1.4 Optimal Binwidth

$\hat{f}(y) = \frac{n_j}{n \cdot h}$ for $y \in (b_j, b_{j+1}]$ and $h = b_{j+1} - b_j$.

count of points in $(b_j, b_{j+1}]$

We will investigate bias and variance of $\hat{f}$ pointwise, because
$\text{MSE}(\hat{f}(y)) = (\text{bias}(\hat{f}(y))^2 + \text{Var}\hat{f}(y)$.

local

$n_j \sim \text{Binomial}(n, p_j)$ where $p_j = P(b_j < Y \le b_{j+1}) = \int_{b_j}^{b_{j+1}} f(y)\,dy$ (if density exists).

$\Rightarrow E[\hat{f}(y)] = \frac{np_j}{n \cdot h} = \frac{p_j}{h}$ and $\text{bias}(\hat{f}(y)) = \frac{p_j}{h} - f(y)$

$\text{Var}[\hat{f}(y)] = \frac{1}{n^2h^2} n p_j(1-p_j) = \frac{1}{nh^2} p_j(1-p_j)$

Assumption: Let's suppose $f(y)$ is Lipschitz continuous over the interval $B_j = (b_j, b_{j+1}]$, i.e. $\exists$ a constant γ_j st. $|f(x) - f(y)| < \gamma_j |x-y|$ $\forall x,y \in B_j$

Then by MVT $p_j = \int_{B_j} f(y)\,dy = h f(\xi_j)$ for some $\xi_j \in B_j$.

$$\Rightarrow \text{Var}[\hat{f}(y)] = \frac{p_j(1-p_j)}{nh^2} \overset{p_j \in [0,1]}{\le} \frac{p_j}{nh^2} = \frac{hf(\xi_j)}{nh^2} = \underbrace{\frac{f(\xi_j)}{nh}}_{\text{increases as } h \to 0 \text{ and decreases w/ } n.} \text{ for some } \xi_j \in B_j$$

$$\text{and } |\text{Bias}[\hat{f}(y)]| = \left|\frac{p_j}{h} - f(y)\right| = |f(\xi_j) - f(y)| \le \gamma_j |\xi_j - y| \le \underbrace{\gamma_j h}_{\text{decrease as } h \to 0 \text{ and unaffected by } n.}$$

Thus $\text{MSE}[\hat{f}(y)] = [\text{bias}\,\hat{f}(y)]^2 + \text{Var}\hat{f}(y) \le \gamma_j^2 h^2 + \frac{f(\xi_j)}{nh} \equiv M$.

So, if f is Lipschitz continuous over B_j, $\hat{f}(y)$ is $\underbrace{\text{mean square consistent}}_{\lim_{n\to\infty} \text{MSE}(\hat{f}(y)) = 0}$ if as $n \to \infty$, $h \to 0$ and $nh \to \infty$.

optimal bin width based on MSE

$$\frac{\partial M}{\partial h} = \frac{-f(\xi_j)}{nh^2} + 2\gamma_j^2 h \overset{\text{set}}{=} 0$$

$$2\gamma_j^2 h^3 = \frac{f(\xi_j)}{n} \Rightarrow h = \left[\frac{f(\xi_j)}{2\gamma_j^2 n}\right]^{1/3} \Rightarrow$$ optimal bin width decreases at a rate proportional to $n^{-1/3}$.

optimal $\text{MSE}[\hat{f}(x)] = \frac{f(\xi_j)}{n(\alpha n^{-1/3})} + \gamma_j^2 (\alpha n^{-1/3})^2 = K n^{-2/3}$ So MSE is not rate n^{-1} (parametric estimators), $\underbrace{\text{but instead } n^{-2/3}}_{\text{cost of being nonparametric.}}$

Global histogram Error: Consider integrated bias & variance separately.

integrated variance

$$IV = \int_{-\infty}^{\infty} \text{Var}(\hat{f}(y))\,dy = \sum_j \int_{B_j} \text{Var}\hat{f}(y)\,dy = \sum_j \int_{B_j} \frac{p_j(1-p_j)}{nh^2}\,dy = \sum_j \frac{p_j(1-p_j)}{nh} = \frac{1}{nh}\Big[\underbrace{\sum_j p_j}_{=1} - \sum_j p_j^2\Big]$$

$$\sum_j p_j^2 \overset{\text{MVT}}{=} \sum_j f(\xi_j)^2 h^2 = h \sum_j f(\xi_j)^2 h = h\Big[\underbrace{\int f^2(y)\,dy + O(1)}\Big]$$

as $h \to 0$, difference between $\hat{f}$ and $f \to 0$

So, $IV = \frac{1}{nh}\left(1 - h\int f^2(y)\,dy + hO(1)\right) = \frac{1}{nh} - \frac{R(f)}{h} + o(n^{-1})$ where $R(f) := \int f^2(y)\,dy$ "roughness"

Consider a typical bin $B_0 = (0, h]$

The bin probability $p_0 = \int_0^h f(t)\,dt = \int_0^h [f(x) + (t-x)f'(x) + \frac{(t-x)^2}{2!} f''(x) + \ldots]\,dt = \left[t f(x) + \frac{(t-x)^2}{2} f'(x) + \frac{(t-x)^3}{3 \cdot 2!} f''(x) + \ldots\right]_0^h$

$$= hf(x) + \left[\frac{(h-x)^2}{2} - \frac{x^2}{2}\right] f'(x) + O(h^3) = hf(x) + h\left(\frac{h}{2} - x\right)f'(x) + O(h^3).$$

$\Rightarrow$ bias at a point in B_0 is $\text{Bias}(\hat{f}(x)) = \frac{p_0}{h} - f(x) = \left(\frac{h}{2} - x\right)f'(x) + O(h^2)$.

Integrated Squared Bias for bin

$$ISB_0 \approx \int_{B_0} \left(\frac{h}{2} - x\right)^2 f'(x)^2\,dx \underset{\text{generalized MVT}}{=} f'(\eta_0) \int_0^h \left(\frac{h}{2} - x\right)dx = \frac{h^3}{12}[f'(\eta_0)]^2$$