1.3 Measures of Performance

(Error).

Squared Error

$$SE(\hat{f}(y)) = [\hat{f}(y) - f(y)]^2$$

local : at a point y

depends on realization $Y_1, \dots, Y_n$ through $\hat{f}$

Mean Squared Error

$$MSE(\hat{f}(y)) = E_f([\hat{f}(y) - f(y)]^2) = Var(\hat{f}(y)) + [bias(\hat{f}(y))]^2$$

local: at a point y

but now describes mean of error (property of dsn).

Integrated Squared Error

$$ISE = \int_{-\infty}^{\infty} [\hat{f}(u) - f(u)]^2 du$$

no longer local

depends on realization $Y_1, \dots, Y_n$

Mean Integrated Squared Error

$$MISE = \int_{-\infty}^{\infty} MSE(\hat{f}(u)) du$$

not local

describes property of dsn of error.

this is what we are estimating => we don't know it!

Of course all theoretical because we have to know f to calculate them

↳ useful for discussing properties of our estimator $\hat{f}$.

Total Integrated squared bias:

$$ISB \approx \frac{h^3}{12} \sum_j \left[f'(\eta_j)\right]^2 = \frac{h^2}{12} \sum_j \left[f'(\eta_j)\right]^2 h$$

$$= \frac{h^2}{12} \left[\int \left[f'(x)\right]^2 dx + o(1) \right]$$

$$= \underbrace{\frac{h^2}{12} \int \left[f'(x)\right]^2 dx}_{\text{called AISB "asymptotic ISB"}} + o(h^2).$$

$$= \frac{1}{12} h^2 R(f').$$

$$MISE = IV + ISB = \frac{1}{nh} - \frac{R(f)}{n} + o(n^{-1}) + \frac{1}{n} h^2 R(f') + o(h^2).$$

$$= \underbrace{\frac{1}{nh} + \frac{1}{n} h^2 R(f')}_{AMISE} + o(n^{-1}) + o(h^2)$$

narrower bins give an estimator that is less biased but more variable. As $h \to 0$ $\hat{f} \to$ set of spikes at each obs w/ 0 bias.

The minimizer of AMISE is $h_0 = \left[\frac{6}{R(f')}\right]^{1/3} n^{-1/3}$ and minimum AMISE is $AMISE_0 = \left[\frac{9R(f')}{16}\right]^{1/3} n^{-2/3}$.

The roughness of the underlying density, as measured by $R(f')$ determines the optimal level of smoothing and the accuracy of the histogram estimate.

Densities w/ few bumps (smaller $R(f')$) require wider bins, while bumpy densities (large $R(f')$) require smaller bins.

We cannot find the optimal binwidth without known the density f itself.

this is the thing we are estimating!

Simple (plug-in) approach: Assume f is a $N(\mu, \sigma^2)$, then

$$h_0 = 3.491\, \sigma\, n^{-1/3}$$

could use sample st. deviation or interquartile range to estimate.

For non-normal data, multiple modes inflate $\sigma^2 \Rightarrow$ Gaussian plugin histogram will be oversmoothed.

No theoretical justification, just something we can do and often passes the "eye test" "cross-validation"

Data driven approach:

$$ISE = \int [f(u) - \hat{f}(u)]^2 du$$

$$= R(f) + \underbrace{R(\hat{f})}_{\text{can be computed in closed form}} - 2\int \hat{f}(u) f(u) du$$

irrelevant for choosing h

this is not wrt the data sample we have.

$$-2\int \hat{f}(u) f(u) du = -2E[\hat{f}(u)], \ U \sim f$$

⇒ one idea is to estimate something proportional to ISE so that we could find h that would minimize.