Test de hipótesis y Test de bondad de ajuste (Borradores, Curso 23)

Sebastian Grynberg

3-12 de junio de 2013

Que no se oiga ya que los ricos devoran a los pobres,
y que la justicia es sólo para los ricos.
(Manuel Belgrano)

Índice

1. Planteo del problema

1.1. Test de hipótesis

Hipótesis estadística. El punto de partida es una muestra aleatoria $\mathbf{X} = (X_1, \dots, X_n)$ de una variable aleatoria X cuya función de distribución $F_X(x) = \mathbb{P}(X \leq x)$ pertenece a una familia paramétrica de distribuciones de probabilidad, $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$.

En este contexto, una *hipótesis estadística* respecto de la distribución de probabilidades de la variable aleatoria X es una afirmación de la forma siguiente:

$$\text{“}F = F_\theta \text{ para algún } \theta \in \Theta_*\text{”}, \tag{1}$$

donde Θ_* es alguna parte del conjunto paramétrico Θ. Para simplificar la escritura, las hipótesis estadísticas (1) serán denotadas

$$H : \theta \in \Theta_*. \tag{2}$$

El problema general consiste en lo siguiente: en base a los resultados arrojados por la muestra aleatoria $\mathbf{X}$ se quiere decidir entre dos hipótesis estadísticas sobre la distribución de probabilidades de la variable aleatoria X.

Test de hipótesis. Sean Θ_0 y Θ_1 dos subconjuntos del espacio paramétrico tales que $\Theta_0 \cap \Theta_1 = \emptyset$. El problema consiste en decidir entre las dos hipótesis

$$H_0 : \theta \in \Theta_0 \qquad \text{contra} \qquad H_1 : \theta \in \Theta_1,$$

basándose en el conocimiento de una muestra aleatoria, $\mathbf{X} = (X_1, \dots, X_n)$.

Como los valores de θ que no pertenecen a $\Theta_0 \cup \Theta_1$ no se examinan, se puede suponer que $\Theta = \Theta_0 \cup \Theta_1$, y que H_1 es la hipótesis *contraria* de H_0. En tal caso, la hipótesis H_1 se puede escribir en la forma $H_1 : \theta \notin \Theta_0$. La hipótesis H_0 será llamada *hipótesis fundamental* o *hipótesis nula* y las hipótesis de la forma $H : \theta = \theta_1$, para $\theta_1 \in \Theta_1$, se llamarán *alternativas.*

Un *test* (o *regla de decisión*) para decidir entre las dos hipótesis H_0 contra H_1 es una aplicación medible $\delta : \mathbb{R}^n \to \{0, 1\}$ que le asigna a cada posible realización de la muestra aleatoria $\mathbf{x}$ una y sólo una de las hipótesis. Concretamente, $\delta(\mathbf{X})$ es una variable aleatoria a valores en el $\{0, 1\}$. Cuando $\delta(\mathbf{X}) = 1$ se rechazará la hipótesis H_0 a favor de la hipótesis H_1. En cambio, cuando, $\delta(\mathbf{X}) = 0$ se aceptará la hipótesis H_0.

Región crítica. Sea $\delta : \mathbb{R}^n \to \{0, 1\}$ un test para decidir entre las hipótesis H_0 contra H_1. La región del espacio $\mathbb{R}^n$ en la que $\delta(\mathbf{x}) = 1$:

$$\mathcal{R} := \{\mathbf{x} \in \mathbb{R}^n : \delta(\mathbf{x}) = 1\} \tag{3}$$

se denomina *región crítica* o *región de rechazo de la hipótesis fundamental*. La región crítica, $\mathcal{R}$, se identifica con la regla de decisión δ debido a que

$$\delta(\mathbf{x}) = \mathbf{1}\{\mathbf{x} \in \mathcal{R}\}. \tag{4}$$

Tipos de error. Todo test para decidir entre las hipótesis H_0 contra H_1 conduce a decisiones erróneas. Hay dos clases de decisiones erróneas.

- Las llamadas *errores de tipo I* que consisten en *RECHAZAR la hipótesis H_0 cuando ésta es verdadera.*
- Las llamadas *errores de tipo II* que consisten en *ACEPTAR la hipótesis H_0 cuando ésta es falsa.*

Nota Bene. Cuando $\theta \in \Theta_0$, la probabilidad de cometer un error de tipo I será

$$\mathbb{P}(\text{Rechazar } H_0|\theta) = \mathbb{P}(\delta(\mathbf{X}) = 1|\theta) = \mathbb{P}(\mathbf{X} \in \mathcal{R}|\theta).$$

Cuando $\theta \in \Theta_1$, la probabilidad de cometer un error de tipo II será

$$\mathbb{P}(\text{Aceptar } H_0|\theta) = \mathbb{P}(\delta(\mathbf{X}) = 0|\theta) = \mathbb{P}(\mathbf{X} \notin \mathcal{R}|\theta) = 1 - \mathbb{P}(\mathbf{X} \in \mathcal{R}|\theta).$$

Ejemplo 1.1. Sea $\mathbf{X} = (X_1, \dots, X_n)$ una muestra aleatoria de una distribución uniforme sobre el intervalo $(0, \theta)$, $\theta > 0$. Para decidir entre las dos hipótesis

$$H_0 : \theta \geq 2 \qquad \text{contra} \qquad H_1 : \theta < 2$$

consideramos el test $\delta(\mathbf{x}) = 1\{x_{(n)} \leq 3/2\}$, donde $x_{(n)} = \text{máx}(x_1, \dots, x_n)$ y queremos determinar, para cada $\theta > 0$, la probabilidad de decidir erróneamente.

Solución. Para calcular las probabilidades de decidir erróneamente estudiaremos la función $\beta : (0, \infty) \to [0, 1]$ definida por

$$\beta(\theta) = \mathbb{P}(\text{Rechazar } H_0|\theta) = \mathbb{P}(\delta(\mathbf{X}) = 1|\theta) = \mathbb{P}_\theta\left(X_{(n)} \leq \frac{3}{2}\right), \qquad \theta > 0. \tag{5}$$

Sabemos que $Q(\mathbf{X}, \theta) = X_{(n)}/\theta$ es un pivote para θ y que su distribución tiene densidad de probabilidades $f_Q(q) = nq^{n-1}\mathbf{1}\{0 < q < 1\}$. En consecuencia,

$$\begin{aligned}
\beta(\theta) &= \mathbb{P}_\theta\left(X_{(n)} \leq \frac{3}{2}\right) = \mathbb{P}\left(\frac{X_{(n)}}{\theta} \leq \frac{3}{2\theta}\right) = \int_0^{\text{mín}\left(1, \frac{3}{2\theta}\right)} nq^{n-1} dq \\
&= \text{mín}\left(1, \frac{3}{2\theta}\right)^n = \mathbf{1}\left\{0 < \theta \leq \frac{3}{2}\right\} + \left(\frac{3}{2\theta}\right)^n \mathbf{1}\left\{\theta > \frac{3}{2}\right\}. \qquad (6)
\end{aligned}$$

Por lo tanto,

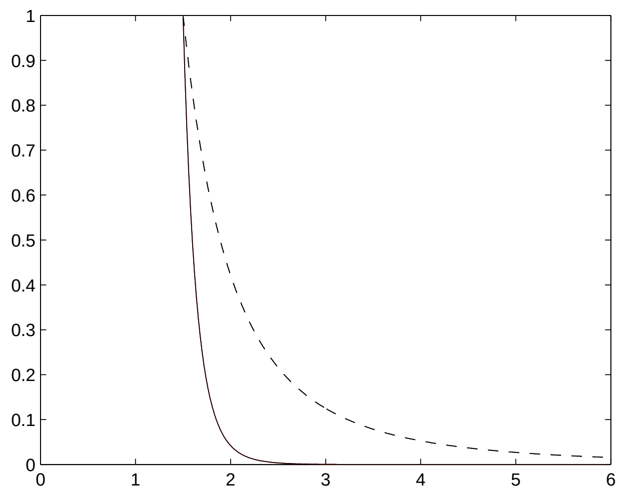

Figura 1: Gráfico de la función $\beta(\theta)$ para distintos volúmenes de muestra: en línea quebrada para volumen $n = 3$; en línea sólida para volumen $n = 11$. Notar que cuando n aumenta disminuyen las probabilidades de los errores de tipo I, pero aumentan las de los errores de tipo II.

- la probabilidad de que ocurra un error de tipo I cuando el verdadero valor del parámetro θ satisface $\theta \geq 2$ es $\beta(\theta) = \left(\frac{3}{2\theta}\right)^n$,
- la probabilidad de que ocurra un error de tipo II cuando el verdadero valor del parámetro θ satisface $\theta \in (0, 3/2]$ es $1 - \beta(\theta) = 1 - 1 = 0$,
- la probabilidad de que ocurra un error de tipo II cuando el verdadero valor del parámetro θ satisface $\theta \in (3/2, 2)$ es $1 - \beta(\theta) = 1 - \left(\frac{3}{2\theta}\right)^n$.

$\square$

1.2. Función de potencia

La calidad de un test de hipótesis $\delta(\cdot)$ se caracteriza por el conjunto de probabilidades de decisiones erróneas (o riesgos de decisión).

Las probabilidades de los errores de un test $\delta(\cdot)$ se pueden representar en el gráfico de la función $\beta : \Theta \to [0, 1]$ definida por

$$\beta(\theta) := \mathbb{P}\left(\text{Rechazar } H_0 | \, \theta\right) = \mathbb{P}\left(\delta(\mathbf{X}) = 1 | \, \theta\right) = \mathbb{P}_\theta\left(\mathbf{X} \in \mathcal{R}\right), \tag{7}$$

llamada la *función de potencia* del test.[1]

[1]En control de calidad, a la función $\mathcal{L}(\theta) = 1 - \beta(\theta)$ se la llama *característica operativa* y su gráfico se llama la *curva característica operativa* del test.

En efecto, la probabilidad de que ocurra un error de tipo I cuando el verdadero valor del parámetro es $\theta \in \Theta_0$ será el valor de la probabilidad $\beta(\theta)$ y la probabilidad de cometer un error de tipo II cuando el verdadero valor del parámetro es $\theta \in \Theta_1$ será el valor de la probabilidad $1 - \beta(\theta)$.

Nota Bene. Una test puede considerarse "bueno" si los valores de su función de potencia están cerca del 0 en la región fundamental Θ_0 y cerca del 1 en la región alternativa Θ_1. En general, establecido el volumen de la muestra, $\mathbf{X} = (X_1, \dots, X_n)$, no es posible construir test capaces de conciliar ambas exigencias.

1.3. Nivel de significación

Sea δ un test para decidir entre las hipótesis $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$. El número

$$\alpha(\delta) = \max_{\theta \in \Theta_0} \beta(\theta) \tag{8}$$

se llama *nivel de significación* del test. Dicho en palabras, el nivel de significación de un test es la máxima probabilidad de rechazar la hipótesis fundamental H_0 cuando ella es verdadera.

Ejemplo 1.2. Sea $\mathbf{X} = (X_1, \dots, X_n)$ una muestra aleatoria de una distribución $\mathcal{U}(0, \theta)$ y sea δ el test definido en el Ejemplo 1.1 para decidir entre las dos hipótesis $H_0 : \theta \geq 2$ contra $H_1 : \theta < 2$.

Debido a que la función de potencia $\beta(\theta)$ es decreciente en θ, el nivel de significación del test es

$$\alpha(\delta) = \max_{\theta \geq 2} \beta(\theta) = \beta(2) = \left(\frac{3}{4}\right)^n.$$

Para que, por ejemplo, el nivel de significación del test sea ≤ 0.05, debe tomarse un volumen de muestra n tal que $(3/4)^n \leq 0.05$. Equivalentemente, $n \geq \log(0.05)/\log(3/4) = 10.413$. Para $n = 11$ el nivel del test resulta $\alpha(\delta) = 0.042...$ $\square$

Comentario sobre el nivel de significación. Utilizar un test de nivel de significación α significa que, en una larga serie de experimentos, no nos equivocaremos al rechazar la hipótesis H_0, siendo que ella es verdadera, más que un $100\,\alpha\,\%$ de los casos. La elección del nivel de significación del test es arbitraria. Habitualmente, en calidad de α se elige alguno de los valores estándar, tales como 0.005, 0.01, 0.05, 0.1. Esta estandarización tiene la ventaja de que permite reducir el volumen de las tablas que se utilizan en el trabajo estadístico.

Nota Bene. La actitud que se tenga hacia la hipótesis fundamental antes de realizar el experimento es una circunstancia importante que puede influir en la elección del nivel de significación. Si se cree firmemente en su veracidad se necesitarán pruebas convincentes en su contra para que se renuncie a ella. En tales condiciones hacen falta criterios de nivel α muy pequeños. Entonces, si la hipótesis fundamental es verdadera, la realización de un valor de muestra perteneciente a la región crítica $\mathcal{R}$ será demasiado inverosímil. La concepción en la que se basa todo el razonamiento es la siguiente: si la probabilidad ϵ de cierto evento A es muy pequeña, consideramos prácticamente imposible el hecho de que este evento ocurra al realizar una sola prueba. Si ocurre, significa que su probabilidad no era tan pequeña.

Máxima potencia. Elegido el nivel de significación α del test de hipótesis, hay que prestarle atención a los valores de su función de *potencia* en la región alternativa Θ_1. Si la potencia en Θ_1 resulta demasiado pequeña, los riesgos de cometer errores de tipo II son muy grandes y tal vez sea conveniente sustituir el nivel de significación por uno mayor. Entre todos los test de nivel α se prefieren aquellos que tengan la *potencia más alta* en toda la región alternativa Θ_1.

1.4. Sobre la construcción de reglas de decisión

En la práctica, las reglas de decisión se construyen basándose en una estadística de la muestra aleatoria $\mathbf{X} = (X_1, \dots, X_n)$, i.e., son de la forma

$$\delta(\mathbf{X}) = \mathbf{1}\{T(\mathbf{X}) \in \mathcal{C}\}, \tag{9}$$

donde $T : \mathbb{R}^n \to \mathbb{R}$ es una función a valores reales y $\mathcal{C}$ es una región de la recta real denominada la *región crítica* o *región de rechazo* del test: si $\delta(\mathbf{X}) = 1$ rechazamos la hipótesis H_0 y si $\delta(\mathbf{X}) = 0$ no la rechazamos.

Nota Bene. La estadística de la muestra, $T(\mathbf{X})$, con la que se construye la regla de decisión (9) debe contener toda la información relevante que hay en la muestra $\mathbf{X}$ para reconstruir el parámetro θ sobre el que recaen las hipótesis H_0 y H_1. Por ejemplo, si se hacen hipótesis sobre la media de la variable aleatoria X, es inútil observar simplemente todos los datos contenidos en la muestra aleatoria $\mathbf{X} = (X_1, \dots, X_n)$. Es intuitivamente claro que si se quiere tomar una decisión entre dos hipótesis sobre la media de una distribución hay que observar el promedio muestral $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$. Si la muestra es suficientemente grande, este valor se no puede desviar demasiado del verdadero valor de la media. Si el desvío fuese desconocido, para tener una idea de su tamaño bastará con observar el valor de la varianza muestral $S^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. Esos dos datos deberían ser suficientes para tomar una decisión sobre una hipótesis sobre la media.

Algunos problemas

1. Dado un test caracterizar su función de potencia, determinar su nivel y los distintos tipos de riesgos estadísticos.

2. Construcción de test prefijando el nivel α y el volumen de la muestra aleatoria n.

3. Construcción de test prefijando el nivel α y la potencia β en alguno de los parámetros alternativos.

Nota Bene. El objetivo de estas notas es presentar una introducción para tratar algunos problemas de carácter muy elemental y el modo de resolverlos mediante razonamientos intuitivos (lo más rigurosos posibles dentro del marco de un curso elemental).[2]

2. Regiones de confianza y test de hipótesis

Supongamos que disponemos de regiones de confianza $\mathcal{S}(\mathbf{X})$ de nivel β para el parámetro θ y queremos construir un test para decidir entre las hipótesis

$$H_0 : \theta = \theta_0 \qquad \text{contra} \qquad H_1 : \theta \neq \theta_0.$$

Debido a que la región de confianza se construye con el objeto de capturar al verdadero valor del parámetro (con alta probabilidad de lograrlo) parece claro que si se observa un resultado $\mathbf{x}$ tal que la región $\mathcal{S}(\mathbf{x})$ contenga a θ_0 deberemos aceptar la hipótesis H_0 y rechazar la contraria H_1. El argumento permite construir el siguiente test

$$\delta(\mathbf{X}) = \mathbf{1}\{\mathcal{S}(\mathbf{X}) \not\ni \theta_0\}.$$

cuyo nivel de significación es

$$\alpha(\delta) = \mathbb{P}\left(\text{Rechazar} H_0 | \theta_0\right) = \mathbb{P}_{\theta_0}\left(\mathcal{S}(\mathbf{X}) \not\ni \theta_0\right) = 1 - \mathbb{P}_{\theta_0}\left(\mathcal{S}(\mathbf{X}) \ni \theta_0\right) = 1 - \beta.$$

$\square$

Usando argumentos similares se obtienen los siguientes resultados.

1. Si $\theta_1(\mathbf{X})$ es una cota inferior de confianza de nivel $1 - \alpha$ para θ, entonces

$$\delta(\mathbf{X}) = \mathbf{1}\{\theta_0 < \theta_1(\mathbf{X})\}$$

es un test de nivel α para decidir entre las hipótesis

$$H_0 : \theta \leq \theta_0 \qquad \text{contra} \qquad H_1 : \theta > \theta_0.$$

2. Si $\theta_2(\mathbf{X})$ es una cota superior de confianza de nivel $1 - \alpha$ para θ, entonces

$$\delta(\mathbf{X}) = \mathbf{1}\{\theta_0 > \theta_2(\mathbf{X})\}$$

es un test de nivel α para decidir entre las hipótesis

$$H_0 : \theta \geq \theta_0 \qquad \text{contra} \qquad H_1 : \theta < \theta_0.$$

[2] Dependiendo de las normas de calidad que se le impongan al test y de la naturaleza de las hipótesis a ser confrontadas, existen metodologías generales para construir test óptimos que pueden consultarse en cualquier libro de *Estadística matemática*. Una exposición rigurosa puede encontrarse en el libro de Borovkov.

3. Si $[\theta_1(\mathbf{X}), \theta_2(\mathbf{X})]$ es un intervalo de confianza de nivel $1-\alpha$ para θ. Entonces

$$\delta(\mathbf{X}) = \mathbf{1}\{[\theta_1(\mathbf{X}), \theta_2(\mathbf{X})] \not\ni \theta_0\}$$

es un test de nivel α para decidir entre las hipótesis

$$H_0 : \theta = \theta_0 \qquad \text{contra} \qquad H_1 : \theta \neq \theta_0.$$

Nota Bene. Notar que en cualquiera de los tres casos se rechaza la hipótesis H_0 cuando y solo cuando los intervalos de confianza están contenidos en la hipótesis alternativa H_1.

3. El método del pivote

Cuando se quieren construir test de hipótesis para el parámetro desconocido θ lo más natural es comenzar la construcción apoyándose en algún estimador puntual del parámetro $\hat{\theta}(\mathbf{X})$ (cuya distribución depende de θ). El *método del pivote* consiste en transformar el estimador $\hat{\theta}(\mathbf{X})$ en un pivote $Q(\hat{\theta}(\mathbf{X}), \theta)$ y utilizarlo para construir el test deseado.

Nota Bene. Por definición, la distribución del pivote $Q(\hat{\theta}(\mathbf{X}), \theta)$ no depende de θ. Para cada $\gamma \in (0,1)$ notaremos mediante q_γ el cuantil-γ del pivote.

En todo lo que sigue vamos a suponer que $Q(\hat{\theta}(\mathbf{X}), \theta)$ es un pivote que goza de las siguientes propiedades:

1. La función de distribución de $Q(\hat{\theta}(\mathbf{X}), \theta)$ es continua y estrictamente creciente.
2. La función $Q(t, \theta)$ es monótona decreciente en θ:
$$\theta_1 < \theta_2 \implies Q(t, \theta_1) > Q(t, \theta_2). \tag{10}$$

3.1. Hipótesis fundamental simple contra alternativa bilateral

Se desea un test de nivel α para decidir entre las hipótesis

$$H_0 : \theta = \theta_0 \qquad \text{contra} \qquad H_1 : \theta \neq \theta_0.$$

Proponemos un test de la forma

$$\delta(\mathbf{X}) = \mathbf{1}\left\{Q(\hat{\theta}(\mathbf{X}), \theta_0) < q_{\gamma_1}\right\} + \mathbf{1}\left\{Q(\hat{\theta}(\mathbf{X}), \theta_0) > q_{\gamma_2}\right\} \tag{11}$$

Como la hipótesis fundamental es de la forma $\theta = \theta_0$ el nivel de significación del test es

$$\begin{aligned}\alpha(\delta) &= \beta(\theta_0) = \mathbb{P}\left(\text{Rechazar } H_0|\theta_0\right) = \mathbb{P}(Q(\hat{\theta}(\mathbf{X}), \theta_0) < q_{\gamma_1}) + \mathbb{P}\left(Q(\hat{\theta}(\mathbf{X}), \theta_0) > q_{\gamma_2}\right) \\ &= \mathbb{P}(Q(\hat{\theta}(\mathbf{X}), \theta_0) \leq q_{\gamma_1}) + 1 - \mathbb{P}\left(Q(\hat{\theta}(\mathbf{X}), \theta_0) \leq q_{\gamma_2}\right) = \gamma_1 + 1 - \gamma_2.\end{aligned}$$

Poniendo $\gamma_1 = \alpha/2$ y $\gamma_2 = 1-\alpha/2$ obtenemos que $\alpha(\delta) = \alpha$. Por lo tanto, el test de hipótesis deseado puede obtenerse de la siguiente manera:

$$\delta(\mathbf{X}) = \mathbf{1}\left\{Q(\hat{\theta}(\mathbf{X}), \theta_0) < q_{\alpha/2}\right\} + \mathbf{1}\left\{Q(\hat{\theta}(\mathbf{X}), \theta_0) > q_{1-\alpha/2}\right\}. \tag{12}$$

□

3.2. Hipótesis fundamental simple contra alternativa unilateral

Se desea un test de nivel α para decidir entre las hipótesis

$$H_0 : \theta = \theta_0 \qquad \text{contra} \qquad H_1 : \theta > \theta_0.$$

Proponemos un test de la forma

$$\delta(\mathbf{X}) = \mathbf{1}\left\{Q(\hat{\theta}(\mathbf{X}), \theta_0) > q_\gamma\right\} \tag{13}$$

Como la hipótesis fundamental es de la forma $\theta = \theta_0$ el nivel de significación del test es

$$\alpha(\delta) = \beta(\theta_0) = \mathbb{P}\left(\text{Rechazar } H_0|\theta_0\right) = \mathbb{P}\left(Q(\hat{\theta}(\mathbf{X}), \theta_0) > q_\gamma\right) = 1 - \gamma.$$

Poniendo $\gamma = 1-\alpha$ obtenemos que $\alpha(\delta) = \alpha$. Por lo tanto, el test deseado puede obtenerse de la siguiente manera:

$$\delta(\mathbf{X}) = \mathbf{1}\left\{Q(\hat{\theta}(\mathbf{X}), \theta_0) > q_{1-\alpha}\right\}. \tag{14}$$

$\square$

3.3. Hipótesis fundamental unilateral contra alternativa unilateral

1.- Como consecuencia de que la función $Q(t, \theta)$ es decreciente en θ, el test definido en (14) también se puede utilizar como test de nivel α para decidir entre las hipótesis

$$H_0 : \theta \leq \theta_0 \qquad \text{contra} \qquad H_1 : \theta > \theta_0.$$

En efecto, si $\theta \leq \theta_0$, entonces $Q(\hat{\theta}(\mathbf{X}), \theta) \geq Q(\hat{\theta}(\mathbf{X}), \theta_0)$ y en consecuencia

$$\beta(\theta) = \mathbb{P}\left(\text{Rechazar } H_0|\theta\right) = \mathbb{P}_\theta\left(Q(\hat{\theta}(\mathbf{X}), \theta_0) > q_{1-\alpha}\right) \leq \mathbb{P}_\theta\left(Q(\hat{\theta}(\mathbf{X}), \theta) > q_{1-\alpha}\right) = \alpha.$$

Por lo tanto,

$$\max_{\theta \leq \theta_0} \beta(\theta) \leq \alpha.$$

Pero como $\beta(\theta_0) = \mathbb{P}_{\theta_0}\left(Q(\hat{\theta}(\mathbf{X}), \theta_0) > q_{1-\alpha}\right) = \alpha$, resulta que

$$\max_{\theta \leq \theta_0} \beta(\theta) = \alpha.$$

$\square$

2.- Si se desea un test de nivel α para decidir entre las hipótesis

$$H_0 : \theta \geq \theta_0 \qquad \text{contra} \qquad H_1 : \theta < \theta_0$$

basta considerar

$$\delta(\mathbf{X}) = \mathbf{1}\left\{Q(\hat{\theta}(\mathbf{X}), \theta_0) < q_\alpha\right\}. \tag{15}$$

En efecto, si $\theta \geq \theta_0$, entonces $Q(\hat{\theta}(\mathbf{X}), \theta) \leq Q(\hat{\theta}(\mathbf{X}), \theta_0)$ y en consecuencia

$$\beta(\theta) = \mathbb{P}\left(\text{Rechazar } H_0|\theta\right) = \mathbb{P}_\theta\left(Q(\hat{\theta}(\mathbf{X}), \theta_0) < q_\alpha\right) \leq \mathbb{P}_\theta\left(Q(\hat{\theta}(\mathbf{X}), \theta) < q_\alpha\right) = \alpha.$$

Por lo tanto,

$$\max_{\theta \geq \theta_0} \beta(\theta) \leq \alpha.$$

Pero como $\beta(\theta_0) = \mathbb{P}_{\theta_0}(Q(\hat{\theta}(\mathbf{X}), \theta_0 < q_\alpha) = \alpha$, resulta que

$$\max_{\theta \geq \theta_0} \beta(\theta) = \alpha.$$

$\square$

3.4. Algunos pivotes

1. **Para media de normales con varianza conocida.** Si $X_1, \ldots, X_n$ es una m.a. de una distribución $\mathcal{N}(\mu, \sigma^2)$, con σ^2 conocida, entonces

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

es un pivote para μ.

2. **Para media de normales con varianza desconocida.** Si $X_1, \ldots, X_n$ es una m.a. de una distribución $\mathcal{N}(\mu, \sigma^2)$, con σ^2 desconocida, entonces

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}$$

es un pivote para μ.

3. **Para varianza de normales con media conocida.** Si $X_1, \ldots, X_n$ es una m.a. de una distribución $\mathcal{N}(\mu, \sigma^2)$, con μ conocida, entonces

$$\frac{n}{\sigma^2}\widehat{\sigma^2}_{mv} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2_n$$

es un pivote para σ^2.

4. **Para varianza de normales con media desconocida.** Si $X_1, \ldots, X_n$ es una m.a. de una distribución $\mathcal{N}(\mu, \sigma^2)$, con μ desconocida, entonces

$$\frac{(n-1)}{\sigma^2} S^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2 \sim \chi^2_{n-1}$$

es un pivote para σ^2.

5. **Para probabilidad de éxito de distribuciones Bernoulli.** Si $X_1, \ldots, X_n$ es una m.a. de una distribución Bernoulli(p) y $n >> 1$, entonces

$$\frac{\sqrt{n}(\bar{X} - p)}{\sqrt{p(1-p)}} \sim \mathcal{N}(0, 1)$$

es un pivote aproximado para p.

6. **Para intensidad de exponenciales.** Si $X_1, \ldots, X_n$ es una m.a. de una distribución Exponencial(λ), entonces

$$2\lambda n \bar{X} = \lambda \sum_{i=1}^{n} X_i \sim \chi^2_{2n}$$

es un pivote para λ.

7. **Para extremo derecho de uniformes.** Si $X_1, \ldots, X_n$ es una m.a. de una distribución $\mathcal{U}(0, \theta)$, entonces

$$\frac{X_{(n)}}{\theta} = \frac{\max(X_1, \ldots, X_n)}{\theta}$$

es un pivote para θ cuya densidad es $f(x) = nx^{n-1}\mathbf{1}\{0 \leq x \leq 1\}$.

8. **Para diferencia de medias de normales con varianzas conocidas.** Si $X_1, \ldots, X_m$ e $Y_1, \ldots, Y_n$ son dos m.a. independientes de distribuciones $\mathcal{N}(\mu_X, \sigma_X^2)$ y $\mathcal{N}(\mu_Y, \sigma_Y^2)$, con σ_X^2 y σ_Y^2 conocidas, entonces

$$\frac{\bar{X} - \bar{Y} - \Delta}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \sim \mathcal{N}(0, 1)$$

es un pivote para la diferencia de medias $\Delta = \mu_X - \mu_Y$.

9. **Para diferencia de medias de normales con varianzas desconocidas pero iguales.** Si $X_1, \ldots, X_m$ e $Y_1, \ldots, Y_n$ son dos m.a. independientes de distribuciones $\mathcal{N}(\mu_X, \sigma^2)$ y $\mathcal{N}(\mu_Y, \sigma^2)$, con varianza común σ^2 desconocida, entonces[3]

$$\frac{\bar{X} - \bar{Y} - \Delta}{\sqrt{S_P^2}\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t_{m+n-2}$$

[3]

$$S_P^2 := \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$$

es un pivote para la diferencia de medias $\Delta = \mu_X - \mu_Y$.

10. **Para cociente de varianzas de normales con medias desconocidas.** Si $X_1, \dots, X_m$ e $Y_1, \dots, Y_n$ son dos m.a. independientes de distribuciones $\mathcal{N}(\mu_X, \sigma_X^2)$ y $\mathcal{N}(\mu_Y, \sigma_Y^2)$, con μ_X y μ_Y desconocidas, entonces

$$\frac{1}{R}\left(\frac{S_X^2}{S_Y^2}\right) \sim F_{m-1,\, n-1}$$

es un pivote para el cociente de las varianzas $R = \sigma_X^2/\sigma_Y^2$.

11. **Para diferencia de probabilidades de éxito de Bernoulli.** Si $X_1, \dots, X_m$ e $Y_1, \dots, Y_n$ son dos m.a. independientes de distribuciones Bernoulli(p_X) y Bernoulli(p_Y). Entonces,

$$\frac{\bar{X} - \bar{Y} - \Delta}{\sqrt{\frac{1}{m}\bar{X}(1-\bar{X}) + \frac{1}{n}\bar{Y}(1-\bar{Y})}} \sim \mathcal{N}(0,1)$$

es un pivote aproximado para la diferencia $\Delta = p_X - p_Y$.

4. Test para media de normales

En esta sección usaremos el método del pivote para construir test de hipótesis sobre la media de distribuciones normales.

4.1. Hipótesis sobre media con varianza conocida

Basados en una muestra aleatoria $\mathbf{X} = (X_1, \dots, X_n)$ de una distribución normal $\mathcal{N}(\mu, \sigma^2)$ con varianza σ^2 conocida queremos construir un test de nivel de significación α para decidir entre las hipótesis

$$H_0 : \mu = \mu_0 \qquad \text{contra} \qquad H_1 : \mu \neq \mu_0,$$

donde μ_0 es un algún valor determinado.

Test de hipótesis

Para distribuciones normales con varianza conocida sabemos que

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

es un pivote para μ basado en $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$.

Es fácil ver que el pivote satisface las dos condiciones enunciadas al principio de la Sección 3. De acuerdo con los resultados expuestos en la sección 3.1

$$\delta(\mathbf{X}) = \mathbf{1}\left\{\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} < z_{\alpha/2}\right\} + \mathbf{1}\left\{\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} > z_{1-\alpha/2}\right\}, \tag{16}$$

es un test de nivel α para decidir entre las hipótesis $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$. Dicho en palabras, el test consiste en rechazar H_0 si $\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} < z_{\alpha/2}$ o $\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} > z_{1-\alpha/2}$ y aceptarla en otro caso.

Nota Bene. Construir un test es la primera fase para decidir entre dos hipótesis. Construido el test es "obligatorio" analizar los riesgos de tomar decisiones erróneas. En otras palabras, el test debe acompañarse con su correspondiente función de potencia.

Función de potencia

Los riesgos de tomar decisiones erróneas utilizando el test de hipótesis definido en (16) pueden evaluarse caracterizando su correspondiente función de potencia: $\beta(\mu) := \mathbb{P}(\text{Rechazar } H_0|\mu)$. Se trata de obtener una expresión "analítica"que nos permita caracterizar cuantitativa y cualitativamente las propiedades de dicha función.

Vale que

$$\beta(\mu) = \Phi\left(z_{\alpha/2} + \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma}\right) + \Phi\left(z_{\alpha/2} + \frac{\sqrt{n}(\mu - \mu_0)}{\sigma}\right). \tag{17}$$

En efecto,

$$\begin{aligned}
\beta(\mu) &= \mathbb{P}(\text{Rechazar } H_0|\mu) \\
&= \mathbb{P}_\mu\left(\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} < z_{\alpha/2}\right) + \mathbb{P}_\mu\left(\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} > z_{1-\alpha/2}\right) \\
&= \mathbb{P}_\mu\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} + \frac{\sqrt{n}(\mu-\mu_0)}{\sigma} < z_{\alpha/2}\right) \\
&\quad + \mathbb{P}_\mu\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} + \frac{\sqrt{n}(\mu-\mu_0)}{\sigma} > z_{1-\alpha/2}\right) \\
&= \mathbb{P}_\mu\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} < z_{\alpha/2} + \frac{\sqrt{n}(\mu_0-\mu)}{\sigma}\right) \\
&\quad + \mathbb{P}_\mu\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} > -z_{\alpha/2} - \frac{\sqrt{n}(\mu-\mu_0)}{\sigma}\right) \\
&= \Phi\left(z_{\alpha/2} + \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma}\right) + \Phi\left(z_{\alpha/2} + \frac{\sqrt{n}(\mu - \mu_0)}{\sigma}\right).
\end{aligned}$$

Notar que la función de potencia dada en (17) satisface las siguientes propiedades

(a) $\beta(\mu)$ es simétrica con respecto a μ_0: $\beta(\mu_0 + m) = \beta(\mu_0 - m)$ para todo $m > 0$.

(b) $\beta(\mu)$ es creciente[4] sobre la semi-recta (μ_0, ∞).

(c) $\beta(\mu_0) = \alpha$.

[4] Derivar con respecto de μ la expresión (17) y hacer cuentas.

(d) $\lim_{\mu\uparrow+\infty}\beta(\mu)=1$

Esto significa que a medida que nos alejamos de la hipótesis $\mu=\mu_0$ disminuye el riesgo de aceptar dicha hipótesis cuando es falsa. La forma típica del gráfico de la función de potencia correspondiente al test de la forma (16) para decidir entre las hipótesis $H_0:\mu=\mu_0$ contra $H_1:\mu\neq\mu_1$ puede observarse en las Figuras 2 y 3.

Nota Bene. La función de potencia es útil para determinar cuan grande debe ser la muestra aleatoria para conseguir ciertas especificaciones relativas a los errores de tipo II. Por ejemplo, supongamos que queremos determinar el volumen de la muestra n necesario para asegurar que la probabilidad de rechazar $H_0:\mu=\mu_0$ cuando el verdadero valor de la media es μ_1 sea aproximadamente β. Esto es, queremos determinar n tal que

$$\beta(\mu_1)\approx\beta.$$

De la expresión (17), esto es equivalente a

$$\Phi\left(z_{\alpha/2}+\frac{\sqrt{n}(\mu_0-\mu)}{\sigma}\right)+\Phi\left(z_{\alpha/2}+\frac{\sqrt{n}(\mu-\mu_0)}{\sigma}\right)\approx\beta. \tag{18}$$

Aunque la ecuación (18) no se pueda resolver analíticamente, se puede conseguir una solución aproximada mediante la siguiente observación.

1. Supongamos que $\mu_1>\mu_0$. En tal caso, el primer término del lado izquierdo de (18) es despreciable, (es fácil ver que está acotado por $\alpha/2\approx 0$) y por lo tanto, el problema se reduce a resolver la ecuación aproximada

$$\Phi\left(z_{\alpha/2}+\frac{\sqrt{n}(\mu_1-\mu_0)}{\sigma}\right)\approx\beta.$$

En consecuencia, basta tomar n tal que $z_{\alpha/2}+\frac{\sqrt{n}(\mu_1-\mu_0)}{\sigma}\approx z_\beta$ ó lo que es equivalente

$$n\approx\left(\frac{\sigma(z_\beta-z_{\alpha/2})}{\mu_1-\mu_0}\right)^2. \tag{19}$$

2. Supongamos que $\mu_1<\mu_0$. En tal caso, el segundo término del lado izquierdo de (18) es despreciable, y por lo tanto, el problema se reduce a resolver la ecuación aproximada

$$\Phi\left(z_{\alpha/2}+\frac{\sqrt{n}(\mu_0-\mu_1)}{\sigma}\right)\approx\beta.$$

En consecuencia, basta tomar n tal que

$$n\approx\left(\frac{\sigma(z_\beta-z_{\alpha/2})}{\mu_0-\mu_1}\right)^2. \tag{20}$$

El resultado obtenido en (19) coincide con el resultado obtenido en (20) y es una aproximación razonable para el volumen de muestra necesario para asegurar que el error de tipo II en el valor $\mu = \mu_1$ es aproximadamente igual a $1 - \beta$.

Ejemplo 4.1. Si se envía una señal de valor μ desde un sitio A, el valor recibido en el sitio B se distribuye como una normal de media μ y desvío estándar 2. Esto es, el ruido que perturba la señal es una variable aleatoria $\mathcal{N}(0, 4)$. El receptor de la señal en el sitio B tiene suficientes motivos para sospechar que recibirá una señal de valor $\mu = 8$. Analizar la consistencia de dicha hipótesis suponiendo que la misma señal fue enviada en forma independientemente 5 veces desde el sitio A y el promedio del valor recibido en el sitio B es $\bar{X} = 9.5$.

Solución. Se trata de construir un test de hipótesis para decidir entre las hipótesis

$$H_0 : \mu = 8 \qquad \text{contra} \qquad H_1 : \mu \neq 8,$$

usando una muestra $\mathbf{X} = (X_1, \dots, X_5)$ de una distribución $\mathcal{N}(\mu, 4)$.

Test de hipótesis. Para un nivel de significación del 5 % el test es de la forma

$$\delta(\mathbf{X}) = \mathbf{1}\left\{\left|\frac{\sqrt{5}(\bar{X} - 8)}{2}\right| > 1.96\right\} \tag{21}$$

Decisión basada en la muestra observada. Calculamos el valor

$$\left|\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}\right| = \left|\frac{\sqrt{5}(9.5 - 8)}{2}\right| = 1.68$$

Como este valor es menor que $z_{1-\alpha/2} = z_{0.975} = 1.96$, se acepta la hipótesis $\mu = 8$. En otras palabras, los datos no son inconsistentes con la hipótesis $\mu = 8$.

Nota Bene. Notar que, si se relaja el nivel de significación al 10 %, entonces la hipótesis $\mu = 8$ debe rechazarse debido a que el valor $z_{0.95} = 1.645$ es menor que 1.68.

Función de potencia. La función de potencia es

$$\beta(\mu) = \Phi\left(-1.96 + \frac{\sqrt{5}(8 - \mu)}{2}\right) + \Phi\left(-1.96 + \frac{\sqrt{5}(\mu - 8)}{2}\right). \tag{22}$$

Si se quiere determinar la probabilidad de cometer un error de tipo II cuando el valor real enviado es 10 basta poner $\mu = 10$ en la expresión (22) y calcular $1 - \beta(10)$:

$$1 - \Phi\left(-1.96 - \sqrt{5}\right) - \Phi\left(-1.96 + \sqrt{5}\right) = \Phi\left(-0.276\right) - \Phi(-4.196) = 0.392.$$

$\square$

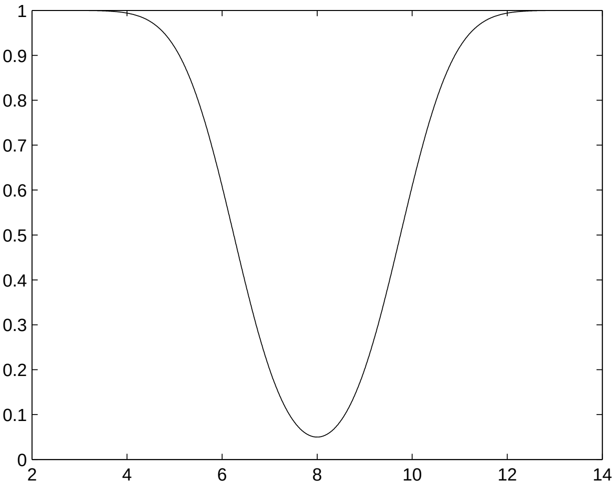

Figura 2: Gráfico de la función de potencia (22) correspondiente al test de hipótesis definido en (21) para decidir entre $H_0 : \mu = 8$ contra $H_1 : \mu \neq 8$ con un nivel de significación del 5 % y basado en una muestra de volumen 5.

Ejemplo 4.2. Volvamos al problema del **Ejemplo 4.1**. Cuántas señales deberían enviarse para que el test de nivel de significación $\alpha = 0.05$ para $H_0 : \mu = 8$ contra $H_1 : \mu \neq 8$ tenga al menos una probabilidad igual a 0.75 de rechazar esa hipótesis cuando $\mu = 9.2$?

Solución. Como $z_{0.025} = -1.96$ y $z_{0.75} = 0.67$, de (19) resulta $n \approx \left(\frac{2(0.67+1.96)}{9.2-8}\right)^2 = 19.21$.

Para una muestra de volumen 20 el test adopta la forma

$$\delta(\mathbf{X}) = \mathbf{1}\left\{\left|\frac{\sqrt{20}(\bar{X}-8)}{2}\right| > 1.96\right\} = \mathbf{1}\left\{\left|\sqrt{5}(\bar{X}-8)\right| > 1.96\right\} \tag{23}$$

y su función de potencia adopta la expresión

$$\beta(\mu) = \Phi\left(-1.96 + \sqrt{5}(8-\mu)\right) + \Phi\left(-1.96 + \sqrt{5}(\mu-8)\right). \tag{24}$$

En consecuencia,

$$\beta(9.2) = \Phi(-4.6433) + \Phi(0.72328) = 0.76525.$$

Dicho en palabras, si el mensaje se envía 20 veces, entonces hay un 76.52 % de posibilidades de que la hipótesis nula $\mu = 8$ sea rechazada cuando la media verdadera es 9.2.

$\square$

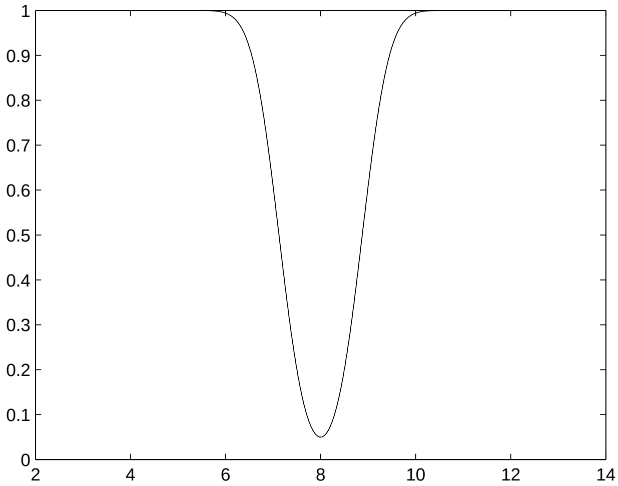

Figura 3: Gráfico de la función de potencia (24) correspondiente al test definido en (23) para decidir entre las hipótesis $H_0 : \mu = 8$ contra $H_1 : \mu \neq 8$ con un nivel de significación del 5 % y basado en una muestra de volumen 20.

Nota Bene. Comparando las Figuras 2 y 3 se puede ver que, fijado el nivel de significación del test, cuando se aumenta el volumen de la muestra disminuyen los errores de tipo II.

4.2. Variaciones sobre el mismo tema

Basados en una muestra $\mathbf{X} = (X_1, \dots, X_n)$ de una distribución normal $\mathcal{N}(\mu, \sigma^2)$ con varianza σ^2 conocida se quiere construir un test de nivel de significación α para decidir entre las hipótesis

$$H_0 : \mu = \mu_0 \qquad \text{contra} \qquad H_1 : \mu > \mu_0,$$

donde μ_0 es un algún valor determinado.

Usando los resultados expuestos en la sección 3.2 tenemos que

$$\delta(\mathbf{X}) = \mathbf{1}\left\{\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} > z_{1-\alpha}\right\}. \tag{25}$$

es un test de nivel α para decidir entre $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$. Dicho en palabras, el test de hipótesis consiste en rechazar H_0 si $\bar{X} > \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$ y aceptarla en otro caso.

Función de potencia. La función de potencia correspondiente al test (25) es

$$\begin{aligned}\beta(\mu) &= \mathbb{P}\left(\text{Rechazar}H_0|\mu\right) = \mathbb{P}_\mu\left(\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} > z_{1-\alpha}\right) \\ &= \mathbb{P}_\mu\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} + \frac{\sqrt{n}(\mu-\mu_0)}{\sigma} > z_{1-\alpha}\right) \\ &= \mathbb{P}_\mu\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} > -z_\alpha - \frac{\sqrt{n}(\mu-\mu_0)}{\sigma}\right) \\ &= \Phi\left(z_\alpha + \frac{\sqrt{n}(\mu-\mu_0)}{\sigma}\right). \end{aligned} \tag{26}$$

De las propiedades de la función $\Phi(\cdot)$ y de la expresión (26) para la función de potencia se deduce que

(a) $\beta(\mu)$ creciente.

(b) $\beta(\mu_0) = \alpha$

(c) $\lim_{\mu\uparrow+\infty}\beta(\mu) = 1$ y $\lim_{\mu\downarrow-\infty}\beta(\mu) = 0$.

Debido a que la función de potencia (26) es creciente, el test definido en (25) también se puede usar para decidir, con un nivel de significación α, entre la hipótesis

$$H_0 : \mu \leq \mu_0 \qquad \text{contra} \qquad H_1 : \mu > \mu_0.$$

Ejemplo 4.3. Volvamos al problema presentado en el **Ejemplo 4.1** pero supongamos que esta vez estamos interesados en testear con nivel de significación, $\alpha = 0.05$, la hipótesis $H_0 : \mu \leq 8$ contra la hipótesis alternativa $H_1 : \mu > 8$. (Recordar que disponemos de muestra aleatoria de volumen 5 de una población normal $\mathcal{N}(\mu, 4)$ cuyo promedio resultó ser $\bar{X} = 9.5$)

En este caso, el test de hipótesis definido en (25) puede enunciarse de la siguiente manera:

$$\text{Rechazar } H_0 \text{ cuando } \bar{X} > 8 + \frac{2}{\sqrt{5}}z_{0.95} = 9.4712 \text{ y aceptarla en otro caso.} \tag{27}$$

Si se observó que $\bar{X} = 9.5$, entonces debe rechazarse la hipótesis $\mu \leq 8$ a favor de la alternativa $\mu > 9$. La función de potencia correspondiente al test de hipótesis (27) es

$$\beta(\mu) = \Phi\left(-1.64 + \frac{\sqrt{5}(\mu-8)}{2}\right) \tag{28}$$

Si se quiere determinar la probabilidad de aceptar la hipótesis $\mu \leq 8$ cuando el valor real enviado es $\mu = 10$ basta poner $\mu = 10$ en la expresión (28) y calculamos:

$$1 - \beta(10) = 1 - \Phi\left(-1.64 + \sqrt{5}\right) = 0.27... \tag{29}$$

□

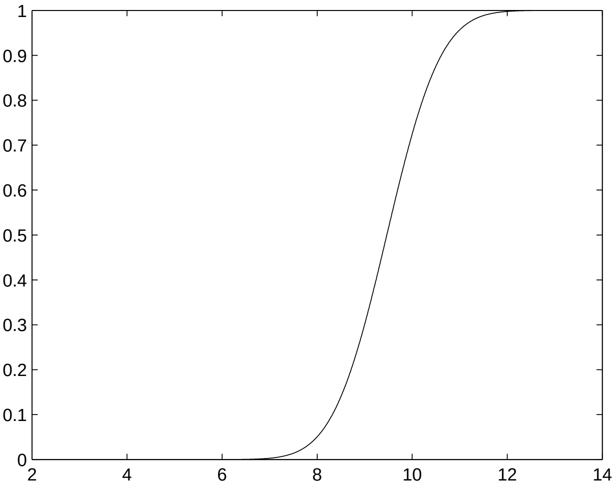

Figura 4: Gráfico de la función de potencia (28) correspondiente al test definido en (27) para decidir entre las hipótesis $H_0 : \mu \leq 8$ contra $H_1 : \mu > 8$ con un nivel de significación del 5 % y basado en una muestra de volumen 5.

4.3. Hipótesis sobre media con varianza desconocida

Basados en una muestra aleatoria $\mathbf{X} = (X_1, \dots, X_n)$ de una distribución normal $\mathcal{N}(\mu, \sigma^2)$ queremos construir un test de nivel de significación α para decidir entre las hipótesis

$$H_0 : \mu = \mu_0 \qquad \text{contra} \qquad H_1 : \mu \neq \mu_0,$$

donde μ_0 es un algún valor determinado.

Test de hipótesis

Para distribuciones normales sabemos que

$$Q(\bar{X}, \mu) = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}$$

es un pivote para μ basado en $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ y $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Es fácil ver que el pivote satisface las dos condiciones enunciadas al principio de la Sección 3. De acuerdo con los resultados expuestos en la sección 3.1

$$\delta(\mathbf{X}) = \mathbf{1}\left\{\frac{\sqrt{n}(\bar{X} - \mu_0)}{S} < t_{n-1,\, \alpha/2}\right\} + \mathbf{1}\left\{\frac{\sqrt{n}(\bar{X} - \mu_0)}{S} > t_{n-1,\, 1-\alpha/2}\right\}, \tag{30}$$

es un test de nivel α para decidir entre las hipótesis $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$. Dicho en palabras, el test en rechazar H_0 si $\frac{\sqrt{n}(\bar{X}-\mu_0)}{S} < t_{n-1,\, \alpha/2}$ o $\frac{\sqrt{n}(\bar{X}-\mu_0)}{S} > t_{n-1,\, 1-\alpha/2}$ y aceptarla en otro caso.

Ejemplo

Ejemplo 4.4. En la siguiente tabla se muestran las mediciones, en segundos de grado, obtenidas por James Short (1761), de la paralaje solar (ángulo bajo el que se ve el radio ecuatorial de la tierra desde el centro del sol) .

8.50	8.50	7.33	8.64	9.27	9.06	9.25	9.09	8.50	8.06
8.43	8.44	8.14	7.68	10.34	8.07	8.36	9.71	8.65	8.35
8.71	8.31	8.36	8.58	7.80	7.71	8.30	9.71	8.50	8.28
9.87	8.86	5.76	8.44	8.23	8.50	8.80	8.40	8.82	9.02
10.57	9.11	8.66	8.34	8.60	7.99	8.58	8.34	9.64	8.34
8.55	9.54	9.07							

Con esos datos tenemos que $\bar{X} = 8.6162$ y $S = 0.749$. En la Figura 5 se muestra un histograma de los datos.

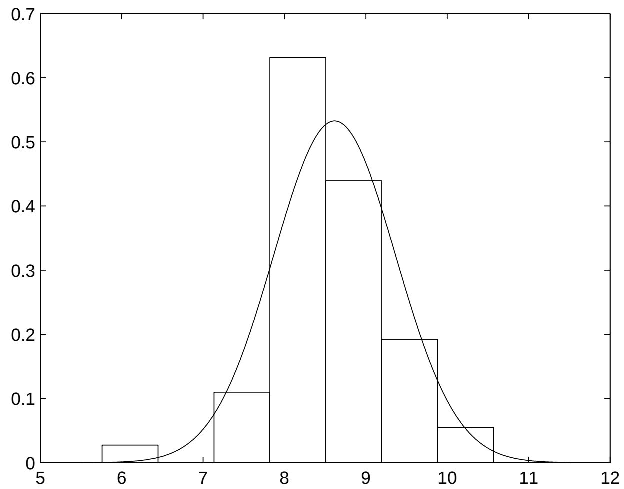

Figura 5: Histograma de las mediciones obtenidas por James Short. Parece razonable asumir que las mediciones de la paralaje solar tienen distribución normal.

Asumiendo que las mediciones tienen distribución $\mathcal{N}(\mu, \sigma^2)$ queremos decidir, con un nivel de significación $\alpha = 0.05$, entre las hipótesis

$$H_0 : \mu = 8.798 \qquad \text{contra} \qquad H_1 : \mu \neq 8.798$$

Como $n = 53$ y $t_{52,\,0.025} = -t_{52,\,0.975} = -2.0066$, el test de hipótesis (30) adopta la forma

$$\delta(\mathbf{X}) = \mathbf{1}\left\{\frac{\sqrt{53}(\bar{X} - 8.798)}{S} < -2.0066\right\} + \mathbf{1}\left\{\frac{\sqrt{53}(\bar{X} - 8.798)}{S} > 2.0066\right\}.$$

Usando los datos de las mediciones tenemos que

$$\frac{\sqrt{53}(\bar{X} - 8.798)}{S} = \frac{\sqrt{53}(8.6162 - 8.798)}{0.749} = -1.7667.$$

Por lo tanto, no hay evidencia suficiente para rechazar que la paralaje solar es $\mu = 8.798$.

Usando como paralaje solar el valor $\mu = 8.798''$ y como radio ecuatorial de la tierra el valor $R = 6378$ km., trigonometría mediante, se puede determinar la distancia D entre la tierra y el sol:

$$\tan\left(\frac{8.798}{3600} \times \frac{\pi}{180}\right) = \frac{6378}{D} \iff D = 1.4953 \times 10^8.$$

Lo que significa que la distancia entre la tierra y el sol es 149.53 millones de km. $\square$

5. Test para probabilidad de éxito de distribuciones Bernoulli

Sea $\mathbf{X} = (X_1, \dots, X_n)$ una muestra aleatoria de una variable aleatoria con distribución Bernoulli(p), $p \in (0,1)$. Basados en la muestra aleatoria, $\mathbf{X}$, queremos construir test para decidir entre dos hipótesis sobre la probabilidad de éxito p.

La cantidad de éxitos en la muestra

$$N = \sum_{i=1}^{n} X_i$$

tiene distribución Binomial(n, p) y resume toda la información relevante sobre el parámetro p contenida en la muestra aleatoria $\mathbf{X}$. La media y la varianza de N son, respectivamente, $\mathbb{E}_p[N] = np$ y $\mathbb{V}_p(N) = np(1-p)$.

Lema 5.1 (Dominación estocástica)**.** Sean $0 < p_1 < p_2 < 1$ arbitrarios pero fijos. Si $N_1 \sim$ Binomial(n, p_1) y $N_2 \sim$ Binomial(n, p_2), entonces para cada $x \in \mathbb{R}$ vale que

$$\mathbb{P}(N_2 \leq x) \leq \mathbb{P}(N_1 \leq x).$$

Demostración Sean $U_1, \dots, U_n$ variables aleatorias independientes cada una con distribución $\mathcal{U}(0,1)$. Para cada $i = 1, \dots, n$ construya las siguientes variables

$$X_{1,i} := \mathbf{1}\{U_i \leq p_1\}, \qquad X_{2,i} := \mathbf{1}\{U_i \leq p_2\}.$$

Por construcción valen las siguientes propiedades:

(a) las variables $X_{1,1}, \dots, X_{1,n}$ son iid Bernoulli(p_1);

(b) las variables $X_{2,1}, \dots, X_{2,n}$ son iid Bernoulli(p_2);

(c) para cada i vale que $X_{2,i} \geq X_{1,i}$.

En consecuencia, las variables

$$\hat{N}_1 := \sum_{i=1}^{n} X_{1,i} \sim \text{Binomial}(n, p_1), \qquad \hat{N}_2 := \sum_{i=1}^{n} X_{2,i} \sim \text{Binomial}(n, p_2)$$

verifican que $\hat{N}_1 \leq \hat{N}_2$. Se deduce entonces que que $\{\hat{N}_2 \leq x\} \subseteq \{\hat{N}_1 \leq x\}$, para cualquier $x \in \mathbb{R}$. Por lo tanto,

$$\mathbb{P}(N_2 \leq x) = \mathbb{P}\left(\hat{N}_2 \leq x\right) \leq \mathbb{P}\left(\hat{N}_1 \leq x\right) = \mathbb{P}\left(N_1 \leq x\right).$$

□

Corolario 5.2. Sea N una variable aleatoria con distribución Binomial(n, p), $p \in (0, 1)$. Fijado un valor $x \in \mathbb{R}^+$, la función polinómica de grado n, $h : (0, 1) \to [0, 1]$, definida por

$$h(p) = \mathbb{P}_p(N \leq x) = \sum_{k=0}^{[x]} \binom{n}{k} p^k (1-p)^{n-k}$$

es decreciente. □

5.1. Test para moneda honesta (de lo simple a lo complejo)

Se quiere decidir si una moneda es honesta o no lo es. Formalmente, se trata de construir un test para decidir entre las hipótesis

$$H_0 : p = \frac{1}{2} \quad \text{contra} \quad H_1 : p \neq \frac{1}{2}.$$

1.- Se quiere decidir tirando la moneda 6 veces. ¿Qué hacer? Observamos la cantidad N de caras obtenidas en los 6 tiros. Para cada p tenemos que $N \sim$ Binomial$(6, p)$. Cuando la moneda es honesta, $\mathbb{E}_{1/2}[N] = 3$. Teniendo en cuenta la existencia de fluctuaciones parece razonable aceptar que la moneda es honesta cuando observamos que $2 \leq N \leq 4$. Proponemos entonces el siguiente test

$$\delta(\mathbf{X}) = 1 - \mathbf{1}\{2 \leq N \leq 4\} = \mathbf{1}\{N < 2\} + \mathbf{1}\{N > 4\},$$

cuya función de potencia des

$$\beta(p) = \mathbb{P}_p(N \leq 1) + \mathbb{P}_p(N \geq 5) = (1-p)^6 + 6p(1-p)^5 + 6p^5(1-p) + p^6.$$

Dada una moneda honesta, ¿qué riesgo se corre de rechazarla como falsa? Esta pregunta se contesta calculando el nivel de significación del test $\alpha = \beta(1/2) = \frac{14}{64} = 0.21875$. □

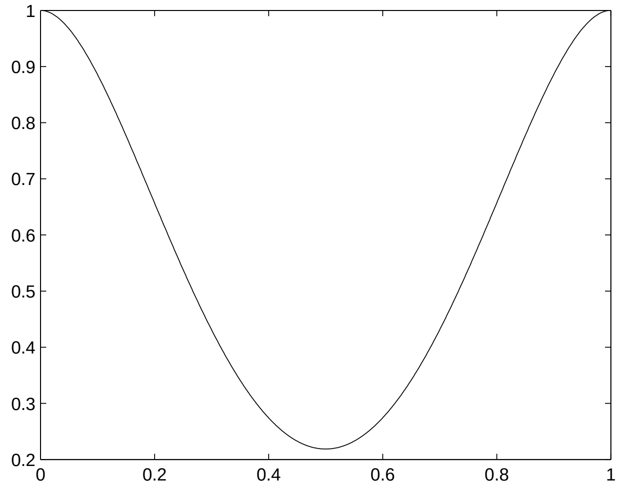

Figura 6: Gráfico de la función de potencia del test $\delta(\mathbf{X}) = \mathbf{1}\{N < 2\} + \mathbf{1}\{N > 4\}$.

2.- Se propone el siguiente *test*: lanzar la moneda 100 veces y contar la cantidad de caras observadas N. Si $40 \leq N \leq 60$ se decide que la moneda es honesta. En caso contrario, se decide que no lo es.

Definido el test lo único que queda por hacer es evaluar los riesgos de decisiones erróneas. Para ello calculamos la función de potencia

$$\beta(p) = \mathbb{P}(\text{Rechazar} H_0 |\, p) = \mathbb{P}_p(N < 40) + \mathbb{P}_p(N > 60).$$

Para cada p la cantidad de caras observadas en 100 lanzamientos se distribuye como una Binomial: $N \sim \text{Binomial}(100, p)$. En consecuencia,

$$\beta(p) = \sum_{k=0}^{39} \binom{100}{k} p^k (1-p)^{100-k} + \sum_{k=61}^{100} \binom{100}{k} p^k (1-p)^{100-k}. \tag{31}$$

Sin una herramienta computacional a la mano es insensato calcular riesgos utilizando la expresión obtenida en (31). Como el volumen de la muestra es 100 usando el teorema central del límite, $N \sim \mathcal{N}(100p, 100p(1-p)$, podemos obtener una buena aproximación de la función de potencia, (al menos para valores de p contenidos en el intervalo abierto $(0.12, 0.88)$)

$$\begin{aligned}
\beta(p) &\approx \Phi\left(\frac{40 - 100p}{\sqrt{100p(1-p)}}\right) + 1 - \Phi\left(\frac{60-100p}{\sqrt{100p(1-p)}}\right) \\
&= \Phi\left(\frac{4-10p}{\sqrt{p(1-p)}}\right) + \Phi\left(\frac{10p-6}{\sqrt{p(1-p)}}\right) \qquad (32)
\end{aligned}$$

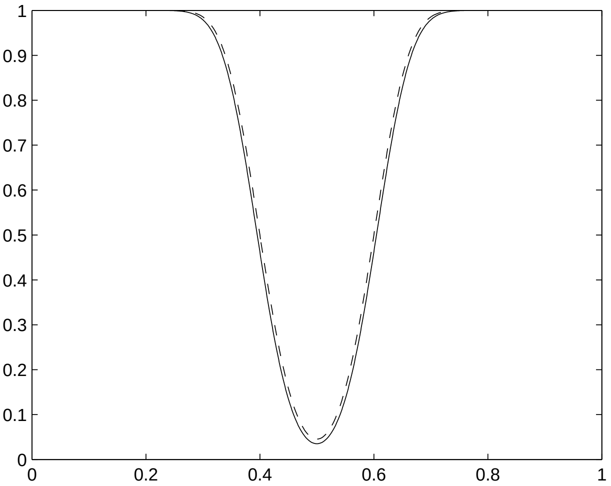

Figura 7: Gráfico de la función de potencia del test $\delta(\mathbf{X}) = \mathbf{1}\{N < 40\} + \mathbf{1}\{N > 60\}$. En línea quebrada aproximación usando el TCL.

Es más o menos claro que la función de potencia es simétrica respecto de $p = 1/2$. Esto es, para cada $q \in (0, 1/2))$, vale que $\beta(1/2 - q) = \beta(1/2 + q)$.

Riesgos:

1. El *nivel de significación del test* es $\alpha = \beta(1/2)$. Calculamos $\beta(1/2)$ utilizando la aproximación obtenida en (32)

$$\beta(1/2) \approx \Phi\left(\frac{4-5}{\sqrt{1/4}}\right) + \Phi\left(\frac{5-6}{\sqrt{1/4}}\right) = \Phi(-2) + \Phi(-2) \approx 0.0455$$

   Esto significa que la probabilidad de rechazar que la moneda es honesta, cuando en verdad lo es, será 0.0455. En palabras: de cada 100 monedas honestas sometidas a verificación (en promedio) serán rechazadas como falsas 4 o 5 de ellas.

2. ¿Qué riesgo se corre de aceptar como honesta una moneda falsa, con carga 0.7 hacia el lado de la cara? Para contestar esta pregunta tenemos que calcular el valor de $1 - \beta(0.7)$. Usando (32) obtenemos

$$1 - \beta(0.7) \approx 1 - \Phi\left(\frac{4-7}{\sqrt{0.21}}\right) - \Phi\left(\frac{7-6}{\sqrt{0.21}}\right) \approx 0.0146.$$

   Grosso modo el resultado se interpreta de la siguiente manera: de cada 100 monedas cargadas con 0.7 para el lado de cara sometidas a verificación (en promedio) serán aceptadas como honestas 1 o 2 de ellas. □

3.- Queremos un test de nivel de significación $\alpha = 0.05$, basado en 64 lanzamientos de la moneda. Parece razonable proponer un test de la forma

$$\delta(\mathbf{X}) = \mathbf{1}\{N < 32 - k\} + \mathbf{1}\{N > 32 + k\}.$$

El problema consiste en determinar el valor de k. El nivel de significación del test es

$$\beta(1/2) = \mathbb{P}_{1/2}(N < 32 - k) + \mathbb{P}_{1/2}(N > 32 + k)$$

Para $p = 1/2$, $N \sim \text{Binomial}(64, 1/2)$ y usando el teorema central de límite obtenemos que la distribución de N es aproximadamente normal de media $\mathbb{E}_{1/2}[N] = (1/2)64 = 32$ y varianza $V_{1/2}(N) = (1/2)(1/2)64 = 16$.

$$\begin{aligned}
\beta(1/2) &= \mathbb{P}_{1/2}(N < 32 - k) + \mathbb{P}_{1/2}(N > 32 + k) \\
&\approx \mathbb{P}_{1/2}\left(\frac{N-32}{4} < -\frac{k}{4}\right) + \mathbb{P}_{1/2}\left(\frac{N-32}{4} > \frac{k}{4}\right) \\
&= \Phi\left(-\frac{k}{4}\right) + \Phi\left(-\frac{k}{4}\right) = 2\Phi\left(-\frac{k}{4}\right)
\end{aligned}$$

En consecuencia,

$$\beta(1/2) = 0.05 \iff \Phi\left(-\frac{k}{4}\right) = 0.025 \iff -\frac{k}{4} = z_{0.025} = -1.96 \iff k = 7.84.$$

Por lo tanto, el test adopta la forma

$$\delta(\mathbf{X}) = \mathbf{1}\{N < 32 - 7.84\} + \mathbf{1}\{N > 32 + 7.84\} = \mathbf{1}\{N < 25\} + \mathbf{1}\{N > 39\}.$$

En palabras, el *test* consiste en lo siguiente: lanzar la moneda 64 veces; si la cantidad de caras observadas es menor que 25 o mayor que 39, se decide que la moneda está cargada; en caso contrario, se decide que la moneda es honesta.

¿Qué riesgo se corre de aceptar como honesta una moneda con carga 0.7 hacia el lado de la cara? La respuesta se obtiene calculando $1 - \beta(0.7)$. Para $p = 0.7$ el TCL establece que $(N - 0.7(64))/\sqrt{(0.7)(0.3)64} \sim \mathcal{N}(0, 1)$, en consecuencia,

$$\beta(0.7) \approx \Phi\left(\frac{25 - 0.7(64)}{\sqrt{(0.21)64}}\right) + \Phi\left(\frac{0.7(64) - 39}{\sqrt{(0.21)64}}\right) \approx \Phi(1.5821) = 0.94318.$$

Por lo tanto, $1 - \beta(0.7) = 0.0568...$ □

4.- Queremos un test de nivel de significación $\alpha = 0.05$, cuya potencia cuando la carga difiere de 0.5 en más de 0.1 sea como mínimo 0.90. Parece razonable proponer una regla de la forma

$$\delta(\mathbf{X}) = \mathbf{1}\{N < n(1/2) - k\} + \mathbf{1}\{N > n(1/2) + k\}.$$

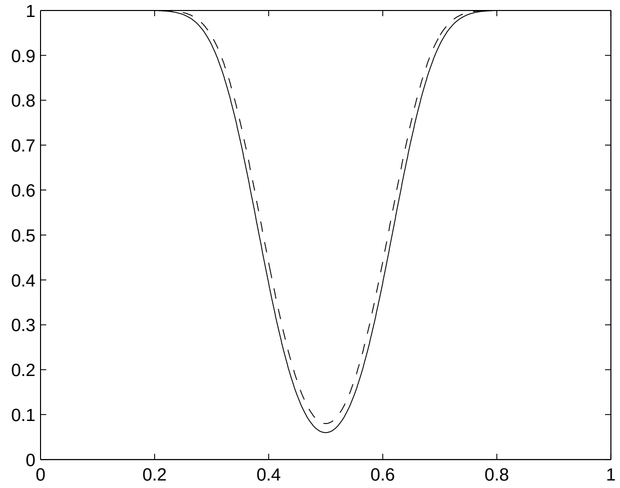

Figura 8: Gráfico de la función de potencia del test $\delta(\mathbf{X}) = \mathbf{1}\{N < 25\} + \mathbf{1}\{N > 39\}$. En línea quebrada aproximación usando el TCL.

El problema consiste en determinar el volumen de la muestra, n, y el valor de k. Las condiciones impuestas al test pueden expresarse de la siguiente manera

$$\alpha(\delta) \leq 0.05 \qquad \text{y} \qquad \beta(0.6) \geq 0.90, \tag{33}$$

donde $\alpha(\delta) = \beta(1/2)$ es en nivel del test y $\beta(0.6)$ es la potencia en $p = 0.6$.

Ambos problemas se resuelven caracterizando la función de potencia del test

$$\beta(p) = \mathbb{P}_p\left(N < n(1/2) - n\epsilon\right) + \mathbb{P}_p\left(N > n(1/2) + n\epsilon\right)$$

De acuerdo con el el TCL tenemos que para cada p

$$Z = \frac{N - np}{\sqrt{np(1-p)}} \sim \mathcal{N}(0,1),$$

en consecuencia,

$$\begin{aligned} \beta(p) &\approx \mathbb{P}_p\left(Z < \frac{n(1/2-p) - n\epsilon}{\sqrt{np(1-p)}}\right) + \mathbb{P}_p\left(Z > \frac{n(1/2-p) + n\epsilon}{\sqrt{np(1-p)}}\right) \\ &= \Phi\left(\frac{\sqrt{n}(1/2 - p - \epsilon)}{\sqrt{p(1-p)}}\right) + \Phi\left(\frac{\sqrt{n}(p - 1/2 - \epsilon)}{\sqrt{p(1-p)}}\right) \end{aligned}$$

Notar que para $p > 1/2$ el primer término del lado derecho de la igualdad es despreciable y entonces

$$\beta(0.6) \approx \Phi\left(\frac{\sqrt{n}(0.1 - \epsilon)}{\sqrt{0.24}}\right)$$

Por otra parte,

$$\beta(1/2) \;\approx\; 2\Phi\left(\frac{-\sqrt{n}\epsilon}{\sqrt{1/4}}\right) = 2\Phi\left(-2\sqrt{n}\epsilon\right)$$

En consecuencia, las desigualdades (33) son equivalentes a las siguientes:

$$2\Phi\left(-2\sqrt{n}\epsilon\right) \leq 0.05 \qquad \text{y} \qquad \Phi\left(\frac{\sqrt{n}(0.1-\epsilon)}{\sqrt{0.24}}\right) \geq 0.90.$$

Por lo tanto, n y ϵ deben ser tales que

$$2\epsilon\sqrt{n} \geq z_{0.975} \qquad \text{y} \qquad \frac{\sqrt{n}(0.1-\epsilon)}{\sqrt{0.24}} \geq z_{0.90} \tag{34}$$

Recurriendo a una tabla de la distribución normal, usando una calculadora de almacenero (que tenga una tecla con el símbolo $\sqrt{\cdot}$), y operando con las desigualdades (34) se pueden obtener soluciones particulares. Por ejemplo, $n = 259$ y $\epsilon = 0.061$.

Tomando $n = 259$ y $\epsilon = 0.061$ obtenemos la siguiente regla de decisión:

$$\delta(\mathbf{X}) = \mathbf{1}\{N < 114\} + \mathbf{1}\{N > 145\}.$$

En palabras, el test establece que hay que lanzar la moneda 259 veces y contar la cantidad de caras observadas. Si la cantidad de caras observadas es menor que 114 o mayor que 145 se decide que la moneda está cargada. En caso contrario, se decide que es honesta.

Una cuenta. Para obtener el resultado particular $n = 259$ y $\epsilon = 0.061$ hay que hacer lo siguiente: En primer lugar, hay que observar que

$$\begin{aligned}\frac{\sqrt{n}(0.1-\epsilon)}{\sqrt{0.24}} \geq z_{0.90} \quad &\iff \quad \sqrt{n}(0.1-\epsilon) \geq z_{0.90}\sqrt{0.24}\\ &\iff \quad 0.1\sqrt{n} - z_{0.90}\sqrt{0.24} \geq \epsilon\sqrt{n}\\ &\iff \quad 2\left(0.1\sqrt{n} - z_{0.90}\sqrt{0.24}\right) \geq 2\epsilon\sqrt{n}\end{aligned} \tag{35}$$

La última desigualdad de (35) combinada con la primera de (34) implican que n debe satisfacer las desigualdades

$$\begin{aligned}0.2\sqrt{n} - 2z_{0.90}\sqrt{0.24} \geq z_{0.975} \quad &\iff \quad \sqrt{n} \geq 5\left(z_{0.975} + 2z_{0.90}\sqrt{0.24}\right)\\ &\iff \quad n \geq 25\left(z_{0.975} + 2z_{0.90}\sqrt{0.24}\right)^2\end{aligned}$$

Tabla de la distribución normal ($z_{0.975} = 1.96$, $z_{0.90} = 1.28$) y calculadora mediante, se obtiene que $n \geq 259$. Poniendo $n = 259$ en la tercera desigualdad de (35) se puede ver que ϵ debe ser tal que

$$\epsilon \leq 0.1 - z_{0.90}\frac{\sqrt{0.24}}{\sqrt{259}} \approx 0.061.$$

Podemos elegir $\epsilon = 0.061$. $\square$

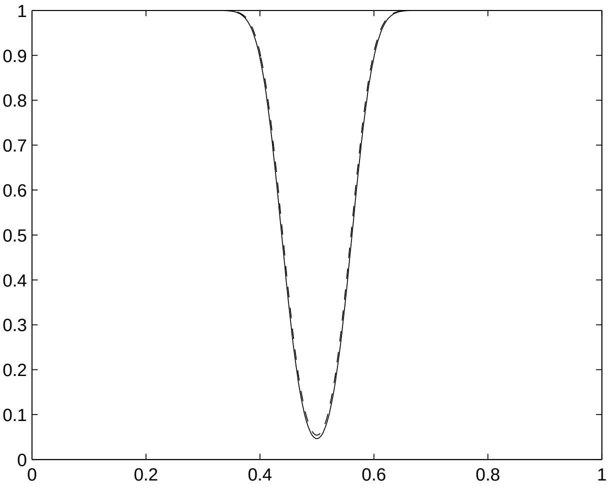

Figura 9: Gráfico de la función de potencia del test $\delta(\mathbf{X}) = \mathbf{1}\{N < 114\} + \mathbf{1}\{N > 145\}$. En línea quebrada aproximación usando el TCL.

5.2. Hipótesis fundamental simple

Sea $\mathbf{X} = (X_1, \dots, X_n)$ una muestra aleatoria de una variable aleatoria con distribución Bernoulli(p), $p \in (0, 1)$. Basados en la muestra aleatoria $\mathbf{X}$ queremos construir test para decidir entre las hipótesis

$$H_0 : p = p_0 \qquad \text{contra} \qquad H_1 : p \neq p_0,$$

donde $p_0 \in (0, 1)$ es un valor arbitrario pero fijo.

Primera fase: diseñar un test de hipótesis

Cuando la hipótesis H_0 es verdadera, la cantidad de éxitos $N = \sum_{i=1}^n X_i$ tiene distribución binomial de media np_0 y desvío $\sqrt{np_0(1 - p_0)}$. Parece razonable construir reglas de decisión de la forma

$$\delta(\mathbf{X}) = \mathbf{1}\left\{N < np_0 - n\epsilon\right\} + \mathbf{1}\left\{N > np_0 + n\epsilon\right\}, \tag{36}$$

donde $n \in \mathbb{N}$ y $\epsilon > 0$ son arbitrarios pero fijos.

En castellano, el test de hipótesis definido en (36) establece el siguiente procedimiento de decisión:

1. Examinar una muestra de tamaño n de la variable aleatoria Bernoulli, $\mathbf{X} = (X_1, \dots, X_n)$ y contar la cantidad de éxitos observados: $N = \sum_{i=1}^n X_i$.

2. Si la cantidad de éxitos observados es menor que $np_0 - n\epsilon$ o mayor que $np_0 + n\epsilon$ se rechaza la hipótesis $p = p_0$ y se decide que $p \neq p_0$. En caso contrario, se no se rechaza la hipótesis $p = p_0$.

Segunda fase: caracterizar la función de potencia

La *segunda fase* del programa consiste en "calcular" la función de potencia. Esta función permite calcular los riesgos de tomar decisiones erróneas:

$$\begin{aligned}
\beta(p) &= \mathbb{P}\,(\text{Rechazar } H_0 |\, p) = \mathbb{P}_p(\delta(\mathbf{X}) = 1) \\
&= \mathbb{P}_p\,(N < np_0 - n\epsilon) + \mathbb{P}_p\,(N > np_0 + n\epsilon) \\
&= \sum_{k=0}^{[np_0-n\epsilon]} \binom{n}{k} p^k (1-p)^{n-k} + \sum_{k=[np_0-n\epsilon]+1}^{n} \binom{n}{k} p^k (1-p)^{n-k}. \qquad (37)
\end{aligned}$$

Notar que la función de potencia resultó ser un complicado polinomio de grado n y no es fácil capturar a simple vista su comportamiento cualitativo.

Nivel de significación. Debido a que la hipótesis fundamental es de la forma $p = p_0$, para cada n y ϵ, *el nivel de significación del test* es

$$\alpha(\delta) = \beta(p_0) = \sum_{k=0}^{[np_0-n\epsilon]} \binom{n}{k} p_0^k (1-p_0)^{n-k} + \sum_{k=[np_0-n\epsilon]+1}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-k}. \qquad (38)$$

Nota Bene 1. Notar que los test (36) contienen un juego de dos parámetros, n y ϵ. Estos parámetros determinan la calidad de cada test y deben ajustarse de acuerdo con las prescripciones impuestas al test sobre su nivel de significación y su potencia en alguna hipótesis alternativa.

Nota Bene 2. Notar que si la muestra tiene volumen prefijado n, por más que se mueva el valor de ϵ, el nivel de significación del test $\alpha(\delta)$ puede tomar a lo sumo $n + 1$ valores distintos. Por lo tanto, si se prescribe que el nivel de significación del test $\delta(\mathbf{X})$ debe ser α, casi seguramente la ecuación $\alpha(\delta) = \alpha$ no tendrá solución.

Aproximación por TCL para muestras "grandes"

La función de potencia (37) se puede aproximar utilizando el teorema central del límite. Si la muestra es suficientemente grande, para cada valor de p, tenemos que

$$Z = \frac{N - np}{\sqrt{np(1-p)}} \sim \mathcal{N}(0, 1).$$

Esto permite aproximar el valor de $\beta(p)$ de la siguiente manera

$$\begin{aligned}\beta(p) &= \mathbb{P}_p\left(Z < \frac{n(p_0 - p - \epsilon)}{\sqrt{np(1-p)}}\right) + \mathbb{P}_p\left(Z > \frac{n(p_0 - p + \epsilon)}{\sqrt{np(1-p)}}\right)\\ &\approx \Phi\left(\frac{\sqrt{n}(p_0 - p - \epsilon)}{\sqrt{p(1-p)}}\right) + \Phi\left(\frac{\sqrt{n}(p - p_0 - \epsilon)}{\sqrt{p(1-p)}}\right). \qquad (39)\end{aligned}$$

Aunque la aproximación (39) pueda resultar "grosera" y no sea lo suficientemente buena para todos los posibles valores de p, permite capturar el comportamiento cualitativo de la función de potencia.

Nivel de significación. Poniendo $p = p_0$, la aproximación (39) permite observar que

$$\alpha(\delta) = \beta(p_0) = 2\Phi\left(\frac{-\sqrt{n}\epsilon}{\sqrt{p_0(1-p_0)}}\right). \qquad (40)$$

Esto indica que basta tomar n suficientemente grande para que $\beta(p_0)$ se ubique todo lo cerca del 0 que uno quiera. En otras palabras, el test puede construirse para garantizar que la probabilidad de rechazar la hipótesis $p = p_0$ cuando ella es verdadera sea todo lo chica que uno quiera.

La aproximación (40) se puede utilizar para ajustar los valores de los parámetros n y ϵ para que valga la desigualdad $\alpha(\delta) \leq \alpha$. Para ello basta observar que la desigualdad aproximada

$$2\Phi\left(\frac{-\sqrt{n}\epsilon}{\sqrt{p_0(1-p_0)}}\right) \leq \alpha \iff \frac{-\sqrt{n}\epsilon}{\sqrt{p_0(1-p_0)}} \leq z_{\alpha/2}. \qquad (41)$$

Por lo tanto, las soluciones de la desigualdad (41) serán todos los valores de $n \in \mathbb{N}$ y todos los valores de $\epsilon > 0$ que satisfagan

$$\frac{\sqrt{n}\epsilon}{\sqrt{p_0(1-p_0)}} \geq z_{1-\alpha/2}. \qquad (42)$$

Fijada una solución particular de (42), una alta dosis de paciencia permite calcular a mano el valor exacto del nivel de significación $\alpha(\delta)$ obtenido en (38) y comprobar si efectivamente satisface $\alpha(\delta) \leq \alpha$.

Test de hipótesis con nivel de significación aproximado. Basados en los argumentos y razonamientos anteriores, podemos diseñar test para decidir entre las hipótesis $H_0 : p = p_0$ contra $H_1 : p \neq p_0$ con nivel de significación "aproximadamente" α. Usando el diseño (36) para valores de n y ϵ que verifiquen la desigualdad (42) obtenemos

$$\delta(\mathbf{X}) = \mathbf{1}\left\{N < np_0 - z_{1-\alpha/2}\sqrt{np_0(1-p_0)}\right\} + \mathbf{1}\left\{N > np_0 + z_{1-\alpha/2}\sqrt{np_0(1-p_0)}\right\}. \qquad (43)$$

Potencia en una alternativa. El mismo problema se presenta cuando se prescribe una potencia β para una alternativa p_1. En esta situación trataremos de resolver la desigualdad $\beta(p_1) \geq \beta$. Nuevamente la aproximación (39) permite resolver el problema:

- Si $p_1 < p_0$ el segundo término en (39) es despreciable respecto del primero y entonces obtenemos la siguiente aproximación:

$$\beta(p_1) \approx \Phi\left(\frac{\sqrt{n}(p_0 - p_1 - \epsilon)}{\sqrt{p_1(1-p_1)}}\right). \tag{44}$$

- Si $p_1 > p_0$ el primer término es despreciable respecto del segundo y entonces obtenemos la siguiente aproximación:

$$\beta(p_1) \approx \Phi\left(\frac{\sqrt{n}(p_1 - p_0 - \epsilon)}{\sqrt{p_1(1-p_1)}}\right). \tag{45}$$

Para fijar ideas supongamos que $p_1 > p_0$. Razonando del mismo modo que antes se obtiene la siguiente solución "aproximada" de la inecuación $\beta(p_1) \geq \beta$:

$$\frac{\sqrt{n}(p_1 - p_0 - \epsilon)}{\sqrt{p_1(1-p_1)}} \geq z_\beta. \tag{46}$$

El razonamiento anterior muestra que, prefijados dos valores α y β, se pueden diseñar test de hipótesis de la forma (36) con prescripciones del siguiente tipo: nivel de significación menor o igual que α y/o potencia en una alternativa particular superior a β. $\square$

5.3. Hipótesis fundamental compuesta

Sea $\mathbf{X} = (X_1, \dots, X_n)$ una muestra aleatoria de una variable aleatoria con distribución Bernoulli(p), $p \in (0, 1)$. Basados en la muestra aleatoria $\mathbf{X}$ queremos construir test para decidir entre las hipótesis

$$H_0 : p \leq p_0 \qquad \text{contra} \qquad H_1 : p > p_0,$$

donde $p_0 \in (0, 1)$ es un valor arbitrario pero fijo.

Programa de actividades. Adaptaremos los argumentos y razonamientos desarrollados en la sección 5.2. La primera fase del programa consiste en construir test de hipótesis basados en la cantidad de éxitos de la muestra $N = \sum_{i=1}^n X_i$. La segunda fase del programa consiste en evaluar los riesgos de tomar decisiones erróneas con los test construidas: se trata de caracterizar analíticamente la función de potencia y estudiar sus propiedades cualitativas y cuantitativas: cálculo del nivel de significación y de la potencia en las hipótesis alternativas simples.

Test de hipótesis. En este caso resulta intuitivamente claro proponer test de forma

$$\delta(\mathbf{X}) = \mathbf{1}\{N > np_0 + n\epsilon\}, \tag{47}$$

donde n y ϵ son parámetros ajustables.

Función de potencia. Fijados n y ϵ la función de potencia del test es

$$\begin{aligned}\beta(p) &= \mathbb{P}\left(\text{rechazar} H_0 |\, p\right) = \mathbb{P}_p(\delta(\mathbf{X}) = 1) = \mathbb{P}_p(N > np_0 + n\epsilon) \\ &= \sum_{k=[np_0+n\epsilon]+1}^{n} \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned} \tag{48}$$

De acuerdo con el **Corolario 5.2** la función de potencia es creciente. Esto es intuitivamente claro si se piensa que cuando aumenta la probabilidad de cada éxito, la cantidad de éxitos debe aumentar.

Aproximación por TCL. Si el volumen de muestra es suficientemente grande, usando el teorema central del límite podemos obtener la siguiente expresión aproximada de la función de potencia

$$\beta(p) = \mathbb{P}_p\left(\frac{N - np}{\sqrt{np(1-p)}} > \frac{np_0 + n\epsilon - np}{\sqrt{np(1-p)}}\right) \approx \Phi\left(\frac{\sqrt{n}(p - p_0 - \epsilon)}{\sqrt{p(1-p)}}\right). \tag{49}$$

Nivel de significación. Como la función de potencia es creciente, el nivel de significación del test se obtiene de la siguiente manera

$$\alpha(\delta) = \max_{p \le p_0} \beta(p) = \beta(p_0) = \sum_{k=[np_0+n\epsilon]+1}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-k} \approx \Phi\left(\frac{-\sqrt{n}\epsilon}{\sqrt{p_0(1-p_0)}}\right). \tag{50}$$

La aproximación en (50) presupone que el volumen de muestra es suficientemente grande (por ejemplo, $np_0(1-p_0) > 10$).

Prefijados un volumen de muestra suficientemente grande y un nivel de significación α para el test de hipótesis, la aproximación (50) permite hallar el valor de ϵ

$$z_{1-\alpha}\sqrt{p_0(1-p_0)} = \sqrt{n}\epsilon. \tag{51}$$

Test de hipótesis con nivel de significación aproximado. Usando el diseño (47) y el resultado obtenido en (51) se deduce que, para n suficientemente grande y fijo, la forma del test de hipótesis de nivel de significación α para decidir entre $H_0 : p \le p_0$ contra $H_1 : p > p_0$ es

$$\delta(\mathbf{X}) = \mathbf{1}\left\{N > np_0 + z_{1-\alpha}\sqrt{np_0(1-p_0)}\right\}. \tag{52}$$

Potencia en una alternativa. El análisis de la potencia en las hipótesis alternativas simples $p = p_1$, con $p_1 > p_0$, se realiza siguiendo las mismas líneas desarrolladas en la sección anterior. □

Ejemplo 5.3. Un productor de chips afirma que no más del 2 % de los chips que produce son defectuosos. Una compañía electrónica (impresionada por dicha afirmación) le compra una gran cantidad de chips. Para determinar si la afirmación del productor se puede tomar literalmente, la compañía decide testear una muestra de 300 de esos chips. Si se encuentra que 10 de los 300 chips son defectuosos, debería rechazarse la afirmación del productor?

Solución. Formalmente, el problema consiste en construir un test de hipótesis para decidir entre

$$H_0 : p \leq 0.02 \qquad \text{contra} \qquad H_1 : p > 0.02.$$

sobre la base de una muestra de volumen 300.

Fijado un nivel de significación, por ejemplo $\alpha = 0.05$, el test de hipótesis (52) adopta la forma

$$\begin{aligned}\delta(\mathbf{X}) &= \mathbf{1}\left\{N > 300(0.02) + z_{0.95}\sqrt{300(0.02)(0.98)}\right\} = \mathbf{1}\{N > 9.9886\} \\ &= \mathbf{1}\{N \geq 10\}. \end{aligned} \tag{53}$$

Dicho en palabras, al nivel del 5 % de significación, un test para decidir entre las hipótesis $H_0 : p \leq 0.02$ contra $H_1 : p > 0.02$, basado en una muestra de volumen 300, consiste en rechazar la hipótesis H_0 siempre que se observen 10 o más éxitos.

Traducido al problema que estamos examinando, el criterio de decisión puede enunciarse de la siguiente manera: "*examinar 300 componentes. Si se observan 10 o más defectuosos debe rechazarse la afirmación del productor de que produce con una calidad de a lo sumo un* 2 %, *si se observan menos de* 10 *defectuosos no hay evidencia suficiente para rechazar su afirmación.*"

En conclusión, como en la muestra examinada se observaron 10 chips defectuosos, al nivel del 5 % de significación, la afirmación del productor debe rechazarse. □

6. Test para varianza de normales

El objetivo de esta sección es ilustrar cómo se pueden obtener test de hipótesis usando intervalos de confianza.

6.1. Hipótesis sobre varianza con media conocida

Usando intervalos de confianza para la varianza de una distribución normal $\mathcal{N}(\mu, \sigma^2)$ con media μ conocida vamos a construir test de hipótesis de nivel de significación α para decidir entre

$$H_0 : \sigma^2 = \sigma_0^2 \qquad \text{contra} \qquad H_1 : \sigma^2 \neq \sigma_0^2,$$

para algún valor σ_0^2 determinado.

Dada una muestra aleatoria $\mathbf{X} = (X_1, \dots, X_n)$ de la distribución normal $\mathcal{N}(\mu, \sigma^2)$ con media μ conocida, sabemos que

$$I(\mathbf{X}) = \left[\frac{n\widehat{\sigma^2}_{mv}}{\chi^2_{n,\,(1+\beta)/2}}, \frac{n\widehat{\sigma^2}_{mv}}{\chi^2_{n,\,(1-\beta)/2}}\right],$$

donde $n\widehat{\sigma^2}_{mv} = \sum_{i=1}^n (X_i - \mu)^2$, es un intervalo de confianza para σ^2 de nivel β. Poniendo $\beta = 1 - \alpha$ se obtiene el siguiente test de nivel α para decidir entre las hipótesis $H_0: \sigma^2 = \sigma_0^2$ contra $H_1: \sigma^2 \neq \sigma_0^2$

$$\begin{aligned} \delta(\mathbf{X}) &= \mathbf{1}\{I(\mathbf{X}) \not\ni \sigma_0^2\} \\ &= \mathbf{1}\left\{\frac{1}{\sigma_0^2}\sum_{i=1}^n (X_i - \mu)^2 < \chi^2_{n,\,\alpha/2}\right\} + \mathbf{1}\left\{\frac{1}{\sigma_0^2}\sum_{i=1}^n (X_i - \mu)^2 > \chi^2_{n,\,1-\alpha/2}\right\}. \end{aligned} \tag{54}$$

Función de potencia. Para calcular y analizar el comportamiento de la función de potencia,

$$\beta(\sigma^2) = \mathbb{P}(\text{Rechazar} H_0 | \sigma^2),$$

debe recordarse que cuando el verdadero valor de la varianza es σ^2, la variable aleatoria $\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu)^2$ tiene distribución $\chi^2_n = \Gamma(n/2, 1/2)$. Multiplicando por $\frac{\sigma_0^2}{\sigma^2}$ en las desigualdades dentro de las llaves en la fórmula del test (54), y "calculando" las correspondientes probabilidades, obtenemos la siguiente expresión

$$\beta(\sigma^2) = \int_0^{a(\sigma^2)} \frac{(1/2)^{n/2}}{\Gamma(n/2)} x^{(n/2)-1} e^{-\frac{1}{2}x} dx + \int_{b(\sigma^2)}^{\infty} \frac{(1/2)^{n/2}}{\Gamma(n/2)} x^{(n/2)-1} e^{-\frac{1}{2}x} dx,$$

donde

$$a(\sigma^2) = \frac{\sigma_0^2}{\sigma^2}\chi^2_{n,\,\alpha/2}, \qquad b(\sigma^2) = \frac{\sigma_0^2}{\sigma^2}\chi^2_{n,1-\alpha/2}.$$

□

Ejemplo 6.1. Dada una muestra aleatoria de volumen 10 de una población normal de media 0 se quiere construir un test de nivel $\alpha = 0.05$ para decidir entre las hipótesis $H_0: \sigma^2 = 1$ contra $H_1: \sigma^2 \neq 1$.

Solución. Como $\chi^2_{10,\,0.025} = 3.247$ y $\chi^2_{10,\,0.975} = 20.483$, el test de hipótesis (54) adopta la forma

$$\delta(\mathbf{X}) = \mathbf{1}\left\{\sum_{i=1}^n X_i^2 < 3.247\right\} + \mathbf{1}\left\{\sum_{i=1}^n X_i^2 > 20.483\right\}. \tag{55}$$

□

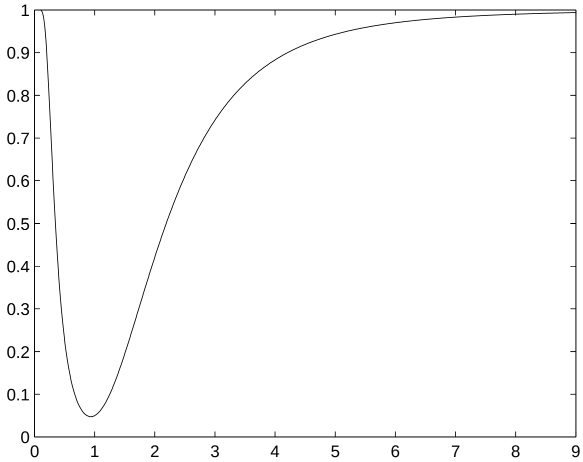

Figura 10: Gráfico de la función de potencia del test (55).

6.2. Hipótesis sobre varianza con media desconocida

Usando intervalos de confianza para la varianza de una distribución normal $\mathcal{N}(\mu, \sigma^2)$ vamos a construir test de hipótesis de nivel de significación α para decidir entre

$$H_0 : \sigma^2 = \sigma_0^2 \qquad \text{contra} \qquad H_1 : \sigma^2 \neq \sigma_0^2,$$

para algún valor σ_0^2 determinado.

Dada una muestra aleatoria $\mathbf{X} = (X_1, \dots, X_n)$ de la distribución normal $\mathcal{N}(\mu, \sigma^2)$ sabemos que

$$I(\mathbf{X}) = \left[\frac{(n-1)S^2}{\chi^2_{n-1,\,1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,\,\alpha/2}}\right],$$

es un intervalo de confianza para σ^2 de nivel β. Poniendo $\beta = 1-\alpha$ se obtiene el siguiente test de nivel α para decidir entre las hipótesis $H_0 : \sigma^2 = \sigma_0^2$ contra $H_1 : \sigma^2 \neq \sigma_0^2$

$$\begin{aligned} \delta(\mathbf{X}) &= \mathbf{1}\{I(\mathbf{X}) \not\ni \sigma_0^2\} \\ &= \mathbf{1}\left\{\frac{(n-1)S^2}{\sigma_0^2} < \chi^2_{n-1,\,\alpha/2}\right\} + \mathbf{1}\left\{\frac{(n-1)S^2}{\sigma_0^2} > \chi^2_{n-1,\,1-\alpha/2}\right\}. \end{aligned} \tag{56}$$

Función de potencia. Notar que el análisis de función de potencia de test (56) es completamente análogo al desarrollado para el caso en que suponíamos que la media μ es conocida. $\square$

Nota Bene. Notar que los test de hipótesis definidas en (54) y (56) son inmediatamente útiles para tomar decisiones.

Ejemplo 6.2. En la Sección dedicada al estudio de intervalos de confianza mostramos que cuando una muestra aleatoria $\mathbf{X}$ (de volumen 8) de una población normal $\mathcal{N}(\mu, \sigma^2)$ arroja los valores $9, 14, 10, 12, 7, 13, 11, 12$, el intervalo $I_{\sigma^2} = [2.248,\ 21.304]$ es un intervalo de confianza de nivel $\beta = 0.95$ para la varianza σ^2.

Si se quiere decidir al 5 % de significación entre las hipótesis

$$H_0 : \sigma^2 = 4 \qquad \text{contra} \qquad H_1 : \sigma^2 \neq 4.$$

el test de hipótesis (56) conduce a no rechazar la hipótesis $\sigma^2 = 4$. $\square$

7. Comparación de dos muestras

7.1. Test para medias de dos muestras normales.

Sean $\mathbf{X} = (X_1, \dots, X_m)$ e $\mathbf{Y} = (Y_1, \dots, Y_n)$ dos muestras aleatorias independientes de distribuciones normales $\mathcal{N}(\mu_X, \sigma_X^2)$ y $\mathcal{N}(\mu_Y, \sigma_Y^2)$, respectivamente. Sea $\Delta = \mu_X - \mu_Y$. Queremos un test para decidir entre las hipótesis

$$H_0 : \Delta = 0 \qquad \text{contra} \qquad H_1 : \Delta > 0.$$

7.1.1. Varianzas conocidas

Supongamos que las varianzas σ_X^2 y σ_Y^2 son conocidas. Para construir el test de hipótesis usaremos los estimadores de media: $\bar{X}$ y $\bar{Y}$. Puesto que

$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\Delta, \frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}\right)$$

el test de nivel α decidir entre $H_0 : \Delta = 0$ contra $H_1 : \Delta > 0$ es

$$\delta(\mathbf{X}, \mathbf{Y}) = \mathbf{1}\left\{\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} > z_{1-\alpha}\right\}$$

$\square$

7.1.2. Varianzas desconocidas pero iguales.

Supongamos las varianzas $\sigma_X^2 = \sigma_Y^2 = \sigma^2$. En tal caso, bajo la hipótesis $\Delta = 0$ tenemos que

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma^2}\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim \mathcal{N}(0, 1).$$

Para estimar la varianza σ^2 ponderamos "adecuadamente" los estimadores de varianza S_X^2 y S_Y^2,

$$S_P^2 := \frac{m-1}{m+n-2} S_X^2 + \frac{n-1}{m+n-2} S_Y^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}.$$

Se puede mostrar que

$$U = \frac{(n+m-2)}{\sigma^2} S_P^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{\sigma^2} \sim \chi_{n+m-2}.$$

Debido a que las variables Z y U son independientes, tenemos que

$$T = \frac{Z}{\sqrt{U/(m+n-2)}} = \frac{\bar{X} - \bar{Y}}{\sqrt{S_P^2}\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t_{m+n-2}$$

Por lo tanto,

$$\delta(\mathbf{X}, \mathbf{Y}) = \mathbf{1}\left\{ \frac{\bar{X} - \bar{Y}}{\sqrt{S_P^2}\sqrt{\frac{1}{m} + \frac{1}{n}}} > t_{m+n-2,\,1-\alpha} \right\}.$$

es un test de nivel de significación α para decidir entre las hipótesis $H_0 : \Delta = 0$ contra $H_1 : \Delta > 0$. □

7.2. Test F para varianzas de normales.

Sean $\mathbf{X} = (X_1, \dots, X_m)$ e $\mathbf{Y} = (Y_1, \dots, Y_n)$ dos muestras aleatorias independientes de distribuciones normales $\mathcal{N}(\mu_X, \sigma_X^2)$ y $\mathcal{N}(\mu_Y, \sigma_Y^2)$, respectivamente. Sea $R = \sigma_X^2/\sigma_Y^2$. Queremos un test para decidir entre las hipótesis

$$H_0 :\ R = 1 \qquad \text{contra} \qquad H_1 :\ R \neq 1.$$

Las varianzas σ_X^2 y σ_Y^2 se pueden estimar mediante sus estimadores insesgados S_X^2 y S_Y^2. Las variables

$$U = \frac{(m-1)}{\sigma_X^2} S_X^2 \sim \chi_{m-1}^2 \qquad \text{y} \qquad V = \frac{(n-1)}{\sigma_Y^2} S_Y^2 \sim \chi_{n-1}^2$$

son independientes.

Test de hipótesis. Bajo la hipótesis $H_0 :\ R = 1$, vale que

$$F = \frac{S_X^2}{S_Y^2} = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{m-1,\,n-1}.$$

Por lo tanto,

$$\delta(\mathbf{X},\, \mathbf{Y}) = \mathbf{1}\left\{F \notin [\phi_1, \phi_2]\right\}, \tag{57}$$

donde ϕ_1 y ϕ_2 son tales que $\mathbb{P}(F < \phi_1) = \mathbb{P}(F > \phi_2) = \alpha/2$, es un test de nivel α para decidir entre las hipótesis $H_0 :\ R = 1$ contra $H_1 :\ R \neq 1$.

Ejemplo 7.1. Queremos construir un test de nivel $\alpha = 0.05$ para decidir entre $H_0 : \mathbb{R} = 1$ contra $H_1 : R \neq 1$ usando muestras $\mathbf{X}$ y $\mathbf{Y}$ de volumen $m = n = 10$.

Proponemos un test de la forma (57). El problema se reduce determinar valores ϕ_1 y ϕ_2 tales que

$$\mathbb{P}(F_{9,9} > \phi_2) = 0.025 \qquad \text{y} \qquad \mathbb{P}(F_{9,9} < \phi_1) = 0.025.$$

Usando las tablas de las distribuciones F resulta que $\phi_2 = 4.5362$ y que $\phi_1 = 1/\phi_2 = 0.2204$.

Finalmente, se obtiene el test

$$\delta(\mathbf{X}, \mathbf{Y}) = \{F \notin [0.2204, 4.5362]\}.$$

□

7.3. Planteo general

Supongamos que tenemos dos muestras aleatorias independientes $\mathbf{X} = (X_1, \dots, X_m)$ e $\mathbf{Y} = (Y_1, \dots, Y_n)$ con distribuciones dependientes de los parámetros ξ y η, respectivamente. Sea $\Delta = \xi - \eta$.

Se quiere decidir entre la hipótesis fundamental

$$H_0 : \Delta = \delta_0$$

contra cualquiera de las hipótesis alternativas:

$$\begin{array}{ll} \text{(a)} & H_1 : \Delta > \delta_0; \\ \text{(b)} & H_1 : \Delta < \delta_0; \\ \text{(c)} & H_1 : \Delta \neq \delta_0. \end{array}$$

Sabemos que si dos estimadores para ξ y η, $\hat{\xi}_m$ y $\hat{\eta}_n$, tienen la propiedad de normalidad asintótica

$$\begin{array}{rl} \sqrt{m}(\hat{\xi}_m - \xi) \to \mathcal{N}(0, \sigma^2) & \qquad \text{cuando } m \to \infty, \\ \sqrt{n}(\hat{\eta}_n - \eta) \to \mathcal{N}(0, \tau^2) & \qquad \text{cuando } n \to \infty, \end{array}$$

donde σ^2 y τ^2 pueden depender de ξ y η, respectivamente y ninguna de las variables está sobre-representada (i.e., m y n son del mismo orden de magnitud), entonces

$$\frac{(\hat{\xi}_m - \hat{\eta}_n) - (\xi - \eta)}{\sqrt{\frac{\sigma^2}{m} + \frac{\tau^2}{n}}} \to \mathcal{N}(0, 1) \tag{58}$$

Si σ^2 y τ^2 son conocidas, de (58) resulta que las regiones de rechazo:

$$\text{(a)}\quad \frac{(\hat{\xi}_m - \hat{\eta}_n) - \delta_0}{\sqrt{\frac{\sigma^2}{m} + \frac{\tau^2}{n}}} > z_{1-\alpha};$$

$$\text{(b)}\quad \frac{(\hat{\xi}_m - \hat{\eta}_n) - \delta_0}{\sqrt{\frac{\sigma^2}{m} + \frac{\tau^2}{n}}} < z_{\alpha};$$

$$\text{(c)}\quad \left|\frac{(\hat{\xi}_m - \hat{\eta}_n) - \delta_0}{\sqrt{\frac{\sigma^2}{m} + \frac{\tau^2}{n}}}\right| > z_{1-\alpha/2}$$

producen un test para H_0 contra H_1 de nivel asintótico α, para cada uno de los casos considerados, respectivamente.

Si σ^2 y τ^2 son desconocidas y $\widehat{\sigma^2}$ y $\widehat{\tau^2}$ son estimadores consistentes para σ^2 y τ^2, se puede demostrar que las regiones de rechazo conservan su validez cuando σ^2 y τ^2 se reemplazan por $\widehat{\sigma^2}$ y $\widehat{\tau^2}$, respectivamente y entonces el test con región de rechazo

$$\text{(a)}\quad \frac{(\hat{\xi}_m - \hat{\eta}_n) - \delta_0}{\sqrt{\frac{\widehat{\sigma^2}}{m} + \frac{\widehat{\tau^2}}{n}}} > z_{1-\alpha};$$

$$\text{(b)}\quad \frac{(\hat{\xi}_m - \hat{\eta}_n) - \delta_0}{\sqrt{\frac{\widehat{\sigma^2}}{m} + \frac{\widehat{\tau^2}}{n}}} < z_{\alpha};$$

$$\text{(c)}\quad \left|\frac{(\hat{\xi}_m - \hat{\eta}_n) - \delta_0}{\sqrt{\frac{\widehat{\sigma^2}}{m} + \frac{\widehat{\tau^2}}{n}}}\right| > z_{1-\alpha/2}$$

también tiene nivel asintótico α.

Para mayores detalles se puede consultar el libro Lehmann, E. L. (1999) *Elements of Large-Sample Theory*. Springer, New York.

Nota Bene. Notar que el argumento anterior proporciona un método general de naturaleza asintótica. En otras palabras, en la práctica los resultados que se obtienen son aproximados. Dependiendo de los casos particulares existen diversos refinamientos que permiten mejorar esta primera aproximación.

7.4. Problema de dos muestras binomiales

Sean $\mathbf{X} = (X_1, \dots, X_m)$ e $\mathbf{Y} = (Y_1, \dots, Y_n)$ dos muestras aleatorias independientes de dos variables aleatorias X e Y con distribución Bernoulli de parámetros p_X y p_Y, respectivamente. Sea $\Delta = p_X - p_Y$. Queremos un test para decidir entre las hipótesis

$$H_0 : \Delta = 0 \qquad \text{contra} \qquad H_1 : \Delta > 0$$

Para construir el test usaremos los estimadores de máxima verosimilitud para las probabilidades p_x y p_Y, $\hat{p}_X = \bar{X}$ y $\hat{p}_Y = \bar{Y}$.

Vamos a suponer que los volúmenes de las muestras, m y n, son suficientemente grandes y que ninguna de las dos variables está sobre representada.

Puesto que $\bar{X}$ y $\bar{Y}$ son estimadores consistentes para las probabilidades p_X y p_Y, resulta que los estimadores $\bar{X}(1-\bar{X})$ y $\bar{Y}\left(1-\bar{Y}\right)$ son consistentes de las varianzas $p_X(1-p_X)$ y $p_Y(1-p_Y)$, respectivamente. Por lo tanto,

$$\delta(\mathbf{X}, \mathbf{Y}) = \mathbf{1}\left\{\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{1}{m}\bar{X}\left(1-\bar{X}\right)+\frac{1}{n}\bar{Y}\left(1-\bar{Y}\right)}} > z_{1-\alpha}\right\}$$

es un test, de nivel aproximado α, para decidir entre las hipótesis $H_0 : \Delta = 0$ contra $H_1 : \Delta > 0$. □

Nota Bene. Observar que el nivel del test se calcula bajo la hipótesis $p_X = p_Y$, en tal caso la desviación estándar de la diferencia $\bar{X}-\bar{Y}$ es de la forma

$$\sqrt{\frac{p_X(1-p_X)}{m}+\frac{p_Y(1-p_Y)}{n}} = \sqrt{p_X(1-p_X)}\sqrt{\frac{1}{m}+\frac{1}{n}}$$

y podemos estimarla mediante

$$\sqrt{\frac{m\bar{X}+n\bar{Y}}{m+n}\left(1-\frac{m\bar{X}+n\bar{Y}}{m+n}\right)}\sqrt{\frac{1}{m}+\frac{1}{n}}.$$

Lo que produce el test

$$\delta(\mathbf{X}, \mathbf{Y}) = \mathbf{1}\left\{\frac{\left(\bar{X}-\bar{Y}\right)\sqrt{mn}}{\sqrt{(m\bar{X}+n\bar{Y})\left(1-\frac{m\bar{X}+n\bar{Y}}{m+n}\right)}} > z_{1-\alpha}\right\} \tag{59}$$

□

Ejemplo 7.2. Se toma una muestra aleatoria de 180 argentinos y resulta que 30 están desocupados. Se toma otra muestra aleatoria de 200 uruguayos y resulta que 25 están desocupados. ¿Hay evidencia suficiente para afirmar que la tasa de desocupación de la población Argentina es superior a la del Uruguay?

Solución. La población desocupada de la Argentina puede modelarse con una variable aleatoria $X \sim$ Bernoulli(p_X) y la del Uruguay con una variable aleatoria $Y \sim$ Bernoulli(p_Y).

Para resolver el problema utilizaremos un test de nivel de significación $\alpha = 0.05$ para decidir entre las hipótesis

$$H_0 : p_X = p_Y \qquad \text{contra} \qquad H_1 : p_X > p_Y$$

basada en dos muestras aleatorias independientes $\mathbf{X}$ e $\mathbf{Y}$ de volúmenes $m = 180$ y $n = 200$, respectivamente.

El test de hipótesis dado en (59) adopta la forma

$$\delta(\mathbf{X}, \mathbf{Y}) = \mathbf{1}\left\{ \frac{\left(\bar{X} - \bar{Y}\right)\sqrt{36000}}{\sqrt{(180\bar{X} + 200\bar{Y})\left(1 - \frac{180\bar{X}+200\bar{Y}}{380}\right)}} > 1.64 \right\} \tag{60}$$

De acuerdo con los datos observados $\bar{X} = 30/180$ y $\bar{Y} = 25/200$:

$$\frac{\left(\frac{30}{180} - \frac{25}{200}\right)\sqrt{36000}}{\sqrt{55\left(1 - \frac{55}{380}\right)}} = 1.152\ldots$$

Debido a que $1.152\ldots < 1.64$, no hay evidencia suficiente para rechazar la hipótesis $p_X = p_Y$. Por lo tanto, con un $5\,\%$ de nivel de significación, no hay evidencia suficiente para afirmar que la tasa de desocupación en la Argentina sea superior a la del Uruguay. □

8. Test de la χ^2 para bondad de ajuste

8.1. Planteo del problema

Los test de bondad de ajuste tienen por objeto decidir si los datos observados se ajustan a una determinada distribución de probabilidades. Más precisamente, se formula una hipótesis, H, que afirma que los datos observados constituyen una muestra aleatoria $\mathbf{X} = (X_1, \ldots, X_n)$ de una distribución F. La distribución F puede estar completamente especificada (hipótesis simple) o puede pertenecer a una familia paramétrica (hipótesis compuesta).

Algunos ejemplos (para fijar ideas):

Ejemplo 8.1 (Moneda honesta)**.** En una sucesión de 100 lanzamientos independientes de una moneda se observaron 55 caras y 45 cecas ¿Estos datos son compatibles con la hipótesis de que la moneda es honesta?

Ejemplo 8.2 (Multinomial)**.** Para identificar las obras de su serie titulada *Los paisajes binarios* el artista digital Nelo las firma con una imagen aleatoria de 10×10 pixels: por cada pixel lanza un dado equilibrado: si sale $1, 2$ o 3 lo pinta de rojo; si sale 4 o 5 lo pinta de verde y si sale 6 lo pinta de azul. Se somete a examen la firma de una obra digital titulada *Cordillera binaria* y se obtienen los siguientes resultados: 46 pixels rojos, 37 verdes y 17 azules. ¿La obra *Cordillera binaria* pertenece a la serie *Los paisajes binarios*?

Ejemplo 8.3 (Números aleatorios)**.** Se producen 10000 números con un generador de "números aleatorios". Para economizar espacio se registra la cantidad de números de la forma $0.d...$, donde $d = 0, 1, \ldots, 9$. Se obtuvieron los resultados siguientes:

d	0	1	2	3	4	5	6	7	8	9
$\#\{0.d...\}$	1008	1043	1014	1027	952	976	973	1021	998	988

(61)

¿Los datos se ajustan a una distribución uniforme $\mathcal{U}[0, 1]$?

Ejemplo 8.4 (Poisson)**.** Una partícula de polen suspendida en agua es bombardeada por moléculas en movimiento térmico. Se la observa durante una hora y se registra la cantidad de impactos que recibe por segundo. Sea X la variable aleatoria que cuenta la cantidad de impactos por segundo recibidos por la partícula. Se obtuvieron los siguientes datos

X	0	1	2	3	4	5	6
# de s. con X impactos	1364	1296	642	225	55	15	3

(62)

Se quiere decidir si los datos provienen de una distribución de Poisson.

Ejemplo 8.5 (Velocidad de la luz)**.** En la siguiente tabla se muestran las mediciones de la velocidad de la luz realizadas por el físico Albert Michelson entre el 5 de junio y el 5 de julio de 1879. Los valores dados + 299.000 son las mediciones de Michelson en km/s.

850	740	900	**1070**	930	850	950	980	980	880
1000	980	930	650	760	810	1000	1000	960	960
960	940	960	940	880	800	850	880	900	840
830	790	810	880	880	830	800	790	760	800
880	880	880	860	720	720	**620**	860	970	950
880	910	850	870	840	840	850	840	840	840
890	810	810	820	800	770	760	740	750	760
910	920	890	860	880	720	840	850	850	780
890	840	780	810	760	810	790	810	820	850
870	870	810	740	810	940	950	800	810	870

(63)

Las mediciones de la velocidad de la luz de Michelson, ¿se ajustan a una distribución normal?

8.2. Test de bondad de ajuste para hipótesis simples

La hipótesis nula afirma que

$$H_0 : F_X = F,$$

donde F es una distribución de probabilidades completamente determinada.

Si la hipótesis H_0 es verdadera, la función de distribución empírica, F_n de los n valores observados debe ser parecida a la función de distribución F. Lo que sugiere introducir alguna *medida de la discrepancia* entre ambas distribuciones y basar el test de hipótesis en las propiedades de la distribución de dicha medida.

Hay varias formas de construir esas medidas. La que sigue fue introducida por Karl Pearson.

Se divide el rango de la variable aleatoria X en una cantidad finita k de partes disjuntas dos a dos, $C_1, \dots, C_k$, llamadas *clases*[5] tales que las probabilidades $p_i = \mathbb{P}(X \in C_i | H_0) > 0$. Las k clases, C_i, serán los k conjuntos en los que agruparemos los datos para tabularlos. Se consideran $n_1, \dots, n_k$ las frecuencias de aparición de las clases $C_1, \dots, C_n$ en la muestra aleatoria $\mathbf{X} = (X_1, \dots, X_n)$,

$$n_i = \sum_{j=1}^{n} \mathbf{1}\{X_j \in C_i\} \qquad \text{y} \qquad \sum_{i=1}^{k} n_i = n.$$

Bajo la distribución hipotética la cantidad de valores muestrales n_i pertenecientes a la clase C_i se distribuye como una Binomial(n, p_i), y en consecuencia, para valores grandes de n, las frecuencias relativas $\frac{n_i}{n}$ deben tener valores muy próximos a las probabilidades p_i. La dispersión entre las frecuencias relativas $\frac{n_i}{n}$ y las probabilidades p_i se puede medir del siguiente modo

$$D^2 = \sum_{i=1}^{k} w_i \left(\frac{n_i}{n} - p_i\right)^2 = \sum_{i=1}^{k} w_i \frac{(n_i - np_i)^2}{n^2}, \tag{64}$$

donde los coeficientes $w_i > 0$ se pueden elegir de manera más o menos arbitraria. Cuando la hipótesis H_0 es verdadera los valores de la medida de dispersión D^2 deben ser pequeños, lo que sugiere diseñar un test de hipótesis que decida *rechazar la hipótesis H_0 cuando y solo cuando se observa que $D^2 > M$*, donde M es una constante arbitraria pero fija.

Karl Pearson demostró que cuando n es grande y la hipótesis H_0 es verdadera, poniendo $w_i = \frac{n}{p_i}$ en (64), la distribución de la medida de dispersión

$$D^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}, \tag{65}$$

es aproximadamente igual a una chi cuadrado con $k-1$ grados de libertad. (Una demostración de este resultado puede consultarse en: Cramer, H.: Métodos matemáticos de estadística. Aguilar, Madrid. (1970).)

Test de bondad de ajuste χ^2. Para decidir si la muestra aleatoria $\mathbf{X} = (X_1, \dots, X_n)$ proviene de la distribución F se puede adoptar el siguiente criterio:

$$\delta(\mathbf{X}) = \mathbf{1}\{D^2 > \chi^2_{k-1,\,1-\alpha}\}, \tag{66}$$

donde $\alpha \in (0, 1)$. Dicho en palabras, *rechazar que $F_X = F$ cuando y solo cuando la medida de dispersión D^2 definida en* (65) *supera al cuantil* $1-\alpha$ *de la distribución chi cuadrado con $k-1$ grados de libertad*. En tal caso, la probabilidad de rechazar H_0 cuando H_0 es verdadera es aproximadamente α. □

[5]Los valores de la variable aleatoria X pertenecen a una y solo a una de las clases $C_1, \dots, C_k$.

8.3. Ejemplos (1^a parte)

El siguiente ejemplo tiene la virtud de mostrar, en un caso particular, una línea de demostración del resultado de Pearson sobre la distribución asintótica de D^2.

Ejemplo 8.6 (Bernoulli)**.** Sea $\mathbf{X} = (X_1, \dots, X_n)$ una muestra aleatoria de una distribución Bernoulli con probabilidad de éxito p. Queremos testear la hipótesis $H_0 : p = p_0$ contra $H_1 : p \neq p_0$, donde $p_0 \in (0, 1)$ es un valor determinado.

La medida de dispersión definida en (65) entre las frecuencias observadas

$$n_1 = \sum_{i=1}^{n} X_i \qquad \text{y} \qquad n_2 = n - n_1$$

y las frecuencias esperadas

$$np_0 \qquad \text{y} \qquad n(1 - p_0)$$

tiene la siguiente expresión

$$D^2 = \frac{(n_1 - np_0)^2}{np_0} + \frac{(n - n_1 - n(1 - p_0))^2}{n(1 - p_0)}.$$

Observando que

$$\begin{aligned} \frac{(n_1 - np_0)^2}{np_0} + \frac{(n - n_1 - n(1 - p_0))^2}{n(1 - p_0)} &= \frac{(n_1 - np_0)^2}{np_0} + \frac{(np_0 - n_1)^2}{n(1 - p_0)} \\ &= \frac{(1 - p_0)(n_1 - np_0)^2 + p_0(n_1 - np_0)^2}{np_0(1 - p_0)} \\ &= \frac{(n_1 - np_0)^2}{np_0(1 - p_0)}, \end{aligned}$$

se obtiene que

$$D^2 = \left(\frac{n_1 - np_0}{\sqrt{np_0(1 - p_0)}} \right)^2 \tag{67}$$

Cuando la hipótesis H_0 es verdadera, $n_1 \sim$ Binomial (n, p_0), y de acuerdo con el teorema central del límite la distribución de la variable aleatoria

$$\frac{n_1 - np_0}{\sqrt{np_0(1 - p_0)}}$$

es asintóticamente normal $\mathcal{N}(0, 1)$. Por lo tanto, para valores grandes de n, D^2 tiene una distribución aproximadamente igual a χ^2_1. $\square$

Ejemplo 8.1. (Continuación) Se trata de un caso particular del esquema anterior, donde $p_0 = 1/2$ y $n = 100$. En consecuencia, la medida de dispersión (67) es

$$D^2 = \left(\frac{n_1 - 50}{5}\right)^2,$$

y para un nivel de significación α el test de hipótesis (66) adopta la forma

$$\delta(\mathbf{X}) = \mathbf{1}\left\{\left(\frac{n_1 - 50}{5}\right)^2 > \chi^2_{1,\,1-\alpha}\right\}.$$

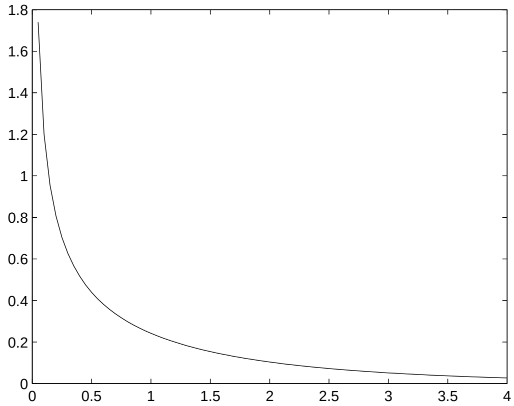

Figura 11: La densidad χ^2_1.

Consultado la tabla de cuantiles de la distribución χ^2_1 vemos que $\chi^2_{1,\,0.95} = 3.841$.

De acuerdo con los datos observados $n_1 = 55$, de donde sigue que como $D^2 = \left(\frac{55-50}{5}\right)^2 = 1$. En vista de que $1 < \chi^2_{1,\,0.95}$, a un nivel de significación del 5 % el test no rechaza la hipótesis de que se la moneda sea honesta. □

Ejemplo 8.2. (Continuación) El color en cada pixel se modela con una variable aleatoria X a valores $\{r, g, b\}$ cuya distribución está completamente determinada por los valores de las probabilidades $\mathbb{P}(X = r) = p_r$, $\mathbb{P}(X = g) = p_g$ y $\mathbb{P}(X = b) = p_b$. Queremos decidir si los datos obtenidos son compatibles (o no) con la hipótesis

$$H_0 : p_r = 3/6,\ p_g = 2/6,\ p_b = 1/6.$$

Para ello construimos un test de bondad de ajuste basado en una muestra aleatoria, $\mathbf{X} = (X_1, \dots, X_n)$ de volumen $n = 10 \times 10 = 100$. Prescrito el nivel de significación α y clasificando los datos de acuerdo con el color observado obtenemos un test de la forma

$$\delta(\mathbf{X}) = \mathbf{1}\{D^2 > \chi^2_{2,\,1-\alpha}\},$$

donde

$$D^2 = \frac{(n_r - 100(3/6))^2}{100(3/6)} + \frac{(n_g - 100(2/6))^2}{100(2/6)} + \frac{(n_b - 100(1/6))^2}{100(1/6)}.$$

Por ejemplo, si se prescribe un nivel de significación del 1 % (i.e., $\alpha = 0.01$) tenemos que $\chi^2_{2,\,1-\alpha} = \chi^2_{2,\,0.99} = 9.2103$ y el test adopta la forma

$$\delta(\mathbf{X}) = \mathbf{1}\left\{\frac{(n_r - 50)^2}{50} + \frac{(n_g - 33.33...)^2}{33.33...} + \frac{(n_b - 16.66...)^2}{16.66...} > 9.2103\right\},$$

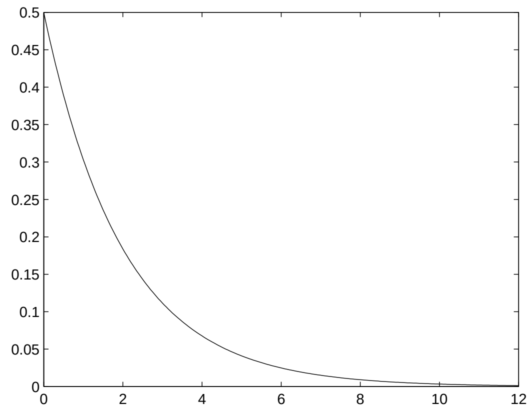

Figura 12: La densidad χ^2_2.

De acuerdo con los datos observados: $n_r = 46$, $n_g = 37$ y $n_b = 17$ y la medida de dispersión de Pearson vale

$$D^2 = \frac{(46 - 50)^2}{50} + \frac{(37 - 33.33...)^2}{33.33...} + \frac{(17 - 16.66...)^2}{16.66...} = 0.73$$

Motivo por el cual, no hay evidencia que permita rechazar que la obra *Cordillera binaria* pertenece a la serie *Los paisajes binarios* del artista Nelo.

Notar que para rechazar que la obra citada pertenece al artista se necesitaba un test de la forma $\delta(\mathbf{X}) = \{D^2 \geq 0.73\}$. Bajo la hipótesis H_0, $D^2 \sim \chi^2_2$ y $p = \mathbb{P}(D^2 \geq 0.73) = 0.694...$ y en ese caso, la probabilidad de equivocarse al rechazar que la obra pertenece a Nelo es del orden del 69 %. $\square$

Ejemplo 8.3. (Continuación) En este caso las clases C_i son los intervalos de la forma $\left[\frac{i-1}{10}, \frac{i}{10}\right)$, $i = 1, \dots, 10$. Si la variable aleatoria X tuviese distribución $\mathcal{U}[0,1]$, $p_i = \mathbb{P}(X \in C_i) = 1/10$. El volumen de la muestra es $n = 10000$. Las frecuencias observadas, n_i, son los valores que se muestran en la tabla (61). Las frecuencias esperadas, np_i, son todas iguales y valen 1000. Por lo tanto, la medida de dispersión de Pearson vale

$$D^2 = \frac{1}{1000}\left(8^2 + 43^2 + 14^2 + 27^2 + 48^2 + 24^2 + 27^2 + 21^2 + 2^2 + 12^2\right) = 7.036$$

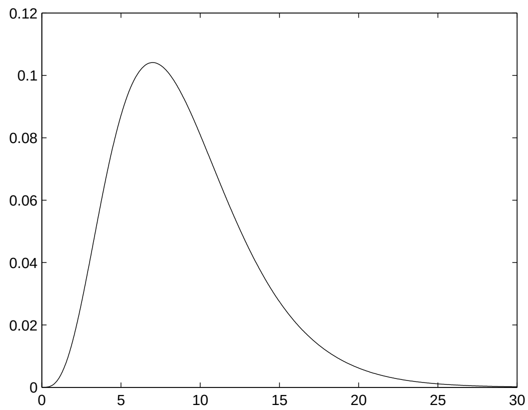

Figura 13: La densidad χ^2_9. El área bajo la curva a la derecha del valor 7.036 es 0.6336....

Bajo la hipótesis $X \sim \mathcal{U}[0,1]$, la medida de dispersión D^2 se distribuye como una chi cuadrado con 9 grados de libertad. Si se observa la Figura 13 se puede ver que un valor de 7.036 para D^2 no es inusual, lo que indica que no hay evidencia suficiente para rechazar la hipótesis $X \sim \mathcal{U}[0,1]$. Para rechazar dicha hipótesis se necesita un test de la forma $\delta(\mathbf{X}) = \{D^2 \geq 7.036\}$. Bajo la hipótesis $X \sim \mathcal{U}[0,1]$, $p = \mathbb{P}(D^2 \geq 7.036) = 0.6336...$ y en tal caso, la probabilidad de equivocarse al rechazar que los datos provienen de una distribución uniforme es del orden del 63 %. $\square$

8.4. Comentarios sobre el método

En la sección 8.2 presentamos el test de bondad de ajuste χ^2 de Pearson. En la sección 8.3 ilustramos su implementación en algunos ejemplos muy simples. Esos ejemplos comparten una característica en común: las clases en que dividimos el rango de la variable X estaban condicionadas por el modo en que estaban tabulados los datos observados.

Esos ejemplos podrían oscurecer el siguiente hecho que no puede pasar desapercibido: *el procedimiento de construcción de las clases $C_1, \dots, C_k$ en que se divide el rango de la variable es (más o menos) arbitrario.* En la descripción del método presentada en la sección 8.2 no se indica cuántas clases deben considerarse ni se indica cómo deben ser esas clases.

Sobre la cantidad de clases (1). Un lector desprevenido podría pensar que para implementar el método basta dividir el rango de la variable en dos clases. Ese modo de proceder no es recomendable. ¿Usando las clases, $C_1 = [-1, 0]$ y $C_2 = (0, 1]$, podrían distinguirse la distribución uniforme sobre el [-1,1] de la distribución triangular con el mismo soporte? Evidentemente no. Sin embargo, en cuanto aumentamos la cantidad de clases, a 4 por ejemplo, la diferencia se podría percibir.

Cuando agrupamos los datos en clases y conservamos solamente la frecuencia con que se observa cada clase destruimos información sobre la variable muestreada. Si la cantidad de partes es muy chica, se pierde mucha información y la resolución del test es bastante mala. □

Sobre la cantidad y la forma de las clases (2). Se podría pensar que al aumentar la cantidad de clases en que se divide el rango de la variable mejora la resolución del test, esto es parcialmente correcto. Si nos excedemos en la cantidad de clases la distribución de la medida de dispersión D^2 deja de parecerse a la χ^2.

Debido a su naturaleza asintótica, el test de bondad de ajuste χ^2 funciona bien solamente cuando las frecuencias esperadas en todas las clases es relativamente grande. En la Bibliografía consultada no se comenta ningún método "óptimo" para determinar la cantidad de clases en que debe dividirse el rango de la variable aleatoria. Aunque sobre este asunto parece no existir acuerdo entre los especialistas, todos coinciden en que la cantidad de clases está limitada por una condición del siguiente tipo:

- $np_i \geq 5$ para $i = 1, \dots, k$ (Fisher);
- $np_i \geq 10$ para $i = 1, \dots, k$ (Cramer);
- $np_i \geq 8$ para $i = 1, \dots, k$ (Borovkov).

DeGroot indica que la condición de Fisher es suficiente para que la distribución χ^2 sea una *buena aproximación* de la distribución de D^2. Incluso afirma que, poniendo $np_i > 1.5$ la aproximación continua siendo satisfactoria. □

En todo lo que sigue adoptaremos la condición de Cramer sobre la cantidad y forma de las clases: $np_i \geq 10$ para $i = 1, \dots, k$. De este modo, si para algún i ocurriese que $np_i < 10$ redefinimos la partición $C_1, \dots, C_k$ del rango de la variable. Por ejemplo, uniendo C_i con C_{i+1}. Esta condición implica que si el volumen de la muestra no es muy grande, la partición del rango de la variable no puede ser muy fina.

Ejemplo 8.7 (Exponencial)**.** Se dispone de los siguientes datos sobre la duración en horas de 100 baterías:

3.9662191	0.5819433	0.1842986	0.5977917	1.9781844
0.6048519	0.7259459	1.5896094	0.2411217	2.4502631
1.6993148	0.9884268	0.4281823	2.0079459	0.0022114
0.0422904	1.6384416	0.2214073	0.4350003	0.1934794
0.3548681	0.7775309	0.1052627	0.6497803	0.7227835
3.0542040	3.4097021	0.3577800	1.4532404	2.2825177
1.4903543	0.6062705	0.9444304	0.1119637	1.2789623
0.3598502	0.8901427	0.1282656	0.3331565	1.6096607
1.3348741	3.1158026	0.4525998	0.4554032	0.8698826
0.0215405	0.7115861	0.4859616	1.3781469	0.0979241
0.8608390	0.1999889	0.6616866	0.6960469	1.4041375
1.6087253	0.2149426	0.4833662	2.3159498	1.0346222

0.2056717	0.5228204	1.8704697	0.2166610	0.9409121
3.4983549	0.3543629	1.5233421	0.1877053	0.3911424
0.1840173	1.1453108	0.0161651	1.7702696	1.0397349
0.0772446	0.0421012	0.4814322	2.5107661	1.6500077
1.2448903	0.1030540	0.4572152	0.6299386	0.1021735
0.2197928	1.1234052	0.0936486	1.6546837	3.1267264
1.4791009	0.3132625	1.0092715	1.2217523	3.2381804
0.1215625	0.7677260	0.2124635	2.2532736	0.7156024

¿Puede afirmarse a un nivel del 1 % que la duración de las baterías se ajusta a una distribución exponencial de media 2 horas?

Solución.
1. Construyendo una partición. Lo primero que tenemos que hacer es determinar la cantidad y la forma de las clases en que agruparemos los datos.

Con la indicación de Cramer ($np_i \geq 10$, para $i = 1, \dots, k$) la máxima cantidad de clases que podemos elegir es 10. Para simplificar un poco las cuentas elegiremos una partición en 7 clases, $C_1, \dots, C_7$, que sean equiprobables bajo la distribución hipotética: $X \sim$ Exponencial$(1/2)$.[6]

Cuando la función de distribución de una variable aleatoria es continua la construcción de la partición en k clases equiprobables se resuelve utilizando los cuantiles. La clase C_i será el intervalo $\left[x_{\frac{i-1}{k}}, x_{\frac{i}{k}}\right)$, donde $x_{\frac{i}{k}}$ es el cuantil-$\frac{i}{k}$ de la distribución hipotética.

La función de distribución de la exponencial de media 2 es $F(x) = (1 - e^{-x/2})\mathbf{1}\{x \geq 0\}$ y su cuantil-γ es la única solución de la ecuación $F(x_\gamma) = \gamma$. En consecuencia, $x_\gamma = -2\log(1-\gamma)$. En consecuencia, para obtener 7 clases equiprobables basta poner

$$C_i = \left[-2\log\left(1 - \frac{i-1}{7}\right), -2\log\left(1 - \frac{i}{7}\right)\right), \qquad i = 1, \dots, 7,$$

lo que produce: $C_1 = [0, 0.3083)$, $C_2 = [0.3083, 0.6729)$, $C_3 = [0.6729, 1.1192)$, $C_4 = [1.1192, 1.6946)$, $C_5 = [1.6946, 2.5055)$, $C_6 = [2.5055, 3.8918)$ y $C_7 = [3.8918, \infty)$.

2. Agrupando los datos. Determinadas las clases agrupamos los datos. En la siguiente tabla se muestran las frecuencias observadas y la cantidad que aporta cada clase a la medida de dispersión D^2:

n_i	26	23	16	18	9	7	1
$(n_i - np_i)^2/np_i$	9.60571	5.31571	0.20571	0.96571	1.95571	3.71571	12.35571

3. Decisión al 1 %. Finalmente comparamos el valor obtenido para $D^2 = 34.12$ con el cuantil 0.99 de la distribución $\chi^2_{6,0.99} = 16.812$. Como $D^2 > \chi^2_{6,0.99}$ concluimos que la duración de las pilas no se ajusta a la distribución exponencial de media 2 horas. $\square$

[6]Notar que al elegir el criterio de las clases "equiprobables" para construir la partición, garantizamos de entrada que no habrá partes sub o sobre dimensionadas y no vamos a encontrarnos con el problema de tener que unir dos clases porque quedaron muy "flacas".

Nota Bene. No siempre se puede dividir el rango de la variable en clases de igual probabilidad. Las variables discretas no lo permiten. En tal caso habrá que conformarse con algunas partes suficientemente "gorditas" como para que valga la condición $np_i \geq 10$ $\square$

8.5. Test de bondad de ajuste para hipótesis compuestas

La hipótesis nula afirma que

$$H_0 : F_X = F_{\theta_1,\dots,\theta_r},$$

donde $F_{\theta_1,\dots,\theta_r}$ es una distribución de probabilidades perteneciente a una familia paramétrica completamente determinada y los valores de los parámetros $\theta_1, \dots, \theta_r$ son desconocidos.

En este caso los r parámetros desconocidos se estiman usando el método de máxima verosimilitud. Los valores de las r estimaciones se "enchufan" en la distribución paramétrica como si fuesen los verdaderos valores de los parámetros y se aplica el test χ^2 desarrollado en la sección 8.2. Solo que ahora se perderá un grado de libertad por cada parámetro estimado. Si para construir la medida de dispersión D^2 se recurrió a una partición del rango de la variable X en k clases, la distribución de D^2 será aproximadamente una χ^2_{k-1-r}.

Ejemplo 8.4. (Continuación) La hipótesis H_0 afirma que la cantidad de impactos por segundo recibidos por la partícula de polen sigue una distribución de Poisson, pero no indica cuál es su media (el parámetro λ).

El estimador de máxima verosimilitud para la media de una distribución de Poisson es $\hat{\lambda}_{mv} = \bar{X}$. Usando los datos que aparecen en la tabla (62) obtenemos

$$\hat{\lambda}_{mv} = \frac{0(1364) + 1(1296) + 2(642) + 3(225) + 4(55) + 5(15) + 6(3)}{3600} = \frac{3568}{3600} = 0.9911 \approx 1.$$

Las clases C_i se pueden construir usando como criterio que $3600\mathbb{P}(X \in C_i) \geq 10$. Si suponemos que $X \sim$ Poisson(1), su función de probabilidades será $\mathbb{P}(X = n) = e^{-1}/n!$, $n = 0, 1, \dots$.

Usaremos como partición las siguientes clases: $C_1 = \{0\}$, $C_2 = \{1\}$, $C_3 = \{2\}$, $C_4 = \{3, 4, 5, \dots\}$, cuyas probabilidades son $p_1 = p_2 = 0.3678$, $p_3 = 0.1839$ y $p_4 = 0.0805$. Obtenemos que

$$\begin{aligned} D^2 &= \frac{(1364 - 3600p_1)^2}{3600p_1} + \frac{(1296 - 3600p_2)^2}{3600p_2} + \frac{(642 - 3600p_3)^2}{3600p_3} + \frac{(298 - 3600p_4)^2}{3600p_4} \\ &= \frac{1593.6064}{1324.08} + \frac{788.4864}{1324.08} + \frac{401.6016}{662.04} + \frac{67.24}{289.8} = 2.6376 \end{aligned}$$

Si se observa la Figura 12 se puede ver que un valor de 2.6376 para D^2 no es inusual para una distribución χ^2_2, lo que indica que la cantidad de impactos recibidos por la partícula de polen se puede considerar como una variable aleatoria con distribución Poisson. $\square$

Ejemplo 8.5. (Continuación) La hipótesis nula es de la forma $H_0 : X \sim \mathcal{N}(\mu, \sigma^2)$. Informalmente, se puede ver usando un histograma que los datos "obedecen" a una distribución normal.

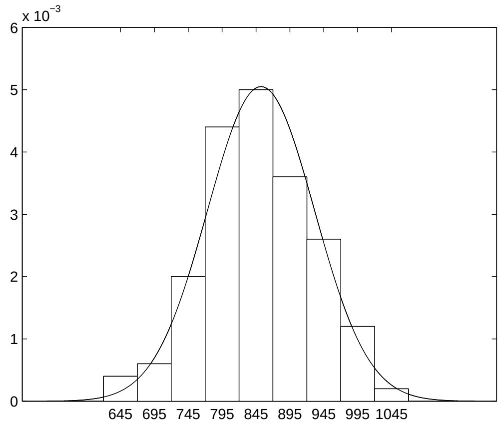

Figura 14: Histograma de los mediciones de Michelson y gráfico de la densidad de la distribución de media $\bar{X} = 852.4$ y varianza $S^2 = 79.0105$.

Usando los cuantiles de la distribución normal de media 852.4 y varianza 79.0105, construimos 9 clases equiprobables delimitadas por los valores: 756, 792, 818, 841, 863, 886, 913 y 949. Las frecuencias observadas en cada una de las 9 clases son, respectivamente, $9, 11, 15, 12, 11, 14, 7, 6$ y 15. Con esos datos, la medida de dispersión resulta $D^2 = 7.82 < \chi^2_{6,\, 0.90} \cdots$ □

9. Bibliografía consultada

Para redactar estas notas se consultaron los siguientes libros:

1. Bolfarine, H., Sandoval, M. C.: Introdução à Inferência Estatística. SBM, Rio de Janeiro. (2001).

2. Borovkov, A. A.: Estadística matemática. Mir, Moscú. (1984).

3. Cramer, H.: Métodos matemáticos de estadística. Aguilar, Madrid. (1970).

4. DeGroot, M. H.: Probability and Statistics. Addion-Wesley, Massachusetts. (1986).

5. Fisher, R. A.: Statistical methods for research workers. Hafner, New York (1954).

6. Hoel P. G.: Introducción a la estadística matemática. Ariel, Barcelona. (1980).

7. Lehmann, E. L.: Elements of Large-Sample Theory. Springer, New York. (1999)

8. Maronna R.: Probabilidad y Estadística Elementales para Estudiantes de Ciencias. Editorial Exacta, La Plata. (1995).

9. Meyer, P. L.: Introductory Probability and Statistical Applications. Addison-Wesley, Massachusetts. (1972).

10. Rice, J. A.: Mathematical Statistics and Data Analysis. Duxbury Press, Belmont. (1995).

11. Ross, S. M.: Introduction to Probability and Statistics for Engieneers and Scientists. Elsevier Academic Press, San Diego. (2004)

12. Walpole, R. E.: Probabilidad y estadística para ingenieros, 6a. ed., Prentice Hall, México. (1998)