Nonparametric estimation

- Problem sets for the first half. Nothing for the second half.

- Give us complete flexibility
- Good for graphs in a couple dimensions.

E.g. Deaton ('89)

- p - price of rice
- q - amount purchased
- y - amount sold
- change in benefits from dp is $dB = (q-y)dp$

$$= p(q-y)\,d\ln(p)$$

$$\Rightarrow \quad \frac{dB}{x}\,d\ln(p) = \left(w - \frac{py}{x}\right)$$

- can graph densities using nonparametric estimation techniques.

Empirical distribution function.

- The CDF of Z is $F_Z(z) = \Pr[Z \le z]$. Let $Z_1, \ldots, Z_n$ be iid data, $1(A)$ indicator function. Then, we have

$$F_Z(z) = E[1(Z_i \le z)]$$

$$\hat{F}_Z(z) = \frac{\#\{i : Z_i \le z\}}{n} = \frac{1}{n}\sum_{i=1}^{n} 1(Z_i \le z)$$

- puts weight $\frac{1}{n}$ on each observation
- consistent and asymptotically normal
- nonparametrically efficient

- No good for densities, because it will be a bunch of point masses.

$\theta_0 = E[\mathbb{1}(Z_i \leq z)]$

$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(Z_i \leq z)$

$Var(\hat{\theta}) = \frac{1}{n} Var(\mathbb{1}(Z_i \leq z)) = \frac{F_{Z_i}(z)[1 - F_{Z_i}(z)]}{n}$

Thus, $\sqrt{n}(\hat{F}(z) - F(z)) \xrightarrow{d} N(0, F(z)(1-F(z)))$

can estimate variance covariance matrix:

$\widehat{F(z)[1-F(z)]} = \hat{F}(z)[1-\hat{F}(z)]$

The human eye is much better at comprehending density estimates.

- Let us smooth out our cdf by adding some continuous noise

Kernel Density Estimator

Let $\bar{Z}_n$ have empirical cdf, u be continuous with pdf $K(u)$, independent of $\bar{Z}_n$

Let $h > 0$, Define $\tilde{Z} = \bar{Z}_n + hu$

- empirical CDF plus noise. Kernel density estimator is density of $\tilde{Z}$

Derivation: Let $F_u(u) = \int_{-\infty}^{u} K(t)dt$ be the CDF of u

By iterated expectations,

$$E[1(\tilde{z} \le z)] = E[E[1(\tilde{z} \le z) | \bar{z}_n]]$$
$$= E[E[1(u \le \frac{z - \bar{z}_n}{h}) | \bar{z}_n]]$$
$$= E[F_u(\frac{z - \bar{z}_n}{h})]$$

$$\Rightarrow F_{\tilde{z}}(z) = \frac{1}{n} \sum_{i=1}^{n} F_u\left(\frac{z - z_i}{h}\right)$$

$$\Rightarrow \frac{dF_{\tilde{z}}(z)}{d\tilde{z}} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{z - z_i}{h}\right) = \hat{f}_n(z).$$

h controls the amount of smoothing
- h small $\Rightarrow$ spiky density
- h large $\Rightarrow$ more bias, smoother density

$\hat{f}_n(z)$ consistent if $h \to 0$ and $nh \to +\infty$

Examples: $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$. Gaussian kernel

$K(u) = 1(|u| \le 1)\frac{3}{4}(1-u^2)$ Epanechnikov kernel

Bias and variance

Suppose $(z_1, \dots, z_n)$ are iid

$$\hat{f}_n(z) = \frac{1}{n} \sum_{i=1}^{n} K_h(z - z_i), \quad K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$$

Thus, $E[\hat{f}_h(z)] = \int K_h(z-t) f_0(t)dt = \frac{1}{h} \int K\left(\frac{z-t}{h}\right) f_0(t)dt$

$$= \int K(u) f_0(z - hu)du$$

Taylor expansion around $h=0$:

$$f_0(z-hu) = f_0(z) - f'(z)hu + \Gamma(h,u,z)h^2$$

$$\Gamma(h,u,z) = f_0''(z+\bar{h}(z,u)u)\,u^2/2$$

where $|\bar{h}| \leq |h|$

$$\Rightarrow E[\hat{f}_h(z)] = f_0(z) + h^2 f_0''(z)\int K(u)u^2\,du/2 + o(h^2)$$

Thus, $E[\hat{f}_h(z)] - f_0(z) = h^2 f_0''(z)\int K(u)u^2\,du/2 + o(h^2)$ under regularity conditions

$$\begin{aligned} Var(\hat{f}_h(z)) &= E[K_h(z-Z_i)^2] - [E[K_h(z-Z_i)]]^2 \cdot \frac{1}{n} \\ &= \frac{1}{h^2}\int K\left(\frac{z-t}{h}\right)^2 f_0(t)\,dt \cdot \frac{1}{n} + O\left(\frac{1}{n}\right) \\ &= \frac{1}{h}\int K(u)^2 f_0(z-hu)\,du \cdot \frac{1}{n} + O\left(\frac{1}{n}\right) \end{aligned}$$

$$\int K(u)^2 f_0(z-hu)\,du \to f_0(z)\int K(u)^2\,du$$

assuming $h \to 0$, $O\left(\frac{1}{n}\right) = o\left(\frac{1}{nh}\right)$, so we have

$$Var(\hat{f}_h(z)) = f_0(z)\int K(u)^2\,du \cdot \frac{1}{nh} + o\left(\frac{1}{nh}\right)$$

Consistency and convergence rates

$h \to 0$: bias goes to zero

$nh \to \infty$: variance goes to zero

Bandwidth shrinks slower than $\frac{1}{n}$.

The mean-squared error is given by:

$$MSE(\hat{f}_h(z)) = Var(\hat{f}_h(z)) + Bias^2(\hat{f}_h(z))$$
$$= f_0(z)\int K(u)^2 du \cdot \frac{1}{nh}$$
$$+ h^4 \left(f_0''(z)\int K(u)u^2 du \cdot \frac{1}{2}\right)^2$$
$$+ o\left(h^4 + \frac{1}{nh}\right)$$

For nonparametric estimators, we will get something like: $n^\alpha(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V)$

Bandwidth choice

- Choose one that looks good. Report several
- Or you can minimize asymptotic MSE

  $\Rightarrow h = \left[\frac{C_1}{n C_2 \int f_0''(z)^2 dz}\right]^{1/5}$, but $f_0''(z)$ needs to be estimated.
- Silverman's rule of thumb
  - when $f_0(z)$ is Gaussian and $K(u)$ is $N(0,1)$,

    $h = 1.06\sigma n^{-1/5}$, $\sigma = Var(z_i)^{1/2}$
- Can minimize estimate of integrated MSE.
  - $\hat{CV}(h) = \int \hat{f}_h(z)^2 dz - \frac{2}{n(n-1)} \sum_{i \neq j} K_h(z_i - z_j)$

  (cross-validation)

Multivariate Density Estimation

Let z be $r \times 1$, $K(u)$ be a pdf for an $r \times 1$ rv.

e.g. $K(u) = \prod_{j=1}^{r} k(u_j)$

Let $\hat{\Sigma}$ be sample covariance matrix of Z_i. Let

$$K_h(u) = h^{-r} \det(\hat{\Sigma})^{-1/2} K(\hat{\Sigma}^{-1/2} u/h)$$

$$\Rightarrow \hat{f}_h(z) = \frac{1}{n} \sum_{i=1}^{n} K_h(z - Z_i)$$

Can use $K(u) = (2\pi)^{-r/2} \exp\{-\frac{u'u}{2}\}$ Gaussian

$K(u) = C_r (1 - u'u) 1(u'u \leq 1)$

as r gets big, this becomes more and more difficult, and the rate of convergence decreases. (ie at rate $\frac{1}{nh^r}$, the variance shrinks)