

Capítulo 7

Contrastes de hipótesis

7.1. Introducción	1
7.1.1. Contraste de media de normal, con varianza conocida	2
7.1.2. Contraste de proporción	7
7.2. Contrastes bilaterales	8
7.2.1. Test bilaterales más usuales	8
7.3. Contrastes unilaterales	10
7.3.1. Diseño de test unilaterales. Elección de hipótesis nula	10
7.3.2. Test unilaterales más usuales	12
7.4. Test generales, función de potencia	13
7.5. Test de razón de verosimilitudes	19
7.6. Test uniformemente más potentes	29
7.6.1. Lema de Neyman–Pearson	29

7.1. Introducción

Continuamos en este capítulo dentro del marco general de los dos anteriores: una cierta variable X tiene función de masa/densidad $f(x; \theta)$, donde θ es un parámetro desconocido. El objetivo de la tecnología de inferencia (paramétrica) consiste en obtener información sobre θ a partir de una muestra empírica (x_1, \dots, x_n) .

En el capítulo 5 analizamos *estimadores* del parámetro θ (métodos de construcción y propiedades). Con un estimador T , la muestra empírica da lugar a una *estimación* $\hat{\theta}$ del parámetro θ .

En el capítulo 6 nos ocupamos de establecer *intervalos de confianza* I para el parámetro θ , de manera que, con alta confianza, el verdadero (y desconocido) valor de θ esté en el intervalo I (que depende de la muestra).

En este capítulo cambiaremos el punto de vista, y desarrollaremos una tecnología para *contrastar hipótesis* sobre el parámetro, como que valga 3, que sea mayor

que 5, etc. Avisamos al lector de que, aunque la cuestión de interés sea distinta, la metodología que diseñaremos se va a apoyar en gran medida en los desarrollos que dieron lugar a los intervalos de confianza.

Digamos, por ejemplo, que tenemos una muestra (x_1, \dots, x_n) , donde los x_i son ceros o unos, que sabemos que ha sido obtenida lanzando repetidamente una moneda. La hipótesis que queremos contrastar (qué otra podía ser) es que la moneda en cuestión es *regular*. En términos más técnicos, si X sigue una $\text{BER}(1/2)$ o no. Observamos la muestra, contamos ceros y unos, calculamos promedios y obtenemos que en la muestra aparece un 30 % de unos. Esto, desde luego, parece ir en contra de la hipótesis; no es que sea imposible que una moneda regular produzca tal proporción de caras, pero se nos antoja bastante improbable. ¿Cuánto?, ¿lo suficiente como para descartar que la moneda sea regular?

Veamos. Si la muestra fuera de tamaño 10 y se tuvieran tres caras y siete cruces, quizás la alarma fuera injustificada: no parece que haya suficiente evidencia como para descartar, sin miramientos, la posibilidad de que la moneda sea 50 %-50 %. Sin embargo, si en 1000 lanzamientos obtenemos 300 caras, entonces sí, la intuición (y si no, nuestro sabio consejo) nos dice que es prácticamente imposible que tal resultado provenga de una moneda regular, lo que nos induciría a rechazar la hipótesis: definitivamente, la moneda estaba cargada.

Así que de lo que se trata es de diseñar un cierto procedimiento, o test, cuya expresión explícita dependerá de la variable X , del estimador del parámetro que se considere, y del tamaño de la muestra, que racionalice el “mosqueo” que la muestra parece producir al respecto de la hipótesis; que cuantifique el peso que la evidencia estadística (la muestra empírica) aporta en contra de la hipótesis de partida.

Note, lector, cómo en esta ilustración estamos insistiendo siempre en la posibilidad de *rechazar* la hipótesis.

7.1.1. Contraste de media de normal, con varianza conocida

Comenzamos, como para los intervalos de confianza, con el ejemplo básico de

contrastar la hipótesis de que la media (esperanza) μ
de una determinada cantidad X en una población que sigue $X \sim \mathcal{N}(\mu, \sigma^2)$
toma un determinado valor μ_0
suponiendo que se conoce la varianza σ^2 .

Aprovecharemos este ejemplo concreto para ir introduciendo la terminología y notación generales.

A. Hipótesis nula

La hipótesis que queremos contrastar, que se conoce como **hipótesis nula** y se denota por H_0 , es en este caso que $\mu = \mu_0$. Escribiremos

$$H_0: \mu = \mu_0.$$

Aquí, μ_0 es un dato. La hipótesis H_0 puede venir de estudios previos, de modelo teórico, de...

B. Test de hipótesis

Para contrastar esta hipótesis realizaremos un test, que se dice **de hipótesis**.

1) Tomamos un **estadístico** para el test, que en este caso va a ser \bar{X} .

2) **Estudio previo.** Si la hipótesis H_0 fuera cierta, entonces la media muestral \bar{X} de una muestra de tamaño n seguiría una normal $\mathcal{N}(\mu_0, \sigma^2/n)$.

Si fijamos α pequeño (digamos $\alpha = 1\%$), entonces

$$\bar{X} \in \left(\mu_0 - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

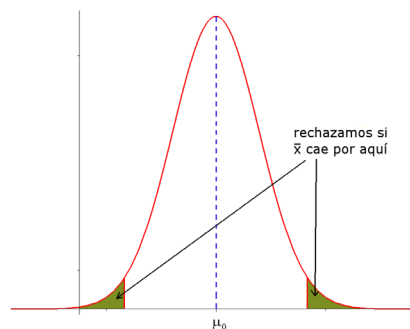
ocurriría con probabilidad de $1 - \alpha$ (digamos, probabilidad de 99%).

3) **Aplicación del test.** Ahora obtenemos una muestra concreta de tamaño n , pongamos (x_1, \dots, x_n) . Su media muestral es \bar{x} .

Si ocurriera que \bar{x} no se encuentra en el intervalo anterior, es decir, si

$$|\bar{x} - \mu_0| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}},$$

habría sucedido un evento extremo, algo que sólo ocurre con probabilidad α (digamos 1 de cada 100). Si así fuera, desconfiaríamos de la hipótesis H_0 . No es que no pueda ocurrir, pero es tan inusual que ocurra (si la hipótesis es cierta) que desconfiamos, además cuantificablemente.



■ Así que, si

$$|\bar{x} - \mu_0| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}},$$

RECHAZAMOS la hipótesis H_0 y decimos que lo hacemos con **nivel de significación** α .

■ Por el contrario, si hubiera ocurrido que $|\bar{x} - \mu_0| \leq z_{\alpha/2} \sigma/\sqrt{n}$, no tendríamos argumentos (evidencia estadística) para rechazar H_0 y ACEPTARÍAMOS H_0 .

Observe, lector, que en este caso, si \bar{x} está significativamente alejado de μ_0 , consideramos que hay evidencia estadística (suficiente) como para rechazar la hipótesis de que $\mu = \mu_0$.

En general, un test de hipótesis rechaza la hipótesis H_0 cuando sucede algo inusualmente extremo, que va muy en contra (probabilísticamente hablando) de la hipótesis.

Los niveles de significación usuales son $\alpha = 5\%, 1\%, 0.1\%$, etc. Una significación del 5% es un estándar de referencia en contrastes de hipótesis de estudios científicos.

C. Hipótesis alternativa

Cuando rechazamos H_0 , en la práctica nos estamos quedando con la llamada **hipótesis alternativa** H_1 , que en este caso es

$$H_1 : \mu \neq \mu_0.$$

Cuando el test rechaza la hipótesis H_0 (y acepta H_1), lo hace con fundamento, pero cuando acepta H_0 lo hace porque no queda más remedio. Más sobre esto, en la discusión sobre tipos de errores en los test¹ de hipótesis del apartado F de esta misma sección.

D. Región de rechazo

La región $\mathcal{R}_\alpha = \{|\bar{x} - \mu_0| > z_{\alpha/2} \sigma / \sqrt{n}\}$ es conocida como la *región de rechazo* del test. Al complementario de \mathcal{R}_α nos referiremos como la *región de aceptación* \mathcal{A}_α del test.

EJEMPLO 7.1.1. *Contraste para la media de una normal de varianza $\sigma^2 = 4$ para una muestra de tamaño 100.*

Tenemos una normal $\mathcal{N}(\mu, \sigma^2)$, con $\sigma^2 = 4$ (y por tanto $\sigma = 2$). La hipótesis nula es $H_0: \mu = \mu_0$, donde $\mu_0 = 1$. Siguiendo el análisis anterior, rechazamos la hipótesis

$$\text{si } \bar{x} \text{ cumple que } |\bar{x} - \mu_0| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}},$$

es decir, si \bar{x} cae en la región de rechazo.

Supongamos que tenemos 100 muestras de X , y que la media muestral es $\bar{x} = 1.3$. Por un lado,

$$|\bar{x} - 1| = 0.3.$$

Por otro lado, si tomamos $\alpha = 1\%$, entonces

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 2.58 \frac{2}{10} = 0.516.$$

La conclusión es que no podemos rechazar la hipótesis nula y, por tanto, no queda más remedio, aceptamos H_0 .

Si tomamos $\alpha = 5\%$, entonces

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1.96 \frac{2}{10} = 0.392,$$

y de nuevo tendríamos que aceptar H_0 .

¹Esto es un plural. La RAE recomienda usar “test” tanto para el singular como para el plural. Aunque, ejem, también recomienda utilizar términos equivalentes como “prueba” o “examen”.

Sin embargo, si hubiéramos fijado $\alpha = 15\%$, entonces tendríamos que

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1.44 \frac{2}{10} = 0.288,$$

y en este caso sí que rechazaríamos H_0 (con nivel de significación del 15%). ♣

F. Significado de α . Tipos de errores

En un test de hipótesis, se pueden producir cuatro situaciones, recogidas en la siguiente tabla:

	Aceptamos H_0	Rechazamos H_0
H_0 cierta	OK	error tipo 1
H_0 falsa	error tipo 2	OK

Las dos situaciones en las que el test comete error se conocen tradicionalmente (y poco nemotécnicamente) como:

- **Error de tipo 1.** Éste es el error que se comete si *se rechaza la hipótesis H_0 cuando es cierta*.
- **Error de tipo 2.** Éste es el error que se comete cuando *se acepta (o no se rechaza) H_0 cuando es falsa*.

En el test que estamos considerando, la probabilidad de cometer un error del tipo 1 es justamente α , pues si la hipótesis nula $H_0 : \mu = \mu_0$ fuera cierta, entonces \bar{X} se distribuiría como una $\mathcal{N}(\mu_0, \sigma^2/n)$, y la probabilidad de que \bar{X} cayera en la región de rechazo sería exactamente α .

Como habitualmente tomamos α pequeño, el test está diseñado para que este tipo de error sea pequeño: *la probabilidad de equivocarse al rechazar es muy pequeña*.

También interesaría controlar la probabilidad de que el error de tipo 2 fuera pequeño, pero el diseño del test anterior no contempla esta posibilidad.

EJEMPLO 7.1.2. Inocentes a la cárcel o culpables en libertad.

En un juicio se intenta dilucidar si un sospechoso es culpable o inocente de un cierto crimen. Nos preguntamos si la hipótesis que se debe plantear es H_0 : “el sospechoso es culpable”, o H_0 : “el sospechoso es inocente”. Parece asunto intrascendente, pero no.

En el sistema judicial español (y en cualquiera razonable, con la presunción de inocencia como guía) se plantea como hipótesis H_0 : “el sospechoso es inocente”, de manera que la evidencia aportada (las pruebas) han de ser muy contundentes para rechazar H_0 , y condenar al sospechoso. Con una tabla análoga a la anterior,

	Se acepta H_0	Se rechaza H_0
H_0 cierta	Se absuelve a un inocente	Se condena a un inocente
H_0 falsa	Se absuelve a un culpable	Se condena a un culpable

Y se pretende que la probabilidad de cometer un error de tipo 1 (condenar a un inocente) sea muy pequeña. Esto exige que el sistema sea muy garantista, que las pruebas de culpabilidad hayan de ser muy sólidas... lo que conduce, claro, a que en ocasiones algunos culpables queden libres (error de tipo 2). Como es bien sabido. ♣

E. Nivel de significación y p -valor

Si, en el ejemplo numérico 7.1.1 fijamos, digamos, $\alpha = 5\%$, solo rechazaríamos H_0 cuando $|\bar{x} - 1| > 0.39$, es decir, cuando la media muestral estuviera por encima de 1.39 o por debajo de 0.61. Sin embargo, si pusiéramos $\alpha = 1\%$, entonces rechazaríamos solo en el caso en que $|\bar{x} - 1| > 0.52$ (media muestral por encima de 1.52 o por debajo de 0.48). Ahora, para rechazar H_0 , se necesita que \bar{x} se aparte de 1 mucho más que en el caso anterior.

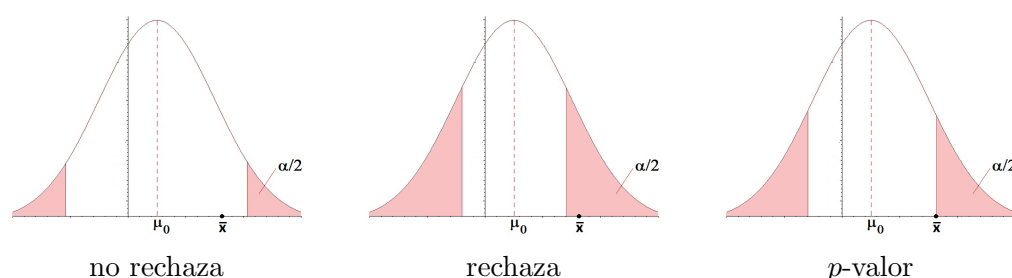
Es decir,

- cuanto más pequeño es el valor de la significación α , más difícil es que el test rechace la hipótesis H_0 ;
- y cuanto mayor sea α , más fácil es que se rechace la hipótesis H_0 .

O, dándole la vuelta, dada la muestra (y el consiguiente valor de \bar{x}): si α es muy grande (es decir, estamos dispuestos a equivocarnos al rechazar en una proporción alta de los casos), entonces es bastante probable que rechacemos la hipótesis. Mientras que si α es muy pequeño, de manera que no queramos equivocarnos casi nunca al rechazar, entonces es muy posible que tengamos que aceptar la hipótesis.

Nos gustaría determinar el *menor* valor de la significación α para el que rechazaríamos la hipótesis H_0 .

Dada la muestra (x_1, x_2, \dots, x_n) de X , al valor de α , digamos α_0 , tal que H_0 se rechaza para $\alpha \geq \alpha_0$ y se acepta para $\alpha < \alpha_0$, se le denomina **p -valor** (de la muestra). Es decir, el p -valor es el más pequeño valor de α para el que la hipótesis se rechaza.



Para una muestra dada, especificar el p -valor informa, de un plumazo, sobre el rechazo o la aceptación de la hipótesis nula para todos los valores de α .

Un rechazo está más justificado cuanto mas pequeño sea el nivel de significación con que se hace. Así que, si el p -valor es muy pequeño, entonces la muestra aporta una evidencia estadística muy contundente *en contra* de la hipótesis que se está contrastando.

EJEMPLO 7.1.3. *Cálculo explícito del p -valor para el contraste de la media en una muestra normal.*

Seguimos con una normal de varianza σ^2 conocida. La hipótesis nula se escribe $H_0: \mu = \mu_0$. Tenemos n muestras, y la media muestral es \bar{x} . La dicotomía aceptación/rechazo es

$$\text{rechazo si } |\bar{x} - \mu_0| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}}; \quad \text{aceptación si } |\bar{x} - \mu_0| \leq z_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

Así que para calcular el p -valor, debemos localizar el valor de α para el que el signo de la desigualdad se invierte. Es decir, hallar el valor de α para el que

$$|\bar{x} - \mu_0| = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \implies z_{\alpha/2} = \frac{\sqrt{n}}{\sigma} |\bar{x} - \mu_0|.$$

Como $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, podemos “despejar” α y obtener una fórmula explícita para el p -valor α_0 :

$$\alpha_0 = 2 \left(1 - \Phi \left(\frac{\sqrt{n}}{\sigma} |\bar{x} - \mu_0| \right) \right).$$

Para la muestra del ejemplo 7.1.1, de tamaño 100 y con media muestral $\bar{x} = 1.3$, tendríamos un p -valor

$$2 \left(1 - \Phi \left(\frac{10}{2} 0.3 \right) \right) = 2(1 - \Phi(3/2)) = 13.36 \%,$$

que es (medianamente) alto. Esto significaría que la evidencia estadística que la muestra aporta en contra de la hipótesis no es muy alta.

Sin embargo, con muestras de tamaño 100, si ocurriera que $\bar{x} = 1.5$, entonces el p -valor sería 1.24 %, y el rechazo estaría más que justificado. ♣

7.1.2. Contraste de proporción

Tratamos ahora otro caso de interés, en el que se desea contrastar la hipótesis de que, en una determinada población, la proporción de individuos con una determinada característica (votantes por un partido, seguidores de un equipo, etc.) es p_0 .

Es decir, tenemos una variable $X \sim \text{BER}(p)$ y queremos contrastar la hipótesis

$$H_0: p = p_0,$$

contra la hipótesis alternativa

$$H_1: p \neq p_0.$$

Si H_0 fuera cierta, es decir, si p_0 fuera el verdadero valor de p , en una muestra aleatoria (X_1, X_2, \dots, X_n) de tamaño n (grande) se tendría que cumplir que \bar{X} se

distribuiría (aproximadamente) como una $\mathcal{N}(p_0, p_0(1-p_0)/n)$, y por tanto se habría de cumplir, con (aproximadamente) probabilidad $1 - \alpha$, que

$$|\bar{X} - p_0| \leq \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{p_0(1-p_0)}.$$

El test de hipótesis consiste entonces en rechazar H_0 cuando en la muestra realizada (x_1, \dots, x_n) se tiene que

$$|\bar{x} - p_0| > \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{p_0(1-p_0)}.$$

La región de rechazo sería la unión de dos intervalos semiinfinitos:

$$\mathcal{R}_\alpha = \left(-\infty, p_0 - \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{p_0(1-p_0)} \right) \cup \left(p_0 + \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{p_0(1-p_0)}, +\infty \right)$$

El estadístico del test es \bar{X} . Obsérvese que, en este caso, \bar{x} es la proporción de la característica en estudio en la muestra.

El test compara la proporción observada \bar{x} con la proporción p_0 esperada/supuesta, y si son muy diferentes, rechaza la suposición.

7.2. Contrastes bilaterales

En los llamados test de contrastes bilaterales, la hipótesis que se desea contrastar es que el parámetro de interés tome un determinado valor. Es decir, la hipótesis nula H_0 es $H_0: \theta = \theta_0$ y la hipótesis alternativa es $H_1: \theta \neq \theta_0$.

Más adelante se analiza el contraste de hipótesis nulas como $H_0: \theta \leq \theta_0$, conocidas como hipótesis unilaterales, que tienen algún ingrediente adicional.

7.2.1. Test bilaterales más usuales

Argumentando como en los dos ejemplos vistos (que, de hecho, son contrastes bilaterales), se obtienen los siguiente test de hipótesis, todos con nivel de significación α . Compare, lector, con los intervalos de confianza análogos de la sección 6.5.1.

Contrastes para una población

Variable	Hipótesis nula	Estad.	Región de rechazo
$\mathcal{N}(\mu, \sigma^2)$	$H_0: \mu = \mu_0$ (σ conocida)	\bar{X}	$ \bar{x} - \mu_0 > z_{\alpha/2} \sigma / \sqrt{n}$
$\mathcal{N}(\mu, \sigma^2)$	$H_0: \mu = \mu_0$ (σ desc.)	\bar{X}	$ \bar{x} - \mu_0 > t_{\{n-1; \alpha/2\}} s / \sqrt{n}$
$\mathcal{N}(\mu, \sigma^2)$	$H_0: \sigma^2 = \sigma_0^2$	S^2	$\frac{(n-1)s^2}{\sigma_0^2} \notin (\chi_{\{n-1; 1-\alpha/2\}}^2, \chi_{\{n-1; \alpha/2\}}^2)$
BER(p)	$H_0: p = p_0$	\bar{X}	$ \bar{x} - p_0 > z_{\alpha/2} \sqrt{p_0(1-p_0)} / \sqrt{n}$
POISS(λ)	$H_0: \lambda = \lambda_0$	\bar{X}	$ \bar{x} - \lambda_0 > z_{\alpha/2} \sqrt{\lambda_0} / \sqrt{n}$

(Nota: el tamaño n de la muestra ha de ser grande en el caso de la Bernoulli y la Poisson.)



Nota 7.2.1. Por si lector quisiera comprobar la validez de la segunda y la tercera regiones de rechazo, le recordamos que

- si X se distribuyera como una normal con media μ_0 , entonces $(\bar{X} - \mu_0)/(S/\sqrt{n})$ se distribuiría como una t de Student con $n - 1$ grados de libertad;
- si X se distribuyera como una normal con varianza σ_0^2 , entonces $(n - 1)S/\sigma_0^2$ se distribuiría como una χ^2 con $n - 1$ grados de libertad.

El argumento para el caso de la Poisson es análogo al de la Bernoulli.

Contrastes para dos poblaciones

Variables	Hipótesis nula	Región de rechazo
$\mathcal{N}(\mu_1, \sigma_1^2), \mathcal{N}(\mu_2, \sigma_2^2)$	$H_0: \mu_1 = \mu_2$ (σ_1, σ_2 conocidas)	$ \bar{x}_1 - \bar{x}_2 > z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$\mathcal{N}(\mu_1, \sigma_1^2), \mathcal{N}(\mu_2, \sigma_2^2)$	$H_0: \mu_1 = \mu_2$ ($\sigma_1 = \sigma_2$ desconocidas)	$ \bar{x}_1 - \bar{x}_2 > t_{\{n_1+n_2-2; \alpha/2\}} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
$\mathcal{N}(\mu_1, \sigma_1^2), \mathcal{N}(\mu_2, \sigma_2^2)$	$H_0: \mu_1 = \mu_2$ ($\sigma_1 \neq \sigma_2$ desconocidas)	$ \bar{x}_1 - \bar{x}_2 > t_{\{f; \alpha/2\}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
$\mathcal{N}(\mu_1, \sigma_1^2), \mathcal{N}(\mu_2, \sigma_2^2)$	$H_0: \sigma_1^2 = \sigma_2^2$	$\frac{s_1^2}{s_2^2} \notin (F_{\{n_1-1; n_2-1; 1-\alpha/2\}}, F_{\{n_1-1; n_2-1; \alpha/2\}})$
$\text{BER}(p_1), \text{BER}(p_2)$	$H_0: p_1 = p_2$	$ \bar{x}_1 - \bar{x}_2 > z_{\alpha/2} \sqrt{\bar{p}(1-\bar{p})} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

donde

- n_1 y n_2 son los tamaños muestrales (que han de ser grandes para el contraste de medias de Bernoullis);
- \bar{x}_1 y \bar{x}_2 son las medias muestrales; s_1^2 y s_2^2 son las cuasivarianzas muestrales;
- $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$;
- f es el entero más próximo a $\frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$;
- $\bar{p} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$.



Nota 7.2.2. Las cuatro primeras regiones de rechazo se obtienen argumentando como en el apartado 6.4.1. Se trata, en particular, de interpretar los resultados de (6.7), (6.9) y (6.10) (para las medias) y (6.11) (para la varianza) tras recitar el mantra: “si la hipótesis nula fuera cierta, entonces...”.

Solo el caso de la diferencia de proporciones requiere un argumento adicional. Veamos. Ponemos n_1 y n_2 grandes. Si suponemos, en (6.12), que $p_1 = p_2$ (digamos que ambos son iguales a p), entonces tendríamos que

$$\bar{X} - \bar{Y} \stackrel{d}{\approx} \sqrt{p(1-p)} \sqrt{1/n_1 + 1/n_2} \mathcal{N}, \quad \text{con } \mathcal{N} \text{ una normal estándar.}$$

Como resulta habitual en el caso de la Bernoulli, queremos sustituir p por una cantidad calculable con la muestra. Pero es que cuando $p_1 = p_2$, los n_1 datos de X_1 y los n_2 dados de X_2 forman una muestra de tamaño $n_1 + n_2$ (de una única Bernoulli) cuya media muestral es $(n_1\bar{x}_1 + n_2\bar{x}_2)/(n_1 + n_2)$.

7.3. Contrastes unilaterales

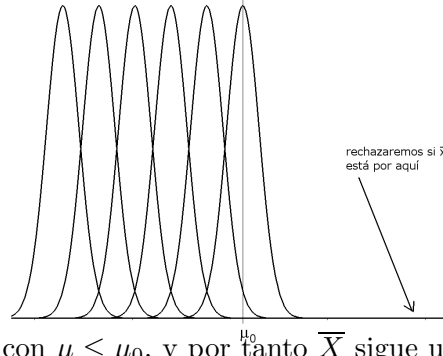
Volvemos al ejemplo básico de una variable $X \sim \mathcal{N}(\mu, \sigma^2)$, con σ^2 conocida y donde estamos contrastando una hipótesis sobre la esperanza μ .

En particular, planteamos ahora, como hipótesis nula,

$$H_0: \mu \leq \mu_0, \quad \text{para un cierto } \mu_0.$$

La hipótesis alternativa es $H_1: \mu > \mu_0$. El estadístico es de nuevo \bar{X} .

Si la hipótesis H_0 es cierta, X sigue una normal $\mathcal{N}(\mu, \sigma^2)$, con una media μ cuyo valor es menor o igual que el μ_0 de referencia. Es decir, \bar{X} es una $\mathcal{N}(\mu, \sigma^2/n)$, con $\mu \leq \mu_0$, y por tanto \bar{X} sigue uno de los modelos representados en la figura.



Esto sugiere rechazar H_0 cuando \bar{x} sea *mucho mayor* que μ_0 . En concreto, y por analogía con el caso bilateral, la región \mathcal{R}_α de rechazo será

$$\bar{x} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}.$$

Ahora la significación α se define como

$$\alpha = \sup_{\mu \leq \mu_0} \mathbf{P}_\mu(\mathcal{R}_\alpha) = \mathbf{P}_{\mu_0}(\mathcal{R}_\alpha),$$

y mide la máxima probabilidad de cometer error de tipo 1, que se alcanza justamente en $\mu = \mu_0$. Véase la nota 7.3.1 siguiente, y también la sección 7.4.

El p -valor es, de nuevo, el valor α_0 de la significación α para el que, *dada la muestra concreta*, se rechaza H_0 si $\alpha \geq \alpha_0$ y no se rechaza para $\alpha \leq \alpha_0$.



Nota 7.3.1. Detalle del significado de α en este test unilateral. Queremos medir la probabilidad de cometer un error de tipo 1, es decir, de rechazar H_0 cuando la hipótesis es cierta. Ahora, nótese, H_0 incluye un rango completo de valores de μ .

Si X tuviera una media μ_0 , entonces la probabilidad de cometer un error de tipo 1 sería

$$\mathbf{P}_{\mu_0}(\mathcal{R}_\alpha) = \mathbf{P}_{\mu_0}(\bar{X} > \mu_0 + z_\alpha \sigma / \sqrt{n}) = \mathbf{P}_{\mu_0}\left(\frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} > z_\alpha\right) = \alpha.$$

Mientras que si X tuviera una media $\tilde{\mu}$, con $\tilde{\mu} < \mu_0$, entonces

$$\mathbf{P}_{\tilde{\mu}}(\mathcal{R}_\alpha) = \mathbf{P}_{\tilde{\mu}}(\bar{X} > \mu_0 + z_\alpha \sigma / \sqrt{n}) = \mathbf{P}_{\tilde{\mu}}\left(\frac{\bar{X} - \tilde{\mu}}{\sigma / \sqrt{n}} > z_\alpha + \frac{\mu_0 - \tilde{\mu}}{\sigma / \sqrt{n}}\right) < \alpha.$$

Nótese el menor estricto.

7.3.1. Diseño de test unilaterales. Elección de hipótesis nula

Seguimos analizando el caso en el que $X \sim \mathcal{N}(\mu, \sigma^2)$, con σ^2 conocida. Tenemos como valor de referencia un cierto μ_0 .

Mantenemos, además, la firme convicción de que el verdadero μ es mayor que ese μ_0 de referencia. Como los test de contraste de hipótesis funcionan “bien” en caso de rechazo, lo natural es plantear en este caso la hipótesis $H_0 : \mu \leq \mu_0$, con la esperanza de que, cuando tengamos la muestra, y por tanto el valor de \bar{x} , podamos rechazar H_0 con un nivel de significación pequeño, y por tanto concluir, confiada y finalmente, que μ es mayor que μ_0 .

Como situación alternativa, digamos que *ya disponemos* de una muestra, y que el valor de \bar{x} es bastante mayor que μ_0 . Esto nos inclina a pensar que, plausiblemente, el μ verdadero sea mayor que μ_0 . De nuevo, el test natural que deberíamos plantear tiene como hipótesis nula a $H_0 : \mu \leq \mu_0$. Si \bar{x} es significativamente mayor que μ_0 , es decir, si \bar{x} cae en la región de rechazo para un α pequeño (o mejor, si el p -valor es pequeño), entonces rechazaremos H_0 con garantía, con confianza, es decir, confirmaremos con fiabilidad que $\mu > \mu_0$.

Nota 7.3.2. Bilateral vs unilateral. Supongamos que $X \sim \mathcal{N}(\mu, \sigma^2)$. La hipótesis de referencia es $\mu = 0$ (y σ es conocido). Pero no nos hemos decantado por aplicar un contraste bilateral o unilateral.

Tenemos en mente el estándar de significación de 5 %. Y disponemos de una muestra con media \bar{x} , positiva y relativamente grande.

Si planteamos un test bilateral, rechazaremos la hipótesis $H_0 : \mu = 0$ si

$$\bar{x} > z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1.96 \frac{\sigma}{\sqrt{n}}.$$

Mientras que si planteamos un test unilateral, rechazaremos la hipótesis $H_0 : \mu \leq 0$ si

$$\bar{x} > z_{\alpha} \frac{\sigma}{\sqrt{n}} = 1.64 \frac{\sigma}{\sqrt{n}}.$$

Como se aprecia, es más fácil rechazar en el unilateral y concluir que $\mu > 0$, que rechazar en el bilateral y concluir que $\mu \neq 0$.

Digamos, por ejemplo, que $\sigma = 2$ y $n = 100$. Así que $1.96\sigma/\sqrt{n} \approx 0.392$, mientras que $1.64\sigma/\sqrt{n} \approx 0.328$. Si la muestra nos diera $\bar{x} = 0.35$, entonces no podríamos rechazar la hipótesis $\mu = 0$, pero sí la hipótesis $\mu \leq 0$.

7.3.2. Test unilaterales más usuales

Contrastes para una población

Variable	Hipótesis nula	Estad.	Región de rechazo
$\mathcal{N}(\mu, \sigma^2)$	$H_0: \mu \geq \mu_0$ (σ conocida)	\bar{X}	$\bar{x} < \mu_0 - z_\alpha \sigma / \sqrt{n}$
$\mathcal{N}(\mu, \sigma^2)$	$H_0: \mu \leq \mu_0$ (σ conocida)	\bar{X}	$\bar{x} > \mu_0 + z_\alpha \sigma / \sqrt{n}$
$\mathcal{N}(\mu, \sigma^2)$	$H_0: \mu \geq \mu_0$ (σ desc.)	\bar{X}	$\bar{x} < \mu_0 - t_{\{n-1; \alpha\}} s / \sqrt{n}$
$\mathcal{N}(\mu, \sigma^2)$	$H_0: \mu \leq \mu_0$ (σ desc.)	\bar{X}	$\bar{x} > \mu_0 + t_{\{n-1; \alpha\}} s / \sqrt{n}$
$\mathcal{N}(\mu, \sigma^2)$	$H_0: \sigma^2 \geq \sigma_0^2$	S^2	$\frac{(n-1)s^2}{\sigma_0^2} < \chi_{\{n-1; 1-\alpha\}}^2$
$\mathcal{N}(\mu, \sigma^2)$	$H_0: \sigma^2 \leq \sigma_0^2$	S^2	$\frac{(n-1)s^2}{\sigma_0^2} > \chi_{\{n-1; \alpha\}}^2$
BER(p)	$H_0: p \geq p_0$	\bar{X}	$\bar{x} < p_0 - z_\alpha \sqrt{p_0(1-p_0)} / \sqrt{n}$
BER(p)	$H_0: p \leq p_0$	\bar{X}	$\bar{x} > p_0 + z_\alpha \sqrt{p_0(1-p_0)} / \sqrt{n}$
POISS(λ)	$H_0: \lambda \geq \lambda_0$	\bar{X}	$\bar{x} < \lambda_0 - z_\alpha \sqrt{\lambda_0} / \sqrt{n}$
POISS(λ)	$H_0: \lambda \leq \lambda_0$	\bar{X}	$\bar{x} > \lambda_0 + z_\alpha \sqrt{\lambda_0} / \sqrt{n}$

(Nota: el tamaño n de la muestra ha de ser grande en el caso de la Bernoulli y la Poisson.)

Contrastes para dos poblaciones

Variables	Hipótesis nula	Región de rechazo
$\mathcal{N}(\mu_1, \sigma_1^2), \mathcal{N}(\mu_2, \sigma_2^2)$	$H_0: \mu_1 \leq \mu_2$ (σ_1, σ_2 conocidas)	$\bar{x}_1 - \bar{x}_2 > z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$\mathcal{N}(\mu_1, \sigma_1^2), \mathcal{N}(\mu_2, \sigma_2^2)$	$H_0: \mu_1 \leq \mu_2$ ($\sigma_1 = \sigma_2$ desconocidas)	$\bar{x}_1 - \bar{x}_2 > t_{\{n_1+n_2-2; \alpha\}} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
$\mathcal{N}(\mu_1, \sigma_1^2), \mathcal{N}(\mu_2, \sigma_2^2)$	$H_0: \mu_1 \leq \mu_2$ ($\sigma_1 \neq \sigma_2$ desconocidas)	$\bar{x}_1 - \bar{x}_2 > t_{\{f; \alpha\}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
$\mathcal{N}(\mu_1, \sigma_1^2), \mathcal{N}(\mu_2, \sigma_2^2)$	$H_0: \sigma_1^2 \leq \sigma_2^2$	$\frac{s_1^2}{s_2^2} > F_{\{n_1-1; n_2-1; \alpha\}}$
BER(p_1), BER(p_2)	$H_0: p_1 \leq p_2$	$\bar{x}_1 - \bar{x}_2 > z_\alpha \sqrt{\bar{p}(1-\bar{p})} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

donde

- n_1 y n_2 son los tamaños muestrales (que han de ser grandes para el contraste de medias de Bernoullis);
- \bar{x}_1 y \bar{x}_2 son las medias muestrales; s_1^2 y s_2^2 son las cuasivarianzas muestrales;
- $s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1) + (n_2-1)}$; $\bar{p} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$.
- f es el entero más próximo a $\frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2 + (s_2^2/n_2)^2}$;

7.4. Test generales, función de potencia

Planteamos a continuación un marco general para analizar test de contrastes de hipótesis. Partimos, como es habitual, de una variable X con función de densidad/masa $f(x; \theta)$, con $\theta \in \Theta$. El espacio de parámetros Θ se divide en Θ_0 y su complementario Θ_1 . La hipótesis nula, que deseamos contrastar, es

$$H_0: \theta \in \Theta_0,$$

y la hipótesis alternativa es

$$H_1: \theta \in \Theta_1.$$

A. Test generales

En un test de hipótesis,

- se selecciona un estadístico $T = h(X_1, \dots, X_n)$;
- y se determina un subconjunto $R \subset \mathbb{R}$ (habitualmente un intervalo, que suele depender de un cierto nivel o umbral).

La operativa del test va como sigue: dada una muestra empírica (x_1, \dots, x_n) de X , obtenemos un valor $t = h(x_1, \dots, x_n)$ del estadístico. Entonces,

- si $t \in R$, entonces se rechaza H_0 (y se acepta H_1);
- si $t \notin R$, entonces se acepta H_0 .

La **región de rechazo** del test es el subconjunto de \mathbb{R}^n dado por

$$\mathcal{R} := \{(x_1, \dots, x_n) : h(x_1, \dots, x_n) \in R\},$$

que abreviaremos a veces como $\mathcal{R} = \{T \in R\}$.

Ya hemos visto, en páginas anteriores, los estadísticos y regiones de rechazo correspondientes a los test más habituales.

Pero, ¿cómo se diseña un test de contraste de hipótesis? Primero, claro, hay que seleccionar un estadístico, uno que creamos que nos puede dar información valiosa sobre el parámetro, y del que conozcamos sus propiedades. Y luego habrá que determinar el subconjunto R , que define la región de rechazo \mathcal{R} del test, de manera que, si la hipótesis fuera cierta, entonces sería poco probable que el estadístico tomara valores en R . Veremos algunas ilustraciones en este mismo apartado. En cualquier caso, en la sección 7.5 presentamos un método general de construcción de test de contraste de hipótesis.

B. Función de potencia

Vamos ahora a dar un paso más allá en el análisis y obtener, para cada posible valor de parámetro θ , la probabilidad de que el test *rechace* la hipótesis nula *suponiendo que ese θ fuera el verdadero valor del parámetro*. Esto es lo que se conoce

como la **función de potencia del test**, $\beta(\theta)$, que a cada $\theta \in \Theta$ le asocia el número (entre 0 y 1)

$$(7.1) \quad \beta(\theta) = \mathbf{P}_\theta(\text{rechazar } H_0) = \mathbf{P}_\theta(T \in R) = \mathbf{P}_\theta(\mathcal{R}).$$

Como ilustraremos en los ejemplos de esta sección, el cálculo explícito de la función de potencia requiere un conocimiento preciso del estadístico, pues necesitamos calcular la probabilidad de que ese estadístico tome valores en una determinada región.



Nota 7.4.1. Observe, lector, que la función de potencia definida en (7.1) se puede escribir en términos de la función de verosimilitud de las muestras de la siguiente manera:

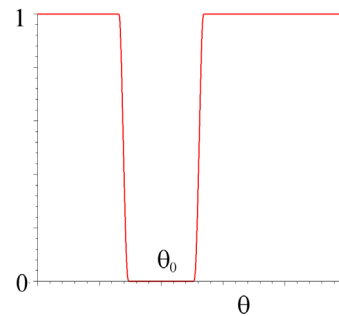
$$(7.2) \quad \beta(\theta) = \mathbf{P}_\theta(\mathcal{R}) = \int_{\mathcal{R}} f(\mathbf{x}; \theta) d\mathbf{x} = \int_{\mathcal{R}} \text{VERO}(\theta; \mathbf{x}) d\mathbf{x},$$

donde $\mathbf{x} = (x_1, \dots, x_n)$. Usaremos esta expresión, por ejemplo, en la discusión del apartado 7.6.

Nótese que con la función $1 - \beta(\theta)$ codificamos la probabilidad de que el test *acepte* la hipótesis suponiendo que ese θ fuera el verdadero valor del parámetro.

De manera que lo ideal sería, claro, que $\beta(\theta)$ tuviera un aspecto como el que se representa en la figura, en la que

- $\beta(\theta)$ es muy próxima a 0 para todo $\theta \in \Theta_0$ (es decir, si la hipótesis es cierta, el test la rechaza con baja probabilidad);
- y que $\beta(\theta)$ es cercana a 1 para todo $\theta \in \Theta_1$ (es decir, si la hipótesis nula es falsa, el test la rechaza con probabilidad cercana a 1).



La primera situación se corresponde con errores de tipo 1. De hecho, la **significación** α del test se define como

$$(7.3) \quad \alpha = \sup_{\theta \in \Theta_0} \beta(\theta),$$

es decir, la “máxima” probabilidad de rechazo suponiendo que la hipótesis fuera cierta (máxima probabilidad de cometer error de tipo 1).

En la segunda situación, cuando $\theta \in \Theta_1$ (y la hipótesis H_0 es falsa), entonces $1 - \beta(\theta)$ nos da la probabilidad, para cada $\theta \in \Theta_1$, de cometer error de tipo 2. Como veremos en los ejemplos que siguen, en los test con un nivel de significación pequeño, la probabilidad de cometer errores de tipo 2 puede llegar a ser relativamente grande.

Los test de las páginas anteriores, que se escriben en términos de un α de referencia, tienen nivel de significación justamente α , en el sentido técnico dado en (7.3), como ilustramos en los dos siguientes ejemplos.



Nota 7.4.2. En un test de contraste de hipótesis general, se plantea una región de rechazo \mathcal{R} , que habitualmente depende de uno o varios umbrales. A posteriori, se calcula su nivel de significación α siguiendo (7.3). Véanse, en los ejemplos 7.4.3 y 7.4.4, un par de ilustraciones.

En los test estándar que hemos descrito en las primeras secciones de este capítulo, se pone por adelantado el valor de α , y se diseña la región de rechazo \mathcal{R}_α para que tenga justamente nivel de significación α .

EJEMPLO 7.4.1. *Función de potencia para el test habitual de la hipótesis $H_0: \mu = \mu_0$ para $X \sim \mathcal{N}(\mu, \sigma^2)$, con σ conocida.*

Para muestras (X_1, \dots, X_n) de X de tamaño n , consideramos el estadístico \bar{X} . Dado un $\alpha \in (0, 1)$, y como ya hemos visto, la región de rechazo es

$$\mathcal{R}_\alpha = \left\{ |\bar{X} - \mu_0| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\}$$

Para calcular la función de potencia, recordamos que, si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces

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

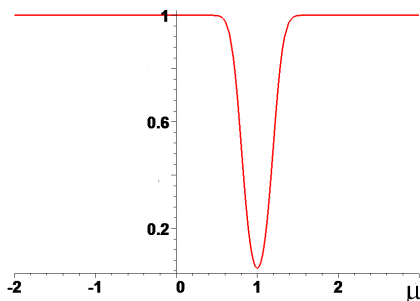
Usando esto, tenemos que

$$\begin{aligned} \beta(\mu) &= \mathbf{P}_\mu(\mathcal{R}_\alpha) = \mathbf{P}_\mu(|\bar{X} - \mu_0| > z_{\alpha/2} \sigma/\sqrt{n}) \\ &= 1 - \mathbf{P}_\mu(|\bar{X} - \mu_0| \leq z_{\alpha/2} \sigma/\sqrt{n}) \\ &= 1 - \mathbf{P}_\mu(-z_{\alpha/2} \sigma/\sqrt{n} \leq \bar{X} - \mu_0 \leq z_{\alpha/2} \sigma/\sqrt{n}) \\ &= 1 - \mathbf{P}_\mu((\mu_0 - \mu) - z_{\alpha/2} \sigma/\sqrt{n} \leq \bar{X} - \mu \leq (\mu_0 - \mu) + z_{\alpha/2} \sigma/\sqrt{n}) \\ &= 1 - \mathbf{P}_\mu\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\alpha/2}\right) \\ &= 1 - \left[\Phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\alpha/2}\right) - \Phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - z_{\alpha/2}\right) \right]. \end{aligned}$$

La función $\beta(\mu)$ tiende a 1 cuando $\mu \rightarrow \pm\infty$. Y además

$$\beta(\mu_0) = 1 - \Phi(z_{\alpha/2}) + \Phi(-z_{\alpha/2}) = 1 - (1 - \alpha/2) + \alpha/2 = \alpha,$$

de manera que el nivel de significación del test es α , como debe ser.



La figura de la izquierda muestra el aspecto de la función de potencia $\beta(\mu)$ del test para los valores $\mu_0 = 1$, $\sigma = 1$, $\alpha = 5\%$ y $n = 100$. Obsérvese cómo β es pequeño en el entorno de μ_0 . Nótese también que si el verdadero μ estuviera cerca de μ_0 , pero no fuera *exactamente* μ_0 , la hipótesis nula sería falsa, pero el test la aceptaría con una (relativa) alta probabilidad. Si por el contrario μ estuviera lejos de μ_0 , entonces el test cometería un error de tipo 2 con baja probabilidad. ♣

EJEMPLO 7.4.2. *Función de potencia para el test habitual de la hipótesis $H_0: \mu \leq \mu_0$ para $X \sim \mathcal{N}(\mu, \sigma^2)$, con σ conocida.*

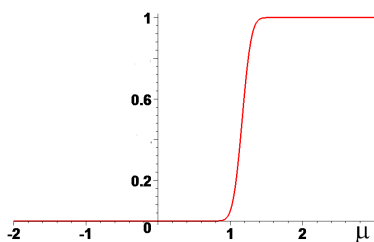
El estadístico es de nuevo \bar{X} . Dado un $\alpha \in (0, 1)$, la región de rechazo es

$$\mathcal{R}_\alpha = \left\{ \bar{X} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \right\}$$

Procediendo de manera análoga a la del ejemplo anterior, obtenemos la siguiente expresión para la función de potencia del test:

$$\begin{aligned} \beta(\mu) &= \mathbf{P}_\mu(\mathcal{R}_\alpha) = \mathbf{P}_\mu(\bar{X} > \mu_0 + z_\alpha \sigma / \sqrt{n}) = \mathbf{P}_\mu(\bar{X} - \mu > (\mu_0 - \mu) + z_\alpha \sigma / \sqrt{n}) \\ &= \mathbf{P}_\mu\left(\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} > \frac{\mu_0 - \mu}{\sigma / \sqrt{n}} + z_\alpha\right) = 1 - \Phi\left(\frac{\mu_0 - \mu}{\sigma / \sqrt{n}} + z_\alpha\right), \end{aligned}$$

en términos, de nuevo, de la función Φ de distribución de una normal estándar.



Representamos a la izquierda el aspecto de la función $\beta(\mu)$ (de nuevo para los valores $\mu_0 = 1$, $\sigma = 1$, $\alpha = 5\%$ y $n = 100$). La función β es pequeña, casi 0, a la izquierda de μ_0 , como esperábamos. El nivel de significación viene dado por $\sup_{\mu \leq \mu_0} \beta(\mu) = \beta(\mu_0) = 1 - \Phi(z_\alpha) = \alpha$, como (de nuevo) debe ser.

Como en el ejemplo anterior, si μ fuera ligeramente mayor que μ_0 , la hipótesis sería falsa, pero el test la aceptaría con alta probabilidad. ♣

EJEMPLO 7.4.3. *Control de calidad: un test para proporciones.*

En un proceso de tallado de diamantes, una proporción p de ellos resultan ser defectuosos. En el control de calidad de la empresa se considera intolerable una proporción de piezas defectuosas superior al 5%.

La hipótesis nula que planteamos es

$$H_0: p \leq 5\%.$$

Si esta hipótesis se rechazara (con un nivel de significación pequeño), entonces la empresa debería proceder a revisar todo su proceso de fabricación.

Para contrastar la hipótesis anterior sólo podremos analizar cinco piezas. Observe, lector, que no estamos en territorio de la normal, y que tampoco el tamaño de la muestra permite estimaciones asintóticas como las de los contrastes de proporciones del apartado 7.3.2.

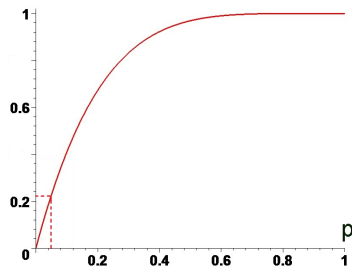
Diseñamos el test como sigue: se analizan cinco piezas, y si en esa muestra encontramos *al menos* una defectuosa, entonces rechazaremos H_0 .

El test parece razonable, pues una pieza (o más) defectuosas de cada cinco apunta a una proporción mucho mayor del 5% (o menor) supuesto.

Digamos que X_1, \dots, X_5 son las variables que registran si cada piedra analizada es defectuosa o no. Son variables independientes, que siguen una distribución $\text{BER}(p)$. Su suma $Z = X_1 + \dots + X_5$, que registra el número de piezas defectuosas, es una variable $\text{BIN}(5, p)$.

La región de rechazo se escribe ahora

$$\mathcal{R} = \{Z \geq 1\}.$$



La función de potencia del test es

$$\beta(p) = \mathbf{P}_p(Z \geq 1) = 1 - \mathbf{P}_p(Z = 0) = 1 - (1 - p)^5.$$

A la izquierda hemos dibujado esta función de potencia. Como $\beta(p)$ es una función creciente, la significación del test es $\sup_{p \leq 5\%} \beta(p) = \beta(5\%) = 22.62\%$, un poco (bastante) alta.

Buscando un mejor nivel de significación, rediseñamos el test y decidimos, alternativamente, rechazar la hipótesis si encontramos en la muestra de cinco piezas *al menos* dos defectuosas. La región de rechazo sería ahora $\mathcal{R} = \{Z \geq 2\}$, y la función de potencia del test sería

$$\beta(p) = \mathbf{P}_p(Z \geq 2) = 1 - \mathbf{P}_p(Z = 0) - \mathbf{P}_p(Z = 1) = 1 - (1 - p)^5 - 5p(1 - p)^4,$$

cuyo aspecto es similar a la de la gráfica anterior, pero que nos daría un nivel de significación de $\sup_{p \leq 5\%} \beta(p) = \beta(5\%) = 2.26\%$, mucho menor que el anterior.

Obsérvese que, para los dos test, pero más pronunciadamente en el segundo, la probabilidad de error de tipo 2 puede llegar a ser muy grande. Por ejemplo, si p fuera del 6%, la hipótesis sería falsa, pero el test la aceptaría con una (altísima) probabilidad del $1 - \beta(6\%) = 96.81\%$. ♣

EJEMPLO 7.4.4. Test para la uniforme.

Digamos que $X \sim \text{UNIF}[0, a]$, donde a es un parámetro positivo. Tratamos primero la hipótesis

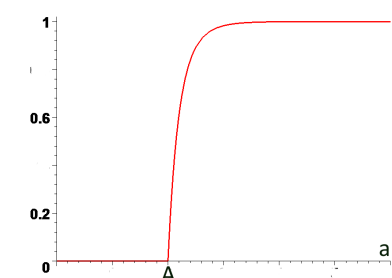
$$H_0 : a \leq A,$$

donde A es un cierto valor dado. Digamos que tenemos una muestra (x_1, \dots, x_n) . Obsérvese que en cuanto haya *algún* valor x_i mayor que A , la hipótesis no puede ser cierta. Esto sugiere definir la siguiente region de rechazo:

$$\mathcal{R} = \{\max(X_1, \dots, X_n) > A\},$$

que describimos usando el máximo de las X_i . La función de potencia asociada es

$$\beta(a) = \mathbf{P}_a(\max(X_1, \dots, X_n) > A) = 1 - \mathbf{P}_a(\max(X_1, \dots, X_n) \leq A) = 1 - \mathbf{P}_a(X \leq A)^n.$$



Obsérvese que $\mathbf{P}_a(X \leq A) = 1$ si $a < A$, y que $\mathbf{P}_a(X \leq A) = A/a$ si $a \geq A$. Lo que nos da, finalmente, que

$$\beta(a) = \begin{cases} 0, & \text{si } a \leq A, \\ 1 - (A/a)^n, & \text{si } a > A. \end{cases}$$

A la izquierda dibujamos la función de potencia del test. El nivel de significación es $\alpha = 0$, pues $\beta(a) = 0$ si $a \leq A$; es decir, no es posible cometer errores de tipo 1 con este test.

Planteamos ahora la hipótesis

$$H_0 : a \geq A.$$

Ahora la hipótesis H_0 sugiere que el parámetro a es “grande” (mayor que A), así que parece razonable rechazarla si en la muestra observáramos “muchos” valores relativamente “pequeños”.

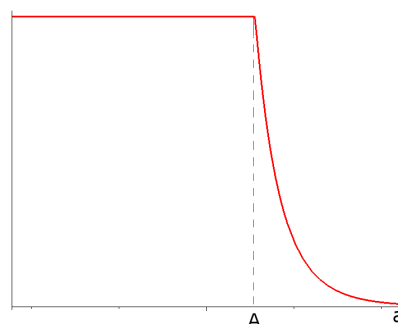
Por ejemplo, podríamos considerar la siguiente región de rechazo:

$$\mathcal{R}_1 = \{\max(X_1, \dots, X_n) < A\}.$$

Su función de potencia es

$$\begin{aligned} \beta(a) &= \mathbf{P}_a(\max(X_1, \dots, X_n) < A) \\ &= \begin{cases} 1, & \text{si } a \leq A, \\ (A/a)^n, & \text{si } a > A. \end{cases} \end{aligned}$$

cuyo aspecto dibujamos a la derecha. Esta región de rechazo tendría nivel de significación $\alpha = \sup_{a \geq A} \beta(a) = \beta(A) = 1$, que es a todas luces insensato.



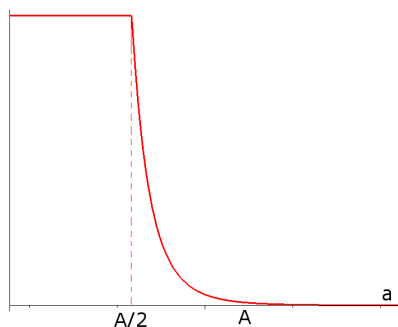
Nos ponemos más exigentes, proponiendo, como región de rechazo, a

$$\mathcal{R}_2 = \{\max(X_1, \dots, X_n) < A/2\}.$$

Ahora la función de potencia es

$$\begin{aligned} \beta(a) &= \mathbf{P}_a(\max(X_1, \dots, X_n) < A/2) \\ &= \begin{cases} 1, & \text{si } a \leq A/2, \\ (A/(2a))^n, & \text{si } a > A/2. \end{cases} \end{aligned}$$

cuyo aspecto se dibuja a la derecha, y que nos daría un nivel de significación $\alpha = \sup_{a \geq A} \beta(a) = \beta(A) = 1/2^n$, que es muy pequeño en cuanto n sea moderadamente grande.



Quizás demasiado pequeño, ¿verdad, lector? Quizás ese factor $1/2$ que hemos incluido en la región de rechazo haya sido excesivo, y baste únicamente con exigir que el máximo esté “un poco” a la izquierda de A para obtener niveles de significación razonables, del orden del 5%, o del 1%. Animamos al lector a que se haga ya sus cálculos, o que visite el ejemplo 7.5.4. ♣

7.5. Test de razón de verosimilitudes

En las ilustraciones anteriores, los estadísticos para el test y las correspondientes regiones de rechazo venían dadas, bien por estudios previos teóricos, bien por análisis *ad hoc* de la variable y de su dependencia del parámetro en cuestión.

Presentamos a continuación un método general para la construcción de test para contrastes de hipótesis.

Partimos, como es habitual, de una variable X con función de densidad/masa $f(x; \theta)$, donde $\theta \in \Theta$. El espacio de parámetros se parte (partición disjunta) en $\Theta = \Theta_0 \cup \Theta_1$. La hipótesis que deseamos contrastar es

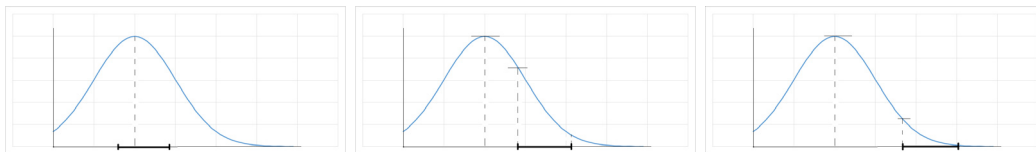
$$H_0: \theta \in \Theta_0.$$

Supongamos que disponemos de una muestra (x_1, \dots, x_n) de X . Asociada a esa muestra, tenemos la función de verosimilitud,

$$\text{VERO}(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta),$$

definida para cada $\theta \in \Theta$. Imaginemos que esta función de verosimilitud de la muestra tiene un aspecto como en el que se muestra en las figuras bajo estas líneas, con un único máximo, al que designaríamos como $\hat{\theta}$ si estuviéramos en el apartado 5.2.2 sobre estimación por máxima verosimilitud.

En las tres figuras se plantean tres posibles situaciones. En la primera, el intervalo Θ_0 que se desea contrastar, marcado en horizontales con una línea gruesa, incluye a ese máximo: esta situación apunta a que la hipótesis H_0 es plausible, dada la muestra. En las otras dos, la muestra (y su verosimilitud) parecen indicar que la hipótesis no es muy creíble. Muy poco, de hecho, en la última figura.



Para cuantificar esta impresión, podemos comparar el valor máximo de la verosimilitud en todo el espacio de parámetros (la altura máxima de la gráfica) con la máxima verosimilitud obtenida cuando θ se mueve en Θ_0 . El cociente de estas cantidades es 1 en la primera figura, algo más pequeño en la segunda, y muy pequeño en la tercera. El cociente al que nos estamos refiriendo, y que vamos a llamar *razón de verosimilitudes*² RV se escribe como

$$\text{RV}(x_1, \dots, x_n) = \frac{\sup_{\theta \in \Theta_0} \text{VERO}(\theta; x_1, \dots, x_n)}{\sup_{\theta \in \Theta} \text{VERO}(\theta; x_1, \dots, x_n)}.$$

²Aunque en realidad es un cociente entre *supremos* de verosimilitudes. Por cierto, en todas las ilustraciones que veremos, estos supremos son *máximos*.

Obsérvese que $RV \in (0, 1]$ y que RV depende de la muestra; pero no del parámetro θ . Un valor de RV grande, cercano a 1 (o directamente 1), apunta a que H_0 es aceptable; mientras que si RV es pequeño, tenderemos a rechazar H_0 . Sólo queda establecer ese umbral de aceptación/rechazo en función del valor de RV .

Fijamos un cierto valor $c \in (0, 1)$, al que nos referiremos como el **calibre**. El test de razón de verosimilitudes (de calibre c) viene determinado por la siguiente **región de rechazo**:

$$\mathcal{R}_c = \{RV(x_1, \dots, x_n) \leq c\}.$$

Es decir, que rechazamos H_0 si la muestra es poco verosímil (en términos relativos); cuando RV sea menor que el umbral c fijado a priori.

Obsérvese que hay un test distinto para cada valor de c .

Analizamos ahora las propiedades del test: su función de potencia y su nivel de significación, y para ello

- elevamos notación, pasando de muestras (x_1, \dots, x_n) numéricas concretas a vectores (X_1, \dots, X_n) , que recogen todo el universo de posibles muestras;
- consideramos la función (aleatoria) de verosimilitud $\text{VERO}(\theta; X_1, \dots, X_n)$;
- seleccionamos como estadístico del test a la “razón de verosimilitud(es)”

$$RV(X_1, \dots, X_n) = \frac{\sup_{\theta \in \Theta_0} \text{VERO}(\theta; X_1, X_2, \dots, X_n)}{\sup_{\theta \in \Theta} \text{VERO}(\theta; X_1, X_2, \dots, X_n)}$$

(que es realmente un estadístico, pues en su definición no interviene el parámetro θ);

- definimos la región de rechazo

$$\mathcal{R}_c = \{RV(X_1, \dots, X_n) \leq c\}.$$

para el calibre $c \in (0, 1)$;

- y finalmente calculamos su función de potencia

$$\beta_c(\theta) = \mathbf{P}_\theta(RV(X_1, \dots, X_n) \leq c),$$

y el nivel de significación asociado, dado por $\sup_{\theta \in \Theta_0} \beta_c(\theta)$.

Vemos a continuación unas cuantas ilustraciones. Algunas de ellas no son más que los test de secciones anteriores.

Cada uno de estos ejemplos consta de dos partes: primero, dada una muestra (x_1, \dots, x_n) , habremos de escribir una expresión explícita para el estadístico $RV(x_1, \dots, x_n)$. Lo más simplificada posible para facilitar, segunda parte, el cálculo de la función de potencia (y del nivel de significación) del test; para este segundo cálculo, claro, trataremos vectores aleatorios (X_1, \dots, X_n) , y no muestras concretas (x_1, \dots, x_n) .

EJEMPLO 7.5.1. *Test de razón de verosimilitudes para la hipótesis $H_0 : \mu = \mu_0$ en una variable $X \sim \mathcal{N}(\mu, \sigma^2)$, con σ^2 conocida.*

En la notación general, $\Theta = \mathbb{R}$ y $\Theta_0 = \{\mu_0\}$. Dada una muestra aleatoria (x_1, \dots, x_n) de X ,

$$\text{VERO}(\mu; x_1, \dots, x_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2\right).$$

Como (ejemplo 5.2.11) la estimación por máxima verosimilitud de μ es \bar{x} ,

$$\begin{aligned} \sup_{\mu \in \mathbb{R}} \text{VERO}(\mu; x_1, \dots, x_n) &= \text{VERO}(\bar{x}; x_1, \dots, x_n) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \bar{x})^2\right). \end{aligned}$$

Por consiguiente,

$$\text{RV}(x_1, \dots, x_n) = \frac{\text{VERO}(\mu_0; x_1, \dots, x_n)}{\text{VERO}(\bar{x}; x_1, \dots, x_n)} = \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{j=1}^n [(x_j - \mu_0)^2 - (x_j - \bar{x})^2]\right)\right),$$

que tras álgebra simplificadora se reduce a

$$\text{RV}(x_1, \dots, x_n) = \exp\left(-\frac{n}{2\sigma^2} (\bar{x} - \mu_0)^2\right).$$

Con calibre $c \in (0, 1)$ obtenemos el test con región de rechazo

$$\mathcal{R}_c = \{\text{RV}(x_1, \dots, x_n) \leq c\} = \left\{|\bar{x} - \mu_0| \geq \frac{\sigma}{\sqrt{n}} \sqrt{2\ln(1/c)}\right\}.$$

Así que el test de razón de verosimilitudes con calibre c es el test bilateral usual con $z_{\alpha/2} = \sqrt{2\ln(1/c)}$.

Aprovechamos los cálculos del ejemplo 7.4.1, y en particular que el test usual tiene nivel de significación justamente α .

La función de potencia del test es la que aparecía allí, salvo que cambiamos $z_{\alpha/2}$ por $\sqrt{2\ln(1/c)}$.

En cuanto al nivel de significación, obsérvese que la relación entre α (del test bilateral usual) y el calibre c de este test se puede escribir como

$$(\star) \quad c = e^{-\frac{1}{2} z_{\alpha/2}^2} \quad \text{o bien} \quad (\star\star) \quad \alpha = 2[1 - \Phi(\sqrt{2\ln(1/c)})],$$

usando que $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$. La expresión $(\star\star)$ nos da el nivel de significación del test para un c dado. O, quizás más interesante, si fijamos un valor, por ejemplo el 5%, entonces la expresión (\star) nos dice qué valor debemos elegir para el calibre c de forma que el test tenga justamente nivel de significación del 5%. ♣

EJEMPLO 7.5.2. *Test de razón de verosimilitudes para la hipótesis $\mu \leq \mu_0$ en una variable $X \sim \mathcal{N}(\mu, \sigma^2)$, con σ^2 conocida.*

Ahora $\Theta_0 = (-\infty, \mu_0]$. Dada una muestra (x_1, \dots, x_n) , la función

$$\text{VERO}(\mu; x_1, \dots, x_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2\right)$$

tiene un máximo en \bar{x} . El exponente de VERO, como función de μ , es una parábola (invertida) con máximo en \bar{x} .

Si $\bar{x} \leq \mu_0$, entonces el lugar donde alcanza el máximo global la función VERO está en Θ_0 y, por tanto, numerador y denominador de RV coinciden. Por tanto, $\text{RV} = 1$ si $\bar{x} \leq \mu_0$.

Por el contrario, si $\bar{x} > \mu_0$, el supremo del numerador se alcanza en μ_0 , lo que nos dice que, en este caso,

$$\text{RV}(x_1, \dots, x_n) = \frac{\text{VERO}(\mu_0, x_1, \dots, x_n)}{\text{VERO}(\bar{x}, x_1, \dots, x_n)}.$$

En suma,

$$\text{RV}(x_1, \dots, x_n) = \begin{cases} 1, & \text{si } \bar{x} \leq \mu_0, \\ \exp\left(-\frac{n}{2\sigma^2} (\bar{x} - \mu_0)^2\right), & \text{si } \bar{x} > \mu_0. \end{cases}$$

Fijamos un calibre $c \in (0, 1)$. Obsérvese que para que $\text{RV}(x_1, \dots, x_n)$ sea $\leq c$ han de cumplirse dos condiciones: primero, que $\bar{x} > \mu_0$ (pues en caso contrario $\text{RV} = 1$), y además que la exponencial en la expresión anterior es $\leq c$.

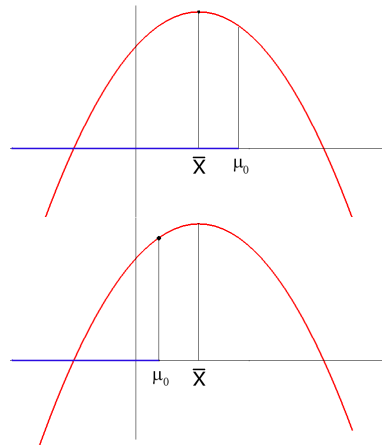
Así que la región de rechazo se puede escribir como

$$\begin{aligned} \mathcal{R}_c &= \{\text{RV}(x_1, \dots, x_n) \leq c\} = \{\bar{x} > \mu_0, \exp\left(-\frac{n}{2\sigma^2} (\bar{x} - \mu_0)^2\right) \leq c\} \\ &= \{\bar{x} > \mu_0, |\bar{x} - \mu_0| \geq \frac{\sigma}{\sqrt{n}} \sqrt{2 \ln(1/c)}\} = \{\bar{x} - \mu_0 \geq \frac{\sigma}{\sqrt{n}} \sqrt{2 \ln(1/c)}\} \end{aligned}$$

(en el último paso quitamos el valor absoluto aprovechando la condición $\bar{x} > \mu_0$). Así que, de nuevo, el test de razón de verosimilitudes con calibre c es el test unilateral usual con $z_\alpha = \sqrt{2 \ln(1/c)}$. Véase el ejemplo 7.4.2. ♣

EJEMPLO 7.5.3. *Test de razón de verosimilitud para la hipótesis $a \leq A$ en una variable $X \sim \text{UNIF}[0, a]$.*

Queremos contrastar la hipótesis $H_0: a \leq A$, para un $A > 0$ dado. Es decir, $\Theta_0 = \{a \leq A\}$. Dada una muestra aleatoria (x_1, \dots, x_n) , llamemos $M_n = \max(x_1, \dots, x_n)$.



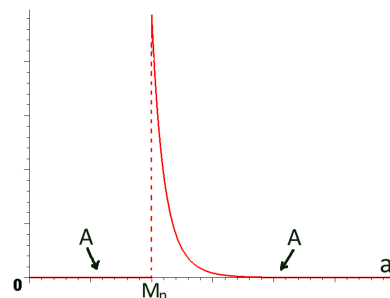
La función de verosimilitud de la muestra es

$$\text{VERO}(a; x_1, \dots, x_n) = \begin{cases} \frac{1}{a^n}, & \text{si } M_n \leq a. \\ 0, & \text{si } M_n > a. \end{cases}$$

cuyo aspecto aproximado dibujamos a la derecha.

Obsérvese que

$$\sup_{a>0} \text{VERO}(a; x_1, \dots, x_n) = \frac{1}{M_n^n},$$



mientras que, argumentando sobre la posición relativa entre A y M_n (véase la figura),

$$\sup_{a \leq A} \text{VERO}(a; x_1, \dots, x_n) = \begin{cases} \frac{1}{M_n^n}, & \text{si } M_n \leq A, \\ 0, & \text{si } M_n > A. \end{cases}$$

Concluimos entonces que

$$\text{RV}(x_1, \dots, x_n) = \begin{cases} 1, & M_n \leq A, \\ 0, & M_n > A. \end{cases}$$

La región de rechazo correspondiente un calibre $c \in (0, 1)$ es

$$\{\text{RV} \leq c\} = \{\text{RV} = 0\} = \{M_n > A\},$$

de manera que sólo hay un test, y se rechaza si $M_n > A$ y se acepta si $M_n \leq A$. Se trata, por cierto, del test del ejemplo 7.4.4; véase allí el aspecto de su función de potencia. ♣

EJEMPLO 7.5.4. *Test de razón de verosimilitud para la hipótesis $a \geq A$ en una variable $X \sim \text{UNIF}[0, a]$.*

Queremos contrastar ahora la hipótesis $H_0: a \geq A$, para un $A > 0$ dado. Escribimos, como antes, $M_n = \max(x_1, \dots, x_n)$.

La función de verosimilitud es la del ejemplo anterior 7.5.3. Argumentando como allí, tenemos que ahora

$$\sup_{a \geq A} \text{VERO}(a; x_1, \dots, x_n) = \begin{cases} \frac{1}{M_n^n}, & \text{si } M_n \geq A, \\ \frac{1}{A^n}, & \text{si } M_n < A. \end{cases}$$

De manera que

$$\text{RV}(x_1, \dots, x_n) = \begin{cases} 1, & \text{si } M_n \geq A, \\ \left(\frac{M_n}{A}\right)^n, & \text{si } M_n < A. \end{cases}$$

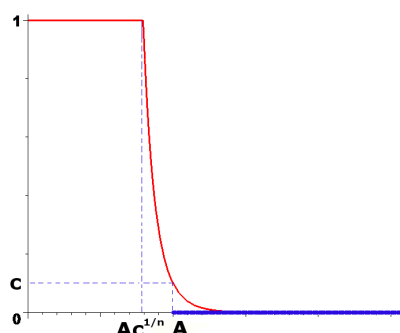
Dado un calibre $c \in (0, 1)$, la región de rechazo del test es

$$\{\text{RV} \leq c\} = \{M_n < A, (M_n/A)^n \leq c\} = \{M_n \leq A c^{1/n}\}.$$

Obsérvese que, para un $c \in (0, 1)$, en cuanto n sea moderadamente grande, $c^{1/n}$ es muy cercano a 1, de manera que rechazaremos la hipótesis cuando el máximo de la muestra esté (un poco) por debajo de A .

La función de potencia resulta ser

$$\begin{aligned} \beta(a) &= \mathbf{P}_a(\text{RV} \leq c) = \mathbf{P}_a(\max(X_1, \dots, X_n) \leq A c^{1/n}) = \mathbf{P}_a(X \leq A c^{1/n})^n \\ &= \begin{cases} 1, & \text{si } a \leq A c^{1/n}, \\ \left(\frac{A c^{1/n}}{a}\right)^n, & \text{si } a > A c^{1/n}. \end{cases} = \begin{cases} 1, & \text{si } a \leq A c^{1/n}, \\ c \left(\frac{A}{a}\right)^n, & \text{si } a > A c^{1/n}. \end{cases} \end{aligned}$$



Dibujamos a la izquierda de estas líneas el aspecto de esta función de potencia $\beta(a)$, que es (vagamente) similar al de la función de potencia del test del ejemplo 7.4.4. El nivel de significación de este test resulta ser justamente el valor del calibre c , puesto que

$$\alpha = \sup_{a \geq A} \beta(a) = \beta(A) = c,$$

sin más que observar que la función $\beta(a)$ es decreciente a partir de $a = A c^{1/n} < A$. ♣

EJEMPLO 7.5.5. *Test de razón de verosimilitudes para la hipótesis $H_0 : \theta \leq \theta_0$ en una variable con función de densidad*

$$f(x; \theta) = \begin{cases} e^{-(x-\theta)}, & \text{si } x > \theta, \\ 0, & \text{si } x \leq \theta, \end{cases}$$

donde θ es un parámetro positivo.

La función de distribución de X viene dada por

$$\mathbf{P}_\theta(X \leq t) = \begin{cases} 0, & \text{si } t \leq \theta, \\ 1 - e^{\theta-t}, & \text{si } t > \theta. \end{cases}$$

Dada una muestra (x_1, \dots, x_n) , llamamos $m_n = \min(x_1, \dots, x_n)$. La función de verosimilitud de la muestra es

$$\text{VERO}(\theta; x_1, \dots, x_n) = \begin{cases} e^{n\theta - \sum_{j=1}^n x_j}, & \text{si } \theta < m_n, \\ 0, & \text{si } \theta \geq m_n. \end{cases}$$

Así que

$$\sup_{\theta \in \mathbb{R}} \text{VERO}(\theta; x_1, \dots, x_n) = e^{-\sum_{j=1}^n x_j + n m_n}$$

Además,

$$\text{si } m_n \leq \theta_0 \quad \text{entonces} \quad \sup_{\theta \leq \theta_0} \text{VERO}(\theta; x_1, \dots, x_n) = e^{-\sum_{j=1}^n x_j + n m_n}$$

$$\text{si } m_n \geq \theta_0 \quad \text{entonces} \quad \sup_{\theta \leq \theta_0} \text{VERO}(\theta; x_1, \dots, x_n) = e^{-\sum_{j=1}^n x_j + n \theta_0}$$

Por tanto,

$$\text{RV}(x_1, \dots, x_n) = \begin{cases} 1, & \text{si } m_n < \theta_0, \\ e^{n(\theta_0 - m_n)}, & \text{si } m_n \geq \theta_0. \end{cases}$$

La región de rechazo en el test de razón de verosimilitudes con calibre c es entonces

$$\mathcal{R}_c = \{\text{RV}(x_1, \dots, x_n) \leq c\} = \left\{ m_n \geq \theta_0, e^{n(\theta_0 - m_n)} \leq c \right\} = \left\{ m_n \geq \theta_0 + \frac{\ln(1/c)}{n} \right\}.$$

La función de potencia β de este test, llamando $h = \theta_0 + \ln(1/c)/n$ y $m_n = \min(X_1, \dots, X_n)$, es Entonces

$$\beta(\theta) = \mathbf{P}_\theta(\text{RV}(X_1, \dots, X_n) \leq c) = \mathbf{P}_\theta(m_n \geq h) = \mathbf{P}_\theta(X \geq h)^n.$$

Como concluimos que

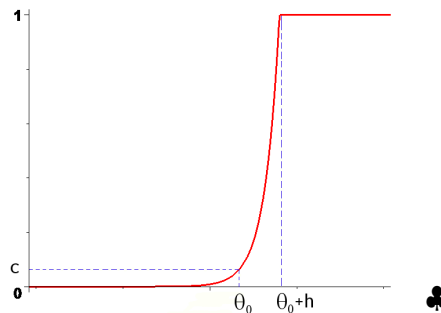
$$\mathbf{P}_\theta(X \geq h) = \begin{cases} 1, & \text{si } \theta \geq h, \\ e^{(\theta - h)}, & \text{si } \theta \leq h. \end{cases}$$

Por consiguiente,

$$\beta(\theta) = \begin{cases} 1, & \text{si } \theta \geq \theta_0 + \ln(1/c)/n, \\ c e^{n(\theta - \theta_0)}, & \text{si } \theta \leq \theta_0 + \ln(1/c)/n. \end{cases}$$

La significación del test resulta ser, de nuevo, el valor del calibre c , pues

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \beta(\theta_0) = c.$$



EJEMPLO 7.5.6. Test de razón de verosimilitudes para la hipótesis $\sigma^2 \leq \sigma_0^2$ en una variable $X \sim \mathcal{N}(\mu, \sigma^2)$, con μ conocida.

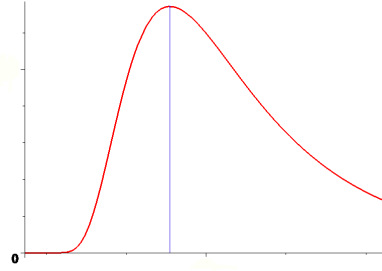
Llamemos, por simplicidad notacional, $\theta = \sigma^2$ al parámetro, y $\theta_0 = \sigma_0^2$ al valor de referencia, de manera que $\Theta_0 = \{\theta \leq \theta_0\}$ es la hipótesis que queremos contrastar. La función de densidad de la variable es

$$f(x; \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2}(x-\mu)^2/\theta}.$$

Dada una muestra (x_1, \dots, x_n) , su función de verosimilitud es

$$\begin{aligned} \text{VERO}(\theta; x_1, \dots, x_n) &= \frac{1}{(2\pi)^{n/2}} \frac{1}{\theta^{n/2}} e^{-\frac{1}{2\theta} \sum_{i=1}^n (x_i - \mu)^2} \\ &= \frac{1}{(2\pi)^{n/2}} \frac{1}{\theta^{n/2}} e^{-\frac{n}{2\theta} \tilde{V}_x}, \end{aligned}$$

donde hemos denotado $\tilde{V}_x = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$. Observe, lector, que *no* se trata de la varianza muestral (que llevaría sumandos del tipo $(x_i - \bar{x})^2$). El aspecto típico de esta función se representa a la derecha. Si el lector se entretiene calculando derivadas, el (único) máximo de esta función se alcanza en $\hat{\theta} = \tilde{V}_x$ (ésta sería la estimación por máxima verosimilitud de θ), de manera que



$$\sup_{\theta} \text{VERO}(\theta; x_1, \dots, x_n) = \text{VERO}(\tilde{V}_x; x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\tilde{V}_x^{n/2}} e^{-n/2}.$$

Por otro lado, como en ejemplos anteriores, argumentando sobre la posición relativa entre θ_0 y \tilde{V}_x , se deduce que

$$\sup_{\theta \leq \theta_0} \text{VERO}(\theta; x_1, \dots, x_n) = \begin{cases} \frac{1}{(2\pi)^{n/2}} \frac{1}{\tilde{V}_x^{n/2}} e^{-n/2} & \text{si } \theta_0 > \tilde{V}_x, \\ \frac{1}{(2\pi)^{n/2}} \frac{1}{\theta_0^{n/2}} e^{-\frac{n}{2} \tilde{V}_x / \theta_0} & \text{si } \theta_0 \leq \tilde{V}_x, \end{cases}$$

lo que nos da que

$$\text{RV}(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } \tilde{V}_x / \theta_0 < 1, \\ (\tilde{V}_x / \theta_0)^{n/2} e^{-\frac{n}{2} (\tilde{V}_x / \theta_0 - 1)} & \text{si } \tilde{V}_x / \theta_0 \geq 1. \end{cases}$$

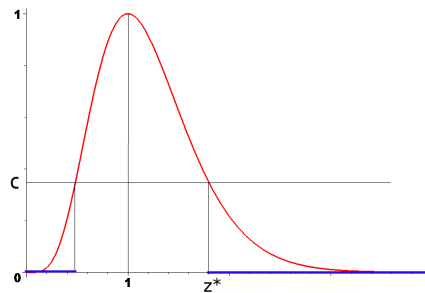
Ahora, dado un calibre $c \in (0, 1)$, obtenemos la región de rechazo

$$\mathcal{R}_c = \{\text{RV} \leq c\} = \left\{ \frac{\tilde{V}_x}{\theta_0} \geq 1, \left(\frac{\tilde{V}_x}{\theta_0} \right)^{n/2} \exp \left(-\frac{n}{2} \left(\frac{\tilde{V}_x}{\theta_0} - 1 \right) \right) \leq c \right\}.$$

Para calcular la función de potencia del test, consideramos la función

$$f(z) = z^{n/2} e^{-\frac{n}{2}(z-1)}$$

que tiene el aspecto que se muestra en la figura de la derecha, con el máximo situado en $z = 1$ (para el que $f(1) = 1$). El conjunto de los valores de z para los que $f(z) \leq c$ es la unión de dos intervalos, marcados en la figura. Si



llamamos z^* a la solución (mayor que 1) de $f(z) = c$, entonces resulta que la región de rechazo anterior se puede escribir como

$$\mathcal{R}_c = \left\{ \frac{\tilde{V}_x}{\theta_0} \geq z^* \right\},$$

una expresión bastante más manejable, salvo por el detalle, no menor, de que el valor de z^* (que depende únicamente de n y c) ha de calcularse numéricamente.

Vamos con la función de potencia. Si $X \sim \mathcal{N}(\mu, \theta)$, si (X_1, \dots, X_n) designa una muestra aleatoria de X , y si llamamos $\tilde{V}_X = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$, entonces

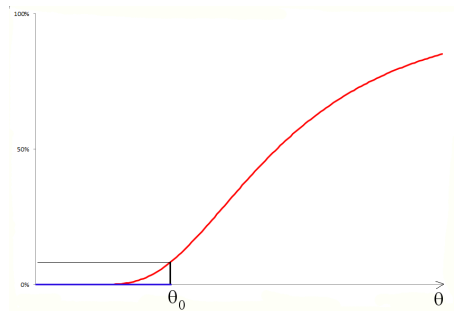
$$n \frac{\tilde{V}_X}{\theta} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sqrt{\theta}} \right)^2 \stackrel{d}{=} \chi_n^2,$$

pues las $(X_i - \mu)/\sqrt{\theta}$ son normales estándar independientes.

De manera que

$$\begin{aligned} \beta(\theta) &= \mathbf{P}_\theta(\mathcal{R}_c) = \mathbf{P}_\theta\left(\frac{\tilde{V}_X}{\theta_0} \geq z^*\right) \\ &= \mathbf{P}_\theta\left(n \frac{\tilde{V}_X}{\theta} \geq n \frac{\theta_0}{\theta} z^*\right) = 1 - F_{\chi_n^2}\left(n \frac{\theta_0}{\theta} z^*\right), \end{aligned}$$

donde $F_{\chi_n^2}$ designa la función de distribución de una variable χ_n^2 . Mostramos a la derecha el aspecto de esta función de potencia. La significación del test es $\sup_{\theta \leq \theta_0} \beta(\theta) = \beta(\theta_0) = \mathbf{P}(\chi_n^2 \geq n z^*)$. Por ejemplo, si $n = 10$ y tomamos un calibre $c = 10\%$, entonces $z^* \approx 2.23$, y la significación es $\alpha = 1.36\%$.♣



EJEMPLO 7.5.7. Test de razón de verosimilitudes para la hipótesis $\sigma^2 \leq \sigma_0^2$ en una variable $X \sim \mathcal{N}(\mu, \sigma^2)$, con μ desconocida.

En el ejemplo anterior hemos supuesto que μ es conocida. Si no disponemos de esta información a priori, el argumento va como sigue. Llamamos, de nuevo, $\theta = \sigma^2$. La función de verosimilitud de una muestra (x_1, \dots, x_n) es la del ejercicio anterior:

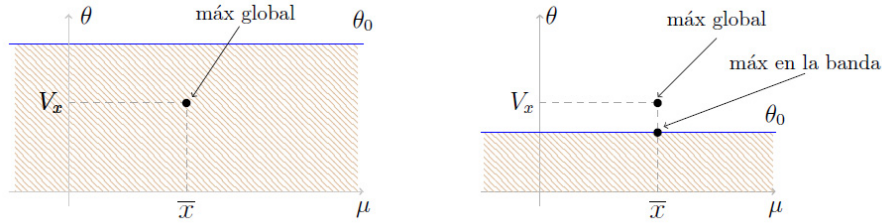
$$(\star) \quad \text{VERO}(\mu, \theta; x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\theta^{n/2}} e^{-\frac{1}{2\theta} \sum_{i=1}^n (x_i - \mu)^2}.$$

Se trata ahora de una función de dos variables, los dos parámetros μ y θ , que “viven” en el semiplano $\mathbb{R} \times \mathbb{R}^+$. La hipótesis nula, por cierto, se refiere sólo a uno de los parámetros, θ . Para calcular la RV, necesitamos calcular

$$\sup_{\mu \in \mathbb{R}, \theta > 0} \text{VERO}(\mu, \theta; x_1, \dots, x_n) \quad \text{y} \quad \sup_{\mu \in \mathbb{R}, 0 < \theta \leq \theta_0} \text{VERO}(\mu, \theta; x_1, \dots, x_n).$$

Veamos. En cuanto a la primera cantidad, como ya sabemos desde el lejano ejemplo 5.2.11, la función de verosimilitud (\star) alcanza su máximo para $\mu = \bar{x}$ y $\theta = V_x = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$. Lector, ahora *sí* que se trata de la varianza muestral.

Para el segundo, distinguimos de nuevo si $\theta_0 > V_x$ o $\theta_0 < V_x$. En el primer caso, el segundo supremo coincide con el primero, como se aprecia en el primer dibujo debajo de estas líneas. El caso en el que $\theta_0 < V_x$ requiere un análisis algo más detallado. Para θ fijo, como sabemos, el máximo de la verosimilitud (\star) se alcanza para $\mu = \bar{x}$; y para $\mu = \bar{x}$, la función (\star) crece desde $\theta = 0$ hasta $\theta = \theta_0$.



De todo este análisis se deduce que

$$RV(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } V/\theta_0 < 1, \\ (V_x/\theta_0)^{n/2} e^{-\frac{n}{2}(V_x/\theta_0 - 1)} & \text{si } V_x/\theta_0 \geq 1. \end{cases}$$

Es decir, como antes, pero sustituyendo \tilde{V}_x por V_x . La región de rechazo del test con calibre c es, pues,

$$\mathcal{R}_c = \{V_x \geq z^* \theta_0\}.$$



Nota 7.5.1. Podemos reescribir la región de rechazo en términos de la cuasivarianza muestral:

$$\mathcal{R}_c = \{V_x \geq z^* \theta_0\} = \{nV_x/\theta_0 \geq nz^*\} = \{(n-1)s^2/\theta_0 \geq nz^*\},$$

para comprobar que se trata de la región de rechazo estándar para la hipótesis $H_0 : \theta \leq \theta_0$ del apartado 7.3.2, cambiando el percentil $\chi^2_{\{n-1; \alpha\}}$ por la cantidad nz^* . ¿Por qué aparece esta χ^2 con $n-1$ grados de libertad? Siga leyendo.

Para la función de potencia, observamos que, por Fisher–Cochran, teorema 4.6,

$$\frac{nD^2}{\theta} = \frac{(n-1)S^2}{\theta} \stackrel{d}{=} \chi^2_{n-1}.$$

Aquí, como hemos hecho desde el capítulo 4, usamos D^2 para referirnos a la variable aleatoria “varianza muestral”: $D^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$. Y sí, lector, la variable de arriba se distribuye como una χ^2 con un grado de libertad menos que antes.

La conclusión es que

$$\beta(\theta) = \mathbf{P}_{\mu, \theta}(D^2 \geq z^* \theta_0) = \mathbf{P}_{\mu, \theta}\left(\frac{nD^2}{\theta} \geq nz^* \frac{\theta_0}{\theta}\right) = \mathbf{P}\left(W \geq nz^* \frac{\theta_0}{\theta}\right),$$

donde W se distribuye como una χ^2_{n-1} . El nivel de significación del test es

$$\sup_{\theta \leq \theta_0} \beta(\theta) = \beta(\theta_0) = \mathbf{P}(W \geq nz^*).$$

Si escogemos c para que nz^* coincida con $\chi^2_{\{n-1; \alpha\}}$, entonces el nivel de significación es justamente α . ♣

7.6. Test uniformemente más potentes

Tenemos dos test para contrastar la hipótesis $\theta \in \Theta_0$. Sus respectivas funciones de potencia son $\beta_1(\theta)$ y $\beta_2(\theta)$.

Para una comparación homogénea, suponemos que ambos test tienen el mismo nivel de significación, es decir, que

$$\sup_{\theta \in \Theta_0} \beta_1(\theta) = \alpha = \sup_{\theta \in \Theta_0} \beta_2(\theta).$$

Decimos que el test 1 es **uniformemente más potente** que el test 2 si

$$\beta_1(\theta) > \beta_2(\theta), \quad \text{para todo } \theta \notin \Theta_0.$$

En otras palabras, cuando la probabilidad de rechazar cuando se debe es siempre mayor en el primer test que en el segundo.

7.6.1. Lema de Neyman–Pearson

En este apartado discutimos un resultado que se conoce como el *lema de Neyman–Pearson*, y que ilustra el principio de que los test de razón verosimilitud son óptimos en el sentido de ser uniformemente más potentes que cualesquiera otros.

Digamos que queremos contrastar una hipótesis nula $H_0: \theta = \theta_0$ contra una hipótesis alternativa $H_1: \theta = \theta_1$. No hay más valores del parámetro, es decir, en este caso, $\Theta = \{\theta_0, \theta_1\}$.

Como referencia tenemos el test de razón de verosimilitudes con un calibre dado $c \in (0, 1)$, cuyo nivel de significación se denota por α .

Dada una muestra (x_1, \dots, x_n) , la región de rechazo (de la hipótesis H_0) de este test es $\mathcal{R} = \{\text{RV}(x_1, \dots, x_n) \leq c\}$, es decir, se rechaza H_0 si

$$\frac{\text{VERO}(\theta_0; x_1, \dots, x_n)}{\max\{\text{VERO}(\theta_0; x_1, \dots, x_n), \text{VERO}(\theta_1; x_1, \dots, x_n)\}} \leq c,$$

que a su vez equivale a rechazar si

$$\text{VERO}(\theta_0; x_1, \dots, x_n) \leq c \text{VERO}(\theta_1; x_1, \dots, x_n).$$

Como α es el nivel de significación, tenemos que

$$\mathbf{P}_{\theta_0}(\text{RV}(X_1, \dots, X_n) \leq c) = \alpha,$$

o en otros términos,

$$\mathbf{P}_{\theta_0}(\text{VERO}(\theta_0; X_1, \dots, X_n) \leq c \text{VERO}(\theta_1; X_1, \dots, X_n)) = \alpha.$$

Denotemos por β a la función de potencia de este test de razón de verosimilitudes, que sólo toma dos valores, $\beta(\theta_0)$ y $\beta(\theta_1)$. Además $\beta(\theta_0) = \alpha$.

Digamos que tenemos un test alternativo para el mismo contraste, con región de rechazo $\tilde{\mathcal{R}}$ y función de potencia $\tilde{\beta}$, y supongamos que *también tiene significación* α ; es decir, que $\tilde{\beta}(\theta_0) = \alpha$.

Pues bien: el LEMA DE NEYMAN–PEARSON afirma que

$$\beta(\theta_1) \geq \tilde{\beta}(\theta_1)$$

sea cual sea ese test alternativo. Es decir, **el test de razón de verosimilitudes es uniformemente más potente que (cualquier) test alternativo.**

Para la DEMOSTRACIÓN, denotemos con $\mathbf{1}_{\mathcal{R}}$ a la función indicadora de la región \mathcal{R} de \mathbb{R}^n de rechazo del test de razón de verosimilitudes, con $\mathbf{1}_{\tilde{\mathcal{R}}}$ a la función análoga, pero para el test alternativo.

Consideremos la integral

$$(\star) \quad I := \int_{\mathbb{R}^n} (\mathbf{1}_{\mathcal{R}}(\mathbf{x}) - \mathbf{1}_{\tilde{\mathcal{R}}}(\mathbf{x})) (c \text{VERO}(\theta_1; \mathbf{x}) - \text{VERO}(\theta_0; \mathbf{x})) d\mathbf{x}.$$

Obsérvese que

- Si $\mathbf{1}_{\mathcal{R}}(\mathbf{x}) = 1$, entonces $c \text{VERO}(\theta_1; \mathbf{x}) - \text{VERO}(\theta_0; \mathbf{x}) \geq 0$, y por tanto

$$(\mathbf{1}_{\mathcal{R}}(\mathbf{x}) - \mathbf{1}_{\tilde{\mathcal{R}}}(\mathbf{x})) (c \text{VERO}(\theta_1; \mathbf{x}) - \text{VERO}(\theta_0; \mathbf{x})) \geq 0.$$

- Si $\mathbf{1}_{\mathcal{R}}(\mathbf{x}) = 0$, entonces $c \text{VERO}(\theta_1; \mathbf{x}) - \text{VERO}(\theta_0; \mathbf{x}) \leq 0$ y por tanto de nuevo

$$(\mathbf{1}_{\mathcal{R}}(\mathbf{x}) - \mathbf{1}_{\tilde{\mathcal{R}}}(\mathbf{x})) (c \text{VERO}(\theta_1; \mathbf{x}) - \text{VERO}(\theta_0; \mathbf{x})) \geq 0.$$

En suma, el integrando de (\star) es siempre no negativo, así que $I \geq 0$.

De manera que, desarrollando la integral que define I , usando la expresión (7.2) de la función de potencia y que $\beta(\theta_0) = \tilde{\beta}(\theta_0) = \alpha$, se tiene que

$$\begin{aligned} 0 &\leq c \int_{\mathcal{R}} \text{VERO}(\theta_1; \mathbf{x}) d\mathbf{x} - \int_{\mathcal{R}} \text{VERO}(\theta_0; \mathbf{x}) d\mathbf{x} \\ &\quad - c \int_{\tilde{\mathcal{R}}} \text{VERO}(\theta_1; \mathbf{x}) d\mathbf{x} + \int_{\tilde{\mathcal{R}}} \text{VERO}(\theta_0; \mathbf{x}) d\mathbf{x} \\ &= c\beta(\theta_1) - \beta(\theta_0) - c\tilde{\beta}(\theta_1) + \tilde{\beta}(\theta_0) = c(\beta(\theta_1) - \tilde{\beta}(\theta_1)), \end{aligned}$$

de donde se concluye que $\beta(\theta_1) \geq \tilde{\beta}(\theta_1)$. ■