# Estimación de función de densidad

Gabriel Montes-Rojas

# Histogramas

- Supongamos $\{x_i\}_{i=1}^{n}$ una muestra aleatoria de la variable aleatoria $X$ con dominio en $\mathcal{X}$.
- Si $X$ es una variable aleatoria **discreta**[1] la densidad o probabilidad se puede estimar como

$$\hat{f}(x) = \frac{\sum_i 1\{x_i = x\}}{n},$$

  para todo $x$ en $\mathcal{X}$. Esto es un estimador de frecuencias o proporción tal que $\sum_{x \in \mathcal{X}} \hat{f}(x) = 1$.
- Se puede probar que $\hat{f}(x) \xrightarrow{p} f(x)$ y $\sqrt{n}(\hat{f}(x) - f(x)) \xrightarrow{d} Normal(0, f(x)(1 - f(x)))$ cuando $n \to \infty$.

[1] Una variable aleatoria es discreta si su dominio es un conjunto contable, finita o infinita.

# Histogramas

- Si $X$ es una variable aleatoria **continua** con dominio en $\mathcal{X}$ entonces debemos considerar intervalos que cubran el dominio de $\mathcal{X}$, que por ahora lo asumimos acotado. Sea $M$ el número de intervalos (bins), indexado por $m$, cada uno con el mismo **ancho de banda** (bandwidth) $h$. Entonces tenemos intervalos de forma

$$[x_m - h/2, x_m + h/2), \ m = 1, 2, ..., M$$

$$x_1 < x_2 < ... < x_m < ... < x_M, x_m + h/2 = x_{m+1} - h/2$$

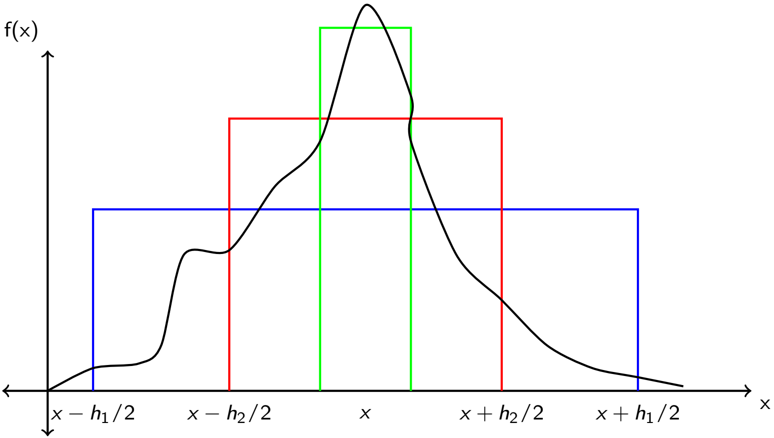
Definimos entonces el estimador de histograma,

$$\hat{f}(x) = \frac{\sum_{i=1}^{n} 1\left[x_i \text{ en intervalo que contiene } x\right]}{nh}.$$

# Histogramas

- Notar que $h \times \hat{f}(x) \xrightarrow{p} P[x - h/2 \leq X < x + h/2]$, cuando $n \to \infty$. También $P[x - h/2 \leq X < x + h/2] = \int_{x-h/2}^{x+h/2} f(z)dz$ que es la probabilidad del intervalo.
- Entonces $\hat{f}(x) \approx \frac{1}{h} \int_{x-h/2}^{x+h/2} f(z)dz \approx f(x)$ es un estimador de la densidad promedio en el intervalo, que es igual a $f(x)$ cuando $h \to 0$.
- ¿Cómo elegir $h$ óptimamente?
- Si $h \to 0$, $\hat{f}(x)$ es la proporción de observaciones con $x$. El resultado va a tener muchos saltos y ceros.
- Si $h \to \infty$, $\hat{f}(x) = 1, \forall x$. Demasiada suavización.

# Histogramas $h_2 < h_1$



# Kernel

Un kernel $K$ es una función simétrica tal que:

1. $\int K(\psi)d\psi = 1$;
2. $\int \psi K(\psi)d\psi = 0$;
3. $\int \psi^2 K(\psi)d\psi = \mu_2 < \infty$.

Nota 1: A menos que se especifique lo contrario (ej. si el dominio de $\mathcal{X}$ es acotado con bordes conocidos), los límites de la integración son $(-\infty, \infty)$.
Nota 2: En general $K(\psi) \geq 0$, $\forall\psi$, con lo que se puede asociar a una función de densidad (ej. normal). Sin embargo se pueden considerar kernels de orden alto donde puede haber valores negativos.

# Estimación de densidad por kernel

*Sea $h$ el ancho de banda (bandwith). Sea $x$ un valor particular donde queremos estimar $f(x)$.*
*El estimador de densidad de kernel es $\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)$.*

Nota 1: Si $\phi(x)$ es la función de densidad de una normal $\psi \sim N(0,1)$, entonces $\psi \sim N(\mu, \sigma^2)$ tiene función de densidad $\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})$. Así $h$ se asocia al desvío estándar. Notar que aplica también a los histogramas.
Nota 2: $\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} w_i$ donde $w_i = \frac{1}{h}K\left(\frac{x_i - x}{h}\right)$.

- En la práctica, estimamos la densidad de $x$ para todo el dominio, $\mathcal{X}$.
- Considermos una grilla $X = [x_1, x_2, ..., x_M]$, donde $\#X = M$ puntos en los cuales vamos a estimar la densidad.
- Entonces, una estimación de la densidad es el gráfico $\{\hat{f}_h(x(m)), x(m)\}_{m=1}^{M}$.
- Podríamos tener anchos de banda diferentes $h(x(m))$. En general deberíamos tener más ancho de banda para regiones con menor densidad.

# Kernels

| Kernel | Formula |
|---|---|
| Biweight | $K[z] = \begin{cases} \frac{15}{16}(1-z^2)^2 & \text{if } \lvert z\rvert < 1 \\ 0 & \text{otherwise} \end{cases}$ |
| Cosine | $K[z] = \begin{cases} 1+\cos(2\pi z) & \text{if } \lvert z\rvert < 1/2 \\ 0 & \text{otherwise} \end{cases}$ |
| Epanechnikov | $K[z] = \begin{cases} \frac{3}{4}(1-\frac{1}{5}z^2)/\sqrt{5} & \text{if } \lvert z\rvert < \sqrt{5} \\ 0 & \text{otherwise} \end{cases}$ |
| Epan2 | $K[z] = \begin{cases} \frac{3}{4}(1-z^2) & \text{if } \lvert z\rvert < 1 \\ 0 & \text{otherwise} \end{cases}$ |
| Gaussian | $K[z] = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$ |
| Parzen | $K[z] = \begin{cases} \frac{4}{3}-8z^2+8\lvert z\rvert^3 & \text{if } \lvert z\rvert \leq 1/2 \\ 8(1-\lvert z\rvert)^3/3 & \text{if } 1/2 < \lvert z\rvert \leq 1 \\ 0 & \text{otherwise} \end{cases}$ |
| Rectangular | $K[z] = \begin{cases} 1/2 & \text{if } \lvert z\rvert < 1 \\ 0 & \text{otherwise} \end{cases}$ |
| Triangular | $K[z] = \begin{cases} 1-\lvert z\rvert & \text{if } \lvert z\rvert < 1 \\ 0 & \text{otherwise} \end{cases}$ |

# Supuestos

- A1. Las observaciones $x_1, x_2, \ldots, x_n$ son iid.
- A2. El kernel $K$ cumple con las propiedades de arriba.
- A3. Las derivadas de segundo orden de $f$ son continuas y acotadas en un vecindario de $x$.
- A4. $h = h_n \to 0$ cuando $n \to \infty$.
- A5. $nh_n \to \infty$ cuando $n \to \infty$.

# Propiedades asintóticas

- Definamos el error cuadrático medio (mean squared error, MSE):

$$MSE(x) = E[(\hat{f}_h(x) - f(x))^2] = Var(\hat{f}_h(x)) + \left(Sesgo(\hat{f}_h(x))\right)^2,$$

  donde $Sesgo(\hat{f}_h(x)) = E\left[\hat{f}_h(x) - f(x)\right]$.

- Siguiendo a Pagan&Ullah pp.23-24:

$$Sesgo(\hat{f}_h(x)) \approx \frac{h^2}{2} f''(x)\mu_2$$

$$Var(\hat{f}_h(x)) \approx \frac{f(x)}{nh} \int K(\psi)^2 d\psi$$

- Notar que para un estimador consistente de $f(x)$ se debe cumplir $MSE \to 0$ cuando $n \to \infty$, y tenemos un trade-off entre sesgo y varianza:
  1. $h \to 0$ cuando $n \to \infty$ para reducir el sesgo.
  2. Pero necesitamos que $nh \to \infty$, o sea, $n$ debe crecer más rápido que $h$ reducirse para eliminar la varianza.

# Propiedades asintóticas

## Pruebas del sesgo y varianza

- Notar que con los supuestos, $w_i = \frac{1}{h} K\left(\frac{x_i - x}{h}\right)$ también forma una muestra aleatoria iid con $E(w_i) = E(w_1)$, $V(w_i) = V(w_1)$, $Cov(w_i, w_j) = 0$, $i \neq j$.
- Sesgo. $E\hat{f} = \frac{1}{n}\sum_{i=1}^{n} Ew_i = Ew_1$
$= h^{-1} E\left[K\left(\frac{x_1 - x}{h}\right)\right] = h^{-1}\int K\left(\frac{x_1 - x}{h}\right) f(x_1) dx_1$. Ahora haciendo cambio de variables $= \int K(\psi) f(h\psi + x) d\psi$. Entonces el sesgo es
$E\hat{f} - f = \int K(\psi)(f(h\psi + x) - f(x)) d\psi$.
- Ahora tomemos la siguiente expansión de Taylor
$f(h\psi + x) = f(x) + h\psi f'(x) + \frac{h^2\psi^2}{2} f''(x) + ....$
- Si lo aplicamos al sesgo, obtenemos $Sesgo(\hat{f}_h(x)) = \frac{h^2}{2} f''(x)\mu_2 + o(h^2)$.
- Ahora calculemos la varianza. $V\hat{f} = n^{-1} Vw_1 = n^{-1} Ew_1^2 - n^{-1}(Ew_1)^2$. Por un lado $E\hat{f}^2 = n^{-1} Ew_1^2 = (nh^2)^{-1} EK(\psi)^2 =$
$(nh)^{-1}\int K(\psi)^2 \left\{f(x) + h\psi f'(x) + \frac{h^2\psi^2}{2} f''(x) + ...\right\} d\psi =$
$(nh)^{-1} f(x)\int K(\psi)^2 d\psi + o((nh)^{-1})$. Entonces,
$V\hat{f} = (nh)^{-1} f(x)\int K(\psi)^2 d\psi + o((nh)^{-1}) + o(n^{-1})$.
Notar que necesitamos que $E[w_1^2] < \infty$ lo que implica $E[w_1] < \infty$.

# Propiedades asintóticas
## MISE

- Dado que nos interesa la densidad en todo el dominio $\mathcal{X}$, debemos considerar el MSE integrado (mean integrated squared error, MISE):

$$MISE = \int_{\mathcal{X}} MSE(x)dx = \frac{1}{nh}\int K(\psi)^2 d\psi + \mu_2^2 \frac{h^4}{4}\int_{\mathcal{X}} (f''(x))^2 dx$$

- Tomemos derivadas con respecto a $h$, para obtener

$$h_{opt} = \left(\int K(\psi)^2 d\psi\right)^{1/5} \mu_2^{-2/5} \left(\int_{\mathcal{X}} (f''(x))^2 dx\right)^{-1/5} n^{-1/5}$$

  Esto es, $h$ debería ir a 0 a tasa $-1/5$ con respecto a $n$. Así debemos considerar $h_n$, como una función de $n$.

- Esto significa que la **tasa de convergencia** es $O_p(n^{-2/5})$ lo que nos da una tasa más *lenta* que la estimación paramétrica de $f(x)$, $O_p(n^{-1/2})$, o sea $\sqrt{n}$-consistencia. Para obtener esto tenemos que calcular $\frac{1}{\sqrt{nh}}$ que es la tasa que se aplica a la sumatoria para estandarizar la varianza.
- Por otro lado $MISE = max((nh)^{-1}, h^4) = O(n^{-4/5})$.

# Estimación de densidad por kernel

- La parte más importante de los estimadores no paramétricos es seleccionar $h$.
- Si $h \to 0$, entonces no hay suavización. La densidad va a tener muchos saltos y ceros.
- Si $h \to \infty$, entonces hay demasiado suavizamiento.
- Notar que para $h_{opt}$ necesitamos
  - $\int K(\psi)^2 d\psi$. Depende del kernel.
  - $\mu_2 = \int \psi^2 K(\psi) d\psi$. Depende del kernel.
  - $\int_{\mathcal{X}} f''(x) dx$. Depende de $f$ que no es conocida.

# Selección de ancho de banda

- **Rule-of-thumb**: Usar la distribución normal estándar $\int_{\mathcal{X}}(f''(x))^2 dx = \frac{3}{8\sqrt{\pi}\sigma^5}$. Por otro lado, supongamos también que $K$ es el kernel gaussiano. Entonces,

$$\hat{h}_{opt} = 1.059\hat{\sigma}n^{-1/5},\ con\ \hat{\sigma} = \sqrt{\frac{1}{n}\sum_i (x_i - x)^2}$$

- Silverman (1986) propone $h = 0.79Rn^{-1/5}$ y $h = 0.9An^{-1/5}$ donde $R$ es el rango intercuartiles y $A = min(\sigma, R/1.34)$. Esto es lo que usa `STATA`.
- **Plug-in**: Usar un estimador de $f$, luego computar $f''$ y re-estimar.
- **Cross-validation**: Ver Pagan&Ullah pp.50-52.
- Usar la intuición! **La econometría no paramétrica es un arte.**

# Otras propiedades asintóticas
## Normalidad asintótica

- Si a los supuestos A1-A5 le agregamos A15. $(nh)^{1/2}h^2 \to 0$ cuando $n \to \infty$, entonces
$$(nh)^{1/2}(\hat{f} - f) \xrightarrow{d} Normal\left(0, f\int K^2(\psi)d\psi\right).$$
- Para armar los intervalos de confianza al nivel de significatividad $1-\alpha$:
$$\hat{f} \pm Z_{1-\alpha/2}(nh)^{-1/2}\left[\hat{f}\int K^2(\psi)d\psi\right]^{1/2},$$
donde $\Phi(Z_{1-\alpha/2}) = 1-\alpha/2$.

# STATA

- Histogramas `hist`:
  `www.stata.com/manuals13/rhistogram.pdf`
- Kernel `kdensity`:
  `www.stata.com/manuals13/rkdensity.pdf`

# Referencias


- *Pagan, A. and Ullah, A. (1999), Nonparametric Econometrics. Cambridge: Cambridge University Press.*
- *Racine, J. (2008), "Nonparametric Econometrics: A Primer", Foundations and Trends in Econometrics, 3(1), 1-88.*