Estimadores de máxima verosimilitud: distribución conjunta de la muestra

- Los conceptos de funciones de verosimilitud se deben a Fisher, y es fundamental en inferencia estadística.
- Este concepto se define a partir de la distribución conjunta de la muestra.
- Supongamos una variable discreta x con distribución $P(x; \vartheta)$ que es conocida.
- Supongamos que tomamos muestras independientes de tamaño n, representando está por el vector **X**.
- Así podemos definir la distribución conjunta de la muestra en función de esta variable, y para el caso de una muestra aleatoria simple tenemos:
  - $P(\mathbf{X} = \mathbf{X}_0) = P(x_1 = x_{10}, x_2 = x_{20}, \ldots, x_n = x_{n0}) = P(x_{10}) \ldots P(x_{n0})$
- Así conociendo la distribución $P(x; \vartheta)$ podemos calcular fácilmente la probabilidad de cualquier muestra (**ver ejemplo 7.6**, para una Poison, $P(x; \lambda)$, **7.7 para una binomial** y **7.8 para una exponencial**).

Estimadores de máxima verosimilitud: distribución conjunta de la muestra

- En el **caso continuo** (función de densidad $f(x; \vartheta)$), la probabilidad del intervalo $x_i - 1/2, x_i + 1/2$, la podemos aproximar por el rectángulo de altura $f(x_i)$ y base unidad:
  - $\mathbf{P(x_i) = f(x_i) \cdot 1}$
- Por tanto la probabilidad de la muestra aleatoria simple:
  - $P(x_1, \ldots, x_n) = \prod f(x_i)$
- Así la función de densidad conjunta de la muestra $f(x_1, \ldots, x_n)$ se interpreta como la probabilidad de obtener los valores muestrales $x_1 \pm 0{,}5, \ldots, x_n \pm 0{,}5$.

Estimadores de máxima verosimilitud: la función de verosimilitud

- Sea una variable aleatoria continua x con función de densidad que $f(x|\vartheta)$ para indicar que depende de un vector de parámetros ϑ. Es decir dado que conozco ϑ, representa cual es la función de densidad de la variable aleatoria x.
- Si tenemos una muestra aleatoria simple $\mathbf{X} = (x_1, \ldots, x_n)$, entonces la función de densidad conjunta de la muestra es:
  - $f(\mathbf{X}|\vartheta) = \prod f(x_i|\vartheta)$
- Es decir cuando conozco ϑ la expresión anterior determina la probabilidad de aparición de cada muestra.

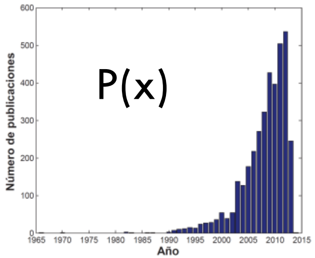
Estimadores de máxima verosimilitud: la función de verosimilitud

- En inferencia para un problema de estimación se **conoce un valor particular de una muestra X**, siendo **desconocido el parámetro** ϑ.
- Así si sustituimos **X** por el valor observado de una muestra, $\mathbf{X}_0 = (x_{10}, \ldots, x_{n0})$, entonces la función $f(\mathbf{X}_0|\vartheta)$ puede ser vista como una función del parámetro.
- Es decir $f(\mathbf{X}_0|\vartheta)$ puede ser visto y proporciona, para cada valor de ϑ, la probabilidad de obtener el valor muestral $\mathbf{X}_0$ para ese valor del parámetro ϑ.

Estimadores de máxima verosimilitud: la función de verosimilitud

- Así esta nueva función que obtenemos cuando variamos ϑ, mientras mantenemos $\mathbf{X}_0$ fijo (no variamos la muestra), define la **función de verosimilitud**, $\ell(\vartheta \mid \boldsymbol{X})$ (dado que conozco la muestra, entonces como varia la probabilidad en función del parámetro ϑ):
  - $\ell(\vartheta \mid \boldsymbol{X})$, o $\ell(\vartheta)$: Es decir $\ell(\vartheta \mid \boldsymbol{X}) = \ell(\vartheta) = f(\mathbf{X}_0 \mid \vartheta)$, con $\mathbf{X}_0$ fijo y ϑ variable.
- La óptica cambia, en vez de tener un parámetro fijo ϑ y calcular para ese parámetro la probabilidad de obtener distintas muestras **X**, lo que fijamos es una determinada muestra $\mathbf{X}_0$ y estimamos que valor del parámetro ϑ hace más verosímil la muestra que se observa $\mathbf{X}_0$.

Estimadores de máxima verosimilitud: la función de verosimilitud



Población

$f(\mathbf{X}_0 \mid \vartheta)$

Muestra

$\mathbf{X}_0 = (x_{10}, \ldots, x_{n0})$

$\ell(\vartheta \mid \boldsymbol{X}_0)$

Estimadores de máxima verosimilitud: la función de verosimilitud

- Este enfoque cambia completamente la forma de la función y ya no se parece en nada a la densidad o distribución de probabilidad, aunque sigue teniendo unidades de probabilidad (pensar porque tiene unidades de probabilidad).
- Si tenemos una variable x que distribuye según una Poisson:
  - $P(x=r)=\frac{\lambda^r}{r!}e^{-\lambda}, r=0,1,2,\ldots$, y observamos el valor de la muestra x=5, entonces $\ell(\lambda)=\frac{\lambda^5}{5!}e^{-\lambda}$, es la función de verosimilitud para una muestra de un solo valor de x=5.

Estimadores de máxima verosimilitud: la función de verosimilitud

- Esta función $\ell(\lambda)$ es continua en λ y **proporcional a la probabilidad de observar x=5 para cada valor posible de λ.**
- El valor de la verosimilitud no es único: $\ell(\vartheta_1) = f(\boldsymbol{X}_0|\vartheta_1) > f(\boldsymbol{X}_0|\vartheta_2) = \ell(\vartheta_2)$.
- Esto quiere decir que **a la vista de los datos muestrales el valor del parámetro $\boldsymbol{\vartheta_1}$ es más verosímil** que el valor del parámetro ϑ_2 ya que la probabilidad de obtener la muestra observada $\mathbf{X}_0$ es mayor con ϑ_1 que con ϑ_2.

Estimadores de máxima verosimilitud: la función de verosimilitud

- Observar que la verosimilitud tiene unidades, las de la variable x, entonces la diferencia de verosimilitudes no tiene sentido ya que varia arbitrariamente en función de las unidades de la variable x.
- Por lo tanto para comparar verosimilitudes lo mejor es el cociente de las mismas ya que este es invariante frente a las diferentes unidades de la variable:
  - $\ell(\vartheta_1 \mid \boldsymbol{X})/\ell(\vartheta_2 \mid \boldsymbol{X})$, este cociente es invariante hacia cambios de escalas en la variable x que se esta observando.
  - El cociente $\ell(\vartheta_1)/\ell(\vartheta_2)$ se puede sustituir por la diferencia de logaritmos: $\ln\ell(\vartheta_1) - ln\ell(\vartheta_2)$.
- Así se puede definir la **función soporte** por el logaritmo de la verosimilitud: $\mathrm{L}(\vartheta) = ln\ell(\vartheta)$.

Estimadores de máxima verosimilitud: la función de verosimilitud

- Se define como la discriminación contenida en la muestra **X** entre ϑ_1 y ϑ_2 a la siguiente expresión (diferencia de soporte de ambos valores): $\mathrm{L}(\vartheta_2) - \mathrm{L}(\vartheta_1) = ln\ell(\vartheta_2) - ln\ell(\vartheta_1)$.
- Si ϑ es un parámetro cuyos valores posibles pertenecen a un intervalo ϑ_1 y ϑ_2, llamaremos **discriminación relativa** entre ϑ_2 y ϑ_1 a:
  - $(\mathrm{L}(\vartheta_2) - \mathrm{L}(\vartheta_1))/(\vartheta_2 - \vartheta_1) = (ln\ell(\vartheta_2) - ln\ell(\vartheta_1))/\,(\vartheta_2 - \vartheta_1)$.
- En el limite cuando $\vartheta_2 \to \vartheta_1$ obtenemos la tasa de discriminación para la muestra **X** respecto el parámetro ϑ valorada en el punto ϑ_1
- $\mathrm{d}(\vartheta_1) = \lim\limits_{\vartheta_2 \to \vartheta_1} \frac{\mathrm{L}(\vartheta_2) - L(\vartheta_1)}{\vartheta_2 - \vartheta_1} = \lim\limits_{h \to 0} \frac{\mathrm{L}(\vartheta_1 + h) - L(\vartheta_1)}{h} = \left.\frac{\mathrm{dL}(\vartheta)}{\mathrm{d}\vartheta}\right|_{\vartheta\, =\, \vartheta_1}$, fue introducida por Fisher, que la denominó “Score”.

Estimadores de máxima verosimilitud: la función de verosimilitud

- Si este "score" cumple que $d(\vartheta_1)>0$, la verosimilitud aumenta para valores superiores a ϑ_1,
  - es decir, la muestra tiene mayor probabilidad de ocurrir con valores mayores que ϑ_1,
- Mientras que si $d(\vartheta_1)< 0$ el razonamiento es el contrario, la verosimilitud aumenta para valores inferiores de ϑ_1,
  - es decir, la muestra tiene mayor probabilidad de ocurrir con valores menores que ϑ_1.

Estimadores de máxima verosimilitud: la función de verosimilitud

- Resumiendo:

1. La función de verosimilitud es la herramienta básica que nos permite juzgar la compatibilidad entre los valores muestrales observados y los posibles valores del parámetro de la distribución de probabilidad.
2. Si queremos comparar dos posibles valores del parámetro, ϑ, se debe utilizar el cociente de sus verosimilitudes, y no su diferencia, ya que la diferencia depende de la escala de medida de las variables, como hemos dicho antes.

Estimadores de máxima verosimilitud: la función de verosimilitud

- Por ejemplo para estimar el parámetro, λ, de Poisson de la muestra observada $x_1, \ldots, x_n$ hacemos:
  - $P(x) = \frac{\lambda^x}{x!} e^{-\lambda}$, distribución de Poisson.
  - $\ell(\lambda) = \Pi \, P(xi|\lambda) = P(x_1|\lambda)P(x_2|\lambda) \ldots P(xn|\lambda) = \frac{\lambda^{\Sigma \mathbf{x}_i}}{\Pi \mathbf{x}_i!} e^{-n\lambda}$
- Como el término $1/(\Pi x_i!)$ es una constante la podemos eliminar y escribir la función de verosimilitud para la distribución de Poisson como:
  - $\ell(\lambda) = \; e^{-n\lambda} \; \lambda^{\Sigma \mathbf{x}_i} = e^{-n\lambda} \; \lambda^{n\bar{x}}$
  - Así la función soporte será para la distribución de Poisson vendrá dada por: $\mathrm{L}(\lambda) = -n\lambda + n\bar{x} \; ln\lambda$ (mirar los ejercicios para el resto de distribuciones).

Estimadores de máxima verosimilitud: el método de máxima verosimilitud

- Una vez que tenemos calculada una función de verosimilitud para un vector de parámetros $\boldsymbol{\vartheta}$, $\ell(\boldsymbol{\vartheta})$, un procedimiento intuitivo para estimar los parámetros de la distribución a partir de los valores observados muestralmente es maximizar el valor del parámetro que sea más verosímil, es decir el que maximice la verosimilitud.
- Así podemos resolver el sistema de ecuaciones:
  - $\partial\ell(\boldsymbol{\vartheta})/\partial\vartheta_1 = 0, \ldots, \partial\ell(\boldsymbol{\vartheta})/\partial\vartheta_p = 0$
- El valor que resuelve el sistema de ecuaciones, $\widehat{\boldsymbol{\vartheta}}$, corresponderá a un máximo si el valor de la matriz hessiana de segundas derivadas en ese punto es definida negativa: $\mathrm{H}(\widehat{\boldsymbol{\vartheta}}) = (\partial^2\ell(\boldsymbol{\vartheta})/\partial\vartheta_i\partial\vartheta_j)_{\boldsymbol{\vartheta}\,=\,\widehat{\boldsymbol{\vartheta}}}$.

Estimadores de máxima verosimilitud: el método de máxima verosimilitud

- A la hora de hacer los cálculos de estimadores máximo-verosímiles (MV) se obtienen derivando la función soporte: $\mathrm{L}(\vartheta) = ln\ell(\vartheta)$, ya que la transformación logarítmica es monótona y por tanto tiene el mismo máximo.
- Recordemos que la derivada de la función soporte la habíamos definido como la tasa de discriminación: $\mathrm{d}(\vartheta_1) = \lim_{\vartheta_2 \to \vartheta_1} \frac{\mathrm{L}(\vartheta_2) - L(\vartheta_1)}{\vartheta_2 - \vartheta_1} = \frac{\mathrm{dL}(\vartheta)}{\mathrm{d}\vartheta}\Big|_{\vartheta\, =\, \vartheta_1}$, así podemos definir el estimador máximo-verosímil como aquel valor de los parámetros para los que se anulan la tasa de discriminación de la muestra.

Estimadores de máxima verosimilitud: Ejemplo

- Para estimar los parámetros de una normal de la muestra observada $x_1, \ldots, x_n$:
  - $f(x)=1/\sigma(2\pi)^{0.5} \exp\{-(1/2\sigma^2)(x-\mu)^2\}$, densidad de probabilidad normal.
  - $\ell(\mu,\sigma^2) = \Pi\, f(xi|\mu,\sigma^2) = f(x_1|\mu,\sigma^2) f(x_2|\mu,\sigma^2) \ldots f(xn|\mu,\sigma^2) =$
    $\Pi_i (1/\sigma(2\pi)^{0.5}) \exp\{-(1/2\sigma 2)(x_i-\mu)^2\} =$
    $(1/(\sigma(2\pi)^{0.5})^n) \exp\{-(1/2\sigma^2)\Sigma(xi-\mu)2\}.$
- Así la función soporte será para la distribución normal vendrá dada por: $L(\mu,\sigma^2) = -n \ln(\sqrt{2\pi}\,\sigma) - (1/2\sigma^2)\Sigma(xi-\mu)2 =$
  $-(n/2) \ln(2\pi\sigma^2) - (1/2\sigma^2)\Sigma(xi-\mu)2.$
- Así ahora derivamos la función soporte de la muestra e igualamos a cero para sacar cuales son los estimadores de máxima verosimilitud para μ, σ^2 de la población.

Estimadores de máxima verosimilitud: Ejemplo

- Así tenemos $L(\mu,\sigma^2) = -(n/2)\ln(2\pi\sigma^2) - (1/2\sigma^2)\Sigma(xi-\mu)^2$ y la optimizamos en función de μ y σ^2:
- $\partial L(\mu,\sigma^2)/\partial\mu = 0 = (2/2\sigma^2)\Sigma(xi-\mu) = \Sigma(xi-\mu)/\sigma^2 = (n\bar{x} - n\mu)/\sigma^2$, así $\boxed{\hat{\mu} = \bar{x}}$, por lo tanto parece lógico que el estimador máxima verosimilitud para la media de la normal es la media aritmética de la muestra observada.
- $\partial L(\mu,\sigma^2)/\partial\sigma^2 = 0 = -(n/2)(2\pi/2\pi\sigma^2) + (2\,\Sigma(xi-\mu)^2)/(2\sigma^2)^2 = -(n/2)(1/\sigma^2) + \Sigma(xi-\mu)^2/2\sigma^4 = -(n\sigma^2/2\sigma^4) + \Sigma(xi-\mu)^2/2\sigma^4 = (-n\sigma^2 + \Sigma(xi-\mu)^2)/2\sigma^4$ así despejando $\boxed{\hat{\sigma}^2 = \Sigma(xi-\mu)^2/n = s^2}$, por lo tanto parece lógico que el estimador máxima verosimilitud para la varianza de la normal es la desviación típica de la muestra observada.

Estimadores de máxima verosimilitud: propiedades de los estimadores máximo-verosímiles

- Si $\widehat{\boldsymbol{\vartheta}}_{MV}$ son los estimadores de máxima verosimilitud de un modelo de una población a través de sus muestras estos cumplen las siguientes propiedades:
  - Asintóticamente centrados.
  - Asintóticamente normales.
  - Asintóticamente eficientes.
  - Suficiencia.
  - Invariancia.
  - Robustez.

Estimadores de máxima verosimilitud: propiedades de los estimadores máximo-verosímiles - invariancia

- Invariancia:
  - Si $\hat{\boldsymbol{\vartheta}}_{MV}$ es el estimador máximo verosímil de $\boldsymbol{\vartheta}$, entonces h($\hat{\boldsymbol{\vartheta}}_{MV}$) es el estimador máximo verosímil de h($\boldsymbol{\vartheta}$).
  - Ejemplo:
    - Sea $x_1, \ldots, x_n$ una muestra aleatoria simple de $x \sim N(\mu, \sigma)$.
    - Sabemos que $\hat{\mu}_{MV} = \bar{x}$, lo hemos demostrado anteriormente.
    - ¿Quiénes serán los estimadores de máxima verosimilitud para 3μ, μ^2 y $1/\mu^2$?
    - Por el principio de invariancia tenemos que:
    - $\widehat{3\mu}_{MV} = 3\bar{x}$
    - $\widehat{\mu^2}_{MV} = \bar{x}^2$
    - $\widehat{1/\mu}_{MV} = 1/\bar{x}$

Estimadores de máxima verosimilitud: propiedades de los estimadores máximo-verosímiles – consistencia y centrado

- Consistencia:
  - Bajo ciertas condiciones generales, $\widehat{\boldsymbol{\vartheta}}_{MV}$ es un estimador consistente de $\boldsymbol{\vartheta}$.
    - $\lim\limits_{n\to\infty} E[\widehat{\vartheta}_n] \to \vartheta.$
    - $\lim\limits_{n\to\infty} Var[\widehat{\vartheta}_n] \to 0.$
- Asintóticamente centrado:
  - Se verifica que el $\lim\limits_{n\to\infty} E\left[\widehat{\boldsymbol{\vartheta}}_{MV}\right] = \boldsymbol{\vartheta}.$

Estimadores de máxima verosimilitud: propiedades de los estimadores máximo-verosímiles – **normalidad asintótica**

- Normalidad asintótica :

Recordad $f(a)+\cancel{\frac{f'(a)}{1!}(x-a)}+\frac{f''(a)}{2!}(x-a)^2+\frac{f^{(3)}(a)}{3!}(x-a)^3+\cdots$

  - Para tamaños muestrales grandes, desarrollando en serie la función soporte en un entorno del estimador $\widehat{\boldsymbol{\vartheta}}_{MV}$ para 2° orden:

Recordad que una normal es:

$1/\sigma(2\pi)^{0.5}\exp\{-(1/2\sigma^2)(x-\mu)^2\}$

  - $L(\boldsymbol{\vartheta}) \cong L\left(\widehat{\boldsymbol{\vartheta}}_{MV}\right) + (1/2)\left(\frac{d^2L\left[\widehat{\boldsymbol{\vartheta}}_{MV}\right]}{d\vartheta^2}\right)(\boldsymbol{\vartheta} - \widehat{\boldsymbol{\vartheta}}_{MV})^2.$
  - Si llamamos $\widehat{\boldsymbol{\sigma}}_{MV}^2 = \left(-\frac{d^2L\left[\widehat{\boldsymbol{\vartheta}}_{MV}\right]}{d\vartheta^2}\right)^{-1}$, entonces la verosimilitud se puede escribir: $\ell(\boldsymbol{\vartheta}) = \ell(\boldsymbol{\vartheta}|\boldsymbol{X}) = \mathrm{k}\exp\left(-\frac{1}{2\widehat{\boldsymbol{\sigma}}_{MV}^2}(\boldsymbol{\vartheta} - \widehat{\boldsymbol{\vartheta}}_{MV})^2\right)$, como el soporte es el log de la verosimilitud, la verosimilitud es la exp, la constante k es $\exp(L(\widehat{\boldsymbol{\vartheta}}_{MV}))$.
  - Así la verosimilitud tiene la forma de una normal, con media $\widehat{\boldsymbol{\vartheta}}_{MV}$, y varianza $\widehat{\boldsymbol{\sigma}}_{MV}^2$.

Ejemplos para realizar en casa

- Realizar para casa los ejemplos del libro:
  - 7-1, 7-2, 7-3, 7-4, 7-6, 7-7, 7-8, 7-9, 7-10, 7-11 y 7-12.

Inferencia Estadística: Estimación por intervalos

Estimación por intervalos

- Introducción
- Un ejemplo simple
- El método del pivote
- Principales estadísticos pivote
- Estimación autosuficiente de intervalos de confianza (bootsrap)

Estimación por intervalos: introducción

- Ya sabemos cómo obtener **estimadores** para un parámetro y cómo calcular una **medida de la precisión del estimador**: es decir su desviación típica en el muestreo.
- Siempre es conveniente dar junto al estimador un intervalo de valores entre los cuales deberá estar el valor del parámetro que se estima con alta probabilidad.
- Éste es el objetivo de la **estimación por intervalos** de confianza.
- Para ilustrar el problema vamos a ver como ejemplo la estimación de la media con una muestra de tamaño 25 en una población normal de desviación típica conocida e igual a 10.
- Antes de observar la muestra y calcular el estimador, por ejemplo $\bar{x}$, se pueden predecir las discrepancias esperadas entre el estimador ($\bar{x}$) y el parámetro (μ): el 95% de las veces: $|\bar{x} - \mu| \leq 1{,}96 \frac{\sigma}{\sqrt{n}} = 1{,}96 \frac{10}{\sqrt{25}} = 3{,}92$

Estimación por intervalos: introducción

- Esto lo veremos en la metodología a continuación como se calcula con detalle.
- Pero lo importante es que podemos predecir que el 95% de las veces $|\bar{x} - \mu|$ no será mayor de 3,92 unidades.
- Por tanto si observamos $\bar{x}= 40$, podemos asegurar que μ estará previsiblemente en el intervalo $40 \pm 3{,}92$, el 95% de las veces que saquemos una muestra del mismo tamaño 25.
- **Ésta es la idea central de construcción de intervalos de confianza.**
- Así Llamaremos intervalo de confianza para el parámetro ϑ a un nivel de confianza $(1 - \alpha)$, a una expresión del tipo $\vartheta_1 \leq \vartheta \leq \vartheta_2$ donde los límites ϑ_1 y ϑ_2 (40-3,92 y 40+3,92 en el ejemplo) dependen de la muestra (a través de $\bar{x}= 40$), con $\alpha \leq 1$.

Estimación por intervalos: introducción

- Estos intervalos de confianza se calculan de manera tal que si tomamos muchas muestras, todas del mismo tamaño, y construimos un intervalo con cada muestra, podemos afirmar que el $100(1-\alpha)\%$ (en el ejemplo 95%) de los intervalos así construidos contendrán el verdadero valor del parámetro. En este caso en el ejemplo $\alpha = 0.05$.
- El otro $100\alpha\ \%$ (en el ejemplo 5%) de los intervalos así construidos no contendrán el valor verdadero del parámetro.

Estimación por intervalos: introducción

- La idea general del procedimiento es que si estimamos $\boldsymbol{\vartheta}$ con el estimador máximo verosímil $\widehat{\boldsymbol{\vartheta}}_{MV}$, el error relativo (error cometido/error promedio) de la estimación se define de la siguiente forma:
  - $\omega = \boldsymbol{\vartheta} - \widehat{\boldsymbol{\vartheta}}_{MV}/\sigma(\widehat{\boldsymbol{\vartheta}}_{MV})$, donde $\sigma(\widehat{\boldsymbol{\vartheta}}_{MV})$ es la desviación típica asintótica de la distribución muestral del estadístico máximo-verosímil, que sigue asintóticamente una distribución normal estándar (por el TCL).

Estimación por intervalos: introducción

- Este resultado indica que, sea cual sea $\boldsymbol{\vartheta}$, podemos conocer aproximadamente la distribución del error relativo que cometeremos al estimar este parámetro por $\widehat{\boldsymbol{\vartheta}}_{MV}$.
- Esto nos lleva al **método del pivote** para la estimación de intervalos de confianza que veremos en detalle después.
- Pero antes vamos a ver este ejemplo sencillo con más profundidad para entender el problema de una manera más clara, y por tanto el método del pivote.

Estimación por intervalos: un ejemplo simple (para entender el problema)

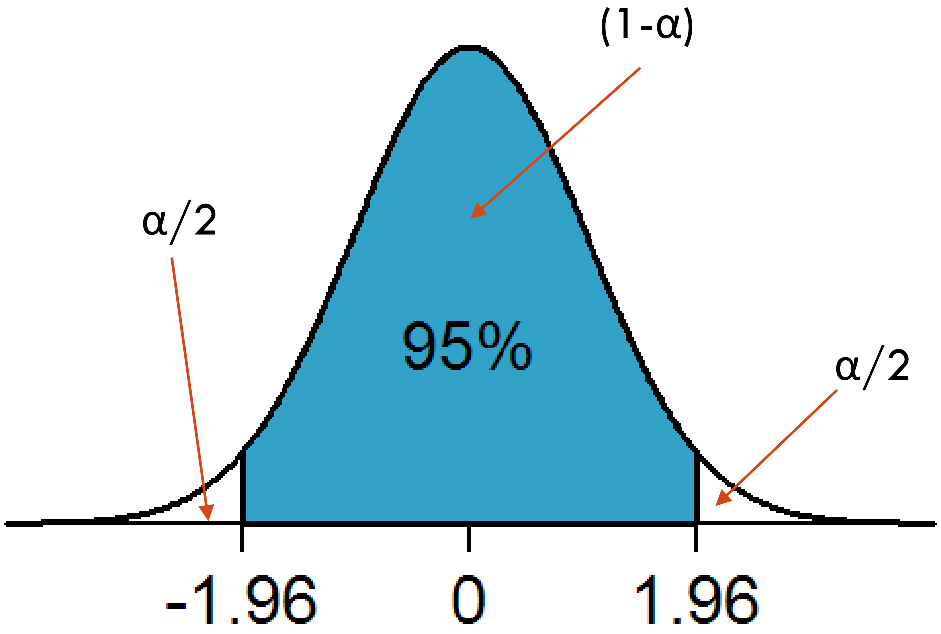
- Vamos a tomar el ejemplo que acabamos de ver con más detalle:
  - Supongamos que sabemos que tenemos una variable aleatoria que sigue una distribución N(μ, σ) donde el valor de varianza es conocido $\sigma^2 = \sigma^2_0$.
  - El objetivo es, dada una m.a.s. de tamaño n de la variable obtener un intervalo a un nivel de confianza del 95 % para el parámetro μ de la distribución.
  - Así tenemos una probabilidad del 95% de encontrar μ en ese intervalo.
- Recordemos que al tomar una muestra de tamaño n de una variable con media μ y varianza σ^2 y distribución cualquiera (en particular una normal también), la distribución muestral de la media verifica que $E[\bar{x}]=\mu$ y $Var[\bar{x}]= \sigma^2/n$.
- En este caso la distribución muestral de la media muestral $\bar{x}$ sigue una normal $N(\mu, \sigma_0/(n)^{0.5})$ (transparencia 244).
- Esto lo podemos aprovechar para construir el siguiente estadístico como variable aleatoria z que sigue una N(0,1), que viene determinado por
  - $z = (\bar{x} - \mu)/(\sigma_0/\sqrt{n})$, que es la variable estandarizada.

Estimación por intervalos: un ejemplo simple (para entender el problema)

- Así la nueva variable aleatoria z se distribuye según una N(0,1).
- Por favor notar que es importante que el nuevo estadístico tenga una distribución conocida, siendo esta distribución conocida independiente del parámetro que queremos estimar en este caso que es μ a través de la media muestral $\bar{x}$:
  - Ya que esto nos permite dada la nueva variable z construir una expresión del siguiente tipo (sabiendo que z es una normal):
    - $P(-z_{\alpha/2} \leq z \leq z_{\alpha/2}) =$
      $$P(-z_{\alpha/2} \leq (\bar{x} - \mu)/(\sigma_0/\sqrt{n}) \leq z_{\alpha/2}) = 1 - \alpha$$
    - Donde $z_{\alpha/2}$ es un valor de la normal estándar tal que cumple:
      - $P(z > z_{\alpha/2}) = 1 - \Phi(z_{\alpha/2}) = \alpha/2$, siendo Φ la función distribución normal.
- Esto nos permite estimar el parámetro μ a través de un intervalo de confianza de la siguiente manera.

Estimación por intervalos: un ejemplo simple (para entender el problema)

- Así la expresión anterior nos permite determinar, de manera independiente de μ, el valor $z_{\alpha/2}$ que delimita una probabilidad $(1-\alpha)$ dentro del intervalo centrado en cero $(-z_{\alpha/2}; z_{\alpha/2})$. Es decir $z_{\alpha/2}=z_{0.975}$
- En este caso, para la distribución N(0, 1), y $(1-\alpha)=95\%$ y el valor $|z_{\alpha/2}|$ es aproximadamente 1,96 (mirar las tablas del final del libro, Tabla 4 Distribución normal estandarizada, *N*(0,1)).



Estimación por intervalos: un ejemplo simple (para entender el problema)

- Por lo tanto podemos poner para α=5% según la figura anterior de la normal, teniendo en cuenta que 1.96 es el valor aproximado del punto percentil 97.5 de la distribución normal:
  - $P(-1.96 \leq (\bar{x} - \mu)/(\sigma_0/\sqrt{n}) \leq 1.96) = 0.95$
  - Tabla libro, pag 618, el área mas cercana es 0,97500 (que es el valor exacto que buscamos) que corresponde z=1.96 para el percentil $z_{0.975}$ de la normal.
- Así si despejamos μ obtenemos el intervalo de confianza siguiente:
  - $P(\bar{x} - 1.96\frac{\sigma_0}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96\frac{\sigma_0}{\sqrt{n}}) = 0.95$

Estimación por intervalos: un ejemplo simple (para entender el problema)

- Así la expresión $P(\bar{x} - 1.96\frac{\sigma_0}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96\frac{\sigma_0}{\sqrt{n}}) = 0.95$ lo que quiere decir básicamente es que:
  - Para la **estimación de la media** con una **muestra de tamaño n** en una **población normal** de **desviación** típica conocida e igual a $\boldsymbol{\sigma_0}$.
  - **Podemos decir** que antes de observar la muestra y calcular el estimador, $\bar{x}$, se pueden predecir las **discrepancias** esperadas entre el estimador ($\boldsymbol{\bar{x}}$) y el parámetro ($\boldsymbol{\mu}$) que vienen dadas por el siguiente intervalo de confianza el 95% de las veces que se observa la muestra:
    - $\mathbf{|\bar{x} - \mu| \leq 1,96\frac{\sigma_0}{\sqrt{n}}}$

Estimación por intervalos: método del pivote

- El método anterior se puede generalizar dando origen al conocido **método del pivote** para la construcción de intervalos de confianza.
- Así este se basa en la **elección de una variable aleatoria** que sea **función de la muestra** y del **parámetro a estimar**, con una serie de condiciones para la función:
  - Que sea una **función continua y monótona** del parámetro.
  - Que su **distribución sea conocida e independiente del parámetro** a estimar.

Estimación por intervalos: método del pivote

- Llamemos **g(ϑ, X)** a la variable escogida y que recibe el nombre de **estadístico pivote.**
- Vemos que esta variable depende del parámetro a estimar, ϑ, y la muestra escogida, X.
- Bajo estas condiciones, fijado el nivel de confianza $(1-\alpha)100\%$, siempre es posible encontrar los valores a y b tales que $P(a \leq g(\vartheta, X) \leq b) = (1-\alpha)$, siempre que la distribución sea conocida para la probabilidad, P, anterior.
- Por las condiciones exigidas sobre el estadístico, será posible despejar ϑ y por tanto obtener los límites para el intervalo.

Estimación por intervalos: método del pivote

- Recordemos que las condiciones que se exigen para el estadístico pivote son:
  1. Que sea una función continua y monótona del parámetro.
  2. Que su distribución sea conocida e independiente del parámetro a estimar.
- Debido a la primera condición podemos despejar el estimador para deducir el intervalo de confianza:
  - $P(g^{-1}(a, X) \leq \vartheta \leq g^{-1}(b, X)) = (1-\alpha)$
- Siendo $\vartheta_1 = g^{-1}(a, X)$ y $\vartheta_2 = g^{-1}(b, X)$ los límites del intervalo deseado, para el estimador ϑ.

Estimación por intervalos: método del pivote

- Debido a la primera condición podemos asegurar que existen $g^{-1}(a, X)$ y $g^{-1}(b, X)$.
- Se puede demostrar que si una función f es continua y monótona en un intervalo [a, b], entonces existe la inversa en el intervalo [f(a),f(b)], y es también monótona y continua.
- Una función monótona es aquella que o bien es creciente o bien es decreciente.
- Debido a la segunda condición podemos calcular la confianza a través de la función de probabilidad (ya que la distribución es conocida):
  - $P(g^{-1}(a, X) \leq \vartheta \leq g^{-1}(b, X))=(1-\alpha)$

Estimación por intervalos: método del pivote

- Dado $P(a \leq g(\vartheta, X) \leq b) = (1-\alpha)$, notar que **los valores a y b que la verifican en general no son únicos.**
- La elección se hace generalmente buscando que el **intervalo tenga la máxima precisión**, es decir, la longitud mínima que caiga la máxima densidad de probabilidad.
- Esta elección depende de la distribución del estadístico pivote.
- Para distribuciones simétricas y unimodales (distribución Normal o t de Student, por ejemplo) se consigue tomando el intervalo centrado, es decir, dejando una probabilidad de $\alpha/2$ a cada lado, como hemos visto en la figura anterior.

Estimación por intervalos: principales estadísticos pivote

- **Intervalos para medias de poblaciones normales (varianza conocida de la población):**
  - Para poblaciones normales que se conoce la varianza, σ,ya sabemos cual es el estadístico pivote para el cálculo de intervalos de confianza para la media μ :
    - $z = (\bar{x} - \mu)/(\sigma/\sqrt{n})$, en este caso z se distribuye como una N(0,1).
    - $P(-z_{\alpha/2} \leq z \leq z_{\alpha/2}) =$

$$P(-z_{\alpha/2} \leq (\bar{x} - \mu)/(\sigma/\sqrt{n}) \leq z_{\alpha/2}) = 1 - \alpha$$

    - Donde el percentil $z_{\alpha/2}$ es un valor de la normal estándar tal que:
    - $P(z > z_{\alpha/2}) = 1 - \Phi(z_{\alpha/2}) = \alpha/2$, siendo Φ la función distribución normal.
    - Así los intervalos de confianza para μ a un nivel de confianza $(1 - \alpha)$ serán :
    - $\boxed{\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$

Estimación por intervalos: principales estadísticos pivote

- **Intervalos para medias de poblaciones normales (varianza conocida de la población):**
  - Recordar que los valores que se han escogido son simétricos para generar el intervalo más corto posible:
    - En el intervalo más corto posible que caiga la máxima densidad de probabilidad.
  - Para la distribución de confianza despejamos de $z = (\bar{x} - \mu)/(\sigma/\sqrt{n})$, $\mu = \bar{x} + z(\sigma/\sqrt{n})$, así al variar α, como $\bar{x}$ es constante, la única variable aleatoria es z.
  - Así, la distribución generada es normal, con media $\bar{x}$ y varianza $\sigma/\sqrt{n}$.
  - Esta distribución resume la incertidumbre existente respecto al valor desconocido μ.

Estimación por intervalos: principales estadísticos pivote

- Cual es el mínimo tamaño de muestra necesario para obtener una precisión dada para estimar el la media en un intervalo de confianza dado. **Vamos a hacer un ejemplo**.
- Supongamos que la altura de los individuos de cierta población sigue una distribución N(μ, 7,5) estando las unidades en cm. Hallar el mínimo tamaño muestral necesario para estimar la altura media con un error inferior a 2 cm y con una confianza del 90%.
- Lo que queremos es que $|\bar{x} - \mu| < 2$ y como queremos una confianza del 90% entonces $z_{\alpha/2} = z_{0.05} = 1{,}645$ (tabla libro, pag 618, el área mas cercana está entre 0,94950 que corresponde a 1,64 y 0.95053 que corresponde a 1,65: así elegimos $1{,}645$).
- Así sabemos que el intervalo de confianza de para la media de la altura viene determinado por: $\bar{x} - 1{,}645\frac{7{,}5}{\sqrt{n}} \leq \mu \leq \bar{x} + 1{,}645\frac{7{,}5}{\sqrt{n}}$ (el valor de 1,645 está calculado de manera más exacta).
- Una expresión equivalente es: $-1{,}645\frac{7{,}5}{\sqrt{n}} \leq \bar{x} - \mu \leq 1{,}645\frac{7{,}5}{\sqrt{n}}$
- O lo que es lo mismo: $|\bar{x} - \mu| \leq 1{,}645\frac{7{,}5}{\sqrt{n}}$ (recordar que $|\bar{x} - \mu| < 2$, así $1{,}645\frac{7{,}5}{\sqrt{n}}$ tiene que ser <2)
- Sustituyendo por su valor $1{,}645\frac{7{,}5}{\sqrt{n}} < 2$, y despejando n obtenemos n>38,05
- Es decir con una muestra de n=39, el 90% de las veces el error en la estimación de la media de la altura para la población a través de la muestra es menor que 2 cm.

Estimación por intervalos: principales estadísticos pivote

- **Intervalos para medias de poblaciones normales (varianza desconocida de la población):**
  - Para este caso se utiliza una variable de Student para estimar el intervalo de confianza de μ: $t = (\bar{x} - \mu)/(\hat{s}/\sqrt{n})$ **con n-1 grados de libertad.**
  - Recordad que $\hat{s}$ es el estimador varianza corregido para la muestra.
  - Además, el estadístico obtenido no depende de μ, y es función monótona de μ (esto es justo lo que necesitamos para el método del pivote), por lo tanto podemos decir que:
    - $P(-t_{\alpha/2} \leq t \leq t_{\alpha/2}) =$
      $$P(-t_{\alpha/2} \leq (\bar{x} - \mu)/(\hat{s}/\sqrt{n}) \leq t_{\alpha/2}) = 1 - \alpha$$
    - Donde el percentil $t_{\alpha/2}$ es un valor de una t de Student tal que cumple:
    - $P(t > t_{\alpha/2}) = \alpha/2$, y por tanto el intervalo de confianza para μ a un nivel de confianza $(1 - \alpha)$ será:
    - $\boxed{\bar{x} - t_{\alpha/2} \frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2} \frac{\hat{s}}{\sqrt{n}}}$ (Ver ejemplo 8.2)

Estimación por intervalos: principales estadísticos pivote

- **Ejemplo (8-1)**: En una empresa se anunciado que los salarios el año pasado crecieron un promedio del 3,5%. Un grupo de trabajadoras toma una muestra de los incrementos que han recibido una muestra de 10 mujeres obteniendo los siguientes incrementos: 3%, 3%, 5%, 1%, 1%, 2%, 1%, 1,5%, 2%, 2%. Construir un intervalo de confianza para el incremento medio experimentado por la remuneración de las mujeres en esta empresa. ESTE EJEMPLO ESTA MAL en el libro, la media y dispersión están mal calculadas.
- La media de los incrementos es $\bar{x} = 2{,}15$ y la desviación típica de los incrementos es $\hat{s} = 1{,}2483$ (desviación corregida, n-1).
- Para un intervalo de confianza del 95% se requiere el percentil de la distribución *t* de Student con 9 grados de libertad, $t_{0,975}$, que es 2,26 (tabla 5 del libro, pag 619).
- Por lo tanto el intervalo es $2{,}15 - 2{,}26\frac{1{,}2483}{\sqrt{10}} \leq \mu \leq 2{,}15 + 2{,}26\frac{1{,}2483}{\sqrt{10}}$
- Saliendo un intervalo de $1.2579 \leq \mu \leq 3.0421$, como este intervalo no incluye el 3,5% como valor posible, se puede concluir que existe una fuerte evidencia (al 95%) de que las mujeres han recibido un **incremento salarial menor que la media de los trabajadores.**

Estimación por intervalos: principales estadísticos pivote

- Intervalo para varianzas de poblaciones normales:
  - Para construir un intervalo de confianza para la estimación de la varianza de una población normal, tenemos en cuenta que
  - $\frac{ns^2}{\sigma^2} = \frac{(n-1)\hat{s}^2}{\sigma^2}$ se distribuye como χ^2_{n-1}.
  - Por lo tanto, determinando dos valores χ^2_a y χ^2_bque dejen entre si $(1-\alpha)$ de la distribución χ^2_{n-1} ya lo tenemos:
  - $P(\chi^2_a \leq \frac{ns^2}{\sigma^2} \leq \chi^2_b) = 1-\alpha$, o expresado como $P(1/\chi^2_a \geq \frac{\sigma^2}{ns^2} \geq 1/\chi^2_b) = 1-\alpha$
  - Por tanto podemos encontrar el intervalo para estimar σ^2 a un nivel de confianza $(1-\alpha)$ de la siguiente forma:
  - $\boxed{ns^2/\chi^2_a \geq \sigma^2 \geq ns^2/\chi^2_b}$ o
    $\boxed{(n-1)\hat{s}^2/\chi^2_a \geq \sigma^2 \geq (n-1)\hat{s}^2/\chi^2_b}$ (Ver ejemplo 8.2)
  - **Existen muchos más estadísticos pivote que pueden consultar en el capítulo 8 del libro.**