Unidad 1

- Estimación

9.1. PRIMEROS CONCEPTOS: ESTIMADOR PUNTUAL E INTERVALO DE CONFIANZA

En los capítulos previos hemos visto determinados *modelos de probabilidad* o *familias de distribuciones* que pueden ajustarse a las observaciones empíricas que se han obtenido acerca de un fenómeno que está siendo objeto de estudio. Al proceder así, suponemos que las características del fenómeno son similares a las del modelo de probabilidad que hemos escogido. Para poder proceder a evaluar tales ajustes y, en particular, para poder asignar probabilidades a sucesos, hemos de dar valores numéricos a los parámetros que definen la familia de distribuciones de probabilidad utilizada. En este capítulo deducimos, de manera rigurosa, el proceso de utilización de la información muestral para asignar valores a los parámetros desconocidos en una población.

En primer lugar, estudiaremos la *estimación puntual*, procedimiento para asociar un valor numérico a cada parámetro desconocido. En la mayoría de los casos, no nos quedaremos satisfechos con asignar un valor numérico a cada parámetro desconocido, sino que querremos avanzar más, proporcionando un *intervalo de confianza*, lo que constituye otra solución al problema de *estimación*. Un intervalo de confianza es todo un rango de valores posibles, dentro de los cuales se encuentre el verdadero valor del parámetro con un determinado *nivel de confianza*, fijado de antemano.

Por último, en muchas ocasiones querremos utilizar la información muestral para *contrastar* un valor numérico determinado para un parámetro desconocido. Por ejemplo, un doctor que quiere analizar la efectividad de un medicamento para regular la presión sanguínea querrá contrastar si la presión sanguínea media de un grupo de enfermos bajo tratamiento es igual o, por el contrario, inferior, a la presión sanguínea media de un grupo de control al que no se ha suministrado el medicamento. Se tratará, por tanto, de estimar el valor esperado de la distribución de los valores de presión sanguínea media de ambos grupos de pacientes, los que han estado o no bajo tratamiento, y querremos *contrastar* si las esperanzas matemáticas en ambos grupos son iguales. Analizaremos la contrastación de hipótesis estadísticas en el Capítulo 10, reservando el presente capítulo al estudio de la estimación. Ambos constituyen los procedimientos de la *inferencia estadística*.

Un *parámetro* es una constante desconocida: así, los parámetros que aparecen en las funciones de densidad o de cuantía de la distribución que ajustamos a una determinada muestra, son constantes, pero generalmente desconocidas. Por ejemplo, aunque supongamos que la distribución de la renta de las familias españolas tiene distribución $N(\mu, \sigma^2)$, desconoceremos los valores numéricos de ambos parámetros. El investigador querrá disponer de procedimientos mediante los cuales asignar valores numéricos a μ y σ^2 si dispone de una muestra de 1.000 familias, cada una de las cuales ha declarado su renta. Trabajando con muestras no experimentales, como es el caso en Economía, nunca averiguaremos el verdadero valor numérico de los parámetros desconocidos, a diferencia de lo que ocurre cuando realizamos pruebas con un determinado diseño experimental, cuyos parámetros son controlados por el investigador. Nuestro objetivo, por tanto, estriba en *estimar* los valores paramétricos de la distribución de probabilidad de la que se extrajo la muestra de que disponemos. Para ello asociamos a cada parámetro desconocido un *estadístico muestral*. Un *estadístico* es una función de las observaciones muestrales: la media y la varianza muestrales son estadísticos. También lo son la mediana de los valores numéricos observados en la muestra, su mínimo o máximo, el primer valor recogido, etc. Cuando decidimos asignar a un parámetro desconocido el valor numérico que toma un determinado estadístico muestral, decimos que dicho estadístico es un *estimador* del parámetro poblacional. El valor numérico del estimador en una muestra concreta constituye una *estimación* del parámetro.

En un proceso de estimación efectuamos, por tanto, dos aproximaciones. En primer lugar, hacemos un supuesto acerca de la familia de distribuciones de probabilidad que generó la muestra. En segundo lugar, se asigna un valor numérico, mediante un estimador, a cada uno de los parámetros desconocidos que caracterizan dicha familia de distribuciones.

En principio, la definición de *estimador* es bastante flexible, por cuanto que lo es cualquier *estadístico*, es decir, cualquier función de las observaciones muestrales. Sin embargo, es fácil apreciar que la mayoría de los posibles *estadísticos*, utilizados como estimadores, son muy pobres. Por ello, dedicamos buena parte de este capítulo a definir las propiedades que consideramos deseables para un estimador. En general, dado un problema de estimación, pocos estadísticos satisfarán las propiedades deseadas. Se trata, por consiguiente, de encontrar, para cada problema de estimación, la expresión analítica del estadístico o estimador que satisface unas ciertas condiciones.

Supongamos, por ejemplo, que un investigador quiere hacer un estudio de las alturas de los hombres y mujeres mayores de 25 años, de una determinada ciudad, para lo que acude a una céntrica plaza, donde pregunta la altura de las primeras 1.000 personas con dicha condición, de las cuales 520 resultan ser hombres y 480 mujeres. Tiene así dos muestras: $\mathbf{X} = \{x_1, x_2, ..., x_{520}\}$, e $\mathbf{Y} = \{y_1, y_2, ..., y_{480}\}$ de las tallas de hombres y mujeres, que puede utilizar por separado, o conjuntamente. Puesto que cree *a priori* que la altura de los hombres es algo mayor que la de las mujeres, y quizá está interesado en contrastar posteriormente tal hipótesis, utiliza las dos muestras por separado.

En primer lugar, debe decidir el objeto de su estudio: supongamos que el investigador decide estimar la altura media de la población, es decir, la esperanza matemática de la distribución de alturas de cada grupo y también la varianza de cada distribución. Se trata, por tanto, de estimar la esperanza matemática μ_h es decir, la altura media de los varones, para luego hacer lo mismo con la distribución de alturas de las mujeres. En segundo lugar, ha de hacer un supuesto acerca del tipo de distribución de probabilidad que sigue la variable aleatoria objeto de estudio: *Altura de los habitantes varones, mayores de 25 años, en dicha ciudad.* Supongamos que el investigador considera que los valores numéricos que toma dicha variable pueden aproximarse por una distribución $N(\mu,\sigma^2)$. En tal caso, cualquier función de los 520 valores observados x_i, es decir, cualquier operación que efectúe con las alturas observadas en los hombres, es un estimador potencial de μ_h. También tendrá que estimar la varianza σ_h^2 aunque, por simplicidad, nos concentramos en la estimación de μ_h.

Para ello, el primer registro muestral de este tipo, es decir, la altura observada al primer varón, x_1, siendo una función de las observaciones muestrales, es un posible estimador. Sin embargo, intuimos que sus propiedades no serán muy buenas, pues el primer transeúnte a quien preguntamos pudo ser una persona de 2,05, o de 1,55, valores posiblemente no muy *representativos* del parámetro que queremos estimar. Al construir un estimador queremos, por tanto, utilizar el máximo de información muestral, no restringirnos a una parte de ella, pues ignorar información disponible puede no ser *eficiente*, en un sentido que haremos concreto más adelante. La suma de las 520 observaciones utiliza toda la información muestral, pero sería un pésimo estimador, que no se corresponde tampoco con el parámetro μ, esperanza matemática de la *altura de los varones mayores de 25 años.* Si dividimos dicha suma por 520, el número de observaciones muestrales, tenemos un estadístico conocido, la *media muestral*, que enseguida reconocemos como un estimador potencialmente interesante de la esperanza matemática.

Supongamos que la media muestral obtenida entre las tallas de los varones, 1,82, es utilizada como estimación de μ_h, del mismo modo que la media muestral similar obtenida para las 480 mujeres, 1,74, es utilizada como estimación de μ_m. También podría pensarse en asociar los valores numéricos obtenidos para las varianzas de las muestras de alturas de hombres y mujeres, a los parámetros σ_h^2 y σ_m^2 Si, posteriormente, procedemos a calcular una *estimación por intervalo*, obtendremos resultados del tipo: *la estimación de la altura de los varones mayores de 25 años en la ciudad es de 1,82 y, con una confianza del 95 %, está situada entre 1,74 y 1,90; la estimación puntual de la altura media de las mujeres mayores de 25 años, en la ciudad objeto de estudio, es de 1,74 y, con una confianza del 95 %, está en el intervalo (1,68, 1,80).* Si el investigador quiere contrastar si la altura media de los hombres es superior a la de las mujeres mayores de 25 años, en dicha ciudad, tiene un problema, puesto que, si bien la estimación puntual de la primera es claramente superior a la de la segunda, ambos intervalos de confianza del 95 % tienen una parte en común. El problema sería más sencillo si los intervalos hubiesen sido: (1,79; 1,85) para los hombres y (1,71; 1,77) para las mujeres, pues los valores que creemos posibles, con una confianza del 95 % para la altura de los varones no comprenden ninguno de los valores del intervalo de igual confianza obtenido para las mujeres. Concluiríamos entonces que la altura de los varones es significativamente superior a la de las mujeres.

Comenzamos describiendo en este capítulo algunas de las propiedades que son deseables para un estimador, para analizar posteriormente algunos de los procedimientos para la obtención de estimadores.

Como vamos a tratar simultáneamente la estimación de parámetros tanto en distribuciones de probabilidad discretas como continuas, necesitamos una notación genérica que nos permita discutir cualquiera de estos problemas de estimación. Representaremos indistintamente por $\mathbf{X} = \{x_1, x_2, ..., x_n\}$ o por $\mathbf{Y} = \{y_1, y_2, ..., y_n\}$ muestras de tamaño n, y usaremos la notación común: $f(x_i/\theta)$ o $f(y_j/\theta)$ para denotar, tanto la función de probabilidad en el punto x_i o y_j en el caso de distribuciones discretas, como la función de densidad en dicho punto, en el caso de tratar con distribuciones continuas.

Denotamos por θ el parámetro que se desea estimar, y por $\hat{\theta}$ un estimador del mismo. Como $\hat{\theta}$ es un estadístico, es decir, una función de los elementos muestrales, su valor numérico específico depende de la información muestral y variaría si tomásemos otra muestra diferente de la misma población. Un estimador es, por tanto, *una variable aleatoria, no un parámetro*, y nos vamos a preocupar de caracterizar la distribución de probabilidad a la que obedece $\hat{\theta}$, así como de calcular su esperanza matemática y su varianza. Si $\hat{\theta}$ fuese constante, su esperanza matemática sería su valor numérico observado en una muestra cualquiera, y no variaría al tomar otra muestra diferente, y su varianza sería cero.

9.2. PROPIEDADES DE UN ESTIMADOR

9.2.1. Estimador insesgado

Parece razonable el requerimiento de que la distribución de probabilidad del estimador utilizado esté centrada alrededor del parámetro que pretendemos estimar. Es decir, queremos que el valor numérico del estimador, si se calcula en distintas muestras, no esté sistemáticamente por debajo o por encima del verdadero valor, desconocido, del parámetro que se estima. Un modo de concretar esta idea es:

Definición 9.1. Sea $\mathbf{X} = \{x_1, x_2, ..., x_n\}$ una muestra aleatoria simple extraída de una distribución de probabilidad dependiente de un parámetro θ cuyo valor se quiere estimar. Sea $\hat{\theta} = g(\mathbf{X}) = g(\{x_1, x_2, ..., x_n\})$ un *estimador* de θ. Decimos que $\hat{\theta}$ es un *estimador insesgado* de θ si: $E(\hat{\theta}) = \theta$. Cuando un estimador no es insesgado, decimos que es un *estimador sesgado.* El *sesgo* de un estimador es la diferencia entre su esperanza matemática y el verdadero valor del parámetro que se pretende estimar: $Sesgo(\hat{\theta}) = E(\hat{\theta}) - \theta$.

¿Cuál es el significado de esta propiedad? Supongamos que se han extraído muchas muestras independientes de una misma distribución de probabilidad, y que en cada una de ellas calculamos el valor numérico del estimador $\hat{\theta}$, es decir, la *estimación* de θ. A pesar de utilizar siempre el mismo estadístico como estimador, su valor numérico diferirá de unas muestras a otras. Pues bien, si el estimador es insesgado, su distribución de probabilidad, que podríamos aproximar mediante un *histograma de frecuencias,* al ir calculándolo en diversas muestras, tendrá como esperanza matemática θ, el verdadero valor numérico, desconocido, del parámetro que se pretende estimar. La propiedad de *insesgo* garantiza que el promedio de los valores numéricos de las estimaciones que se obtendrían con *todas las muestras posibles de un tamaño n* coincidiría con el verdadero valor de θ, pero es imposible obtener todas las estimaciones posibles. Lo que es importante, sin embargo, es reconocer que la propiedad de insesgo no garantiza nada acerca de si la estimación obtenida *a partir de una única muestra* se halla cerca o lejos del verdadero valor de θ. Esta consideración es especialmente relevante en Economía, donde, a diferencia de las ciencias experimentales, sólo disponemos, generalmente, de una muestra a partir de la cual obtener estimaciones de los parámetros desconocidos.

Ejemplo 9.1. Para determinar la implantación de un canal privado de televisión en un determinado municipio, se quiere estimar la proporción p de hogares que reciben dicho canal. Para ello, se visita una muestra de 1.275 hogares, preguntando si la familia es subscriptora del canal, obteniendo 345 respuestas afirmativas, y 930 respuestas negativas. El modelo de probabilidad que parece adecuado para este estudio es el de representar cada hogar por una variable de Bernouilli Y, $B(1,p)$ siendo p el parámetro que queremos estimar. La variable Y toma el valor $y_i = 1$ si la familia está suscrita al canal de televisión, e $y_i = 0$ en caso contrario, aunque también podría definirse al contrario. Un «éxito» en la prueba consiste, en este caso, en estar suscrito al canal, y el número de éxitos, suma de las variables y_i, tiene distribución $B(1275,p)$. Supongamos que tomamos:

$$\hat{p} = \frac{1}{1275} \sum_{i=1}^{1275} y_i$$

como estimador de p. Dada la naturaleza dicotómica $\{0,1\}$ de la variable y_i, en el numerador tenemos el total de hogares que en la muestra han resultado ser suscriptores del canal. Al dividir por 1.275, estamos calculando la media muestral de las y_i que, en este caso, coincide con el número de suscriptores en la muestra dividido por el tamaño muestral, es decir, *la proporción de hogares muestrales que son suscriptores del canal de TV.* Tenemos:

$$E(\hat{p}) = E\left(\frac{1}{1275}\sum_{i=1}^{1275} y_i\right) = \frac{1}{1275}E\left(\sum_{i=1}^{1275} y_i\right) = \frac{1}{1275}\sum_{i=1}^{1275} E(y_i) =$$
$$= \frac{1}{1275}\sum_{i=1}^{1275} p = \frac{1}{1275}(1275\,p) = p$$

donde hemos utilizado el hecho de que la esperanza matemática de una variable aleatoria $Y \sim B(1,p)$ es precisamente igual a p. En consecuencia, *la proporción muestral*, que se obtiene como la media muestral de las variables y_i, *es un estimador insesgado de la proporción poblacional* de hogares suscritos al canal de TV, que queríamos estimar.

La proporción muestral no es el único estimador insesgado de p. Supongamos que, cayendo en el defecto antes mencionado de ignorar una buena cantidad de información muestral, utilizamos el resultado del primer hogar encuestado, como estimador $\hat{p}_2$ de p. Es decir, estimamos $\hat{p}_2 = 100\ \%$ si $y_1 = 1$ (el hogar resultó ser un suscriptor) y por $\hat{p}_2 = 0\ \%$ si $y_1 = 0$ (el hogar resultó no ser suscriptor). Es claro que y_1 no es *tan buen* estimador de p como la proporción muestral, pues da origen tan sólo a dos valores posibles extremos, $\hat{p}_2 = 100\ \%$ y $\hat{p}_2 = 0\ \%$. Sin embargo, es un estimador insesgado, pues por ser y_1 una variable $B(1,p)$, tenemos: $E(\hat{p}) = E(y_1) = p$. Este ejemplo ilustra que la ausencia de sesgo, por sí sola, no garantiza que un estimador sea aceptable.

Generalmente vamos a encontrar una situación como la anterior: Para resolver un problema de estimación, disponemos de varios estimadores insesgados, alguno de los cuales debería ser preferible a los demás. Sólo si incorporamos criterios adicionales además del de ser insesgado, podremos seleccionar de entre todos los estimadores insesgados, alguno como preferido.

Por otra parte, hemos de tener presente que determinar que un estimador es sesgado puede conducirnos a un estimador insesgado: supongamos que $\hat{\theta}$ es un estimador sesgado de θ: $E(\hat{\theta}) = \theta + a$, y que podemos cuantificar la magnitud del *sesgo a*; entonces, el nuevo estimador $\hat{\theta}' = \hat{\theta} - a$ es insesgado. En otras ocasiones, podemos cuantificar que: $E(\hat{\theta}) = k\theta$ con $k \neq 1$. En tal caso, el estimador $\hat{\theta}' = (1/k)\hat{\theta}$ es insesgado. Ello significa que, conocer el sesgo de un estimador no debe hacernos rechazarlo, puesto que nos puede permitir obtener otro estimador que sea, a diferencia del primero, insesgado.

Un caso de especial importancia es el de la estimación de la esperanza matemática de la población de la que se extrajo una muestra aleatoria simple. En el caso en que la población sea $N(\mu,\sigma^2)$, ya sabemos [Teorema 8.7] que la media muestral sigue una distribución $N(\mu,\sigma^2/n)$, de modo que: $E(\bar{x}) = \mu$, por lo que la media muestral es un estimador insesgado de la esperanza poblacional, μ. En realidad, éste no es un resultado específico de la distribución Normal. Al contrario, el siguiente teorema prueba que es válido con total generalidad.

Teorema 9.1. Para estimar la esperanza matemática μ de una distribución de probabilidad, la media muestral $\bar{x}$ de una muestra aleatoria simple es siempre un estimador insesgado.

Demostración. En efecto, por definición de *esperanza matemática* de una distribución de probabilidad, tenemos que, para cualquier elemento muestral: $E(x_i) = \mu$. Por tanto:

$$E(\bar{x}) = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}n\mu = \mu$$

resultado que ya conocemos de la Sección 8.2.

Teorema 9.2. Cuando se conoce el verdadero valor de la esperanza matemática, el estadístico: $\dot{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$, calculado en una muestra aleatoria simple, es un estimador insesgado de la varianza poblacional σ^2 en cualquier distribución de probabilidad.

Demostración.

$$E(\dot{\sigma}^2) = E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n}E(x_i - \mu)^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n}n\sigma^2 = \sigma^2$$

■

Esta proposición, interesante desde el punto de vista teórico apenas tiene, sin embargo, relevancia práctica, por cuanto en el trabajo aplicado se desconoce el valor numérico de la esperanza matemática μ. Por tanto, no es posible obtener el valor muestral del estadístico $\dot{\sigma}^2$. Debe notarse que este estadístico *no es* la varianza muestral, pues aparece μ, no la media muestral, restando de cada observación x_i. El resultado siguiente nos proporciona un resultado de mayor utilidad práctica.

Teorema 9.3. La varianza muestral es un estimador sesgado de la varianza poblacional. La cuasivarianza muestral es un estimador insesgado de la varianza poblacional.

Demostración. En primer lugar, si extraemos una muestra aleatoria simple de una población con esperanza matemática μ y varianza σ^2, ambas finitas, se tiene:

$$E(x_i^2) = \sigma^2 + \mu^2; \qquad E(\bar{x}^2) = Var(\bar{x}) + [E(\bar{x})]^2 = \frac{\sigma^2}{n} + \mu^2$$

por lo que la *cuasivarianza muestral* $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ tiene esperanza matemática:

$$E(s^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)\right] =$$

$$= \frac{1}{n-1}E\left[\sum_{i=1}^{n}x_i^2 - 2\bar{x}\sum_{i=1}^{n}x_i + n\bar{x}^2\right] =$$

$$= \frac{1}{n-1}E\left[\sum_{i=1}^{n}x_i^2 - 2\bar{x}(n\bar{x}) + n\bar{x}^2\right] = \frac{1}{n-1}\sum_{i=1}^{n}E(x_i^2) - nE(\bar{x}^2) =$$

$$= \frac{1}{n-1}\left[n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right] = \frac{1}{n-1}(n\sigma^2 - \sigma^2) = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$$

■

Que la varianza muestral es un estimador sesgado de la varianza poblacional se deduce de la relación entre varianza y cuasivarianza muestrales. En efecto:

$$E(S^2) = E\left[\frac{n-1}{n} s^2\right] = \frac{n-1}{n} E(s^2) = \frac{n-1}{n} \sigma^2 \neq \sigma^2$$

Estos resultados son de suma utilidad, precisamente por ser válidos para cualquier distribución de probabilidad con momentos finitos: la media muestral es siempre un estimador insesgado del parámetro poblacional básico, la esperanza matemática, mientras que la cuasivarianza muestral es un estimador insesgado de la varianza poblacional.

Disponemos, por tanto, de estimadores *insesgados* de la esperanza matemática y varianza poblacionales de cualquier distribución de probabilidad, pero ello no excluye la existencia de otros estimadores asimismo insesgados [como enseguida veremos en algunos ejemplos]. Se hace preciso introducir criterios adicionales para seleccionar un estimador. Entre los estimadores insesgados, aquél de menor varianza tendrá mayor masa de probabilidad en un determinado entorno del parámetro desconocido θ. Este es el concepto de *eficiencia* de un estimador y parece un criterio de selección razonable. No podemos excluir tampoco el uso de un estimador sesgado, pues si su varianza es reducida, su distribución tendrá una masa importante de probabilidad en la proximidad de θ [*Menor Error Cuadrático Medio*]. Por último, puede preferirse un estimador que en una sola muestra se aproxime en un cierto sentido al parámetro desconocido [*consistencia*].

9.2.2. Eficiencia de un estimador

Habiendo comentado acerca de la limitación que supone trabajar con una sola muestra, pasamos ahora a caracterizar una propiedad de un estimador que, de cumplirse, proporciona ciertas garantías de que su valor numérico, incluso en una sola muestra, no se alejará *demasiado* del verdadero valor del parámetro que se pretende estimar. Denominaremos *eficiente* a un estimador insesgado que tenga una varianza mínima, en un cierto sentido que concretaremos a continuación. Teniendo la menor varianza posible, que su valor numérico se aleje del centro de su distribución que, por ser insesgado, coincide con el verdadero valor (desconocido) del parámetro que se pretende estimar, es un suceso de probabilidad reducida. Al menos, podremos decir que es el estimador para el que dicho suceso ocurre con una menor probabilidad.

Definición 9.2. Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados de un parámetro desconocido θ. Decimos que $\hat{\theta}_1$ es *más eficiente* que $\hat{\theta}_2$, si: $Var(\hat{\theta}_2) > Var(\hat{\theta}_1)$. La *eficiencia relativa* de $\hat{\theta}_1$ respecto de $\hat{\theta}_2$ en una muestra se mide por el ratio: $Var(\hat{\theta}_2)/Var(\hat{\theta}_1)$.

En el caso de los dos estimadores $\hat{p}_1$ y $\hat{p}_2$ que propusimos en el sondeo de audiencia televisiva, tenemos:

$$Var(\hat{p}_1) = Var\left[\frac{1}{1275}\sum_{i=1}^{1275} y_i\right] = \frac{1}{(1275)^2} Var\left(\sum_{i=1}^{1275} y_i\right) = \frac{1}{(1275)^2}\sum_{i=1}^{1275} Var(y_i) =$$

$$= \frac{1}{(1275)^2}\sum_{i=1}^{1275} p(1-p) = \frac{(1275)\,p(1-p)}{(1275)^2} = \frac{p(1-p)}{1275}$$

mientras que:

$$Var(\hat{p}_2) = Var(y_1) = p(1-p)$$

donde hemos utilizado el hecho de que la varianza de una variable $B(1,p)$ es igual a $p(1-p)$, así como la propiedad de que la varianza de una suma de variables aleatorias independientes es igual a la suma de sus respectivas varianzas.

Vemos que la eficiencia de la proporción poblacional, en relación a la del primer elemento muestral, utilizado como estimador, es: $Var(\hat{p}_1)/Var(\hat{p}_2) = 1.275$ en esta muestra. El primer estimador es preferido al segundo por ser relativamente más eficiente (es más de 1.000 veces más eficiente).

Que dos estimadores insesgados $\hat{\theta}_1$ y $\hat{\theta}_2$, con $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$, puedan ordenarse de acuerdo con su eficiencia relativa implica que el primero es preferible al segundo, pero no que sea un estimador *suficientemente* bueno, pues podría encontrarse otro estimador θ_3 con una varianza inferior a $Var(\theta_1)$. Así, la cuestión natural cuando tratamos de estimar un parámetro poblacional θ de una determinada familia de distribuciones es: ¿existirá un estimador insesgado que sea relativamente más eficiente que cualquier otro estimador insesgado que podamos considerar? Contestaremos enseguida a esta cuestión, pero antes necesitamos introducir un concepto importante, el de cota mínima para la varianza de un estimador.

Dada una muestra $\mathbf{X} = \{x_1, x_2, ..., x_n\}$, extraída de una población con función de probabilidad $f(x;\theta)$, denotemos por $L(\mathbf{X}) = L(\{x_1, x_2, ..., x_n, \theta\})$ o $L(\mathbf{X};\theta)$ la *función de probabilidad muestral conjunta*. Esta función admite una doble interpretación, que le confiere una naturaleza u otra. Si consideramos unos valores paramétricos θ, desconocidos o no, tenemos la *probabilidad de extraer una muestra*, fijados unos determinados valores paramétricos, y podríamos comparar las probabilidades de dos muestras diferentes, para tales valores paramétricos. Dichas probabilidades serán diferentes si cambiamos los valores de θ que hemos fijado. Alternativamente, si tomamos los elementos muestrales como fijos, y consideramos todos los valores posibles de los parámetros, tenemos la *función de verosimilitud* que, dada una muestra, va variando según vamos cambiando el valor de θ. La función de verosimilitud es una función de los valores paramétricos θ.

Definición 9.3. Dada una muestra $\mathbf{X} = \{x_1, x_2, ..., x_n\}$, extraída de una población con función de probabilidad $f(x,\theta)$, la *función de verosimilitud* $L(\theta/\mathbf{X}) = L(\theta/x_1, x_2, ..., x_n)$ es la función que se obtiene al fijar los valores muestrales $\{x_1, x_2, ..., x_n\}$ y variar θ:

$$L(\theta/\mathbf{X}) = L(\theta/x_1, x_2, ..., x_n) = P(\theta/X = x_1, X = x_2, ..., X = x_n)$$

Teorema 9.4. Sea $\hat{\theta}$ un estimador de un parámetro poblacional θ desconocido, con sesgo igual a $b(\theta)$, es decir: $E(\hat{\theta}) = \theta + b(\theta)$. La varianza de θ satisface:

$$Var(\hat{\theta}) \geqslant \frac{[1 + b'(\theta)]}{E\left(\dfrac{\partial \ln L(\mathbf{X};\theta)}{\partial\theta}\right)^2} \tag{9.1}$$

La demostración es compleja, por lo que no entramos en su discusión.

Cuando la muestra es aleatoria simple, entonces la función de probabilidad muestral es, por la independencia en el muestreo, igual al producto de los valores de la función de densidad teórica en cada elemento muestral, pero todos ellos tienen igual distribución, de modo que:

$$E\left(\frac{\partial \ln L(\mathbf{X};\theta)}{\partial\theta}\right)^2 = E\left[\frac{\partial}{\partial\theta}\ln\left(\prod_{i=1}^{n} f(x_i;\theta)\right)\right]^2 = E\left(\frac{\partial}{\partial\theta}\ln [f(x;\theta)]^n\right)^2 = nE\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2$$

En consecuencia, se tiene para cualquier estimador $\hat{\theta}$ que:

$$Var(\hat{\theta}) \geqslant \frac{[1 + b'(\theta)]}{nE\left(\dfrac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2} \tag{9.1}$$

La cota (9.1) es válida para cualquier estimador obtenido en cualquier tipo de muestra. Sin embargo, en un tipo de muestreo genérico, obtener la expresión de la función de verosimilitud y su esperanza matemática puede ser sumamente complejo. Cuando el muestreo es aleatorio simple, la cota mínima para la varianza se simplifica a la expresión (9.2).

El siguiente teorema, que es un corolario del anterior, proporciona uno de los resultados más importantes en inferencia estadística, pues proporciona la menor varianza posible de un estimador insesgado. Si un estimador *insesgado* tiene una varianza igual a la proporcionada por este teorema, podremos estar seguros de que no existe otro estimador insesgado con una varianza menor. Diremos entonces que es un *estimador eficiente.*

Teorema 9.5. (Desigualdad de Cramer-Rao). Sea $\mathbf{X} = \{x_1, x_2,..., x_n\}$ una *muestra aleatoria simple* de una distribución de probabilidad poblacional representada por la función de densidad $f(x;\theta)$, que tiene derivadas primeras continuas, excepto en un número finito de puntos. Supongamos asimismo que el conjunto de puntos en que $f(x;\theta) \neq 0$ no depende de θ. Sea $\theta = h(\mathbf{X}) = h(\{x_1, x_2,..., x_n\})$ un estimador *insesgado* del parámetro θ. Entonces se tiene:

$$Var(\hat{\theta}) \geqslant \frac{1}{n\displaystyle\int_{-\infty}^{\infty}\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2 f(x;\theta)\,dx} = \frac{1}{n\,E\left(\dfrac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2}$$

Si la función de densidad tiene derivadas segundas contínuas, entonces se tiene:

$$E\left(\frac{\partial \ln f(x;\theta)}{\partial \theta}\right)^2 = -E\left(\frac{\partial^2 \ln f(x;\theta)}{\partial \theta^2}\right)$$

por lo que la cota de Cramer-Rao puede también escribirse:

$$Var(\theta) \geqslant \frac{1}{-n\displaystyle\int_{-\infty}^{\infty} \frac{\partial^2 \ln f(x;\theta)}{\partial \theta^2} f(x;\theta)\, dx} = \frac{1}{-n\, E\left(\dfrac{\partial^2 \ln f(x;\theta)}{\partial \theta^2}\right)}$$

siendo en unas ocasiones más sencillo calcular una que otra, y viceversa.

En estimadores insesgados, la cota de Cramer-Rao depende tan sólo de dos factores: 1) el tamaño muestral n, y 2) la densidad de probabilidad de la familia con la que se trabaja. La cota afirma que hay una cantidad mínima de imprecisión (varianza) asociada a trabajar con una cierta cantidad de información (tamaño muestral) procedente de una determinada familia de distribuciones de probabilidad.

Ejemplo 9.2. Si tenemos una muestra aleatoria simple de una población de Bernouilli, tendremos elementos muestrales que son, 0 o 1. La función de probabilidad de cada uno de ellos es:

$$f(y_i/p) = p^{y_i}(1-p)^{1-y_i} \qquad y_i = 0,1 \qquad , \qquad 0 < p < 1$$

Ya hemos probado [Ejemplo 9.1] que la proporción muestral, que coincide con la media muestral en este caso, es un estimador insesgado de la proporción poblacional p. Su varianza es:

$$Var(p) = Var\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(y_i) =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} p(1-p) = \frac{n\,p(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Calculemos ahora el valor mínimo que puede tomar la varianza de un estimador insesgado $\hat{p}$ según aparece en el teorema anterior. En primer lugar:

$$\ln f(y_i/p) = y_i \ln(p) + (1 - y_i)\ln(1-p)$$

de modo que:

$$\frac{\partial \ln f(y_i;p)}{\partial p} = \frac{y_i}{p} - \frac{1-y_i}{1-p}$$

$$\frac{\partial^2 \ln f(y_i;p)}{\partial p^2} = -\frac{y_i}{p^2} - \frac{1-y_i}{(1-p)^2}$$

y tomando esperanzas en la segunda derivada:

$$E\left[\frac{\partial^2 \ln f(y_i;p)}{\partial p^2}\right] = -\frac{p}{p^2} - \frac{1-p}{(1-p)^2} = -\frac{1}{p(1-p)}$$

de modo que, de acuerdo con el teorema, se tiene, para cualquier estimador insesgado $\dot{p}$:

$$Var(\dot{p}) \geqslant \frac{p(1-p)}{n}$$

siendo el término de la derecha, por tanto, la menor varianza que posiblemente puede tomar un estimador insesgado de p.

Ahora bien, acabamos de probar que la varianza de la proporción muestral, estimador insesgado de la proporción poblacional p es, precisamente: $p(1-p)/n$, de modo que este estimador tiene la mínima varianza posible entre todos los estimadores insesgados.

Ejemplo 9.3. El logaritmo de la función de densidad $N(\mu;\sigma^2)$ es:

$$\ln f(x;\mu,\sigma^2) = -\frac{1}{2}\ln\sigma^2 - \ln(\sqrt{2\pi}) - \frac{1}{2\sigma^2}(x-\mu)^2$$

a partir de la cual, tenemos:

$$\frac{\partial \ln f(x;\mu,\sigma^2)}{\partial \mu} = \frac{x-\mu}{\sigma^2}$$

y, por tanto:

$$E\left(\frac{\partial \ln f(x;\mu,\sigma^2)}{\partial \mu}\right)^2 = E\left(\frac{(x-\mu)^2}{(\sigma^2)^2}\right) = \frac{Var(x)}{(\sigma^2)^2} = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$$

y la cota de Cramer-Rao resulta ser:

$$\frac{1}{nE\left(\frac{\partial \ln f(x;\mu,\sigma^2)}{\partial \theta}\right)^2} = \frac{1}{n\frac{1}{\sigma^2}} = \frac{\sigma^2}{n}$$

pero ésta es, precisamente, la varianza de la media muestral, que es un estimador insesgado de la esperanza matemática μ.

Definición 9.4. Un estimador insesgado $\hat{\theta}$ de un parámetro poblacional desconocido θ se dice *eficiente* si su varianza es tan baja como la cota de Cramer-Rao:

$$Var(\hat{\theta}) = \frac{1}{nE\left(\dfrac{\partial \ln f(x_i;\theta)}{\partial \theta}\right)^2} = \frac{1}{-nE\left(\dfrac{\partial^2 \ln f(x;\theta)}{\partial \theta^2}\right)}$$

La *eficiencia* de un estimador insesgado $\hat{\theta}$ se define como:

$$\text{Eficiencia}(\hat{\theta}) = \frac{\dfrac{1}{nE\left(\dfrac{\partial \ln f(x_i;\theta)}{\partial \theta}\right)^2}}{Var(\hat{\theta})}$$

siendo siempre inferior a 1.

Como consecuencia del análisis que efectuamos en el Ejemplo 9.2, *la proporción muestral es un estimador eficiente de la proporción poblacional.* Asimismo, por el Ejemplo 9.3, la media muestral es un estimador *eficiente* de la esperanza matemática en una población Normal.

R. A. Fisher, un gran estadístico contemporáneo, propuso utilizar como indicador de la información que una muestra contiene respecto a un parámetro, la función: $\partial \ln L(\mathbf{X};\theta)/\partial\theta$, que refleja el modo en que varía la función de probabilidad conjunta muestral al variar los valores de dicho parámetro. Puede probarse que este indicador, que es una variable aleatoria, tiene esperanza matemática igual a cero. Si, cambiando el valor de θ no varía la función de probabilidad, entonces diríamos con toda justificación que la muestra no contiene ninguna información acerca de θ. Para medir la cantidad de información, Fisher sugirió utilizar la varianza del indicador propuesto, llamándose *información de Fisher* $I(\theta)$:

$$I(\theta) = Var\left(\frac{\partial \ln L(\mathbf{X};\theta)}{\partial \theta}\right) = E\left(\frac{\partial \ln L(\mathbf{X};\theta)}{\partial \theta}\right)^2$$

donde la igualdad proviene del hecho que la derivada parcial tiene esperanza nula.

Puede probarse que:

$$E\left(\frac{\partial \ln L(\mathbf{X};\theta)}{\partial \theta}\right)^2 = -E\left(\frac{\partial^2 \ln L(\mathbf{X};\theta)}{\partial \theta^2}\right)$$

y, con una muestra aleatoria simple:

$$E\left(\frac{\partial \ln L(\mathbf{X};\theta)}{\partial \theta}\right)^2 = nE\left(\frac{\partial \ln f(x;\theta)}{\partial \theta}\right)^2$$

de modo que:

$$I(\theta) = -E\left(\frac{\partial^2 \ln L(\mathbf{X};\theta)}{\partial\theta^2}\right) = nE\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2 = -nE\left(\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right)$$

Para que estos indicadores tengan plena justificación, es preciso que se satisfagan unas propiedades de regularidad:

1) el soporte de la distribución de probabilidad no depende del parámetro θ que se estima.
2) existen las dos primeras derivadas de la función de probabilidad muestral respecto de θ.
3) pueden intercambiarse las propiedades de derivación e integración (en el caso continuo), o derivación y suma (en el caso discreto).

Si no se cumplen estas condiciones, entonces no se tienen las equivalencias que estamos usando entre la esperanza del cuadrado de la derivada y la esperanza del hessiano de la función de probabilidad muestral, de modo que todo lo anterior deja de ser válido.

Ejemplo 9.4. En una muestra aleatoria simple de una población $N(\mu;\sigma^2)$, la cantidad de información acerca del parámetro μ es $I(\mu) = n/\sigma^2$ [ver Ejercicio 9.10], que aumenta con el tamaño muestral, y es menor cuanto mayor sea la varianza de la población. Ambos efectos coinciden con la idea intuitiva de contenido informativo de una muestra.

Podemos ahora volver a examinar el concepto de eficiencia: el denominador de la cota de Cramer-Rao es, precisamente, la *información de Fisher*. La cota afirma, por tanto, que cuanto menor sea la información, mayor será la varianza y, con ella, la imprecisión del estimador insesgado de mínima varianza y recíprocamente, cuanto mayor sea la información de Fisher, menor será la imprecisión asociada al estimador insesgado de mínima varianza.

Definición 9.5. Dado el conjunto U formado por todos los estimadores insesgados de un parámetro poblacional θ, se dice que el estimador insesgado $\hat{\theta}$ es *el estimador insesgado de mínima varianza* si se tiene:

$$Var(\hat{\theta}) \leqslant Var(\hat{\theta}^*) \qquad \text{para todo } \hat{\theta}^* \text{ en } U$$

De modo análogo puede definirse el estimador de mínima varianza dentro de una clase cualquiera de estimadores. La familia de los estimadores insesgados es especialmente interesante, y es la única que consideramos aquí.

Hay que hacer notar que los conceptos de varianza mínima y eficiencia no son equivalentes. Si la varianza de un estimador insesgado alcanza la cota de Cramer-Rao, entonces necesariamente dicho estimador será de mínima varianza dentro de la clase de estimadores insesgados. Ya hemos probado que la proporción muestral es un estimador eficiente de la proporción poblacional, y es también, por tanto, un estimador insesgado de varianza mínima de p. Sin embargo, el recíproco no es cierto: existen casos en que ningún estimador insesgado alcanza la cota de Cramer-Rao y, por tanto, ninguno es eficiente o de mínima varianza. Sin embargo, generalmente existe en dicha situación, un estimador insesgado que, aun teniendo varianza superior a la cota de Cramer-Rao, es igual o inferior a la de cualquier otro estimador insesgado, por lo que dicho estimador es de mínima varianza, dentro de la clase de los estimadores insesgados.

9.2.3. Error Cuadrático Medio de un estimador

Ya hemos comentado que, en todo rigor, la ausencia de sesgo en un estimador sólo significa que, cuando éste se calcula, utilizando diversas muestras, su promedio tiende a estar próximo al verdadero valor del parámetro desconocido. Sin embargo, calculado a partir de una sola muestra, como es la situación habitual en el trabajo práctico de inferencia estadística en Economía, su valor numérico puede estar cerca o lejos del verdadero valor de θ. La desviación típica del estimador, es decir, la raíz cuadrada de su varianza, es la única indicación acerca de la distancia media a la que esperamos que esté cada estimación del verdadero valor de θ. En definitiva, cuando se dispone de una sola muestra, no es obvio que la ausencia de sesgo sea una propiedad crucial.

Analicemos el problema de esta manera: la cota de Cramer-Rao que antes vimos nos proporciona la menor varianza posible de un estimador insesgado de θ. Ahora bien, es perfectamente posible que dispongamos de un estimador *sesgado* de θ, pero con una varianza inferior a la cota de Cramer-Rao. ¿Es evidente que debamos preferir el estimador insesgado, con mayor varianza, que el estimador sesgado, que tiene una varianza inferior?

Salvo que tuviésemos múltiples muestras en que calcular valores numéricos de $\hat{\theta}$ para luego promediarlos, la respuesta es que no es evidente que debamos preferir el estimador insesgado. En definitiva, como el mismo interrogante sugiere, se trata de ponderar conjuntamente el posible sesgo con la varianza, de modo que, englobando ambos en una sola medida, pueda esta aplicarse a todo tipo de estimador, para escoger uno como preferido a otro.

Definimos el *Error Cuadrático Medio* de un estimador, y lo denotamos por $ECM(\hat{\theta})$ o, simplemente ECM:

$$ECM(\hat{\theta}) = E\,[(\hat{\theta} - \theta)^2]$$

Teorema 9.6. *a*) El *ECM* de todo estimador es igual a su sesgo al cuadrado, más su varianza. *b*) Si un estimador es insesgado, entonces su *ECM* coincide con su varianza.

Demostración. *a*) en primer lugar:

$$ECM(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = E[[(\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \theta)]^2] =$$

$$= E(\hat{\theta} - E(\hat{\theta}))^2 + E(E(\hat{\theta}) - \theta)^2 + 2E[(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)] =$$

$$= Var(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2 + 2\,(E(\hat{\theta}) - \theta)\,E[\hat{\theta} - E(\hat{\theta})] = Var(\hat{\theta}) + [Sesgo(\hat{\theta})]^2$$

donde hemos utilizado que: $E[\hat{\theta} - E(\hat{\theta})] = E(\hat{\theta}) - E(\hat{\theta}) = 0$, así como la definición de *sesgo*, como diferencia entre la esperanza del estimador y el verdadero valor (desconocido) del parámetro que se pretende estimar.

b) si un estimador es insesgado, su sesgo es cero, por lo que el resultado se sigue a partir de la expresión anterior para el *ECM* de todo estimador. ■

Si un estimador es sesgado, su promedio, supuesto que pudiésemos calcularlo a través de todas las muestras posibles, *no coincidiría* con el verdadero valor θ del parámetro desconocido. Si embargo, si el sesgo es pequeño, la distancia entre tal promedio y θ sería reducida. Si además, el estimador tiene varianza pequeña, entonces su valor en cada muestra no diferirá mucho de su valor esperado que, a su vez, no difiere mucho de θ. Ello puede hacer que podamos preferir en un momento determinado un estimador sesgado a un estimador insesgado, que tiene mayor varianza que el primero. Un investigador podría penalizar más el sesgo que la varianza de un estimador, definiendo una versión modificada del *ECM*, que introdujera una ponderación relativa.

9.2.4. Estimador consistente

Estamos tomando como propiedad básica de un estimador el que no difiera sustancialmente del parámetro que pretende estimar. Un estimador insesgado satisface esa propiedad sólo cuando se poseen muchas realizaciones (estimaciones) del mismo que puedan promediarse. El promedio estaría próximo al verdadero valor del parámetro. La eficiencia, por otra parte, nos dice que la distancia media de cada estimación al verdadero valor del parámetro es la menor que podemos lograr, si bien no nos garantiza que dicha distancia sea pequeña. Examinamos ahora una propiedad de enorme interés cuando, como ocurre en Economía, disponemos tan sólo de una muestra. Si un estimador es *consistente*, la estimación que genera tenderá, en un sentido que vamos a especificar, al verdadero valor del parámetro cuando el tamaño de la muestra crece sin límite. La *consistencia* es muy relevante porque, a diferencia del insesgo, es una propiedad de una muestra. Si un estimador es consistente, basta con disponer de una muestra para poder afirmar que no nos desviamos *mucho* del parámetro que se estima. Ahora bien, para ello necesitamos que la muestra sea *suficientemente* grande.

Es preciso tener en cuenta que la distribución de probabilidad de un estimador generalmente depende del tamaño muestral aunque sólo sea a través de sus momentos. Un estimador, calculado para distintos tamaños muestrales, puede verse como una sucesión de variables aleatorias.

Definición 9.6. Un estimador es *asintóticamente insesgado* si su posible sesgo tiende a cero al aumentar el tamaño muestral con que se calcula. Evidentemente, todo estimador insesgado es asintóticamente insesgado, puesto que su sesgo es cero para cualquier tamaño muestral.

Ejemplo 9.5. Ya vimos que la varianza muestral es un estimador sesgado de la varianza poblacional, por cuanto que:

$$E(S^2) = \frac{n-1}{n}\sigma^2$$

Sin embargo, cuando n tiende a ∞, entonces el cociente $(n-1)/n$ tiende a 1, y el sesgo de la varianza muestral tiende a desaparecer, por cuanto que $E(S^2)$ converge a σ^2. Por tanto, *la varianza muestral es un estimador asintóticamente insesgado de la varianza poblacional.*

Definición 9.7. Un estimador $\hat{\theta}_n$ es *consistente* si converge en probabilidad al verdadero valor numérico (desconocido) del parámetro que se pretende estimar, es decir, si:

$$\operatorname*{plím}_n \hat{\theta}_n = \theta$$

donde el subíndice hace referencia al tamaño muestral utilizado. Utilizando la definición de convergencia en probabilidad, podemos decir que $\hat{\theta}_n$ es consistente si:

$$\lim_n P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$$

o también, si: para todo $\varepsilon > 0$ y todo $\delta > 0$ existe un tamaño muestral $n_0(\varepsilon,\delta)$ tal que:

$$n > n_0(\varepsilon,\delta) \Rightarrow P(|\hat{\theta}_n - \theta| < \varepsilon) > 1 - \delta$$

Ejemplo 9.6. Ya sabemos que un estimador razonable del parámetro p de una distribución de Bernouilli,a partir de una muestra de tamaño n, es la proporción muestral: $\dot{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$. También hemos visto que la esperanza matemática de $\dot{p}$ es p, y su varianza es: $Var(\dot{p}) = p(1-p)/n$. De este modo, si aplicamos la desigualdad de Chebychev, tenemos:

$$P(|\dot{p} - p| < \varepsilon) > 1 - \frac{Var(\dot{p})}{\varepsilon^2} = 1 - \frac{p(1-p)}{n\varepsilon^2}$$

Ahora bien, el máximo valor numérico que puede tomar el producto $p(1-p)$ es $1/4$, cuando $p = 1/2$. Por tanto, tenemos:

$$P(|\dot{p} - p| < \varepsilon) > 1 - \frac{1}{4n\varepsilon^2}$$

Dado un número $\delta > 0$, si tomamos: $n_0 = \frac{1}{4\varepsilon^2\delta}$ tenemos:

$$n > n_0(\varepsilon,\delta) \Rightarrow P(|\dot{p} - p| < \varepsilon) > 1 - \frac{1}{4\varepsilon^2 n} > 1 - \frac{1}{4\varepsilon^2}4\varepsilon^2\delta = 1 - \delta$$

por lo que la proporción poblacional es un estimador consistente del parámetro p en una población de Bernouilli.

Teorema 9.7. La media muestral es siempre un estimador consistente de la esperanza matemática.

Demostración. Tomando límites en la desigualdad de Chebychev:

$$\lim_{n\to\infty} P(|\bar{x} - \mu| \geqslant \varepsilon) \leqslant \lim_{n\to\infty} \frac{E(\bar{x} - \mu)^2}{\varepsilon^2} = \lim_{n\to\infty} \frac{\sigma^2/n}{\varepsilon^2} = 0$$

pues ε y σ son constantes. ∎

Teorema 9.8. Un estimador $\hat{\theta}$ *asintóticamente insesgado*, cuya varianza tiende a cero al aumentar el tamaño muestral con que se calcula, es *consistente*.

Demostración. [Ver ejercicio 9.3.2]. ∎

Las propiedades del límite en probabilidad garantizan el siguiente teorema, muy importante para la construcción de estimadores consistentes.

Teorema 9.9. Si $\hat{\theta}$ es una estimador consistente de θ y g es una función continua, entonces $g(\hat{\theta})$ es un estimador consistente de $g(\theta)$.

Por tanto, si $\hat{\theta}$ es un estimador consistente de θ, entonces $\sqrt{\hat{\theta}}$ es un estimador consistente de $\sqrt{\theta}$, $\ln(\hat{\theta})$ es un estimador consistente de $\ln(\theta)$, $\hat{\theta}$ lo es de θ, $\hat{\theta}^2$ es estimador consistente de θ^2, y $3\hat{\theta} - 5$ lo es de $3\theta - 5$.

9.2.5. Estimador suficiente

Definición 9.8. Dada una muestra $\mathbf{X} = \{x_1, x_2, ..., x_n\}$, sea $u(x_1, x_2, ..., x_n)$ un estadístico, y supongamos que la función de probabilidad conjunta muestral puede descomponerse como el producto:

$$L(x_1, x_2, ..., x_n; \theta) = \phi[u(x_1, x_2, ..., x_n); \theta]\, h(x_1, x_2 ..., x_n)$$

entonces $u(x_1, x_2, ..., x_n)$ es un *estadístico suficiente*. Un *estimador suficiente* de θ es un estimador que es estadístico suficiente.

Nótese que ha de ser posible escribir la función de probabilidad muestral como producto de dos factores (cada factor puede englobar varios). En el único de ellos en que aparezca el parámetro que se pretende estimar, la información muestral sólo debe entrar a través del estadístico suficiente. En el otro factor, la información muestral puede aparecer en la forma de cualquier función, pero no puede aparecer el parámetro a estimar.

Si extraemos una muestra de tamaño n de una población de Bernouilli $B(1,p)$: $\mathbf{X} = (x_1, x_2, ..., x_n)$, en la que cada elemento muestral x_i es igual a 0 o 1, la función de verosimilitud es:

$$L(p/x_1, x_2, ..., x_n) = \prod_{i=1}^{n} p^{x_i}(1 - p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1 - p)^{n-\sum_{i=1}^{n} x_i}$$

que, considerado como un solo factor, muestra que la suma de los elementos muestrales es un estadístico suficiente. Un estadístico suficiente resume *toda la información muestral* para la estimación del parámetro de interés, lo que queda patente en el hecho de que la información de Fisher, por ser la derivada segunda del logaritmo de la función de probabilidad muestral con respecto al parámetro, sólo depende del factor en que está el parámetro y, por tanto, sólo depende de la muestra a través del estadístico suficiente:

$$\ln L(x_1, x_2, ..., x_n; \theta) = \ln(\phi[u(x_1, x_2, ..., x_n); \theta]) + \ln(h(x_1, x_2 ..., x_n))$$

que implica:

$$\frac{\partial \ln L(x_1, x_2, ..., x_n; \theta)}{\partial \theta} = \frac{\partial \ln(\phi[u(x_1, x_2, ..., x_n); \theta])}{\partial \theta} = g(u(x_1, x_2 ..., x_n); \theta)$$

y lo mismo ocurre con la segunda derivada.

Evidentemente, la *suficiencia* de un estadístico tiene implicaciones importantes en cuanto a delimitar el tipo de información muestral que es relevante para la estimación de θ, como veremos a continuación. En la muestra del ejemplo anterior de la distribución de Bernouilli, una vez que conocemos el número total de unos y ceros, es decir, la suma de los elementos muestrales, conocer su ordenación en la muestra es irrelevante. Tan sólo conocer su suma es útil.

Una función biunívoca de un estadístico suficiente que no haga intervenir la información muestral, es asimismo un estadístico suficiente. El caso más útil en el trabajo aplicado es que si la suma de los elementos muestrales es un estadístico suficiente, también lo es su cociente por n, la media muestral.

9.3. PROCEDIMIENTOS DE ESTIMACIÓN

Habiendo definido las principales propiedades que puede tener un estimador, queda por resolver una cuestión de suma importancia: ¿cómo obtenemos estimadores? y especialmente: ¿cómo obtenemos estimadores que tengan buenas propiedades? Sería especialmente interesante si pudiésemos disponer de algún procedimiento de obtención de estimadores que fuera aplicable a una amplia gama de problemas, y que garantizase además, unas buenas propiedades para los mismos. Ese es el tema que tratamos en esta sección.

9.3.1. El método de máxima verosimilitud

Si la muestra es aleatoria simple, entonces cada extracción es independiente de las demás, por lo que la probabilidad del suceso compuesto anterior es igual al producto de las probabilidades de los sucesos que lo componen, de modo que podemos escribir la función de verosimilitud:

$$L(\theta/\mathbf{X}) = L(\theta/x_1, x_2, ..., x_n) = \prod_{i=1}^{n} P(X_i = x_i; \theta)$$

Un criterio posible de estimación se basa en asignar a los parámetros desconocidos θ precisamente aquellos valores que asignan el mayor valor a la función de verosimilitud $L(\theta/\mathbf{X})$.

Esta estrategia se basa en asignar el valor *más verosímil* a los parámetros, *dadas las observaciones muestrales*. Se trata, por tanto, de maximizar la función $L(\{\theta/x_1, x_2, ..., x_n\})$ con respecto al vector θ, y obtener la expresión del estadístico que es *estimador de máxima verosimilitud*. En muchas ocasiones, es mucho más sencillo maximizar el logaritmo de la función de verosimilitud, que la propia función de verosimilitud. Es evidente que el valor máximo que alcanzan ambas funciones no es el mismo, pero el conjunto de valores paramétricos para los cuales alcanzan su máximo sí es el mismo, por ser una función una transformación monótona de la otra. Dichos valores de los parámetros, obtenidos como función de los elementos muestrales, constituyen el estimador de máxima verosimilitud.

Definición 9.9. Dada una muestra $\mathbf{X} = \{x_1, x_2, ..., x_n\}$ extraída de una distribución con función de densidad $f(x;\theta)$, el *estimador de máxima verosimilitud* es un valor numérico $\hat{\theta}_{MV}$ tal que:

$$L(\hat{\theta}_{MV}/x_1, x_2, ..., x_n) \geqslant L(\hat{\theta}/x_1, x_2, ..., x_n) \qquad \text{para todo } \hat{\theta}$$

Ejemplo 9.7. Supongamos que hemos extraído una muestra de tamaño n de una población de Bernouilli $B(1,p)$: $\mathbf{X} = (x_1, x_2, ..., x_n)$, en la que cada elemento muestral x_i es igual a 0 o 1. Sea k el número de unos en la muestra, es decir: $k = \sum_{i=1}^{n} x_i$. La función de verosimilitud es:

$$L(p/x_1, x_2, ..., x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{\sum_{i=1}^{n}(1-x_i)}$$

de modo que:

$$\ln L(p/x_1, x_2, ..., x_n) = \left(\sum_{i=1}^{n} x_i\right)\ln(p) + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-p)$$

Por tanto, el estimador de máxima verosimilitud de p se obtiene igualando a cero la derivada del logaritmo de la función de verosimilitud:

$$\frac{\partial \ln L(p/\mathbf{X})}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = 0$$

es decir:

$$\hat{p}_{MV} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

el cociente entre el número de *unos* o *éxitos* obtenidos en la muestra, y el tamaño muestral, es decir, la *proporción muestral*. Este es el *estimador de máxima verosimilitud del parámetro p*.

La derivada segunda:

$$\frac{\partial^2 \ln L(\mathbf{X};p)}{\partial p^2} = -\frac{\sum_{i=1}^{n} x_i}{p^2} - \frac{n - \sum_{i=1}^{n} x_i}{(1-p)^2}$$

que, evaluada en el estimador de máxima verosimilitud que antes hallamos, conduce a:

$$\frac{\partial^2 \ln L(\mathbf{X};\hat{p}_{MV})}{\partial p^2} = -\frac{n}{\bar{x}} - \frac{n}{1-\bar{x}} < 0$$

por lo que, efectivamente, hemos maximizado la función de verosimilitud.

Antes vimos que la suma de los elementos muestrales es, en el caso de una muestra extraída de una población Bernouilli, un estadístico suficiente. Ahora que conocemos cómo obtener estimadores de máxima verosimilitud, podemos afirmar que *el estimador de máxima verosimilitud es siempre una función tan sólo de uno o varios estadísticos suficientes*. Ello es debido a que ningún otro estadístico aparece en la función de verosimilitud acompañando al parámetro que se estima, que es con respecto a quien maximizamos la función. Al derivar respecto al parámetro e igualar a cero, el segundo factor de la definición de *estimador suficiente* desaparece, por lo que el estimador de máxima verosimilitud sólo depende del estadístico suficiente.

Ejemplo 9.8. Consideremos una variable aleatoria X, con distribución binomial $B(n,p)$, de la que hemos extraído una muestra de tamaño k con valores numéricos $x_1, x_2, ..., x_k$. Cada uno de ellos es un número entre 0 y n. Supongamos que el parámetro n es conocido, y queremos estimar p. La función de verosimilitud para este resultado muestral, es:

$$L(p/n, x_1, x_2, ..., x_k) = \binom{n}{x_1} p^{x_1}(1-p)^{n-x_1} \binom{n}{x_2} p^{x_2}(1-p)^{n-x_2} \cdots \binom{n}{x_k} p^{x_k}(1-p)^{n-x_k} =$$

$$= \prod_{i=1}^{k} \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i} = \left[\prod_{i=1}^{k} \binom{n}{x_i}\right] p^{\sum_{i=1}^{k} x_i}(1-p)^{nk - \sum_{i=1}^{k} x_i}$$

y su logaritmo:

$$\ln L(p/n, x_1, x_2, ..., x_k) = \sum_{i=1}^{k} \ln \binom{n}{x_i} + \sum_{i=1}^{k} x_i \ln(p) + (nk - \sum_{i=1}^{k} x_i) \ln(1-p)$$

en el que tan sólo los dos últimos términos dependen del parámetro p. Derivando respecto a éste dos veces, tenemos:

$$\frac{\partial \ln L(p/n, x_1, x_2, ..., x_k)}{\partial p} = \frac{\sum_{i=1}^{k} x_i}{p} - \frac{nk - \sum_{i=1}^{k} x_i}{1-p} = 0$$

$$\frac{\partial^2 \ln L(p/n, k)}{\partial p^2} = -\frac{\sum_{i=1}^{k} x_i}{p^2} - \frac{nk - \sum_{i=1}^{k} x_i}{(1-p)^2}$$

Igualando a cero la primera derivada y resolviendo, obtenemos el estimador de máxima verosimilitud de p:

$$p_{MV} = \frac{\sum_{i=1}^{k} x_i}{nk} = \frac{1}{n}\bar{x}$$

el cociente entre la suma de los elementos muestrales y el producto del parámetro n por el tamaño muestral, k. Si sustituimos en la segunda condición p por la expresión del estimador, obtenemos:

$$\frac{\partial^2 \ln L(p/n, k)}{\partial p^2} = -\frac{kn^2}{\bar{x}} - \frac{kn^2}{n - \bar{x}} < 0$$

por lo que el estimador propuesto maximiza, efectivamente, la función de verosimilitud.

Ejemplo 9.9. Sea $\mathbf{X} = (x_1, x_2, ..., x_n)$, una muestra aleatoria simple de una distribución de probabilidad Poisson (λ): $f(x/\lambda) = e^{-\lambda}\lambda^x/x!$. La función de verosimilitud es:

$$L(\lambda/x_1, x_2, ..., x_n) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\left(\lambda^{\sum_{i=1}^{n} x_i}\right)}{\prod_{i=1}^{n} x_i!}$$

de modo que:

$$\ln L(\lambda/x_1, x_2, ..., x_n) = -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\ln \lambda - \ln\left(\prod_{i=1}^{n} x_i!\right)$$

y derivando respecto de λ e igualando a cero:

$$\frac{\partial \ln L(\lambda/x_1, x_2, ..., x_n)}{\partial \lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i = 0$$

es decir:

$$\hat{\lambda}_{MV} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

La derivada segunda del logaritmo de la función de verosimilitud es:

$$\frac{\partial^2 \ln L(\lambda/x_1, x_2, ..., x_n)}{\partial \lambda^2} = -\frac{1}{\lambda^2}\sum_{i=1}^{n} x_i \leqslant 0$$

ya que la distribución de Poisson sólo toma valores enteros no negativos. Por tanto, la media muestral es el estimador de máxima verosimilitud del parámetro λ de una distribución de Poisson de la que es, como sabemos, su esperanza matemática y su varianza.

Ejemplo 9.10. Supongamos que de una función de densidad exponencial: $f(x;\theta) = (1/\theta)e^{-x/\theta}$, $x > 0$ se han extraído tres observaciones, resultando: $x_1 = 3{,}1$; $x_2 = 4{,}2$; $x_3 = 5{,}3$. La función de verosimilitud de esta muestra es:

$$L(\theta/x_1 = 3{,}1; x_2 = 4{,}2; x_3 = 5{,}3) = \prod_{i=1}^{3} \frac{1}{\theta} e^{-\frac{x_i}{\theta}} = \frac{1}{\theta^3} e^{-\frac{1}{\theta}\sum_1^3 x_i} = \frac{1}{\theta^3} e^{-\frac{12{,}6}{\theta}}$$

observando en el Gráfico 9.1 que alcanza un máximo para $\theta = 4{,}2$. Este es el valor numérico que obtendremos asimismo si maximizamos la función anterior. Ello quiere decir que dicho valor numérico del parámetro θ selecciona, de entre toda la familia de distribuciones exponenciales, aquella función de densidad específica para la cual es más probable que, al extraer una muestra de 3 elementos, obtengamos precisamente los que hemos observado: 3,1, 4,2, 5,3. En ese sentido, el valor numérico 4,2 de θ es el que *mejor explica* los datos, por ser el que hace *más probable* a la muestra concreta que hemos obtenido.

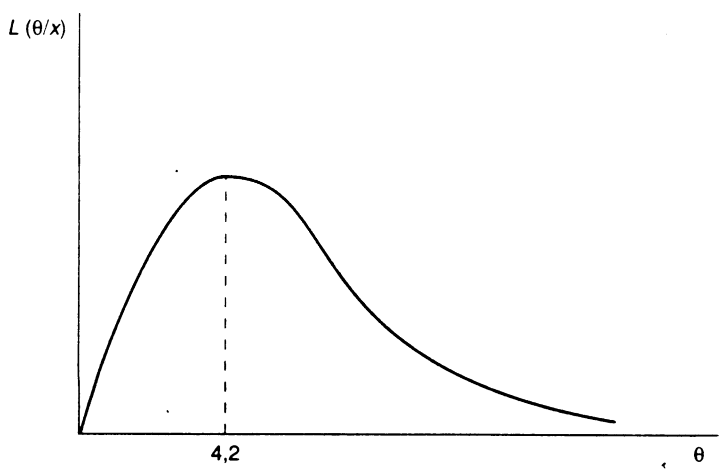

Gráfico 9.1.

9.3.1.1. *Estimación de máxima verosimilitud en una población Normal*

Cuando la función de probabilidad, discreta o continua, depende de más de un parámetro, el estimador de máxima verosimilitud se obtiene derivando la función de verosimilitud, o su logaritmo, con respecto a cada uno de dichos parámetros, e igualando a cero. Se tiene así un sistema de tantas ecuaciones como parámetros desconocidos. El hessiano, matriz de derivadas segundas, debe ser definido negativo cuando se sustituyen en él los parámetros desconocidos por sus estimadores máximo-verosímiles.

Consideremos una muestra aleatoria simple: $\mathbf{X} = (x_1, x_2, ..., x_n)$ de una población $N(\mu, \sigma^2)$. Su función de verosimilitud es:

$$L(\mu, \sigma^2 / x_1, x_2, ..., x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i - \mu)^2} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2}$$

de modo que:

$$\ln L(\mu, \sigma^2 / x_1, x_2, ..., x_n) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

por lo que, derivando respecto a μ y σ^2 e igualando a cero, tenemos:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

de donde se deducen los estimadores de máxima verosimilitud:

$$\hat{\mu}_{MV} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

$$\hat{\sigma}^2_{MV} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = S^2$$

es decir, la media y la varianza muestral, respectivamente.

Las derivadas segundas, son:

$$\frac{\partial^2 \ln L}{\partial \mu^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(-1) = -\frac{n}{\sigma^2}$$

$$\frac{\partial^2 \ln L}{\partial \mu \partial \sigma^2} = -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i - \mu) \tag{9.3}$$

$$\frac{\partial^2 \ln L}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i - \mu)^2$$

y, al sustituir μ y σ^2 por los estimadores de máxima verosimilitud, que son la media y la varianza muestral, producen el hessiano:

$$H = \begin{vmatrix} -\dfrac{n}{S^2} & 0 \\ 0 & -\dfrac{n}{2S^4} \end{vmatrix}$$

pues la derivada cruzada se hace cero al sustituir μ por la media muestral. Este hessiano es definido negativo, por lo que las soluciones que hemos hallado como estimador de máxima verosimilitud, efectivamente, maximizan dicha función.

9.3.1.2. *Propiedades del estimador de Máxima Verosimilitud*

1. Los estimadores de máxima verosimilitud *no son, en general, insesgados*. Por ejemplo, ya hemos visto que el estimador de máxima verosimilitud de σ^2 en una población $N(\mu;\sigma^2)$ es la varianza muestral S^2, que es, como sabemos, un estimador sesgado de dicho parámetro. Sin embargo, los estimadores de máxima-verosimilitud son, generalmente, consistentes, y el estimador de la varianza de una población Normal es, nuevamente, un ejemplo de esta propiedad. Por tanto, aunque pueden ser sesgados, son generalmente, asintóticamente insesgados.
2. Si existe un estadístico suficiente, el estimador de máxima verosimilitud es función únicamente de dicho estadístico, por las razones que ya expusimos en la Sección 9.3.1.
3. El estimador de máxima verosimilitud es *invariante*: si $\hat{\theta}$ es el estimador de máxima verosimilitud de θ, y g es una función inyectiva, entonces $g(\hat{\theta})$ es el estimador de máxima verosimilitud de $g(\theta)$. Esta propiedad tiene implicaciones prácticas importantes: por ejemplo, ya hemos visto que la *proporción muestral* es el estimador de máxima verosimilitud de la probabilidad p en una población $B(n,p)$. Por otra parte, la varianza de dicha distribución es: $np(1-p)$. Pues bien, el estimador de máxima verosimilitud de dicha varianza se obtiene al sustituir en su expresión teórica, el parámetro p por su estimador, es decir: $n\hat{p}(1-\hat{p})$, que podemos utilizar como varianza del estimador $\hat{p}$.
4. Las propiedades de los estimadores de máxima verosimilitud en términos de eficiencia son notables: si se cumplen las condiciones de regularidad que mencionamos tras la Definición 9.4, y si existe un estimador eficiente del parámetro que se pretende estimar, entonces dicho estimador es el estimador de máxima verosimilitud. Por tanto, si bien no es cierto que todo estimador de máxima verosimilitud es eficiente, sí es cierto que si existe un estimador eficiente del parámetro en estudio, éste es necesariamente el estimador de máxima verosimilitud.
5. Por último, enunciamos una importante propiedad acerca de la distribución de probabilidad del estimador de máxima verosimilitud. Como función que es de la muestra, el estimador es una variable aleatoria. Generalmente, nos hemos preocupado únicamente de los momentos (esperanza y varianza) del estimador, pero es generalmente muy importante poder caracterizar completamente su distribución de probabilidad. A este efecto, se tiene que, si el estimador de máxima verosimilitud es *consistente*, entonces al aumen-

tar el tamaño muestral, su distribución de probabilidad converge a una Normal, con esperanza θ y varianza igual a la inversa de la Información de Fisher, evaluada en el estimador. Por tanto, el estimador de máxima verosimilitud iguala, en tales casos, la cota de Cramer -Rao en un sentido límite, por lo que es *asintóticamente eficiente.*

Por ejemplo, ya sabemos que la media y varianza muestrales son los estimadores de máxima verosimilitud de μ y σ^2 en una población Normal. Su distribución se aproxima, al tender el tamaño muestral a infinito, a una Normal bivariante, con vector de esperanzas $(\mu;\sigma^2)$ y matriz de covarianzas igual a la inversa de la matriz de información, $I(\mu;\sigma^2)$. Para obtener dicha matriz, tomamos esperanzas en las derivadas segundas del logaritmo de la función de verosimilitud (9.3) tras cambiar de signo, y teniendo en cuenta que: $E(x_i - \mu) = 0$, $E(x_i - \mu)^2 = \sigma^2$, obtenemos:

$$Var(\hat{\mu}_{MV};\hat{\sigma}^2_{MV}) = [I(\hat{\mu}_{MV};\hat{\sigma}^2_{MV})]^{-1} = \begin{vmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{vmatrix}$$

pues la derivada cruzada se hace cero al sustituir μ por la media muestral. Este hessiano es definido negativo, por lo que las soluciones que hemos hallado como estimador de máxima verosimilitud, efectivamente, maximizan dicha función.

9.3.2. El método de momentos

El método de momentos para la obtención de estimadores tiene como principal ventaja la de su sencillez. Este procedimiento consiste en igualar momentos poblacionales respecto al origen a los correspondientes momentos muestrales, formando de este modo tantas ecuaciones como parámetros poblacionales se pretenden estimar:

$$E(X) = \mu_1(\theta_1, \theta_2, ..., \theta_k) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$E(X^2) = \mu_2(\theta_1, \theta_2, ..., \theta_k) = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

$$\cdots\cdots\cdots$$

$$E(X^k) = \mu_k(\theta_1, \theta_2, ..., \theta_k) = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

que se resuelven para encontrar los estimadores de momentos de $(\theta_1, \theta_2, ..., \theta_k)$.

Ejemplo 9.11. Consideremos una población $N(\mu,\sigma^2)$ cuyos dos parámetros queremos estimar. En este caso: $k = 2$, $\theta_1 = \mu$, $\theta_2 = \sigma^2$. Extraída una muestra de tamaño n, obtenemos los primeros dos momentos muestrales respecto al origen, y formamos las ecuaciones:

$$E(X) = \mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$E(X^2) = \sigma^2 + \mu^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

de modo que:

$$\hat{\mu} = \sum_{i=1}^{n} \frac{x_i}{n} = \bar{x}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \hat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = S^2$$

los estimadores de momentos de μ y σ^2 son la media y la varianza muestrales, respectivamente. El último es sesgado.

Ejemplo 9.12. Recordemos que la esperanza y varianza de una distribución Gamma(λ, a) son:

$$E(X) = \frac{a}{\lambda} \; ; \; Var(X) = \frac{a}{\lambda^2}$$

de modo que: $E(X^2) = \dfrac{a(a+1)}{\lambda^2}$. Para estimar por el método de momentos los parámetros a y β, formamos las ecuaciones:

$$E(X) = \frac{a}{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$E(X^2) = \frac{a(a+1)}{\lambda^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

de modo que:

$$\hat{a} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\hat{\lambda} = \bar{x}\hat{\lambda}$$

y resolviendo el sistema:

$$\hat{\lambda} = \frac{\bar{x}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2} = \frac{\bar{x}}{S^2}$$

$$\hat{\alpha} = \bar{x}\hat{\lambda} = \frac{\bar{x}^2}{S^2}$$

9.3.2.1. *Propiedades de los estimadores del método de momentos*

1. En primer lugar, debe notarse que es natural extender el método de momentos para obtener un estimador de una función cualquiera de los momentos poblacionales. Este método sugiere utilizar como estimador la misma función evaluada en los momentos muestrales. Por ejemplo, para estimar la varianza: $\sigma^2 = \alpha_2 - \alpha_1^2$, puede utilizarse: $\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 = S^2$.
2. El método de momentos utiliza la información muestral contenida en sus momentos, pero no utiliza ninguna información acerca de la distribución de probabilidad de la misma, lo que sí es utilizado por el estimador de máxima verosimilitud, por ejemplo. Ello hace que, si bien es un procedimiento muy sencillo de utilizar, no podemos garantizar que tenga siempre buenas propiedades.
3. Los estimadores obtenidos por el método de momentos son asintóticamente insesgados, pero pueden ser sesgados en muestras finitas. Son, sin embargo, insesgados, cuando se estima un parámetro que es un momento respecto al origen, como ocurre con la esperanza matemática.
4. La consistencia de los estimadores del método de momentos está garantizada por el teorema de Khintchine [Teorema 8.24], que garantiza que, bajo condiciones muy generales, los momentos muestrales convergen en probabilidad a los momentos poblacionales, y también que toda función continua de los momentos muestrales converge en probabilidad a la misma función de los momentos poblacionales, siendo, por tanto, un estimador consistente de la misma.
5. Por último, una versión del Teorema Central del Límite, aplicado al momento de orden k respecto al origen en una muestra aleatoria simple a_k, como promedio que es de variables aleatorias independientes e igualmente distribuidas: $a_k = \sum_{i=1}^{n} x_i^k$, tiene una distribución límite Normal, con esperanza matemática igual al correspondiente momento poblacional, α_k, y varianza: $(\alpha_{2k} - \alpha_k^2)/n$. Por tanto, el estimador de momentos de α_k, que es el momento muestral análogo a_k, tiene tal distribución límite.

9.4. INTERVALOS DE CONFIANZA PARA LA ESPERANZA MATEMÁTICA

9.4.1. Población Normal con varianza conocida

Dada una muestra aleatoria simple $\mathbf{X} = \{x_1, x_2, ..., x_n\}$ procedente de una población $N(\mu, \sigma^2)$, con esperanza desconocida y varianza conocida, ya sabemos que la media muestral es el estimador de máxima verosimilitud de μ, y es insesgado. Sin embargo, esto no nos dice mucho acerca de la precisión con la que conseguimos estimar μ mediante la media muestral. Para resolver esta cuestión, hemos de utilizar un aspecto que hasta ahora hemos ignorado, y es el hecho de que, en general, conocemos la distribución de probabilidad de un estimador, y no sólo su esperanza matemática y su varianza. En el caso de la media de la muestra anterior, sabemos que su distribución es $N(\mu, \sigma^2/n)$, por lo que la variable tipificada:

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Por tanto, fijada una probabilidad, por ejemplo, de 0,95, la simple inspección de las tablas de la distribución $N(0,1)$ nos dice que:

$$0{,}95 = P\left(-1{,}96 \leqslant \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leqslant 1{,}96\right)$$

del mismo modo que si hubiésemos escogido una probabilidad de 0,90, habríamos tomado $(-1{,}645;1{,}645)$, y con una probabilidad de 0,99, hubiésemos escogido el intervalo $(-2{,}33;2{,}33)$. En general, tendremos:

$$1 - a = P\left(-z_{a/2} \leqslant \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leqslant z_{a/2}\right) \tag{9.4}$$

donde $z_{a/2}$ es la abscisa que deja a su derecha una probabilidad de $a/2$ en una distribución $N(0,1)$. Por simetría, a la izquierda de $-z_{a/2}$ hay también una probabilidad de $a/2$.

Ahora bien, la desigualdad anterior es equivalente a:

$$-1{,}96\frac{\sigma}{\sqrt{n}} \leqslant \bar{x}-\mu \leqslant 1{,}96\frac{\sigma}{\sqrt{n}} \quad \Leftrightarrow \quad \bar{x} - 1{,}96\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + 1{,}96\frac{\sigma}{\sqrt{n}}$$

de modo que todas ellas son sucesos que ocurren con la misma probabilidad que la primera desigualdad, que es 0,95, por haber utilizado el factor 1,96 en vez de otro alternativo.

Es decir, podemos escribir:

$$P\left[\bar{x} - 1{,}96\left(\frac{\sigma}{\sqrt{n}}\right) \leqslant \mu \leqslant \bar{x} + 1{,}96\left(\frac{\sigma}{\sqrt{n}}\right)\right] = 0{,}95$$

que significa que la probabilidad de que el *intervalo aleatorio*:

$$\left[\bar{x} - 1{,}96\left(\frac{\sigma}{\sqrt{n}}\right); \bar{x} + 1{,}96\left(\frac{\sigma}{\sqrt{n}}\right)\right]$$

contenga al verdadero valor, desconocido, del parámetro μ, es de 0,95.

En general, tendremos un intervalo probabilístico como (9.4), del que deduciremos que con *probabilidad* igual a 100 $(1 - a)$ % el intervalo aleatorio:

$$\left[\bar{x} - z_{a/2}\frac{\sigma}{\sqrt{n}}; \bar{x} + z_{a/2}\frac{\sigma}{\sqrt{n}}\right]$$

contiene al verdadero valor de μ.

El intervalo estará centrado en la media muestral, que no es el centro de la distribución de probabilidad, que corresponde al parámetro desconcido, μ. Al cambiar de muestra, tendríamos otra media muestral diferente, situada en otro punto en un entorno de μ.

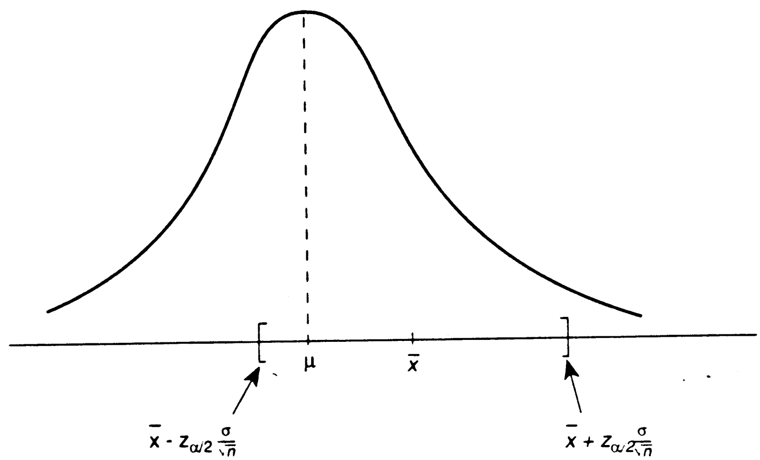

Gráfico 9.2.

La anterior afirmación probabilística es válida antes de extraer la muestra. En ese momento, los resultados muestrales son variables aleatorias, aún por realizar, y si pudiésemos tomar 100 muestras diferentes de tamaño n, deberíamos esperar que 95 de los intervalos que podríamos construir con cada una de las medias muestrales que obtuviésemos, contendrían en su interior el verdadero valor de μ. Sin embargo, una vez que tenemos una muestra concreta, con una determinada media muestral, tendremos un intervalo numérico del tipo anterior, con extremos cuantificados numéricamente, que contendrá o no a μ, lo cual es un suceso seguro, no probabilístico.

Sin embargo, el investigador no sabrá si su intervalo numérico contiene o no a μ. Tampoco puede efectuar ninguna afirmación probabilística rigurosa, pues ya hemos dicho que la probabilidad de dicho suceso es 0 o 1, sólo que desconocida. El investigador dirá que tiene un *intervalo de confianza del 95 %* para la esperanza matemática μ. Como vemos, el intervalo de confianza está centrado en torno al estimador de μ, $\bar{x}$, y hemos tomado a su izquierda y derecha un número de veces la desviación típica del estimador $\sigma/\sqrt{n}$. Si quisiéramos un nivel de confianza diferente del 95 %, no tendríamos más que cambiar el factor utilizado en su construcción. Nótese que hablamos de confianza, y no de probabilidad.

Es claro cuál es el interés de un intervalo de confianza: no sólo el investigador puede proporcionar una *estimación puntual* de μ mediante el valor numérico de $\bar{x}$, sino que, además, puede dar un rango de valores entre los cuales se halla el verdadero valor de μ con un determinado nivel de confianza. De hecho puede dar varios de tales intervalos de valores, si así lo desea. Sin embargo, al construir un intervalo de confianza, debe ponderar dos objetivos contradictorios entre sí: por un lado, querría tener la máxima confianza posible en que el intervalo que proporcione contenga el valor desconocido de μ. Por otra parte, al aumentar el nivel de confianza, el intervalo se amplía, con lo cual puede llegar a proporcionar una información irrelevante, pues de poco serviría saber que con una confianza de 0,999, μ se halla dentro del intervalo $(-34{,}7\ ;\ 254{,}1)$, por ejemplo. Fijado un nivel de confianza, la longitud del intervalo es un indicador inverso de la *precisión* con que hemos llevado a cabo la estimación. Cuanto más corto sea el intervalo de confianza, mayor será la precisión, pues el rango de valores dentro del cual se halla μ con una determinada confianza, es más estrecho. Estamos, por tanto, muy interesados en que *los intervalos sean reducidos, siempre que proporcionen un mínimo* *nivel de confianza*. Notemos que, fijado un nivel de confianza, es posible disminuir la longitud del intervalo correspondiente a dicho nivel de confianza, si aumentamos el tamaño muestral.

Debe notarse, además, que hemos propuesto un intervalo de confianza simétrico, teniendo igual longitud a la izquierda que a la derecha de la media muestral. Esta no es la única estrategia posible, pero nos proporciona el intervalo de confianza del 95 % de menor longitud. Por ejemplo, el lector puede comprobar fácilmente que el intervalo:

$$\left[\bar{x} - 2{,}33\left(\frac{\sigma}{\sqrt{n}}\right); \bar{x} + 1{,}75\left(\frac{\sigma}{\sqrt{n}}\right)\right]$$

es también un intervalo de confianza del 95 %, pero es algo más amplio que el anterior.

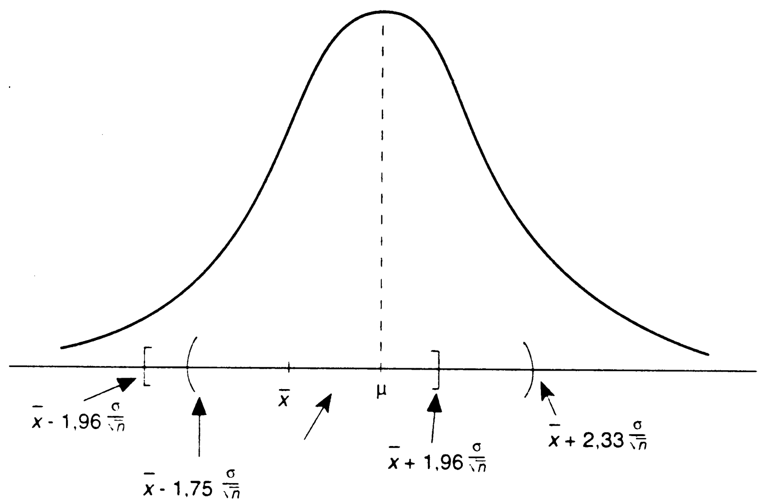

Gráfico 9.3.

Ejemplo 9.13. Supongamos que la distribución del gasto mensual en consumo de bienes alimenticios, en miles de ptas., por cada persona mayor de 16 años sigue, en una determinada ciudad, una distribución $N(\mu; 15000^2)$. Al encuestar a 100 personas mayores de 16 años y preguntarles por su gasto en alimentos durante el último mes, se ha obtenido un gasto medio de 35.650 ptas. Un intervalo de confianza del 95 % para la esperanza matemática es:

$$\left[\bar{x} - z_{0.025}\frac{\sigma}{\sqrt{n}}; \bar{x} + z_{0.025}\frac{\sigma}{\sqrt{n}}\right] =$$

$$= \left[\bar{x} - z_{0.025}\frac{15000}{\sqrt{100}}; \bar{x} + z_{0.025}\frac{15000}{\sqrt{100}}\right] =$$

$$= [35650 - (1{,}96)(1500); 35650 + (1{,}96)(1500)] = [32710; 38590]$$

en ptas., como gasto mensual por persona en bienes alimenticios en dicha ciudad.

9.4.2. Población Normal con varianza desconocida

Es poco probable que el investigador desconozca la esperanza matemática de una población, pero conozca su varianza. Sólo en rarísimas ocasiones estará dispuesto a apelar a experimentos previos en que midió la desviación típica, y suponer que ésta continúa siendo la misma y, por tanto, conocida. Así, vamos a discutir en esta sección el caso más frecuente en que se desconoce la varianza de la población, por lo que el procedimiento de la sección anterior no es aplicable. Sin embargo, ya probamos en la Sección 8.6 que:

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

donde s denota la cuasivarianza muestral. Por tanto, si escogemos la abscisa $t_{\alpha/2}$ que en una distribución t de Student deja a su derecha una probabilidad de α, teniendo en cuenta que esta distribución es simétrica, como ocurría con la Normal, tendremos:

$$1 - \alpha = P\left(-t_{n-1,\alpha/2} \leqslant \frac{\bar{x} - \mu}{s/\sqrt{n}} \leqslant t_{n-1,\alpha/2}\right)$$

con un *intervalo de confianza del* 100 (1 − α) %:

$$\left[\bar{x} - (t_{n-1,\alpha/2})\frac{s}{\sqrt{n}}\,;\bar{x} + (t_{n-1,\alpha/2})\frac{s}{\sqrt{n}}\right]$$

Para muestras de tamaño $n = 30$ o superiores, se tiene que la distribución t con tales grados de libertad coincide con la Normal, por lo que puede utilizarse entonces la tabla de la $N(0,1)$, incluso cuando se desconoce la varianza poblacional.

Ejemplo 9.14. En el análisis del gasto en consumo de bienes alimenticios, supongamos que desconocemos la varianza muestral, y que al encuestar a 20 personas, obtenemos unas cifras de gastos, en miles de ptas.: $\mathbf{X}$ = {28,4; 31,2; 36,0; 29,4; 32,5; 37,4; 24,2; 26,6; 34,1; 22.3; 28.5; 36,3; 24,3; 27,3; 21,4; 23,8; 26,2; 24,7; 29,5; 34,2}. Con estos datos, se tiene: $\bar{x} = 28.915$; $s^2 = 23{,}877$; $s = 4{,}886$, de modo que una estimación del gasto medio por persona en toda la población es de 28.915 ptas. Por otra parte, el valor crítico de la distribución $t_{19,0.025}$ es de 2.093, por lo que el intervalo de confianza de 95 % es:

$$\left[28{,}915 - (2{,}093)\frac{4{,}37}{\sqrt{20}}\,;28{,}915 + (2{,}093)\frac{4{,}37}{\sqrt{20}}\right] = (26{,}87\,;30{,}96)$$

en miles de ptas.

Si se comparan las tablas de la distribución t de Student para un número reducido de grados de libertad, se observa que $t_{n-1,\alpha/2} > z_{\alpha/2}$ para cualquier α, siendo $z_{\alpha/2}$ la abscisa de la distribución $N(0,1)$. Ello significa que el intervalo de confianza del 100(1−α)% tenderá a ser más amplio cuando se utiliza la distribución t, es decir, cuando no se conoce la varianza poblacional, que cuando se conoce. Decimos que *tenderá a ser más amplio*, sin afirmarlo categóricamente, porque la amplitud de dicho intervalo depende también de la cuasivarianza muestral. Puede probarse que, en promedio, la longitud del intervalo construido con la t será mayor que la del intervalo construido con la distribución Normal, como cabría esperar, pues en el segundo caso contamos con más información, y podemos lograr más precisión, es decir, un intervalo más estrecho.

9.4.3. Intervalos de confianza con poblaciones distintas de la Normal

Cuando no podemos suponer que la población de la que se extrajo la muestra sigue una distribución Normal, podemos utilizar el Teorema Central del Límite para garantizar que el cociente: $\dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ tiene una *distribución aproximada* $N(0,1)$, si el tamaño muestral es suficientemente grande. En este caso, tendremos:

$$1 - \alpha \simeq P\left(-z_{\alpha/2} \leqslant \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leqslant z_{\alpha/2}\right)$$

y tenemos el *intervalo de nivel aproximado de confianza* del $100(1-\alpha)\%$:

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\,;\,\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

El grado de aproximación con que se satisface el nivel de confianza fijado depende del tipo de distribución poblacional, y del tamaño muestral. Si la distribución poblacional es simétrica, unimodal y continua, la aproximación es bastante buena incluso para tamaños muestrales muy reducidos. Cuanto más se distinga la distribución poblacional de una Normal, mayor deberá ser el tamaño muestral para lograr una buena aproximación al nivel de confianza fijado.

Si, además de no poder suponer que la población es Normal, se desconoce la varianza, entonces puede utilizarse una *distribución t aproximada* para el ratio $\dfrac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$. Si el tamaño muestral es mayor de 25, esta aproximación es suficientemente buena en la mayoría de los casos. Si la distribución poblacional es simétrica, unimodal y continua, la aproximación es buena incluso con un tamaño muestral apreciablemente menor.

9.4.4. Intervalos para la diferencia de esperanzas en poblaciones Normales

Supongamos que queremos comparar las esperanzas matemáticas de dos poblaciones Normales, $N(\mu_X;\sigma_X^2)$ y $N(\mu_Y;\sigma_Y^2)$. Sean $\mathbf{X} = \{x_1, x_2, ..., x_n\}$ e $\{y_1, y_2, ..., y_m\}$ muestras aleatorias simples, independientes entre sí, de tamaños respectivos n y m, y varianzas conocidas. Si denotamos por $\bar{x}$ e $\bar{y}$ las medias muestrales, entonces ya vimos en la Sección 8.2 que su diferencia es una variable aleatoria con distribución:

$$\bar{x} - \bar{y} \sim N\left(\mu_X - \mu_Y; \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)$$

por lo que:

$$1 - a = P\left(-z_{a/2} \leqslant \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}}} \leqslant z_{a/2}\right)$$

que puede escribirse:

$$1 - a = P\left[(\bar{x} - \bar{y}) - z_{a/2}\sigma; \leqslant M_x - M_y \leqslant (\bar{x} - \bar{y}) + z_{a/2}\sigma\right]$$

donde $\sigma = \sqrt{\sigma_X^2/n + \sigma_Y^2/m}$, por lo que tenemos como *intervalo de confianza del* $100(1-a)\%$:

$$\left[(\bar{x} - \bar{y}) - z_{a/2}\sigma ; (\bar{x} - \bar{y}) + z_{a/2}\sigma\right]$$

que es simétrico y centrado alrededor del estimador puntual de la diferencia $\mu_X - \mu_Y$, que es la diferencia de medias $\bar{x} - \bar{y}$, tomando a la izquierda y derecha de éste la desviación típica de dicha diferencia, multiplicada por $z_{a/2}$.

Ejemplo 9.15. En el ejemplo 9.13 consideramos una población $N(\mu_X; 15.000^2)$ de la que se había extraído una muestra de tamaño 100, con una media de 35.650 ptas. de gasto. Supongamos ahora que disponemos de una segunda población, con distribución $N(\mu_Y; 800^2)$, en la que se encuesta a 36 personas, obteniendo una media de 32.400 ptas. de gasto.

El parámetro σ^2, cuyo cuadrado es una suma ponderada de las dos varianzas, es:

$$\sigma = \sqrt{\frac{15.000^2}{100} + \frac{800^2}{36}} = 1.505,9$$

mientras que la diferencia de medias muestrales es: 35.650 − 32.400 = 3.250. Al 95 % de confianza, el valor crítico de la Normal es 1,96, de modo que tenemos el intervalo para la diferencia de las dos esperanzas matemáticas:

$$[3.250 - (1,96)(1.505,9) ; 3.250 + (1,96)(1.505,9)] = (298,5 ; 6.201,6)$$

en pesetas. La diferencia entre los niveles medios de gasto en ambas poblaciones *parece ser* significativa, puesto que el intervalo de confianza del 95 % no contiene al 0.

Cuando las varianzas son desconocidas, obtener un intervalo de confianza para la diferencia de las esperanzas matemáticas es un problema difícil si las muestras son de reducido tamaño. Sin embargo, si podemos suponer que las varianzas, aun siendo desconocidas, son iguales, entonces podemos proceder a construir el intervalo incluso con tamaños muestrales reducidos.

Vimos en la Sección 8.6 que, con igualdad de varianzas:

$$\frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}}} \sim t_{n+m-2}$$

por lo que podemos determinar la probabilidad de un intervalo del tipo:

$$P\left[(\bar{x} - \bar{y}) - (t_{n+m-2, a/2})\left(s\sqrt{\frac{m+n}{nm}}\right) \leqslant \right.$$

$$\left. \leqslant \mu_X - \mu_Y \leqslant (\bar{x} - \bar{y}) + (t_{n+m-2, a/2})\left(s\sqrt{\frac{m+n}{nm}}\right)\right] = 1 - a$$

donde:

$$s = \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}}$$

y tenemos por tanto, el *intervalo de confianza* del $100(1-a)\%$:

$$\left((\bar{x} - \bar{y}) - (t_{n+m-2, a/2})\, s\sqrt{\frac{m+n}{nm}} \quad ; \quad (\bar{x} - \bar{y}) + (t_{n+m-2, a/2})\, s\sqrt{\frac{m+n}{nm}}\right)$$

Ejemplo 9.16. Supongamos que, en la situación del Ejemplo 9.14, se ha obtenido una muestra de las cifras de gasto mensual en alimentos en otra ciudad, tras encuestar a 12 personas: $\mathbf{Y} = \{32{,}1;\ 36{,}1;\ 35{,}4;\ 42{,}4;\ 36{,}5;\ 34{,}1;\ 27{,}5;\ 28{,}3;\ 30{,}0;\ 35{,}8;\ 29{,}8;\ 38{,}4\}$, que tiene: $\bar{x}_Y = 33{,}867$; $s_Y^2 = 19{,}779$; $s_Y = 4{,}447$, y queremos construir un intervalo de confianza del 90 % para la diferencia entre el gasto medio en consumo alimenticio en ambas ciudades, la del ejemplo 9.14 y ésta. En primer lugar, hemos de obtener el parámetro s:

$$s = \sqrt{\frac{(19)(23{,}877) + (11)(19{,}779)}{20 + 12 - 2}} = 4{,}730$$

Por otra parte, la raíz que depende de los tamaños muestrales es: $\sqrt{32/240} = 0{,}365$, y el valor crítico $t_{30:0{,}05} = 1{,}697$, mientras que la diferencia de medias muestrales es: $\bar{x} - \bar{y} = 28{,}915 - 33{,}867 = -4{,}952$, todo lo cual conduce al intervalo: $(-7{,}88, -2{,}02)$. No hay nada incorrecto en obtener ambos signos negativos. Ello se debe a que estamos tomando la diferencia en el sentido de la muestra que produjo el menor gasto medio *menos* la que generó el mayor gasto medio. Lo que el intervalo sugiere es que parece que deberíamos creer que la diferencia entre las esperanzas matemáticas respectivas toma signo negativo, lo que sería consistente con la posible creencia de que el gasto alimenticio en la segunda ciudad es significativamente superior al de la primera.

Si no podemos afirmar que las varianzas son igual entre sí, pero sí estamos bastante confiados en conocer su valor relativo: $\sigma_X^2/\sigma_Y^2 = k$, podemos proceder de un modo similar al de la Sección 8.6. En estas condiciones, se tienen las distribuciones:

$$\bar{x} - \bar{y} \sim N\left[\mu_X - \mu_Y; \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right] = N\left[\mu_X - \mu_Y; \sigma_Y^2\left(\frac{k}{n} + \frac{1}{m}\right)\right]$$

$$\frac{(n-1)\,s_X^2}{k\sigma_Y^2} + \frac{(m-1)\,s_Y^2}{\sigma_Y^2} \sim \chi^2_{n+m-2}$$

y la segunda es independiente de la diferencia de las medias muestrales, pues cada cuasivarianza es independiente de la media de su muestra y, además, las dos poblaciones son independientes entre sí, luego también lo son las muestras. Por tanto, el cociente de la distribución Normal tipificada, por la variable chi-cuadrado, dividida por sus grados de libertad, sigue una distribución t de Student:

$$\frac{[(\bar{x}-\bar{y}) - (\mu_X - \mu_Y)]/\sqrt{\sigma_Y^2\left(\dfrac{k}{n} + \dfrac{1}{m}\right)}}{\sqrt{\dfrac{1}{m+n-2}\dfrac{1}{\sigma_Y^2}\left(\dfrac{(n-1)s_X^2}{k} + (m-1)s_Y^2\right)}} = \frac{(\bar{x}-\bar{y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{mk+n}{mn}}\sqrt{\dfrac{(n-1)s_X^2 + k(m-1)s_Y^2}{m+n-2}}} \sim t_{m+n-2}$$

independiente del valor desconocido de σ_Y^2, que ha desaparecido de la expresión.

9.5. INTERVALOS DE CONFIANZA PARA LA VARIANZA Y PARA EL COCIENTE DE VARIANZAS EN POBLACIONES NORMALES

Recordemos que si s_X^2 es la *cuasivarianza* muestral de X, entonces [Teorema 8.17]:

$$\frac{(n-1)s_X^2}{\sigma^2} \sim \chi^2_{n-1}$$

de modo que podemos encontrar constantes a y b tales que:

$$1 - a = P\left(a \leqslant \frac{(n-1)s_X^2}{\sigma^2} \leqslant b\right) \tag{9.5}$$

para cualquier nivel de confianza $1-\alpha$ que fijemos. En este caso, tenemos una complicación con respecto a las situaciones previas, basadas en distribuciones Normales. La distribución chi-cuadrado no es simétrica, por lo que no es evidente el modo de escoger las constantes a y b. Por analogía con el caso simétrico, parece razonable escogerlas de modo que:

$$a = \chi^2_{n-1,\,1-\alpha/2} \quad ; \quad b = \chi^2_{n-1,\,\alpha/2}$$

que dejan la misma probabilidad, $\alpha/2$ en ambas colas fuera de intervalo. Este intervalo no es, sin embargo, el más corto en el caso de la distribución χ^2.

Una vez que se han seleccionado las constantes, el intervalo probabilístico puede asimismo escribirse:

$$1-\alpha = P\left(\frac{(n-1)s_X^2}{b} \leqslant \sigma^2 \leqslant \frac{(n-1)s_X^2}{a}\right)$$

que se obtiene despejando σ^2 en (9.5). Debe apreciarse la reversión en el orden de los extremos a y b a ambos lados de la cadena de desigualdades. A partir de aquí, obtenemos el intervalo de confianza del $100(1-\alpha)\%$:

$$\left(\frac{(n-1)s_X^2}{b} \; ; \; \frac{(n-1)s_X^2}{a}\right)$$

Ejemplo 9.17. En la primera muestra de gasto en alimentos, obtuvimos una cuasivarianza: $s_X^2 = 23{,}877$ lo que implica: $(n-1)s_X^2 = 453{,}663$. Por otra parte, tenemos los valores críticos: $\chi_{19:0.95} = 7{,}633$ y $\chi_{19:0.05} = 30.14$, que delimitan una probabilidad de 0,90. Por tanto, el intervalo del 90 % de confianza para la varianza de esta población es:

$$\left(\frac{453{,}663}{30{,}14} \leqslant \sigma^2 \leqslant \frac{453{,}663}{7{,}633}\right) = (15{,}05;59{,}434)$$

por lo que podemos tomar asimismo como intervalo de confianza del 90 % para la desviación típica: (3,88;7,71).

Si se quieren comparar las varianzas de dos poblaciones Normales independientes, tomaremos una muestra de cada una de ellas, con lo que podremos construir dos estadísticos del tipo que hemos utilizado en el caso anterior: $(n-1)s_X^2/\sigma_X^2$ y $(m-1)s_Y^2/\sigma_Y^2$, donde n y m denotan los tamaños muestrales respectivos. El primero de estos estadísticos tiene una distribución chi-cuadrado con $n-1$ grados de libertad, mientras que el segundo sigue una distribución chi-cuadrado con $m-1$ grados de libertad. Ambos son independientes, por serlo las poblaciones de las que se extrajeron las muestras.

Por tanto, el cociente:

$$\frac{(m-1)s_Y^2/\sigma_Y^2}{(n-1)s_X^2/\sigma_X^2}\,\frac{1/(m-1)}{1/(m-1)}=\frac{s_Y^2/\sigma_Y^2}{s_X^2/\sigma_X^2}\sim F_{m-1,n-1}$$

de modo que podemos encontrar en las tablas de esta distribución de probabilidad dos contrastes a y b tales que:

$$1-a=P\left(a\leqslant\frac{s_Y^2/\sigma_Y^2}{s_X^2/\sigma_X^2}\leqslant b\right)=P\left(a\frac{s_X^2}{s_Y^2}\leqslant\frac{\sigma_X^2}{\sigma_Y^2}<b\frac{s_X^2}{s_Y^2}\right)$$

que genera el intervalo de confianza del $100(1-a)\%$:

$$\left(a\frac{s_X^2}{s_Y^2}\ ;\ b\frac{s_X^2}{s_Y^2}\right)$$

Dadas las limitaciones de la tabla de la distribución F, solemos escoger: $a = F_{m-1,n-1;\,1-a/2} = 1/F_{n-1,m-1;a/2}$ y $b = F_{m-1,n-1;a/2}$.

Ejemplo 9.18. Para comparar las varianzas (desconocidas) de los gastos en bienes alimenticios de los ejemplos 9.14 y 9.16, recordemos que $s_X^2 = 23{,}877$ y $s_Y^2 = 19{,}779$. No es fácil encontrar valores críticos para $F_{19,11}$, pero podemos aproximarlos por los de la $F_{20,10}$, teniendo: $F_{20:10:0.975} = 2{,}77$, a la vez que: $F_{10:20:0.975} = 3{,}42$, de modo que el intervalo de confianza para el cociente de varianzas es:

$$\left(a\frac{s_X^2}{s_Y^2}\,;\,b\,\frac{s_X^2}{s_Y^2}\right)=\left(\frac{1}{3{,}42}\,\frac{19{,}78}{23{,}88}\quad;\quad(2{,}77)\,\frac{19{,}78}{23{,}88}\right)=(0{,}242\,;\,2{,}29)$$

no rechazando la igualdad de σ_X^2 y σ_Y^2, pues el intervalo contiene a 1.

9.6. INTERVALOS DE CONFIANZA PARA PROPORCIONES

En muchas ocasiones, el investigador está interesado en estimar una proporción y, al igual que en los casos anteriores, es de sumo interés que pueda proporcionar una indicación de la precisión con que ha llevado a cabo dicha estimación. Este indicador consiste en un intervalo de un determinado nivel de confianza. Para su construcción, utilizamos el hecho de que cuando observamos n repeticiones de un fenómeno Bernouilli, el número X de *éxitos* sigue una distribución $B(n,p)$, y, para tamaños muestrales n suficientemente grandes, puede utilizarse la implicación del Teorema Central del Límite:

$$\frac{x-np}{\sqrt{np(1-p)}}=\frac{x/n-p}{\sqrt{p(1-p)/n}}\sim N(0,1)$$

Esto implica que podemos hallar un punto $z_{a/2}$ tal que:

$$1 - a \simeq P\left(-z_{a/2} \leqslant \frac{x/n - p}{\sqrt{p(1-p)/n}} \leqslant z_{a/2}\right)$$

es decir:

$$1 - a \simeq P\left(\frac{x}{n} - z_{a/2}\sqrt{\frac{p(1-p)}{n}} \leqslant p \leqslant \frac{x}{n} + z_{a/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

lo cual no nos resuelve el problema del modo en que lo hacían los intervalos de secciones anteriores. Nótese que, en esta ocasión, aunque tuviéramos la muestra y contabilizásemos el número de éxitos x en ella, no podríamos construir el intervalo de confianza a partir del intervalo probabilístico anterior, debido a la presencia del parámetro desconocido p en ambos extremos del mismo. Para evitar esta dificultad, se sustituye dicho parámetro por su estimador, la proporción muestral x/n, obteniendo el intervalo de confianza del $100(1-a)$ %:

$$\left(\frac{x}{n} - z_{a/2}\sqrt{\frac{(x/n)(1-(x/n))}{n}}\,;\,\frac{x}{n} + z_{a/2}\sqrt{\frac{(x/n)(1-(x/n))}{n}}\right)$$

donde hemos hecho una aproximación adicional, que será válida si la muestra es suficientemente grande.

Ejemplo 9.19. En la primera muestra de gasto en consumo alimenticio, 7 personas declararon un gasto superior a las 30.000 ptas. mensuales, una proporción de $7/20 = 0{,}35$. El intervalo del 90 % de confianza para la proporción poblacional es:

$$\left(\frac{7}{20} - (1{,}645)\sqrt{\frac{(7/20)(1-(7/20))}{20}}\,;\,\frac{7}{20} + (1{,}645)\sqrt{\frac{(7/20)(1-(7/20))}{20}}\right) =$$

$$= (0{,}1745\,;\,0{,}526)$$

Realmente este intervalo es muy poco informativo, pues afirmar que la proporción de personas que gastan al mes más de 30.000 ptas. en alimentos está entre el 17,5 % y el 52,6 % no es muy útil. Ello se debe al reducido tamaño muestral. El lector puede comprobar fácilmente que si la misma proporción muestral, 0,35, se hubiera obtenido en una muestra de 2.000 personas, el intervalo de confianza del 90 % para la proporción poblacional sería: (0,3325 ; 0,3675), mucho más informativo que el que pudimos delimitar con un tamaño muestral muy inferior.

Por último, estaremos interesados en ocasiones en contrastar la igualdad de proporciones de dos poblaciones Bernouilli independientes. Para ello, tomaremos dos muestras de tamaño n y m, y calcularemos las proporciones de *éxitos* en cada caso: x/n e y/m. Vimos en la Sección 8.8.3 que, para tamaños muestrales suficientemente grandes, tenemos la distribución de la diferencia de proporciones muestrales:

$$\frac{x}{n} - \frac{y}{m} \simeq N\left(p_x - p_y\,; \frac{p_x(1-p_x)}{n} + \frac{p_y(1-p_y)}{m}\right)$$

por lo que la variable tipificada:

$$\frac{\left(\frac{x}{n} - \frac{y}{m}\right) - (p_x - p_y)}{\sqrt{\frac{p_x(1-p_x)}{n} + \frac{p_y(1-p_y)}{m}}} \simeq N(0,1)$$

y podemos recurrir a las tablas de esta distribución para obtener el punto $z_{\alpha/2}$ de modo que:

$$1 - \alpha \simeq P\left(-z_{\alpha/2} \leqslant \frac{\left(\frac{x}{n} - \frac{y}{m}\right) - (p_x - p_y)}{\sqrt{\frac{p_x(1-p_x)}{n} + \frac{p_y(1-p_y)}{m}}} \leqslant z_{\alpha/2}\right)$$

donde tenemos un problema similar al de la construcción de un intervalo para la proporción. en el sentido de que los parámetros desconocidos apareeen en los extremos del intervalo. Lo resolvemos de modo análogo, sustituyendo los parámetros por sus estimadores, llegando al intervalo de confianza del $100(1-\alpha)$ % para la diferencia de proporciones $p_x - p_y$ en la forma:

$$\left(\left(\frac{x}{n} - \frac{y}{m}\right) - z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n} + \frac{\frac{y}{m}\left(1-\frac{y}{m}\right)}{m}}\;;\right.$$

$$\left.\left(\frac{x}{n} - \frac{y}{m}\right) + + z_{\alpha/2}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n} + \frac{\frac{y}{m}\left(1-\frac{y}{m}\right)}{m}}\right)$$

Ejemplo 9.20. En la segunda muestra hubo 8 de 12 personas con gasto superior a las 30.000 ptas., una proporción de $8/12 = 0{,}666$. El intervalo de confianza para la diferencia de proporciones entre ambas poblaciones es:

$$\left((0{,}35 - 0{,}666) - (1{,}645)\sqrt{\frac{(0{,}35)(1 - 0{,}35)}{20} + \frac{(0{,}666)(1 - 0{,}666)}{12}}\,;(0{,}35 - 0{,}666) + \right.$$

$$\left. + (1{,}645)\sqrt{\frac{(0{,}35)(1 - (0{,}35))}{20} + \frac{(0{,}666)(1 - 0{,}666)}{12}}\right) = (0{,}033\,;0{,}601)$$

donde el intervalo es amplio, nuevamente, debido a los reducidos tamaños muestrales. No contiene, sin embargo, a 0, lo que debería hacernos pensar acerca de posibles diferencias relevantes entre los hábitos de consumo en ambas ciudades.

EJERCICIOS

9.1. Se ha extraído una muestra de tamaño 3 de una población con esperanza matemática y varianza desconocidas: μ y σ^2. Pruebe que el estimador: $Z = (x_1 + 2x_2 + 3x_3)/6$ es un estimador insesgado de μ, y compare su eficiencia con la de la media muestral.

9.2. Considere como estimador de la esperanza matemática la siguiente combinación lineal convexa de las observaciones muestrales: $\hat{\mu} = \sum_{i=1}^{n} c_i x_i$ donde la suma de los coeficientes c_i es igual a 1. Pruebe que es un estimador insesgado.

9.3. A partir de una muestra de tamaño n de una población $N(\mu;\sigma^2)$, considere los estimadores de μ: $\hat{\mu}_1 = \frac{1}{n-1}\sum_{i=1}^{n} x_i$ y $\hat{\mu}_2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i$. ¿Cuál de los dos preferiría en términos del ECM?

9.4. Dada una muestra aleatoria simple extraída de una población con esperanza matemática μ y varianza σ^2, pruebe directamente, utilizando un argumento similar al del Teorema 9.3, que el cuadrado de la media muestral tiene esperanza matemática:

$$E(\bar{x}^2) = \frac{\sigma^2}{n} + \mu^2$$

9.5. Halle la cota de Cramer-Rao para la estimación del parámetro λ de una población de Poisson.

9.6. Halle la cota de Cramer-Rao para la estimación del parámetro λ de una población exponencial:

$$f(x) = \frac{1}{\theta} e^{-x/\theta} \qquad 0 < x < \infty \qquad 0 < \theta < \infty$$

9.7. Pruebe que la media muestral es un estimador insesgado del parámetro θ en una distribución exponencial, y que su varianza es θ^2/n. ¿Es asimismo un estimador consistente?

9.8. Demuestre que la media aritmética es un estadístico suficiente para la función de densidad exponencial.

9.9. Demuestre que la media aritmética es un estadístico suficiente en una distribución de Poisson.

9.10. Pruebe que en una muestra aleatoria simple de una población $N(\mu;\sigma^2)$, la cantidad de información acerca del parámetro μ es $I(\mu) = n/\sigma^2$.

9.11. A partir de una muestra de una población $B(n,p)$, considere el cociente $\bar{x}/n$ como estimador del parámetro p. Halle la cota de Cramer-Rao y pruebe que este estimador es eficiente.

9.12. Pruebe que la media muestral es un estimador eficiente del parámetro λ en una distribución de Poisson $P(\lambda)$.

9.13. Suponga que la proporción de votantes que votan por el Partido Blanco en una determinada elección se distribuye de acuerdo con la función de densidad:

$$f(p) = \theta p^{\theta-1} \quad 0 < p < 1 \quad 0 < \theta < \infty$$

Justo antes de celebrarse las elecciones, se ha encuestado a votantes de 5 distritos, obteniendo unos porcentajes de votantes declarando su intención de votar por el Partido Blanco, de: 0,43, 0,46, 0,52, 0,34, 0,32, respectivamente. Estime el parámetro θ.

9.14. Considere que una función de densidad Gamma $(a;q)$ parametrizada:

$$f(x;a,q) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \quad 0 \leqslant x$$

y tome la media muestral como estimador del parámetro λ. Pruebe que es un estimador sesgado, calcule su varianza y compárela con la cota de Cramer-Rao. Muestre que es de mínima varianza dentro de los estimadores con su sesgo.

9.15. Obtenga el estimador de máxima verosimilitud del parámetro μ en una población $N(\mu;25)$.

9.16. Obtenga el estimador de máxima verosimilitud del parámetro σ^2 en una población $N(1;\sigma^2)$.

9.17. Obtenga el estimador de máxima verosimilitud del parámetro θ de una distribución exponencial.

9.18. Obtenga el estimador de máxima verosimilitud del parámetro θ de una distribución geométrica:

$$f(x/\theta) = (1-\theta)^{x-1}\theta \quad x = 1,2,3,...$$

9.19. Sea X_i una variable aleatoria que puede tomar k valores numéricos: $x_1, x_2, ..., x_k$ con probabilidades: $p_1, p_2, ..., p_k$, con $\sum_{i=1}^{n} p_i = 1$. Esta es la *distribución multinomial.* Si se extrae una muestra de tamaño n, en la que se obtienen n_1 valores iguales a x_1, n_2 valores iguales a x_2, ..., n_k valores iguales a x_k, con $\sum_{i=1}^{n} x_i = n$.Probar que el estimador de máxima verosimilitud del vector $p = p_1$, $p_2, ..., p_k$, es:

$$\hat{p} = \left(\frac{n_1}{n}, \frac{n_2}{n}, ..., \frac{n_k}{n}\right)$$

9.20. Una variable aleatoria tiene como función de densidad: $f(x) = kx^2/\theta^2$ sobre el intervalo $0 \leqslant x \leqslant \theta$. Calcule la esperanza y la varianza del estimador y pruebe que su varianza es inferior a la propuesta en la cota de Cramer-Rao. ¿Cómo puede suceder esto?

9.21. Represente gráficamente la función de densidad:

$$f(x/\theta) = \theta x^{\theta-1} \quad 0 < x < 1 \quad 0 < \theta < \infty$$

para: *a*) $\theta = 1/2$, *b*) $\theta = 1$, *c*) $\theta = 2$. Pruebe que el estimador de máxima verosimilitud de θ es:

$$\hat{\theta}_{MV} = -\frac{n}{\ln \prod_{i=1}^{n} x_i}$$

9.22. Dada una muestra aleatoria simple de tamaño n de una población con función de densidad:

$$f(x) = \frac{1}{\theta^2} x e^{-x/\theta} \qquad 0 < x < \infty, \qquad 0 < \theta < \infty$$

hallar el estimador de máxima verosimilitud del parámetro θ.

9.23. Los estimadores de máxima verosimilitud satisfacen una propiedad denominada de *invariancia*: el estimador de máxima verosimilitud de una función $g(\theta)$ es: $g(\theta_{MV})$. Utilizar esta propiedad para hallar el estimador de máxima verosimilitud de la desviación típica de la distribución de Poisson.

9.24. Dada una muestra aleatoria simple de tamaño n, extraída de una población representada por la función de densidad:

$$f(x/\theta) = (\theta + 1)x^\theta \qquad 0 < x < 1$$

Hallar el estimador de θ por el método de momentos.

9.25. Obtenga el estimador de momentos del parámetro θ en una distribución Uniforme sobre el intervalo $(0,\theta)$, y pruebe que es consistente.

9.26. Obtenga el estimador de momentos del parámetro p en una población $B(n,p)$.

9.27. Obtenga los estimadores de momentos de los parámetros a y λ en una distribución Gamma $G(a,\lambda)$.

9.28. Utilizar el método de momentos para obtener el estimador del parámetro θ en una función de densidad:

$$f(\theta) = \theta x^{\theta-1} \quad , \quad 0 < x < 1$$

9.29. ¿Es el estimador $\theta = \bar{x}$ consistente en la estimación del parámetro λ de una distribución de Poisson?

9.30. Demuestre que la suma de los elementos muestrales no es un estimador consistente en la estimación del parámetro θ de una distribución exponencial.

9.31. Utilice la desigualdad de Chebychev y la descomposición del *ECM* en suma de varianza y sesgo al cuadrado para probar que si la varianza de un estimador asintóticamente insesgado tiende a cero al aumentar el tamaño muestral, dicho estimador es consistente.

9.32. Pruebe que la media muestral es un estimador suficiente de μ en una población $N(\mu;1)$. Pruebe que su cubo también lo es, mientras que $\bar{x}^2$ no lo es.