

Estimación y Predicción en Series Temporales

Estimadores óptimos

Departamento de Procesamiento de Señales

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería

2022

- 1 (Breve) Repaso de probabilidad
- 2 Estimación de parámetros
- 3 Modelado de los datos
- 4 Estimación insesgada de mínima varianza (MVU)

Breve repaso de Probabilidad

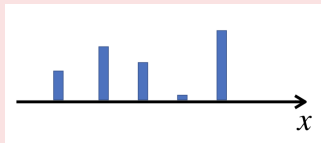
Breve repaso de Probabilidad

- **Probabilidad:**

- Caracteriza la frecuencia relativa o incertidumbre sobre la variable aleatoria X

- X –**Variable Aleatoria Discreta**

$$p_X(x_i) := \Pr(X = x_i) \quad (p_X : \text{pmf})$$



$$\sum_i \Pr(X = x_i) = 1,$$
$$\Pr(X = x_i) \geq 0$$

- **Valor Esperado**

$$\mathbb{E}(X) = \sum_i x_i p_X(x_i)$$

$$\mathbb{E}(g(X)) = \sum_i g(x_i) p_X(x_i)$$

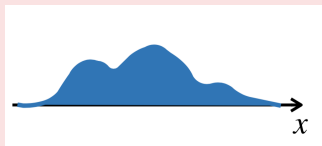
Breve repaso de Probabilidad

- **Probabilidad:**

- Caracteriza la frecuencia relativa o incertidumbre sobre la variable aleatoria X

- **X –Variable Aleatoria Continua**

$$\int_S p_X(x) = \Pr(x \in S) \quad (p_X : \text{pdf})$$



$$\int p_X(x) dx = 1, \quad p_X(x) \geq 0$$

- **Valor Esperado**

$$\mathbb{E}(X) = \int x p_X(x) dx$$

$$\mathbb{E}(g(X)) = \int g(x) p_X(x) dx$$

Distribuciones conjunta y condicional

- Regla del Producto

$$p_{X,Y}(x, y) = p_{Y|X}(y | x)p_X(x) = p_{X|Y}(x | y)p_Y(y)$$

- Regla de la Suma

$$p_X(x) = \int_y p_{X,Y}(x, y)$$

$$p_Y(y) = \int_x p_{X,Y}(x, y)$$

Breve repaso de Probabilidad

La **varianza** de una variable aleatoria X es el segundo momento central,

$$\text{var}(X) = \mathbb{E} \left[(X - \mathbb{E}(X))^2 \right].$$

Ejercicio: demostrar las siguientes propiedades

- 1 Una formulación alternativa de la varianza es,

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$$

- 2 Si X es una variable aleatoria con varianza finita, para cualquier constantes a y b se cumple que,

$$\text{var}(aX + b) = a^2 \text{var}(X)$$

- 3 Si X_1 y X_2 son variables aleatorias independientes,

$$\text{var}(X_1 + X_2) = \text{var}(X_1) + \text{var}(X_2)$$

Estimación de Parámetros

Estimación de parámetros

Planteo del Problema:

- Dadas N muestras de una señal discreta $x[n]$ que depende de cierto parámetro θ desconocido.
- Estimar θ a partir de las N muestras $x[0], x[1], \dots, x[N-1]$

Para ello se define un estimador de θ que es función de los datos:

$$\hat{\theta} = g(x[0], x[1], \dots, x[N-1])$$

- g : función a determinar
- $\hat{\theta}$: estimador de θ

Objetivo: Encontrar función g de forma que $\hat{\theta}$ sea buen estimador de θ .

- Estimador $\hat{\theta}$ debe ser cercano (en algún sentido a definir) al valor verdadero de θ .
- El criterio de cercanía debe ser especificado teniendo en cuenta que $\hat{\theta}$ es una Variable Aleatoria (función de V.As).

Modelado de los datos

Modelado de los datos

- Se dispone de un **conjunto de N datos** $x[i] \in \mathbb{R}^n$:

$$\mathcal{D} = \{x[0], x[1], \dots, x[N-1]\}$$

y un modelo que depende de un parámetro θ desconocido.

- Debido a la complejidad del fenómeno a caracterizar, **modelamos los datos estadísticamente**, mediante la **función de densidad de probabilidad o pdf**,

$$p(x[0], x[1], \dots, x[N-1]; \theta)$$

- La PDF está parametrizada por el parámetro desconocido θ , es decir define una familia de funciones.
- Puede interpretarse como que los datos son “aleatorios”
- **Notación:** se utiliza el punto y coma para denotar esa dependencia con el parámetro θ (determinístico). No confundir con una eventual densidad de probabilidad conjunta $p(x[0], \theta)$.

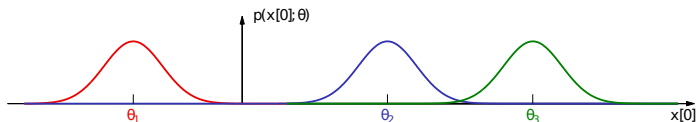
Ejemplo: PDF paramétrica (Gaussiana)

- Supongamos que el parámetro desconocido θ es la media. Si $N = 1$, los datos se modelan como:

$$x[0] \sim \mathcal{N}(\theta, \sigma^2), \quad \text{con } \sigma^2 \text{ conocido.}$$

- la PDF sería:

$$p(x[0]; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (x[0] - \theta)^2 \right]$$



- Como el valor de θ afecta la probabilidad de $x[0]$, debería ser posible inferir el valor θ a partir del valor observado de $x[0]$.
- Ejemplo.** Si el valor observado de $x[0]$ es negativo es poco probable que $\theta = \theta_3$, es más probable que $\theta = \theta_1$.

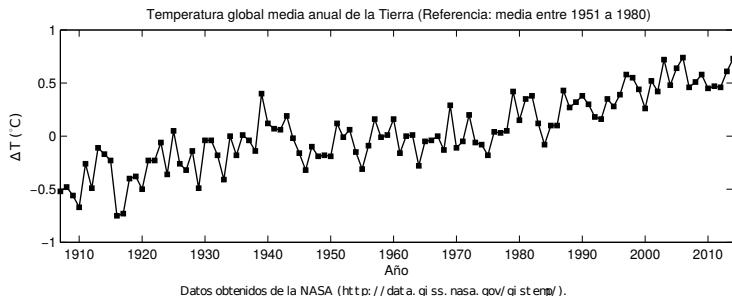
- **Especificación de la PDF** es crucial para obtener un buen estimador.
- En un problema real, la PDF de los datos no es conocida. Debe ser elegida de forma que:
 - Sea consistente con las restricciones del problema
 - Refleje el conocimiento previo de los datos (e.g., ruido Gaussiano)
 - Sea matemáticamente tratable

Ruido en imágenes digitales

- Ruido Gaussiano (electrónica)
- *shot noise* (Fotones, Poisson)
- Ruido impulsivo (píxeles muertos)
- Ruido estructurado (ganancia variable en cada columna del captor)

Modelado de los datos

Ejemplo: Temperatura global media de la Tierra



- Los datos son de naturaleza ruidosa, pero en promedio muestran una tendencia creciente.
- Por ejemplo, un modelo razonable sería una recta en ruido,

$$x[n] = A + Bn + w[n], \quad n = 0, 1, \dots, N - 1.$$

Ejemplo: Temperatura global media de la Tierra

- Si asumimos que el ruido es **blanco** y **Gaussiano** (*WGN, White Gaussian Noise*).
 - **blanco**: cada muestra no está correlacionada con las demás muestras
 - **Gaussiano**: cada muestra $w[n]$ tiene PDF $\mathcal{N}(0, \sigma^2)$.
- La PDF conjunta de las muestras de ruido es,

$$\begin{aligned} p(w[0], w[1], \dots, w[N-1]) &= \prod_{n=0}^{N-1} p(w[n]) \\ &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{w^2[n]}{2\sigma^2}\right] \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} w^2[n]\right]. \end{aligned}$$

Ejemplo: Temperatura global media de la Tierra

- En este ejemplo, los parámetros desconocidos son A y B .
- Arreglando los datos y los parámetros como un vector,

$$\theta = [A, B]^T, \quad \mathbf{x} = [x[0], x[1], \dots, x[N-1]]^T$$

la PDF de los datos es:

$$\begin{aligned} p(\mathbf{x}; \theta) &= p(x[0], x[1], \dots, x[N-1]; \theta) \\ &= p_w(x[0] - A, x[1] - A - B, \dots, x[N-1] - A - B(N-1)) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn)^2 \right]. \end{aligned}$$

Algunas observaciones:

- La hipótesis de AWGN es justificada por la necesidad de obtener un **modelo matemáticamente tratable** que conduzca a estimadores que puedan expresarse en forma cerrada.
- La hipótesis también es razonable a menos que haya evidencia de otra cosa (muestras correlacionadas).
- El desempeño del estimador tiene dependencia fuerte con las hipótesis de la PDF de los datos.
- A lo sumo, se puede esperar que el estimador obtenido sea **robusto**, en el sentido en que pequeños cambios en la PDF de los datos no afecten demasiado el desempeño del estimador.

Estimadores insesgados

Definición (estimador insesgado). Un estimador de cierto parámetro desconocido es *insesgado* si en promedio conduce al valor verdadero del parámetro.

Formalmente, un estimador $\hat{\theta}$ del parámetro $\theta \in (a, b)$ es **insesgado** si:

$$\mathbb{E}(\hat{\theta}) = \theta, \quad \forall \theta \in (a, b).$$

Ejemplo: Estimador insesgado del nivel de DC en WGN.

- Se consideran las observaciones,

$$x[n] = A + w[n], \quad n = 0, 1, \dots, N-1,$$

con $w[n]$ WGN, $w[n] \sim \mathcal{N}(0, \sigma^2)$ y $A \in \mathbb{R}$ es el parámetro a estimar.

- Un estimador *razonable* de A es la media muestral,

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n].$$

Estimadores insesgados

Ejemplo: Estimador insesgado del nivel de DC en WGN.

- Se quiere ver si el estimado es insesgado.

$$\mathbb{E}(\hat{A}) = \mathbb{E} \left[\frac{1}{N} \sum_{n=0}^{N-1} x[n] \right]$$

$$\stackrel{(a)}{=} \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{E}(x[n])$$

$$\stackrel{(b)}{=} \frac{1}{N} \sum_{n=0}^{N-1} A = A$$

(a) Linealidad de la esperanza.

(b) Como A es determinístico,

$$\begin{aligned} \mathbb{E}(x[n]) &= \mathbb{E}(A + w[n]) \\ &= A + \mathbb{E}(w[n]) \\ &= A. \end{aligned}$$

- En este problema, el estimador media muestral es insesgado.

¿Cuál es la PDF de \hat{A} ?

- Suma de variables aleatorias Gaussianas independientes, es una variable aleatoria gaussiana
- \hat{A} es un V.A. Gaussiana (queda especificada por su media y varianza).

Estimadores insesgados

Ejemplo: Estimador insesgado del nivel de DC en WGN.

- La media es $\mathbb{E}(\hat{A}) = A$. Sólo falta calcular la varianza.

$$\begin{aligned}\text{var}(\hat{A}) &= \text{var} \left[\frac{1}{N} \sum_{n=0}^{N-1} x[n] \right] \\ &\stackrel{(a)}{=} \frac{1}{N^2} \sum_{n=0}^{N-1} \text{var}(x[n]) \\ &\stackrel{(b)}{=} \frac{1}{N^2} \sum_{n=0}^{N-1} \sigma^2 = \frac{\sigma^2}{N}\end{aligned}$$

- (a) Si X e Y son V.A. independ.,

$$\text{var}(aX+bY) = a^2\text{var}(X)+b^2\text{var}(Y)$$

- (b) Como A es constante,

$$\begin{aligned}\text{var}(x[n]) &= \text{var}(A + w[n]) \\ &= \text{var}(w[n]) = \sigma^2.\end{aligned}$$

¿Cuál es la PDF de \hat{A} ?

- Se concluye que $\hat{A} \sim \mathcal{N}(A, \sigma^2/N)$.
- La varianza del estimador decrece un factor de N respecto a la a varianza de las muestras individuales.

Estimadores insesgados: Comentarios

- La restricción de que $\mathbb{E}(\hat{\theta}) = \theta$ para todo $\theta \in (a, b)$ es importante. Significa que si,

$$\hat{\theta} = g(\mathbf{x}), \text{ con } \mathbf{x} = [x[0], x[1], \dots, x[N-1]]^T,$$

se tiene que cumplir que

$$\mathbb{E}(\hat{\theta}) = \int g(\mathbf{x})p(\mathbf{x}; \theta)d\mathbf{x} = \theta \quad \forall \theta \in (a, b)$$

Podría ocurrir que se cumpla la igualdad únicamente para algunos valores de θ pero no para otros.

- Estimadores **insesgados**: no son necesariamente buenos estimadores. En promedio alcanzan el valor verdadero del parámetro.
- Estimadores **sesgados**: introducen error sistemático en la estimación pero pueden lograr reducir su varianza (menos variabilidad).
- Compromiso **sesgo-varianza** (*bias-variance tradeoff*).

Combinación de estimadores **insesgados**

- La propiedad de insesgado tiene implicancias importantes al combinar estimadores.
- Supongamos que disponemos de p estimadores del mismo parámetro θ , es decir: $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p$.
- Promedio de estimadores da un nuevo estimador:

$$\hat{\theta} = \frac{1}{p} \sum_{i=1}^p \hat{\theta}_i$$

- Asumiendo que los estimadores son insesgados, de igual varianza y no correlacionados, tenemos que:

$$\mathbb{E}(\hat{\theta}) = \theta, \quad \text{var}(\hat{\theta}) = \frac{\text{var}(\hat{\theta}_1)}{p}.$$

- Cuantos más estimadores se combinan, más decrece la varianza y obtenemos un mejor estimador:

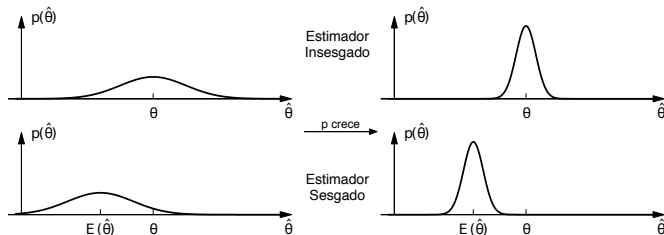
$$\lim_{p \rightarrow \infty} \text{var}(\hat{\theta}) \rightarrow 0, \quad \text{entonces } \Pr \left(\lim_{p \rightarrow \infty} \hat{\theta} = \theta \right) = 1.$$

Combinación de estimadores **sesgados**

- En el caso en que los estimadores $\hat{\theta}_i$ son sesgados (mismo sesgo), es decir, $\mathbb{E}(\hat{\theta}_i) = \theta + b(\theta)$, se tiene que,

$$\mathbb{E}(\hat{\theta}) = \frac{1}{p} \sum_{i=1}^p \mathbb{E}(\hat{\theta}_i) = \theta + b(\theta).$$

- Sin importar cuántos estimadores se promedien, $\hat{\theta}$ no converge al valor verdadero θ .



- Sesgo de un estimador $b(\theta) = \mathbb{E}(\hat{\theta}) - \theta$.

Criterio de Mínima Varianza

- En la búsqueda de **estimadores óptimos** es necesario utilizar algún **criterio de optimalidad**.
- Uno natural es la minimización del **Error Cuadrático Medio** (*MSE, Mean Square Error*)

$$\text{MSE}(\hat{\theta}) = \mathbb{E} \left[(\hat{\theta} - \theta)^2 \right].$$

- Análisis (descomposición) del error cuadrático medio:

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= \mathbb{E} \left[(\hat{\theta} - \theta)^2 \right] \\ &= \mathbb{E} \left\{ \left[(\hat{\theta} - \mathbb{E}(\hat{\theta})) + (\mathbb{E}(\hat{\theta}) - \theta) \right]^2 \right\} \\ &= \mathbb{E} \left[(\hat{\theta} - \mathbb{E}(\hat{\theta}))^2 \right] + \underbrace{2 \mathbb{E} \left[(\hat{\theta} - \mathbb{E}(\hat{\theta})) (\mathbb{E}(\hat{\theta}) - \theta) \right]}_{(\mathbb{E}(\hat{\theta}) - \theta) \mathbb{E}(\hat{\theta} - \mathbb{E}(\hat{\theta})) = 0} + \mathbb{E} \left[(\mathbb{E}(\hat{\theta}) - \theta)^2 \right] \\ &= \mathbb{E} \left[(\hat{\theta} - \mathbb{E}(\hat{\theta}))^2 \right] + (\mathbb{E}(\hat{\theta}) - \theta)^2 \\ &= \text{var}(\hat{\theta}) + b^2(\theta) \end{aligned}$$

- Descomposición sumamente útil *bias-variance*.

Criterio de Mínima Varianza

Ejemplo: Estimador MSE del nivel de DC en WGN

$$x[n] = A + w[n], \quad n = 0, 1, \dots, N-1, \quad w[n] \text{ i.i.d con } w[n] \sim \mathcal{N}(0, \sigma^2).$$

- Se considera como estimador la media muestral modificada,

$$\check{A} = \frac{a}{N} \sum_{n=0}^{N-1} x[n], \quad \text{para una constante } a.$$

- Se desea el valor de a que minimiza el MSE.

Media	Sesgo	Varianza
$\mathbb{E}(\check{A}) = aA$	$b(A) = (a - 1)A$	$\text{var}(\check{A}) = \frac{a^2 \sigma^2}{N}$

- Sustituyendo $b(A)$ y $\text{var}(\check{A})$ en la ecuación del MSE obtenemos

$$\text{MSE}(\check{A}) = \frac{a^2 \sigma^2}{N} + (a - 1)^2 A^2$$

Criterio de Mínima Varianza

Ejemplo: Estimador MSE del nivel de DC en WGN

- Sustituyendo $b(A)$ y $\text{var}(\check{A})$ en la ecuación del MSE obtenemos

$$\text{MSE}(\check{A}) = \frac{a^2 \sigma^2}{N} + (a - 1)^2 A^2.$$

- Diferenciando respecto a a , se obtiene

$$\frac{d\text{MSE}(\check{A})}{da} = \frac{2a\sigma^2}{N} + 2(a - 1)A^2,$$

- e igualando a cero para encontrar el valor de a_{opt} se obtiene que

$$a_{\text{opt}} = \frac{A^2}{A^2 + \sigma^2/N}.$$

- El estimador que produce el menor error cuadrático medio es

$$\check{A} = \left(\frac{A^2}{A^2 + \sigma^2/N} \right) \left(\frac{1}{N} \sum_{n=0}^{N-1} x[n] \right).$$

- Problema:** El estimador del parámetro desconocido depende del valor del parámetro desconocido. No se puede realizar.

Criterio de Mínima Varianza

- En general los estimadores que minimizan el error cuadrático medio (MSE) dependen del parámetro desconocido y por lo tanto no son realizables.
- Esto es porque el MSE es función del sesgo y el sesgo en general depende del parámetro desconocido.
- Comparación de los estimadores del nivel de DC en WGN

Sesgo	Varianza	MSE
$b(\check{A}) = (a - 1)A$	$\text{var}(\check{A}) = \frac{a^2 \sigma^2}{N}$	$\text{MSE}(\check{A}) = a^2 \sigma^2 / N + (a - 1)^2 A^2$

donde

$$a_{\text{opt}} = \frac{A^2}{A^2 + \sigma^2 / N} < 1, \quad y \quad a_{\text{unbiased}} = 1.$$

- Estimador insesgado tiene mayor varianza y error cuadrático medio.
- **Compromiso sesgo-varianza** (*bias-variance tradeoff*): En general, reducir la varianza de un estimador tiene el costo de hacerlo sesgado.

Estimador insesgados de varianza mínima (MVU)

- El enfoque de **minimizar el MSE** debe ser abandonado ya que (en general) conduce a estimadores irrealizables.
- **Alternativa:** restringirse a estimadores insesgados y minimizar la varianza
- Recordar que, $\text{MSE}(\hat{\theta}) = \text{var}(\hat{\theta}) + b^2(\theta)$.

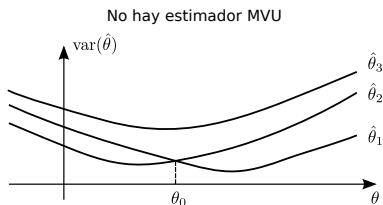
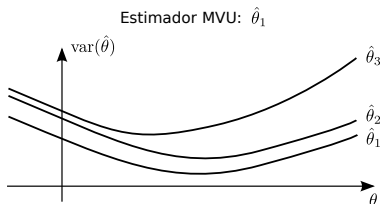
Si, $b(\theta) = 0$, entonces, $\text{MSE}(\hat{\theta}) = \text{var}(\hat{\theta})$.

- Como el error cuadrático medio de un estimador insesgado es su varianza, minimizar la varianza equivale a minimizar el MSE.
- Estimador insesgados de varianza mínima o *MVU*, *Minimum-variance Unbiased*.

Estimador insesgados de varianza mínima (MVU)

Existencia de estimadores MVU.

- Se dice que **existe un estimador MVU** si hay un estimador de **menor varianza** que el resto de los posibles estimadores **para todo θ** .



- Dos ejemplos: izquierda (existe MVU); derecha (no existe MVU).
- El estimador MVU no tiene porqué existir (Ejemplo a continuación).

Estimador insesgados de varianza mínima (MVU)

Ejemplo: no existencia de estimador MVU

- Si la *forma* de la PDF cambia con θ es esperable que el mejor estimador también dependa de θ .
- Se dispone de dos observaciones independientes $x[0]$ y $x[1]$ con PDF,

$$x[0] \sim \mathcal{N}(\theta, 1) \quad x[1] \sim \begin{cases} \mathcal{N}(\theta, 1) & \text{si } \theta \geq 0 \\ \mathcal{N}(\theta, 2) & \text{si } \theta < 0, \end{cases}$$

y se quiere estimar el parámetro θ .

- Se proponen los siguientes estimadores:

$$\hat{\theta}_1 = \frac{1}{2} (x[0] + x[1]), \quad \hat{\theta}_2 = \frac{2}{3}x[0] + \frac{1}{3}x[1].$$

- Es fácil ver que ambos estimadores son insesgados.

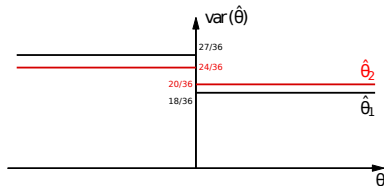
Estimador insesgados de varianza mínima (MVU)

Ejemplo: no existencia de estimador MVU

- La varianza de los estimadores es,

$$\text{var}(\hat{\theta}_1) = \frac{1}{4} (\text{var}(x[0]) + \text{var}(x[1])) = \begin{cases} \frac{18}{36} & \text{si } \theta \geq 0 \\ \frac{27}{36} & \text{si } \theta < 0, \end{cases}$$

$$\text{var}(\hat{\theta}_2) = \frac{4}{9} \text{var}(x[0]) + \frac{1}{9} \text{var}(x[1]) = \begin{cases} \frac{20}{36} & \text{si } \theta \geq 0 \\ \frac{24}{36} & \text{si } \theta < 0, \end{cases}$$



- Se cumple que si:
 $\theta \geq 0$, $\hat{\theta}_1$ menor varianza
 $\theta < 0$, $\hat{\theta}_2$ menor varianza
- No existe un estimador MVU entre $\hat{\theta}_1$ y $\hat{\theta}_2$.

Estimador insesgados de varianza mínima (MVU)

- No solo puede no existir el estimador MVU, sino que incluso puede suceder que no exista ni un sólo estimador insesgado.
- En este caso no tiene sentido buscar el estimador MVU.

Ejemplo: No existencia de estimador insesgado.

- Se dispone de una única observación $x[0]$, y se sabe que $x[0] \sim \mathcal{U}[0, 1/\theta]$ con $\theta > 0$. Se quiere estimar θ .
- Sea $\hat{\theta} = g(x[0])$ un estimador genérico, y se buscan funciones g de manera de que el estimador sea insesgado. Se necesita que,

$$\mathbb{E}(\hat{\theta}) = \mathbb{E}(g(x[0]))$$

$$\stackrel{(a)}{=} \int g(u) p_{x[0]}(u) du$$

$$\stackrel{(b)}{=} \theta \int_0^{\frac{1}{\theta}} g(u) du$$

$$\stackrel{(c)}{=} \theta.$$

(a) Definición de esperanza.

(b) Como $x[0] \sim \mathcal{U}[0, 1/\theta]$,

$$p_{x[0]}(u) = \begin{cases} \theta & \text{si } 0 \leq u \leq 1/\theta \\ 0 & \text{en otro caso.} \end{cases}$$

(c) Condición de insesgado.

Estimador insesgados de varianza mínima (MVU)

Ejemplo: No existencia de estimador **insesgado**.

- Se dispone de una única observación $x[0]$, y se sabe que $x[0] \sim \mathcal{U}[0, 1/\theta]$ con $\theta > 0$. Se quiere estimar θ .
- Sea $\hat{\theta} = g(x[0])$ un estimador genérico, y se buscan funciones g de manera de que el estimador sea insesgado. Se necesita que,
- Se llegó a que para que el estimador sea insesgado se tiene que cumplir que:

$$\int_0^{\frac{1}{\theta}} g(u) du = 1, \quad \forall \theta > 0.$$

- No existe una función g que cumpla la condición $\forall \theta > 0$.
- Se concluye entonces que no existe un estimador insesgado para este problema de estimación.

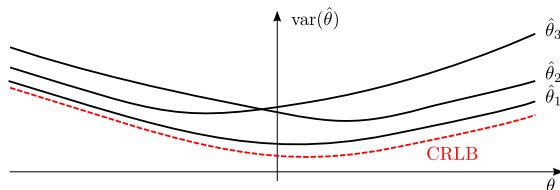
Búsqueda de estimadores MVU

- Aún si existe un estimador MVU, puede no ser posible encontrarlo. No hay ninguna *receta infalible* para encontrar estimadores MVU.

Enfoques de búsqueda de estimadores MVU:

1 Utilizando la cota de inferior de Cramér-Rao (CRLB, *Cramér-Rao Lower Bound*)

- Determinar la CRLB y ver si algún estimador la alcanza.
- CRLB determina un límite inferior en la varianza de cualquier estimador insesgado (Capítulo 3 Kay)
- Si un estimador tiene varianza igual a la CRLB para todos los valores de θ , es el estimador MVU.



Enfoques de búsqueda de estimadores MVU:

2 **Buscar estadísticos suficientes y aplicar el teorema de Rao-Blackwell-Lehmann-Scheffé (RBLs)**

- Puede existir un estimador MVU que no alcance la CRLB.
- Capítulo 5 (Kay)

3 **Restringir la clase de estimadores (e.g., lineales)**

- Restringir la clase de estimadores no sólo a los insesgados, sino también a los insesgados que sean lineales con los datos, y encontrar el MVU en esta clase.
- Este estimador no será óptimo, a menos que el estimador MVU sea lineal en ese problema en particular.
- Capítulo 6 (Kay)

Extensión a vector de parámetros

En el problema general de estimación de parámetros, los parámetros desconocidos pueden ser varios.

Estimador insesgado

- Si hay p parámetros desconocidos, se construye el vector de parámetros desconocidos, $\theta = [\theta_1, \theta_2, \dots, \theta_p]^T$.
- Se dice que un estimador $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p]^T$ es insesgado, si

$$\mathbb{E}(\hat{\theta}_i) = \theta_i, \quad a_i < \theta_i < b_i,$$

para todo $i = 1, 2, \dots, p$.

- Si definimos la esperanza de un vector de variables aleatorias como

$$\mathbb{E}(\hat{\theta}) = [\mathbb{E}(\hat{\theta}_1), \mathbb{E}(\hat{\theta}_2), \dots, \mathbb{E}(\hat{\theta}_p)]^T,$$

un estimador insesgado cumple la igualdad vectorial

$$\mathbb{E}(\hat{\theta}) = \theta.$$

Estimador $\hat{\theta}$ de parámetro vectorial $\theta \in \mathbb{R}^p$ es MVU si:

- Es insesgado, es decir cumple la igualdad vectorial

$$\mathbb{E}(\hat{\theta}) = \theta;$$

- cumple la propiedad de que

$$\text{var}(\hat{\theta}_i) \text{ es mínima, para } i = 1, 2, \dots, p,$$

entre todos los estimadores insesgados.

- **Kay, S. M.** (1993)
Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory, Capítulo 2.