10.1 Estimating the Probability Density Function

- Consider the problem of estimating a continuous pdf $f(x)$.

Histogram:

- Suggested width, $d = 3.5 \frac{S}{\sqrt[3]{n}}$, S = sample standard deviation

10.1.1 Kernel Method

- involves taking a certain weighted average of data points near X to estimate $f(x)$.
- Kernel Estimate of $f(x)$,

$$\hat{f}(x) = \frac{1}{n\Delta}\sum_{i=1}^{n} W\left(\frac{X - X_i}{\Delta}\right)$$, where $W(z)$ = kernel, Δ = bandwidth $= \frac{1.06}{\sqrt[5]{n}} S$

* Assume $W(z)$ is the standard normal pdf

10.2 Nonparametric Curve Smoothing

- In many cases, we may not know the functional form of a model, meaning we don't know if the actual model is linear, quadratic, or other forms. Then we may use nonparametric regression, smoothing specifically.
- Many smooth nonparametric estimates of $\phi(X)$ involve giving greater weight to pairs of observations (X_i, Y_i) for which X_i is near X and lesser weight to other pairs.

splines, loess, kernel

10.2.1 Loess Method

- note that $\phi(x)$ can often be adequately approximated by a linear function $l(x) = \beta_0 + \beta_1(X - X_0)$ when X is near X_0.
- first determine the k values of the X_i's nearest to X_0 (Span $= \frac{k}{n}$)

* Span determines the smoothness of the approximation

$\uparrow$ Span, $\downarrow$ complexity (smooth or constant model)

$\downarrow$ Span, $\uparrow$ complexity (Complex model)

- The loess approximation uses weighted least squares to find the $l(x)$ that minimizes,

$$\sum_{X_i \in N_k(X_0)} [Y_i - l(X_i)]^2 \cdot W\left(\frac{|X_0 - X_i|}{\Delta_{X_0}}\right)$$, where

$l(x)$ = can be either linear, quadratic, or higher terms.

$N_k(X_0)$ = k nearest X_i's to the point X_0

$W(u) = (1-u^3)^3$

$\Delta_{X_0} = \max_{X_i \in N_k(X_0)} |X_i - X|$, meaning it is the farthest point X_i among k nearest points from X_0, so u becomes the standardization.

10.2.2 Kernel Method

- Kernel Method is determined by the choice of h = bandwidth.

  $\uparrow h$, $\downarrow$ complexity $\downarrow h$ $\uparrow$ complexity

- Kernel Method uses points within the range of $X_0 \pm h$, whereas KNN and Loess uses a predetermined k number of points, universally.
- The conditional expectation of Y given $X = x$ is,

$$\phi(x) = \int y \cdot \frac{f(x,y)}{f(x)}\, dy$$

, so estimating $\phi(x)$ by $\hat{\phi}(x)$

is obtained by estimating $f(x,y)$ and $f_X(x)$ by the kernel method and then performing the integration to get an approximation of the conditional expectation.

- $\hat{f}(x,y)$ can be obtained,

$$\hat{f}(x,y) = \frac{1}{n \Delta_x \Delta_y} \sum_{i=1}^{n} w\left(\frac{x-X_i}{\Delta_x}\right) w\left(\frac{y-Y_i}{\Delta_y}\right)$$

$$\Rightarrow \hat{\phi}(x) = \frac{\sum_{i=1}^{n} Y_i\, w\left(\frac{x-X_i}{\Delta_x}\right)}{\sum_{i=1}^{n} w\left(\frac{x-X_i}{\Delta_x}\right)}$$

, where n is the number of points within the bandwidth.

KNN	Loess	Kernel
- take the averange of k nearest X_i's from the X_0 - more like step function	- Each point X_0 has unique function with k nearest X_i's - The closer, the more weights	- Each point X_0 has unique function within the range of $X_0 \pm h$. - Uses the conditional pdf