# 6. Estimación de Parámetros

## Estadística Computacional - San Joaquín - 2022

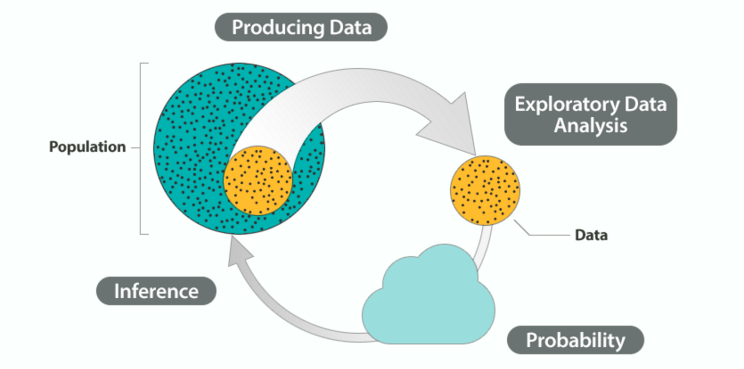
Ricardo Ñanculef, Manuel Goyo
*jnancu@inf.utfsm.cl*, *manuel.goyo@usm.cl*

Departamento de Informática UTFSM

Departamento de Informática
Universidad Técnica Federico Santa María

# ¿Cómo estudiaremos la Estadística?

- 3 Partes: Estadística Descriptiva, Teoría de Probabilidades, Inferencia.
- Obtener conclusiones acerca de un fenómeno aleatorio desde **datos** experimentales.



# Inferencia Estadística

## Objetivo de la Inferencia

Obtener conclusiones acerca de un fenómeno aleatorio a partir de **datos** experimentales.

1. Ejemplo: ¿Un cierto equipo durará más de $T = 2$ años antes de fallar?
2. Ejemplo: ¿Cuál es la temperatura máxima a la que se calienta el dispositivo?

## Objetivo de la Teoría de Probabilidades

Obtener conclusiones acerca de un fenómeno aleatorio a partir de un **modelo** para el fenómeno.

1. Ejemplo: Si el tiempo que dura el equipo se denota por $X$ y asumimos que un modelo razonable para $X$ es $X \sim \text{Poi}(2)$ la pregunta anterior (1) se traduce en calcular $p = P(X > 2) = 1 - F_{\text{Poi}}(2)$.

# Inferencia Estadística

## Método Clásico de Inferencia

1. Construir un modelo del fenómeno a partir de los datos y luego obtener conclusiones a partir del modelo.
2. Más concretamente, elegimos una familia de modelos y luego encontramos el modelo de la familia que mejor explica los datos observados.

## Ejemplo

Para modelar el tiempo $X$ que dura el equipo asumimos que $X$ una v.a. de la "familia" Poisson, es decir, $T \sim \text{Poi}(\theta)$, con $\theta$ libre, desconocido (por esto estamos especificando la familia, no el modelo específico). Usamos los datos recogidos para estimar $\theta$.

# Estimación de Parámetros

## Parámetro

Característica desconocida cuyo valor nos interesa conocer o describir. Tradicionalmente lo denotamos $\theta$.

## Muestra Aleatoria

Conjunto de datos experimentales disponibles $S_n = \{X_1, X_2, \ldots X_n\}$ obtenido de modo aleatorio simple.

- Aleatorio simple (*IID*):
  - Las v.a. $X_i$ son independientes entre sí.
  - Cada $X_i$ tiene la misma distribución de probabilidad (fdp): $X_i \sim f_X(x)$.

## Estimador

Función de $S_n$ que nos permite estimar el valor de $\theta$ a partir de los datos que recogemos. Tradicionalmente lo denotamos $\hat{\theta}_n$.

# Ejemplo

## Parámetro $\theta$

Si para modelar el tiempo $X$ que dura el equipo asumimos que $X \sim \text{Poi}(\theta)$, con $\theta$ libre, desconocido, el **parámetro** por estimar es justamente $\theta$.

## Muestra Aleatoria

Duración de una serie de equipos elegidos al azar del tipo que estamos estudiando $S_n = \{X_1, X_2, \ldots X_n\}$.

## Estimador

Proponemos estimar $\theta$ como $\hat{\theta}_n = \frac{1}{n} \sum_i X_i$.

# Ejemplo Estimadores

## Ejemplo

Considere las siguiente 20 observaciones del rendimiento de una maquinaria en una empresa en diferentes días:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 24.46 | 25.61 | 26.25 | 26.42 | 26.66 | 27.15 | 27.31 | 27.54 | 27.74 | 27.94 |
| 27.98 | 28.04 | 28.28 | 28.49 | 28.50 | 28.87 | 29.11 | 29.13 | 29.50 | 30.88 |

- Estimador de $\theta = \mu$: $\bar{X}_n = \sum X_i/n = 555.86/20 = 27.793$
- Estimador de $\theta = m$: $\tilde{X}_n = (27.94 + 27.98)/2 = 27.960$
- Estimador de $\theta =$ Rango: $\hat{\mathrm{R}}_n = \text{máx}(S_n) - \text{mín}(S_n) = 30.88 - 24.46 = 6.420$
- Estimador de media trucada (recortando el 10 % más pequeño y más grande de la muestra, luego proemdiar): $\bar{X}_{tr(10)} = 27.838$

# Inferencia Estadística

## Métodos

Dado un parámetro de interés $\theta$, ¿Cómo construir el estimador $\hat{\theta}_n$ sobre una muestra $S = \{X_1, X_2, \ldots X_n\}$?

- **Métodos Paramétricos:** Asumen que la v.a. de interés sigue un modelo de probabilidad $f_X(\Theta)$ que depende de parámetros desconocidos $\Theta$ que se deben estimar desde la muestra.
- **Métodos No-Paramétricos:** Intentan responder preguntas acerca de $\theta$ sin asumir un determinado modelo $f_X(\Theta)$ o un modelo suficientemente flexible. Frecuentemente basados en simulaciones computacionales intensivas.

# Método de los Momentos

# Método de los Momentos

## Idea

Un método simple para estimar las características de una distribución desconocida $f(x;\Theta)$ a partir de una muestra aleatoria $S = \{X_1, X_2, \ldots, X_n\}$ consiste en expresar el sistema de ecuaciones:

$$\mu^k = E[X^k] \overset{\text{Ecuación}}{=} \frac{1}{n}\sum_{i=1}^{n} X_i^k = \hat{\mu}^k \quad k = 1, 2 \ldots \; .$$

como función de la cantidad desconocida y resolverlo. En este sistema $\mu^k$ denota el $k$-ésimo momento **poblacional** y $\hat{\mu}^k$ el $k$-ésimo momento **muestral**.

# Método de los Momentos

## Más precisamente ...

Sea $X$ una v.a. de interés y supongamos que su distribución es $f(x;\Theta)$ con $\Theta$ desconocido. Como la distribución depende de $\Theta$, los momentos (poblacionales) de $X$ serán funciones de $\Theta$, digamos

$$\mu^k = E[X^k] = f_k(\Theta) \quad k = 1, 2, \ldots \ .$$

Supongamos ahora que disponemos de una muestra IID de $X$, $S = \{X_1, X_2, \ldots, X_n\}$. El "**Método de los Momentos**" (M.M.) para encontrar un estimador $\hat{\Theta}$ de $\Theta$ consiste en reemplazar $\Theta$ por $\hat{\Theta}$ en las ecuaciones anteriores, igualar cada ecuación con el correspondiente momento muestral $\hat{\mu}^k$

$$f_k(\hat{\Theta}) = \hat{\mu}^k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \quad k = 1, 2 \ldots \ .$$

y resolver el sistema en $\hat{\Theta}$ (despejar $\hat{\Theta}$).

# Método de los Momentos - Ejemplo

## Ejemplo

Determinemos los estimadores por el M.M. de los parámetros de una distribución $\exp(\theta)$ a partir de una muestra $S = \{X_1, X_2, \ldots, X_n\}$.

Acá el estimador que estamos interesados en saber es $\theta$. Como $X$ sigue una distribución exponencial, tenemos que $E(X) = 1/\theta$, además por el M.M. tenemos que $E(X) = \frac{1}{n}\sum_i x_i = \overline{X}$. Por tanto $\theta = \frac{1}{\overline{X}}$

# Método de los Momentos - Ejemplo

**Ejemplo**

Determinemos los estimadores M.M. de los parámetros de una distribución Nomal$(\mu, \sigma^2)$ a partir de una muestra $S = \{X_1, X_2, \ldots, X_n\}$.

Acá necesitamos estimar los parámetros $\mu$, $\sigma^2$. Como $X$ se distribuye normal, tenemos que $E(X) = \mu$ y $E(X^2) = \sigma^2 + \mu^2$. Usando M.M. tenemos que $E(X) = \overline{X}$ y además que $E(X^2) = \dfrac{1}{n}\sum_i x_i^2$, por lo que $\mu = \overline{X}$. Despejando $\sigma^2 = \dfrac{1}{n}\sum_i x_i^2 - \overline{X}^2$

# Método de los Momentos - Ejemplo

## Ejemplo

Determinemos los estimadores M.M. de los parámetros de una distribución Gamma $(\alpha, \beta)$ a partir de una muestra $S = \{X_1, X_2, \ldots, X_n\}$.

En este caso necesitamos estimar $\alpha$ y $\beta$. Como $X$ se distribuye gamma tenemos que $E(X) = \alpha\beta$ y además que $E(X^2) = \beta^2\alpha(\alpha+1) = \alpha^2\beta^2 + \beta^2\alpha$. Por M.M. tenemos que $\alpha\beta = E(X) = \overline{X}$ y que $E(X^2) = \dfrac{1}{n}\sum_i x_i^2 = \alpha^2\beta^2 + \beta^2\alpha$. Dividiendo en la ecuación anterior por $\alpha\beta$ tenemos que

$$\beta = \frac{\frac{1}{n}\sum_i x_i^2 - \overline{X}^2}{\overline{X}}$$

Sustituyendo ese valor de $\beta$ en el primer momento tenemos que

$$\alpha = \frac{\overline{X}^2}{\frac{1}{n}\sum_i x_i^2 - \overline{X}^2}$$

# Método de los Momentos

## Observaciones

- El sistema de ecuaciones no siempre se puede resolver.
- En ocasiones pueden dar lugar a estimadores sin sentido.
- Simple y con buenas propiedades.

# Método de Máxima Verosimilitud

# Estimación de Máxima Verosimilitud

## Lógica del Procedimiento

- Sea $S = \{X_1, X_2, \ldots, X_n\}$ una muestra IID de una v.a. $X$
- Formulamos cómo hipótesis que la distribución de $X$ es $f(x;\Theta)$: una f.d.p. que depende de ciertos parámetros $\Theta$.
- Como cada $X_i$ es una v.a., tiene sentido hablar de la FDP conjunta de la muestra:

$$f_S(\Theta) = f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) \ .$$

- Como $X_i$ es independiente de $X_j$, $\forall i \neq j$ y además, cada $X_i$ se distribuye idénticamente a la variable $X$ estudiada, $f_S$ es:

$$f_S(\Theta) = \prod_{i=1}^{n} f(x_i) \ .$$

- Si damos a $\Theta$ valores cercanos a los verdaderos, $f_S(\Theta)$ debería ser grande.
- Si damos a $\Theta$ valores incorrectos, $f_S(\Theta)$ debería ser pequeña.

# Estimación de Máxima Verosimilitud

## Ejemplo

Sea $S = \{X_1, X_2, \ldots, X_n\}$ una muestra aleatoria IID de la población. Supongamos que la variable que estamos midiendo $(X)$ sigue una fdp Normal $\mathcal{N}(\mu, \sigma^2)$. Entonces, $\Theta = (\mu, \sigma)$

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

Además,

$$f_S(\Theta) = \prod_{i=1}^{n} f(X_i; \Theta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left( \frac{(x_i-\mu)^2}{2\sigma^2} \right) .$$

# Estimación de Máxima Verosimilitud

## Verosimilitud

Supongamos ahora que observamos una realización especifica de la muestra $S = \{x_1, x_2, \ldots, x_n\}$. Al valor de $f_S(\Theta)$ sobre esta realización especifica de la muestra la denominamos **verosimilitud** (*likelihood*) y la denotamos como $L(\Theta)$,

$$L(\Theta) := \prod_{i=1}^{n} f(x_i; \Theta) \;= P(S|\Theta).$$

Para nuestro ejemplo,

$$L(\Theta) = \prod_{i=1}^{n} f_X(x_i; \Theta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \; .$$

# Verosimilitud

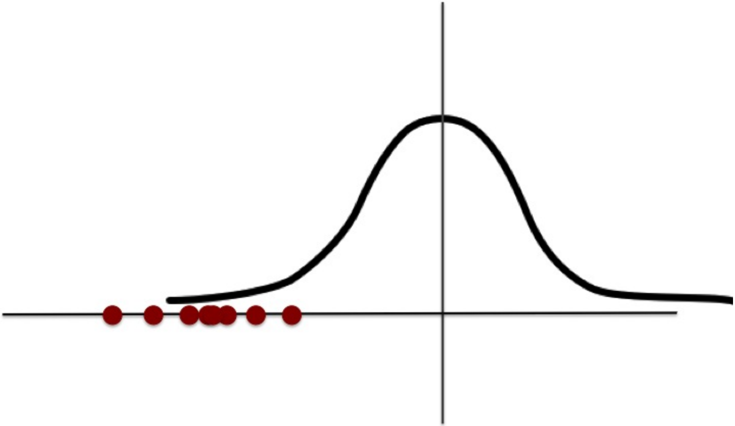
## Verosimilitud

Consideremos una realización especifica de la muestra $S = \{x_1, x_2, \ldots, x_n\}$ ...

# Verosimilitud

## Verosimilitud

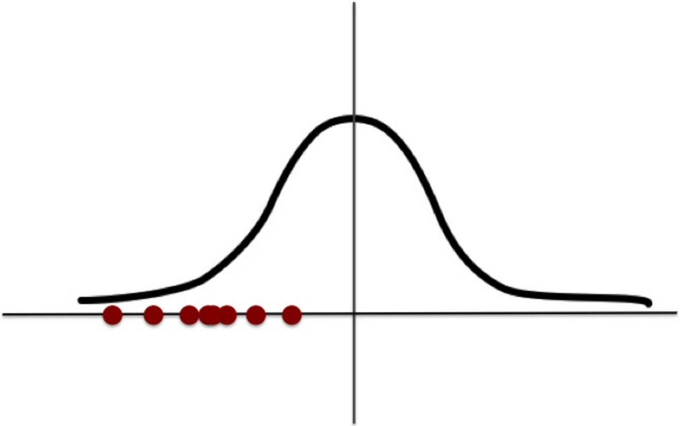
¿Cuál es la **verosimilitud** $L(\Theta)$ del modelo correspondiente a la curva negra 1 (obtenido dando ciertos valores a $\Theta$)?...



# Verosimilitud

## Verosimilitud

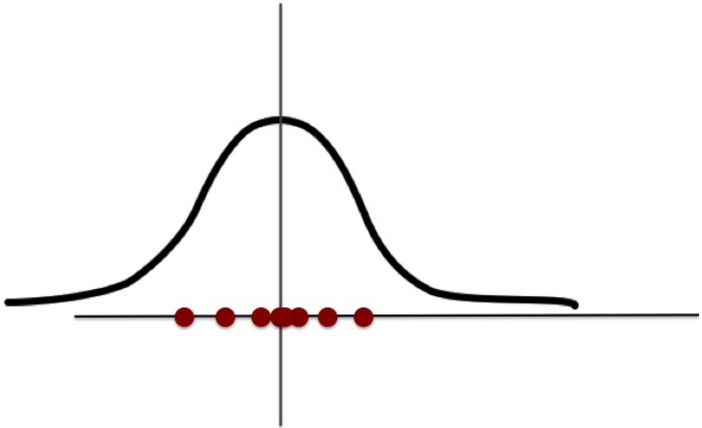
¿Cuál es la **verosimilitud** $L(\Theta)$ de otro modelo correspondiente a la curva negra 2 (obtenido dando otros valores a $\Theta$)?...



# Verosimilitud

## Verosimilitud

¿Cuál es la **verosimilitud** $L(\Theta)$ de otro modelo correspondiente a la curva negra 3 (obtenido dando otros valores a $\Theta$)?...



# Método de Máxima Verosimilitud

R. Fisher (1890-1962), Londres.
Padre de la Estadística Moderna.

**Idea ...**

- ¿Que tal si elegimos el modelo que maximiza la verosimilitud de la muestra?

# Método de Máxima Verosimilitud

## Definición

Sea $S = \{X_1, X_2, \ldots, X_n\}$ una muestra IID de una v.a. $X$ y supongamos que la f.d.p. de $X$ es $f(x; \Theta)$ con $\Theta$ desconocido. El método de máxima verosimilitud corresponde a estimar $\Theta$ para maximizar la verosimilitud $L(\Theta)$, es decir, elegir como estimador

$$\hat{\Theta}_{MV} = \arg\max_{\Theta} L(\Theta) \ .$$

## Log-verosimilitud

Como la función logaritmo preserva orden, lo anterior es equivalente a elegir como estimador

$$\hat{\Theta}_{MV} = \arg\max_{\Theta} \ell(\Theta) = \ln L(\Theta) \ .$$

# Estimación de Máxima Verosimilitud

## Ejemplo - Exponencial

Sea $S = \{X_1, X_2, \ldots, X_n\}$ una muestra aleatoria IID de la población Exp$(\theta)$ y supongamos que queremos estimar precisamente el valor esperado $E[X] = \theta$.

- La verosimilitud de la muestra (condicional al modelo que asumimos) es

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \exp\left(-\frac{x_i}{\theta}\right) \ .$$

- Y la log-verosimilitud correspondiente es:

$$\ell(S;\theta) = \log L(\theta) = -n\ln(\theta) - \sum_{i=1}^{n} \frac{x_i}{\theta} \ .$$

- Entonces,

$$\hat{\theta}_n = \arg\max_{\theta} \ell(S;\theta) = \arg\max_{\theta} - n\ln(\theta) - \sum\nolimits_{i=1}^{n} \frac{x_i}{\theta}$$

- ¿Cómo encontramos el valor de $\theta$ que maximiza esta función?

## ¿Cómo encontrar máximos?

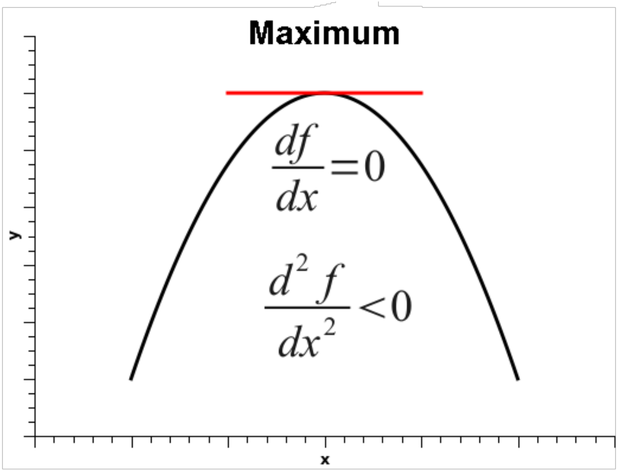
**Teorema**

Sea $\ell(\theta)$ una función diferenciable en $\theta$. Si el gradiente de $\ell(S;\theta)$ se anula en $\hat{\theta}^{\star}$, i.e.

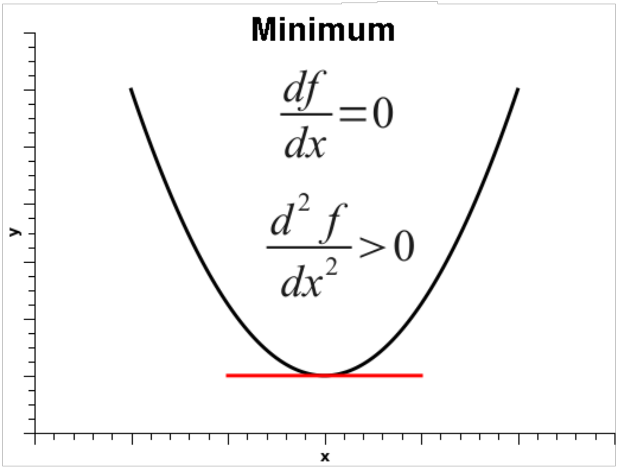
$$\left.\frac{\partial \ell(S;\theta)}{\partial \theta}\right|_{\theta^{\star}} = 0 \ .$$

- Entonces $\hat{\theta}^{\star}$ es un **máximo local** de $\ell(\theta)$. Si además, la función admite un único máximo, éste debe ser $\hat{\theta}^{\star}$. Un modo de asegurar que la función tiene un único máximo es mostrar que ésta es cóncava i.e. la segunda derivada es siempre negativa.

# ¿Cómo encontrar máximos?



## ¿Cómo encontrar máximos?



# Estimación de Máxima Verosimilitud

## Ejemplo - Exponencial

- Si encontramos $\hat{\theta}$ que satisfaga las condiciones anteriores tendremos el estimador candidato a ser el máximo verosímil (máximo local).
- La condición de primer orden es:

$$\left.\frac{\partial \ell(S;\theta)}{\partial \theta}\right|_{\hat{\theta}} = 0$$

- Derivando $\ell(\cdot)$ en $\theta$ tenemos,

$$\frac{\partial}{\partial \theta}\left(-n\ln(\theta) - \sum_{i=1}^{n}\frac{x_i}{\theta}\right) = \frac{-n}{\theta} + \sum_{i=1}^{n}\frac{x_i}{\theta^2} \ .$$

$$\left.\frac{\partial \ell(S;\theta)}{\partial \theta}\right|_{\hat{\theta}} = 0 \Rightarrow \hat{\theta}_n = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X} \ .$$

# Estimación de Máxima Verosimilitud

## Ejemplo - Exponencial

- Para concluir que se trata de un máximo debemos chequear la segunda derivada

$$\left.\frac{\partial^2 \ell(S;\theta)}{\partial\theta^2}\right|_{\hat{\theta}} < 0$$

- Derivando por segunda vez en $\theta$ tenemos,

$$\begin{aligned}\frac{\partial^2}{\partial\theta^2}\left(-n\ln(\theta) - \sum_{i=1}^{n}\frac{x_i}{\theta}\right) &= \frac{\partial}{\partial\theta}\left(\frac{-n}{\theta} + \sum_{i=1}^{n}\frac{x_i}{\theta^2}\right) \\ &= \left(\frac{n}{\theta^2} - \frac{\sum_{i=1}^{n} x_i}{2\theta^3}\right) = \left(\frac{n}{\theta^2} - \frac{n\bar{X}}{2\theta^3}\right) .\end{aligned}$$

- Evaluando en $\hat{\theta}_n = \bar{X}$

$$\left.\left(\frac{n}{\theta^2} - \frac{n\bar{X}}{2\theta^3}\right)\right|_{\hat{\theta}} = \left(\frac{n}{\bar{X}^2} - \frac{n\bar{X}}{2\bar{X}^3}\right) = -\frac{n}{\bar{X}^2} < 0$$

# Estimación de Máxima Verosimilitud

## Ejemplo - Exponencial

- Por lo tanto, el estimador máximo verosímil de $\theta = E[X]$ para una población exponencial es:

$$\hat{\theta}_n = \bar{X} \ .$$

- La idea de Fisher correspondió a elegir $\hat{\theta}$ para maximizar la probabilidad de haber observado una muestra. En este caso $\hat{\theta}$ debe ser la media muestral de tamaño $n$.

# Estimación de Máxima Verosimilitud

## Ejemplo 2 - Normal

- Sea $S = \{X_1, X_2, \ldots, X_n\}$ una muestra aleatoria IID de la población. Supongamos que la variable que estamos midiendo ($X$) sigue una FDP normal $\mathcal{N}(\mu, \sigma^2)$ y que queremos estimar precisamente el valor esperado $E[X] = \mu$ y la varianza $Var[X] = \sigma^2$.
- En este caso, vimos que

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \,.$$

- Por lo tanto,

$$\begin{aligned} \hat{\mu}, \hat{\sigma}^2 &= \arg\max_{\mu,\sigma^2} L(S; \mu, \sigma^2) \\ &= \arg\max_{\mu,\sigma^2} \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \,. \end{aligned}$$

# Estimación de Máxima Verosimilitud

## Ejemplo 2 - Normal

- Utilizando la log-verosimilitud...

$$
\begin{aligned}
\hat{\mu}, \hat{\sigma}^2 &= \arg \max_{\mu, \sigma^2} \ell(S, \mu, \sigma^2) = \\
&= \arg \max_{\mu, \sigma^2} \sum_{i=1}^{n} \ln \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \\
&= \arg \max_{\mu, \sigma^2} - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} - n \ln(\sqrt{2\pi}\sigma) \\
&= \arg \min_{\mu, \sigma^2} \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} + n \ln(\sqrt{2\pi}\sigma) \ .
\end{aligned}
$$

# Estimación de Máxima Verosimilitud

## Ejemplo 2 - Normal

- Condiciones suficientes para un máximo local:

$$\left.\frac{\partial \ell(S;\theta)}{\partial \theta}\right|_{\hat{\theta}} = 0 \Rightarrow \begin{aligned} \left.\frac{\partial \ell(S;\theta)}{\partial \mu}\right|_{\hat{\mu},\hat{\sigma}^2} &= 0 \\ \left.\frac{\partial \ell(S;\theta)}{\partial \sigma^2}\right|_{\hat{\mu},\hat{\sigma}^2} &= 0 \end{aligned}$$

- Si encontramos $\hat{\mu}, \hat{\sigma}^2$ que satisfagan las condiciones anteriores tendremos un máximo local, candidato a ser el par de estimadores máximo verosímiles.

# Estimación de Máxima Verosimilitud

## Ejemplo - continuado

- Derivando la log-verosimilitud en $\mu$

$$\frac{\partial}{\partial \mu} \ell(S;\theta) = -\frac{\partial}{\partial \mu}\left(\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2} - n\ln(\sqrt{2\pi}\sigma)\right)$$
$$= +\sum_{i=1}^{n} \frac{(x_i-\mu)}{\sigma^2} \; .$$

- Si la derivada se anula en $(\hat{\mu}, \hat{\sigma}^2)$ tenemos

$$\sum_{i=1}^{n}(x_i-\hat{\mu}) = 0 \Rightarrow \hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}$$

# Estimación de Máxima Verosimilitud

## Ejemplo - Normal

- Derivando la log-verosimilitud en $\sigma^2$

$$\frac{\partial}{\partial \sigma^2} \ell(S;\theta) = -\frac{\partial}{\partial \sigma^2} \left( \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2} - n \ln(\sqrt{2\pi}\sigma) \right)$$
$$= \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^4} - \frac{n}{2\sigma^2}$$

- Si la derivada se anula en $(\hat{\mu}, \hat{\sigma}^2)$ tenemos

$$\sum_{i=1}^{n} \frac{(x_i-\hat{\mu})^2}{2\hat{\sigma}^4} - \frac{n}{2\hat{\sigma}^2} = 0 \Rightarrow \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i-\hat{\mu})^2}{n}$$

- Por lo tanto, el par $(\hat{\mu}, \hat{\sigma})$ es un candidato a estimador máximo verosímil de $(\mu, \sigma)$ para la distribución Normal.
- Debemos chequear ahora que la matriz de segundas derivadas es **definida negativa**.

# Estimación de Máxima Verosimilitud

## Ejemplo - Normal

- Es fácil deducir que

$$\left[ \begin{array}{cc} \frac{\partial^2}{\partial \mu^2}\ell(\cdot) & \frac{\partial^2}{\partial \mu \partial \sigma^2}\ell(\cdot) \\ \frac{\partial^2}{\partial \sigma^2 \partial \mu}\ell(\cdot) & \frac{\partial^2}{\partial (\sigma^2)^2}\ell(\cdot) \end{array} \right]\Bigg|_{\hat{\mu},\hat{\sigma}^2} = \left[ \begin{array}{cc} \frac{-n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{-n}{2\hat{\sigma}^4} \end{array} \right]$$

- ¿Es esta matriz definida negativa?

## Teorema

Sea $M \in \mathcal{M}(2 \times 2)$ (una matrix de 2 por 2)

$$M = \left[ \begin{array}{cc} a & b \\ c & d \end{array} \right] .$$

Si $a < 0$ y $\det(M) > 0$, entonces, $M$ es definida negativa.

# Estimación de Máxima Verosimilitud

## Ejemplo - Normal

- Nuestra matriz

$$M = \left[ \begin{array}{cc} \frac{\partial^2}{\partial \mu^2}\ell(\cdot) & \frac{\partial^2}{\partial \mu \partial \sigma^2}\ell(\cdot) \\ \frac{\partial^2}{\partial \sigma^2 \partial \mu}\ell(\cdot) & \frac{\partial^2}{\partial (\sigma^2)^2}\ell(\cdot) \end{array} \right] \Bigg|_{\hat{\mu},\hat{\sigma}^2} = \left[ \begin{array}{cc} \frac{-n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{-n}{2\hat{\sigma}^4} \end{array} \right]$$

  es tal que $\frac{-n}{\hat{\sigma}^2} < 0$ y $det(M) = \frac{n^2}{2\hat{\sigma}^6} > 0$.
- Por lo tanto, $M$ es definida negativa.

# Estimación de Máxima Verosimilitud

## Ejemplo - Normal

- Conclusión: El par

$$\hat{\mu}^2 = \frac{\sum_{i=1}^{n} x_i}{n}$$
$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_i - \hat{\mu})^2}{n}$$

representa un máximo de $\ell(\cdot)$.

- Es posible mostrar que $\ell(\cdot)$ es cóncava, por lo que el par anterior corresponde a los estimadores máximo verosímiles de la distribución Normal.

# Estimación de Máxima Verosimilitud

## Tarea

Calcular los estimadores máximo verosímiles de:

- de $\lambda$ en una distribución de Poisson.
- de $(\alpha, \beta)$ en una distribución Gamma.
- de $b$ en una distribución Uniforme $U(0, b)$.
- de $p$ en una distribución Binomial $Bin(n, p)$.

# Inferencia Estadística

## Preguntas básicas

1. ¿Tiene sentido el estimador anterior?
2. ¿Cómo obtener buenos estimadores de $\theta$ a partir de la muestra?
3. ¿Qué es un buen estimador? Propiedades deseables?
4. ¿Cuál es la distribución de probabilidad de $\hat{\theta}_n$?
5. ¿Cuál es probabilidad de que nuestro estimador $\hat{\theta}_n$ este "cerca" (vecindad $\pm\epsilon$) de $\theta$?

# Propiedades de los Estimadores

# Formalización del Problema

R. Fisher (1890-1962), Londres.
Padre de la Estadística Moderna.

**Población de Interés**

Sea $X : \Omega \to \mathbb{R}^d$ una v.a. que representa una variable de interés sobre la población o fenómeno que estudiamos (peso, tiempo, temperatura, etc). Sea $f_X(x)$ la f.d.p. de $X$. En la práctica esta función es desconocida.

# Formalización del Problema

## Definición (Parámetro)

Un parámetro $\theta$ es cualquier función de $f_X(x)$ que nos gustaría conocer, i.e., una característica desconocida de $f_X(x)$, como su máximo, valor medio, mínimo, varianza, etc. Usualmente es una cantidad que individualiza un modelo dentro de una familia, como en el ejemplo anterior, pero en general es cualquier función de $f_X(x)$.

## Definición (Muestra Aleatoria IID)

Una muestra IID $S = \{X_1, X_2, \ldots X_n\}$ de $X$ es formalmente una colección de $n$ variables aleatorias independientes cuya distribución es $f_X(x)$. Representa lo que obtenemos cuando tomamos una muestra de la población y medimos $X$ sobre cada elemento de la muestra.

## Definición (Estimador)

Una función de $S = \{X_1, X_2, \ldots X_n\}$, digamos $\hat{\theta}_n = T(X_1, \ldots, X_n)$ que permite estimar el valor verdadero y desconocido de $\theta$.

# Formalización del Problema

R. Fisher (1890-1962), Londres.
Padre de la Estadística Moderna.

## Distribución de Muestreo

Dada una muestra IID $S = \{X_1, X_2, \dots X_n\}$ de $X \sim f_X$ y un estimador $\hat{\theta}_n = T(X_1, \dots, X_n)$, la distribución de probabilidad de $\hat{\theta}_n$ se denomina, *distribución de muestreo de* $\hat{\theta}_n$. Preguntas básicas

1. ¿$E(\hat{\theta}_n) = \theta$?
2. ¿$P(|\hat{\theta}_n) - \theta| \geq \epsilon$?
3. ¿$\mathrm{Var}(\hat{\theta}_n)$?

## Inferencia Estadística

### Teorema

Si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces,

$$\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}) \; .$$

### Teorema (Ley de los Grandes Números)

Aunque el supuesto $X \sim \mathcal{N}(\mu, \sigma^2)$ no se cumpla, $\forall \epsilon$

$$P\left(|\bar{X}_n - \mu| \geq \epsilon\right) \underset{n \to \infty}{\longrightarrow} 0 \; .$$

### Teorema (del Límite Central)

Aunque el supuesto $X \sim \mathcal{N}(\mu, \sigma^2)$ no se cumpla, si $\text{Var}(X) = \sigma^2 < \infty$

$$\bar{X}_n \underset{n \to \infty}{\xrightarrow{\mathcal{D}}} \mathcal{N}(\mu, \frac{\sigma^2}{n}) \; .$$

# Propiedades de un Estimador

## ECM, Sesgo y Varianza

Consideremos un cierto parámetro $\theta$ y un estimador propuesto $\hat{\theta}$. Definimos

- **Error Cuadrático Medio**:

$$\mathrm{EC}(\hat{\theta}) = E\left[\hat{\theta} - \theta\right]^2 .$$

- **Sesgo (Bias)**:

$$\mathrm{Bias}(\hat{\theta}) = E\left[\hat{\theta} - \theta\right] = E\left[\hat{\theta}\right] - \theta .$$

- **Varianza**:

$$\mathrm{Var}(\hat{\theta}) = E\left(\hat{\theta} - E\left[\hat{\theta}\right]\right)^2 .$$

- Nos interesa que nuestros estimadores minimicen estas medidas de "error".

# Propiedades de un Estimador

**Definición**

Sea $\theta$ un parámetro y $\hat{\theta}$ un estimador propuesto. Entonces, si $\text{Bias}(\hat{\theta}) = 0$, decimos que el estimador es **Insesgado**.

**Definición**

Sea $\theta$ un parámetro y $\hat{\theta}_1, \hat{\theta}_2$ dos estimadores propuestos. Decimos que $\hat{\theta}_1$ es **más Eficiente** que $\hat{\theta}_2$ si

$$\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2) \ .$$

# Propiedades de un Estimador

## Objetivos

- En general, entre un estimador sesgado y uno insesgado, ambos con la misma varianza, preferiremos aquél insesgado.
- En general, entre dos estimadores con el mismo sesgo, preferiremos aquél más eficiente.
- ¿Qué sucede si tenemos una situación en que un estimador tiene mayor sesgo, pero es más eficiente?

## Teorema

Sea $\theta$ un parámetro y $\hat{\theta}$ un estimador propuesto. Entonces,

$$\mathsf{EC}(\hat{\theta}) = \mathsf{Bias}^2(\hat{\theta}) + \mathsf{Var}(\hat{\theta}) \ .$$

# Propiedades de un Estimador

## Ejemplo

Supongamos que queremos estimar el valor esperado de una v.a. $\mu = E[X]$ y usamos la media muestral como estimador

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Determine el sesgo, varianza y ECM del estimador.

# Propiedades de un Estimador

## Recordando Algunas Propiedades ...

Sean $\{X_1, X_2, \ldots, X_n\}$ $n$ variables aleatorias. Entonces

$$E\left(X_1 + X_2 + \ldots X_n\right) = E(X_1) + E(X_2) + \ldots E(X_n)\ .$$

## Recordatorio de Algunas Propiedades ...

Sean $\{X_1, X_2, \ldots, X_n\}$ $n$ variables aleatorias independientes. Entonces

$$Var\left(X_1 + X_2 + \ldots X_n\right) = Var(X_1) + Var(X_2) + \ldots Var(X_n)\ .$$

# Estimación de Parámetros

## Ejemplo

Supongamos que queremos estimar el valor esperado de una v.a. $\mu = E[X]$ y usamos la media muestral como estimador

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Determine el sesgo, varianza y ECM del estimador.

## Solución

Asumiendo que la muestra $\{X_1, X_2, \ldots, X_n\}$ es IID, tenemos

$$E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu \ .$$

Por lo tanto Bias$[\bar{X}] = E\left[\bar{X} - \mu\right] = \mu - \mu = 0$

# Estimación de Parámetros

## Solución - continuación

Por otro lado, tenemos que la varianza de variables aleatorias independientes es la suma de las varianzas individuales. Si denotamos por $\sigma^2$ la varianza de $X$, tenemos

$$\mathrm{Var}[\bar{X}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}[X_i] = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n} \ .$$

Por lo tanto

$$\mathrm{Var}[\bar{X}] = \frac{\sigma^2}{n} \ .$$

Lo que nos deja,

$$\mathrm{EC}[\bar{X}] = 0^2 + \frac{\sigma^2}{n} = \frac{\sigma^2}{n} \ .$$

# Estimación de Parámetros

## Solución - comentarios

Por lo tanto, el promedio es un estimador razonable de $E[X] = \mu$ ya que su sesgo es 0 y su varianza decrece monótonamente cuando $n \to \infty$. En efecto

$$\lim_{n \to \infty} \mathrm{Var}[\bar{X}] = \lim_{n \to \infty} \frac{\sigma^2}{n} = 0 \ .$$

Chequear si estos resultados son consistentes con el Teorema de Límite Central y la Ley de los Grandes Números.

# Propiedades de un Estimador

## Teorema (Cota) de Cramer Rao

Sea $X$ una v.a. con fdp $f(x;\theta)$ donde $\theta$ es un parámetro desconocido. Sea $\hat{\theta}$ un estimador insesgado de $\theta$. Entonces, bajo ciertas condiciones de regularidad,

$$\text{Var}(\hat{\theta}) \geq \frac{1}{nE\left[\left(\frac{\partial f(x;\theta)}{\partial \theta}\right)^2\right]} = \frac{1}{-nE\left[\left(\frac{\partial^2 f(x;\theta)}{\partial \theta^2}\right)\right]}$$

## Eficiencia

Un estimador insesgado $\hat{\theta}$ de $\theta$ se dice eficiente si

$$\text{Var}(\hat{\theta}) = \frac{1}{nE\left[\left(\frac{\partial f(x;\theta)}{\partial \theta}\right)^2\right]}$$

# Propiedades de un Estimador de los Momentos

## Propiedades en el Método de los Momentos

- Sea $\hat{\Theta}_n$ un estimador MM de un parámetro $\Theta$, obtenido sobre una muestra de tamaño $n$. Entonces (consistencia),

$$\lim_{n\to\infty} P\left(\left|E[\hat{\Theta}_n] - \Theta\right| > \epsilon\right) = 0 \ .$$

- Si bien los estimadores MM son generalmente insesgados, rara vez son suficientes y eficientes.

# Propiedades de un Estimador M.V.

## Teorema

Sea $\hat{\theta}$ un estimador máximo verosímil de $\theta$ obtenido con una muestra de tamaño $n$. Entonces $\hat{\theta}$ es asintóticamente insesgado,

$$\mathrm{Bias}(\hat{\theta}_n) \underset{n\to\infty}{\longrightarrow} 0$$

## Teorema

Sea $\hat{\theta}$ un estimador máximo verosímil de $\theta$ obtenido con una muestra de tamaño $n$. Entonces, $\hat{\theta}$ es asintóticamente eficiente

$$Var(\hat{\theta}_n) \underset{n\to\infty}{\longrightarrow} \frac{1}{E\left[\left(\frac{\partial f(x;\theta)}{\partial \theta}\right)^2\right]}$$

# Estimación de Parámetros

## Ejemplo 2

Supongamos que queremos estimar la varianza de una v.a. $\sigma^2 = \text{Var}[X]$ y usamos la varianza muestral como estimador

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

Determine el sesgo, varianza y ECM del estimador.

## Solución

Hemos visto en capítulos anteriores que

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right) - \bar{X}^2 \ .$$

Por lo tanto,

$$S^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right) - \frac{n}{n-1}\bar{X}^2 = \left(\frac{1}{n-1}\sum_{i=1}^{n}X_i^2\right) - \frac{n}{n-1}\bar{X}^2 \ .$$

# Estimación de Parámetros

## Solución - continuación

Asumiendo que la muestra $\{X_1, X_2, \ldots, X_n\}$ es IID, tenemos

$$
\begin{aligned}
E[S^2] &= E\left[\frac{1}{n-1}\sum_{i=1}^{n} X_i^2\right] - \frac{n}{n-1}E\left[\bar{X}^2\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} E[X_i^2] - \frac{n}{n-1}E\left[\bar{X}^2\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} \left(\text{Var}[X_i] + E[X_i]^2\right) - \frac{n}{n-1}\left(\text{Var}[\bar{X}] + E[\bar{X}]^2\right) \\
&= \frac{1}{n-1}\sum_{i=1}^{n} (\sigma^2 + \mu^2) - \frac{n}{n-1}\left(\frac{\sigma^2}{n} + \mu^2\right) \\
&= \frac{n}{n-1}\sigma^2 + \frac{n}{n-1}\mu^2 - \frac{1}{n-1}\sigma^2 - \frac{n}{n-1}\mu^2 \\
&= \frac{n-1}{n-1}\sigma^2 = \sigma^2 \ .
\end{aligned}
$$

Por lo tanto

# Estimación de Parámetros

## Solución - continuación

Por lo tanto

$$\text{Bias}[S^2] = E\left[S^2 - \sigma^2\right] = \sigma^2 - \sigma^2 = 0 \ .$$

La varianza muestral (con el factor $n-1$ en el denominador) es un estimador insesgado de la varianza poblacional. ¿Cuál es el sesgo si dividimos por $n$ en vez de $n-1$?

## Varianza

El cálculo de la varianza de la varianza muestral es un poco engorroso, lo cual llega al siguiente resultado

$$Var[S^2] = \frac{1}{n}\left(\bar{\mu}_4 - \frac{n-3}{n-1}\sigma^4\right) \ ,$$

con $\bar{\mu}_4 = E\left[(X - E(X))^4\right]$.

Entonces $EC[S^2] = 0^2 + \frac{1}{n}\left(\bar{\mu}_4 - \frac{n-3}{n-1}\sigma^4\right) = \frac{1}{n}\left(\bar{\mu}_4 - \frac{n-3}{n-1}\sigma^4\right)$