Cap. 6 : Estimación de parámetros

Alexandre Blondin Massé

Departamento de Informática y Matematica
Université du Québec à Chicoutimi

19 de junio del 2015
Modelado de sistemas aleatorios
Ingeniería de sistemas, producción y ambiental

Tabla de contenidos

Estimación de parámetros

- En este capítuolo, se considera una **muestra** de n variables aleatorias
$$(X_1, X_2, \ldots, X_n)$$
que siguen una **distribución conocida**;
- Sin embargo, los **parámetros** de la distribución son **desconocidos**;
- **Ejemplo:** Una muestra de una **distribución de Poisson** cuya **media** λ es desconocida (**1 parámetro**);
- **Otro ejemplo:** Una muestra de una **distribución normal** cuyas **media** μ y **varianza** σ^2 son desconocidas (**2 parámetros**).

Estimación puntual o por intervalo

Estudiamos **2 típos** de estimadores de un parámetro θ : **puntual** y **por intervalo**.

- **Puntual**. Se asigna una **cantidad exacta** a θ;
- **Por intervalo**. Se da
  1. un **intervalo** $[a, b]$ en el que pensamos que se encuentra el parámetro θ et
  2. un **nivel de confianza** que θ esté en $[a, b]$.

Definición

Definición

Cualquiera estadística que se utiliza para ***estimar*** *el valor de un* ***parámetro desconocido*** θ *se llama* ***estimador****. El valor del* ***estimador*** *se llama* ***estimación****.*

- Por ejemplo, un estimador de la **media de una población normal** es la **media muestral**

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

- Si tenemos una muestra de tamaño $n = 3$ de tal manera que $X_1 = 2$, $X_2 = 3$ et $X_3 = 4$, entonces el valor $\overline{X} = 3$ es una **estimación** de la media de la población.

Tabla de contenidos

Ejemplo (1/2)

- Suponemos que una muestra X_1, X_2, ..., X_n tiene una distribución **conjunta exponencial** de **media desconocida θ**.
- Entonces

$$f_{X_i}(x) = \lambda e^{-\lambda x} = \frac{1}{\theta} e^{-x/\theta}, \quad \text{para } i = 1, 2, \ldots, n,$$

porque $\lambda = 1/\theta$.

- **Nota**: si X y Y son dos variables **aleatorias independentes** de densidades f_X y f_Y, entonces sus **densidades individuales** son

$$f(x, y) = f_X(x) f_Y(y).$$

Ejemplo (2/2)

- Si los X_i son independientes, la **densidad conjunta** es

$$f(x_1, x_2, \ldots, x_n)$$

Ejemplo (2/2)

- Si los X_i son independientes, la **densidad conjunta** es

$$f(x_1, x_2, \ldots, x_n) \quad = \quad f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

Ejemplo (2/2)

- Si los X_i son independientes, la **densidad conjunta** es

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdots f_{X_n}(x_n) \\ &= \frac{1}{\theta} e^{-x_1/\theta} \cdot \frac{1}{\theta} e^{-x_2/\theta} \cdots \frac{1}{\theta} e^{-x_n/\theta} \end{aligned}$$

Ejemplo (2/2)

- Si los X_i son independentes, la **densidad conjunta** es

$$
\begin{aligned}
f(x_1, x_2, \ldots, x_n) &= f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdots f_{X_n}(x_n) \\
&= \frac{1}{\theta} e^{-x_1/\theta} \cdot \frac{1}{\theta} e^{-x_2/\theta} \cdots \frac{1}{\theta} e^{-x_n/\theta} \\
&= \frac{1}{\theta} \exp(-x_1/\theta) \cdot \frac{1}{\theta} \exp(-x_2/\theta) \cdots \frac{1}{\theta} \exp(-x_n/\theta)
\end{aligned}
$$

Ejemplo (2/2)

- Si los X_i son independientes, la **densidad conjunta** es

$$\begin{aligned}
f(x_1, x_2, \ldots, x_n) &= f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdots f_{X_n}(x_n) \\
&= \frac{1}{\theta} e^{-x_1/\theta} \cdot \frac{1}{\theta} e^{-x_2/\theta} \cdots \frac{1}{\theta} e^{-x_n/\theta} \\
&= \frac{1}{\theta} \exp(-x_1/\theta) \cdot \frac{1}{\theta} \exp(-x_2/\theta) \cdots \frac{1}{\theta} \exp(-x_n/\theta) \\
&= \frac{1}{\theta^n} \exp\left(-\frac{x_1 + x_2 + \cdots + x_n}{\theta}\right)
\end{aligned}$$

Ejemplo (2/2)

- Si los X_i son independientes, la **densidad conjunta** es

$$
\begin{aligned}
f(x_1, x_2, \ldots, x_n) &= f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdots f_{X_n}(x_n) \\
&= \frac{1}{\theta} e^{-x_1/\theta} \cdot \frac{1}{\theta} e^{-x_2/\theta} \cdots \frac{1}{\theta} e^{-x_n/\theta} \\
&= \frac{1}{\theta} \exp(-x_1/\theta) \cdot \frac{1}{\theta} \exp(-x_2/\theta) \cdots \frac{1}{\theta} \exp(-x_n/\theta) \\
&= \frac{1}{\theta^n} \exp\left(-\frac{x_1 + x_2 + \cdots + x_n}{\theta}\right)
\end{aligned}
$$

- **Propósito**: estimar el parámetro θ a partir de los valores x_1, x_2, ..., x_n utilizando la **densidad conjunta**.

Máxima verosimilitud

La densidad conjunta $f(x_1, x_2, \ldots, x_n)$ se puede ver como una función del **parámetro desconocido θ**

Definición

*La **función de verosimilitud** del parámetro θ para una densidad conjunta $f(x_1, x_2, \ldots, x_n)$ es*

$$L(\theta) = f(x_1, x_2, \ldots, x_n).$$

*El **estimador de máxima verosimilitud (EMV)**, denotado por $\hat{\theta}$, es el valor de θ que **maximiza** $L(\theta)$.*
***Nota**: No es siempre **único**.*

Distribución de Bernoulli (1/3)

- Hacemos n experimentos **independentes** de Bernoulli, con probabilité de **éxito** p;
- ¿Cuál es el **EMV** para el **parámetro** p si tenemos una muestra X_1, X_2, ..., X_n?
- Es razonable suponer que la **proporción** de los variables X_i que son iguales a 1 en comparación con los que son iguales a 0 tenda hacia n;
- En otras palabras, esperamos obtener

$$\hat{\mathbf{p}} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

Distribución de Bernoulli (2/3)

- Primero, tenemos

$$
\begin{aligned}
P(X_i = 1) &= p \\
P(X_i = 0) &= 1 - p \\
P(X_i = x) &= \mathbf{p^x(1-p)^{1-x}}, \text{para } x = 0, 1.
\end{aligned}
$$

Distribución de Bernoulli (2/3)

- Primero, tenemos

$$
\begin{aligned}
P(X_i = 1) &= p \\
P(X_i = 0) &= 1 - p \\
P(X_i = x) &= \mathbf{p^x(1-p)^{1-x}}, \text{para } x = 0, 1.
\end{aligned}
$$

- Consecuamente, la **función de verosimilitud** es

$L(p)$

Distribución de Bernoulli (2/3)

- Primero, tenemos

$$\begin{aligned} P(X_i = 1) &= p \\ P(X_i = 0) &= 1 - p \\ P(X_i = x) &= \mathbf{p^x(1-p)^{1-x}}, \text{para } x = 0, 1. \end{aligned}$$

- Consecuamente, la **función de verosimilitud** es

$$L(p) = p^{x_1}(1-p)^{1-x_1} \cdot p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n}$$

Distribución de Bernoulli (2/3)

- Primero, tenemos

$$\begin{aligned} P(X_i = 1) &= p \\ P(X_i = 0) &= 1 - p \\ P(X_i = x) &= \mathbf{p^x(1-p)^{1-x}}, \text{para } x = 0, 1. \end{aligned}$$

- Consecuamente, la **función de verosimilitud** es

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

Distribución de Bernoulli (2/3)

- Primero, tenemos

$$\begin{aligned} P(X_i = 1) &= p \\ P(X_i = 0) &= 1 - p \\ P(X_i = x) &= \mathbf{p^x(1-p)^{1-x}}, \text{para } x = 0, 1. \end{aligned}$$

- Consecuamente, la **función de verosimilitud** es

$$\begin{aligned} L(p) &= p^{x_1}(1-p)^{1-x_1} \cdot p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n} \\ &= p^{x_1+x_2+\ldots+x_n}(1-p)^{n-(x_1+x_2+\ldots+x_n)} \\ &= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}. \end{aligned}$$

- ¿Cuál es el valor de p que **maximiza** $L(p)$?

Distribución de Bernoulli (3/3)

- Para calcular el **máximo** de una función que contiene **productos**, es más fácil utilizar **logarithmo**:

$$\ln L(p) = \ln\left(p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}\right)$$

- Para calcular el **máximo** de una función que contiene **productos**, es más fácil utilizar **logarithmo**:

$$
\begin{aligned}
\ln L(p) &= \ln\left(p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}\right) \\
&= \sum_{i=1}^n x_i \ln(p) + \left(n - \sum_{i=1}^n x_i\right)\ln(1-p)
\end{aligned}
$$

Distribución de Bernoulli (3/3)

- Para calcular el **máximo** de una función que contiene **productos**, es más fácil utilizar **logarithmo**:

$$\begin{aligned}\ln L(p) &= \ln\left(p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}\right)\\ &= \sum_{i=1}^n x_i \ln(p) + \left(n - \sum_{i=1}^n x_i\right)\ln(1-p)\end{aligned}$$

- Luego, derivamos

$$\frac{d}{dp}\ln L(p) = \frac{\sum_{i=1}^n x_i}{p} - \frac{n-\sum_{i=1}^n x_i}{1-p}$$

Distribución de Bernoulli (3/3)

- Para calcular el **máximo** de una función que contiene **productos**, es más fácil utilizar **logarithmo**:

$$
\begin{aligned}
\ln L(p) &= \ln\left(p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}\right) \\
&= \sum_{i=1}^{n} x_i \ln(p) + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-p)
\end{aligned}
$$

- Luego, derivamos

$$\frac{d}{dp}\ln L(p) = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n-\sum_{i=1}^{n} x_i}{1-p}$$

- La derivada es **nula** si

$$p = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Aplicación

- Se producen **microchips RAM** en una fábrica.
- Cada microchip es **aceptable** con probabilidad p, de manera **independiente**.
- Suponemos que en una muestra de 1000 **microchips**, 921 son aceptables y 79 **son defectuos**.
- ¿Cuál es el **EMV** para p?

Aplicación

- Se producen **microchips RAM** en una fábrica.
- Cada microchip es **aceptable** con probabilidad p, de manera **independiente**.
- Suponemos que en una muestra de 1000 **microchips**, 921 son aceptables y 79 **son defectuos**.
- ¿Cuál es el **EMV** para p?
- Sea
$$X_i = \begin{cases} 1, & \text{si el microchip } i \text{ es aceptable;} \\ 0, & \text{si no es aceptable.} \end{cases}$$

Aplicación

- Se producen **microchips RAM** en una fábrica.
- Cada microchip es **aceptable** con probabilidad p, de manera **independiente**.
- Suponemos que en una muestra de 1000 **microchips**, 921 son aceptables y 79 **son defectuos**.
- ¿Cuál es el **EMV** para p?
- Sea
$$X_i = \begin{cases} 1, & \text{si el microchip } i \text{ es aceptable;} \\ 0, & \text{si no es aceptable.} \end{cases}$$
- Entonces
$$\hat{p} = \sum_{i=1}^{n} x_i/1\,000 = 921/1000 = \mathbf{0{,}921}.$$

Distribución de Poisson (1/3)

- Sean X_1, X_2, ..., X_n variables aleatorias **independentes** de **Poisson** de parámetro **desconocido** λ;
- ¿Cuál es el **EMV** para λ?
- Como λ es la **media**, esperamos que

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

- La **función de probabilidad de Poisson** es

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}.$$

- La **función de verosimilitud** del parámetro λ para una muestra de tamaño n es

$$L(\lambda) = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda}\lambda^{x_2}}{x_2!} \cdots \frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

Distribución de Poisson (2/3)

- La **función de probabilidad de Poisson** es

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}.$$

- La **función de verosimilitud** del parámetro λ para una muestra de tamaño n es

$$\begin{aligned} L(\lambda) &= \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda}\lambda^{x_2}}{x_2!} \cdots \frac{e^{-\lambda}\lambda^{x_n}}{x_n!} \\ &= \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i}}{x_1!x_2!\cdots x_n!} \end{aligned}$$

- ¿Cuál es el **EMV** $\hat{\lambda}$?

Distribución de Poisson (3/3)

- Se calcula otra vez el **logaritmo**:

$$\ln L(\lambda) = -n\lambda + \sum_{i=1}^{n} x_i \ln(\lambda) - \ln(x_1! x_2! \cdots x_n!).$$

Distribución de Poisson (3/3)

- Se calcula otra vez el **logaritmo**:

$$\ln L(\lambda) = -n\lambda + \sum_{i=1}^{n} x_i \ln(\lambda) - \ln(x_1!x_2! \cdots x_n!).$$

- Luego, **derivamos**

$$\frac{d}{d\lambda} \ln L(\lambda) = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda}.$$

Distribución de Poisson (3/3)

- Se calcula otra vez el **logaritmo**:

$$\ln L(\lambda) = -n\lambda + \sum_{i=1}^{n} x_i \ln(\lambda) - \ln(x_1! x_2! \cdots x_n!).$$

- Luego, **derivamos**

$$\frac{d}{d\lambda} \ln L(\lambda) = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda}.$$

- La derivada es **nula** cuando

$$\lambda = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Aplicación (1/2)

- Los números de **accidentes** en una ciudad dada durante 10 **días** eligidos al azar son

$$4, 0, 6, 5, 2, 1, 2, 0, 4, 3.$$

- Utilizen estos datos para **estimar la proporción** de días del año que no tienen más de 2 **accidentes**.

Aplicación (2/2)

- Es razonable suponer que el número de accidentes sigue una **distribución de Poisson**.

Aplicación (2/2)

- Es razonable suponer que el número de accidentes sigue una **distribución de Poisson**.

- También, tenemos

$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 2{,}7.$$

Aplicación (2/2)

- Es razonable suponer que el número de accidentes sigue una **distribución de Poisson**.
- También, tenemos

$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 2{,}7.$$

- Entonces $\hat{\lambda} = 2{,}7$.

Aplicación (2/2)

- Es razonable suponer que el número de accidentes sigue una **distribución de Poisson**.
- También, tenemos
$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 2{,}7.$$
- Entonces $\hat{\lambda} = 2{,}7$.
- Ahora, sea X el número de accidentes en un día dado;

Aplicación (2/2)

- Es razonable suponer que el número de accidentes sigue una **distribución de Poisson**.
- También, tenemos

$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 2{,}7.$$

- Entonces $\hat{\lambda} = 2{,}7$.
- Ahora, sea X el número de accidentes en un día dado;
- Entonces la estimación buscada es

$$P(X \leq 2)$$

Aplicación (2/2)

- Es razonable suponer que el número de accidentes sigue una **distribución de Poisson**.
- También, tenemos
$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 2{,}7.$$
- Entonces $\hat{\lambda} = 2{,}7$.
- Ahora, sea X el número de accidentes en un día dado;
- Entonces la estimación buscada es
$$P(X \leq 2) \quad = \quad P(X = 0) + P(X = 1) + P(X = 2)$$

Aplicación (2/2)

- Es razonable suponer que el número de accidentes sigue una **distribución de Poisson**.
- También, tenemos
$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 2{,}7.$$
- Entonces $\hat{\lambda} = 2{,}7$.
- Ahora, sea X el número de accidentes en un día dado;
- Entonces la estimación buscada es
$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= e^{-2{,}7}\left(1 + 2{,}7 + \frac{(2{,}7)^2}{2}\right) \end{aligned}$$

Aplicación (2/2)

- Es razonable suponer que el número de accidentes sigue una **distribución de Poisson**.
- También, tenemos
$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 2{,}7.$$
- Entonces $\hat{\lambda} = 2{,}7$.
- Ahora, sea X el número de accidentes en un día dado;
- Entonces la estimación buscada es
$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= e^{-2{,}7}\left(1 + 2{,}7 + \frac{(2{,}7)^2}{2}\right) \\ &\approx \mathbf{0{,}4936}. \end{aligned}$$

Distribución normal (1/3)

- Sean X_1, X_2, ..., X_n variables **normales** de media **desconocida** μ y de desviación estándar **desconocida** σ;
- ¿Cuál es el **EMV** para μ y σ?

Distribución normal (1/3)

- Sean X_1, X_2, $\ldots$, X_n variables **normales** de media **desconocida** μ y de desviación estándar **desconocida** σ;
- ¿Cuál es el **EMV** para μ y σ?
- Vamos a ver que

$$\begin{aligned}\hat{\mu} &= \overline{X}\\ \hat{\sigma} &= S\sqrt{\frac{n-1}{n}},\end{aligned}$$

- En otras palabras, son la **media muestral** y la **desviación estándar muestral no corrigida**.

Distribución normal (2/3)

- La **función de densidad** de una distribución normal es

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right].$$

Distribución normal (2/3)

- La **función de densidad** de una distribución normal es

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right].$$

- Entonces, la **función de verosimilitud** de los parámetros μ y σ para une muestra de tamaño n es

$$L(\mu,\sigma) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sigma^n} \exp\left[\frac{-\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right].$$

Distribución normal (2/3)

- La **función de densidad** de una distribución normal es

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right].$$

- Entonces, la **función de verosimilitud** de los parámetros μ y σ para une muestra de tamaño n es

$$L(\mu, \sigma) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sigma^n} \exp\left[\frac{-\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right].$$

- Buscamos los valores μ y σ que **maximizan** L.

Distribución normal (3/3)

- El **logaritmo** de verosimilitud es

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

Distribución normal (3/3)

- El **logaritmo** de verosimilitud es

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

- Derivando **respecto a** μ y a σ, obtenemos

$$\begin{aligned}
\frac{\partial}{\partial\mu}\ln(L(\mu,\sigma)) &= \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2}\\
\frac{\partial}{\partial\sigma}\ln(L(\mu,\sigma)) &= -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^3}.
\end{aligned}$$

Distribución normal (3/3)

- El **logaritmo** de verosimilitud es

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

- Derivando **respecto a** μ y a σ, obtenemos

$$\begin{aligned}\frac{\partial}{\partial\mu}\ln(L(\mu,\sigma)) &= \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2}\\ \frac{\partial}{\partial\sigma}\ln(L(\mu,\sigma)) &= -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^3}.\end{aligned}$$

- Entonces

$$\hat{\mu} = \sum_{i=1}^{n} x_i/n \quad \text{y} \quad \hat{\sigma} = \sqrt{\sum_{i=1}^{n}(x_i-\hat{\mu})^2/n}.$$

Distribución uniforme

- Suponemos que X_1, X_2, ..., X_n siguen una distribución uniforme en el intervalo $[0, \theta]$, donde θ **es desconocido**;

Distribución uniforme

- Suponemos que X_1, X_2, …, X_n siguen una distribución uniforme en el intervalo $[0, \theta]$, donde θ **es desconocido**;
- Entonces la función de verosimilitud es

$$L(\theta) = \begin{cases} 1/\theta^n, & \text{si } 0 < x_i < \theta \text{ para } i = 1, 2, \ldots, n; \\ 0, & \text{de otra manera.} \end{cases}$$

Distribución uniforme

- Suponemos que X_1, X_2, …, X_n siguen una distribución uniforme en el intervalo $[0, \theta]$, donde θ **es desconocido**;
- Entonces la función de verosimilitud es

$$L(\theta) = \begin{cases} 1/\theta^n, & \text{si } 0 < x_i < \theta \text{ para } i = 1, 2, \ldots, n; \\ 0, & \text{de otra manera.} \end{cases}$$

- El máximo se llega cuando θ **es pequeño**.

Distribución uniforme

- Suponemos que X_1, X_2, …, X_n siguen una distribución uniforme en el intervalo $[0, \theta]$, donde θ **es desconocido**;
- Entonces la función de verosimilitud es

$$L(\theta) = \begin{cases} 1/\theta^n, & \text{si } 0 < x_i < \theta \text{ para } i = 1, 2, \ldots, n; \\ 0, & \text{de otra manera.} \end{cases}$$

- El máximo se llega cuando θ **es pequeño**.
- Pero $\theta \geq x_1, x_2, \ldots, x_n$, de tal manera que

$$\hat{\theta} = \text{máx}\{x_1, x_2, \ldots, x_n\}.$$

Distribución uniforme

- Suponemos que X_1, X_2, ..., X_n siguen una distribución uniforme en el intervalo $[0, \theta]$, donde θ **es desconocido**;
- Entonces la función de verosimilitud es

$$L(\theta) = \begin{cases} 1/\theta^n, & \text{si } 0 < x_i < \theta \text{ para } i = 1, 2, \ldots, n; \\ 0, & \text{de otra manera.} \end{cases}$$

- El máximo se llega cuando θ **es pequeño**.
- Pero $\theta \geq x_1, x_2, \ldots, x_n$, de tal manera que

$$\hat{\theta} = \text{máx}\{x_1, x_2, \ldots, x_n\}.$$

- La media de la distribución es

$$\frac{\text{máx}\{X_1, X_2, \ldots, X_n\}}{2} \neq \overline{X}.$$

Tabla de contenidos

- Sea X_1, X_2, …, X_n una muestra de una **población normal** de **media** μ y de **varianza** σ^2;
- Sabemos que $\overline{X}$ es el **EMV** para μ;
- No esperamos que $\overline{X}$ sea **exactamente igual** a μ, pero que esté **cerca**;
- Consecuamente, en lugar de dar una **estimación puntual**, se prefiere de vez en cuando dar un **intervalo** donde se encuentra μ con un **nivel de confianza**.

Tabla de contenidos

Intervalo de confianza (1/3)

- Deseamos estimar la **media** μ de una población **normal** asumiendo que **conocemos** σ;
- Sabemos que la variable aleatoria $\overline{X}$ sigue una distribución **normal** con **media** μ y **varianza** σ^2/n;
- Buscamos un **intervalo** que contiene μ con probabilidad $0{,}95$;
- Leyendo la **tabla normal**, obtenemos

$$P\left(-1{,}96 < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < 1{,}96\right) = 0{,}95.$$

Intervalo de confianza (2/3)

- **Hallando** μ en la mitad de las desigualdades, obtenemos

$$P\left(\overline{X} - 1{,}96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1{,}96\frac{\sigma}{\sqrt{n}}\right) = 0{,}95.$$

- Si $\overline{x}$ es el vector de los **datos** en la muestra, entonces dicemos que

$$\left(\overline{x} - 1{,}96\frac{\sigma}{\sqrt{n}}, \quad \overline{x} + 1{,}96\frac{\sigma}{\sqrt{n}}\right)$$

  es un **intervalo de confianza (bilateral)** del parámetro μ à **95 %**;

Intervalo de confianza (3/3)

En general:

Definición

Sea X_1, X_2, ..., X_n una muestra aleatoria de una población normal de ***media desconocida*** μ *y de* ***desviación estándar conocida*** σ*. Sea* $\alpha \in [0, 1]$ *un número real y* z_α *el número que satisface* $P(Z > z_\alpha) = \alpha$*, donde* Z *es una variable aleatoire normal estándar.*
Entonces

$$\left(\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

es un ***intervalo de confianza (bilateral)*** *de nivel* $1 - \alpha$ *para la* ***media*** μ*.*

Ejemplo (1/2)

- Suponemos que una información se transmite por un señal con un cierto **ruido**.
- Más precisamente, cuando el valor μ se transmite a partir de **un punto** A, vuelve $\mu + N$ al punto B, donde N es una variable normal de **media** 0 y de **varianza** 4.
- Para reducir los errores, se envia el mismo valor 9 **veces**.
- Suponemos que los valores son

$$5; 8{,}5; 12; 15; 7; 9; 7{,}5; 6{,}5; 10{,}5.$$

- Calculamos un intervalo de confianza de (a) $95\,\%$ y (b) $99\,\%$ para la media μ.

Ejemplo (2/2)

- Obtenemos $\overline{x} = 81/9 = \mathbf{9}$.

Ejemplo (2/2)

- Obtenemos $\overline{x} = 81/9 = \mathbf{9}$.
- Consecuamente, el interval de confianza es

$$\left(9 - 1{,}96\frac{\sigma}{3}, \quad 9 + 1{,}96\frac{\sigma}{3}\right) \approx (7{,}69; 10{,}31).$$

Ejemplo (2/2)

- Obtenemos $\overline{x} = 81/9 = \mathbf{9}$.
- Consecuamente, el interval de confianza es

$$\left(9 - 1{,}96\frac{\sigma}{3}, \quad 9 + 1{,}96\frac{\sigma}{3}\right) \approx (7{,}69; 10{,}31).$$

- Entonces, estamos 95 % seguros que el valor esta entre 7,69 y 10,31.

Ejemplo (2/2)

- Obtenemos $\overline{x} = 81/9 = \mathbf{9}$.
- Consecuamente, el interval de confianza es

$$\left(9 - 1{,}96\frac{\sigma}{3}, \quad 9 + 1{,}96\frac{\sigma}{3}\right) \approx (7{,}69; 10{,}31).$$

- Entonces, estamos $95\,\%$ seguros que el valor esta entre $7{,}69$ y $10{,}31$.
- Ahora, como $z_{0{,}01/2} = z_{0{,}005} \approx 2{,}575$, hay que

$$\left(9 - 2{,}575\frac{\sigma}{3}, \quad 9 + 2{,}575\frac{\sigma}{3}\right) \approx (7{,}283; 10{,}717).$$

Ejemplo (2/2)

- Obtenemos $\overline{x} = 81/9 = \mathbf{9}$.
- Consecuamente, el interval de confianza es

$$\left(9 - 1{,}96\frac{\sigma}{3}, \quad 9 + 1{,}96\frac{\sigma}{3}\right) \approx (7{,}69; 10{,}31).$$

- Entonces, estamos 95 % seguros que el valor esta entre 7,69 y 10,31.
- Ahora, como $z_{0{,}01/2} = z_{0{,}005} \approx 2{,}575$, hay que

$$\left(9 - 2{,}575\frac{\sigma}{3}, \quad 9 + 2{,}575\frac{\sigma}{3}\right) \approx (7{,}283; 10{,}717).$$

- Entonces, estamos 99 % seguros que el valor esta entre 7,283 y 10,717.

Intervalo unilateral

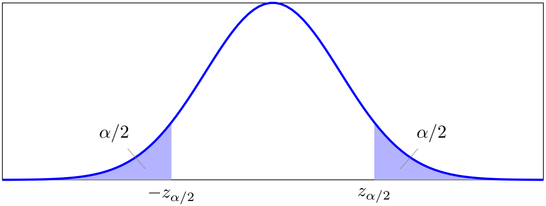
- De vez en cuando, deseamos saber si la media μ **no es mayor** que un valor dado con un **nivel de confianza**;
- En este caso, el intervalo es de la forma

$$\left(\overline{x} - z_\alpha \frac{\sigma}{\sqrt{n}}, +\infty\right) \quad \text{ou} \quad \left(-\infty, \overline{x} + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$$

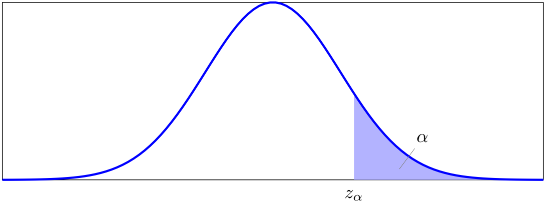
  y decimos que el intervalo es **unilateral** y que su **nivel de confianza** es $1 - \alpha$.

Gráficas

Bilateral :



Unilateral :



Tamaño de una muestra (1/2)

- De vez en cuando, el problema se **invierte**;
- Por ejemplo, se puede preguntar que **tamaño de muestra** hay que utilizar para que el **error** sea con un **nivel de confianza dado**;
- Por ejemplo, el **peso promedio** de salmones **varia** cada año, pero la **desviación estándar** es siempre $0{,}3$.
- Deseamos conocer la media con **confianza** $95\,\%$ y **error** menor que $\pm 0{,}1$.
- ¿Cuál **valor de** n tenemos que utilizar?

Tamaño de una muestra (1/2)

- Buscamos un intervalo de confianza a $95\,\%$.

Tamaño de una muestra (1/2)

- Buscamos un intervalo de confianza a 95 %.
- Sea n el tamaño de la muestra. Entonces

$$\mu \in \left(\overline{x} - 1{,}96 \frac{\sigma}{\sqrt{n}}, \quad \overline{x} + 1{,}96 \frac{\sigma}{\sqrt{n}} \right).$$

Tamaño de una muestra (1/2)

- Buscamos un intervalo de confianza a $95\,\%$.
- Sea n el tamaño de la muestra. Entonces

$$\mu \in \left(\overline{x} - 1{,}96\frac{\sigma}{\sqrt{n}}, \quad \overline{x} + 1{,}96\frac{\sigma}{\sqrt{n}}\right).$$

- Hay que

$$1{,}96\frac{\sigma}{\sqrt{n}} \leq 0{,}1,$$

es decir que $\sqrt{n} \geq 5{,}88$ y entonces $n \geq 34{,}57$.

Tamaño de una muestra (1/2)

- Buscamos un intervalo de confianza a 95 %.
- Sea n el tamaño de la muestra. Entonces
$$\mu \in \left(\overline{x} - 1{,}96\frac{\sigma}{\sqrt{n}}, \quad \overline{x} + 1{,}96\frac{\sigma}{\sqrt{n}}\right).$$
- Hay que
$$1{,}96\frac{\sigma}{\sqrt{n}} \leq 0{,}1,$$
es decir que $\sqrt{n} \geq 5{,}88$ y entonces $n \geq 34{,}57$.
- Así, una muestra de 35 **o más** es suficiente.

Tabla de contenidos

Media y varianza desconocidas (1/2)

- Sea una muestra X_1, X_2, ..., X_n de una población normal de **media desconocida** μ y de **varianza desconocida** σ^2;
- Aquí **no tenemos** el hecho que $\sqrt{n}(\overline{X} - \mu)/\sigma$ es una variable normal estándar;
- Sin embargo, sabemos que

$$\sqrt{n}\frac{\overline{X} - \mu}{S}$$

sigue una **distribución de Student** con $n - 1$ **grados de libertad**;

Media y varianza desconocidas (2/2)

- Consecuamente, para cualquier número $\alpha \in (0, 1/2)$, tenemos

$$P\left(-t_{\alpha/2,n-1} < \sqrt{n}\frac{\overline{X} - \mu}{S} < t_{\alpha/2,n-1}\right) = 1 - \alpha.$$

- Entonces,

$$P\left(\overline{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

- Así que si se obtienen $\overline{x}$ y s, entonces la media es

$$\mu \in \left(\overline{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \quad \overline{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$$

con **nivel de confianza** $1 - \alpha$.

Versión unilateral

- En el caso donde la **varianza es conocida**, tenemos versiones **unilaterales** para los intervalos.
- Podemos decir con **confianza** $1 - \alpha$ que

$$\mu \in \left(\overline{x} - \frac{s}{\sqrt{n}} t_{\alpha,n-1}, \quad +\infty\right)$$

- También,

$$\mu \in \left(-\infty, \quad \overline{x} - \frac{s}{\sqrt{n}} t_{\alpha,n-1}\right)$$

Ejemplo (1/2)

- Calculamos un intervalo de confianza a 95 % para estimar el número de latidos del corazón por minuto de 15 **personas**:

$$54, 63, 58, 72, 49, 92, 70, 73, 69, 104, 48, 66, 80, 64, 77.$$

- Se calculan la **media** y la **varianza muestrales**:

```
>>> X = [54,63,58,72,49,92,70,73,69,104,48,66,80,64,77]
>>> Xb = sum(X) * 1.0 / len(X)
>>> Xc = map(lambda x: x*x, X)
>>> print X; print Xb; print Xc
[54, 63, 58, 72, 49, 92, 70, 73, 69, 104, 48, 66, 80, 64, 77]
69.2666666667
[2916, 3969, 3364, 5184, 2401, 8464, 4900, 5329, 4761, 10816, 2304,
    4356, 6400, 4096, 5929]
>>> S2 = (sum(Xc) - len(X) * Xb**2) * 1.0 / (len(X) - 1)
>>> from math import sqrt
>>> S = sqrt(S2)
>>> print S2; print S
230.066666667
15.1679486638
```

Ejemplo (1/2)

- Luego calculamos los límites del intervalo de confianza a $95\,\%$:

```
>>> from scipy.stats import t
>>> t.interval(0.95, len(X) - 1)
(-2.1447866879169273, 2.1447866879169273)
>>> t.interval(0.95, len(X) - 1, Xb, S/sqrt(len(X)))
(60.866936673272512, 77.666396660060826)
```

- Para obtener un intervalo **unilateral**, se hace así:

```
>>> t.interval(0.90, len(X) - 1, Xb, S/sqrt(len(X)))
(62.368764111375242, 76.164569221958089)
```

de tal manera que los intervalos son

$$(-\infty, 76{,}1646) \quad \text{y} \quad (62{,}3688, +\infty).$$

Tabla de contenidos

Intervalo de confianza para la varianza

- Sea una muestra X_1, X_2, …, X_n de una población normal de **media desconocida** μ y de **varianza desconocida** σ^2;
- Es también posible estimar la **varianza** σ^2 utilizando el hecho que

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

- Entonces

$$\sigma^2 \in \left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \quad \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} \right)$$

con **confianza** $1-\alpha$.

Ejemplo (1/2)

- Supoenmos que una empresa produce **tuercas**;
- La desviación estándar debe ser **muy pequeña** para su espesor;
- Sea una muestra de 10 **tuercas** con esperores:

$$123, 124, 126, 120, 130, 133, 125, 128, 124, 126.$$

- Deseamos calcular un intervalo confianza a $90\,\%$ para la desviación estándar.

Ejemplo (1/2)

- Supoenmos que una empresa produce **tuercas**;
- La desviación estándar debe ser **muy pequeña** para su espesor;
- Sea una muestra de 10 **tuercas** con esperores:

$$123, 124, 126, 120, 130, 133, 125, 128, 124, 126.$$

- Deseamos calcular un intervalo confianza a $90\,\%$ para la desviación estándar.
- La varianza muestral es $\mathbf{S^2} \approx 1{,}366 \times 10^{-5}$;

Ejemplo (2/2)

- Tenemos $\chi^2_{0{,}05,9} \approx \mathbf{16{,}917}$ y $\chi^2_{0{,}95,9} \approx \mathbf{3{,}334}$:

```
>>> from scipy.stats import chi2
>>> chi2.ppf(0.05,9)
3.3251128430668162
>>> chi2.ppf(0.95,9)
16.918977604620448
```

Ejemplo (2/2)

- Tenemos $\chi^2_{0{,}05,9} \approx \mathbf{16{,}917}$ y $\chi^2_{0{,}95,9} \approx \mathbf{3{,}334}$:

```
>>> from scipy.stats import chi2
>>> chi2.ppf(0.05,9)
3.3251128430668162
>>> chi2.ppf(0.95,9)
16.918977604620448
```

- También

```
>>> 9 * 1.366 * 10**(-5) / chi2.ppf(0.95,9)
7.2663965206991e-06
>>> 9 * 1.366 * 10**(-5) / chi2.ppf(0.05,9)
3.6973181303107313e-05
```

Ejemplo (2/2)

- Tenemos $\chi^2_{0,05,9} \approx \mathbf{16{,}917}$ y $\chi^2_{0,95,9} \approx \mathbf{3{,}334}$:

```
>>> from scipy.stats import chi2
>>> chi2.ppf(0.05,9)
3.3251128430668162
>>> chi2.ppf(0.95,9)
16.918977604620448
```

- También

```
>>> 9 * 1.366 * 10**(-5) / chi2.ppf(0.95,9)
7.2663965206991e-06
>>> 9 * 1.366 * 10**(-5) / chi2.ppf(0.05,9)
3.6973181303107313e-05
```

- Un intervalo de confianza a 90 % para σ^2 es

$$(7{,}266 \times 10^{-6}, 36{,}973 \times 10^{-6})$$

Ejemplo (2/2)

- Tenemos $\chi^2_{0,05,9} \approx \mathbf{16{,}917}$ y $\chi^2_{0,95,9} \approx \mathbf{3{,}334}$:

```
>>> from scipy.stats import chi2
>>> chi2.ppf(0.05,9)
3.3251128430668162
>>> chi2.ppf(0.95,9)
16.918977604620448
```

- También

```
>>> 9 * 1.366 * 10**(-5) / chi2.ppf(0.95,9)
7.2663965206991e-06
>>> 9 * 1.366 * 10**(-5) / chi2.ppf(0.05,9)
3.6973181303107313e-05
```

- Un intervalo de confianza a 90 % para σ^2 es

$$(7{,}266 \times 10^{-6}, 36{,}973 \times 10^{-6})$$

- Finalmente, un intervalo de confianza para σ es

$$(2{,}696 \times 10^{-3}, 6{,}072 \times 10^{-3}).$$

Tabla de contenidos

Diferencia de las medias

- A hora se consideran **dos poblaciones normales** de **medias μ_1 y μ_2** y de **varianzas σ_1^2 y σ_2^2**;
- Eligimos dos muestras $(X_i)_{1 \leq i \leq n}$ et $(Y_i)_{1 \leq i \leq m}$
- Deseamos estimar $\mu_1 - \mu_2$;
- Como antes, hay dos casos donde
  - Las varianzas σ_1^2 y σ_2^2 son conocidas;
  - Las varianzas σ_1^2 y σ_2^2 son desconocidas y $\sigma_1 = \sigma_2$.

Distribución de $\overline{X} - \overline{Y}$

- Primero, es fácil verificar que el **EMV** de $\mu_1 - \mu_2$ es $\overline{X} - \overline{Y}$;
- También, sabemos que

$$\begin{aligned} \overline{X} &\sim \mathcal{N}(\mu_1, \sigma_1^2/n) \\ \overline{Y} &\sim \mathcal{N}(\mu_2, \sigma_2^2/m), \end{aligned}$$

- Entonces

$$\overline{X} - \overline{Y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right).$$

- Cuando σ_1 **y** σ_2 son **conocidas**, se utiliza la **distribución normal** para estimar $\mu_1 - \mu_2$.

Tabla de contenidos

Varianzas conocidas

- Si las varianzas σ_1^2 y σ_2^2 son **conocidas**, entonces

$$\mu_1 - \mu_2 \in \left(\overline{x} - \overline{y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} < \mu_1 - \mu_2 < \overline{x} - \overline{y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \right)$$

  con **nivel de confianza** $1 - \alpha$;

- Se obtienen fórmulas similares para intervalos **unilaterales**.
- La distribución es mas **compleja** cuando las varianzas son **desconocidas**.

Tabla de contenidos

- En el caso donde σ_1^2 y σ_2^2 son **desconocidas**, es natural estimarlas con S_1 y S_2;
- Hay que conocer la distribución de

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n + S_2^2/m}}.$$

- Desafortunadamente, esta distribución **depiende de σ_1 y σ_2**;
- Sin embargo, en el caso **special** donde $\sigma_1 = \sigma_2$, es posible dar una **estimación por intervalo**.

El caso $\sigma_1 = \sigma_2$

- Se puede demostrar que

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{S_p^2(1/n + 1/m)}}$$

sigue una **distribución de Student** de $n + m - 2$ **grados de libertad**, donde S_p^2 satisface

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}.$$

- Consecuamente,

$$\mu_1 - \mu_2 \in \left(\overline{x} - \overline{y} - t_{\alpha/2,n+m-2} s_p \sqrt{1/n + 1/m}, \quad \overline{x} - \overline{y} + t_{\alpha/2,n+m-2} s_p \sqrt{1/n + 1/m}\right)$$

con **nivel de confianza** $1 - \alpha$.