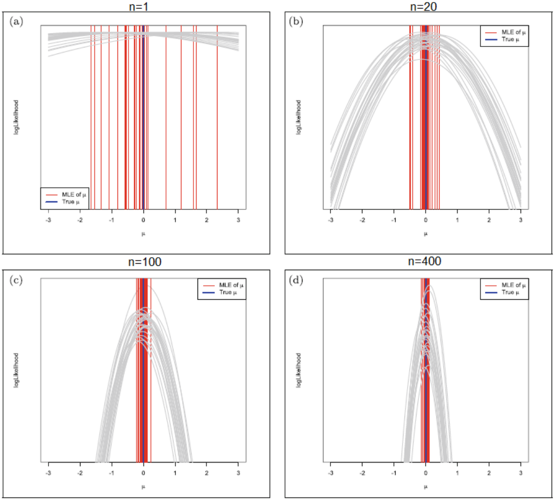
Estimadores: Propiedades

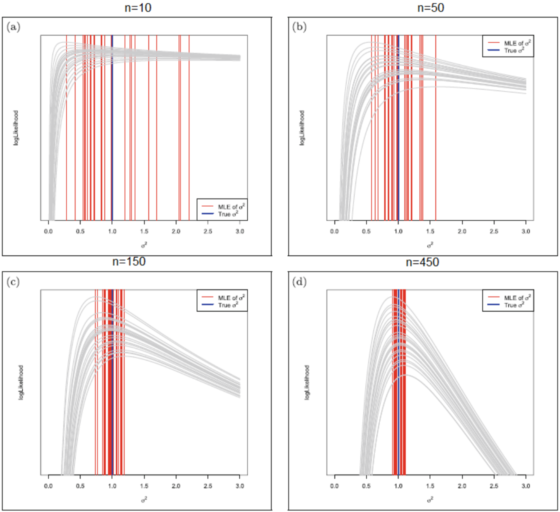
Fluctuación; log–Verosimilitud y su máximo (EMV) $X_i \sim N(\mu, 1)$

Panaretos, V. M. 2016. *Statistics for Mathematicians: A Rigorous First Course*, Birkhaüser, Springer.



25 replicaciones

Fluctuación: log–Verosimilitud y su máximo (EMV) $X_i \sim N(0, \sigma^2)$



25 replicaciones

- En ambos casos se observa cómo las funciones de log–verosimilitud se vuelven gradualmente más curvas a medida que n aumenta, por lo que su máximo fluctúa cada vez menos de una repetición a otra.

- También se observa que los máximos tienden a concentrarse alrededor del verdadero valor del parámetro a medida que n aumenta.

Ejemplo: $\mathcal{U}[0,\theta]$

$X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{U}[0,\theta]$

- Estimador de Momentos: $\widehat{\theta}_n = 2\bar{X}_n$
- Estimador de Máxima Verosimilitud: $\widetilde{\theta}_n = \max\limits_{1\leq i\leq n}(X_i)$

Visitemos la app de shiny

¡Gracias Mariela!!!

`https://glmconr2.shinyapps.io/estimadores_uniforme/`

Visitemos la app de shiny

¡Gracias Mariela!!!

https://glmconr2.shinyapps.io/estimadores_uniforme/

- $\widehat{\theta}_n$ y $\widetilde{\theta}_n$ son variables aleatorias: tienen una distribución

- En la práctica tenemos una sola estimación!!

- Veamos un ejemplo.

Ejemplo: Estimamos una proporción

Queremos estimar a p, la proporción de ciudadanos de Argentina que está a favor de las vacunas. Para ello, elegiremos n ciudadanos al azar y se les preguntará si están a favor de las vacunas o no. ¿Cómo estimamos a p?

Ejemplo: Estimamos una proporción

Queremos estimar a p, la proporción de ciudadanos de Argentina que está a favor de las vacunas. Para ello, elegiremos n ciudadanos al azar y se les preguntará si están a favor de las vacunas o no. ¿Cómo estimamos a p?

$$X_i = \begin{cases} 1 \text{ si la } i \text{ ésimo encuestado dice que sí} \\ 0 \text{ si no} \end{cases}$$

- $X_1, \ldots, X_n$ son variables aleatorias i.i.d. con distribuición $B(1, p)$, siendo p la verdadera proporción que queremos estimar.
- Un estimador puntual (EMM o EMV) de p es $\hat{p}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$
- Una vez realizado el experimento con $n = 1000$, observamos que hay 853 personas a favor. Obtenemos una **estimación** para p: $\hat{p}_{n\,obs} = 0.853$.
- ¿Por qué tomamos $n = 1000$? ¿Sería más conveniente tomar $n = 2000$? ¿Cómo se aprecia si es mejor una elección que la otra?

Criterios de bondad de un estimador: Sesgo

Sea $\mathbf{X}_n = (X_1 \ldots X_n)$ una muestra aleatoria de una distribución $F \in \mathcal{F}$, $\mathcal{F} = \{F(\cdot, \theta)\ , \theta \in \Theta\}$ y supongamos que queremos estimar $q(\theta)$, una función real del parámetro de F

Un estimador de $q(\theta)$, $T_n = T(\mathbf{X}_n)$, se dice **insesgado** para $q(\theta)$ si

$$\mathbb{E}_\theta(T_n) = q(\theta) \text{ para todo } \theta \in \Theta.$$

Caso contrario, decimos que es **sesgado**. Un estimador de $q(\theta)$, T_n, se dice **asintóticamente insesgado** para $q(\theta)$ si

$$\lim_{n\to\infty} \mathbb{E}_\theta(T_n) = q(\theta) \text{ para todo } \theta \in \Theta.$$

Llamamos **sesgo** de un estimador T_n de $q(\theta)$ a

$$\mathbb{B}_\theta(T_n) = \mathbb{E}_\theta(T_n) - q(\theta).$$

Ejemplo: Volviendo a las vacunas...

$X_1, \ldots, X_n$ son variables aleatorias i.i.d. $X_i \sim B(1, p)$, siendo p la verdadera proporción que queremos estimar. Tenemos que $\mathbb{E}_p(X_1) = p$.

- $\hat{p}_n = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} X_i$
- $\mathbb{E}_p(\hat{p}_n) = \mathbb{E}_p(\bar{X}_n) = \frac{1}{n}\, \mathbb{E}_p(\displaystyle\sum_{i=1}^{n} X_i) = \dfrac{n}{n}\, \mathbb{E}_p(X_1) = p$
- En este caso, resulta $\mathbb{E}_p(\hat{p}_n) = p$

Ejemplo: Caso general - F con media y varianza

Sea $\mathbf{X}_n = (X_1 \ldots X_n)$ una muestra aleatoria de una distribución $F \in \mathcal{F}$, y supongamos que queremos estimar $q(F)$.
Ejemplos: $q_1(F) = \mathbb{E}_F(X_1) \quad q_2(F) = \mathbb{V}_F(X_1)$.

- ¿Cómo estimamos a $q_1(F)$ y $q_2(F)$?

- Vimos que los estimadores de momentos son:

$$\begin{aligned} \widehat{q}_1(F) &= \bar{X}_n \\ \widehat{q}_2(F) &= \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n} = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2 \end{aligned}$$

- ¿Son insesgados?

Ejemplo: Calculemos el sesgo de cada uno

- $E_F(\bar{X}_n) = \frac{1}{n}\,\mathbb{E}_F(\sum_{i=1}^{n} X_i) = \dfrac{n}{n}\,\mathbb{E}_F(X_1) = \mathbb{E}_F(X_1) = q_1(F)$

  $\longrightarrow$ insesgado!

Ejemplo: Calculemos el sesgo de cada uno

- $E_F(\bar{X}_n) = \frac{1}{n}\,\mathbb{E}_F(\sum_{i=1}^{n} X_i) = \dfrac{n}{n}\,\mathbb{E}_F(X_1) = \mathbb{E}_F(X_1) = q_1(F)$

  $\longrightarrow$ insesgado!

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

Ejemplo: Calculemos el sesgo de cada uno

- $E_F(\bar{X}_n) = \frac{1}{n}\,\mathbb{E}_F(\sum_{i=1}^{n} X_i) = \dfrac{n}{n}\,\mathbb{E}_F(X_1) = \mathbb{E}_F(X_1) = q_1(F)$

  $\longrightarrow$ insesgado!

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

$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_F\left(X_i^2\right) - \mathbb{E}_F\left(\bar{X}^2\right) = \frac{n}{n} E_F\left(X_1^2\right) - \mathbb{E}_F\left(\bar{X}^2\right)$$

Ejemplo: Calculemos el sesgo de cada uno

- $E_F(\bar{X}_n) = \frac{1}{n}\,\mathbb{E}_F(\sum_{i=1}^{n} X_i) = \dfrac{n}{n}\,\mathbb{E}_F(X_1) = \mathbb{E}_F(X_1) = q_1(F)$

  $\longrightarrow$ insesgado!

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

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_F\left(X_i^2\right) - \mathbb{E}_F\left(\bar{X}^2\right) = \frac{n}{n}E_F\left(X_1^2\right) - \mathbb{E}_F\left(\bar{X}^2\right)$$

$$= \mathbb{V}_F\left(X_1\right) + \mathbb{E}_F\left(X_1\right)^2 - \left\{\mathbb{V}_F\left(\bar{X}\right) + \mathbb{E}_F\left(\bar{X}\right)^2\right\}$$

Ejemplo: Calculemos el sesgo de cada uno

- $E_F(\bar{X}_n) = \frac{1}{n}\,\mathbb{E}_F(\sum_{i=1}^{n} X_i) = \frac{n}{n}\,\mathbb{E}_F(X_1) = \mathbb{E}_F(X_1) = q_1(F)$

  $\longrightarrow$ insesgado!

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

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_F\left(X_i^2\right) - \mathbb{E}_F\left(\bar{X}^2\right) = \frac{n}{n}E_F\left(X_1^2\right) - \mathbb{E}_F\left(\bar{X}^2\right)$$

$$= \mathbb{V}_F\left(X_1\right) + \mathbb{E}_F\left(X_1\right)^2 - \left\{\mathbb{V}_F\left(\bar{X}\right) + \mathbb{E}_F\left(\bar{X}\right)^2\right\}$$

$$= \mathbb{V}_F(X_1) + \mathbb{E}_F\left(X_1\right)^2 - \frac{\mathbb{V}_F(X_1)}{n} - \mathbb{E}_F\left(X_1\right)^2 = \frac{n-1}{n}\mathbb{V}_F(X_1)$$

Ejemplo: Calculemos el sesgo de cada uno

- El sesgo es

$$\begin{aligned}\mathbb{B}_F\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2\right) &= \frac{n-1}{n}\mathbb{V}_F(X_1) - \mathbb{V}_F(X_1) \\ &= -\mathbb{V}_F(X_1)/n < 0\end{aligned}$$

$\longrightarrow$ sesgo: $-\frac{1}{n}\mathbb{V}_F(X_1)$ subestima!

- El sesgo puede corregirse

$$\frac{n}{n-1}\,\widehat{q}_2(F) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = S^2$$

que hemos llamado varianza muestral.

Ejemplo: Caso Normal

Supongamos que $\mathbf{X}_n = (X_1 \ldots X_n)$ i.i.d. $X_i \sim N(\mu, \sigma^2)$, y que queremos estimar a μ y σ^2.

Como sabemos $q_1(F) = \mathbb{E}_F(X_1) = \mu$ y $q_2(F) = \mathbb{V}_F(X_1) = \sigma^2$.

Por lo que hemos visto, $\bar{X}_n$ y $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ son estimadores insesgados de μ y σ^2, respectivamente.

Error cuadrático medio

$\mathbf{X}_n = (X_1 \ldots X_n)$ una m. a. de una distribución $F \in \mathcal{F}$,
$\mathcal{F} = \{F(\cdot, \theta)\ , \theta \in \Theta\}$ y queremos estimar $q(\theta)$.
Sea $T_n = T(\mathbf{X}_n)$ un estimador de $q(\theta)$. El error de estimación es:

$$T_n - q(\theta)$$

Error cuadrático medio

$\mathbf{X}_n = (X_1 \dots X_n)$ una m. a. de una distribución $F \in \mathcal{F}$,
$\mathcal{F} = \{F(\cdot, \theta)\ , \theta \in \Theta\}$ y queremos estimar $q(\theta)$.
Sea $T_n = T(\mathbf{X}_n)$ un estimador de $q(\theta)$. El error de estimación es:

$$T_n - q(\theta) \qquad \text{es una v.a.}$$

Una noción del error cometido está dada por el **Error Cuadrático Medio**

$$ECM_\theta(T_n) = \mathbb{E}_\theta \left((T_n - q(\theta))^2 \right)$$

El **Error Cuadrático Medio** da un criterio para determinar si un estimador T_n de $q(\theta)$ es mejor que otro W_n, pues basta verificar si

$$ECM_\theta(T_n) \leq ECM_\theta(W_n) \quad \forall \theta \in \Theta$$

Compromiso sesgo-varianza

Teorema

$ECM_\theta(T_n) = \mathbb{V}_\theta(T_n) + (\mathbb{B}_\theta(T_n))^2$

Compromiso sesgo-varianza

Teorema

$ECM_\theta(T_n) = \mathbb{V}_\theta(T_n) + (\mathbb{B}_\theta(T_n))^2$

$$\begin{aligned} ECM_\theta(T_n) &= \mathbb{E}_\theta\left[(T_n - q(\theta))^2\right] \\ &= \mathbb{V}_\theta\left[(T_n - q(\theta))\right] + \left[\mathbb{E}_\theta(T_n - q(\theta))\right]^2 \\ &= \mathbb{V}_\theta(T_n) + \left[\mathbb{E}_\theta(T_n) - q(\theta)\right]^2 \\ &= \mathbb{V}_\theta(T_n) + (\mathbb{B}_\theta(T_n))^2, \text{ como queríamos demostrar.} \end{aligned}$$

ECM

Notemos que

- si el estimador T_n es insesgado, $ECM_\theta(T_n) = \mathbb{V}_\theta(T_n)$
- comparar el ECM de dos estimadores insesgados, se reduce a comparar sus varianzas.

ECM

Volvamos al ejemplo de estimar la proporción.

Recordemos que $\hat{p}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ es insesgado como estimador de p.

Usando la independencia, tenemos que

$$\mathbb{V}(\hat{p}_n) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i) = \frac{\mathbb{V}(X_1)}{n} = \frac{p(1-p)}{n}.$$

¿Esto responde la pregunta que nos hicimos en el caso de las vacunas?

ECM

Volvamos al ejemplo de estimar la proporción.

Recordemos que $\hat{p}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ es insesgado como estimador de p.

Usando la independencia, tenemos que

$\mathbb{V}(\hat{p}_n) = \dfrac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i) = \dfrac{\mathbb{V}(X_1)}{n} = \dfrac{p(1-p)}{n}$.

¿Esto responde la pregunta que nos hicimos en el caso de las vacunas?

Como $\hat{p}_n$ es insesgado $ECM_p(\hat{p}_n) = \frac{p(1-p)}{n} \to 0$ si $n \to \infty$

ECM del EMM y EMV de θ para $U[0,\theta]$

(1) EMM: $\hat{\theta} = 2\bar{X}_n$. Para todo θ tenemos

$$E_\theta(\hat{\theta}) = 2E_\theta(\bar{X}_n) = 2\,\frac{\theta}{2} = \theta$$

$$V_\theta(\hat{\theta}) = 4\mathbb{V}_\theta(\bar{X}_n) = 4\,\frac{\theta^2}{12n} = \frac{\theta^2}{3n}$$

Por lo tanto

$$ECM_\theta(\hat{\theta}) = \mathbb{V}_\theta(\hat{\theta}) = \frac{\theta^2}{3n}$$

(2) EMV: $\tilde{\theta} = \max\{X_1, \ldots, X_n\}$

Probarán que

$$ECM_\theta(\tilde{\theta}) = \frac{n}{(n+2)(n+1)^2}\theta^2 + \left(\frac{n}{n+1}\theta - \theta\right)^2 = \frac{2\theta^2}{(n+1)(n+2)}$$

En este caso, tenemos que $\forall n$

$$\frac{2\theta^2}{(n+1)(n+2)} \leq \frac{\theta^2}{3n}$$

$\widehat{f}_h(x)$: estimador de Parzen

`https://glmconr2.shinyapps.io/app_regre2/`

1. Generá tu conjunto de datos: usá tu numero de libreta.
2. Realizá una estimación de la densidad usando el núcleo normal usando los valores de ventana h= 0.25,1, 4. ¿Qué pasa a medida que h crece?
3. Pegá tus gráficos **acá**.

Estimadores de núcleos: Selección de ventana

En los gráficos anteriores se muestra que la elección de la ventana es crucial.

- Una ventana h pequeña dará un estimador muy rugoso, con muchos picos y difícil de interpretar
- una ventana h grande sobresuaviza al estimador de la densidad y enmascara estructuras de los datos.

$\widehat{f_h}(x)$: estimador de Parzen

$$\hat{f}_h(x) = \frac{1}{2h\,n}\sum_{i=1}^{n}\mathcal{I}_{[x-h,x+h]}(X_i) \quad \text{con} \sum_{i=1}^{n}\mathcal{I}_{[x-h,x+h]}(X_i) \sim Bi(n,p)$$

donde $p = \mathbb{P}(x-h \leq X \leq x+h) = F(x+h) - F(x-h) = p_{x,h}$

$\widehat{f_h}(x)$: estimador de Parzen

$$\hat{f}_h(x) = \frac{1}{2h\,n}\sum_{i=1}^{n}\mathcal{I}_{[x-h,x+h]}(X_i) \quad \text{con} \sum_{i=1}^{n}\mathcal{I}_{[x-h,x+h]}(X_i) \sim Bi(n,p)$$

donde $p = \mathbb{P}(x-h \leq X \leq x+h) = F(x+h) - F(x-h) = p_{x,h}$
Luego, usando las propiedades de esperanza y varianza:

$$\mathbb{E}[\hat{f}_h(x)] = \frac{F(x+h) - F(x-h)}{2h}$$

$$\mathbb{V}\text{ar}[\hat{f}_h(x)] = \frac{F(x+h) - F(x-h)}{4nh^2} - \frac{(F(x+h) - F(x-h))^2}{4nh^2}$$

$\widehat{f}_h(x)$: estimador de Parzen

$$\hat{f}_h(x) = \frac{1}{2h\,n}\sum_{i=1}^{n}\mathcal{I}_{[x-h,x+h]}(X_i) \quad \text{con} \sum_{i=1}^{n}\mathcal{I}_{[x-h,x+h]}(X_i) \sim Bi(n,p)$$

donde $p = \mathbb{P}(x-h \leq X \leq x+h) = F(x+h) - F(x-h) = p_{x,h}$
Luego, usando las propiedades de esperanza y varianza:

$$\mathbb{E}[\hat{f}_h(x)] = \frac{F(x+h) - F(x-h)}{2h}$$

$$\mathbb{V}\mathrm{ar}[\hat{f}_h(x)] = \frac{F(x+h) - F(x-h)}{4nh^2} - \frac{(F(x+h) - F(x-h))^2}{4nh^2}$$

Si

- $h \to 0 \Rightarrow \mathbb{E}[\hat{f}_h(x)] \to f(x)$ y $\mathbb{V}\mathrm{ar}[\hat{f}_h(x)] \approx \frac{f(x)}{2nh} - \frac{f(x)^2}{n} \to \infty$
- $nh \to \infty$, $n \longrightarrow \infty$ $h \longrightarrow 0$, entonces sesgo y varianza se reducen.

Sesgo y Varianza del estimador de Rosenblatt–Parzen

Supongamos que:

- la densidad f es de cuadrado integrable, dos veces continuamente diferenciable y su segunda derivada es de cuadrado integrable.
- el núcleo K es simétrico, es una densidad acotada, de cuadrado integrable y con segundo momento finito.
- $h = h_n \to 0$, $n \to \infty$ y $nh \to \infty$.

Luego,

$$\begin{aligned}
\mathbb{B}\left(\hat{f}_h(x)\right) &= \frac{1}{2}\mu_2(K)f''(x)h^2 + o(h^2) \\
\mathbb{V}\left(\hat{f}_h\left(x\right)\right) &= \frac{\|K\|_2^2}{nh}f(x) + o((nh)^{-1})
\end{aligned}$$

Recomendamos:
Tesis de Licenciatura de Romina Cornistein: *Estimadores de Naradaya Watson aplicados a datos funcionales*
Disponible en:
https://web.dm.uba.ar/index.php/509-tesis-de-licenciatura

Sesgo del estimador de Rosenblatt–Parzen (Video 1)

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - X_i}{h}\right)$$

Sesgo del estimador de Rosenblatt–Parzen (Video 1)

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right)$$

$$\mathbb{E}\left(\hat{f}_h(x)\right) = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right)\right]$$

Sesgo del estimador de Rosenblatt–Parzen (Video 1)

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right)$$

$$\begin{aligned}
\mathbb{E}\left(\hat{f}_h(x)\right) &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}\mathbb{E}\left[K\left(\frac{x - X_i}{h}\right)\right]
\end{aligned}$$

Sesgo del estimador de Rosenblatt–Parzen (Video 1)

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right)$$

$$\begin{aligned}
\mathbb{E}\left(\hat{f}_h(x)\right) &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}\mathbb{E}\left[K\left(\frac{x - X_i}{h}\right)\right] \\
&= \mathbb{E}\left[\frac{1}{h}K\left(\frac{x - X_1}{h}\right)\right] \\
&= \int \frac{1}{h}K\left(\frac{x - s}{h}\right) f(s)ds
\end{aligned}$$

Resolvemos la integral (Video 1)

$$\mathbb{E}\left(\hat{f}_h(x)\right) = \int \frac{1}{h} K\left(\frac{x-s}{h}\right) f(s)ds$$

Resolvemos la integral (Video 1)

$$\mathbb{E}\left(\hat{f}_h(x)\right) = \int \frac{1}{h} K\left(\frac{x-s}{h}\right) f(s) ds$$

hacemos el cambio de variables: $s = x - ht \rightarrow ds = -hdt$

$$= \int K(t)\, f(x - ht)\, dt$$

Resolvemos la integral (Video 1)

$$\mathbb{E}\left(\hat{f}_h(x)\right) = \int \frac{1}{h} K\left(\frac{x-s}{h}\right) f(s) ds$$

hacemos el cambio de variables: $s = x - ht \to ds = -hdt$

$$= \int K\left(t\right) f\left(x - ht\right) dt$$

ahora sumamos y restamos $f\left(x\right)$

$$= f\left(x\right) + \int K(t) f\left(x - ht\right) dt - \int f\left(x\right) K(t) dt$$

$$= f\left(x\right) + \int \left[f\left(x - ht\right) - f\left(x\right)\right] K(t) dt$$

Seguimos.... (Video 1)

Mediante un desarrollo de Taylor obtenemos

$$f(x-ht) = f(x) - f'(x)ht + \frac{f''(x)}{2}h^2t^2 + o(h^2t^2).$$

Luego,

$$\begin{aligned}\int K(t)[f(x-ht) - f(x)]\,\mathrm{d}t \\ &= \int K(t)\Big\{-f'(x)ht + \frac{f''(x)}{2}h^2t^2 + o(h^2t^2)\Big\}\,\mathrm{d}t \\ &= \frac{1}{2}\mu_2(K)f''(x)h^2 + o(h^2),\end{aligned}$$

siendo $\mu_2(K) = \int K(t)t^2\mathrm{d}t$ el segundo momento de K.

Varianza del estimador de Rosenblatt–Parzen (Video 1)

$$\mathbb{V}\left(\hat{f}_h(x)\right) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x-X_i}{h}\right)\right) = \frac{1}{n}\mathbb{V}\left(\frac{1}{h}K\left(\frac{x-X_1}{h}\right)\right)$$ por qué?

$\swarrow \mathbb{V}(Z) = \mathbb{E}\left(Z^2\right) - \mathbb{E}^2(Z)$

$$= \frac{1}{n}\int \frac{1}{h^2}K^2\left(\frac{x-s}{h}\right)f(s)ds - \frac{1}{n}\left[\int K(t)f(x-ht)\,dt\right]^2$$

$$= \frac{1}{nh}\int K^2(t)f(x-ht)\,dt - \frac{1}{n}\left[\int K(t)f(x-ht)\,dt\right]^2$$

Continuanción

$$
\begin{aligned}
\frac{1}{h}\int K^2(t) f(x-ht)\,\mathrm{d}t = & \\
&= \frac{1}{h}\int K^2(t)\left\{f(x)+O(ht)\right\}\,\mathrm{d}t \\
&= \frac{\|K\|_2^2}{h} f(x) + O(1).
\end{aligned}
$$

siendo $\|K\|_2^2 = \int K^2(t)\,\mathrm{d}t$. Por lo tanto, reemplazando

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}[\hat{f}_h(x)] &= \frac{1}{n}\left\{\frac{\|K\|_2^2}{h} f(x) + O(1) - O(1)\right\} \\
&= \frac{\|K\|_2^2}{nh} f(x) + O(n^{-1}) \\
&= \frac{\|K\|_2^2}{nh} f(x) + o((nh)^{-1}),
\end{aligned}
$$

ya que $n^{-1} = o((nh)^{-1})$.

Compromiso Sesgo–Varianza

Luego,

$$
\begin{aligned}
\mathbb{B}\left(\hat{f}_h(x)\right) &= \frac{1}{2}\mu_2(K)f''(x)h^2 + o(h^2) \\
\mathbb{V}\left(\hat{f}_h\left(x\right)\right) &= \frac{\|K\|_2^2}{nh}f(x) + o((nh)^{-1})
\end{aligned}
$$

Compromiso Sesgo–Varianza

Luego,

$$
\begin{aligned}
\mathbb{B}\left(\hat{f}_h(x)\right) &= \frac{1}{2}\mu_2(K)f''(x)h^2 + o(h^2)\\
\mathbb{V}\left(\hat{f}_h\left(x\right)\right) &= \frac{\|K\|_2^2}{nh}f(x) + o((nh)^{-1})
\end{aligned}
$$

Para n fijo:
$h \to 0$ implica $E\left(\hat{f}_h(x)\right) \to f\left(x\right)$ y $\mathbb{V}\left(\hat{f}_h\left(x\right)\right) \to \infty$

Sesgo y Varianza de $\widehat{f}_h(x)$

Comprobamos en el caso fácil que para que todo funcione bien deben pasar dos cosas:

- $h \longrightarrow 0$
- $nh \longrightarrow \infty$
- Se puede probar que bajo condiciones generales del núcleo K

$$
\begin{aligned}
\mathbb{B}[\widehat{f}_h(x)] &\approx \frac{h^2}{2}\mu_2(K)\, f''(x) \\
\mathbb{V}\mathrm{ar}[\widehat{f}_h(x)] &\approx \frac{1}{nh}\|K\|_2^2\, f(x)
\end{aligned}
$$

Sesgo y Varianza de $\widehat{f}_h(x)$

Comprobamos en el caso fácil que para que todo funcione bien deben pasar dos cosas:

- $h \longrightarrow 0$
- $nh \longrightarrow \infty$
- Se puede probar que bajo condiciones generales del núcleo K

$$
\begin{aligned}
\mathbb{B}[\widehat{f}_h(x)] &\approx \frac{h^2}{2}\mu_2(K)\, f''(x) \\
\mathbb{V}\mathrm{ar}[\widehat{f}_h(x)] &\approx \frac{1}{nh}\|K\|_2^2\, f(x)
\end{aligned}
$$

- El sesgo es proporcional a $h^2 \Rightarrow$ elijamos h pequeña

Sesgo y Varianza de $\widehat{f}_h(x)$

Comprobamos en el caso fácil que para que todo funcione bien deben pasar dos cosas:

- $h \longrightarrow 0$
- $nh \longrightarrow \infty$
- Se puede probar que bajo condiciones generales del núcleo K

$$\begin{aligned} \mathbb{B}[\widehat{f}_h(x)] &\approx \frac{h^2}{2}\mu_2(K)\, f''(x) \\ \mathbb{V}\mathrm{ar}[\widehat{f}_h(x)] &\approx \frac{1}{nh}\|K\|_2^2\, f(x) \end{aligned}$$

- El sesgo es proporcional a $h^2 \Rightarrow$ elijamos h pequeña
- El sesgo depende de $f''(x)$ que mide la curvatura de f en x

Sesgo y Varianza de $\widehat{f}_h(x)$

Comprobamos en el caso fácil que para que todo funcione bien deben pasar dos cosas:

- $h \longrightarrow 0$
- $nh \longrightarrow \infty$
- Se puede probar que bajo condiciones generales del núcleo K

$$\begin{aligned} \mathbb{B}[\widehat{f}_h(x)] &\approx \frac{h^2}{2}\mu_2(K)\, f''(x) \\ \mathbb{V}\mathrm{ar}[\widehat{f}_h(x)] &\approx \frac{1}{nh}\|K\|_2^2\, f(x) \end{aligned}$$

- La varianza dismimuye a medida que nh crece

Sesgo y Varianza de $\widehat{f}_h(x)$

Comprobamos en el caso fácil que para que todo funcione bien deben pasar dos cosas:

- $h \longrightarrow 0$
- $nh \longrightarrow \infty$
- Se puede probar que bajo condiciones generales del núcleo K

$$\begin{aligned} \mathbb{B}[\widehat{f}_h(x)] &\approx \frac{h^2}{2}\mu_2(K)\, f''(x) \\ \mathbb{V}\mathrm{ar}[\widehat{f}_h(x)] &\approx \frac{1}{nh}\|K\|_2^2\, f(x) \end{aligned}$$

- La varianza dismimuye a medida que nh crece
- Para disminuir la varianza necesitamos h o n grandes.

Error Cuadrático Medio de $\widehat{f}_h(x)$

Compromiso Sesgo-Varianza

Tenemos que

$$\mathsf{MSE}[\widehat{f}_h(x)] = (\mathbb{B}[\widehat{f}_h(x)])^2 + \mathbb{V}\mathrm{ar}[\widehat{f}_h(x)]$$

Por otro lado:

$$\begin{aligned}
\mathbb{B}[\widehat{f}_h(x)] &\approx \frac{h^2}{2}\mu_2(K)\, f''(x) \\
\mathbb{V}\mathrm{ar}[\widehat{f}_h(x)] &\approx \frac{1}{nh}\|K\|_2^2\, f(x)
\end{aligned}$$

$$\Rightarrow \mathsf{MSE}[\widehat{f}_h(x)] \approx \frac{h^4}{4}\mu_2^2(K)\,(f''(x))^2 + \frac{1}{nh}\|K\|_2^2\, f(x)$$

Si $h \longrightarrow 0$ y $nh \longrightarrow \infty$, entonces $\mathsf{MSE}[\widehat{f}_h(x)] \to 0$.

Compromiso Sesgo-Varianza $\longleftrightarrow$ relación con sobreajuste

Error Cuadrático Integrado y Medio Integrado de $\widehat{f}_h$

Como medida global: Error Cuadrático Integrado

$$\mathsf{ISE}[\widehat{f}_h] = \int \left(\widehat{f}_h(x) - f(x)\right)^2 \mathrm{d}x$$

Como medida global: Error Cuadrático Medio Integrado

$$\mathsf{MISE}[\widehat{f}_h] = \int \mathsf{MSE}[\widehat{f}_h(x)]\, dx$$

Error Cuadrático Integrado y Medio Integrado de $\widehat{f}_h$

Como medida global: Error Cuadrático Integrado

$$\mathsf{ISE}[\widehat{f}_h] = \int \left(\widehat{f}_h(x) - f(x)\right)^2 \mathrm{d}x$$

Como medida global: Error Cuadrático Medio Integrado

$$\mathsf{MISE}[\widehat{f}_h] = \int \mathsf{MSE}[\widehat{f}_h(x)]\, dx$$

Asintóticamente: Error Cuadrático Medio Integrado Asintótico

$$\mathsf{AMISE}[\widehat{f}_h(x)] \;\; = \;\; \frac{h^4}{4}\mu_2^2(K)\int (f''(x))^2 dx + \frac{1}{nh}\|K\|_2^2$$

Error Cuadrático Medio Integrado de $\widehat{f}_h$

$$\mathsf{AMISE}[\widehat{f}_h(x)] = \frac{h^4}{4}\mu_2^2(K)\|f''\|_2^2 + \frac{1}{nh}\|K\|_2^2$$

- Notemos que es una función de la ventana h.

- Derivando a AMISE respecto de h e igualando a 0, obtenemos que el valor de h que minimiza el AMISE es

$$h = \left\{ \frac{\|K\|_2^2}{\mu_2^2(K)\|f''\|_2^2\, n} \right\}^{\frac{1}{5}}$$

Error Cuadrático Medio Integrado de $\widehat{f}_h$

$$\mathsf{AMISE}[\widehat{f}_h(x)] \quad = \quad \frac{h^4}{4}\mu_2^2(K)\|f''\|_2^2 + \frac{1}{nh}\|K\|_2^2$$

- Notemos que es una función de la ventana h.

- Derivando a AMISE respecto de h e igualando a 0, obtenemos que el valor de h que minimiza el AMISE es

$$h = \left\{ \frac{\|K\|_2^2}{\mu_2^2(K)\|f''\|_2^2 \; n} \right\}^{\frac{1}{5}} = algo * \left(\frac{1}{n}\right)^{\frac{1}{5}} \sim n^{-\frac{1}{5}}$$

Selección de h: Regla de Silverman

El AMISE$[\widehat{f}_h(x)]$ se minimiza en

$$h = \left\{ \frac{\|K\|_2^2}{\mu_2^2(K)\|f''\|_2^2 \, n} \right\}^{\frac{1}{5}}$$

• Silverman propone reemplazar a $\|f''\|_2^2$ por su valor cuando f es normal:

$$\|f''\|_2^2 = \frac{3}{8}\pi^{-1/2} \, \sigma^{-5} \approx 0.212 \, \sigma^{-5}$$

y a σ por un estimador $\widehat{\sigma}$

Selección de h: Regla de Silverman

El AMISE$[\widehat{f}_h(x)]$ se minimiza en

$$h = \left\{ \frac{\|K\|_2^2}{\mu_2^2(K)\|f''\|_2^2 \, n} \right\}^{\frac{1}{5}}$$

• Silverman propone reemplazar a $\|f''\|_2^2$ por su valor cuando f es normal:

$$\|f''\|_2^2 = \frac{3}{8}\pi^{-1/2} \, \sigma^{-5} \approx 0.212 \, \sigma^{-5}$$

y a σ por un estimador $\widehat{\sigma}$

$$h_{Sil} = \left(\frac{4\widehat{\sigma}^5}{3 \, n} \right)^{\frac{1}{5}} \approx 1.06 \, \widehat{\sigma} \, n^{-\frac{1}{5}}$$

- Si f es normal, la ventana h_{Sil} es óptima.
- Si f no es normal, h_{Sil} dará una ventana no muy alejada de la óptima cuando la distribución no es muy diferente a la normal.

Selección de h: Regla de Silverman

$$h_{Sil} = \left(\frac{4\widehat{\sigma}^5}{3\,n}\right)^{\frac{1}{5}} \approx 1.06\;\widehat{\sigma}\;n^{-\frac{1}{5}}$$

- σ puede estimarse por S (sd(datos) en R).

  o

- σ puede estimarse por la distancia intercuartil IQR (IQR(datos) en R). Para que coincida con σ bajo la distribución normal debe dividirse por 1.349
- La ventana óptima de acuerdo a la regla de Silverman resulta:

$$h_{Sil} = 1.06\;\min\left(S, \frac{IQR}{1.349}\right)\;n^{-\frac{1}{5}}$$

- Algunos autores sugieren reducir el factor 1.06 a 0.90 para aumentar la posibilidad de detectar bimodalidad. (Explorar bw.nrd y bw.nrd0 en R)

¿Qué hace R?

Sheather, S. J. (2004). Density Estimation, *Statistical Science*, Vol. 19, No. 4, 588?597.

- La ventana óptima que vimos resulta de:

$$h_{Sil} = 1.06\ \min(S, \frac{IQR}{1.349})\ n^{-\frac{1}{5}}$$

  (bw.nrd)
- Silverman (1986) propone usar la ventana óptima que resulta de:

$$h_{Sil} = 0.9\ \min(S, \frac{IQR}{1.349})\ n^{-\frac{1}{5}}$$

  así reduciendo el factor se espera poder detectar más fácilmente bimodalidad.

  (bw.nrd0)

Otro camino para determinar h

Convalidación Cruzada por Máxima Verosimilitud (CV)

Podríamos elegir la ventana h maximizando la log–verosimilitud .....pero

- El EMV de h es degenerado, resultando una densidad que da masa $1/n$ a cada uno de los datos.

Otro camino para determinar h

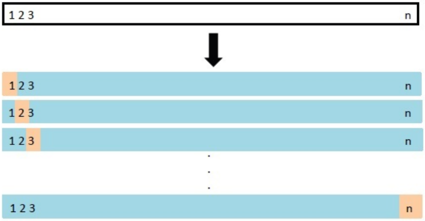
Convalidación Cruzada por Máxima Verosimilitud (CV)

Podríamos elegir la ventana h maximizando la log–verosimilitud .....pero

- El EMV de h es degenerado, resultando una densidad que da masa $1/n$ a cada uno de los datos.
- Alternativa: maximimizar una *pseudo–verosimilutd* computada sacando de los datos una observación a la vez (leave–one–out cross–validation).

Convalidación Cruzada

Para evitar sobreajuste usamos *leave-one-out-Cross-Validation*:
(Fuente: James, Witten, Hastie y Tibshirani, 2013)



Convalidación Cruzada por Máxima Verosimilitud (CV)

- El EMV de h es degenerado: da $h = 0$, resultando una densidad que da masa 1 a cada uno de los datos.
- Alternativa: maximimizar una *pseudo–verosimilutd* computada sacando de los datos una observación a la vez (leave–one–out cross–validation).

  Si observamos los datos $x_1, \ldots, x_n$

  $$h^*_{CV} = \text{argmax}_h \left\{ \frac{1}{n} \sum_{i=1}^{n} \log \widehat{f}_h^{(-i)}(x_i) \right\}$$

  siendo $\widehat{f}_h^{(-i)}(x_i)$ la densidad estimada en el punto x_i sin utilizar al punto x_i:

  $$\widehat{f}_h^{(-i)}(x_i) = \frac{1}{(n-1)h} \sum_{j \neq i} K\left(\frac{x_i - x_j}{h}\right)$$

- En general, buscamos el máximo sobre una grilla $h_1, \ldots, h_q$ y luego, eventualmente, se refina.

Otra forma de implementar Convalidación Cruzada

Integrated Squared Error

$$
\begin{aligned}
\mathrm{ISE}[\widehat{f}_h(\cdot)] &= \int \left(\widehat{f}_h(x) - f(x)\right)^2 \,\mathrm{d}x \\
&= \int \left(\widehat{f}_h(x)\right)^2 \,\mathrm{d}x - 2\int \widehat{f}_h(x) f(x)\,\mathrm{d}x \\
&\quad + \int (f(x))^2 \,\mathrm{d}x.
\end{aligned}
$$

Como el último término no involucra a h, podemos minimizar

$$
A = \int \left(\widehat{f}_h(x)\right)^2 \,\mathrm{d}x - 2\int \widehat{f}_h(x) f(x)\,\mathrm{d}x
$$

Convalidación Cruzada

Bowman (1984) propone estimar a A por

$$\frac{1}{n}\sum_{i=1}^{n}\int\left(\widehat{f}_h^{(-i)}(x)\right)^2 dx-\frac{2}{n}\sum_{i=1}^{n}\widehat{f}_h^{(-i)}(x_i)$$

siendo $\widehat{f}_h^{(-i)}(x_i)$ la densidad estimada en el punto x_i sin utilizar al punto x_i:

$$\widehat{f}_h^{(-i)}(x_i)=\frac{1}{(n-1)h}\sum_{j\neq i}K\left(\frac{x_i-x_j}{h}\right)$$

Una versión más simple que se usa en general es:

$$\mathrm{LSCV}(h)=\int\left(\widehat{f}_h(x)\right)^2 dx-\frac{2}{n}\sum_{i=1}^{n}\widehat{f}_h^{(-i)}(x_i)$$

Convalidación Cruzada

Unbiased (o Least Square) Cross-validation

$$\mathrm{LSCV}(h) = \int \left(\widehat{f}_h(x)\right)^2 dx - \frac{2}{n}\sum_{i=1}^{n} \widehat{f}_h^{(-i)}(x_i)$$

Se busca el mínimo sobre una grilla $h_1, \ldots, h_q$ y luego, eventualmente, se refina.

(bw.ucv)