Inferencia

Virgilio L. Foglia

E-mail address: ing@foglia.com.ar
URL: http://www.ifoglia.com

Contents

CHAPTER 1

ESTIMACION

1. Definición de Estimador

Considérese una v.a. X continua cuya función de densidad es $f(x/\theta)$. Esta densidad es conocida salvo su parámetro θ.

(Si X fuese discreta la densidad sería $p(x/\theta)$. Además si los parámetros desconocidos fuesen dos, escribiriamos $f(x/\theta_1, \theta_2)$, o $p(x/\theta_1, \theta_2)$)

A continuación se toma una muestra $X_1 X_2 \cdot\cdot X_n$ independiente(podría no serlo, pero en todo lo que sigue se considerarán muestras independientes).

Se tendrá entonces:

$$\boxed{f(x/\theta)}$$
$$X_1 X_2 \cdot\cdot X_n$$

En la práctica la muestra son n números reales $x_1 x_2 \cdot\cdot x_n$ totalmente conocidos. La intención es, con estos números, realizar alguna operación que nos proporcione una estimación del valor del parámetro θ. Supongamos que se propone como operación, cierta función g de la muestra, llamada la función estimadora, se tendría

$$\widehat{\theta} = g(x_1 x_2 \cdot\cdot x_n)$$

Aquí $\widehat{\theta}$ sería un número, que representa la estimación de θ, con la muestra $x_1 x_2 \cdot\cdot x_n$ que tenemos.

Este planteo tiene dos problemas: 1°) conocemos la estimación $\widehat{\theta}$, pero como θ es desconocido, no tenemos forma de saber si $\widehat{\theta}$ esta cerca o lejos del θ; 2°) si se toma otra muestra, los $x_1 x_2 \cdot\cdot x_n$ cambiarán, y también la estimación $\widehat{\theta}$. O sea, el valor estimado no es el mismo siempre, cambiará con la muestra.

Por eso para poder **evaluar** una función estimadora g, es necesario tener en cuenta que la muestra son variables aleatorias $X_1 X_2 \cdot\cdot X_n$ y la correspondiente estimación $\widehat{\Theta}$ también lo será, o sea

$$\widehat{\Theta} = g(X_1 X_2 \cdot\cdot X_n)$$

Aquí $\widehat{\Theta}$, el estimador, es una **variable aleatoria**, que depende de la muestra también aleatoria, y tendrá su correspondiente función de densidad $f_{\widehat{\Theta}}(\widehat{\theta}/\theta)$.

El problema general de estimación es el de averiguar la expresión de la función estimadora g, que hace que la variable aleatoria estimador $\widehat{\Theta}$ tenga buenas propiedades.

1.1. Propiedades deseables del estimador $\widehat{\Theta}$.

- Si probamos con varias muestras $X_1 X_2 \cdot\cdot X_n$ y con cada una de ellas calculamos los correspondientes valores del estimador $\widehat{\Theta}$, es deseable que estos valores esten alrededor de θ que es el valor que se quiere estimar, o sea pediremos que

$$E(\widehat{\Theta}) = \theta$$

Si un estimador cumple esta propiedad, se dice que es insesgado.

- Suponiendo que el estimador es insesgado, las estimaciones proporcionadas estarán alrededor de θ. Entonces será deseable que además de estar alrededor de θ, esten muy cerca de el, o sea se pedirá que

$$Var(\widehat{\Theta}) \text{ sea la menor posible}$$

EXAMPLE 1. *Considerese una v.a.* $X \sim N(\mu; 2)$ *y se quiere estimar* μ *con una muestra de tamaño* $n = 4$. *Se tiene entonces*

$$\begin{array}{c} \boxed{N(\mu;2)} \\ X_1X_2X_3X_4 \end{array}$$

Se proponen 3 funciones estimadoras:

$$\widehat{\mu}_1 = \frac{X_1 + X_2 + X_3 - X_4}{2} \qquad \widehat{\mu}_2 = \frac{X_1 + X_2 + X_3 + X_4}{4} \qquad \widehat{\mu}_3 = \frac{X_1 + 2X_2 + 3X_3 + 4X_4}{5}$$

Como $E(\widehat{\mu}_1) = \frac{\mu+\mu+\mu-\mu}{2} = \mu$ *resulta insesgada;* $E(\widehat{\mu}_2) = \frac{\mu+\mu+\mu+\mu}{4} = \mu$ *también es insesgada. Además* $E(\widehat{\mu}_3) = \frac{\mu+2\mu+3\mu+4\mu}{5} = 2\mu$, *entonces* $\widehat{\mu}_3$ *es sesgada, proporciona estimaciones que están alrededor de* 2μ, *o sea el doble del valor a estimar. Por eso la descartamos por el momento. Si a continuación se analizan las varianzas de las dos primeras, se tendrá:*

$$Var(\widehat{\mu}_1) = \frac{2^2 + 2^2 + 2^2 + 2^2}{2^2} = 4 \quad y \quad Var(\widehat{\mu}_2) = \frac{2^2 + 2^2 + 2^2 + 2^2}{4^2} = 1$$

De aquí resulta que el mejor estimador es $\widehat{\mu}_2$. *Si ahora se analizan las distribuciones de estos dos estimadores, teniendo es cuenta que son combinaciones lineales de v.a. normales se tiene*

$$\widehat{\mu}_1 \sim N(\mu; 2) \qquad \widehat{\mu}_2 \sim N(\mu; 1)$$

O sea $\widehat{\mu}_1$ *proporciona estimaciones alrededor de* μ *con desvío* 2, *mientras que el elegido* $\widehat{\mu}_2$ *proporciona estimaciones, tambien alrededor de* μ *pero en general mas cerca ya que el desvío es* 1.

1.2. Corrección del sesgo de un estimador. Si $\widehat{\Theta} = g(X_1X_2 \cdot \cdot X_n)$ es un estimador con $E(\widehat{\Theta}) = cte\ \theta$, este estimador es sesgado. Si la *cte* no depende de θ, se define el correspondiente estimador insesgado $\widehat{\Theta}_I$ mediante:

$$\widehat{\Theta}_I = \frac{\widehat{\Theta}}{cte}$$

ya que $E(\widehat{\Theta}_I) = E(\frac{\widehat{\Theta}}{cte}) = \frac{E(\widehat{\Theta})}{cte} = \frac{cte\ \theta}{cte} = \theta$ (insesgado).

Menos usual, si $\widehat{\Theta} = g(X_1X_2 \cdot \cdot X_n)$ es un estimador con $E(\widehat{\Theta}) = \theta + cte$, este estimador es sesgado. Si la *cte* no depende de θ, se define el correspondiente estimador insesgado $\widehat{\Theta}_I$ mediante:

$$\widehat{\Theta}_I = \widehat{\Theta} - cte$$

ya que $E(\widehat{\Theta}_I) = E(\widehat{\Theta} - cte) = E(\widehat{\Theta}) - cte = \theta + cte - cte = \theta$ (insesgado).

EXAMPLE 2. *En el ejemplo anterior se descartó $\widehat{\mu}_3$ por ser sesgado, ya que $E(\widehat{\mu}_3) = 2\mu$. Se define entonces*

$$\widehat{\mu}_{3I} = \frac{\widehat{\mu}_3}{2} = \frac{\frac{X_1+2X_2+3X_3+4X_4}{5}}{2} = \frac{X_1+2X_2+3X_3+4X_4}{10}$$

Ahora sí las estimaciones de este estimador estarán alrededor de μ. Para compararlo con los dos anteriores se evaluara su varianza

$$Var(\widehat{\mu}_{3I}) = \frac{2^2+2^2 2^2+3^2 2^2+4^2 2^2}{10^2} = 1.2$$

Luego en definitiva, el mejor de los tres estimadores insesgados sigue siendo $\widehat{\mu}_2$, le sigue $\widehat{\mu}_{3I}$, y el peor de todos es $\widehat{\mu}_1$.

2. Método de máxima verosimilitud

Se presentará ahora un método para obtener una función estimadora. Es el mas utilizado, ya que el estimador correspondiente tiene muy buenas propiedades.

Considérese una v.a. X discreta cuya función de densidad es $p(x/\theta)$ (si X es continua se utilizará $f(x/\theta)$), y una muestra $X_1X_2\cdot\cdot X_n$ independiente

$$\boxed{p(x/\theta)}$$
$$X_1X_2\cdot\cdot X_n$$

Como la muestra consiste en n v.a. independientes, cada una con densidad $p(x/\theta)$, y recordando lo visto en VVA, la **función de densidad de la muestra** será el producto de las n densidades marginales, o sea

$$p(x_1x_2\cdot\cdot x_n/\theta) = p(x_1/\theta)\ p(x_2/\theta)\cdot\cdot\ p(x_n/\theta)$$

En la teoría de probabilidad, en VVA, el parámetro θ es conocido, y los $x_1x_2\cdot\cdot x_n$ representan todas las muestras posibles. Además usando este θ conocido y cualquier muestra $x_1x_2\cdot\cdot x_n$, la $p(x_1x_2\cdot\cdot x_n/\theta)$ proporciona la probabilidad de que cuando se tome una muestra, esta asuma los valores $x_1x_2\cdot\cdot x_n$.

En el problema de estimación, ocurre al revés, ya que la muestra la conocemos, (es la muestra que tenemos, es conocida), y lo desconocido es θ.

Entonces cuando a la $p(x_1x_2\cdot\cdot x_n/\theta)$ la evaluamos para la muestra que tenemos, dándole a θ cualquier valor que elegimos, el número obtenido proporcionaría, en el caso de tomar otra muestra, cuanto vale la probabilidad de que salga la muestra que realmente tenemos.

En lugar de usar el término probabilidad, a este concepto se lo designa **verosimilitud de la muestra**.

Como la muestra es fija, se usa la notación $L(\theta/x_1x_2\cdot\cdot x_n)$ (del inglés Likelihood), poniendo después de la barra lo conocido, y antes el θ desconocido. Sin embargo su expresión funcional es la misma

$$L(\theta/x_1x_2\cdot\cdot x_n) = p(x_1/\theta)\ p(x_2/\theta)\cdot\cdot\ p(x_n/\theta)$$

Ahora $L(\theta/x_1x_2\cdot\cdot x_n)$ se la piensa como una función solo de θ, ya que la muestra es conocida. Y entonces los valores de θ que hacen que $L(\theta/x_1x_2\cdot\cdot x_n)$ tome un valor alto, se interpretan como que son θ que hacen a la muestra que tenemos muy verosimil en relación a θ, y al revés cuando toma un valor bajo.

En definitiva el método de Maxima verosimilitud propone como estimador del parámetro, al valor de θ que hace más verosimil a la muestra que tenemos. O sea:

$$\widehat{\theta}_{MV} \text{ es el valor de } \theta \text{ que hace máxima a la } L(\theta/x_1x_2\cdot\cdot x_n)$$

Para obtener el estimador, en general habrá que derivar la verosimilitud respecto de θ, y buscar el máximo.

En el caso que la densidad de la población sea continua ($f(x/\theta)$), los resultados obtenido son similares, solo habría que cambiar $p(x/\theta)$ por $f(x/\theta)$, y modificar algunos detalles de la explicación anterior.

REMARK 1. *Obtenido el estimador, este será función solo de la muestra, o sea* $\widehat{\theta}_{MV} = g(x_1x_2\cdot\cdot x_n)$. *Si aquí se reemplazan las* x_i *por las correspondientes v.a. de la muestra, las* X_i, *se obtiene la expresión del estimador como variable aleatoria*

$$\widehat{\Theta}_{MV} = g(X_1X_2\cdot\cdot X_n)$$

Esto es importante para estudiar las propiedades del estimador obtenido. Por ejemplo, si se quiere averiguar si es insesgado (propiedad que ***no garantiza*** *el estimador de MV), habrá que evaluar si* $E(\widehat{\Theta}_{MV}) = \theta$. *Si esto no ocurre, se tratará de corregirlo si esto es posible, y obtener el correspondiente estimador insesgado.*

REMARK 2. *Si la densidad poblacional tiene mas parámetros, por ejemplo dos como* $p(x/\theta_1,\theta_2)$ *o* $f(x/\theta_1,\theta_2)$; *la verosimilitud entonces será función de dos parámetros como* $L(\theta_1,\theta_2/x_1x_2\cdot\cdot x_n)$, *y al maximizar habrá que utilizar derivadas parciales.*

EXAMPLE 3. *Sea la densidad(continua)* $f(x/\theta) = (\theta+1)x^\theta$ *para* $0 < x < 1$ *y* $\theta > -1$. *Se pide estimar* θ *por MV. Se tendrá entonces*

$$\boxed{f(x/\theta) = (\theta+1)x^\theta \text{ para } 0 < x < 1 \text{ y } \theta > -1}$$
$$X_1X_2\cdot\cdot X_n$$

Se tiene entonces

$$L(\theta/x_1x_2\cdot\cdot x_n) = (\theta+1)x_1^\theta(\theta+1)x_2^\theta\cdot\cdot(\theta+1)x_n^\theta = (\theta+1)^n(x_1x_2\cdot\cdot x_n)^\theta$$

Para evitar errores y facilitar el calculo, se darán dos recomendaciones:

(1) Como las x_i son valores conocidos, llamamos momentaneamente $a = x_1x_2\cdot\cdot x_n$ entonces queda:
$$L(\theta/x_1x_2\cdot\cdot x_n) = (\theta+1)^n a^\theta$$
(2) Habría que derivar esto respecto de θ. Pero antes conviene calcular el $ln(L(\theta/x_1x_2\cdot\cdot x_n))$ y después derivar. El estimador obtenido no cambia, ya que el logaritmo es una función creciente, y entonces obtener el máximo en $L(\theta/x_1x_2\cdot\cdot x_n)$, es equivalente a obtener el máximo de $ln(L(\theta/x_1x_2\cdot\cdot x_n))$.
$$ln(L(\theta/x_1x_2\cdot\cdot x_n)) = n\ln(\theta+1) + \theta\ln(a)$$

Ahora sí derivando, e igualanco a cero

$$\frac{d\ln(L(\theta/x_1x_2\cdot\cdot x_n))}{d\theta} = \frac{n}{\theta+1} + \ln(a) = 0$$

De aquí surge(habría que verificar que es realmente un máximo)

$$\widehat{\theta}_{MV} = -1 - \frac{n}{\ln(a)} = -1 - \frac{n}{\ln(x_1x_2\cdot\cdot x_n)}$$

Finalmente la expresión del estimador como variable aleatoria sería

$$\widehat{\Theta}_{MV} = -1 - \frac{n}{\ln(X_1 X_2 \cdot \cdot X_n)}$$

Este es el estimador de θ que surge del método de máxima verosimilitud. Es un estimado muy bueno, pero no está garantizado que sea insesgado. Para analizar esto habria que evaluar si se cumple

$$E(\widehat{\Theta}_{MV}) = \theta$$

(y por supuesto corregir el estimador para hacerlo insesgado, si es posible)

2.1. Método de los momentos. Este método para encontrar la función estimadora de un estimador, suele ser muy simple de aplicar, pero el estimador obtenido no tiene tan buenas propiedades como el de MV. Se lo presentará para una población con dos parámetros θ_1 y θ_2. Consideremos el caso continuo

$$\begin{array}{c} \boxed{f(x/\theta_1;\theta_2)} \\ X_1 X_2 \cdot \cdot X_n \end{array}$$

Cuando se estudiaron los conceptos de media y varianza de una v.a. se definió:

- Si se tiene la densidad poblacional $f(x/\theta_1;\theta_2)$, la media y varianza "**poblacionales**" serán:

$$\mu = \int_{dom} x f(x/\theta_1;\theta_2)dx = g(\theta_1;\theta_2) \qquad \sigma^2 = \int_{dom} (x-\mu)^2 f(x/\theta_1;\theta_2)dx = h(\theta_1;\theta_2)$$

  Notar que como θ_1 y θ_2 son desconocidos, la μ y σ^2 "poblacional" dependerán de θ_1 y θ_2.
- Si se tiene una muestra $X_1 X_2 \cdot\cdot X_n$, la media y varianza "**muestral**" serán:

$$\overline{X} = \frac{X_1 + X_2 + \cdot \cdot X_n}{n} \qquad S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}$$

Además se comentó que para n "grande", estas medias y varianzas "muestrales" tienden (el algún sentido), a las correspondientes medias y varianzas "poblacionales", o sea

$$\overline{X} \longrightarrow \mu = g(\theta_1;\theta_2) \qquad S^2 \longrightarrow \sigma^2 = h(\theta_1;\theta_2)$$

Este método se basa en suponer que el n es suficientemente grande como para postular la igualdad o sea que

$$\begin{cases} \overline{X} = g(\theta_1;\theta_2) \\ S^2 = h(\theta_1;\theta_2) \end{cases}$$

Finalmente de este sistema se despejan θ_1 y θ_2 en función de $\overline{X}$ y S^2, y resultan los estimadores buscados. Se presentará un ejemplo.

EXAMPLE 4. *Considérese una población $G(\alpha,\beta)$ y una muestra*

$$\begin{array}{c} \boxed{G(\alpha,\beta)} \\ X_1 X_2 \cdot \cdot X_n \end{array}$$

Aquí no hace falta integrar ya que para la gamma la media y variaza "poblacionales" son $\mu = \frac{\alpha}{\beta}$ *y* $\sigma^2 = \frac{\alpha}{\beta^2}$ *y para la muestra la media y varianza "muestrales" son* $\overline{X}$ *y* S^2*. Igualando queda el sistema*

$$\begin{cases} \overline{X} = \frac{\alpha}{\beta} \\ S^2 = \frac{\alpha}{\beta^2} \end{cases}$$

Resolviendo resultan, como v.a. los estimadores

$$\widehat{\alpha} = \frac{\overline{X}^2}{S^2} \quad y \quad \widehat{\beta} = \frac{\overline{X}}{S^2}$$

REMARK 3. *Si la población tiene un solo parámetro a estimar, como por ejemplo la* $P_o(\lambda)$*, no hace falta un sistema con dos ecuaciones, basta con una. En este caso la media "poblacional" es* $\mu = \lambda$*, y la "muestral" es* $\overline{X}$*. Luego igualando* $\overline{X} = \lambda$*, resulta el estimador, como v.a.* $\widehat{\lambda} = \overline{X}$*.*

3. Estadistico muestral y Suficiencia

Consideremos una población, por ejemplo continua, con un parámetro θ desconocido, y una muestra

$$\begin{array}{c} \boxed{f(x/\theta)} \\ X_1X_2 \cdot \cdot X_n \end{array} \tag{3.1}$$

En inferencia estadística, la muestra es conocida, y con ella se pretende averiguar información respecto del θ de la población. Por ejemplo, con la muestra: **1)** estimar θ, o **2)** hallar un intervalo de confianza para θ, o **3)** averiguar si $\theta \leq 10$(estos dos últimos puntos,se verán mas adelante).

Para evaluar todo esto se requiere pensar la muestra como n variables aleatorias independientes, con su correspondiende función de densidad

$$f(x_1x_2 \cdot \cdot x_n/\theta) = f(x_1/\theta)\ f(x_2/\theta) \cdot \cdot\ f(x_n/\theta)$$

Claro que esto suele ser muy complicado, ya que hay que utilizar la teoría de VVA, con integrales múltiples(son n!!).

Sería mucho mas simple definir una nueva variable aleatoria, H, que es función de la muestra, que se llamará **estadístico muestral**

$$H = g(X_1X_2 \cdot \cdot X_n)$$

Esto es un cambio de variable, y si se averigua la densidad de H, esta será $f(h/\theta)$, que es una densidad en una variable, y que seguramente dependerá de θ(esto no siempre ocurre), ya que la muestra $X_1X_2 \cdot\cdot X_n$ tiene una densidad $f(x_1x_2 \cdot\cdot x_n/\theta)$ que depende de θ.

Entonces, para dar respuesta a las tres inquietudes de la inferencia estadística sugeridas al iniciar este tema, la idea sería: en lugar de considerar toda la muestra, que tiene la densidad $f(x_1x_2 \cdot \cdot x_n/\theta)$, como en (3.1); considerar que la muestra es solo H con densidad $f(h/\theta)$ o sea

$$\begin{array}{c} \boxed{f(h/\theta)} \\ H \end{array} \tag{3.2}$$

Esta forma de proceder es habitual en estadística, ya que es considerablemente mas simple.

Pero claro, hasta ahora no se dijo nada respecto de quien es la función $g(X_1X_2\cdot\cdot X_n)$ que define al estadístico muestral; puede ser cualquier función?

Para evaluar este tema se requeriría el concepto de función de densidad condicional(que no fué visto). Así que para evitar esto se definirá, bastante ambiguamente, un concepto sobre información.

En (3.1), diremos que la muestra tiene cierta "información" respecto de θ. Parece razonable, ya que si no la tuviese, no sería posible, solo con la muestra estimar θ. Definimos entonces a la información que contiene la muestra respecto de θ a $Inf_\theta(X_1X_2\cdot\cdot X_n)$.

Por ejemplo si tomamos una muestra de tamaño 1, la información que contiene respecto de θ sería $Inf_\theta(X_1)$, en cambio si tomamos una muestra de tamaño 3, la información sería $Inf_\theta(X_1X_2X_3)$. En este caso es razonable pensar que $Inf_\theta(X_1)\leq Inf_\theta(X_1X_2X_3)$. O sea que cuanto mayor es la muestra, mayor información tendrá respecto de θ. Aunque esto no siempre ocurre, lo habitual es que muestras mas grandes, tengan mayor información respecto de θ.

Si utilizamos un estadistico muestral como en (3.2), H en general tendrá también cierta información respecto de θ. Definimos la información que contiene H respecto de θ, a $Inf_\theta(H)$.

Si la función g con la cual definimos el estadístico es una función cualquiera, lo habitual es que $Inf_\theta(H)\leq Inf_\theta(X_1X_2\cdot\cdot X_n)$. Esto es así ya que la información que tenemos inicialmente es la de toda la muestra, $Inf_\theta(X_1X_2\cdot\cdot X_n)$. Y si por comodidad resolvemos utilizar $H=g(X_1X_2\cdot\cdot X_n)$, que es una sola variable, la información en general será menor.

DEFINITION 1. *Un estadístico muestral es* ***suficiente*** *respecto de un parámetro si*

$$Inf_\theta(H)=Inf_\theta(X_1X_2\cdot\cdot X_n)$$

O sea, si logramos encontrar una función g tal que $H=g(X_1X_2\cdot\cdot X_n)$ sea suficiente respecto de θ, esta resuelto el problema. Descartamos utilizar toda la muestra, y a partir de ahora usaremos solo H para "fabricar" estimaciones de θ, intervalos de confianza sobre θ, o pruebas de hipótesis sobre θ, con la seguridad que lo obtenido será tan bueno como si hubiese sido obtenido con toda la muestra, pero mucho mas simple operativamente.

Solo falta un método para, dada la densidad de una muestra, averiguar si existe un estadístico suficiente para θ. Esto lo proporciona el siguiente teorema.

THEOREM 1. ***Teorema de factorización de Neyman****: Dada la densidad de la muestra $f(x_1x_2\cdot\cdot x_n/\theta)$, entonces $h=g(x_1x_2\cdot\cdot x_n)$ es un estadístico suficiente respecto de θ, si la densidad se puede factorizar así*

$$f(x_1x_2\cdot\cdot x_n/\theta)=u(h,\theta)\ v(x_1x_2\cdot\cdot x_n)$$

O sea, un factor que es función solo a h y θ, y el otro, solo función de la muestra(aquí no puede aparecer θ).

4. Estimación de λ de una población Poisson

EXAMPLE 5. ***Estadístico suficiente, y estimación de λ, para una muestra de una*** $P_o(\lambda)$. *Sea*

$$\boxed{P_o(\lambda)} \qquad (4.1)$$
$$X_1X_2\cdot\cdot X_n$$

La densidad de la Poisson es $p(x/\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}$. Luego la densidad de la muestra

$$\begin{aligned} p(x_1x_2\cdot\cdot x_n/\lambda) &= e^{-\lambda}\frac{\lambda^{x_1}}{x_1!}e^{-\lambda}\frac{\lambda^{x_2}}{x_2!}\cdot\cdot e^{-\lambda}\frac{\lambda^{x_n}}{x_n!} \\ &= e^{-n\lambda}\frac{\lambda^{x_1+x_2+\cdot\cdot+x_n}}{x_1!x_2!\cdot\cdot x_n!} \end{aligned}$$

Inspeccionando, si designamos $h = x_1 + x_2 + \cdot\cdot + x_n$ queda

$$p(x_1x_2\cdot\cdot x_n/\lambda) = \left[e^{-n\lambda}\lambda^h\right]\left[\frac{1}{x_1!x_2!\cdot\cdot x_n!}\right] = u(h,\lambda)\ v(x_1x_2\cdot\cdot x_n)$$

Luego $H = X_1 + X_2 + \cdot\cdot X_n$ es un estadístico suficiente para λ. Y utilizando

Suma de Poisson independientes

(4.2) $$\boxed{\begin{array}{l} X \sim P_o(\lambda_x) \\ Y \sim P_o(\lambda_y) \end{array} Ind \rightrightarrows \boxed{Q = X + Y} \longrightarrow Q \sim P_o(\lambda_x + \lambda_y)}$$

Resulta $H = X_1 + X_2 + \cdot\cdot X_n \sim P_o(n\lambda)$ la distribución del estadístico suficiente para λ.

Luego para hacer inferencias sobre λ, podemos considerar que la población es $P_o(n\lambda)$, y la muestra solo H

(4.3) $$\begin{array}{c} \boxed{P_o(n\lambda)} \\ H \end{array}$$

Por ejemplo para estimar λ por MV: la densidad de H es $p(h/\lambda) = e^{-n\lambda}\frac{(n\lambda)^h}{h!}$, la verosimilitud de H es $L(\lambda/h) = e^{-n\lambda}\frac{(n\lambda)^h}{h!}$, luego $\ln L(\lambda/h) = -n\lambda + h\ln(n\lambda) - \ln(h!)$ y derivando

$$\frac{d\ln L(\lambda/h)}{d\lambda} = -n + h\frac{n}{n\lambda} - 0 = 0 \quad \text{luego resulta } \widehat{\lambda} = \frac{h}{n}$$

En definitiva para una muestra de una Poisson(4.1), se puede trabajar solo con el estadístico suficiente como en (4.3) y obtener el estimador

$$\widehat{\lambda} = \frac{H}{n} = \frac{X_1 + X_2 + \cdot\cdot X_n}{n}$$

Que es insesgado, ya que $E(\widehat{\lambda}) = \frac{E(X_1)+E(X_2)+\cdot\cdot E(X_n)}{n} = \frac{\lambda+\lambda+\cdot\cdot\lambda}{n} = \frac{n\lambda}{n} = \lambda$; y $Var(\widehat{\lambda}) = \frac{Var(X_1)+Var(X_2)+\cdot\cdot Var(X_n)}{n^2} = \frac{\lambda+\lambda+\cdot\cdot+\lambda}{n^2} = \frac{n\lambda}{n^2} = \frac{\lambda}{n}$.

5. Estimación de p de una población Bernouilli

EXAMPLE 6. ***Estadístico suficiente y estimación de*** p***, para una muestra de una*** *Bernouilli*(p). *Recordando, un experimento de Bernouilli tiene solo dos resultados exito y* $\overline{exito}$. *La* $P(exito) = p$, *y* $P(\overline{exito}) = 1 - p$. *Para expresar esto como v.a. definimos* $X = 1$ *si el resultado es exito, y* $X = 0$ *si el resultado es* $\overline{exito}$. *Entonces la densidad de* X *será:* $p(x/p) = p^x(1-p)^{1-x}$ *para* $x = 1, 0$. *Verifiquemos:* $p(0/p) = p^0(1-p)^{1-0} = 1 - p$; $p(1/p) = p^1(1-p)^{1-1} = p$. *Notemos de paso que la densidad de* $X \sim Bernouilli(p)$ *coincide con la* $B_i(1;p)$, *en efecto* $p_{bi}(x/1,p) = \binom{1}{x}p^x(1-p)^{1-x} = p^x(1-p)^{1-x}$ *ya que para* $x = 0, 1$ *el combinatorio* *vale 1. Con esta introducción supongamos que tenemos una muestra de tamaño n de una Bernouilli$(p) \approx B_i(1;p)$*

$$\boxed{B_i(1;p)}$$
$$X_1 X_2 \cdot \cdot X_n$$

La densidad de la muestra es

$$\begin{aligned} p(x_1 x_2 \cdot \cdot x_n/p) &= p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdot \cdot p^{x_n}(1-p)^{1-x_n} \\ &= p^{x_1+x_2+ \cdot\cdot +x_n}(1-p)^{n-x_1+x_2+ \cdot\cdot +x_n} \end{aligned}$$

Inspeccionando, si designamos $h = x_1 + x_2 + \cdot\cdot + x_n$ queda

$$p(x_1 x_2 \cdot \cdot x_n/p) = p^h(1-p)^{n-h} = u(h,p)\ 1 \ (\textit{aquí } v(x_1 x_2 \cdot \cdot x_n) = 1)$$

Luego $H = X_1 + X_2 + \cdot \cdot X_n$ es un estadístico suficiente para p. Y utilizando

Suma de Binomiales independientes de igual p

$$\boxed{\begin{matrix} X \sim B_i(n_x;p) \\ Y \sim B_i(n_y;p) \end{matrix}\ Ind \rightrightarrows \boxed{Q = X + Y} \longrightarrow Q \sim B_i(n_x + n_y;p)} \tag{5.1}$$

Resulta $H = X_1 + X_2 + \cdot \cdot X_n \sim B_i(n;p)$ la distribución del estadístico suficiente para p. (esto lo conocemos: el número de éxitos en n repeticiones es $B_i(n;p)$)

REMARK 4. *Que $H = X_1 + X_2 + \cdot\cdot X_n$ es un estadístico suficiente para p, surgió aquí, del teorema de factorización. Sin embargo pudo haberse intuido. Supongamos que interesa el p (de defectuoso) de una máquina. Para evaluar esto se fabrican $n = 10$ artículos, obteniendo la muestra $X_1 X_2 \cdot \cdot X_{10} = 0,1,0,0,1,0,0,0,1,0$, pero se esconde esta muestra. A continuación se nos pregunta que preferimos (si nuestra intención es averiguar cuanto valdrá p), **Opción 1**: que nos den toda la muestra; **Opción 2**: que nos digan cuanto vale $H = X_1 + X_2 + \cdot \cdot X_n = 3$? Se comprende que tanto la muestra, como conocer solo que $H = 3$ tienen la misma información respecto de p? Porque, si tener toda la muestra fuese superior, esto querria decir que la posición concreta en que aparecen los 3 defectuosos en la muestra aporta algo sobre p. No parece razonable esto. Por eso H es suficiente.*

Luego para hacer inferencias sobre p, podemos considerar que la población es $B_i(n;p)$, y la muestra solo H (el número de defectuosos)

$$\boxed{B_i(n;p)}$$
$$H$$

Si queremos estimar p por MV: la densidad de H es $p_{B_i}(h/n;p) = \binom{n}{h} p^h (1-p)^{n-h}$, la verosimilitud de H es $L(p/h) = \binom{n}{h} p^h (1-p)^{n-h}$, luego $\ln L(p/h) = \ln \binom{n}{h} + h \ln p + (n-h)\ln(1-p)$ derivando

$$\frac{d \ln L(p/h)}{dp} = 0 + h\frac{1}{p} - (n-h)\frac{1}{1-p} = 0 \quad \text{luego resulta } \widehat{p} = \frac{h}{n}$$

En definitiva para una muestra de una *Bernouilli*$(p) \approxeq B_i(1;p)$, se puede trabajar solo con el estadístico suficiente y obtener el estimador

$$\widehat{p} = \frac{H}{n} = \frac{X_1 + X_2 + \cdot \cdot X_n}{n}$$

Notar que es razonable este estimador, o sea $\widehat{p} = \frac{\textit{número de exitos en la muestra}}{n}$. Además como la media de la binomial es np y la varianza $np(1-p)$ resulta:

$$E(\widehat{p}) = \frac{np}{n} = p \text{ (insesgado) } y \ \ Var(\widehat{p}) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

REMARK 5. *Cuando una población tiene 2 parámetros desconocidos, por ejemplo $f(x/\theta_1\theta_2)$, usualmente la información de la muestra respecto de $\theta_1\theta_2$,no se puede resumir en un solo número $H_1 = g_1(X_1, X_2 \cdot\cdot X_n)$ sinó que se requieren más, por ejemplo (H_1, H_2) con $H_1 = g_1(X_1, X_2 \cdot\cdot X_n)$ y $H_2 = g_2(X_1, X_2 \cdot\cdot X_n)$. En este caso el estadístico suficiente sería la pareja (H_1, H_2). Esto ocurre con la gamma y sus dos parámetros.*

6. Estimación de β en una población Gamma

EXERCISE 1. *Estadístico suficiente para una muestra de una $G(\alpha;\beta)$. Sea*

$$\begin{array}{c}\boxed{G(\alpha;\beta)}\\ X_1X_2\cdot\cdot X_n\end{array}$$

Como la densidad de la gamma es $f(x/\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$ la densidad de la muestra es

$$\begin{aligned} f(x_1x_2\cdot\cdot x_n/\alpha,\beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)}x_1^{\alpha-1}e^{-\beta x_1}\cdot\cdot\frac{\beta^\alpha}{\Gamma(\alpha)}x_n^{\alpha-1}e^{-\beta x_n} \\ &= \frac{\beta^{n\alpha}}{\Gamma^n(\alpha)}(x_1x_2\cdot\cdot x_n)^{\alpha-1}e^{-\beta(x_1+x_2+\ \cdot\cdot\ +x_n)} \end{aligned}$$

Entonces llamando $h_1 = x_1x_2\cdot\cdot x_n$ y $h_2 = x_1 + x_2 + \cdot\cdot + x_n$ resulta la factorización

$$f(x_1x_2\cdot\cdot x_n/\alpha,\beta) = \frac{\beta^{n\alpha}}{\Gamma^n(\alpha)}(h_1)^{\alpha-1}e^{-\beta(h_2)}\ 1 = u(h_1,h_2,\alpha,\beta)\ 1$$

Luego la pareja de estadísticos $H_1 = X_1X_2\cdot\cdot X_n$ y $H_2 = X_1 + X_2 + \cdot\cdot X_n$ contienen la misma información que toda la muestra, respecto de los parámetros α,β.

Como caso particular se presentará otro ejemplo relacionado.

EXAMPLE 7. *Estadístico suficiente y estimación de β, para una muestra de una $G(1;\beta)$.Sea*

$$\begin{array}{c}\boxed{G(1;\beta)}\\ X_1X_2\cdot\cdot X_n\end{array}$$

Aquí habría que repetir lo del ejemplo anterior para la $G(1;\beta)$ que es mas simple, y resulta el estadístico suficiente para β

$$H = X_1 + X_2 + \cdot\cdot X_n$$

Y utilizando el resultado

Suma de Gammas independientes de igual β

$$\boxed{\begin{array}{c} X \sim G(\alpha_x;\beta) \\ Y \sim G(\alpha_y;\beta)\end{array}\ Ind \rightrightarrows \boxed{Q = X + Y} \longrightarrow Q \sim G(\alpha_x + \alpha_y;\beta)} \tag{6.1}$$

resulta

$$H = X_1 + X_2 + \cdot\cdot X_n \sim G(n,\beta)$$

Ahora si, si se quiere un estimador de β, en lugar de trabajar con toda la muestra, podemos hacerlo solo con H. Como la densidad de H es $f_\gamma(h/n,\beta) = \frac{\beta^n}{\Gamma(n)}h^{n-1}e^{-\beta h}$, la verosimilitud será $L(\beta/h) = \frac{\beta^n}{\Gamma(n)}h^{n-1}e^{-\beta h} = \frac{h^{n-1}}{\Gamma(n)}\beta^n e^{-\beta h}$ y su logaritmo

$$\begin{aligned}
\ln L(\beta/h) &= \ln(\frac{h^{n-1}}{\Gamma(n)}) + n\ln\beta - \beta h \\
\frac{d\ln L(\beta/h)}{d\beta} &= 0 + \frac{n}{\beta} - h = 0
\end{aligned}$$

De aquí surge que

$$\widehat{\beta} = \frac{n}{H} = \frac{n}{X_1 + X_2 + \cdot\cdot X_n} = \frac{1}{\overline{X}}$$

6.1. Función de un estadístico suficiente. Sea $f(x/\theta)$ una población, $X_1X_2\cdot\cdot X_n$ una muestra de esta población, y $H = g(X_1X_2\cdot\cdot X_n)$ un estadístico suficiente para θ.

Si nuestro interés es averiguar información respecto del θ desconocido, y nos preguntan que preferimos conocer: toda la muestra o solo el valor de H? La respuesta sería "indistinto", ya que los dos tienen la misma información respecto de θ(aunque como vimos es mas práctica la segunda opción, pues trabajamos con solo una v.a.).

Pero que tal si la opción es, que preferimos conocer: solo H, o solo el valor de $T = 3H$? Aquí la respuesta tambiés es "indistinto". En efecto, si nos dan T, lo dividimos por 3, y tenemos H. O sea H y T tienen la misma información respecto de θ. O sea también T es un estadístico suficiente. De aquí el resultado:

LEMMA 1. *Si H es suficiente, y $T = g(H)$ con g inyectiva $\Longrightarrow T$ será suficiente también*

Este resultado se aplicará varias veces en lo que sigue. Por ejemplo es usual que, respecto de un parámetro θ, resulte $H = X_1 + X_2 + \cdot\cdot X_n$ un estadístico suficiente. Pero entonces $\overline{X} = \frac{H}{n}$ también será suficiente(ya que conociendo $\overline{X}$ podemos averiguar H).

Sin embargo $T = \overline{X}^2$ en general no será suficiente(conocido T, no queda unívocamente definido el valor de $\overline{X}$, ya que puede ser $\overline{X}$ o $-\overline{X}$, pues la función cuadrática no es inyectiva).

A continuación se presentará un listado de estadísticos suficientes para distintas distribuciones y parámetros.

6.2. Listado de estadísticos suficientes. A continuación se presenta un listado de estadísticos suficientes(todos surgen del teorema de factorización).

Cuando un parámetro es conocido, se le pondrá un circulo como subíndice("$\circ$"). Además en el caso de dos parámetros desconocidos, aparece como estadístico, una pareja suficiente $(H_1; H_2)$.

- $N(\mu;\sigma_\circ^2)$ $H = X_1 + X_2 + \cdot\cdot X_n$
- $N(\mu_\circ;\sigma^2)$ $H = (X_1 - \mu_\circ)^2 + \cdot\cdot + (X_n - \mu_\circ)^2$
- $N(\mu;\sigma^2)$ $H_1 = X_1 + X_2 + \cdot\cdot X_n$ y $H_2 = (X_1 - \overline{X})^2 + \cdot\cdot + (X_n - \overline{X})^2$
- $P_o(\lambda)$ $H = X_1 + X_2 + \cdot\cdot X_n$
- $G(\alpha;\beta)$ $H_1 = X_1X_2\cdot\cdot X_n$ y $H_2 = X_1 + X_2 + \cdot\cdot X_n$
- $B_i(1;p)$ $H = X_1 + X_2 + \cdot\cdot X_n$

7. Estimación de parámetros en una población normal

7.1. Estimación de μ cuando σ_o es conocido.

$$\boxed{N(\mu;\sigma_\circ)}$$
$$X_1 X_2 \cdot \cdot X_n$$

Del listado, como $H = X_1 + X_2 + \cdot \cdot X_n$ es suficiente para μ, y $H \sim N(n\mu; \sqrt{n}\sigma_\circ)$, podemos utilizar la verosimilitud de H

$$L(\mu/h) = \frac{1}{\sqrt{2\pi}\sqrt{n}\sigma_\circ} e^{-\frac{(h-n\mu)^2}{2n\sigma_\circ^2}}$$

Hallando $\ln L(\mu/h)$ y derivando surge

$\widehat{\mu} = \dfrac{H}{n} = \overline{X}$ (este estimador es suficiente pues es función inyectiva de H suficiente)

y además

$$\widehat{\mu} = \overline{X} \sim N(\mu; \frac{\sigma_\circ}{\sqrt{n}}) \quad \text{o sea insesgado}$$

7.2. Estimación de σ^2 cuando μ_o es conocido.

(consideramos como parámetro $v = \sigma^2$ para evitar radicales)

$$\boxed{N(\mu_\circ;\sigma^2)}$$
$$X_1 X_2 \cdot \cdot X_n$$

Esta vez, en lugar de utilizar el estadístico suficiente para $v = \sigma^2$, (del listado), usaremos toda la muestra. Como la densidad de cada X es

$$f(x/v) = \frac{1}{\sqrt{2\pi}\sqrt{v}} e^{-\frac{(x-\mu_\circ)^2}{2v}}$$

La verosuimilitud de la muestra será

$$L(v/x_1 x_2 \cdot \cdot x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} v^{\frac{n}{2}}} e^{-\frac{(x_1-\mu_\circ)^2 + \cdot\cdot + (x_n-\mu_\circ)^2}{2v}}$$

Luego

$$\ln L(v/x_1 x_2 \cdot \cdot x_n) = -\ln(2\pi)^{\frac{n}{2}} - \frac{n}{2}\ln v - \frac{(x_1-\mu_\circ)^2 + \cdot\cdot + (x_n-\mu_\circ)^2}{2v}$$
$$\frac{d \ln L(v/x_1 x_2 \cdot \cdot x_n)}{dv} = 0 - \frac{n}{2v} + \frac{(x_1-\mu_\circ)^2 + \cdot\cdot + (x_n-\mu_\circ)^2}{2v^2} = 0$$

despejando surge

$$\widehat{v} = S^2 = \frac{H}{n} = \frac{(X_1-\mu_\circ)^2 + \cdot\cdot + (X_n-\mu_\circ)^2}{n}$$

Este es el estimador de varianza obtenido por MV, que al ser función inyectiva del suficiente H, es suficiente. Pero para mas adelante , necesitamos su distribución.

Partiendo de

$$S^2 = \frac{(X_1-\mu_\circ)^2 + \cdot\cdot + (X_n-\mu_\circ)^2}{n}$$

resulta

$$nS^2 = (X_1-\mu_\circ)^2 + \cdot\cdot + (X_n-\mu_\circ)^2$$

y dividiendo por σ^2

$$\frac{nS^2}{\sigma^2} = \left(\frac{X_1-\mu_\circ}{\sigma}\right)^2 + \cdot\cdot + \left(\frac{X_n-\mu_\circ}{\sigma}\right)^2$$

Como cada paréntesis es la estandarización del respectivo X_i que es normal, reempazando cada $\frac{X_i-\mu_\circ}{\sigma}$ por $Z_i \sim N(0;1)$ resulta

$$\frac{nS^2}{\sigma^2} = Z_1^2 + \cdot\cdot + Z_n^2$$

Además como los X_i son independientes, también lo serán los Z_i.

Pero necesitamos un resultado

Cuadrado de una N(0;1)

$$\boxed{Z \sim N(0;1) \longrightarrow \boxed{Y = Z^2} \longrightarrow Y \sim G(\tfrac{1}{2};\tfrac{1}{2}) = \chi_1^2}$$

O sea, el cuadrado de una v.a. $N(0;1)$ se distribuye como una $G(\frac{1}{2};\frac{1}{2})$, que a partir de ahora se llamará chi-cuadrado con 1 grado de libertad(esto se verá a continuación).

Utilizando este resultado, cada $Z_i^2 \sim G(\frac{1}{2};\frac{1}{2})$. Entonces del lado derecho tenemos la suma de n v.a. independientes $G(\frac{1}{2};\frac{1}{2})$. Con el resultado(6.1) sobre suma de gammas independientes de igual beta, resulta

$$Z_1^2 + \cdot\cdot + Z_n^2 \sim G(\frac{n}{2};\frac{1}{2}) = \chi_n^2$$

Esta distribución, que es una gamma particular que tiene $\alpha = \frac{n}{2}$ y $\beta = \frac{1}{2}$ se la designa χ_n^2 con "n" grados de libertad. Lo de grados de libertad tiene que ver con que se puede expresar como la suma de "n" cuadrados de v.a. **independientes** y $N(0;1)$ como en este caso. En definitiva queda

$$\frac{nS^2}{\sigma^2} \sim \chi_n^2$$

Entonces sabemos que $\frac{nS^2}{\sigma^2}$ se distribuye según una χ_n^2, pero ahora queremos analizar como se distribuye S^2. Pero necesitamos otro resultado.

Producto de una Gamma por una cte

$$\boxed{X \sim G(\alpha;\beta) \longrightarrow \boxed{Y = cte\ X} \longrightarrow Y \sim G(\alpha;\tfrac{\beta}{cte})}$$

Usando esto, como $\frac{nS^2}{\sigma^2} \sim G(\frac{n}{2};\frac{1}{2})$, si multiplicamos el lado izquierdo por la constante $\frac{\sigma^2}{n}$, el lado derecho será una $G(\frac{n}{2};\frac{n}{2\sigma^2})$ o sea

$$S^2 \sim G(\frac{n}{2};\frac{n}{2\sigma^2})$$

luego resulta, utilizando $\mu_\gamma = \frac{\alpha}{\beta}$ y $var_\gamma = \frac{\alpha}{\beta^2}$

$$E(S^2) = \frac{n/2}{n/2\sigma^2} = \sigma^2(\text{insesgado}) \text{ y } Var(S^2) = \frac{n/2}{(n/2\sigma^2)^2} = \frac{2\sigma^4}{n}$$

Entonces el estimador de varianza S^2 es insesgado, y la varianza de este estimador es $\frac{2\sigma^4}{n}$(como en todo estimador, la varianza del estimador mide que tan cerca de σ^2 estará S^2).

7.3. Estimación de μ y σ (ambos desconocidos). (ambos parámetro desconocidos μ y $v = \sigma^2$ para evitar radicales)

$$\boxed{N(\mu;\sigma^2)}$$
$$X_1 X_2 \cdot\cdot X_n$$

Aplicando el método de MV se obtienen los estimadores de μ y $v = \sigma^2$

$$\widehat{\mu} = \frac{X_1 + X_2 + \cdot\cdot X_n}{n} = \overline{X} \quad \text{y} \quad \widehat{v} = \frac{(X_1 - \overline{X})^2 + \cdot\cdot + (X_n - \overline{X})^2}{n}$$

Si se consulta el listado de estadísticos suficientes, resulta $\widehat{\mu} = \frac{H_1}{n}$ y $\widehat{v} = \frac{H_2}{n}$ ambos estimadores son suficientes.

Respecto del primero se tiene que

$$\widehat{\mu} = \overline{X} \sim N(\mu; \frac{\sigma}{\sqrt{n}}) \text{ o sea insesgado}$$

Para el segundo, para ver si es insesgado habría que evaluar $E(\widehat{v})$ pero, como da un poco de trabajo, se presenta el resultado

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

O sea es sesgado. El estimador $\widehat{v}$ no estima σ^2, sinó un "poco" menos. Si se corrige por sesgo, designando S^2 al estimador insesgado

$$S^2 = \frac{\widehat{v}}{\frac{n-1}{n}} = \frac{\frac{(X_1 - \overline{X})^2 + \cdot\cdot + (X_n - \overline{X})^2}{n}}{\frac{n-1}{n}} = \frac{(X_1 - \overline{X})^2 + \cdot\cdot + (X_n - \overline{X})^2}{n-1} \text{ este sí es insesgado}$$

O sea, cuando se conoce la media de la normal($\mu_\circ$), como fué analizado anteriormente, en el estimador de varianza se divide por n. Pero al desconocer la media, el divisor debe ser $n-1$ para que el estimador sea insesgado.

Para analizar su distribución, procediendo como en el caso anterior, pasando $n-1$ a la izquierda, y dividiendo por σ^2 resulta

$$\frac{(n-1)S^2}{\sigma^2} = \left(\frac{X_1 - \overline{X}}{\sigma}\right)^2 + \cdot\cdot + \left(\frac{X_n - \overline{X}}{\sigma}\right)^2 \tag{7.1}$$

En el caso anterior(donde $\mu = \mu_\circ$), el lado derecho quedaba expresado como la suma $Z_1^2 + \cdot\cdot + Z_n^2$ de n v.a. $N(0;1)$ independientes, y entonces resultaba una χ^2_n.

Pero ahora ocurren dos problemas: 1°) como en cada paréntesis figura $\overline{X}$ en lugar de la verdadera media μ, la distribución no es $N(0;1)$ ya que por ejemplo

$$\frac{X_i - \overline{X}}{\sigma} \sim N(0; \sqrt{\frac{n^2 - n + 1}{n^2}}) \neq N(0;1)$$

Aunque esto se podría corregir , mas grave es el 2°) las v.a. contenidas en los paréntesis no son v.a. independientes. En efecto el $\overline{X}$ esta presente en todos los paréntesis, o mejor aún, si las sumamos

$$\frac{X_1 - \overline{X}}{\sigma} + \cdot\cdot + \frac{X_n - \overline{X}}{\sigma} = \frac{X_1 + \cdot\cdot + X_n - n\overline{X}}{\sigma} = \frac{n\overline{X} - n\overline{X}}{\sigma} = 0$$

Claramente estas v.a. son dependientes ya que su suma da exactamente 0 siempre. Luego, ni son $N(0;1)$ ni tampoco independientes.

Sin embargo se puede demostrar, y esto no es tan simple, que el lado derecho de (7.1) se puede expresar como la suma de $n-1$ v.a. $N(0;1)$ independientes así

$$\frac{(n-1)S^2}{\sigma^2} = Z_1^2 + \cdot\cdot + Z_{n-1}^2$$

Donde cada Z_i es función de las X_i. Entonces según lo visto antes, resulta

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

Notar que los grados de libertad de la χ^2_{n-1} coincide con el divisor utilizado en la fórmula de S^2 para hacer que sea insesgado.

Entonces, como $\frac{(n-1)S^2}{\sigma^2} \sim G(\frac{n-1}{2}; \frac{1}{2})$, si multiplicamos el lado derecho por la constante $\frac{\sigma^2}{n-1}$, el lado izquierdo será una $G(\frac{n-1}{2}; \frac{n-1}{2\sigma^2})$ o sea

$$S^2 \sim G(\frac{n-1}{2}; \frac{n-1}{2\sigma^2})$$

luego resulta, utilizando $\mu_\gamma = \frac{\alpha}{\beta}$ y $var_\gamma = \frac{\alpha}{\beta^2}$
(7.2)

$$E(S^2) = \frac{(n-1)/2}{(n-1)/2\sigma^2} = \sigma^2(\text{insesgado}) \text{ y } Var(S^2) = \frac{(n-1)/2}{((n-1)/2\sigma^2)^2} = \frac{2\sigma^4}{n-1}$$

Aquí se obtuvieron los estimadores $\overline{X}$ y S^2 para estimar los parámetros μ y σ^2 ambos desconocidos. Ambos estimadores se construyeron utilizando la **misma** muestra $X_1X_2 \cdot\cdot X_n$. Por este motivo, lo esperable es que las v.a. $\overline{X}$ y S^2 sean dependientes. Sin embargo se demuestra(y no es tan simple), que para población **normal** los estimadores $\overline{X}$ y S^2 son v.a. **independientes**. Este resultado se usará más adelante.

8. Estimador en pool

Ahora se analizará una situación que se presenta a veces, cuando se tienen dos poblaciones independientes, que comparten un mismo parámetro, y se quiere, con las dos muestras estimar este parámetro común. Concretamente

$$\underset{X_1X_2\cdot\cdot X_{n_x}}{\boxed{f(x/\theta)}} \qquad\qquad \underset{Y_1Y_2\cdot\cdot Y_{n_y}}{\boxed{f(y/\theta)}}$$

Supongamos, que por MV o el método que sea, tenemos los estimadores insesgados

$$\widehat{\Theta}_x = g_x(X_1X_2\cdot\cdot X_{n_x}) \quad \text{y} \quad \widehat{\Theta}_y = g_y(Y_1Y_2\cdot\cdot Y_{n_y})$$

Entonces para estimar θ, podriamos utilizar $\widehat{\Theta}_x$, o también $\widehat{\Theta}_y$.

Pero ya que tenemos dos muestras, la idea es ver si podemos combinarlos en un solo estimador que llamaremos $\widehat{\Theta}_{pool}$.

Una primera propuesta podría ser promediar ambas estimaciones, o sea

$$\widehat{\Theta}_{pool} = \frac{\widehat{\Theta}_x + \widehat{\Theta}_y}{2}$$

Esto parece atractivo, sin embargo si por ejemplo $n_x = 5$ y $n_y = 1000$, probablemente el estimador $\widehat{\Theta}_y$ será mucho mas confiable(tendrá menor varianza) que el $\widehat{\Theta}_x$, y entonces al promediar un estimador malo con uno bueno, estariamos arruinando el estimador bueno.

Notar que esta propuesta de estimador en pool equivale a

$$\widehat{\Theta}_{pool} = \frac{1}{2}\widehat{\Theta}_x + \frac{1}{2}\widehat{\Theta}_y$$

O sea el estimador asigna igual "peso" a cada estimador, en este caso $\frac{1}{2}$ y $\frac{1}{2}$. Quizás, si el mejor es $\widehat{\Theta}_y$ lo correcto seriá usar otros "pesos", por ejemplo $\frac{1}{10}$ y $\frac{9}{10}$.

Entonces la idea es proponer como estimador en pool una expresión del tipo

$$\widehat{\Theta}_{pool} = \alpha_x\widehat{\Theta}_x + \alpha_y\widehat{\Theta}_y \text{ donde } \alpha_x \text{ y } \alpha_y \text{ son los "pesos" a determinar}$$

Como se quiere que $\widehat{\Theta}_{pool}$ sea insesgado, se exigirá que

$$E(\widehat{\Theta}_{pool}) = \theta$$

Pero $E(\widehat{\Theta}_{pool}) = E(\alpha_x\widehat{\Theta}_x + \alpha_y\widehat{\Theta}_y) = \alpha_x E(\widehat{\Theta}_x) + \alpha_y E(\widehat{\Theta}_y) = \alpha_x\theta + \alpha_y\theta = \theta$

De aquí resulta que una primera condición a cumplir por los "pesos" es

$$\alpha_x + \alpha_y = 1$$

Reemplazando el estimador buscado debera ser

$$\widehat{\Theta}_{pool} = \alpha_x\widehat{\Theta}_x + (1-\alpha_x)\widehat{\Theta}_y$$

Cualquiera sea el valor de α_x este estimador es insesgado. Ahora se buscará α_x de manera que el estimador $\widehat{\Theta}_{pool}$ tenga varianza mínima. Calculando su varianza

$$Var(\widehat{\Theta}_{pool}) = \alpha_x^2 Var(\widehat{\Theta}_x) + (1-\alpha_x)^2 Var(\widehat{\Theta}_y)$$

Ahora se deriva para encontrar el α_x que minimiza esta varianza

$$\frac{dVar(\widehat{\Theta}_{pool})}{d\alpha_x} = 2\alpha_x Var(\widehat{\Theta}_x) - 2(1-\alpha_x)Var(\widehat{\Theta}_y) = 0$$

De aquí sale que

$$\alpha_x = \frac{Var(\widehat{\Theta}_y)}{Var(\widehat{\Theta}_x) + Var(\widehat{\Theta}_y)} \quad \text{y} \quad \alpha_y = \frac{Var(\widehat{\Theta}_x)}{Var(\widehat{\Theta}_x) + Var(\widehat{\Theta}_y)}$$

Si se divide numerador y denominador por el producto $Var(\widehat{\Theta}_x)Var(\widehat{\Theta}_y)$, se obtiene una expresión mas compleja en este caso, pero mas simple de generalizar a varias poblaciones

$$\alpha_x = \frac{Var^{-1}(\widehat{\Theta}_x)}{Var^{-1}(\widehat{\Theta}_x) + Var^{-1}(\widehat{\Theta}_y)} \quad \text{y} \quad \alpha_y = \frac{Var^{-1}(\widehat{\Theta}_y)}{Var^{-1}(\widehat{\Theta}_x) + Var^{-1}(\widehat{\Theta}_y)}$$

Un requisito importante que se debe cumplir para que este estimador sea útil, es que estos "pesos" α_x y α_y, no dependan de θ. Y esto es así ya que un estimador debe depender solo de las muestras, conocidas, y nunca del parámetro θ que se quiere estimar.

Example 8. *Si se tiene tres muestras independientes de poblaciones que comparten el mismo parámetro θ*

$$\underset{X_1X_2\cdot\cdot X_{n_x}}{\boxed{f(x/\theta)}} \qquad \underset{Y_1Y_2\cdot\cdot Y_{n_y}}{\boxed{f(y/\theta)}} \qquad \underset{Z_1Z_2\cdot\cdot Z_{n_z}}{\boxed{f(z/\theta)}}$$

El estimador en pool de θ, $\widehat{\Theta}_{pool} = \alpha_x\widehat{\Theta}_x + \alpha_y\widehat{\Theta}_y + \alpha_z\widehat{\Theta}_z$ tendrá

$$\begin{aligned} \alpha_x &= \frac{Var^{-1}(\widehat{\Theta}_x)}{Var^{-1}(\widehat{\Theta}_x) + Var^{-1}(\widehat{\Theta}_y) + Var^{-1}(\widehat{\Theta}_z)} \\ \alpha_y &= \frac{Var^{-1}(\widehat{\Theta}_y)}{Var^{-1}(\widehat{\Theta}_x) + Var^{-1}(\widehat{\Theta}_y) + Var^{-1}(\widehat{\Theta}_z)} \\ \alpha_z &= \frac{Var^{-1}(\widehat{\Theta}_z)}{Var^{-1}(\widehat{\Theta}_x) + Var^{-1}(\widehat{\Theta}_y) + Var^{-1}(\widehat{\Theta}_z)} \end{aligned}$$

(útil si α_x, α_y y α_z no dependen de θ).

EXAMPLE 9. *Se tienen artículos de peso $A \sim N(\mu;\sigma^2)$. Se venden en cajas de 5 y cajas de 10. Se tienen las muestras:*

- ▶ $X_1X_2\cdot\cdot X_8$ de pesos de 8 cajas chicas
- ▶ $Y_1Y_2\cdot\cdot Y_3$ de pesos de 3 cajas grandes

Se pide estimar μ y σ^2?

SOLUTION 1. *Primero hay que averiguar de que poblaciones provienen las dos muestras. Para esto llamemos X a el peso de una cualquiera de las 8 cajas chicas. Como $X = A_1 + A_2 + \cdot\cdot + A_5 \sim N(5\mu;5\sigma^2)$, entonces la muestra $X_1X_2\cdot\cdot X_8$ proviene de una $N(5\mu;5\sigma^2)$. Además llamando Y al peso de una cualquiera de las cajas grandes, y como $Y = A_1^* + A_2^* + \cdot\cdot + A_{10}^* \sim N(10\mu;10\sigma^2)$, la población de la segunda muestra es $N(10\mu;10\sigma^2)$. En definitiva se tiene*

$$\underset{X_1X_2\cdot\cdot X_8}{\boxed{N(5\mu;5\sigma^2)}} \qquad\qquad \underset{Y_1Y_2Y_3}{\boxed{N(10\mu;10\sigma^2)}}$$

Como $S_x^2 \xrightarrow{I} 5\sigma^2$(es estimador insesgado de $5\sigma^2$), resulta $\frac{S_x^2}{5} \xrightarrow{I} \sigma^2$ (estimador insesgado de σ^2). Además como $S_y^2 \xrightarrow{I} 10\sigma^2$, resulta también $\frac{S_y^2}{10} \xrightarrow{I} \sigma^2$. Entonces, para lograr un mejor estimador de σ^2, la idea es combinar en pool los estimadores insesgados $\frac{S_x^2}{5}$ y $\frac{S_y^2}{10}$. Para obtenerlo se necesitan las varianzas de estos dos estimadores. Se recuerda la expresión de la varianza de un estimador de varianza, vista en (7.2)

$$Var(S^2) = \frac{2\sigma^4}{n-1} \text{ (donde } \sigma^4 = (\sigma^2)^2 \text{ es el cuadrado de la varianza poblacional)}$$

Entonces:

$$Var(\frac{S_x^2}{5}) = \frac{\frac{2(5\sigma^2)^2}{8-1}}{5^2} = \frac{2\sigma^4}{8-1} \quad y \quad Var(\frac{S_y^2}{10}) = \frac{\frac{2(10\sigma^2)^2}{3-1}}{10^2} = \frac{2\sigma^4}{3-1}$$

Luego

$$S_{pool}^2 = \frac{\frac{8-1}{2\sigma^4}\frac{S_x^2}{5} + \frac{3-1}{2\sigma^4}\frac{S_y^2}{10}}{\frac{8-1}{2\sigma^4} + \frac{3-1}{2\sigma^4}} = \frac{(8-1)\frac{S_x^2}{5} + (3-1)\frac{S_y^2}{10}}{(8-1)*(3-1)} = \frac{7\frac{S_x^2}{5} + 2\frac{S_y^2}{10}}{7+2}$$

O sea, y esto es general para un estimador en pool de varianza, los dos estimadores insesgados $\frac{S_x^2}{5}$ y $\frac{S_y^2}{10}$ se combinan en pool con "pesos" proporcionales a los grados de libertad de los estimadores de varianza involucrados, o sea 8-1 y 3-1

$$S_{pool}^2 = \frac{7\frac{S_x^2}{5} + 2\frac{S_y^2}{10}}{7+2}$$

Además vale también que

$$\frac{(7+2)S_{pool}^2}{\sigma^2} \sim \chi_{7+2}^2$$

Notar que al tener la χ_{7+2}^2, 9 grados de libertad, el estimador S_{pool}^2 es mejor que los dos iniciales(tiene menor varianza). Ahora se analizará un estimador para el parámetro μ. Como $\overline{X} \xrightarrow{I} 5\mu$ y $\overline{Y} \xrightarrow{I} 10\mu$, resultan $\frac{\overline{X}}{5} \xrightarrow{I} \mu$ y $\frac{\overline{Y}}{10} \xrightarrow{I} \mu$ ambos estimadores insesgador de μ. Para combinarlos el pool se necesitan sus varianzas

$$Var(\frac{\overline{X}}{5}) = \frac{\frac{5\sigma^2}{8}}{5^2} = \frac{\sigma^2}{40} \quad y \quad Var(\frac{\overline{Y}}{10}) = \frac{\frac{10\sigma^2}{3}}{10^2} = \frac{\sigma^2}{30}$$

Luego

$$\widehat{\mu}_{pool} = \frac{\frac{40}{\sigma^2}\frac{\overline{X}}{5} + \frac{30}{\sigma^2}\frac{\overline{Y}}{10}}{\frac{40}{\sigma^2} + \frac{30}{\sigma^2}} = \frac{40\frac{\overline{X}}{5} + 30\frac{\overline{Y}}{10}}{40 + 30}$$

Que expresado en función de $\overline{X}$ *y* $\overline{Y}$ *queda*

$$\widehat{\mu}_{pool} = \frac{8}{70}\overline{X} + \frac{3}{70}\overline{Y}$$

Por ejemplo si interesa la distribución de este estimador, como es insesgado $E(\widehat{\mu}_{pool}) = \mu$*, y al ser lineal de poblaciones normales*

$$\widehat{\mu}_{pool} \sim N(\mu; \sqrt{\frac{8^2}{70^2}\frac{5\sigma^2}{8} + \frac{3^2}{70^2}\frac{10\sigma^2}{3}}) = N(\mu; \frac{\sigma}{\sqrt{70}})$$

CHAPTER 2

INTERVALOS DE CONFIANZA

1. Definición de Intervalo de Confianza

Si se necesita averiguar el parámetro θ de una densidad $f(x/\theta)$, por ejemplo para utilizar después dicha densidad en otros cálculos, lo apropiado sería, con una muestra $x_1x_2\cdot\cdot x_n$ y un estimador, obtener el estimado $\widehat{\theta}$, reemplazarlo en la densidad, obteniendo $f(x/\widehat{\theta})$, y luego utilizarla en los cálculos deseados.

Pero claro, $\widehat{\theta}$ es solo una estimación que obtuvimos con una muestra, y no está garantizado que coincida con el verdadero valor de θ. Lo conveniente es este caso sería estimar también el desvío de la estimación $\widehat{\sigma}_{\widehat{\theta}}$. Si es pequeño podemos suponer que $\widehat{\theta}\simeq\theta$, lo que justificaría usar $\widehat{\theta}$ para reemplazar en la densidad. Si es grande, convendría tomar una muestra mas grande, para obtener una mejor estimación.

O sea, acompañar la estimación $\widehat{\theta}$ con su desvío $\widehat{\sigma}_{\widehat{\theta}}$, es una manera de evaluar que tan cerca está $\widehat{\theta}$ de θ.

Otra manera de presentar esto es con un intervalo de confianza para θ. En este caso, con la muestra $x_1x_2\cdot\cdot x_n$, se "fabrica" un intervalo, por ejemplo $[2;5]$, que tiene la propiedad de incluir al verdadero θ, con un "nivel de confianza" alto(no se dice "probabilidad", se dice "confianza", por motivos que se verán más adelante).

El nivel de confianza del intervalo, con notación NC, se **impone** antes de obtener el intervalo, y usualmente se usan NC de 0.90, 0.95 o 0.99(a veces se habla de NC al 90%, 95% o 99%).

Ya que el NC se fija antes de obtener el intervalo, podría pensarse que lo ideal sería imponer por ejemplo NC=0.999, ya que de esta manera el intervalo obtenido tendría un altísimo "nivel de confianza" de incluir a θ. Lamentablemente cuando se toma un NC tan alto, el correspondiente intervalo se ensancha, y pierde utilidad. Al revés, si tomamos NC=0.1, muy bajo, el intervalo obtenido será muy angosto, lo cual es deseable, pero con muy poco "nivel de confianza" de incluir a θ. Esto tampoco sirve. Por eso los valores sugeridos.

Por último antes de la definición, si NC es el nivel de confianza impuesto, se designa $\alpha=1-NC$. O sea siempre

$$NC+\alpha=1$$

Definition 2. ***Intervalo de Confianza de nivel NC:*** *Sea la población y una muestra aleatoria*

$$\boxed{f(x/\theta)}$$
$$X_1X_2\cdot\cdot X_n$$

Supongamos que encontramos dos funciones de la muestra, $L=l(X_1X_2\cdot\cdot X_n)$ y $U=u(X_1X_2\cdot\cdot X_n)$ (obviamente L y U son v.a., ya que son funciones de la muestra, que es aleatoria también), de manera que podemos demostrar que se cumple

$$P(L \leq \theta \leq U) = NC \quad \text{o sea:} \tag{1.1}$$

$$P(l(X_1X_2 \cdot \cdot X_n) \leq \theta \leq u(X_1X_2 \cdot \cdot X_n)) = NC \tag{1.2}$$

Entonces estos dos extremos definen el intervalo de confianza aleatorio

$$IC_\theta \text{ al NC} : [l(X_1X_2 \cdot \cdot X_n); u(X_1X_2 \cdot \cdot X_n)] \tag{1.3}$$

Lo que garantiza esta definición, es que, **pensado "a priori"**(o sea, antes de tomar la muestra): cuando tomemos una muestra concreta, la **probabilidad** de que el intervalo obtenido incluya a θ es NC(aquí sí es válido utilizar el término probabilidad).

O de otra manera, si tomamos una muestra, y después otra, y otra, etc. y con cada una de ellas reemplazamos en (1.3) obteniendo los correspondientes intervalos, alrededor del NC% de los intervalos obtenidos incluirá a el parámetro θ.

O sea, es el **método de fabricación** del intervalo, el que garantiza con **probabilidad** NC, que el intervalo incluirá a θ.

Sin embargo, cuando tomamos una muestra concreta $x_1x_2 \cdot \cdot x_n$ y obtenemos el correspondiente intervalo $[l(x_1x_2 \cdot \cdot x_n); u(x_1x_2 \cdot \cdot x_n)]$, **NO** podemos, esta vez "**a posteriori**"(o sea, después de conocer la muestra), afirmar que este intervalo tiene probabilidad NC de incluir a θ. Este intervalo podrá o no incluir a θ. Por eso en este caso decimos solamente que tenemos un "**nivel de confianza NC**" que incluya a θ, ya que fué construído con un método que "a priori" garantiza con **probabilidad NC**, la inclusión de θ.

A continuación se presentarán dos procedimientos para obtener un intervalo de confianza: el método del estadístico pivotal, y el método del estadístico muestral.

En ambos, y esto es similar a lo recomendado en los problemas de estimación, no se trabajará con toda la muestra, sinó que se partirá de un estadistico muestral, que es una sola v.a., lo que hace que los desarrollos para obtener el intervalo sean mas simples.

De esta manera obtendremos un "buen" intervalo de confianza.

Lo de "bueno" quiere decir que para el NC que fijamos, la amplitud del intervalo obtenido es en general mas pequeña. O sea si lo fabricamos partiendo de un estadistico muestral no suficiente, es de esperar intervalos mas anchos.

DEFINITION 3. ***Fractil de orden α de una v.a. continua:*** *Sea X una v.a. continua con función de distribución $F(x)$ estrictamente creciente, α una probabilidad, con $0 < \alpha < 1$, se define fractil de orden α de la v.a. X, con notación x_α, al valor de la variable que deja a la izquierda(incluído x_α), una probabilidad α, o sea:*

$$F(x_\alpha) = \alpha \quad \textit{o también } P(X \leq x_\alpha) = \alpha$$

Por ejemplo si $Z \sim N(0;1)$, se tiene:

a) $z_{0.5} = 0$, ya que $F_z(0) = 0.5$, o también $P(Z \leq 0) = 0.5$

b) $z_{0.05} = -1.645$, ya que $F_z(-1.645) = 0.05$, o tambiéne $P(Z \leq -1.645) = 0.05$

c) $z_{0.90} = 1.282$, ya que $F_z(1.282) = 0.90$, o tambiéne $P(Z \leq 1.282) = 0.90$

Y en general, si se quiere el fractil de orden, por ejemplo 0.27 de una v.a., hay que resolver la ecuación $F_x(x) = 0.27$, y la solución será $x_{0.27}$.

2. Método del estadístico pivotal

Se presentará en varios pasos. Sea la población **continua** y una muestra aleatoria

$$\boxed{f(x/\theta)}$$
$$X_1 X_2 \cdot \cdot X_n$$

1) Sea $H = g(X_1 X_2 \cdot \cdot X_n)$ un estadístico suficiente respecto de θ

2) Averiguamos la distribución de H, supongamos que

$$H \sim f(h/\theta)$$

Notar que en esta expresión, del lado izquierdo esta $H = g(X_1 X_2 \cdot \cdot X_n)$ que es función **solo de la muestra**, y en el lado derecho aparece el parámetro θ, dentro de la densidad de H.

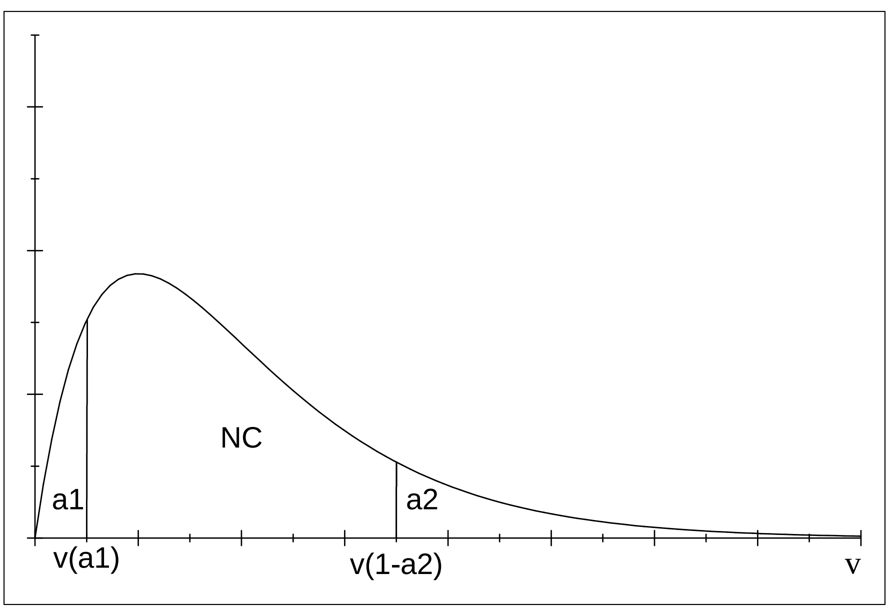
3) Aquí se define un estadístico **"pivotal"** $V = v(H; \theta)$, que es una función de H y del parámetro θ, con la particular propiedad

$$V = v(H; \theta) \sim f(v)$$

Notar que del lado izquierdo aparece la muestra(en H) y el parámetro θ, y del lado derecho una densidad **totalmente conocida,** ya que aquí no aparece el parámetro. Es importante que esta densidad sea totalmente conocida, o sea no debe aparecer en ella ningun otro parámetro desconocido.

Se aclara que para un estadístico muestral, no siempre se puede obtener un estadístico pivotal con las propiedades exigidas. Por eso se presenta después otro método para obtener intervalos de confianza.

4) Como $f(v)$ es conocida, podemos dibujarla. A continuación se buscan dos abcisas, que encierren entre ellas una probabilidad NC. O, lo que es lo mismo, que fuera de ellas la probabilidad sea $1 - NC = \alpha$. Según la figura deberá ser $\alpha_1 + \alpha_2 = \alpha$(esta escrito a1 y a2 en lugar de α_1 y α_2). Entonces una de las abcisas es el fractil v_{α_1} (en la figura v(a1)), que deja a la izquierda una probabilidad α_1; y la otra el fractil $v_{1-\alpha_2}$ (en la figura v(1-a2), que deja a la derecha una probabilidad α_2.



REMARK 6. *Como NC es conocido, también lo es $\alpha = 1 - NC$. Entonces la elección de α_1 y α_2 queda indeterminada, sujeta al cumplimiento $\alpha_1 + \alpha_2 = \alpha$. Usualmente en los problemas se toma $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$, y entonces los fractiles a considerar son $v_{\alpha/2}$ y $v_{1-\alpha/2}$.*

Aquí para mayor generalidad, consideraremos los fractiles v_{α_1} y $v_{1-\alpha_2}$ entonces

$$P(v_{\alpha_1} \leq V \leq v_{1-\alpha_2}) = NC$$

5) Ahora, como $V = v(H;\theta)$, se reemplaza en la expresión anterior V por $v(H;\theta)$ quedando

$$P(v_{\alpha_1} \leq v(H;\theta) \leq v_{1-\alpha_2}) = NC \tag{2.1}$$

Lo que está dentro del paréntesis es todo conocido salvo θ. En realidad es una doble desigualdad, que equivale a la intersección

$$(v_{\alpha_1} \leq v(H;\theta)) \cap (v(H;\theta) \leq v_{1-\alpha_2})$$

Si se despeja θ de la primera, $v_{\alpha_1} \leq v(H;\theta)$, puede quedar así: $l(H;v_{\alpha_1}) \leq \theta$ (a veces queda al revés). Y despejando θ de la segunda $v(H;\theta) \leq v_{1-\alpha_2}$ puede quedar así: $\theta \leq u(H;v_{1-\alpha_2})$ (también, a veces queda al revés), luego es equivalente a escribir

$(l(H;v_{\alpha_1}) \leq \theta) \cap (\theta \leq u(H;v_{1-\alpha_2}))$ que es lo mismo que $l(H;v_{\alpha_1}) \leq \theta \leq u(H;v_{1-\alpha_2})$

Y reemplazando en el parentesis de (2.1) queda

$$P(l(H;v_{\alpha_1}) \leq \theta \leq u(H;v_{1-\alpha_2})) = NC$$

pero esta es la definición del intervao de confianza deseado, o sea

$$IC_\theta \text{ al NC} : [l(H;v_{\alpha_1}) \leq \theta \leq u(H;v_{1-\alpha_2})]$$

Se presentarán un par de ejemplos.

3. IC para μ de una población normal con $\sigma_\circ$ conocida

$$\boxed{N(\mu;\sigma_\circ)}$$
$$X_1 X_2 \cdot\cdot X_n$$

1) Partimos de $H = X_1 + X_2 + \cdot\cdot + X_n$ que es suficiente respecto de μ. Pero por comodidad se usará $\overline{X} = \frac{H}{n}$ que es también suficiente, y de paso es un estimador de μ.

2) Distribución: $\overline{X} \sim N(\mu; \frac{\sigma_\circ}{\sqrt{n}})$

3) Estadístico pivotal: estandarizando resulta $Z = \frac{\overline{X}-\mu}{\sigma_\circ/\sqrt{n}} \sim N(0;1)$ y este es un estadístico pivotal

4) En la N(0;1) se buscan dos fractiles que encierren probabilidad NC(tomamos $z_{\alpha/2}$ y $z_{1-\alpha/2}$). Entonces resultará

$$P(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}) = NC$$

5) Reemplazando aquí $Z = \frac{\overline{X}-\mu}{\sigma_\circ/\sqrt{n}}$ queda

$$P(z_{\alpha/2} \leq \frac{\overline{X}-\mu}{\sigma_\circ/\sqrt{n}} \leq z_{1-\alpha/2}) = NC$$

Trabajando con la doble desigualdad resulta

$$P(\overline{X} - z_{1-\alpha/2}\frac{\sigma_\circ}{\sqrt{n}} \leq \mu \leq \overline{X} - z_{\alpha/2}\frac{\sigma_\circ}{\sqrt{n}}) = NC$$

En el caso particular de la $N(0;1)$, por la simetría, resulta $z_{\alpha/2} = -z_{1-\alpha/2}$, expresando el intervalo en función de $z_{1-\alpha/2}$, queda el intervalo

$$\left[\overline{X} - z_{1-\alpha/2}\frac{\sigma_\circ}{\sqrt{n}}; \overline{X} + z_{1-\alpha/2}\frac{\sigma_\circ}{\sqrt{n}}\right] \text{ o también expresado } \left[\overline{X} \pm z_{1-\alpha/2}\frac{\sigma_\circ}{\sqrt{n}}\right]$$

Notar que el estimador de μ es $\widehat{\mu} = \overline{X}$, y el intervalo de confianza tiene el aspecto $\overline{X} \pm plus$, donde el *plus* es la semi-amplitud del intervalo. Cuanto mayor tomemos el NC, mayor será la semi-amplitud del intervalo. Esto parece lógico. Además la semi-amplitud es constante para un mismo tamaño de muestra(esto no siempre ocurre), o sea, no depende de los valores particulares de la muestra, es decir de los $X_1X_2 \cdot \cdot X_n$.

REMARK 7. ***1)*** *Sea la Población $N(\mu;5)$ y la muestra $X_1X_2 \cdot \cdot X_{16}$ de la que se conoce que $\overline{X} = 30$ que es un estimador de μ. Se quiere un IC_μ con NC=0.90. Como $\alpha = 0.10$, con el PQRS obtenemos $z_{0.05} = -1.645$ y $z_{0.95} = +1.645$ luego* ***un*** *IC_μ al 90% será*

$$\left[30 \pm 1.645\frac{5}{\sqrt{16}}\right] = [27.94; 32.06] \text{ con semi-amplitud} = \frac{32.06 - 27.94}{2} = 2.06$$

REMARK 8. ***2)*** *En el ejemplo anterior los dos fractiles que encierran probabilidad 0.9 fueron $z_{0.05}$ y $z_{0.95}$. Pero que tal si hubiesemos elegido los fractiles $z_{0.01} = -2.326$ y $z_{0.91} = 1.341$. Estos fractiles también encierran una probabilidad 0.9. Y se puede construir un IC con ellos. Pero habrá que partir de*

$$P(z_{0.01} \leq Z \leq z_{0.91}) = 0.90 \text{ y luego } P(z_{0.01} \leq \frac{\overline{X} - \mu}{\sigma_\circ/\sqrt{n}} \leq z_{0.91}) = 0.90$$

Trabajando con la desigualdad, y reemplazando, queda el intervalo

$$\left[\overline{X} - z_{0.91}\frac{\sigma_\circ}{\sqrt{n}}; \overline{X} - z_{0.01}\frac{\sigma_\circ}{\sqrt{n}}\right] = [28.32; 32.91] \text{ con semi-amplitud} = \frac{32.91 - 28.32}{2} = 2.30$$

Notar que los dos intervalos obtenidos tienen el mismo NC=90%, pero el primero es mas angosto, así que es más conveniente. En el caso de la $N(0;1)$, como es simétrica, los fractiles $z_{\alpha/2}$ y $z_{1-\alpha/2}$ son los que están mas cerca entre sí, y los que proporcionan un IC mas angosto.

REMARK 9. ***3)*** *Supongamos que la semi-amplitud obtenida en 1) parece demasiado grande. La pregunta ahora es: que tamaño de muestra se debe tomar, para que el IC tenga, por ejemplo semi-amptud = 1. Deberá ser entonces*

$$1.645\frac{5}{\sqrt{n}} = 1$$

Despejando surge $n = 68$, o sea, si queremos un IC mas angosto, el precio a pagar es un mayor tamaño de muestra

4. IC para σ^2 de una población normal con μ desconocida

$$\boxed{N(\mu;\sigma^2)}$$
$$X_1X_2\cdot\cdot X_n$$

1) Del listado de estadísticos suficientes $H_1 = X_1 + X_2 + \cdot\cdot X_n$ y $H_2 = (X_1 - \overline{X})^2 + \cdot\cdot + (X_n - \overline{X})^2$ son suficientes respecto de la pareja $(\mu;\sigma^2)$. Usaremos aquí $S^2 = \frac{H_2}{n-1} = \frac{(X_1-\overline{X})^2+\cdot\cdot+(X_n-\overline{X})^2}{n-1}$ que es también suficiente, y de paso es un estimador de σ^2.

2) Como $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ estonces $\frac{(n-1)S^2}{\sigma^2}$ es un estadístico pivotal

3) En la χ^2_{n-1} se buscan dos fractiles que encierren probabilidad NC(tomamos $\chi^2_{n-1,\alpha/2}$ y $\chi^2_{n-1,1-\alpha/2}$). Entonces resultará

$$P(\chi^2_{n-1,\alpha/2} \leq \chi^2_{n-1} \leq \chi^2_{n-1,1-\alpha/2}) = NC$$

4) Reemplazando el estadístico pivotal

$$P(\chi^2_{n-1,\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1,1-\alpha/2}) = NC$$

5) Trabajando con la doble desigualdad queda

$$P(\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}) = NC$$

En definitiva queda:

$$IC_{\sigma^2} : \left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}; \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}\right] \text{ y también } IC_{\sigma} : \left[\sqrt{\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}}; \sqrt{\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}}\right]$$

REMARK 10. ***1)*** *Si la población normal tuviese la media conocida, o sea $N(\mu_\circ;\sigma^2)$, lo único que cambiaría es que el estimador de varianza a utilizar debería ser $S^2 = \frac{(X_1-\mu_\circ)^2+\cdot\cdot+(X_n-\mu_\circ)^2}{n}$ con un pivotal $\frac{nS^2}{\sigma^2} \sim \chi^2_n$, y el intervalo para la varianza sería*

$$IC_{\sigma^2} : \left[\frac{nS^2}{\chi^2_{n,1-\alpha/2}}; \frac{nS^2}{\chi^2_{n,\alpha/2}}\right]$$

REMARK 11. ***2)*** *Si interesa el problema inverso: Cuál debe ser el tamaño de muestra n para que el IC para σ^2 tenga una amplitud dada, tropezamos con una dificultad. En efecto si se evalua la amplitud del intervalo*

$$amplitud = \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}} - \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}} = (\frac{1}{\chi^2_{n-1,\alpha/2}} - \frac{1}{\chi^2_{n-1,1-\alpha/2}})(n-1)S^2$$

Notar que la amplitud es una v.a, en efecto aparece S^2, que depende de la muestra(que todavía no hemos tomado!!). Si igualamos esta expresión de amplitud a un valor dado, no podremos despejar n pues nos queda el S^2 que desconocemos. Una alternativa consiste en re-definir el concepto de amplitud del intervalo. Se define amplitud al cociente entre el extremo superior / extremo inferior del intervalo

$$amplitud^* = \frac{ext.\sup}{ext.\inf} = \frac{\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}}{\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}} = \frac{\chi^2_{n-1,1-\alpha/2}}{\chi^2_{n-1,\alpha/2}}$$

Por ejemplo, supongamos que interesa averiguar que tamaño de muestra habrá que tomar, para obtener un IC_σ de $NC = 0.90$ tal que su extremo superior sea un 30% mayor que el inferior. Entonces el IC_{σ^2} para σ^2 deberá tener una amplitud$^* = 1.3^2 = 1.69$. Para esto deberá cumpirse

$$\frac{\chi^2_{n-1,0.95}}{\chi^2_{n-1,0.05}} = 1.69$$

Probando con el pqrs resulta $n-1 = 80$, o sea se necesita una muestra de tamaño 81.

REMARK 12. ***3)*** *En el caso del IC_μ tomamos como fractiles $z_{\alpha/2}$ y $z_{1-\alpha/2}$, y se comentó que esta elección de $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ es la óptima, ya que con ellos se obtiene el intervalo mas angosto. Sin embargo, para un IC_{σ^2}, como la χ^2_{n-1} no es simétria, la elección de los fractiles $\chi^2_{n-1,1-\alpha/2}$ y $\chi^2_{n-1,\alpha/2}$ no es la óptima. O sea, para un α dado, habrían que elegir α_1 y α_2, con $\alpha_1 + \alpha_2 = \alpha$ de manera que el IC tenga amplitud mínima. O sea(ver expresión de la amplitud)*

$$\textit{buscar } \alpha_1 \textit{ tal que: } \frac{1}{\chi^2_{n-1,\alpha_1}} - \frac{1}{\chi^2_{n-1,1-\alpha_1}} \textit{ sea mínimo}$$

Como esto es complicado, se toma usualmente $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$.

5. Método del estadístico muestral

Sea la población(se lo desarrollará para una **discreta**, pero vale también para **continua**) y una muestra aleatoria

$$\boxed{p(x/\theta)}$$
$$X_1 X_2 \cdot \cdot X_n$$

1) Sea $H = g(X_1 X_2 \cdot \cdot X_n)$ un estadístico suficiente respecto de θ
2) Averiguamos la distribución de H, supongamos que

$$H \sim p(h/\theta)$$

3) Sea $h = g(x_1 x_2 \cdot\cdot x_n)$ el valor del estadístico para la muestra concreta que tenemos(este valor lo conocemos). Aunque no se demostrará el método, se justificará intuitivamente el procedimiento.

La expresión de la densidad de H, que es $p(h/\theta)$, la conocemos salvo a θ; y de ella salió el h que conocemos.

Ahora, que tal si a θ le damos valores crecientes $\theta_1 \theta_2 \cdots \theta_{100}$ y graficamos las correspondientes $p(h/\theta_i)$. La densidad irá cambiando de lugar según el valor de θ_i.

- A continuación elegimos cual es el valor de θ_i que hace que $p(h/\theta_i)$ tenga sus "bastones" altos alrededor de h. Supongamos que este valor es θ_{40}. Entonces θ_{40} hace que el h observado muestralmente sea un valor "razonable" de observar. Si nuestro interés fuese estimar θ, entonces θ_{40} serviría como estimador de θ.
- Seguidamente probamos con $\theta_{41}\theta_{42} \cdot \cdot etc$ y supongamos que vemos que $p(h/\theta_i)$ se va corriendo más y más a la derecha. A medida que incrementamos i, la correspondiente densidad $p(h/\theta_i)$ hace que el h obtenido sea cada vez más "irrazonable" en relación al respectivo θ_i. Entonces un

extremo del IC será el valor de θ que hace que la probabilidad $P(H \leq h)$ sea pequeña, y tomaremos $\frac{\alpha}{2}$ a esta probabilidad, o sea

$$\theta_{\text{sup}} : P(H \leq h) = \frac{\alpha}{2} \quad \text{o sea} \quad F(h/\theta_{\text{sup}}) = \frac{\alpha}{2}$$

- Ahora probamos para el otro lado con $\theta_{39}\theta_{38} \cdot \cdot etc$, y la densidad se va corriendo a la izquierda. A medida que disminuimos i, la correspondiente densidad $p(h/\theta_i)$ hace que el h obtenido sea cada vez más "irrazonable" en relación al respectivo θ_i. Entonces otro extremo del IC será el valor de θ que hace que la probabilidad $P(H \geq h)$ sea pequeña, y tomaremos $\frac{\alpha}{2}$ a esta probabilidad, o sea

$$\theta_{\text{inf}} : P(H \geq h) = \frac{\alpha}{2} \quad \text{o sea} \quad G(h/\theta_{\text{inf}}) = \frac{\alpha}{2}$$

En definitiva los dos extremos del IC_θ se obtienen resolviendo

$$\begin{array}{l} F(h/\theta) = \frac{\alpha}{2} \\ G(h/\theta) = \frac{\alpha}{2} \end{array} \quad \text{y de aquí sale el intervalo } [\theta_{\text{inf}}; \theta_{\text{sup}}]$$

CLAIM 1. *En la justificación anterior se asumió que al incrementarse θ la $p(h/\theta)$ se corre a la derecha. Si esto es así, con la primera ecuación surge θ_{sup}, y con la segunda θ_{inf}. Sin embargo a veces ocurre al revés, es decir, al incrementarse θ la $p(h/\theta)$ se corre a la izquierda. Entonces, de la primera surge θ_{inf} y de la segunda θ_{sup}. Esto se verá en los ejemplos.*

6. IC para p de una población Bernouilli

$$\begin{array}{c} \boxed{B_i(1;p)} \\ X_1 X_2 \cdot \cdot X_n \end{array}$$

Para concretar supongamos que se fabricaron 20 artículos con una máquina en que $P(artículo\ defectuoso) = p$, y se encontraron 5 artículos defectuosos. Se quiere un IC_p al 90%. Se tendrá entonces

$$\begin{array}{c} \boxed{B_i(1;p)} \\ X_1 X_2 \cdot \cdot X_{20} \end{array}$$

Cada X_i vale 1 o 0, según el correspondiente artículo sea defectuoso, o no.

1) Utilizaremos $H = X_1 + X_2 + \cdot \cdot + X_{20}$ que es suficiente respecto de p.

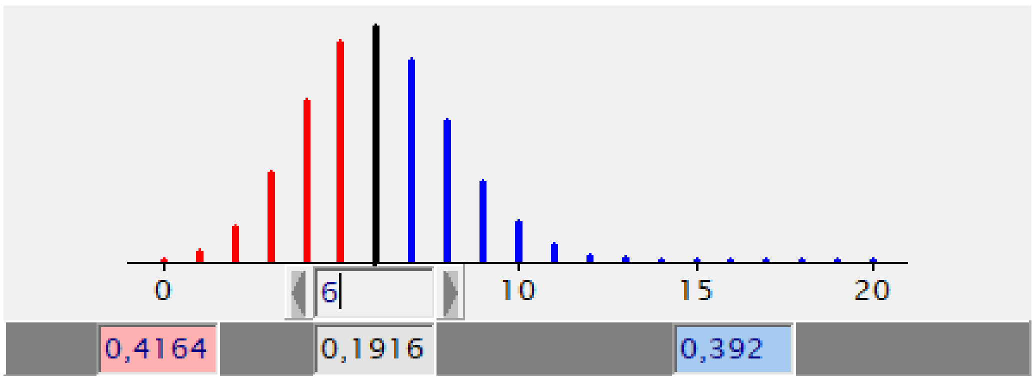
2) Sabemos que $H \sim B_i(20; p)$

3) Como se nos informa que se obtuvieron 5 defectuosos, el valor muestral de H es $h = 5$.

4) Luego las ecuaciones a resolver son:

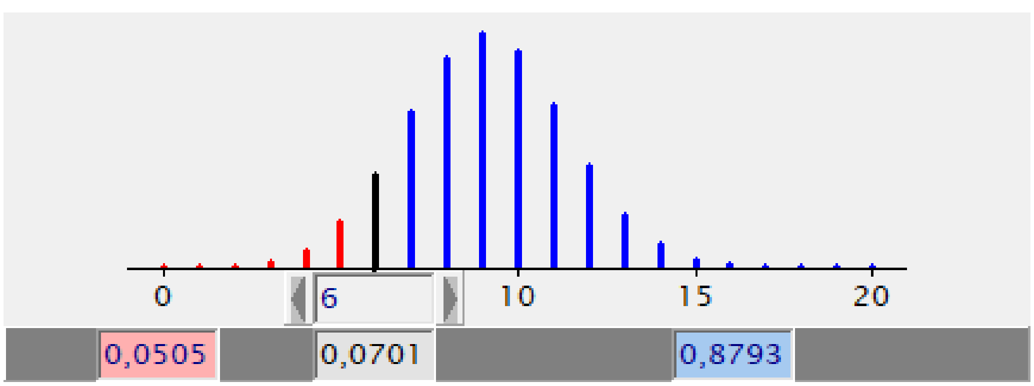
$$\begin{array}{c} P(H \leq 5) = F_{B_i}(5/20; p) = 0.05 \\ P(H \geq 5) = G_{B_i}(5/20; p) = 0.05 \end{array}$$

Ahora sí, para la primera, vamos al pqrs, Binomial, $n = 20$, p elegimos cualquier valor, por ejemplo $p = 0.3$, Botón, y cuando aparece la densidad, en el centro del "tetris" ponemos 6 y luego Botón. De esta manera, en la casilla **izquierda** del "tetris" figurará $P(H \leq 5)$ que en este caso da $0,4164$.



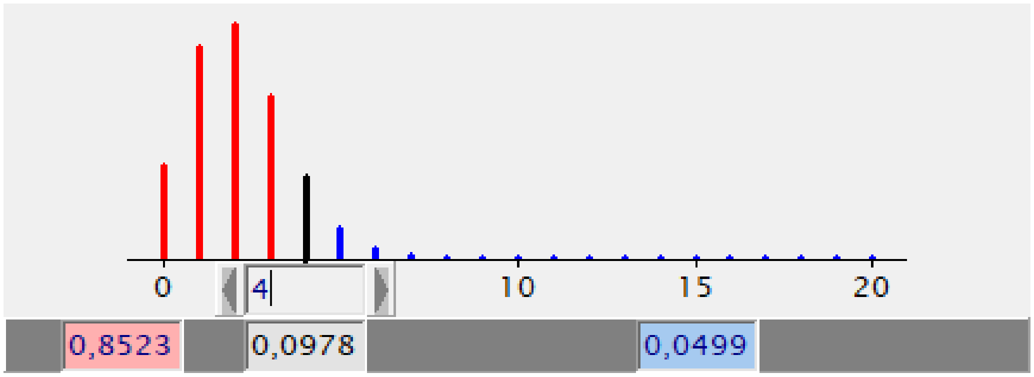
Como $P(H \leq 5) = 0,4164 \neq 0.05$ el $p = 0.3$ elegido inicialmente no es la solución.

En este caso tendremos que elegir un p mayor, de manera que al correrse la densidad a la derecha, resulte $P(H \leq 5) = 0.05$. Seguramente habrá que probar varias veces con diferentes valores de p. Supongamos que ponemos $p = 0,455$, y luego Botón



Ahora sí, para $p = 0,455$, resulta $P(H \leq 5) = 0.0505 \simeq 0.05$, y entonces $p_{\text{sup}} = 0,455$.

Para resolver la segunda ecuación, en el centro del "tetris" ponemos 4 y Botón. De esta manera, en la casilla **derecha** del "tetris" figurará $P(H \geq 5)$. Aquí también hay que probar varias veces. Supongamos que ponemos $p = 0,104$, entonces resulta



Ahora sí, la densidad se corre a la izquierda, y para $p = 0,104$, resulta $P(H \geq 5) = 0.0499 \simeq 0.05$, y entonces $p_{\text{inf}} = 0,104$.

En definitiva el IC_p al 90% es $[0,104; 0,455]$.

7. IC para λ de una Poisson

Example 10. *Ciertos rollos de cable de 10m, presentan fallas de aislación según un proceso de Poisson. Se inspeccionaron 5 rollos, encontrando en c/u de ellos: 2, 1, 1, 0 y 3 fallas. Se pide un IC_β al 90%. Como el número de fallas en cada rollo es $P_o(\lambda)$ con $\lambda = \beta 10$ se tiene*

$$\boxed{P_o(\beta 10)}$$
$$F_1 F_2 \cdot \cdot F_5$$

Usamos como estadístico $H = F_1 + F_2 + \cdot \cdot + F_5$ que es suficiente para λ(y por lo tanto para β), y además su distribución es

$$H \sim P_o(5\lambda) = P_o(50\beta)$$

Por simplicidad llamemos $\lambda^* = 50\beta$, y entonces será $H \sim P_o(\lambda^*)$. El valor de H para la muestra que tenemos es $h = 7$.Entonces planteando las ecuaciones

$$P(H \leq 7) = F_{P_o}(7/\lambda^*) = 0.05$$
$$P(H \geq 7) = G_{P_o}(7/\lambda^*) = 0.05$$

Con el pqrs surge $\lambda^*_{\text{sup}} = 13.15 = 50\beta_{\text{sup}}$ y $\lambda^*_{\text{inf}} = 3.286 = 50\beta_{\text{inf}}$. En definitiva el IC_β al 90% es

$$[0.0657; 0.263]$$

Distribución de Student. A continuación se presenta una v.a. que aparece, cuando en poblaciones normales interesa como parámetro la media, pero se desconoce el desvío estándar. Se define así:

Definición de T_ν de Student con ν grados de libertad

$$\boxed{\begin{array}{c} Z \sim N(0;1) \\ U \sim \chi^2_\nu \end{array} \quad Ind \rightrightarrows \boxed{T = \frac{Z}{\sqrt{\frac{U}{\nu}}}} \longrightarrow T \sim T_\nu}$$

O sea, esta v.a. aparecerá cuando tengamos una expresión, donde en el numerador hay v.a. que es $N(0;1)$ y en el denominador la raíz cuadrada de otra, que es una v.a. χ^2_ν, pero dividida por sus grados de libertad.

Los grados de libertad de la T_ν coinciden con los de la χ^2_ν que aparece en su definición.

Visualmente la T_ν tiene una densidad tipo "campana" como la normal, centrada en cero, y con mayor desvío(para $\nu \geq 3$), pero con otra expresión funcional.

Además cuando $\nu \longrightarrow \infty$, la $T_\nu \longrightarrow N(0;1)$. Esto quiere decir que, por ejemplo si $\nu = 7$, la T_7 es diferente a la $N(0;1)$, sin embargo si $\nu \geq 30$, la $T\nu$ es practicamente muy parecida a la $N(0;1)$. Por eso en los problemas, cuando aparece una $T\nu$ con $\nu \geq 30$, se la suele reemplazar por una $N(0;1)$.

8. IC para μ en una población Normal de σ desconocida

$$\boxed{N(\mu;\sigma)}$$
$$X_1 X_2 \cdot\cdot X_n$$

1) Partimos de $H = X_1 + X_2 + \cdot\cdot +X_n$ que es suficiente respecto de μ. Pero por comodidad se usará $\overline{X} = \frac{H}{n}$ que es también suficiente, y de paso es un estimador de μ.

2) Distribución: $\overline{X} \sim N(\mu; \frac{\sigma}{\sqrt{n}})$

3) Estadístico pivotal: estandarizando resulta $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0;1)$ pero este NO es un estadístico pivotal. En efecto, debería depender solo de la muestra y del parámetro μ, pero depende también de σ que es desconocido.

Una salida "ingenua" a esta dificultad sería, con la muestra estimar σ con el estimador que conocemos

$$S = \sqrt{\frac{(X_1 - \overline{X})^2 + \cdot\cdot +(X_n - \overline{X})^2}{n-1}}$$

y lo reemplazamos en el estadístico anterior, quedando

$$Z^* = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

Pero el problema es que este estadístico, si bién depende solo de la muestra y el parámetro μ, no conocemos su distribución. Para que sea $N(0;1)$, en el denominador debe figurar la σ conocida, NO la v.a. S que es un estimador de σ.

Pero dividiendo por $\sigma/\sqrt{n}$ el numerador y denominador queda

$$Z^* = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\frac{S}{\sigma}} = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{S^2}{\sigma^2}}}$$

Notar que en el numerador figura $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = Z$, que es $N(0;1)$. Además dentro del radical del denominador

$$\frac{S^2}{\sigma^2} = \frac{\frac{(n-1)S^2}{\sigma^2}}{n-1} = \frac{U}{n-1} \text{ donde } U \sim \chi^2_{n-1}$$

Luego(apelando a la independencia entre $\overline{X}$ y S^2), y según la definición de la T de Student

$$Z^* = \frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{Z}{\sqrt{\frac{U}{n-1}}} \sim T_{n-1}$$

En definitiva el estadístico pivotal deseado es

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim T_{n-1}$$

Notar que si conociesemos σ, el estadístico pivotal sería $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim Z$, pero al desconocer σ y tener que estimarlo con la muestra, el estadístico pivotal es $\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim T_{n-1}$. Y el costo por desconocer σ se reflejará en tener que utilizar para los IC una T_{n-1} en lugar de una Z. Y como la T_{n-1} tiene una densidad con mas desvío que la normal, los intervalos obtenidos serán mas anchos. Lo que sigue para obtener el intervalo es como siempre.

4) Buscamos los fractiles $t_{n-1,\frac{\alpha}{2}}$ y $t_{n-1,1-\frac{\alpha}{2}}$ de la T_{n-1} , luego será

$$P(t_{n-1,\frac{\alpha}{2}} \leq T_{n-1} \leq t_{n-1,1-\frac{\alpha}{2}}) = NC$$

y reemplazando

$$P(t_{n-1,\frac{\alpha}{2}} \leq \frac{\overline{X}-\mu}{S/\sqrt{n}} \leq t_{n-1,1-\frac{\alpha}{2}}) = NC$$

y luego de invertir queda(como la T_{n-1} es simétrica respecto de 0, $t_{n-1,\frac{\alpha}{2}} = -t_{n-1,1-\frac{\alpha}{2}}$)

$$IC_\mu \text{ al NC: } \left[\overline{X} \pm t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right]$$

9. IC en dos poblaciones normales independientes

Cuando se tienen dos poblaciones normales independientes, se tienen en general 4 parámetros(dos medias y dos desvíos). Pero suelen ser de interés, otros parámetros que se derivan de estos 4. Por ejemplo si las normales son $N(\mu_x;\sigma_x)$ y $N(\mu_y;\sigma_y)$ puede interesar como parámetro a:

- $\delta = \mu_x - \mu_y$: este parámetro es de utilidad si se quiere evaluar como es μ_x respecto de μ_y. Por ejemplo si $\delta = 0$ quiere decir que $\mu_x = \mu_y$, si $\delta > 0$ significa que $\mu_x > \mu_y$, etc.
- $\varphi^2 = \frac{\sigma_x^2}{\sigma_y^2}$: este se utilizará para comparar varianzas. Si $\varphi^2 = 1$ quiere decir que las varianzas de las dos poblaciones son iguales, y si $\varphi^2 > 1$ significa que $\sigma_x^2 > \sigma_y^2$, etc.

Para δ se analizarán tres casos según los desvíos de las poblaciones sean : conocidos ambos, desconocidos pero iguales, y desconocidos(sin suponer igualdad).

Para φ^2 se analizará un solo caso.

10. IC para $\delta = \mu_x - \mu_y$ (desvíos conocidos)

$$\boxed{N(\mu_x;\sigma_{x\circ})} \qquad\qquad \boxed{N(\mu_y;\sigma_{y\circ})}$$
$$X_1 X_2 \cdot\cdot X_{n_x} \qquad\qquad Y_1 Y_2 \cdot\cdot Y_{n_y}$$

Como estimador de δ tomamos $\widehat{\delta} = \overline{X} - \overline{Y}$ (es función de suficientes), y además como

$$E(\widehat{\delta}) = \mu_x - \mu_y = \delta \text{ y } Var(\widehat{\delta}) = \frac{\sigma_{x\circ}^2}{n_x} + \frac{\sigma_{y\circ}^2}{n_y}$$

resultará

$$\widehat{\delta} = \overline{X} - \overline{Y} \sim N(\delta; \sqrt{\frac{\sigma^2_{x\circ}}{n_x} + \frac{\sigma^2_{y\circ}}{n_y}})$$

y resulta el estadístico pivotal

$$\frac{(\overline{X} - \overline{Y}) - \delta}{\sqrt{\frac{\sigma^2_{x\circ}}{n_x} + \frac{\sigma^2_{y\circ}}{n_y}}} \sim Z$$

Tomando los fractiles $\pm z_{1-\frac{\alpha}{2}}$ y planteando

$$P(-z_{1-\frac{\alpha}{2}} \leq \frac{(\overline{X} - \overline{Y}) - \delta}{\sqrt{\frac{\sigma^2_{x\circ}}{n_x} + \frac{\sigma^2_{y\circ}}{n_y}}} \leq z_{1-\frac{\alpha}{2}}) = NC$$

Y luego de invertir resulta

$$IC_\delta \; al \; NC : \left[(\overline{X} - \overline{Y}) \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma^2_{x\circ}}{n_x} + \frac{\sigma^2_{y\circ}}{n_y}} \right]$$

11. IC para $\delta = \mu_x - \mu_y$ (desvíos desconocidos e iguales)

$$\underset{X_1 X_2 \cdot \cdot X_{n_x}}{\boxed{N(\mu_x; \sigma)}} \qquad\qquad \underset{Y_1 Y_2 \cdot \cdot Y_{n_y}}{\boxed{N(\mu_y; \sigma)}}$$

Como estimador de δ tomamos $\widehat{\delta} = \overline{X} - \overline{Y}$ (es función de suficientes), y además como

$$E(\widehat{\delta}) = \mu_x - \mu_y = \delta \;\; \text{y} \;\; Var(\widehat{\delta}) = \frac{\sigma^2}{n_x} + \frac{\sigma^2}{n_y} = \sigma^2(\frac{1}{n_x} + \frac{1}{n_y})$$

resultará

$$\widehat{\delta} = \overline{X} - \overline{Y} \sim N(\delta; \sigma\sqrt{\frac{1}{n_x} + \frac{1}{n_y}})$$

luego

$$\frac{(\overline{X} - \overline{Y}) - \delta}{\sigma\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim Z$$

Pero este no un estadístico pivotal pues σ es desconocido. Con argumentos similares al utilizado cuando interesaba como parámetro μ en una normal con σ desconocido, buscamos un estimador de σ^2. El mejor es el estimador en pool. Como $S^2_x \xrightarrow{I} \sigma^2$ y $S^2_y \xrightarrow{I} \sigma^2$, luego como los pesos del estimador son proporcionales a los grados de libertad de cada estimador de varianza, resulta

$$S^2_{pool} = \frac{(n_x - 1)S^2_x + (n_y - 1)S^2_y}{n_x + n_y - 2} \;\; \text{y resultando también} \;\; \frac{(n_x + n_y - 2)S^2_{pool}}{\sigma^2} \sim \chi^2_{n_x+n_y-2}$$

Omitiendo detalles, si se reemplaza este estimador en el estadístico anterior, resultará

$$\frac{(\overline{X} - \overline{Y}) - \delta}{S_{pool}\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim T_{n_x+n_y-2}$$

Este si es pivotal respecto del parámetro δ. Buscando los fractiles $\pm t_{n_x+n_y-2,1-\frac{\alpha}{2}}$ y planteando

$$P(-t_{n_x+n_y-2,1-\frac{\alpha}{2}} \leq \frac{(\overline{X}-\overline{Y})-\delta}{S_{pool}\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}} \leq +t_{n_x+n_y-2,1-\frac{\alpha}{2}}) = NC$$

y luego de invertir, resulta

$$IC_\delta \ al \ NC: \left[(\overline{X}-\overline{Y}) \pm t_{n_x+n_y-2,1-\frac{\alpha}{2}} S_{pool}\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}\right]$$

12. IC **para** $\delta = \mu_x - \mu_y$ **(desvíos desconocidos)**

$$\frac{\boxed{N(\mu_x;\sigma_x)}}{X_1X_2\cdots X_{n_x}} \qquad\qquad \frac{\boxed{N(\mu_y;\sigma_y)}}{Y_1Y_2\cdots Y_{n_y}}$$

Como estimador de δ tomamos $\widehat{\delta} = \overline{X} - \overline{Y}$ (es función de suficientes), y además como

$$E(\widehat{\delta}) = \mu_x - \mu_y = \delta \ \text{ y } \ Var(\widehat{\delta}) = \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}$$

resultará

$$\widehat{\delta} = \overline{X} - \overline{Y} \sim N(\delta; \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}})$$

luego

$$\frac{(\overline{X}-\overline{Y})-\delta}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} \sim Z$$

Nuevamente este no un estadístico pivotal pues tanto σ_x^2 como σ_x^2 son desconocidos. Quizás podrámos usar un recurso similar al del ejemplo anterior, es decir estimarlos. Estimariamos σ_x^2 con S_x^2 y σ_y^2 con S_y^2. Entonces el estadístico dependería solo de la muestra y del δ desconocido, como necesitamos. Sin embargo no hay teoría que demuestre cuál es la distribución de este estadístico(no es ni $N(0;1)$ ni T de Student). Sin embargo, sí se puede demostrar, que **aproximadamente** se distribuye como una T_ν, con unos grados de libertad ν que se calculan mediante la fórmula de Welch

$$\nu_{Welch} = \frac{\left(\frac{S_x^2}{n_x}+\frac{S_y^2}{n_y}\right)^2}{\frac{S_x^4}{n_x^2(n_x-1)}+\frac{S_y^4}{n_y^2(n_y-1)}}$$

Remark 13. *En el denominador de esta expresión, $n_x - 1$ y $n_y - 1$ son los grados de libertad de los estimadores de varianza S_x^2 y S_y^2. Como S_x^2 se obtuvo* ***solo*** *con la muestra $X_1X_2\cdots X_{n_x}$, tiene $n_x - 1$ grados de libertad, idem para S_y^2. Pero, si por ejemplo en otro problema, la varianza σ_x^2 se estima haciendo un pool con* ***otra*** *muestra, obteniendo S_{poolX}^2 , en lugar de $n_x - 1$ debemos usar en la fórmula los grados de libertad de este pool, o sea ν_{poolX}. Idem si pasa lo mismo con σ_y^2.*

Entonces resultará aproximadamente

$$\frac{(\overline{X}-\overline{Y})-\delta}{\sqrt{\frac{S_x^2}{n_x}+\frac{S_y^2}{n_y}}} \sim T_{\nu_{Welch}}$$

Buscando los fractiles $\pm t_{\nu_{Welch},1-\frac{\alpha}{2}}$ y planteando

$$P(-t_{\nu_{Welch},1-\frac{\alpha}{2}} \leq \frac{(\overline{X}-\overline{Y})-\delta}{\sqrt{\frac{S_x^2}{n_x}+\frac{S_y^2}{n_y}}} \leq +t_{\nu_{Welch},1-\frac{\alpha}{2}}) = NC$$

y luego de invertir, resulta

$$IC_\delta \ al \ NC : \left[(\overline{X}-\overline{Y}) \pm t_{\nu_{Welch},1-\frac{\alpha}{2}} \sqrt{\frac{S_x^2}{n_x}+\frac{S_y^2}{n_y}}\right]$$

12.0.1. ***Distribución F de Fisher***. A continuación se presenta una v.a. que aparece, cuando en poblaciones normales interesa como parámetro el cociente entre dos varianzas; y también en otros problemas en que interviene el cociente entre dos v.a. Gamma. Se define así:

Definición de F de Fisher con ν_1 y ν_2 grados de libertad

$$\boxed{\begin{array}{l} U_x \sim \chi^2_{\nu_x} \\ U_y \sim \chi^2_{\nu_y} \end{array} \quad Ind \rightrightarrows \boxed{F = \frac{U_x/\nu_x}{U_y/\nu_y}} \longrightarrow F \sim F_{\nu_x,\nu_y}}$$

13. IC para $\varphi^2 = \frac{\sigma_x^2}{\sigma_y^2}$

$$\begin{array}{c} \boxed{N(\mu_x;\sigma_x)} \\ X_1 X_2 \cdot \cdot X_{n_x} \end{array} \qquad\qquad \begin{array}{c} \boxed{N(\mu_y;\sigma_y)} \\ Y_1 Y_2 \cdot \cdot Y_{n_y} \end{array}$$

Como estadísticos suficientes respecto de σ_x^2 y σ_y^2 usaremos S_x^2 y S_y^2. Además llamemos U_x y U_y a los correspondientes estadísticos pivotales para σ_x^2 y σ_y^2, o sea

$$U_x = \frac{(n_x-1)S_x^2}{\sigma_x^2} \sim \chi^2_{n_x-1} \quad \text{y} \quad U_y = \frac{(n_y-1)S_y^2}{\sigma_y^2} \sim \chi^2_{n_y-1}$$

Usando la definición de la F de Fisher

$$\frac{U_x/(n_x-1)}{U_y/(n_y-1)} = \frac{\frac{(n_x-1)S_x^2}{\sigma_x^2}/(n_x-1)}{\frac{(n_y-1)S_y^2}{\sigma_y^2}/(n_y-1)} = \frac{\frac{S_x^2}{\sigma_x^2}}{\frac{S_y^2}{\sigma_y^2}} = \frac{S_x^2/S_y^2}{\sigma_x^2/\sigma_y^2} = \frac{S_x^2/S_y^2}{\varphi^2} \sim F_{n_x-1,n_y-1}$$

O sea

$$\frac{S_x^2/S_y^2}{\varphi^2} \sim F_{n_x-1,n_y-1}$$

es un estadístico pivotal respecto del parámetro $\varphi^2 = \frac{\sigma_x^2}{\sigma_y^2}$. Buscando los fractiles de la F_{n_x-1,n_y-1} resultará

$$P(f_{n_x-1,n_y-1,\frac{\alpha}{2}} \leq F_{n_x-1,n_y-1} \leq f_{n_x-1,n_y-1,1-\frac{\alpha}{2}}) = NC$$

luego reemplazando

$$P(f_{n_x-1,n_y-1,\frac{\alpha}{2}} \leq \frac{S_x^2/S_y^2}{\varphi^2} \leq f_{n_x-1,n_y-1,1-\frac{\alpha}{2}}) = NC$$

e invirtiendo resulta

$$IC_{\varphi^2} \text{ al NC} : \left[\frac{S_x^2/S_y^2}{f_{n_x-1,n_y-1,1-\frac{\alpha}{2}}}; \frac{S_x^2/S_y^2}{f_{n_x-1,n_y-1,\frac{\alpha}{2}}}\right]$$

14. IC para $\theta = \frac{\beta_x}{\beta_y}$en dos poblaciones gamma independientes

$$\begin{array}{c}\boxed{G(1;\beta_x)}\\ X_1X_2\cdot\cdot X_{n_x}\end{array} \qquad\qquad \begin{array}{c}\boxed{G(1;\beta_y)}\\ Y_1Y_2\cdot\cdot Y_{n_y}\end{array}$$

Utilizando los correspondientes estadísticos suficientes para β_x y β_y

$$H_x = X_1 + X_2 + \cdot\cdot + X_{n_x} \sim G(n_x;\beta_x) \quad \text{y } H_y = Y_1 + Y_2 + \cdot\cdot + Y_{n_y} \sim G(n_y;\beta_y)$$

recordando que cuando a una v.a. gamma se la multiplica por una constante, el respectivo beta queda dividido por esta constante(en este caso la constante utilizada es 2β)

$$2\beta_x H_x \sim G(n_x;\frac{\beta_x}{2\beta_x}) \quad \text{y } 2\beta_y H_y \sim G(n_y;\frac{\beta_y}{2\beta_y})$$

o sea

$$2\beta_x H_x \sim G(\frac{2n_x}{2};\frac{1}{2}) \quad \text{y } 2\beta_y H_y \sim G(\frac{2n_y}{2};\frac{1}{2})$$

quedando

$$2\beta_x H_x \sim \chi^2_{2n_x} \quad \text{y } 2\beta_y H_y \sim \chi^2_{2n_y}$$

otro resultado anterior decía: cuando a una v.a. que es $\chi^2_{\nu_x}$ se la divide por sus g.l.; y a otra independiente que es $\chi^2_{\nu_y}$ se la divide también por sus g.l., y luego se hace el cociente, se obtiene una F_{ν_x,ν_y} o sea

$$\frac{\frac{2\beta_x H_x}{2n_x}}{\frac{2\beta_y H_y}{2n_y}} \sim F_{2n_x,2n_y}$$

como $\frac{H_x}{n_x} = \overline{X}$ y $\frac{H_y}{n_y} = \overline{Y}$, y simplificando

$$\frac{\beta_x}{\beta_y}\frac{\overline{X}}{\overline{Y}} \sim F_{2n_x,2n_y} \tag{14.1}$$

llamando $\theta = \frac{\beta_x}{\beta_y}$ queda

$$\theta\frac{\overline{X}}{\overline{Y}} \sim F_{2n_x,2n_y}$$

Como este es un estadístico pivotal para θ resulta

$$P(f_{2n_x,2n_y,\frac{\alpha}{2}} \le F_{2n_x,2n_y} \le f_{2n_x,2n_y,1-\frac{\alpha}{2}}) = NC \quad \text{y} \quad P(f_{2n_x,2n_y,\frac{\alpha}{2}} \le \theta\frac{\overline{X}}{\overline{Y}} \le f_{2n_x,2n_y,1-\frac{\alpha}{2}}) = NC$$

o sea

$$IC_\theta \text{ al NC}: \left[\frac{\overline{Y}}{\overline{X}}f_{2n_x,2n_y,\frac{\alpha}{2}};\frac{\overline{Y}}{\overline{X}}f_{2n_x,2n_y,1-\frac{\alpha}{2}}\right]$$

REMARK 14. *A veces lo que interesa como parámetro es el cociente entre las medias de las dos exponenciales. Como* $\mu_x = \frac{1}{\beta_x}$ *y* $\mu_y = \frac{1}{\beta_y}$*, luego* $\tau = \frac{\mu_x}{\mu_y} = \frac{\beta_y}{\beta_x}$*, luego reemplazando en (14.1) el estadístico pivotal en este caso sería*

$$\frac{1}{\tau}\frac{\overline{X}}{\overline{Y}} \sim F_{2n_x,2n_y}$$

REMARK 15. *Un enfoque parecido se puede aplicar si las poblaciones son* $G(\alpha_x;\beta_x)$ *y* $G(\alpha_y;\beta_x)$ *independientes, con* α_x *y* α_y *conocidos y números naturales, cuando el parámetro de interés es el cociente entre los betas.*

15. Un ejemplo con poblaciones normales

La finalidad del siguiente ejemplo es integrar diversos conceptos, estudiados por separado, en poblaciones normales.

EXAMPLE 11. *Considere artículos de peso $N(\mu;\sigma)$ que se venden en cajas de peso fijo θ. Se tienen las muestras:*

▶ de cajas con 2 artículos:$X_1X_2...X_4$ con $\overline{X} = 36, S_x = 2.9$
▶ de cajas con 4 artículos:$Y_1Y_2...Y_6$ con $\overline{Y} = 54, S_y = 4.1$
▶ de artículos solamente:$A_1A_2..A_4$ con $\overline{A} = 10, S_a = 2$
Hallar los **mejores** IC al 90% para: σ, μ y θ

Solución
Primero, esto es importante, hay que averiguar las poblaciones de las que provienen las muestras.
Si $A \sim N(\mu;\sigma)$, y X es el peso de una caja con 2 artículos, entonces

$$X = \theta + A_1 + A_2 \sim N(\theta + 2\mu; 2\sigma^2)$$

Idem si Y es el peso de una caja con 4 artículos, entonces

$$Y = \theta + A_1 + A_2 + A_3 + A_4 \sim N(\theta + 4\mu; 4\sigma^2)$$

Luego se tienen las poblaciones y muestras

$\boxed{N(\theta + 2\mu; 2\sigma^2)}$	$\boxed{N(\theta + 4\mu; 4\sigma^2)}$	$\boxed{N(\mu;\sigma^2)}$
$X_1X_2\cdot\cdot X_4$	$Y_1Y_2\cdot\cdot Y_6$	$A_1A_2\cdot\cdot A_4$

IC$_\sigma$
Para estimar σ^2 disponemos de los estimadores $S_x^2/2$, $S_y^2/4$ y de S_a^2. El mejor estimador de σ^2 con las muestras disponibles es(utilizando pesos proporcionales a los grados de libertad de cada estimador de σ^2)

$$S_{pool}^2 = \frac{(4-1)\frac{S_x^2}{2} + (6-1)\frac{S_y^2}{4} + (4-1)S_a^2}{(4-1)+(6-1)+(4-1)} \text{ que tiene 11 gl (haciendo cuentas } S_{pool}^2 = 4.15,\ S_{pool} = 2.037)$$

luego

$$\frac{11S_{pool}^2}{\sigma^2} \sim \chi_{11}^2$$

y el correspondiente intervalo será

$$IC_{\sigma^2} \text{ al } 95\% : \left[\frac{11S_{pool}^2}{\chi_{11,0.95}^2}; \frac{11S_{pool}^2}{\chi_{11,0.05}^2}\right] \quad \text{o sea } IC_\sigma \text{ al } 95\% : \left[\sqrt{\frac{11S_{pool}^2}{\chi_{11,0.95}^2}}; \sqrt{\frac{11S_{pool}^2}{\chi_{11,0.05}^2}}\right]$$

Como(pqrs) $\chi_{11,0.95}^2 = 19.68$ y $\chi_{11,0.05}^2 = 4.57$ reemplazando valores queda

$$IC_\sigma \text{ al } 90\% : [1.52; 3.16]$$

Por supuesto, para obtener IC_σ podríamos haber utilizado solo la tercera muestra, con S_a^2 que tiene solo $3gl$. Pero el intervalo no sería el **mejor**, sería mas ancho, ya que estaríamos desperdiciando la información respecto de σ^2 que contienen las otras dos muestras.
IC$_\mu$
Necesitamos un estimador de μ. Hay varias opciones.

La más simple sería $\overline{A}$. Como $E(\overline{A}) = \mu$ y $Var(\overline{A}) = \frac{\sigma^2}{4}$ estandarizando

$$\frac{\overline{A} - \mu}{\sigma/\sqrt{4}} = Z \sim N(0;1)$$

No es pivotal ya que σ es desconocido. Si lo reemplazamos por un estimador, el mejor sin duda es $S^{\cdot}_{pool}$ y entonces quedaría

$$\frac{\overline{A} - \mu}{S^{\cdot}_{pool}/\sqrt{4}} = T_{11} \text{ con el } IC_\mu \text{ al } 90\% : \left[\overline{A} \pm t_{11,0.95}\frac{S^{\cdot}_{pool}}{\sqrt{4}}\right] = \left[10 \pm 1.796\frac{2.037}{\sqrt{4}}\right] = [8.17; 11.83]$$

Pero con esta solución, hemos desperdiciado la información respecto de μ que también tienen las otras dos muestras. Por ejemplo $\overline{X} \xrightarrow{I} \theta + 2\mu$ y $\overline{Y} \xrightarrow{I} \theta + 4\mu$, que no son estimadores de μ, sin embargo $\overline{Y} - \overline{X} \xrightarrow{I} 2\mu$, y entonces $\frac{\overline{Y}-\overline{X}}{2} \xrightarrow{I} \mu$, sí es un estimador insesgado de μ.

Podríamos con este estimador construir un IC_μ como recién hicimos con $\overline{A}$.

Pero como $\frac{\overline{Y}-\overline{X}}{2}$ depende solo de las dos primeras muestras, $\overline{A}$ de la tercera, y son independientes, entonces la idea es obtener el estimador en pool de μ con estos dos.

Tenemos

$$\frac{\overline{Y} - \overline{X}}{2} \xrightarrow{I} \mu \quad \text{y} \quad \overline{A} \xrightarrow{I} \mu$$

con

$$Var(\frac{\overline{Y} - \overline{X}}{2}) = \frac{\frac{4\sigma^2}{6} + \frac{2\sigma^2}{4}}{2^2} = \frac{7}{24}\sigma^2 \quad \text{y} \quad Var(\overline{A}) = \frac{\sigma^2}{4} \tag{15.1}$$

Luego(simplificando el σ^2)

$$\widehat{\mu}_{pool} = \frac{\frac{24}{7}(\frac{\overline{Y}-\overline{X}}{2}) + 4\overline{A}}{\frac{24}{7} + 4} = \frac{6}{13}(\frac{\overline{Y} - \overline{X}}{2}) + \frac{7}{13}\overline{A}$$

Como (utilizando las varianzas calculadas en 15.1)

$$E(\widehat{\mu}_{pool}) = \mu \quad \text{y} \quad Var(\widehat{\mu}_{pool} = \left(\frac{6}{13}\right)^2 (\frac{7}{24}\sigma^2) + \left(\frac{7}{13}\right)^2 \frac{\sigma^2}{4} = \frac{7}{52}\sigma^2$$

Luego

$$\frac{\widehat{\mu}_{pool} - \mu}{\sqrt{\frac{7}{52}}\sigma} = Z \sim N(0;1)$$

Para llegar a la T usamos el $S^{\cdot}_{pool}$ anterior, y reemplazando la expresión de $\widehat{\mu}_{pool}$ queda

$$\frac{(\frac{6}{13}(\frac{\overline{Y}-\overline{X}}{2}) + \frac{7}{13}\overline{A}) - \mu}{\sqrt{\frac{7}{52}}S^{\cdot}_{pool}} = T_{11}$$

y luego

$$IC_\mu \text{ al } 95\% \;: \left[(\frac{6}{13}(\frac{\overline{Y} - \overline{X}}{2}) + \frac{7}{13}\overline{A}) \pm t_{11,0.95}\sqrt{\frac{7}{52}}S^{\cdot}_{pool}\right]$$

reemplazando valores queda

$$IC_\mu \text{ al } 95\% : [8.20; 10.88]$$

de amplitud menor que el obtenido con solo $\overline{A}$(antes 3.66, ahora 2.68), ya que el estimador de μ que utilizamos es superior.

IC$_\theta$

Aquí como $\overline{X} \xrightarrow{I} \theta + 2\mu$ y $\overline{Y} \xrightarrow{I} \theta + 4\mu$, que no son estimadores de θ, sin embargo $2\overline{X} - \overline{Y} \xrightarrow{I} \theta$, y entonces $\widehat{\Theta} = 2\overline{X} - \overline{Y}$, sí es un estimador insesgado de θ. Podriamos construir el IC basandonos en este estimador. Pero claro, no hemos utilizado la tercer muestra. Se propondrá un método mas general para evaluar si esto es posible.

En principio tenemos $\overline{X} \xrightarrow{I} \theta+2\mu$, $\overline{Y} \xrightarrow{I} \theta+4\mu$ y $\overline{A} \xrightarrow{I} \mu$, todos estimadores independientes. Se propone ahora

$$\widehat{\Theta}_{pool} = \alpha\overline{X} + \beta\overline{Y} + \gamma\overline{A}$$

Donde α, β y γ son constantes a determinar. No se trata del concepto de estimador en pool visto antes, donde los estimadores estimaban el mismo parámetro, sinó que ahora estiman combinaciones lineales diferentes de los parámetros. Como queremos que

$$E(\widehat{\Theta}_{pool}) = \theta$$

deberá ser

$$E(\widehat{\Theta}_{pool}) = E(\alpha\overline{X} + \beta\overline{Y} + \gamma\overline{A}) = \alpha(\theta + 2\mu) + \beta(\theta + 4\mu) + \gamma\mu = \theta$$

Y agrupando

$$(\alpha + \beta)\theta + (2\alpha + 4\beta + \gamma)\mu = \theta$$

Para que esta igualdad se cumpla $\forall\theta$ y $\forall\mu$ deberá ser

$$\left\{ \begin{array}{c} \alpha + \beta = 1 \\ 2\alpha + 4\beta + \gamma = 0 \end{array} \right. \quad \text{o sea} \quad \left\{ \begin{array}{c} \alpha + \beta = 1 \\ \alpha + 2\beta = -\frac{\gamma}{2} \end{array} \right.$$

Resolviendo el sistema en α y β resulta

$$\alpha = \frac{\gamma}{2} + 2 \quad \text{y} \quad \beta = -(\frac{\gamma}{2} + 1)$$

Luego el estimador insesgado de θ sería

$$\widehat{\Theta}_{pool} = (\frac{\gamma}{2} + 2)\overline{X} - (\frac{\gamma}{2} + 1)\overline{Y} + \gamma\overline{A} \tag{15.2}$$

Este estimador, para cualquier valor de γ, cumple siempre $E(\widehat{\Theta}_{pool}) = \theta$. Como γ lo desconocemos, elegiremos este valor de manera que

$$Var(\widehat{\Theta}_{pool}) : \text{ sea mínima}$$

Pero

$$\begin{aligned} Var(\widehat{\Theta}_{pool}) &= (\frac{\gamma}{2} + 2)^2 Var(\overline{X}) + (\frac{\gamma}{2} + 1)^2 Var(\overline{Y}) + \gamma^2 Var(\overline{A}) \\ &= (\frac{\gamma}{2} + 2)^2 \frac{2\sigma^2}{4} + (\frac{\gamma}{2} + 1)^2 \frac{4\sigma^2}{6} + \gamma^2 \frac{\sigma^2}{4} \\ &= \left\{ (\frac{\gamma}{2} + 2)^2 \frac{1}{2} + (\frac{\gamma}{2} + 1)^2 \frac{2}{3} + \gamma^2 \frac{1}{4} \right\} \sigma^2 \end{aligned} \tag{15.3}$$

Derivando la llave respecto de γ e igualando a cero

$$2(\frac{\gamma}{2} + 2)(\frac{1}{2})\frac{1}{2} + 2(\frac{\gamma}{2} + 1)(\frac{1}{2})\frac{2}{3} + 2\gamma\frac{1}{4} = 0$$

De aquí resulta $\gamma = -\frac{20}{13}$ para lograr varianza mínima. Reemplazando en (15.2) queda el estimador óptimo deseado

$$\widehat{\Theta}_{pool} = \frac{16}{13}\overline{X} - \frac{3}{13}\overline{Y} - \frac{20}{13}\overline{A}$$

Obviamente $E(\widehat{\Theta}_{pool}) = \theta$, pero veamos su varianza(ver 15.3)

$$Var(\widehat{\Theta}_{pool}) = \left(\frac{16}{13}\right)^2 \frac{2\sigma^2}{4} + \left(\frac{3}{13}\right)^2 \frac{4\sigma^2}{6} + \left(\frac{20}{13}\right)^2 \frac{\sigma^2}{4} = \frac{18}{13}\sigma^2$$

luego

$$\widehat{\Theta}_{pool} \sim N(\theta; \frac{18}{13}\sigma^2)$$

y también

$$\frac{\widehat{\Theta}_{pool} - \theta}{\sqrt{\frac{18}{13}}\sigma} = Z \sim N(0;1)$$

y utilizando $S^{\cdot}_{pool}$

$$\frac{\widehat{\Theta}_{pool} - \theta}{\sqrt{\frac{18}{13}}S^{\cdot}_{pool}} \sim T_{11}$$

luego

$$IC_\theta \text{ al } 95\% : \left[(\frac{16}{13}\overline{X} - \frac{3}{13}\overline{Y} - \frac{20}{13}\overline{A}) \pm t_{11,0.95}\sqrt{\frac{18}{13}}S^{\cdot}_{pool} \right]$$

Aquí $\frac{16}{13}\overline{X} - \frac{3}{13}\overline{Y} - \frac{20}{13}\overline{A} = \widehat{\Theta}_{pool} = 16.46$ es la estimación del peso fijo, θ de las cajas, y el intervalo resulta

$$\left[16.46 \pm 1.796\sqrt{\frac{18}{13}}2.037 \right] = [12.16; 20.76]$$

16. Incorporación del error de un instrumento

Supongamos que ω es el valor de cierta cantidad que queremos medir. Por ejemplo el voltaje entre dos puntos de un circuito, o el peso de un artículo, o la concentración de glucosa en una muestra de sangre. El valor de ω lo desconocemos, y para averiguarlo usaremos un instrumento de medición.

En el primer ejemplo el instrumento sería un voltímetro, en el segundo una balanza, y en el tercero realizar un análisis bioquímico.

Como los instrumentos tienen errores, el valor medido seguramente no coincidirá con ω que es el valor que queremos averiguar.

Si se designa X al valor que proporciona el instrumento al efectuar una medición, supondremos que

$$X = \omega + E \quad \text{con } E \sim N(0;\sigma)$$

O sea, el valor medido(X) es el valor verdadero que buscamos(ω), más un error(E) aleatorio que le agrega el instrumento de medición. De aquí surge que

$$X \sim N(\omega;\sigma)$$

En estos casos, para lograr mayor precisión en el conocimiento de ω, se suelen efectuar n mediciones con el instrumento, teniendo entonces la muestra de mediciones $X_1X_2\cdot\cdot X_n$ y resultará

$$\frac{\boxed{N(\omega;\sigma)}}{X_1X_2\cdot\cdot X_n}$$

Aquí seguramente interesará un IC_ω al NC, que lo obtendremos con la T de Student, como ya fué visto

$$IC_\omega \text{ al NC: } \left[\overline{X} \pm t_{n-1,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right]$$

Por supuesto, cuanto mayor sea n, mas angosto será el intervalo.

REMARK 16. *Lo usual es suponer que $E \sim N(0;\sigma)$, o sea que la media del error del instrumento es cero. Si esto es así se dice que el instrumento esta calibrado. Cuando esto no ocurre supondremos que $E \sim N(\tau;\sigma)$, donde τ es el sesgo del instrumento. En este caso el valor medido será también $X = \omega + E$, y resultará*

$$X \sim N(\omega+\tau;\sigma)$$

Esto quiere decir que las mediciones que hagamos(X) estarán alrededor de $\omega+\tau$, y no de ω que es el valor que nos interesa. Si $\tau > 0$ obtendremos mediciones en general un poco arriba de ω, y al revés si $\tau < 0$. Habría que calibrar el instrumento!

EXAMPLE 12. *La concentración de glucosa en sangre(G) de ciertos animales responde a una $N(\mu_g;\sigma_g)$ mg/dl. Se tomaron al azar 4 animales, se les sacó sangre, obteniendo 4 tubos de ensayo con la sangre de cada animal, se tendrá entonces*

$$\frac{\boxed{N(\mu_g;\sigma_g)}}{G_1G_2\cdot\cdot G_4} \tag{16.1}$$

Aquí los G_i representan la concentración de glucosa en cada tubo. Se pide:

- **(a):** IC al 90% para la concentracion real de glucosa de **cada animal**(o sea, la presente en cada tubo de ensayo)
- **(b):** IC al 90% para μ_g (la concentración media de glucosa de este **tipo de animales**)

Notar que en (16.1) los $G_1G_2\cdot\cdot G_4$, los desconocemos, así que no podemos utilizarlos como muestra para dar respuesta a lo que pide el problema.

Para la parte **(a)**, designaremos g_1, g_2, g_3, g_4 a los valores reales de glucosa en cada tubo(o sea el valor que toma cada G_i en cada tubo). Estos valores son fijos pero desconocidos, y el problema pide justamente hallar intervalos de confianza para cada g_i. Antes tendremos que usar un instrumento(un glucómetro) para determinar la concentración de glucosa en cada tubo. Supondremos que el error de este instrumento es $E \sim N(0;\sigma)$ ya que está calibrado. Y para lograr mayor precisión se harán 3 determinaciones de glucosa en cada tubo. Por ejemplo para el primer tubo, si g_1 es la concentración real de glucosa en este tubo, el primer valor medido por el instrumento será $X_{11} = g_1 + E_{11} \sim N(g_1;\sigma)$, el segundo valor $X_{12} = g_1 + E_{12} \sim N(g_1;\sigma)$, y el tercero $X_{13} = g_1 + E_{13} \sim N(g_1;\sigma)$, o sea las tres mediciones en el primer tubo constituyen **la muestra obtenida** y se tendrá

$$\frac{\boxed{N(g_1;\sigma)}}{X_{11}X_{12}X_{13}} \tag{16.2}$$

Como esto lo hacemos para los 4 tubos, se tendrán(con los respectivos valores muestrales)

$$\begin{array}{cccc} \boxed{N(g_1;\sigma)} & \boxed{N(g_2;\sigma)} & \boxed{N(g_3;\sigma)} & \boxed{N(g_4;\sigma)} \\ X_{11}X_{12}X_{13} & X_{21}X_{22}X_{23} & X_{31}X_{32}X_{33} & X_{41}X_{42}X_{43} \\ \overline{X}_{1\cdot} = 80 \quad S_1 = 2 & \overline{X}_{2\cdot} = 90 \quad S_1 = 1 & \overline{X}_{3\cdot} = 87 \quad S_1 = 3 & \overline{X}_{4\cdot} = 83 \quad S_1 = 2 \end{array}$$

Se obtendrá solo el IC_{g_1} al 90%, para el primer tubo(los otros salen en forma similar).

Como $\overline{X}_{1\cdot} \xrightarrow{I} g_1$ entonces $\overline{X}_{1\cdot} \sim N(g_1; \frac{\sigma}{\sqrt{3}})$ y

$$\frac{\overline{X}_{1\cdot} - g_1}{\frac{\sigma}{\sqrt{3}}} = Z \sim N(0;1)$$

Como estimador de σ^2 usamos

$$S^2_{pool} = \frac{(3-1)S_1^2 + (3-1)S_2^2 + (3-1)S_3^2 + (3-1)S_4^2}{(3-1)+(3-1)+(3-1)+(3-1)} = 4.5 \text{ (o sea } S_{pool} = 2.121)$$

Como tiene 8 gl, resulta

$$\frac{\overline{X}_{1\cdot} - g_1}{\frac{S_{pool}}{\sqrt{3}}} \sim T_8$$

Y el intervalo deseado para la concentración de glucosa del primer animal será:

$$IC_{g_1} \text{ al } 90\%\text{:} \ \left[\overline{X}_{1\cdot} \pm t_{8,0.95}\frac{S_{pool}}{\sqrt{3}}\right] = \left[80 \pm 1.86\frac{2.121}{\sqrt{3}}\right] = [77.72; 82.28] \text{ (de long: 4.55)}$$

Notar que aquí se utilizaron 4 muestras de 4 poblaciones diferentes. Pero solo la primera muestra, a través de $\overline{X}_{1\cdot}$ tiene información respecto del parámetro g_1 de interés. Sin embargo se utilizaron las cuatro muestras para obtener el mejor estimador de σ^2, que es S^2_{pool} con 8 gl. De haber utilizado solo la primera muestra para estimar σ^2, usando S_1^2 con solo 2 gl, el intervalo final obtenido hubiese sido mas ancho(de long: 6.27).

Para la parte (**b**), el parámetro de interés es μ_g de (16.1), que es el valor alrededor del cual están los $G_1G_2 \cdot\cdot G_4$ que desconocemos. Aquí necesitamos pensar a las G_i como variables aleatorias, no como los valores fijos g_1, g_2, g_3, g_4 de la parte (**a**). La primera determinación que hacemos de G_1 con el instrumento es $X_{11} = G_1 + E_{11}$, la segunda $X_{12} = G_1 + E_{12}$ y la tercera $X_{13} = G_1 + E_{13}$ luego el promedio de estas tres es

$$\overline{X}_{1\cdot} = \frac{G_1 + E_{11} + G_1 + E_{12} + G_1 + E_{13}}{3} = G_1 + \frac{E_{11} + E_{12} + E_{13}}{3} \sim N(\mu_g; \sqrt{\sigma_g^2 + \frac{\sigma^2}{3}})$$

Analizando lo mismo con $\overline{X}_{2\cdot}$, $\overline{X}_{3\cdot}$ y $\overline{X}_{4\cdot}$ resulta que los cuatro promedios vienen de la misma normal, se tendrá entonces

$$\begin{array}{c} \boxed{N(\mu_g; \sqrt{\sigma_g^2 + \frac{\sigma^2}{3}})} \\ \overline{X}_{1\cdot}\overline{X}_{2\cdot}\overline{X}_{3\cdot}\overline{X}_{4\cdot} \end{array}$$

Llamando $\sqrt{\sigma_g^2 + \frac{\sigma^2}{3}} = \sigma_{\overline{x}}$ queda

$$\begin{array}{c} \boxed{N(\mu_g; \sigma_{\overline{x}})} \\ \overline{X}_{1\cdot}\overline{X}_{2\cdot}\overline{X}_{3\cdot}\overline{X}_{4\cdot} \end{array}$$

Ahora sí: $\overline{\overline{X}} = \frac{\overline{X}_{1\cdot}+\overline{X}_{2\cdot}+\overline{X}_{3\cdot}+\overline{X}_{4\cdot}}{4} \xrightarrow{I} \mu_g$ con $\overline{\overline{X}} \sim N(\mu_g; \frac{\sigma_{\overline{x}}}{\sqrt{4}})$ luego

$$\frac{\overline{\overline{X}} - \mu_g}{\frac{\sigma_{\overline{x}}}{\sqrt{4}}} = Z \sim N(0;1)$$

Como estimador de $\sigma^2_{\overline{x}}$ usamos(teniendo en cuenta que $\overline{\overline{X}} = 85$)

$$S^2_{\overline{x}} = \frac{(\overline{X}_{1\cdot} - \overline{\overline{X}})^2 + (\overline{X}_{2\cdot} - \overline{\overline{X}})^2 + (\overline{X}_{3\cdot} - \overline{\overline{X}})^2 + (\overline{X}_{4\cdot} - \overline{\overline{X}})^2}{4-1} = 19.33 (\text{o sea } S_{\overline{x}} = 4.4)$$

Como tiene 3 gl, resulta

$$\frac{\overline{\overline{X}} - \mu_g}{\frac{S_{\overline{x}}}{\sqrt{4}}} \sim T_3$$

Y el intervalo deseado para la concentración media glucosa de ese tipo de animales será:

$$IC_{\mu_g} \text{ al } 90\%: \left[\overline{\overline{X}} \pm t_{3,0.95}\frac{S_{\overline{x}}}{\sqrt{4}}\right] = \left[85 \pm 2.35\frac{4.4}{\sqrt{4}}\right] = [79.83; 90.17] \text{ (de long: 10.34)}$$

REMARK 17. *En este problema se hicieron en total 12 mediciones. Para la parte (**a**), se repartieron en 3 por tubo, para lograr mejores estimaciones de los g_i y por lo tanto IC_{g_i} mas angostos. La parte (**b**) se resolvió utilizando las mismas mediciones, con otra teoría. Sin embargo, y manteniendo las 12 mediciones, si el objetivo del problema es obtener **solo** el IC_{μ_g}, lo más apropiado hubiese sido tomar una muestra de 12 animales, con solo **una medición** por animal. En este caso, cada X_i sería $X_i = G_i + E_i \sim N(\mu_g; \sqrt{\sigma_g^2 + \sigma^2})$ teniendo(llamando σ_x a este desvío)*

$$\begin{array}{c}\boxed{N(\mu_g;\sigma_x)}\\ X_1X_2\cdot\cdot X_{12}\\ \overline{X}\ \ ,\ S_x\end{array} \quad \text{y como } \overline{X} \xrightarrow{I} \mu_g,\ S_x^2 \text{ con } 11gl, \text{ resulta } IC_{\mu_g} \text{ al } 90\%: \left[\overline{X} \pm t_{11,0.95}\frac{S_x}{\sqrt{12}}\right]$$

Con este diseño, el intervalo para μ_g sería mas angosto.

17. Intervalos NO exactos ∪ NO óptimos

Casi todos los intervalos analizados hasta ahora eran exactos ∩ óptimos.
Se partía de una población $f(x/\theta)$ y su correspondiente muestra $X_1X_2\cdot\cdot X_n$.

- Un requisito para construir el intervalo, era conocer la distribución del estadístico H(o un pivotal que surge de H)

$$H \sim f(h/\theta) \text{ o } V = v(H;\theta) \sim f(v)$$

  Y el intervalo obtenido es **exacto**, si se usa esta $f(h/\theta)$ o $f(v)$ conocida.
- Además para construirlo, por comodidad, en lugar de usar toda la muestra $X_1X_2\cdot\cdot X_n$, lo hacíamos con un estadístico suficiente respecto de θ

$$H = g(X_1X_2\cdot\cdot X_n) \text{ suficiente respecto de } \theta$$

  Entonces el intervalo obtenido es **óptimo**, ya es tan "bueno" como si lo hubiesemos obtenido con toda la muestra(esto es debido a la suficiencia, ya que H tiene la misma "información" respecto de θ que toda la muestra).

Por diversos motivos, uno o ambos de estos dos requisitos, a veces no se cumple.

El **primer incumplimiento** se presenta en general, cuando la densidad de H o V, o sea $f(h/\theta)$ o $f(v)$, se desconoce, o es complicada. En este caso se la reemplaza por otra que la aproxime. Entonces el intervalo obtenido será **NO exacto.**

REMARK 18. *Esto de reemplazar la densidad de H por otra que la aproxime, tiene un motivo: se recordará que en varios casos ya vistos, el estadístico suficiente es una suma, o sea $H = X_1 + X_2 + \cdot\cdot + X_n$. Supongamos que la exacta densidad de H es $f(h/\theta)$. Sin embargo, si n es "grande", y utilizando el* ***TCL****, la densidad de H será* ***aproximadamente*** *una $N(;)$ con media y sigma a determinar en cada caso. Si se utiliza esta normal como densidad de H, el intervalo será aproximado.*

REMARK 19. *El caso de reemplazar la densidad de V por una que la aproxime, ya se utilizó cuando se estudió el IC para $\delta = \mu_x - \mu_y$ en dos poblaciones normales independientes con desvíos desconocidos(se aproximó la densidad del estadístico pivotal, por una t de Student, usando la fórmula de Welch, cuando en realidad se desconocía su distribución).*

El **segundo incumplimiento** se presenta, si en lugar de construir el intervalo con un estadístico suficiente, se lo hace con otro

$$R = r(X_1 X_2 \cdot\cdot X_n) \textbf{ NO suficiente } \text{respecto de } \theta$$

Entonces el intervalo obtenido **NO será óptimo**, y esto en general se verá reflejado en un intervalo **mas ancho**. Por eso, si se opta por un R no suficiente, este estadístico deberá contener "bastante" información respecto del parámetro, como para que el intervalo obtenido no sea demasiado ancho.

REMARK 20. *Esto usualmente ocurre cuando, con la muestra disponible no se puede calcular el apropiado estadístico suficiente.*

Se analizarán dos ejemplos de intervalos no exactos, y dos de intervalos no óptimos.

EXAMPLE 13. ***Intervalo de confianza para** p **de una muestra de población*** *Bernouilli$(p) \approx B_i(1;p)$,* ***cuando** n **es grande***

$$\boxed{B_i(1;p)}$$
$$X_1 X_2 \cdot\cdot X_n$$

Aquí el estadístico suficiente es $H = X_1 + X_2 + \cdot\cdot + X_n \sim B_i(n;p)$. Pero en lugar de H trabajaremos con

$$\widehat{P} = \frac{H}{n} = \frac{X_1 + X_2 + \cdot\cdot + X_n}{n}$$

que también es suficiente, y además, es un estimador de p. Como en esta expresión aparece una suma, si n es grande, por el TCL resultará aproximadamente $\widehat{P} \sim N(;)$. Para obtener la media y sigma de esta normal, calculamos(utilizando media y sigma de la $H \sim B_i(n;p)$)

$$E(\widehat{P}) = E(\frac{H}{n}) = \frac{np}{n} = p \quad y \quad Var(\widehat{P}) = Var(\frac{H}{n}) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Luego resulta

$$\widehat{P} \approx N(p; \sqrt{\frac{p(1-p)}{n}})$$

y entonces estandarizando, y reemplazando $\widehat{P} = \frac{h}{n} = \widehat{p}$

$$\frac{\widehat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = Z \sim N(0;1)$$

utilizando los fractiles de la normal

$$P(-z_{1-\frac{\alpha}{2}} \leq \frac{\widehat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{1-\frac{\alpha}{2}}) = NC$$

Habría que invertir esta doble desigualdad. Esto es matemáticamente simple(aparece solo una ecuación cuadrática), pero da bastante trabajo. El intervalo obtenido sería

$$\frac{\frac{z^2_{1-\frac{\alpha}{2}}}{2n} + \left[\widehat{p} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right]}{1 + \frac{z^2_{1-\frac{\alpha}{2}}}{n}}$$

Bien entendido: hay que calcular los dos extremos con $\left[\widehat{p} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right]$*, luego a cada extremo se le suma* $\frac{z^2_{1-\frac{\alpha}{2}}}{2n}$*, y luego ambos extremos se dividen por* $1 + \frac{z^2_{1-\frac{\alpha}{2}}}{n}$. *Esto sería muy complicado. Sin embargo notar que como* n *es grande,* $\frac{z^2_{1-\frac{\alpha}{2}}}{2n} \simeq 0$ *y* $1 + \frac{z^2_{1-\frac{\alpha}{2}}}{n} \simeq 1$. *Utilizando esto queda finalmente*

$$IC_p \text{ al } NC: \left[\widehat{p} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right]$$

Una expresión muy simple, y razonable ya que el intervalo es: $\widehat{p}$ *(que es un estimador de p), mas/menos un plus, que depende del desvío del estimador(*$\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$*) y del fractil. Este sería un intervalo óptimo pero aproximado.*

REMARK 21. *El IC para p que vimos anteriormente, con el método del estadístico muestral era óptimo y exacto. Sin embargo tiene el inconveniente que se hace muy difícil evaluar tamaño de muestra. Con esta última expresión, aproximada, es mucho mas simple. Supongamos que interesa un* IC_p *con* $NC = 90\%$*, pero queremos averiguar que tamaño de muestra será necesario para que la semi-amplitud del intervalo sea por ejemplo 0.05. Entonces la ecuación a resolver sería(*$z_{0.95} = 1.645$*)*

$$1.645\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 0.05$$

de aquí hay que despejar n

$$n = \left(\frac{1.645}{0.05}\right)^2 \widehat{p}(1-\widehat{p})$$

Pero claro, como todavía no hemos tomado la muestra, no conocemos $\widehat{p}$. *Entonces esto no tendría solución! Estas situaciones son comunes cuando hay que averiguar tamaño de muestra. Suelen haber algunas cantidades cuyos valores se desconocen, y no se puede despejar el* n. *Lo que se hace es darle algún valor a* $\widehat{p}$. *Pero cual? En esta expresión aparece una función cuadrática del tipo* $y = x(1-x)$. *Es una parábola invertida con raíz en 0 y 1(representarla!). Se busca cual es el valor de* x *que da el máximo. Este valor es* $x = 0.5$. *O sea si usamos* $\widehat{p} = 0.5$*, resulta* $n = 271$.

Cualquier otro valor de $\widehat{p}$ que usemos, dará un n menor. De esta manera, tomar $n = 271$ es la opción mas conservadora. Después, cuando tomemos la muestra, y obtengamos el IC_p, notaremos que su semiamplitud será, como máximo 0.05.

EXAMPLE 14. ***Intervalo de confianza para λ de una muestra de población $P_o(\lambda)$, cuando n es grande***

$$\boxed{P_o(\lambda)}$$
$$X_1X_2 \cdot \cdot X_n$$

Aquí $H = X_1 + X_2 + \cdot \cdot +X_n \sim P_o(n\lambda)$ es suficiente. Igual que en el ejemplo anterior, usaremos $\widehat{\Lambda} = \frac{H}{n} = \frac{X_1+X_2+\cdot\cdot+X_n}{n}$ que también es suficiente, y de paso es un estimador de λ. Como en su expresión aparece una suma, si n es grande, por el TCL resultará

$$\widehat{\Lambda} \approx N(;)$$

y como

$$E(\widehat{\Lambda}) = E(\frac{H}{n}) = \frac{n\lambda}{n} = \lambda \quad y \quad Var(\widehat{\Lambda}) = Var(\frac{H}{n}) = \frac{n\lambda}{n^2} = \frac{\lambda}{n}$$

luego

$$\widehat{\Lambda} \approx N(\lambda; \sqrt{\frac{\lambda}{n}})$$

y entonces estandarizando

$$\frac{\widehat{\Lambda} - \lambda}{\sqrt{\frac{\lambda}{n}}} = Z \sim N(0;1)$$

utilizando los fractiles de la normal

$$P(-z_{1-\frac{\alpha}{2}} \leq \frac{\widehat{\Lambda} - \lambda}{\sqrt{\frac{\lambda}{n}}} \leq z_{1-\frac{\alpha}{2}}) = NC$$

inviertiendo(aquí es mas simple que el ejemplo anterior), y reemplazando $\widehat{\Lambda} = \frac{H}{n} = \widehat{\lambda}$, la estimación de λ con la muestra

$$\left[\left(\widehat{\lambda} + \frac{z^2_{1-\frac{\alpha}{2}}}{2n}\right) \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{\lambda}}{n} + \frac{z^2_{\frac{\alpha}{2}}}{4n^2}}\right]$$

igual que antes, si n es grande, $\frac{z^2_{1-\frac{\alpha}{2}}}{2n} \simeq 0$ y $\frac{z^2_{1-\frac{\alpha}{2}}}{4n^2} \simeq 0$ y entonces

$$IC_\lambda \text{ al } NC\colon \left[\widehat{\lambda} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{\lambda}}{n}}\right]$$

Nuevamente, la expresión del intervalo es atractiva: $\widehat{\lambda}$(que es el estimador de λ) mas/menos el fractil por $\sqrt{\frac{\widehat{\lambda}}{n}}$ que es el desvío del estimador. Este sería un intervalo óptimo pero aproximado.

REMARK 22. *Si interesa que tamaño de muestra habría que tomar, para obtener un* IC_λ *con* $NC = 90\%$ *con semiamplitud, por ejemplo* 0.5*, habría que resolver(*$z_{0.95} = 1.645$*)*

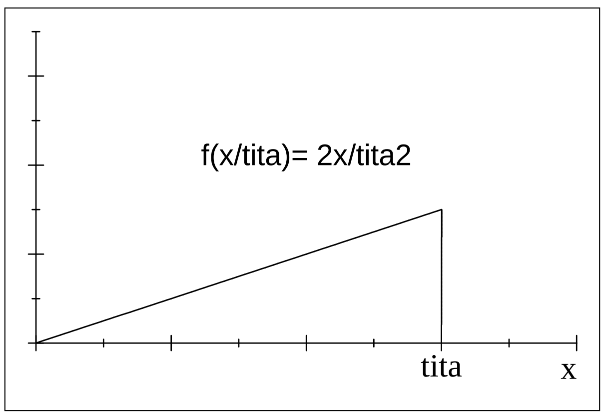
$$1.645\sqrt{\frac{\widehat{\lambda}}{n}} = 0.5 \quad o\ sea\ n = \left(\frac{1.645}{0.5}\right)^2 \widehat{\lambda} = 10.82\widehat{\lambda}$$

Como se desconoce $\widehat{\lambda}$ *habría que asignarle un valor. Pero aquí no podemos tomar el valor que de el n máximo posible, ya que sería* ∞*. En un caso como este, se requiere alguna información adicional sobre la población, por ejemplo que "dificilmente* λ *se mayor que 5". Entonces usando* $\widehat{\lambda} = 5$*, resulta* $n = 54$*.*

EXAMPLE 15. ***Intervalo de confianza para*** θ ***de una muestra de la siguiente población, cuando*** n ***es grande***

$$\boxed{f(x/\theta) = \frac{2x}{\theta^2}\ para\ 0 \leq x \leq \theta}$$
$$X_1 X_2 \cdot \cdot X_n$$

Esta densidad es triangular entre 0 *y* θ*.*



Si se averigua un estadístico suficiente este es

$$H = máx\,\{X_1 X_2 \cdot \cdot X_n\}$$

O sea para construir el IC basta trabajar con el mayor valor de la muestra, y descartar las restantes observaciones. Esto es razonable, pues si se presta atención a la densidad, el máximo de la muestra es la observación que está más cerca de θ*. Con este estadístico obtendremos un* IC_θ *muy angosto. Sin embargo en la materia, no hemos estudiado la distribución del máximo muestral. Así que buscaremos otro estadístico. Si se calcula la media y sigma de la población(integrando!) resulta*

$$E(X) = \frac{2\theta}{3} \quad y\ \ Desv(X) = \frac{\theta}{\sqrt{18}}$$

Veamos como es el siguiente estadístico muestral

$$R = \overline{X} = \frac{X_1 + X_2 + \cdot \cdot + X_n}{n}$$

Sabemos que la media muestral, $R = \overline{X}$ *tiende a la media poblacional,* $\frac{2\theta}{3}$ *cuando* n *es grande. O sea, si el* $\overline{X}$ *muestral nos da pequeño, esto lo interpretaremos como que* θ *es pequeño, y si* $\overline{X}$ *muestral da un valor grande, significará que* θ *será grande. Este análisis intuitivo nos convence que* $\overline{X}$ *tiene "alguna" información respecto de*

θ, así que lo usaremos como estadístico. Como en la expresión de R aparece una suma, por TCL resultará

$$\overline{X} \approx N(\frac{2\theta}{3}; \frac{\theta/\sqrt{18}}{\sqrt{n}})$$

luego

$$\frac{\overline{X} - \frac{2\theta}{3}}{\frac{\theta}{\sqrt{18n}}} = Z \sim N(0;1)$$

utilizando los fractiles de la normal

$$P(-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{X} - \frac{2\theta}{3}}{\frac{\theta}{\sqrt{18n}}} \leq z_{1-\frac{\alpha}{2}}) = NC$$

invirtiendo resulta

$$IC_\theta \quad al\ NC: \left[\frac{\overline{X}}{\frac{2}{3} + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{18n}}}; \frac{\overline{X}}{\frac{2}{3} - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{18n}}}\right]$$

Por ejemplo si la muestra fué de tamaño $n = 72$, y se obtuvo $\overline{X} = 47$, el IC_θ al 95% es $[65.18; 76.78]$. Por otro lado, para la misma muestra, si hubiesemos utilizado el estadístico suficiente $H = máx\{X_1 X_2 \cdot \cdot X_n\}$, y en este caso $h = 69$, el correspondiente intervalo sería IC_θ al 95% : $[69.01; 70.8]$, muchísimo mas angosto. Esto demuestra la importancia de un estadístico suficiente para construir un intervalo.

Example 16. *A veces, las condiciones en que se toma la muestra, determinan la utilización de un estadístico que no es suficiente para la población original. Supongamos que en la etapa final de fabricación de un artículo, se requieren 15min de tallado con un instrumento que tiene un repuesto de duración $G(4;\beta)$min. El inconveniente es que si en estos 15min se rompe el repuesto, deja marcas en el artículo, y queda como defectuoso. Por eso se decide empezar el tallado de cada artículo con un repuesto nuevo. Interesa un IC_μ al 90%, donde μ es la duración media de cada repuesto($\mu = \frac{4}{\beta}$). Lo ideal en este problema(desde el punto de vista estadístico), sería tomar una muestra de duraciones de n repuestos, o sea*

$$\boxed{G(4;\beta)}$$
$$X_1 X_2 \cdot \cdot X_n$$

Aquí el estadístico suficiente es $H = X_1 + X_2 + \cdot \cdot + X_n \sim G(4n;\beta)$. Transformando $2\beta H \sim G(\frac{8n}{2}; \frac{1}{2}) \sim \chi^2_{8n}$, y expresado en función de $\mu = \frac{4}{\beta}$

$$\frac{8H}{\mu} \sim \chi^2_{8n}$$

luego

$$P(\chi^2_{8n,0.05} \leq \frac{8H}{\mu} \leq \chi^2_{8n,0.95}) = 0.9 \quad y\ queda\ IC_\mu\ al\ 90\%: \left[\frac{8H}{\chi^2_{8n,0.95}}; \frac{8H}{\chi^2_{8n,0.05}}\right]$$

El inconveniente aquí es de tipo práctico. Para conseguir la muestra $X_1 X_2 \cdot \cdot X_n$, habría que tallar cada artículo, no 15min, sinó el tiempo necesario para que se rompa el repuesto, y de esta manera conocer cada X_i. Lo común un un problema de este tipo es que la información disponible sea, por ejemplo

"De 80 artículos fabricados, 24 resultaron defectuosos"

O sea, de cada artículo solo se sabe si resultó defectuoso, o bueno. Llamando $P(Def) = p = P(X < 15) = F_\gamma(15/4;\beta)$*, la población y muestra a utilizar será(*$D_i = 1$ *si es defectuoso,* $D_i = 0$ *si es bueno)*

$$\boxed{B_i(1;p}$$
$$D_1 D_2 \cdot\cdot D_{80}$$

Usando $H = D_1 + D_2 + \cdot\cdot +D_{80} \sim B_i(80;p)$ *(que es suficiente para esta nueva población) resulta, como* $h = 24$

$$\begin{cases} F_{B_i}(24/80;p_{\text{sup}}) = 0.05 \\ G_{B_i}(24/80;p_{\text{inf}}) = 0.05 \end{cases} \quad \textit{resultando } IC_p \textit{ al 90\%}: [0.216; 0.395]$$

Como $p = F_\gamma(15/4;\beta)$ *sale* IC_β *al 90%:* $[0.158; 0.213]$*, y como* $\mu = \frac{4}{\beta}$ *resulta*

$$IC_\mu \textit{ al 90\%}: [18.8; 25.3]$$

Por supuesto, si hubiésemos tomado la muestra $X_1X_2\cdot\cdot X_{80}$ *de la población original* $G(4;\beta)$*, el* IC_μ *hubiese sido mucho mas angosto.*

CHAPTER 3

PRUEBAS DE HIPOTESIS

Como en los capítulos anteriores, siempre tendremos una población con un parámetro θ desconocido, y una muestra $X_1X_2 \cdot \cdot X_n$, con X_i independientes. Pero ahora nuestro interés será analizar si se cumple una hipótesis respecto de este parámetro, por ejemplo si $\theta < 5$, o si $\theta > 8$ o si $\theta = 15$.

Como la información que tenemos de la población es a través de la muestra, con ella deberemos decidir si la hipótesis es válida, o no.

Igual que en temas de estimación e intervalos de confianza, por motivos prácticos, en lugar de trabajar con la muestra, lo haremos con un estadístico suficiente respecto del parámetro. Y también aquí existirán dos métodos para construir una prueba de hipótesis: con un estadístico muestral, o con uno pivotal, eligiendo el mas conveniente según el caso.

Por motivos de claridad, se presentará el tema con varios ejemplos, definiendo la terminología particular de las pruebas de hipótesis.

Example 17. *En cierto proceso productivo la concentración de una sustancia química varía de frasco en frasco según una $N(\mu; 2)mg/g$. Con el actual sistema de elaboración, μ no supera los $10mg/g$. Hay una propuesta de un nuevo proceso de elaboración, con el cual se afirma que se obtendrían valores de $\mu > 10$. Esto es importante ya que mejoraría la calidad del producto, pero claro, hay que modificar todo el proceso productivo, lo que tiene sus costos. Por eso, antes de decidirse por el cambio, se decide fabricar 16 frascos con el **nuevo procedimiento**, medir después la concentración de la sustancia en cada frasco, y luego tomar una decisión. O sea, se tendrá*

$$\boxed{N(\mu; 2)}$$
$$X_1X_2 \cdot \cdot X_{16}$$

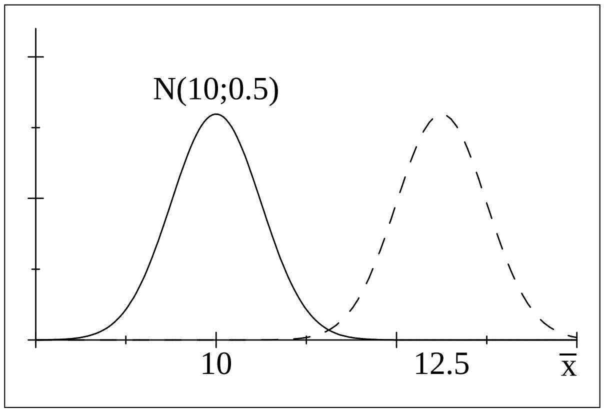
donde el valor de μ con el nuevo procedimiento lo desconocemos, pero queremos decidir entre las hipótesis

$$H_0 : \mu \leq 10 \quad v.s. \quad H_a : \mu > 10$$

*H_0 se designa "**hipótesis nula**", y representa en este caso, la situación en que el nuevo procedimiento es similar al anterior. Y H_a es la "**hipótesis alternativa**", y representa la situación en que el nuevo procedimiento es superior al anterior, y que querriamos detectar(para decidir el cambio). Para evaluar entre estas hipótesis usaremos la muestra. Pero mejor, un estadístico suficiente, que en este caso es $H = X_1 + X_2 + \cdot \cdot + X_{16}$, o también $\overline{X} = \frac{H}{16} \sim N(\mu; \frac{2}{\sqrt{16}}) = N(\mu; 0.5)$, o sea trabajaremos con*

$$\boxed{N(\mu; 0.5)} \tag{0.1}$$
$$\overline{X}$$

Ahora representamos la densidad de $\overline{X}$ *cuando* $\mu = 10$*, o sea la* $N(10;0.5)$



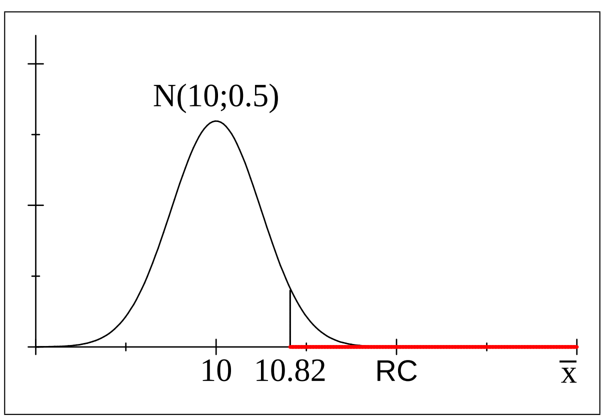
Esta normal nos dice que si $\mu = 10$*, el* $\overline{X}$ *obtenido de la muestra estará alrededor de 10. Pero si* $\mu \in H_a$*, por ejemplo* $\mu = 12.5$*, la normal será otra, una* $N(12.5;0.5)$*, mas corrida a la derecha(representada en punteado), y el* $\overline{X}$ *tomaría en general valores mas grandes(alrededor de* 12.5*). Comprendido esto, ahora se determinará* RC*, la* ***región crítica(o región de rechazo)*** *de esta PH. Es un intervalo de valores que puede tomar* $\overline{X}$ *, tal que, si* $\mu = 10$*, la probabilidad que* $\overline{X}$ *caiga en RC sea muy baja. A esta probabilidad se la designará* α*,* ***nivel de significación*** *de la PH, y usaremos en este ejemplo* $\alpha = 0.05$*. Pero hay muchas regiones con* $\alpha = 0.05$*. Un segundo requisito es elegir la* RC*, en que es mas probable que caiga* $\overline{X}$ *si* $\mu \in H_a$*. Pero como vimos que si* $\mu \in H_a$*, el* $\overline{X}$ *muestral estará mas a la derecha, buscamos una región crítica hacia la derecha, o sea*

$$Si\ \mu = 10,\ buscamos\ \overline{x}_{cr} : P(\overline{X} \geq \overline{x}_{cr}) = \alpha = 0.05$$

O sea, deberá ser $P(\overline{X} < \overline{x}_{cr}) = 0.95$*, luego* $F_z(\frac{\overline{x}_{cr}-10}{0.5}) = 0.95$*, o sea(pqrs),* $\frac{\overline{x}_{cr}-10}{0.5} = 1.645$*, esto es* $\overline{x}_{cr} = 10.82$*, luego la región crítica es*

$$RC : [10.82;\infty)$$

y representando



Ahora sí, llegó el momento de obtener la muestra, calcular el valor de $\overline{X}$*, y tomar una decisión. Emplearemos aquí una regla de decisión que parece natural, aunque* *después la cambiaremos.*

Regla de decisión ***provisoria***
Si $\overline{X} \in RC \Longrightarrow$ ***Aceptamos*** H_a
Si $\overline{X} \notin RC \Longrightarrow$ ***Aceptamos*** H_0

Recordar que planteamos esta PH porque no sabiamos cuál, si H_0 *o* H_a *era cierta. Luego de tomada la muestra, y según si* $\overline{X} \in RC$ *o* $\overline{X} \notin RC$*, concluiremos(con esta regla de decisión "****provisoria****"), cual aceptaremos como cierta. Pero debemos analizar que tan bién funciona esta regla. Lo ideal sería que si* H_0 *fuese cierta, la regla nos conduzca a aceptar* H_0*; y si la cierta fuese* H_a*, con la regla aceptemos* H_a*. Pero a veces esto no ocurre, y cometemos un error.*

- **Error cuando** $\overline{X} \in RC$ **:** si H_a fuese cierta, al aceptar H_a no hay ningún error. Pero si la cierta fuese H_0, al aceptar H_a esto es un error. Este error se designa "**Error tipo I**", y depende del valor que tenga μ dentro de H_0, así que se lo designará $\alpha(\mu)$. Calculemos que tan probable es cometer este error.

$$\begin{aligned} P(\text{Error tipo I}) &= P_{\mu \in H_0}(\overline{X} \in RC) = P_{\mu \in H_0}(\overline{X} \geq 10.82) \\ &= \alpha(\mu) = 1 - F_z(\frac{10.82 - \mu}{0.5}) \leq \alpha = 0.05 \quad \forall \mu \in H_0 \end{aligned}$$

Para $\mu = 10$, $\alpha(\mu) = \alpha = 0.05$. Y si $\mu \in H_0$ con $\mu < 10$, como la $N(\mu; 0.5)$ va a estar mas corrida a la izquierda, entonces en estos casos $\alpha(\mu)$ será todavía menor que $\alpha = 0.05$. O sea, este error esta acotado superiormente por α. Como este error es muy chico para todo $\mu \in H_0$, si $\overline{X} \in RC$ es "casi" válido afirmar que **Aceptamos** H_a.

- **Error cuando** $\overline{X} \notin RC$ **:** si H_0 fuese cierta, al aceptar H_0 no hay ningún error. Pero si la cierta fuese H_a, al aceptar H_0 esto es un error. Este error se designa "**Error tipo II**", y también dependerá del valor de μ dentro de H_a, y lo designaremos $\beta(\mu)$. Calculemos que tan probable es este error.

$$\begin{aligned} P(\text{Error tipo II}) &= P_{\mu \in H_a}(\overline{X} \notin RC) = P_{\mu \in H_a}(\overline{X} < 10.82) \\ &= \beta(\mu) = F_z(\frac{10.82 - \mu}{0.5}) < 1 - \alpha = 0.95 \quad \forall \mu \in H_a \end{aligned}$$

Para $\mu = 12.5$, $\beta(12.5) = 0.0004$, muy chico. Sin embargo, si $\mu = 10.3$, $\beta(10.3) = 0.85$, muy grande. Y si $\mu \in H_a$ con μ muy cercano a 10 por la derecha, por ejemplo $\mu = 10^+$, este error sería todavía mayor $\beta(10^+) = 1 - \alpha = 0.95$. Este error esta acotado superiormente por $1 - \alpha = 0.95$. Pero esta cota es muy alta. Notar entonces, que si H_a es cierta, y $\overline{X} \notin RC$, el error que podemos cometer al Aceptar H_0 puede ser muy probable de ocurrir(dependiendo del μ de H_a). Entonces no parece razonable afirmar que **Aceptamos** H_0.

O sea, las PH presentan una asimetría: cuando $\overline{X} \in RC$ es "casi" válido decir que **Aceptamos** H_a, en cambió si $\overline{X} \notin RC$, **no es correcto** decir que **Aceptamos** H_0, ni tampoco que **Aceptamos** H_a, o sea, en este caso no podremos decidir nada. Por este motivo la regla de decisión que se utiliza, con la terminología usualmente empleada es

(0.2)

Regla de decisión de una PH
Si $\overline{X} \in RC \Longrightarrow$ **Se Rechaza** $H_0 \cong$ **Aceptamos** H_a
Si $\overline{X} \notin RC \Longrightarrow$ **NO** se Rechaza H_0

0.1. Decisión usando el valor de p. Con la regla(0.2), la decisión se tomaba utilizando el estadístico $\overline{X}$ (según caiga o no en la RC).

Para la PH con $RC : [10.82; \infty)$, veamos tres casos, según el valor concreto de $\overline{X}$, y teniendo en cuenta la Regla de decisión.

- **Caso 1:** $\overline{X} = 10.5$. En este caso NO Rechazamos H_0. No se puede decidir nada.
- **Caso 2:** $\overline{X} = 10.9$. En este caso Rechazamos H_0 y Aceptamos H_a. Se decide cambiar el proceso productivo.
- **Caso 3:** $\overline{X} = 11.5$. En este caso Rechazamos H_0 y Aceptamos H_a. Se decide cambiar el proceso productivo.

Esto es correcto.

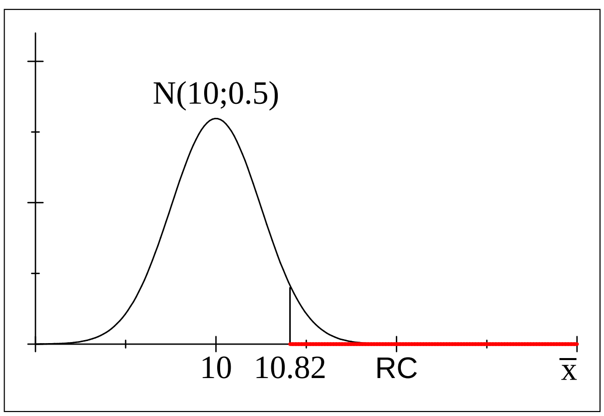
Sin embargo, en los casos 2 y 3, aunque la decisión es la misma, parecería que en el 3, la decisión de cambiar el proceso productivo, es en algún sentido mas "fuerte" que en el 2. Y sin embargo este aspecto no está contemplado en la Regla.

Otra forma, mas completa, de tomar una decisión en una PH es utilizando el "**valor de p**". En lugar de decidir con $\overline{X}$, lo haremos con $p = g(\overline{X})$ que es una función de $\overline{X}$.

Recordar que la condición para hallar el comienzo de la región crítica, $\overline{x}_{cr}$, fué(**imponiendo** $\alpha = 0.05$)

$$\text{Si } \mu = 10, \text{ buscamos } \overline{x}_{cr} : P(\overline{X} \geq \overline{x}_{cr}) = \alpha = 0.05 \qquad (0.3)$$

y de aquí resultó $RC : [10.82; \infty)$ y la figura



Luego comparando el $\overline{X}$ obtenido con $\overline{x}_{cr} = 10.82$, resultaron las conclusiones de los Casos 1, 2 y 3. Ahora procederemos al revés.

En el Caso 2 obtuvimos $\overline{X} = 10.9$(un poco a la derecha de 10.82). Que tal si nos preguntamos "que valor podría haber tenido α de la PH, para que este $\overline{X} = 10.9$ caiga justo al comienzo de la RC", o sea para que $\overline{x}_{cr} = 10.9$? Usando la (0.3), con $\overline{x}_{cr} = 10.9$ resulta

$$\text{Si } \mu = 10, \ P(\overline{X} \geq 10.9) = 1 - F_z(\frac{10.9 - 10}{0.5}) = 0.0359$$

Este es el llamado **valor de p** para $\overline{X} = 10.9$, o sea $p(\overline{X})_{\overline{X}=10.9} = 0.0359$.

Entonces en el Caso 2, diriamos que se rechaza H_0, y por lo tanto que aceptamos H_a con un valor $p = 0.0359$.

Para el Caso 3, donde obtuvimos $\overline{X} = 11.5$ si efectuamos el mismo cálculo

$$\text{Si } \mu = 10,\ P(\overline{X} \geq 11.5) = 1 - F_z(\frac{11.5 - 10}{0.5}) = 0.0013$$

Aquí el valor de p para $\overline{X} = 11.5$, o sea $p(\overline{X})_{\overline{X}=11.5} = 0.0013$.

Y en el Caso 3, diriamos que se rechaza H_0, y por lo tanto que aceptamos H_a con un valor $p = 0.0013$. Notar que como este valor de p es menor que en el caso 2, la decisión de rechazar H_0 y aceptar H_a es mas "fuerte".

Por último si calculamos el valor de p para el caso 1

$$\text{Si } \mu = 10,\ P(\overline{X} \geq 10.5) = 1 - F_z(\frac{10.5 - 10}{0.5}) = 0.1587$$

Aquí el valor de p para $\overline{X} = 10.5$, o sea $p(\overline{X})_{\overline{X}=10.5} = 0.1587$. O sea si tomasemos $\alpha = 0.1587$ también rechazaríamos en este caso. Pero claro, el valor de α de nuestra PH es 0.05, y este valor es fijo, y es el máximo P(error tipo I) admisible. Así que cuando ocurre esto no rechazamos H_0, y no podemos tomar una decisión.

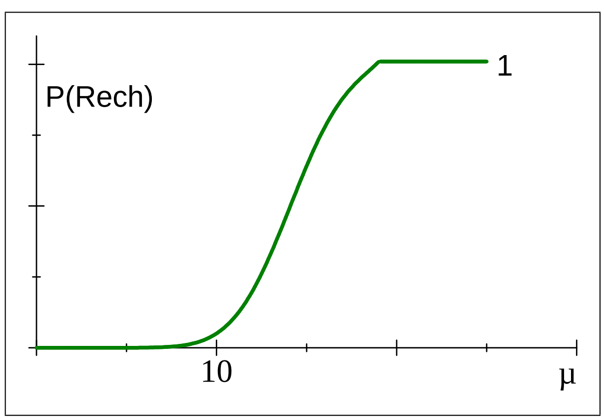
En definitiva, la Regla de decisión usando el valor de p queda

Regla de decisión de una PH
Si $p(\overline{X}) \leq \alpha \Longrightarrow$ **Se Rechaza** $H_0 \cong$ **Aceptamos** H_a
Si $p(\overline{X}) > \alpha \Longrightarrow$ **NO** se Rechaza H_0

0.2. Curva de potencia. Se definirá ahora la "**curva de potencia**" de esta PH. Mide la probabilidad de Rechazar(que el $\overline{X}$ caiga en la RC), en función del valor del parámetro, en este caso μ.

$$\pi(\mu) = P_\mu(\overline{X} \in RC) = P_\mu(\overline{X} \geq \overline{x}_{cr}) = 1 - F_z(\frac{10.82 - \mu}{0.5})$$

Si se representa esta función de μ



La interpretación de esta curva varía dependiendo si $\mu \in H_0$ o $\mu \in H_a$. Notar que:

- Si H_0 es cierta($\mu \leq 10$):

$$P(\text{Error tipo I}) = \alpha(\mu) = P(\text{Rechazar}) = P(\overline{X} \in RC) = \pi(\mu) \text{ para } \mu \in H_0$$

O sea la curva de potencia para $\mu \in H_0$ (o sea $\mu \leq 10$) nos proporciona la P(Error tipo I) para diferentes μ. En particular para $\mu = 10$, resulta $\pi(10) = \alpha$ que es el error tipo I máximo, y decrece para $\mu < 10$. Como vemos, muy bajo error.

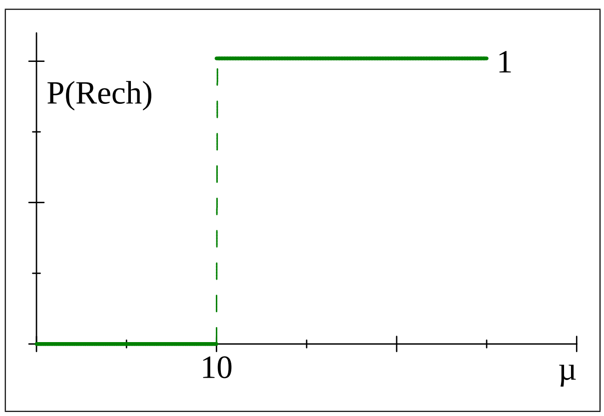
- Si H_a es cierta($\mu > 10$):

$$P(\text{Error tipo II}) = \beta(\mu) = P(\text{NO Rechazar}) = P(\overline{X} \notin RC) = 1 - \pi(\mu) \text{ para } \mu \in H_a$$

O sea, 1 - la curva de potencia, para $\mu \in H_a$ (o sea $\mu > 10$) nos da la P(Error tipo II) para diferentes μ. Como vemos, este error puede ser muy grande, ya que está acotado por $1 - \alpha = 0.95$.

En definitiva la curva de potencia de una PH nos proporciona información completa sobre los dos tipos de errores que se pueden cometer en una PH(debajo de la curva para $\mu \in H_0$, y arriba para $\mu \in H_a$).

0.3. Curva de potencia ideal. La curva de potencia ideal para las hipótesis del ejemplo, sería



Según esta curva, cuando $\mu \in H_0(\mu \leq 10)$, $P(\text{Error tipo I}) = P(\text{Rechazar}) = \pi(\mu) = 0$. Y cuando $\mu \in H_a(\mu > 10)$, $P(\text{Error tipo II}) = P(\text{NO Rechazar}) = 1 - \pi(\mu) = 1 - 1 = 0$. O sea los dos errores serian nulos.

Con esta curva de potencia, Rechazar H_0 equivaldría a Aceptar H_a sin error. Y No Rechazar equivaldría a Aceptar H_0 sin error.

Pero claro, es una curva ideal. Raramente ocurre en las aplicaciones.

En el ejemplo anterior, cuando $\mu \in H_0(\mu \leq 10)$, la curva de potencia $\pi(\mu)$ toma valores muy bajos, concretamente $\pi(\mu) \leq \alpha = 0.05$, o sea se parece bastante a la ideal(el error tipo I es pequeño). Esto se pudo lograr al imponer un valor de $\alpha = 0.05$ muy bajo. Pero para $\mu \in H_a(\mu > 10)$, $\pi(\mu)$ toma valores que difieren mucho de 1, así que el error tipo II es grande. Y este error esta fuera de control.

Pero como se puede hacer para obtener una curva de potencia lo mas parecida a la ideal?

Una forma es utilizar para la PH un estadístico suficiente respecto de parámetro(μ en este caso). Si el estadistico usado **no** fuese suficiente, la $\pi(\mu)$ para $\mu \leq 10$ tomara aún, valores bajos, menores que α, ya que este valor es impuesto. Pero para $\mu > 10$, la $\pi(\mu)$ distará mucho del 1. Por eso se enfatiza el usar un estadístico suficiente, ya que con el se obtiene para $\mu > 10$, una $\pi(\mu)$ lo mas cercana posible a 1.

En nuestro problema utilizamos un estadístico suficiente, y sin embargo para $\mu > 10$, la $\pi(\mu)$ si bién es la mejor posible para la muestra de $n = 16$, todavía dista mucho de 1.

Que podemos hacer entonces?

La solución: usar el estadístico suficiente, pero aumentar el tamaño de la muestra. Cuanto mayor sea la muestra, el estadístico suficiente contiene mayor información respecto del parámetro, lo que se reflejará en una curva de potencia, para $\mu > 10$, cada vez mas cercana a 1(tendiendo a 1 para $n \to \infty$). Pero claro, esto en general no es gratis, tiene un costo.

En definitiva, en inferencia estadística, cuando se aumenta el tamaño de la muestra(n):

- Si un estimador depende de un estadístico suficiente, disminuirá la varianza del estimador.
- Si un IC se construyó con un estadístico suficiente, disminuirá la longitud del intervalo.
- Si una PH se construyó con un estadístico suficiente, mejorará la curva de potencia.

0.4. Mejoramiento de la PH anterior. Como vimos, solo si $\overline{X} \in RC$, al Rechazar H_0(y Aceptar H_a), tomamos la decisión de cambiar el proceso productivo. Pero si $\overline{X} \notin RC$, no se puede tomar ninguna decisión.

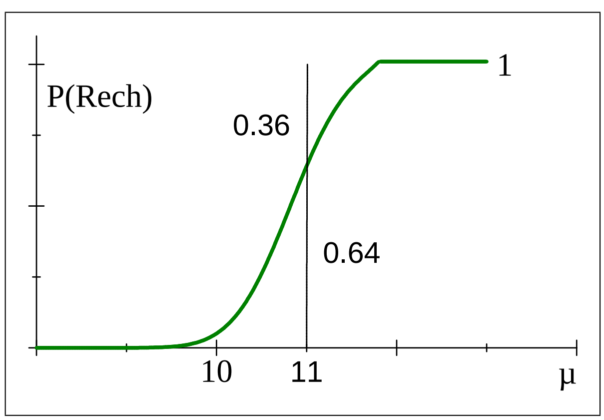
Supongamos ahora que si el nuevo proceso tiene $\mu = 11$, se considera este valor suficientemente interesante como para emprender el cambio del proceso.

Entonces, en este caso, sería deseable que con la PH rechazemos, ya que así efectuaríamos el cambio.

Pero veamos, con la PH anterior, cual es la P(Rechazo) si $\mu = 11$. Usamos la curva de potencia.

$$\pi(11) = P_{\mu=11}(\overline{X} \in RC) = P_{\mu=11}(\overline{X} \geq \overline{x}_{cr}) = 1 - F_z(\frac{10.82 - 11}{0.5}) = 0.64$$

Este valor no es muy alto. O también, visto por el complemento, si $\mu = 11$, P(NO rechazar) = P(Error tipo II) = 0.36



O sea, hay un riesgo del 36% de no rechazar, y por lo tanto no poder tomar ninguna decisión(cuando en realidad querriamos efectuar el cambio).

Nos gustaría para $\mu = 11$ una probabilidad de rechazar de 0.90, y de no rechazar 0.1.

Para resolver este inconveniente habrá que modificar nuestra PH incrementando el tamaño de muestra. O sea consideraremos una muestra $X_1 X_2 \cdot \cdot X_n$, con n a determinar, manteniendo el estadístico suficiente $\overline{X}$, pero ahora su distribución será $N(\mu; \frac{2}{\sqrt{n}})$. O sea tendremos

$$\begin{array}{c}\boxed{N(\mu; \frac{2}{\sqrt{n}})}\\ \overline{X}\end{array}$$

La nueva PH tendrá otro n y otro $\overline{x}_{cr}$(o sea cambiará la RC). Las condiciones que exigiremos a la curva de potencia serán:

$$\begin{cases} \text{Si } \mu = 10 \;\; \text{P(Rechazar)} = 0.05 \\ \text{Si } \mu = 11 \;\; \text{P(Rechazar)} = 0.90 \end{cases}$$

La primera para mantener el error de tipo I máximo en $\alpha = 0.05$, y la segunda para que en 11 el error de tipo II sea 0.10.

La curva de potencia será

$$\pi(\mu) = P_\mu(\overline{X} \in RC) = P_\mu(\overline{X} \geq \overline{x}_{cr}) = 1 - F_z(\frac{\overline{x}_{cr} - \mu}{2/\sqrt{n}})$$

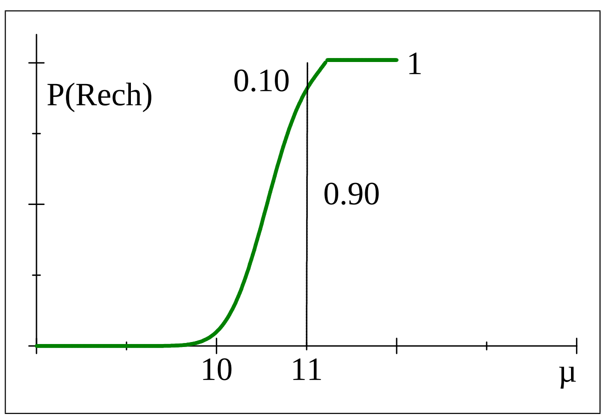
O sea deberá cumplirse

$$\begin{cases} \pi(10) = 1 - F_z(\frac{\overline{x}_{cr} - 10}{2/\sqrt{n}}) = 0.05 \\ \pi(11) = 1 - F_z(\frac{\overline{x}_{cr} - 11}{2/\sqrt{n}}) = 0.90 \end{cases}$$

O sea

$$\begin{cases} F_z(\frac{\overline{x}_{cr} - 10}{2/\sqrt{n}}) = 0.95 \\ F_z(\frac{\overline{x}_{cr} - 11}{2/\sqrt{n}}) = 0.10 \end{cases} \quad \text{o sea}(pqrs) \quad \begin{cases} \dfrac{\overline{x}_{cr} - 10}{2/\sqrt{n}} = 1.645 \\ \dfrac{\overline{x}_{cr} - 11}{2/\sqrt{n}} = -1.282 \end{cases}$$

Resolviendo el sistema resulta $n \simeq 35$ y $\overline{x}_{cr} = 10.56$, con curva de potencia



Notar que se mantiene el nivel de significación $\alpha = 0.05$(ya que $\pi(10) = 0.05$), pero, aunque no se ve muy claro en la figura, debajo de 10, la curva decrece más rápidamente, lo que significa que para $\mu < 10$, la P_μ(error tipo I) será aún menor. Lo que sí se nota claramente es que para $\mu > 10$ la curva sube más rápido, y en $\mu = 11$ alcanza una P_μ(Rechazo)=0.90, o sea un P(error tipo II)=0.10.

0.5. PH usando un estadístico pivotal.