Stat 550 Virtual Whiteboard
Chapter 3 Work Sheets

David W Scott

Rice University

September 14, 2023

Variance Stabilizing (Poisson)

$X_i \sim \mathrm{Pois}(1)$

$Y = \sum_1^n X_i \sim \mathrm{Pois}(n) \approx N(n, \sqrt{n}^2)$ (CLT)

$\approx n + Z\sqrt{n} \quad = n\left(1 + \frac{1}{\sqrt{n}} Z\right)$

$\therefore \sqrt{Y} \approx \sqrt{n}\sqrt{1 + \frac{1}{\sqrt{n}} Z}$

$\approx \sqrt{n} \cdot \left(1 + \frac{1}{2\sqrt{n}} Z\right)$

$= \sqrt{n} + \frac{1}{2} Z \sim N\left(\sqrt{n}, \frac{1}{2^2}\right)$

i.e. (variance indep of n)

$Bin(n, p)$ $\quad p = \frac{1}{2}$

$$\binom{0}{0}$$

$$\binom{1}{0}\binom{1}{1}$$

$$\binom{2}{0}\binom{2}{1}\binom{2}{2}$$

$$
\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & 1 & & \\
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1
\end{array}
$$

in the k^{th} row

$$\binom{k}{0}\binom{k}{1}\binom{k}{2}\cdots\binom{k}{k-2}\binom{k}{k-1}\binom{k}{k}$$

$$n = \sum_{i=0}^{k}\binom{k}{i} = (x+y)^k = \sum_{i=0}^{k} x^i y^{k-i}\binom{k}{i}$$

$$= 2^k$$

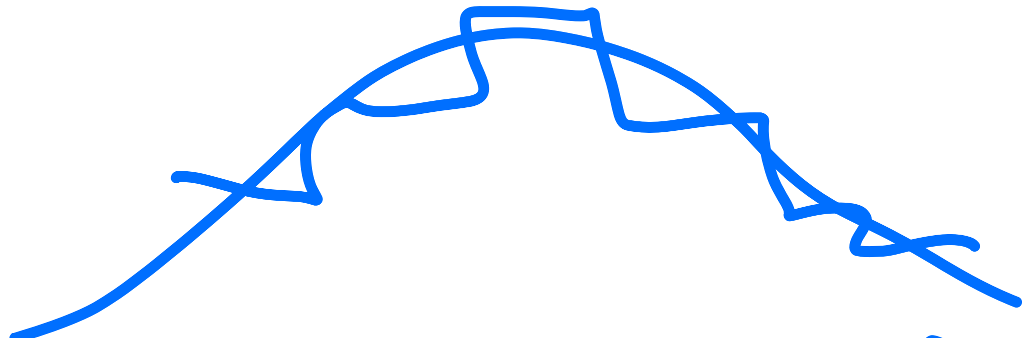
$$k = \log_2 n$$

$$\#\text{bins} = k+1 = 1+\log_2 n$$

L_2 Theory

ISE → IMSE, MISE

$\hat f_h$ vs g



$$ISE = \int (\hat f_h - g)^2$$

$$= \sum_k \int_{B_k} \left(\hat f_h^2 - 2\hat f_h g + g^2\right)$$

$$\left(\frac{y_k}{nh}\right)^2 \cdot h \qquad -2 \hat f_h \int_{B_k} g \qquad R(g)$$

roughness

$$\sum_k \frac{y_k^2}{n^2 h} - 2 \sum_k \frac{y_k p_k}{nh} + R(g)$$

II IMSE $= \int \mathrm{mse}\left(\hat{f}_h(x)\right)$

$\operatorname{mse}\left(\hat{f}_h(x)\right) = \operatorname{Var}\hat{f}_h(x) + \left(E\hat{f}_h(x) - f(x)\right)^2$

Bk

$\Rightarrow$ IMSE $=$ IV $+$ ISB

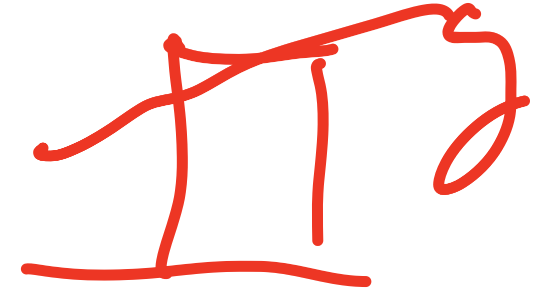
$$\hat{f}_h(x) = \frac{V_k}{nh} \begin{cases} E\hat{f}_h(x) = \dfrac{np_k}{nh} \\ \operatorname{Var} = \dfrac{np_k(1-p_k)}{n^2h^2} \geq \dfrac{p_k}{nh} \end{cases}$$

Bias

$$E\,\hat f_h(x) = \frac{p_k}{h}$$

$$\Downarrow$$

$$g(m_k)$$



B_k

$$p_k \doteq h\, g(m_k)$$

$$\boxed{p_k = h\, g(\xi_k)}, \quad \xi_k \in B_k$$

$$\text{Bias} \doteq g(m_h) - g(x)$$

$$\text{Bias}^2 \overset{?}{=} O(h^2)$$

$$\frac{o(h)}{h} \qquad \frac{O(h)}{h}$$

$$\mathrm{MSE}\ \hat{f}_h(x) = \frac{p_k(1-p_k)}{nh^2} + \left(g(\xi_k) - g(x)\right)^2$$

IMSE? — try $h = \frac{1}{\sqrt{n}}$

$$IV = \int = \sum_k \int_{B_k} \frac{p_k(1-p_k)}{nh^2} - \frac{R(g)}{n}$$

$n \to \infty$

$h \to 0$??

$nh \to \infty$

$$= \sum_k \frac{p_k(1-p_k)}{nh} = \frac{1}{nh} - \sum \frac{p_k^2}{nh}$$

for the 2nd term

$$\frac{1}{nh} \sum_k \left(h\, g(m_k)\right)^2 = \boxed{\frac{1}{n} \sum_k g(m_k)^2 \cdot h}$$

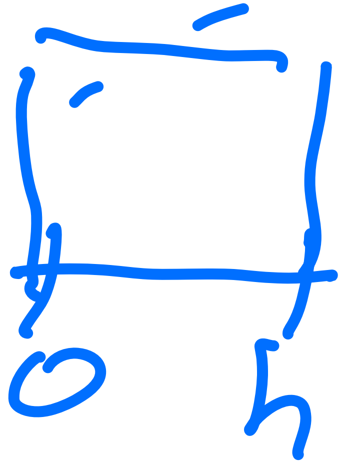
Steps $h = \dfrac{1}{1+\log_2 n}$

$$nh = \frac{n}{1+\log_2 n} \to \infty$$

Works!!

Back to bias — g bias

w.l.o.g.



$E\hat{f} - g = \frac{P_0}{h} - g(x)$

$$P_0 = \int_0^h g(x)\,dx = \int_0^h \left[g(0) + x g'(0) + \frac{1}{2} x^2 g''(0) + \dots \right]$$

$$= h\, g(0) + \frac{1}{2} h^2 g'(0) + \frac{1}{6} h^3 g''(0) + \dots$$

$$\text{Bias}(x) = g(0) + \frac{1}{2} h g'(0) + \frac{1}{6} h^2 g''(0).$$

$$E\hat{f} - g(x) = \dots - \left[g(0) + x g'(0) + \frac{1}{2} x^2 g''(0) \right]$$

$$\left(\frac{h}{2} - x\right) s'(0) = B_{x,y}$$

$$\int_0^h \left(\frac{h}{2} - x\right)^2 s'(0)^2 \, dx$$

$$y = x - h/2$$

$$\int_{-h/2}^{h/2} y^2 \, dy = 2 \cdot \left(\frac{h}{2}\right)^3 \frac{1}{3} = \frac{1}{12} h^3 s'(0)^2$$

$$ISB = \sum_k \frac{1}{12} h^{2} \, s'(t_k)^2 \cdot h \to \frac{1}{12} h^2 R(s')$$

$$\mathrm{IMSE} \simeq \frac{1}{nh} + \frac{1}{12} h^2 R(g')$$

$n \to \infty$

$h \to 0$

$nh \to \infty$

$$h = \frac{c}{\sqrt{n}} \qquad \frac{c}{n^{1/3}}$$

$$= \frac{1}{c\sqrt{n}} + \frac{1}{12} \frac{c^2}{n} R(g')$$

$$\frac{1}{c n^{2/3}} + \frac{1}{12} \frac{c^2}{n^{2/3}} R(g')$$