenierías, existe un curso sobre Cálculo de Probabilidades y Estadística, cada vez es más frecuente el abordar el estudio de la Inferencia Estadística en un curso propio separado del Cálculo de Probabilidades.

En esta línea este texto está pensado para estudiantes de Estadística, en Matemáticas, Biología, Economía, Informática, Ingenierías, Medicina, Psicología, Sociología, y en definitiva para los que estén interesados en modelizar la incertidumbre mediante la probabilidad para extraer conclusiones de esta modelización. El libro recoge lo que en mi opinión es la naturaleza de la Estadística moderna: un equilibrio entre el análisis de datos -con especial atención a las representaciones gráficas-, y los modelos matemáticos, ambas cosas realizadas con el creciente concurso del ordenador.

Los prerrequisitos incluyen nociones de probabilidad básica, además de los resultados usuales de cálculo, álgebra matricial y análisis real.

Los teoremas y demostraciones se incluyen cuando su enunciado y su demostración son necesarios para una correcta comprensión del texto y siempre que se mantengan dentro del nivel de complejidad y abstracción que se ha querido dar al libro. Al final de cada capítulo se incluye una colección de aproximadamente 25 problemas, cuya solución detallada ha sido realizada en colaboración con la profesora Paloma Maín Yaque.

En cuanto al contenido del texto, los dos primeros capítulos constituyen la introducción a la "Teoría de la Inferencia Estadística", estudian los estadísticos muestrales, los diagramas de tallos y hojas y de cajas, las distribuciones asociadas al muestreo con la distribución normal, los estadísticos ordenados y los principios de reducción de datos. Se recomienda especialmente la atenta lectura de los principios de verosimilitud, suficiencia y condicionalidad; el significado de los mismos, junto al teorema de Birnbaum, son fundamentales para una correcta interpretación de los distintos procedimientos que integran la Inferencia Estadística. También se ha incluido el concepto de intercambiabilidad como una manera más razonable que la independencia para recoger la información suministrada por la muestra. Este concepto es de capital importancia pues justifica la aproximación bayesiana a la inferencia.

Los Capítulos 3, 4 y 5, se dedican a la estimación por punto, la estimación por regiones y el contraste de hipótesis, que constituyen el núcleo central de la Inferencia Estadística. Un aspecto importante en la manera de desarrollar estos capítulos, es la diferenciación entre los métodos de obtención de las distintas técnicas y los métodos de evaluación de las mismas, que son tratados a continuación. Este tratamiento permite que los tres métodos usados para eliminar la incertidumbre presenten una estructura uniforme.

Los tres Capítulos finales contienen aplicaciones específicas, así el Capítulo 6 es una introducción a la Teoría de la Decisión Estadística, el Capítulo 7 aborda el Análisis de la Varianza y la Regresión Lineal y el Capítulo 8 la Inferencia no Paramétrica.

Tres aspectos hacen singular, en mi opinión, a este libro. El primero es que los métodos bayesianos y los clásicos son desarrollados a la vez sin ningún tipo de menoscabo de unos frente a otros, aunque el autor es partidario de la aproximación bayesiana. Así a continuación de los estimadores por punto clásicos se estudian los estimadores por punto bayesianos. Después de los intervalos de confianza los intervalos creíbles y siguiendo a los contrastes de hipótesis mediante la teoría de Neyman y Pearson, los contrastes bayesianos y los factores Bayes. Además, tras cada método, según corresponda, se estudia su comportamiento frente a la distribución en el muestreo, recogido en el criterio del error cuadrático medio, y frente a la distribución inicial, recogido en la pérdida final esperada. El estudio conjunto de estas dos aproximaciones hace que se haya incluido en el Capítulo 7 el análisis bayesiano de la varianza, que no está en otros textos de nivel parecido, ya que o sólo utilizan la aproximación clásica o sólo la aproximación bayesiana. Así mismo, en el Capítulo 6, sobre la "Teoría de la Decisión Estadística", se incluye la paradoja de Stein, presentada de la manera más elemental posible, para poner de manifiesto las anomalías que presentan los métodos clásicos cuando la dimensión del parámetro es mayor o igual que tres.

El segundo aspecto es que al final de cada Capítulo se incluye una aproximación histórica de la evolución de las ideas que han sido desarrolladas en el mismo, incluyendo unas pinceladas biográficas de quienes más han contribuido al desarrollo de las mismas.

El tercer aspecto es la publicación de la solución detallada de los ejercicios de final de capítulo como un eficaz complemento de las ideas desarrolladas. Con frecuencia, el alumno echa de menos una colección de ejercicios resueltos que le permitan aclarar y comprobar si ha asimilado correctamente los conocimientos expuestos.

No se discute en el texto ningún paquete de programas, pero en la resolución de los problemas se ha utilizado S, SPSS y Statgraphics; pudiendo utilizarse otros paquetes también. El libro ha sido empleado para impartir durante todo un curso 4 horas de clase y una de laboratorio por semana. La utilización de un paquete estadístico permite tratar problemas más reales y ayuda a aclarar aspectos que de otra manera son más difíciles de comprender.

Quiero expresar mi agradecimiento a todas las personas a las que debo mis conocimientos en el campo de la Estadística, desde aquella primera versión traducida al castellano del libro de Cramér en el que por primera vez leí sobre estas ideas y aquellas primeras clases que, impartidas por el profesor Sixto Ríos, me pusieron en contacto con la Estadística en la Licenciatura de Matemáticas de la Universidad Complutense. A todos los compañeros, tanto los anteriores como los actuales, del Departamento de Estadística e Investigación Operativa de la Complutense de Madrid, muchas gracias, y en particular a los profesores Jesús Artalejo, Antonio Gómez, Eusebio Gómez, Paloma Maín, Javier Montero, Luis Sanz y Teófilo Valdés, que me han sugerido cambios y aclaraciones. Finalmente quiero dar las gracias a Pedro Ortega por su trabajo de verter el manuscrito inicial al Latex.

Madrid, enero de 2005.

Índice general

Capítulo 1

INTRODUCCIÓN A LA ESTADÍSTICA MATEMÁTICA Y ESTADÍSTICOS MUESTRALES

1.1. INTRODUCCIÓN

Dentro del Cálculo de Probabilidades se pretende calcular la probabilidad de ciertos sucesos a partir de probabilidades calculadas de otros sucesos más simples. Así, si se supone que un determinado experimento aleatorio se puede modelizar mediante una variable aleatoria de Bernoulli de parámetro θ conocido, la probabilidad de que el suceso éxito se presente en dos repeticiones consecutivas del experimento es θ^2.

En esta línea, se dice que el Cálculo de Probabilidades utiliza un razonamiento deductivo. Así, tras modelizar la incertidumbre mediante la probabilidad, es posible utilizar dicho razonamiento para llegar a obtener probabilidades de sucesos más complicados a partir de las probabilidades de sucesos más simples.

En cuanto a la Inferencia Estadística, Cramér (1945) dice:

> *El objetivo fundamental de la Teoría Estadística consiste en investigar la posibilidad de extraer de los datos estadísticos inferencias válidas.*

Se admite que los datos estadísticos deben de poseer una característica que tiene un significado preciso, que es la aleatoriedad. Fisher (1958) dice que la ciencia estadística es esencialmente una rama de la matemática aplicada:

> *La Estadística puede ser mirada como el estudio de poblaciones, como el estudio de la variación y como el estudio de métodos de reducción de datos.*

y añade Fisher que el significado original de la palabra Estadística sugiere que ésta surge con el objetivo de estudiar colectivos humanos viviendo en unión política, aunque más tarde ha ido ampliando su campo de aplicación primero a la Biología y, después, a la Economía y a todas las restantes ciencias experimentales.

Lindley (1965) ve la Estadística como el análisis de cómo los grados de opinión son alterados por los datos.

Kendall y Stuart (1969) afirman:

> *La Estadística es la rama del método científico que se enfrenta con datos obtenidos contando o midiendo elementos de poblaciones asociadas a fenómenos naturales.*

Breiman (1973) dice:

> *La diferencia entre un modelo probabilístico y un modelo estadístico es que en el primero la probabilidad está totalmente especificada, mientras que en el segundo la distribución de probabilidad es desconocida. Todo lo que se conoce son algunas características de esta distribución de probabilidad.*

Cox (1973) establece:

> *La Estadística pretende elegir un elemento de entre una familia $\mathcal{P}$ de modelos de probabilidad, que en la mayor parte de los casos es conocida excepto por un número limitado de parámetros desconocidos, para responder a ¿son los datos consistentes con la familia $\mathcal{P}$? Suponiendo que los datos han sido obtenidos de uno de los modelos de $\mathcal{P}$, ¿qué puede concluirse sobre los valores de los parámetros desconocidos o acerca de los valores de observaciones extraídas del mismo modelo?*

Por último, según Lehmann (1978):

> *La Estadística trata de la recogida de datos así como de su análisis e interpretación.*

Basados en las definiciones anteriores se puede decir que la Inferencia Estadística pretende, dados unos datos sujetos a incertidumbre, obtener el conocimiento de los parámetros, en un sentido amplio, del modelo de Cálculo de Probabilidades que aceptablemente se ajuste a dichos datos. En este sentido, la Inferencia Estadística utiliza un razonamiento de tipo inductivo.

Un proceso de fabricación, a lo largo del tiempo, puede producir a veces algún elemento defectuoso. La probabilidad θ de que un elemento cualquiera sea defectuoso es desconocida. De toda la producción se extraen al azar n elementos y con estos datos se trata de conjeturar el valor de θ.

El ejemplo explicita los elementos con los que se va a trabajar: a saber, una población, una muestra: los n elementos observados; y una distribución de probabilidad. La población se puede considerar que es una variable aleatoria con distribución de Bernoulli de parámetro θ, donde θ es la probabilidad de producir un elemento defectuoso, los n elementos son la repetición con independencia, de n variables aleatorias de Bernoulli y lo que se pretende, basados en la distibución de probabilidad de la muestra $\theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i}$, es conocer el valor de θ. En la misma línea se puede considerar un ejemplo un poco más complicado. Se suponen almacenadas gran cantidad de semillas en condiciones constantes de temperatura y humedad. Las semillas se van deteriorando a lo largo del tiempo, de tal forma que existirá una función $r(t)$ que indique la proporción de semillas que en el instante t continúan siendo válidas. Si en sucesivos instantes de tiempo $t_1, ..., t_s$ se toman cada vez n semillas y se observa que continúan válidas $r_1, ..., r_s$; se trataría de determinar la función $r(t)$ al menos en los instantes t_i con $i = 1, ..., s$. En este caso la distribución de probabilidad de la muestra viene dada por $\prod_{i=1}^{s} \binom{n}{r_i}(r(t_i))^{r_i}(1-r(t_i))^{n-r_i}$. En esta situación no existe el parámetro θ o más bien el parámetro, en sentido amplio, es la función $r(t)$. Por último, un experimentador realiza n observaciones $x_1, ..., x_n$ con vistas a determinar una cantidad θ, que él desconoce, pero supone que estas observaciones pueden ponerse en la forma $x_i = \theta + \varepsilon_i \quad i = 1, ..., n$. Donde ε_i son los errores contenidos en cada observación que se suponen independientes y distribuidos mediante una distribución $N(0, \sigma)$ normal de media cero y de desviación típica σ desconocida. La distribución de probabilidad de la muestra viene dada a partir del producto de n variables $N(0, \sigma)$. En este caso se desea determinar el valor del parámetro desconocido θ. Todas estas situaciones pueden resumirse en que se tiene una v.a. X y sobre la σ-álgebra de Borel de $\mathbb{R}$, está definida una familia de distribuciones de probabilidad $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$. A lo largo de lo que sigue se irá concretando el problema en distintas direcciones.

Así, se hablará de la *estimación por punto* cuando sólo se esté interesado en la elección de un determinado valor para θ, se hablará de la *estimación por intervalo* cuando se quiera obener un subconjunto de Θ al que pertenezca la distribución con la que se está trabajando, y se hablará del *contraste* *de hipótesis* cuando se pretenda únicamente elegir entre dos subfamilias de distribuciones de probabilidad $\mathcal{P}$ y $\mathcal{P}'$. A continuación se estudiará la *Teoría de la Decisión* como un modelo más general que unifica las aproximaciones anteriores, para acabar trasladando estos problemas al caso no paramétrico.

1.2. ESTADÍSTICOS MUESTRALES

El modelo matemático de la Inferencia Estadística viene dado por una variable aleatoria X a la que se llama *universo* o *población* cuya función de distribución F viene determinada por algún elemento de una familia $\mathcal{P}$ de distribuciones de probabilidad. En los ejemplos dados en la sección 1.1 F era la distribución de *Bernoulli* de parámetro θ, o la normal de media θ y varianza desconocida, o una $Binomial(n, \pi(t))$ con $\pi(t)$ variable. El problema es determinar el elemento de la familia de distribuciones de probabilidad para X. Para esto de dispone de cierta información que se concre-, mediante la observación de una muestra.

Definición 1.2.1 (muestra aleatoria simple)
Dada una población X se llama muestra aleatoria simple de tamaño n a la repetición de $X_1, ..., X_n$ variables aleatorias independientes con distribución igual a la de X. Es decir, la función de distribución de la muestra $(x_1, ..., x_n)$ es

$$F(x_1, ..., x_n) = \prod_{i=1}^{n} F(x_i)$$

donde $F(x)$ es la función de distribución de la población X que como se ha dicho sigue dada por un elemento de $\mathcal{P}$.
■

Cuando la población X tiene un número finito de elementos, $x_1, ..., x_N$ se dispone, básicamente de dos procedimientos para obtener una muestra de tamaño n, $n \leq N$. En el primero se puede suponer que cada elemento tiene la misma probabilidad de ser escogido y por tanto ésta es $\frac{1}{N}$, es decir sería equivalente a suponer que los N elementos están metidos en una urna y que se extrae sucesivamente con reposición cada elemento de la urna con probabilidad $\frac{1}{N}$, por tanto

$$P\{X_2 = x_2 | X_1 = x_1\} = \frac{1}{N}$$

siendo x_1 y x_2 cualesquiera elementos pertenecientes a la colección $\{x_1, ..., x_N\}$.

En el símil de la urna esto equivale a ir sacando valores *con reeemplazamiento.* Observamos por tanto que se cumple

$$P\{X_1 = x_1, ..., X_n = x_n\} = \frac{1}{N^n}$$

y que esta manera de muestrear está de acuerdo con la Definición 1.2.1. No obstante, utilizar este procedimiento con el fin de conocer la población $\{x_1, ..., x_N\}$ no parece lo más razonable, ya que se podría observar varias veces el mismo elemento. El otro procedimiento posible consiste en suponer que los elementos se extraen *sin reemplazamiento*. Así el primer elemento es escogido con probabilidad $\frac{1}{N}$, y para la obtención del segundo todos los elementos tienen la misma probabilidad pero el elemento a escoger ha de ser diferente del primero; en el símil de la urna sería equivalente a dejar fuera el primer elemento observado y extraer cualquiera de los restantes con equiprobabilidad, con lo que

$$P\{X_2 = x_2 | X_1 = x_1\} = \begin{cases} 0 & \text{si} \quad x_2 = x_1 \\ \frac{1}{N-1} & \text{si} \quad x_2 \neq x_1 \end{cases}$$

En el caso de obtención de la muestra de una población finita *sin reemplazamiento* no se satisfacen las condiciones de la Definición 1.2.1. Es curioso entonces que el procedimiento de obtener una m.a.s. de una población sea el no intuitivo cuando la población es finita. También es importante el observar que aunque las n variables en el muestreo de una población finita sin reemplazamiento no son independientes, sin embargo las variables están idénticamente distribuidas con distribución marginal $P\{X_i = x\} = \frac{1}{N}$ para cada $x \in \{x_1, ..., x_N\}$ y cada $i = 1, ..., n$. En efecto, basta observar que $P\{X_1 = x\} = \frac{1}{N}$ y que por el teorema de la probabilidad total

$$P\{X_2 = x\} = \sum_{i=1}^{N} P\{X_1 = x_i\} P\{X_2 = x | X_1 = x_i\}$$

pero $P\{X_2 = x | X_1 = x_i\} = 0$ si $x = x_i$ y en otro caso $P\{X_2 = x | X_1 = x_i\} = \frac{1}{(N-1)}$ ya que todos los elementos disponibles se escogen con la misma probabilidad, luego

$$P\{X_2 = x\} = \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{1}{N} \frac{1}{N-1} = \frac{1}{N}.$$

El mismo razonamiento puede aplicarse de forma sucesiva hasta alcanzar n. Si N es suficientemente grande con respecto a n el muestreo *sin reemplazamiento* puede aproximarse por el muestreo *con reemplazamiento* en el sentido de que las probabilidades calculadas tanto en un tipo de muestreo como en el otro, son prácticamente iguales.

Ejemplo 1.2.1
Dada una población finita de $N = 500$ elementos $\{1, 2, ..., 500\}$ de la que se extraen muestras de tamaño $n = 5$, la probabilidad de que todos los valores

muestrales sean mayores que 100 cuando se extraen *con reemplazamiento* es

$$\begin{aligned} P\{X_1 > 100, ..., X_5 > 100\} &= \prod_{i=1}^{5} P\{X_i > 100\} \\ &= \left(\frac{400}{500}\right)^s = 0'32768. \end{aligned}$$

Cuando se extraen *sin reemplazamiento* la probabilidad es

$$P\{X_1 > 100, ..., X_5 > 100\} = \frac{\binom{400}{5}\binom{100}{0}}{\binom{500}{5}} = 0'326035,$$

lo que parece una aproximación aceptable.
■

Existen otros tipos de muestreo, que en determinadas situaciones prácticas pueden parecer más interesantes, de hecho existen libros enteros dedicados al estudio de los distintos tipos de muestreo. A lo largo del resto del libro, se usará la Definición 1.2.1 como definición de muestra aleatoria correspondiente a una población. Frecuentemente no se trabaja con toda la muestra, tanto por resumir la información como para mimetizar ciertos elementos de la población, por lo que se introduce la siguiente definición.

Definición 1.2.2 (estadístico muestral)
Dada $(X_1, ..., X_n)$ una muestra aleatoria simple de la población X, a toda $f : \mathbb{R}^n \to \mathbb{R}^k$ función conocida medible, se la llamará estadístico muestral.

Ejemplo 1.2.2
$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ y $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ son estadísticos muestrales. El primero es, el conocido como la media muestral y al segundo se le denomina la cuasivarianza muestral.
■

A continuación se introducen estadísticos muestrales particulares.

Definición 1.2.3 (momentos centrales)
Se llama *momento muestral respecto al origen* de orden k a

$$a_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \qquad k > 0.$$

El momento de orden uno, es la media muestral, ya citado en el Ejemplo 1.2.2, que se acostumbra a notar por $\overline{X}$. Se llama *momento muestral respecto al centro* de orden k a

$$b_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k \qquad k > 0.$$

El momento respecto al centro más utilizado es b_2, la varianza muestral.
■

Es fácil ver que

$$b_k = a_k - \binom{k}{1}\overline{x}a_{k-1} + ... + (-1)^k\overline{x}^k.$$

Por consiguiente los momentos muestrales respecto al centro, pueden calcularse en función de los momentos muestrales respecto al origen. En particular la varianza muestral puede ponerse como

$$b_2 = a_2 - \overline{X}^2.$$

Intuitivamente los momentos respecto al origen son medidas de centralización, indican alrededor de qué valores se concentran las observaciones, mientras que los momentos respecto al centro son medidas de dispersión, valores altos indican una mayor dispersión de las observaciones.

Proposición 1.2.1
Dada una m.a.s. de tamaño n de una población X con momento de segundo orden finito, se tiene que $E[\overline{X}] = \mu$ y $V[\overline{X}] = \frac{\sigma^2}{n}$; siendo $\mu = E[X]$ y $\sigma^2 = V[X]$.

Demostración:

$$E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

$$\begin{aligned}
V[\overline{X}] &= E[(\overline{X}-\mu)^2] = E\left[\left(\tfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\right)^2\right] \\
&= \tfrac{1}{n^2}E\left[\sum_{i=1}^{n}(X_i-\mu)^2 + 2\sum_{i=1}^{n}\sum_{\substack{j=1\\ j>i}}^{n}(X_i-\mu)(X_j-\mu)\right] \\
&= \tfrac{1}{n^2}\left[\sum_{i=1}^{n}V[X_i] + 2\sum_{i=1}^{n}\sum_{j>i}E[X_i-\mu]E[X_j-\mu]\right] = \tfrac{\sigma^2}{n}
\end{aligned}$$

ya que el segundo sumando es nulo. La hipótesis $E[X^2] < +\infty$ es necesaria para asegurar la existencia de μ y σ^2.

Proposición 1.2.2
Dada una m.a.s. de tamaño n de una población con momento poblacional de cuarto orden finito, $\alpha_4 = E[X^4] < +\infty$, se tiene que

$$\begin{aligned}
E[b_2] &= \tfrac{n-1}{n}\sigma^2 \\
V[b_2] &= \tfrac{\beta_4-\beta_2^2}{n} - 2\tfrac{\beta_4-2\beta_2^2}{n^2} + \tfrac{\beta_4+3\beta_2^2}{n^3}
\end{aligned} \tag{1.2.1}$$

donde β_k es el momento poblacional de orden k respecto al centro, $\beta_k = E[(X-\alpha_1)^k]$.

Demostración: Sólo se comprobará la primera expresión. Obsérvese que

$$E[b_2] = \frac{1}{n}\sum_{i=1}^{n}\left[E[X_i^2] + E[\overline{X}^2] - 2E[X_i\overline{X}]\right]$$

siendo inmediato el comprobar que

$$E[X_i^2] = \sigma^2 + \mu^2 \qquad E[\overline{X}^2] = \frac{\sigma^2 + n\mu^2}{n} \qquad E[X_i\overline{X}] = \frac{\sigma^2 + n\mu^2}{n}.$$

Ahora es inmediato el comprobar que $E[S^2] = \sigma^2$, la esperanza de la cuasi-varianza muestral S^2 dada por

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{1.2.2}$$

es la varianza poblacional.

■

En la Sección 3.2 se verá la conveniencia de utilizar estadísticos muestrales que tengan de media el parámetro que se pretende estimar. Los momentos muestrales, también pueden introducirse en el caso de muestras de variables n dimensionales. Tan solo se va a abordar en este libro el caso de poblaciones bivariantes. Así, para una m.a.s. $(x_1, y_1), ..., (x_n, y_n)$ se pueden definir los *momentos marginales respecto al origen*

$$a_{k0} = \frac{1}{n}\sum_{i=1}^{n} x_i^k \qquad a_{0k} = \frac{1}{n}\sum_{i=1}^{n} y_i^k \qquad k > 0 \tag{1.2.3}$$

y los *momentos muestrales marginales respecto al centro*

$$b_{k0} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^k \qquad b_{0k} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^k \qquad k > 0 \tag{1.2.4}$$

pero esta vez se pueden definir además *momentos cruzados,* el más utilizado es

$$b_{11} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \tag{1.2.5}$$

la covarianza muestral y que también acostumbra a notarse como b_{xy}. Se define el *coeficiente de correlación muestral* mediante

$$r = \frac{b_{xy}}{b_x b_y} \tag{1.2.6}$$

donde $b_x^2 = b_{20}$ y $b_y^2 = b_{02}$.

1.3. ESTADÍSTICA DESCRIPTIVA

Antes de continuar con el estudio de los momentos, para inferir a partir de ellos características de la población en estudio, se debe tener en cuenta que también es posible no pasar de la etapa de obtención de la muestra y realizar únicamente lo que se llama *Estadística Descriptiva*. Se entiende ésta como el resumen de los datos, o en terminología más actual el *Análisis de Datos*. Estos procedimientos, muchos de los cuales incluyen representaciones gráficas, son útiles para revelar la estructura de los datos, que frecuentemente vendrán escritos en columnas dentro de una página o registrados en una salida de un ordenador.

Lo primero que se debe tener en cuenta es que, desde este punto de vista, se debe empezar diferenciando entre variable cualitativa, cuando las observaciones sólo pueden clasificarse en categorías no numéricas, como por ejemplo el grupo sanguíneo, tipo de estudios, etc., y variables cuantitativas que son las que toman valores numéricos, como ingresos, número de llamadas a una centralita, altura, etc. De las primeras variables sólo se pueden hacer representaciones gráficas, que permitan hacerse idea de la variable en estudio, existiendo una gran variedad de representaciones; las más frecuentes suelen ser el *diagrama de rectángulos* y el *diagrama de sectores.*

Ejemplo 1.3.1
Se clasifica una colección de 155 coches, 85 europeos, 26 americanos y 44 japoneses, de acuerdo con el lugar de fabricación. Estos datos pueden representarse gráficamente mediante el *diagrama de rectángulos* y el *diagrama de sectores* contenidos en la Figura 1.1

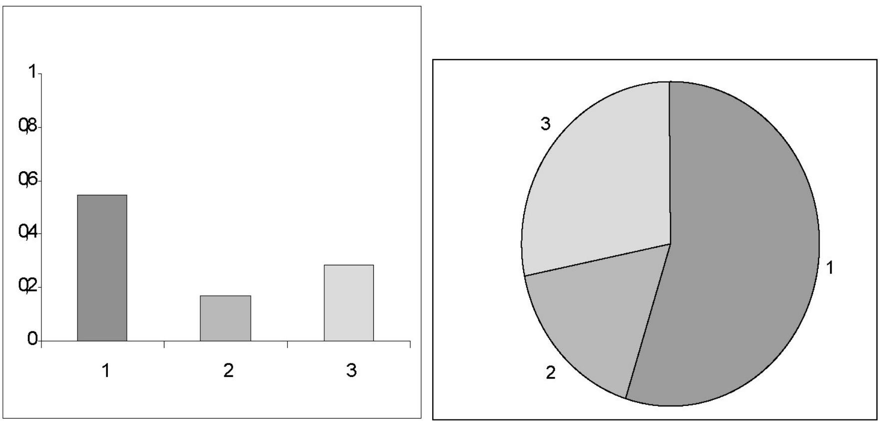

Figura 1.1. Diagrama de rectángulos y diagrama de sectores correspondientes al Ejemplo 1.3.1.

Las variables cuantitativas, por su parte, pueden clasificarse en discretas y continuas, las primeras sólo pueden tomar una cantidad finita o numerable de valores y las segundas pueden tomar cualquier valor en un intervalo acotado o no. Con independencia de si la variable cuantitativas es discreta o continua, y dependiendo básicamente del número de datos de que se disponga, a veces conviene agrupar los n elementos observados en distintas clases $C_1, C_2, ..., C_k$, siendo $x_1, x_2, ..., x_k$ los representantes de las clases y $n_1, n_2, ..., n_k$ las frecuencias absolutas, esto es, el número de observaciones dentro de cada clase. Se denota por $f_1, f_2, ..., f_k$ a las frecuencias relativas dentro de cada clase, donde $f_i = n_i/n$, se verifica que $\sum_{i=1}^{k} n_i = n$ y $\sum_{i=1}^{k} f_i = 1$. Alternativamente si el número de valores es grande se pueden construir las clases, para llegar a la *tabla de frecuencias*.

Otro problema que hay que resolver es el decidir el número de clases distintas a considerar, generalmente igual a $\sqrt{n}$.

También hay que tener presente que cuando los datos se observan pasando antes por una *tabla de frecuencias* se pierde algo de la información que se tiene cuando se consideran los datos directamente, a cambio se facilita el manejo y la interpretación de los mismos.

Al mismo tiempo se deben escoger los límites que definen los intervalos, de forma que las clases sean de la misma longitud, al menos la mayor parte de ellas, y de manera que cada observación pueda ponerse con precisión en una única clase.

Ejemplo 1.3.2
Los siguientes datos, tomados de McPherson (1990), recogen los niveles de colesterol en sangre medidos en miligramos de 135 personas:

145	145	146	158	158	163	168	168	168	168	172	174	175
175	175	175	175	175	178	180	180	180	181	181	187	187
192	194	194	195	196	196	198	198	198	198	201	201	201
204	204	204	205	205	205	205	205	206	206	206	207	208
209	211	211	211	212	212	212	214	214	214	216	217	217
217	217	218	218	218	218	221	221	221	221	223	224	225
225	227	277	227	227	228	230	235	235	235	236	237	238
238	239	241	242	242	243	243	245	245	245	247	248	252
253	253	253	254	254	254	256	256	257	257	261	262	266
267	267	268	268	268	268	273	274	278	283	284	285	289
292	294	296	300	345								

Estos datos se clasifican en 11 clases en la siguiente *tabla de frecuencias:*

Clases	Marcas de clase	frecuencias	frecuencias relativas
130 – 150	140	3	0'022
150 – 170	160	7	0'052
170 – 190	180	16	0'118
190 – 210	200	27	0'200
210 – 230	220	32	0'237
230 – 250	240	18	0'133
250 – 270	260	20	0'148
270 – 290	280	7	0'052
290 – 310	300	4	0'030
310 – 330	320	0	0'000
330 – 350	340	1	0'007

La tabla permite extraer consecuencias elementales de la colección de datos, en primer lugar el único dato contenido en la última clase, parece ser inusual y puede deberse quizás a una errata de transcripción, un tratamiento estadístico de estas observaciones puede verse en la Sección 8.4.4. En segundo lugar, más de la mitad de los valores tienden a concentrarse sobre las 5 clases más bajas, por lo que la distribución parece asimétrica.

■

Algunos autores reservan el nombre de *variables estadísticas* para las colecciones de datos cuando no se hace referencia a la distribución de probabilidad poblacional sobre la que se quieren inferir resultados, no considerándose por lo tanto a éstos datos como una muestra. Aquí no se va a establecer esta distinción.

1.3.1. Medidas de centralización

Una medida de centralización o localización indica el centro de una colección de números. Si los números provienen de diferentes medidas de la misma cantidad una medida de localización se utiliza para *conjeturar* el valor de esa cantidad. En otras situaciones la medida de localización representa un resumen de la colección de números, como cuando se dice, *la nota media de una clase en un examen es 6'5.*

Ya se han visto una colección de medidas de localización, los *momentos muestrales respecto al origen* incluidos en la Definición 1.2.3, el que se usa con mayor frecuencia es el primero $\overline{x} = a_1$ que cuando los datos están agrupados pasa a ser

$$\overline{x} = \sum_{i=1}^{k} f_i x_i.$$

La media muestral presenta el inconveniente de depender de todas las observaciones y se ve muy afectada por los valores extremos. Por este motivo se recurre a definir otras medidas de centralización que sean más *robustas*, es decir menos vulnerables en el sentido antes citado.

Definición 1.3.1.1 (mediana muestral)
Dada una colección finita de elementos la *mediana muestral* M es el valor central, entendiendo como tal el valor $M = x_{([\frac{n}{2}]+1)}$ si n es impar y $M = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right)$ si n es par.

Por $[x]$ se entiende la parte entera de x y $x_{(1)} \leq ... \leq x_{(n)}$ es la muestra ordenada.

Definición 1.3.1.2 (media truncada)
Para una colección finita de elementos se llama *media truncada* de orden α con $0 < \alpha < 0'5$, a la media $\overline{x}_\alpha$ de los elementos que quedan de quitar el $100\alpha\,\%$ más bajo y el $100\alpha\,\%$ más alto de los mismos. Formalmente

$$\overline{x}_\alpha = \frac{1}{n - 2[n\alpha]}\left(x_{([n\alpha]+1)} + ... + x_{(n-[n\alpha])}\right)$$

Las medias truncadas más usadas son $\overline{x}_{0'1}$ y $\overline{x}_{0'2}$.

Ejemplo 1.3.1.1
Para la colección de datos

$$35 \quad 69 \quad 72 \quad 77 \quad 79 \quad 79 \quad 84 \quad 85 \quad 86 \quad 87 \quad 95$$

la mediana muestral es $M = 79$ y la *media truncada* correspondiente a $\alpha = 0'2$ es

$$\begin{aligned} \overline{x}_{0'2} &= \tfrac{1}{7}(72 + 77 + ... + 86) \\ &= 80'2857 \end{aligned}$$

1.3.2. Medidas de dispersión

Una medida de dispersión o escala indica lo alejados que están los datos respecto a un valor central. Los *momentos muestrales respecto al centro*, ver Definición 1.2.3, son todos ellos medidas de dispersión. El más usado es la *cuasivarianza muestral*, véase la Expresión 1.2.2, que cuando los datos están agrupados toma la forma

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i(x_i - \overline{x})^2$$

donde f_i es la *frecuencia relativa* con que aparece el valor x_i, es decir el cociente entre el número de veces que aparece x_i partido por el número total de observaciones n. La cuasivarianza muestral también presenta el inconveniente de ser muy sensible a errores en los datos; por lo que en homogeneidad con las medidas de localización se recurre a definir otras medidas de dispersión.

Definición 1.3.2.1 (cuantiles; recorrido intercuartílico)
Dada una colección finita de n números se llama *cuantil muestral* de orden p, con $p \in (0,1)$ al valor x entendiendo como tal el valor $x_{([np]+1)}$ si np no es un número entero y $\frac{1}{2}(x_{(np)} + x_{(np+1)})$ si np es un número entero. Los *cuantiles muestrales* más usados son la *mediana* que se corresponde con $p = 1/2$ (véase la Definición 1.3.1.1) y los tres cuartiles que se corresponden el *primer cuartil* con $p = 1/4$, el *segundo cuartil* con $p = 2/4$ (coincide con la mediana), y el *tercer cuartil* con $p = 3/4$. Se denomina *recorrido intercuartílico muestral* (RIM), a la diferencia entre el tercer y primer cuartil muestrales de la misma forma que se denomina *recorrido intercuartílico poblacional* a la diferencia entre el tercer y primer cuartil poblacional.

Se recuerda que dada una función de distribución F se define *cuantil poblacional* de orden p, al valor x_p tal que se verifique

$$F(x_p) \geq p \quad \text{y} \quad P\{X \geq x_p\} \geq 1 - p \tag{1.3.2.1}$$

o de forma equivalente

$$p \leq F(x_p) \leq p + P\{X = x_p\}.$$

La definición de cuantil muestral puede hacerse de la misma forma cambiando la función de distribución poblacional por la función de distribución empírica, (véase la Definición 1.4.1). Aquí se ha preferido dar una definición esencialmente idéntica, pero que cambie todo un intervalo por un único valor, la semisuma de los extremos.

Definición 1.3.2.2 (mediana de las desviaciones absolutas)
Dada la colección de n números se llama *mediana de las desviaciones absolutas* MDA, a la mediana de los números $|x_i - M|$, donde M es la mediana muestral. Es también una medida *robusta* de dispersión en el sentido de que no es demasiado sensible a observaciones erróneas o atípicas.

Ejemplo 1.3.2.1
Para la colección de datos

$$35 \quad 69 \quad 72 \quad 77 \quad 79 \quad 79 \quad 84 \quad 85 \quad 86 \quad 87 \quad 95$$

la *cuasivarianza muestral* es $s^2 = 247'89$, el *primer cuartil* es 72 el *tercer cuartil* es 86, el *recorrido intercuartílico* (RIM) es 14, la *mediana muestral* es 79 y la *mediana de las desviaciones absolutas* (MDA) es 7.

1.3.3. Diagramas de tallos y hojas y de caja

Como ya se ha dicho, la representación más extendida de una colección de datos es mediante la *tabla de frecuencias,* sin embargo existen otras representaciones adecuadas, que están a medio camino entre la representación numérica y gráfica. Éstas son el *diagrama de tallos y hojas* y el *diagrama de caja*, principalmente. El *diagrama de tallos y hojas* se construye realizando los siguientes pasos:

a) se redondea la colección de datos a dos o tres cifras
b) las cifras de orden superior determinan los *tallos* normalmente de una o dos cifras
c) a continuación se escribe una raya vertical y se pone la cifra de las unidades de orden inferior que determina las *hojas*

así los datos 1236, 1278 darían lugar, redondeados a tres cifras, a 124, 128 ambos se corresponderían con el *tallo* 12 y darían lugar a las *hojas* 4 y 8 con lo que su representación en el diagrama sería

$$12|48$$

Ejemplo 1.3.3.1
Para los datos

256	256	257	257	261	262	266	267	267
268	268	268	268	273	274	278	283	284

el *diagrama de tallos y hojas* correspondiente viene dado por

4	25	6677
(13)	26	126778888
5	27	348
2	28	34

Siguiendo una práctica usual se han incluido delante de los *tallos* las frecuencias acumuladas hasta llegar al *tallo* que contiene a la mediana, que se indica con paréntesis. Se observa que las *hojas* juegan gráficamente el mismo papel, que las frecuencias absolutas correspondientes a las clases en la tabla de frecuencias, con la ventaja de que informan de los valores que realmente toma la variable.
■

El *diagrama de caja* es también una representación a medio camino entre numérica y gráfica y recoge la mediana, el recorrido intercuartílico, la presencia de posibles valores discordantes o atípicos, junto con una indicación de la posible asimetría de los datos. El *diagrama de caja* se construye de la siguiente forma:

a) se representan tres líneas horizontales correspondientes a la mediana y al primer y al tercer cuartil, que dan lugar a la caja;
b) se trazan dos rectas horizontales, de menor tamaño que las que delimitan la caja y que cubran las observaciones que queden dentro del primer cuartil menos 1’5 RIM y del tercer cuartil más 1’5 RIM;
c) se representan por un punto las observaciones que quedan por debajo del primer cuartil menos 1’5 RIM y por encima del tercer cuartil más 1’5 RIM si es que existen.

Los *diagramas de caja* no están estandarizados, la estructura que se acaba de explicar es la más extendida con modificaciones mínimas. Esta representación da una idea de donde se encuentra el centro de la colección, mediante la mediana; de la dispersión de los datos, mediante el RIM; de la asimetría de los mismos, mediante la distancia de los cuartiles a la mediana y por último informa de la posible existencia de observaciones discordantes, o que desentonan demasiado con los restantes datos en el sentido establecido en c).

Ejemplo 1.3.3.2
Para los datos del Ejemplo 1.3.2 el *diagrama de caja* está recogido en la Figura 1.3.3.1 los datos para su constucción son: la mediana 218, el primer cuartil 198 y el tercer cuartil 248. El dato atípico 345 ha dado lugar al

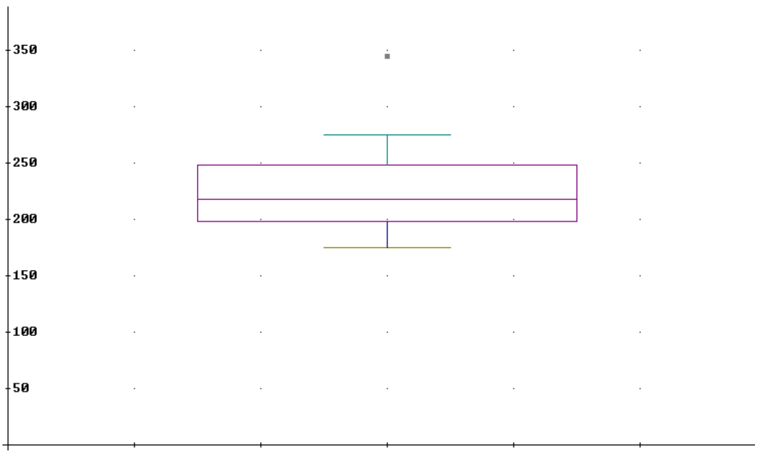

Figura 1.3.3.1. Diagrama de caja correspondiente al Ejemplo 1.3.3.2.

punto aislado por encima de las rectas horizontales. Se observa la ligera asimetría de los puntos por encima de la mediana.

1.4. PROPIEDADES ASINTÓTICAS DE LOS MOMENTOS MUESTRALES

El objetivo de esta sección es demostrar que los momentos muestrales convergen casi seguro a sus correspondientes momentos poblacionales, lo que justificará la utilización de los primeros para estimar los segundos en una población. La aproximación será entonces mayor cuanto mayor sea el tamaño muestral. Además se incluye la función de distribución empírica.

Teorema 1.4.1
Si una población tiene momentos de orden k, el momento muestral respecto al origen de orden k, converge, casi seguro, al momento poblacional respecto al origen de orden k

$$a_k \xrightarrow{c.s.} \alpha_k.$$

Demostración: Basta ver que por la ley fuerte de Kintchine (Loève (1963) pág. 239) dada una sucesión de v.a. $\{Y_n\}_{n\in\mathbb{N}}$ con $E[y_i] = \mu < +\infty$ se sigue que

$$\overline{Y}_n \xrightarrow{c.s.} \mu$$

donde $\overline{Y}_n = \frac{1}{n}\sum Y_i$. El resultado se obtiene tomando $Y_i = X_i^k$.

Teorema 1.4.2
Si una población tiene momentos de orden k, el momento muestral respecto al centro de orden k, converge, casi seguro, al momento poblacional respecto al centro de orden k

$$b_k \xrightarrow{c.s.} \beta_k.$$

Demostración: Si se tiene en cuenta que

$$b_k = a_k - \binom{k}{1}\overline{x}\ a_{k-1} + ... \pm \overline{x}^k$$

y que si $\{X_n^{(i)}\}_{n\in\mathbb{N}}$ $i = 1, ..., p$ converge c.s. a $X^{(i)}$ y es g una función continua de $\mathbb{R}^p$ en $\mathbb{R}$, entonces

$$g(x_n^{(1)}, ..., x_n^{(p)}) \xrightarrow{c.s.} g(x^{(1)}, ..., x^{(p)}),$$

resultado que se demuestra fácilmente, ya que si se denota por $\overrightarrow{X}_n = (X_n^{(1)}, ..., X_n^{(p)})$ y $\overrightarrow{X} = (X^{(1)}, ..., X^{(p)})$, por $\overrightarrow{X}_n \xrightarrow{c.s.} \overrightarrow{X}$ $P\{\lim_{n\to\infty} \overrightarrow{X}_n = \overrightarrow{X}\} = 1$. Pero el suceso $\lim_{x\to\infty} \overrightarrow{X}_n = \overrightarrow{X}$ está contenido en el suceso $\lim_{n\to\infty} g(\overrightarrow{X}_n) = g(\overrightarrow{X})$ puesto que si $\lim_{n\to\infty} \overrightarrow{X}_n = \overrightarrow{X}$ aplicando g se sigue que $g(\lim_{n\to\infty} \overrightarrow{X}_n) = g(\overrightarrow{X})$ y por ser g continua $\lim_{n\to\infty} g(\overrightarrow{X}_n) = g(\lim_{n\to\infty} \overrightarrow{X}_n)$; luego

$$P\{\lim_{n\to\infty} g(\overrightarrow{X}_n) = g(\overrightarrow{X})\} = 1$$

tomando como g la expresión para b_k y como $X_n^{(i)}$ los momentos respecto al origen al variar i, se obtiene que $b_k \xrightarrow{c.s.} \beta_k$.

■

También se puede probar la convergencia en ley de los momentos muestrales hacia la distribución normal.

Teorema 1.4.3
Si la población tiene momentos de orden $2k$ respecto al origen entonces se tiene la siguiente convergencia en ley

$$\frac{a_k - \alpha_k}{\sqrt{\alpha_{2k} - \alpha_k^2}}\sqrt{n} \xrightarrow{\mathcal{L}} Z \sim N(0,1).$$

Demostración: Basta utilizar el teorema central del límite, $\{Y_n\}$ con v.a.i.i.d. con momento de segundo orden y $E[Y_n] = \mu \quad V[Y_n] = \sigma^2 < +\infty$ entonces $\frac{\overline{Y}-\mu}{\sigma}\sqrt{n} \longrightarrow Z \sim N(0,1)$. Se puede considerar entonces $Y_n = X_n^k$, $a_k = \overline{Y}$ con lo que $\mu = \alpha_k$ y $\sigma^2 = \alpha_{2k} - \alpha_k^2$.

■

La convergencia en ley de los momentos muestrales respecto al centro queda recogida en el siguiente resultado:

Teorema 1.4.4
Si existe el momento poblacional respecto al origen de orden $2k$ entonces

$$\sqrt{n}\frac{b_k - \beta_k}{\sqrt{\beta_{2k} - \beta_k^2 - 2k\beta_{k-1}\beta_{k+1} + k^2\beta_{k-1}^2\beta_2}} \xrightarrow{\mathcal{L}} Z\, N(0,1).$$

Demostración: Puede verse en Cramér (1973) pág. 419 y se apoya en el Lema 1.4.1 que tiene interés por sí mismo por lo que se va a demostrar a continuación. Además el siguiente resultado, que relaciona la convergencia en ley con la convergencia en probabilidad, también será de utilidad en el estudio de convergencias a lo largo del texto.

Teorema 1.4.5 (Slutsky)
Si $X_n \xrightarrow{\mathcal{L}} X$ y $Y_n \xrightarrow{P} a$ entonces se tiene que

I) $Y_n X_n \xrightarrow{\mathcal{L}} aX$

II) $X_n + Y_n \xrightarrow{\mathcal{L}} X + a$

Demostración: puede verse en Bickel y Doksun(1977) pág. 461.

Lema 1.4.1
Si $\{a_n\}$ es una sucesión de constantes con $\lim\limits_{n\to\infty} a_n = +\infty$ y a es un número fijo, tal que $a_n(X_n - a) \xrightarrow{\mathcal{L}} X$, entonces para cualquier función $g : \mathbb{R} \to \mathbb{R}$ con derivada continua y no nula en a, se tiene que

$$a_n(g(X_n) - g(a)) \xrightarrow{\mathcal{L}} g'(a)X$$

Demostración: Si se desarrolla $g(X_n)$ alrededor de $g(a)$ en serie de Taylor, se tiene

$$g(X_n) = g(a) + (X_n - a)g'(a'_n)$$

donde $|a'_n - a| \leq |X_n - a|$. Ahora bien de las hipótesis se sigue que $X_n \xrightarrow{P} a$ ya que si se considera $X_n - a = \frac{1}{a_n}[a_n(X_n - a)]$ como la expresión entre corchetes converge en ley a X y $\frac{1}{a_n} \to 0$, por el Teorema de Slutsky se tiene que $X_n - a \xrightarrow{P} 0$. Por tanto $a'_n \xrightarrow{P} a$. Como $\alpha_n(g(X_n) - g(a)) = a_n(X_n - a)g'(a'_n)$ y la función g' es continua en a, entonces $g'(a'_n) \xrightarrow{P} g'(a)$ por lo que aplicando otra vez el Teorema de Slutsky se obtiene el resultado deseado.

■

Debido a su carácter más elemental, se incluye la obtención de la distribución asintótica del momento muestral respecto al centro más sencillo: la varianza muestral.

Teorema 1.4.6
Para m.a.s. de una población con momento α_4 finito, la distribución límite de $b_2 = \frac{1}{n}\sum_1^n (X_i - \overline{X})^2$ es tal que

$$\sqrt{n}(b_2 - \sigma^2) \xrightarrow{\mathcal{L}} Z \sim N(0, \sqrt{\beta_4 - \sigma^4})$$

Demostración: Se puede poner

$$b_2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu - (\overline{X} - \mu))^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y})^2$$

donde las variables $Y_i = X_i - \mu$ son independientes, tienen de media 0 y de varianza σ^2. Como

$$b_2 = \frac{1}{n}\sum_{i=1}^n Y_i^2 - \overline{Y}^2$$

se tiene que

$$\sqrt{n}(b_2 - \sigma^2) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n Y_i^2 - \sigma^2\right) - \sqrt{n}\overline{Y}^2.$$

Pero

$$\sqrt{n}\overline{Y}^2 = (n^{1/4}\overline{Y})^2 \qquad \text{y} \qquad n^{1/4}\overline{Y} = n^{-1/4}\sqrt{n}\overline{Y};$$

como

$$\sqrt{n}\overline{Y} \xrightarrow{\mathcal{L}} Z \sim N(0, \sigma)$$

(en virtud del Teorema 1.4.3) y $n^{-1/4} \xrightarrow{P} 0$; luego por el Teorema de Slutsky, $n^{1/4}\overline{Y} \xrightarrow{\mathcal{L}} 0$ y por tanto converge también en probabilidad, con lo que su cuadrado, $\sqrt{n}\overline{Y}^2$ converge en probabilidad hacia 0. Por el Teorema 1.4.3,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \sigma^2\right) \xrightarrow{\mathcal{L}} Z \sim N(0, \sqrt{\beta_4 - \sigma^4}),$$

puesto que

$$E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i^2\right] = \sigma^2 \ y \ \ V\left[\frac{1}{n}\sum_{i=1}^{n} Y_i^2\right] = \frac{V[Y_i^2]}{n} = \frac{\beta_4 - \sigma^4}{n}.$$

Otra vez más por el Teorema de Slutsky

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \sigma^2\right) - \sqrt{n}\overline{Y}^2 \xrightarrow{\mathcal{L}} Z \sim N(0, \sqrt{\beta_4 - \sigma^4}).$$

Ejemplo 1.4.1
Si se sabe que

$$\sqrt{n}(Z_n - \theta) \xrightarrow{\mathcal{L}} X$$

donde $X \sim N(0, \sigma(\theta))$, lo que ocurre por ejemplo para $Z_n = \overline{X}$ en una población con momento de orden 2 finito y media θ, entonces

$$\sqrt{n}(g(Z_n) - g(\theta)) \xrightarrow{\mathcal{L}} Y$$

con $Y \sim N(0, |g'(\theta)|\sigma(\theta))$, sin más que exigir que g tenga derivada continua y no nula en el punto θ.

Ejemplo 1.4.2 (estabilización de la varianza)
Dada una v.a., si se conoce la distribución asintótica de una función sencilla de ella, como ocurre en el Ejemplo 1.4.1, se puede tratar de obtener una función cuya varianza se independiente de θ. Por ejemplo, si X sigue una distribución de Poisson $\mathcal{P}(\theta)$, entonces, por el teorema central del límite

$$\sqrt{n}\left(\overline{X} - \theta\right) \xrightarrow{\mathcal{L}} Z$$

con $Z \sim N(0, \sqrt{\theta})$,

$$\sqrt{n}(g(\overline{X}) - g(\theta)) \xrightarrow{\mathcal{L}} Y$$

con $Y \sim N(0, |g'(\theta)|\sqrt{\theta})$. Se trata de obtener una función g tal que $\sqrt{\theta}g'(\theta)$ no dependa de θ. Esto se consigue, por ejemplo, con $g(\theta) = \sqrt{\theta}$, pues $\sqrt{\theta}g'(\theta) = \frac{1}{2}$. Por tanto $g(\overline{X}) = \sqrt{\overline{X}}$ es tal que

$$\sqrt{n}\left(\sqrt{\overline{X}} - \sqrt{\theta}\right) \xrightarrow{\mathcal{L}} Y$$

con $Y \sim N(0, 1/2)$.

Ejemplo 1.4.3
Para una población de Bernoulli de parámetro θ, la distribución asintótica de $\sqrt{n}\overline{X}(1-\overline{X})$ puede calcularse, debidamente centrada, a partir de la distribución de $\overline{X}$. Por el teorema central del límite se sabe que

$$\sqrt{n}\left(\overline{X}-\theta\right) \xrightarrow{\mathcal{L}} Z \sim N(0, \sqrt{\theta(1-\theta)})$$

luego si se considera $g(\theta)=\theta(1-\theta)$ se sigue, aplicando el Lema 1.4.1, que

$$\sqrt{n}(\overline{X}(1-\overline{X})-\theta(1-\theta)) \xrightarrow{\mathcal{L}} Z \in N(0, |1-2\theta|\sqrt{\theta(1-\theta)}).$$

■

Otro estadístico muestral importante es la *función de distribución empírica* o *muestral.*

Definición 1.4.1 (distribución empírica)
Dada una realización particular de una muestra $(x_1, ..., x_n)$ viene definida mediante

$$F_n^*(x) = \begin{cases} 0 & \text{si} \quad x < x_{(1)} \\ \frac{k}{n} & \text{si} \quad x_{(k)} \leq x < x_{(k+1)} \\ 1 & \text{si} \quad x \geq x_{(n)} \end{cases} \qquad k = 1, ..., n-1$$

donde se recuerda que $(x_{(1)}, ..., x_{(n)})$ es la muestra ordenada de menor a mayor.

■

Alternativamene la función $F_n^*(x)$ puede ponerse como

$$F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(x_i). \tag{1.4.1}$$

Para todo x se verifica que $F_n^*(x) \xrightarrow{c.s.} F(x)$, siendo $F(x)$ la función de distribución de la población. En efecto, haciendo

$$Y_i = I_{(-\infty,x]}(x_i)$$

se tiene que $E[Y_i]=F(x)$, $V[Y_i]=F(x)[1-F(x)]$ y $F_n^*(x)=\overline{Y}$. Aplicando ahora la ley fuerte de los grandes números se obtiene la convergencia c.s. citada. Además, de la Expresión (1.4.1) se sigue, por el teorema central del límite, que

$$\sqrt{n}\frac{F_n^*(x)-F(x)}{\sqrt{F(x)[1-F(x)]}} \xrightarrow{\mathcal{L}} Z \sim N(0,1).$$

La convergencia casi seguro puede refinarse aún más como indica el siguiente teorema, que es llamado por algunos el teorema central de la estadística.

Teorema 1.4.7 (Glivenko-Cantelli)
Si se tienen m.a.s. de tamaño n de una población X, con función de distribución $F(x)$, para cualquier número real positivo arbitrario ε, se tiene que

$$\lim_{n\to\infty} P\left\{\sup_{x\in\mathbb{R}}|F_n^*(x) - F(x)| \geq \varepsilon\right\} = 0.$$

Demostración: La demostración puede verse en Fisz (1963) pág. 391.
■

Es importante señalar que en este teorema la probabilidad del primer miembro está obtenida a partir de la distribución F; por lo que sería más correcto emplear la notación P_F en lugar de P.

En virtud del Teorema 1.4.7 si se quiere determinar de forma aproximada la función de distribución, se puede obtener una m.a.s. de tamaño n gande, construir a partir de ella la función de distribución empírica F_n^* y esperar que $F_n^*(x)$ sea una buena aproximación de $F(x)$.

Ejemplo 1.4.4
Es inmediato, mediante un programa de ordenador, construir $F_{100}^*(0)$ a partir de una m.a.s. de una población $N(0,1)$. Por ejemplo, simulando una muestra de 100 valores de la distribución $N(0,1)$ se observa que son 52 menores o iguales que 0, por lo tanto es $F_{100}^*(0) = 0,52$. Es sabido que el valor de la función de distribución de la $N(0,1)$ en el punto 0 es $0'5$, por lo que la aproximación es bastante buena.

1.5. DISTRIBUCIONES EN EL MUESTREO ASOCIADAS A LA DISTRIBUCIÓN NORMAL

El Teorema de Glivenko-Cantelli justifica de alguna manera, la sustitución de los parámetros desconocidos de la función de distribución de la población, por sus características análogas calculadas a partir de la muestra. Es por tanto interesante estudiar las distribuciones de los estadísticos muestrales cuando la población de partida es conocida. Debido al papel central de la distribución normal en el Cálculo de Probabilidades, se estudia a continuación la distribución de los momentos muestrales cuando la población es normal.

1.5.1. Distribución de la media muestral

Dada una m.a.s. de tamaño n de una población X con distribución $N(\mu,\sigma)$, se sigue que la distribución de la media muestral $\overline{X}$ es una $N(\mu, \frac{\sigma}{\sqrt{n}})$.

En efecto, la función característica de la media muestal es

$$\begin{aligned}\varphi_{\overline{X}}(t) &= E[e^{i\frac{t}{n}\sum\limits_{1}^{n} X_i}] = [\varphi_X(\tfrac{t}{n})]^n = e^{\left(i\frac{t}{n}\mu - \frac{1}{2}\frac{\sigma^2}{n^2}t^2\right)n} \\ &= e^{it\mu - \frac{1}{2}\frac{\sigma^2}{n}t^2}\end{aligned}$$

que es la función característica de la distribución $N(\mu, \frac{\sigma}{\sqrt{n}})$.

1.5.2. Distribución del momento muestral respecto al origen de orden dos

Para m.a.s. de tamaño n de una población $N(0,\sigma)$ la distribución de $a_2 = \frac{1}{2}\sum\limits_{1}^{n} X_i^2$ viene dada por

$$(1.5.2.1) \qquad g(y) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})}\left(\frac{n}{\sigma^2}\right)^{n/2} e^{-\frac{n}{2\sigma^2}y} y^{\frac{n}{2}-1} \qquad \text{con} \qquad y \geq 0.$$

Procediendo por pasos sucesivos, si X tiene distribución $N(0,1)$ la distribución de X^2 es una gamma ($a = \frac{1}{2}, p = \frac{1}{2}$), ya que para $y \geq 0$

$$P\{X^2 \leq y\} = P\{-\sqrt{y} \leq X \leq \sqrt{y}\} = 2\Phi(\sqrt{y}) - 1$$

con Φ la función de distribución de la v.a. $N(0,1)$. Derivando respecto de y

$$f_{X^2}(y) = \varphi(\sqrt{y})y^{-1/2}$$

donde φ es la función de densidad de la $N(0,1)$. Por tanto, la densidad de X^2 es $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x}x^{-\frac{1}{2}}$, con $x \geq 0$. Como la función característica de la distribución *Gamma* (a,p) es $\left(1 - \frac{2it}{a}\right)^{-p}$, se sigue que la distribución de $\sum\limits_{1}^{n} X_i^2$, cuando X_i se distribuye $N(0,1)$, es una *Gamma* ($a = \frac{1}{2}, p = \frac{n}{2}$), es decir

$$(1.5.2.2) \qquad f(x) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})}e^{-\frac{1}{2}x}x^{\frac{n}{2}-1} \qquad \text{con} \qquad x \geq 0.$$

A una v.a. con esta densidad se la denomina variable aleatoria χ_n^2, donde el parámetro n recibe el nombre de *número de grados de libertad*. Es fácil establecer (ver Ejercicio 1.9.14) que la media de la distribución χ_n^2 es n y que su varianza es $2n$. En la Figura 1.5.2.1 pueden verse distintas representaciones de la función de densidad al variar el número de grados de libertad.

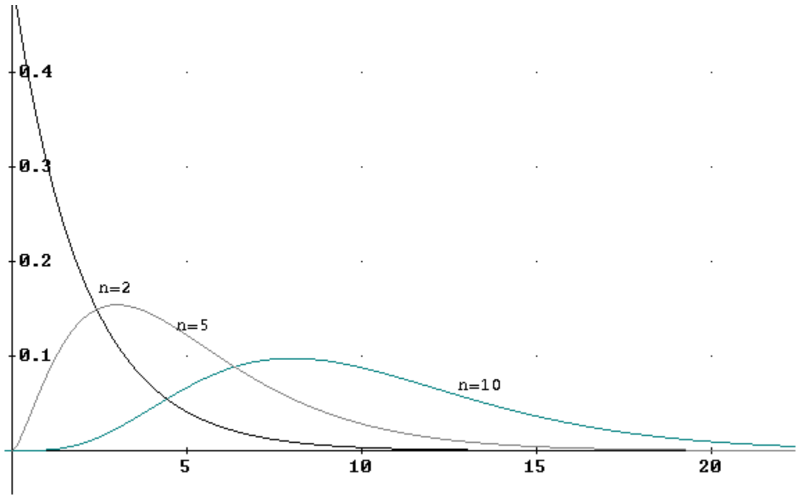

Figura 1.5.2.1. Funciones de densidad de la v.a. χ^2_n para $n = 2, 5, 10$.

Se puede entonces introducir la siguiente distribución.

Definición 1.5.1 (distribución χ^2_n)
Para m.a.s. de tamaño n de una población $N(0,1)$ $(X_1, ..., X_n)$ la v.a. $Y = \sum_{i=1}^{n} X_i^2$ tiene una distribución χ^2_n con n el número de grados de libertad, o de forma equivalente cuando su función de densidad es

$$f_{\chi^2_n}(x) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} e^{-\frac{1}{2}x} x^{\frac{n}{2}-1} \qquad \text{con} \qquad x \geq 0 \tag{1.5.2.2}$$

Al final del libro, en el Apéndice titulado Tablas, la Tabla 4 recoge cuantiles útiles de la distribución χ^2_n para diferentes valores de n. A título de ejemplo la probabilidad de que la distribución χ^2_{24} tome un valor superior a $36'4$ es $1 - 0'95 = 0'05$. La distribución χ^2_n puede generalizarse al caso en que las variables X_i sean independientes pero tengan distribución $N(\mu_i, 1)$ con $\mu_i \neq 0$. Ya que la función característica de $\sum_{i=1}^{n} X_i^2$ es

$$\begin{aligned}
E[e^{it\sum_{i=1}^{n} X_i^2}] &= \int ... \int e^{it\sum_{1}^{n} x_j^2} e^{-\frac{1}{2}\sum_{j=1}^{n}(x_j-\mu_j)^2} \frac{1}{(2\pi)^{n/2}} dx_1 ... dx_n \\
&= \frac{1}{(2\pi)^{n/2}} \int ... \int e^{-\frac{1-2it}{2}\sum_{j=1}^{n}(x_j - \frac{\mu_j}{1-2it})^2 - \sum_{j=1}^{n}\frac{\mu_j^2}{2}(1-\frac{1}{1-2it})} dx_1 ... dx_n \\
&= \frac{1}{(2\pi)^{n/2}} e^{-\lambda(1-\frac{1}{1-2it})} \int ... \int e^{-\frac{1-2it}{2}\sum_{j=1}^{n}(x_j - \frac{\mu_j}{1-2it})^2} dx_1 ... dx_n
\end{aligned}$$

donde $\lambda = \frac{1}{2}\sum_{1}^{n}\mu_j^2$. Como

$$
\begin{aligned}
e^{-\lambda(1-\frac{1}{1-2it})}(1-2it)^{-n/2} &= e^{-\lambda}(1-2it)^{-n/2}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!(1-2it)^j} \\
&= e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}(1-2it)^{-(\frac{n}{2}+j)}
\end{aligned}
$$

mediante el teorema de inversión de la función característica, (véase Loève(1963) pág. 188), se obtiene la función de densidad, que resulta ser

$$
\begin{aligned}
f(x) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\varphi(t)dt \\
&= \sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{j!}\frac{1}{2^{\frac{n}{2}+j}\Gamma(\frac{n}{2}+j)}e^{-\frac{x}{2}}x^{\frac{n}{2}+j-1} \qquad \text{con} \qquad x \geq 0
\end{aligned}
\tag{1.5.2.3}
$$

densidad correspondiente a la v.a. χ_n^2 no centrada con *parámetro de no centralidad* $\lambda = \frac{1}{2}\sum_{1}^{n}\mu_j^2$ y n número de grados de libertad.

■

Para obtener la distribución de otros momentos muestrales, se necesita el siguiente resultado que recoge la distribución de ciertas funciones de v.a. normales.

Lema 1.5.1
Dada una m.a.s. de tamaño n $(X_1, ..., X_n)$ de una población $N(0, \sigma)$, si se construyen las v.a.

$$
Y_i = c_i' X = (c_{i_1}, ..., c_{i_n}) \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \qquad i = 1, ..., p
$$

donde p es menor que n, con $c_i' c_j = \delta_{ij}$ $(\delta_{ij} = 1$ si $i = j$ y $\delta_{ij} = 0$ si $i \neq j)$, la v.a.

$$
Z = \sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{p} Y_i^2,
$$

es tal que $\frac{Z}{\sigma^2} \sim \chi_{n-p}^2$, y además la v.a. Z es independiente de las variables $Y_1, ..., Y_p$.

Demostración: Inicialmente se determinan $n-p$ vectores más c_i de dimensión $n \times 1$ para $i = p+1, ..., n$, que sigan cumpliendo la restricción $c_i' c_j = \delta_{ij}$; esto siempre es posible ya que se puede encontrar en $\mathbb{R}^n$ una base ortonormal de n elementos cuyos p primeros vectores sean los vectores c_i, $i = 1, ..., p$. Se tiene entonces

$$
Y_i = c_i' \mathbf{X} \qquad i = 1, ..., n
$$

o escrito matricialmente $\mathbf{Y} = C\mathbf{X}$. Es claro que $E[Y_i] = c_i' E[X] = 0$ y por ser las variables Y_i combinaciones lineales de variables normales, tienen distribución normal de media 0, varianza σ^2 y son independientes, pues

$$Cov(Y_i, Y_j) = \sigma^2 \sum_{k=1}^{n} c_{ik} c_{jk} = \sigma^2 \delta_{ij}.$$

Si ahora se expresa la variable aleatoria Z en función de las v.a. Y_i

$$Z = \mathbf{X}'\mathbf{X} - \sum_{i=1}^{p} Y_i^2 = \mathbf{Y}'CC'\mathbf{Y} - \sum_{i=1}^{p} Y_i^2 = \sum_{i=p+1}^{n} Y_i^2$$

con lo que Z es independiente de $Y_1, ..., Y_p$ y además $\frac{Z}{\sigma^2}$, de acuerdo con la Definición 1.5.2.1, tiene distribución χ^2_{n-p}.

Teorema 1.5.1 (Fisher)
Para m.a.s. de tamaño n de una población $N(\mu, \sigma)$ se tiene que b_2, momento muestral respecto al centro de orden 2 y $\overline{X}$, media muestral, son variables aleatorias independientes, además $\overline{X}$ tiene distribución $N(\mu, \frac{\sigma}{\sqrt{n}})$ y $n\frac{b_2}{\sigma^2}$ tiene distribución χ^2_{n-1}.

Demostración: Basta observar que

$$\begin{aligned} nb_2 &= \sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n}(X_i - \mu - (\overline{X} - \mu))^2 \\ &= \sum_{i=1}^{n}(X_i - \mu)^2 - n\left(\overline{X} - \mu\right)^2. \end{aligned}$$

Dado que $Y_i - \mu \sim N(0, \sigma)$, se puede aplicar el Lema 1.5.2.1 con $p = 1$,

$$Y_1 = \left(\frac{1}{\sqrt{n}}, ..., \frac{1}{\sqrt{n}}\right) \begin{pmatrix} X_1 - \mu \\ \vdots \\ X_n - \mu \end{pmatrix}$$

y $Z = nb_2$. Entonces $n\frac{b_2}{\sigma^2} \sim \chi^2_{n-1}$ es independiente de Y_1 y, por tanto, también de $\overline{X}$. En definitiva, en m.a.s. de una población $N(\mu, \sigma)$, la media muestral $\overline{X}$ y la cuasivarianza muestral son v.a. independientes y además son conocidas sus distribuciones.

1.5.3. Distribución del cociente entre la media muestral y el momento respecto al origen de orden dos

Para encontrar la distribución de este cociente es necesario introducir la distribución de Student, la segunda distribución en el muestreo asociada