The Bias-Variance Tradeoff

PUBH 8442: Bayes Decision Theory and Data Analysis

Eric F. Lock
UMN Division of Biostatistics, SPH
elock@umn.edu

02/3/2021

Unbiased rules

- A rule $d(\mathbf{y})$ is unbiased if risk under a given $\theta \in \Theta$ is minimized at θ

$$E_{\mathbf{y} \mid \theta}\, l(\theta', d(\mathbf{y})) \geq E_{\mathbf{y} \mid \theta}\, l(\theta, d(\mathbf{y}))$$

  for all $\theta', \theta \in \Theta$.

- For $l(\theta, d(\mathbf{y})) = (\theta - d(\mathbf{y}))^2$, unbiased implies

$$E_{\mathbf{y} \mid \theta}\, d(\mathbf{y}) = \theta.$$

- The *bias* of d is

$$E_{\mathbf{y} \mid \theta} d(\mathbf{y}) - \theta.$$

$E := E_{y|\theta}$

$$E(\theta' - d(y))^2 = E\left(\theta' - Ed(y) + Ed(y) - d(y)\right)^2$$

$$= (\theta' - Ed(y))^2 + E(Ed(y) - d(y))^2$$

$$+ 0$$

minimized at $\theta' = \theta$ only if

$Ed(y) = \theta$

Example: silly unbiased estimator

- Let X be the number of times an event occurs in an hour

$$X \sim \text{Poisson}(\lambda).$$

- Using X, estimate $P(A)$, where $A :=$ no events in two hours:

$$P(A|\lambda) \qquad P(A) = e^{-2\lambda}.$$

- Would like an unbiased estimate for $P(A)$ under squared loss:

$$E\left[d(X)\right] = e^{-2\lambda}.$$

Example: silly unbiased estimator

- d is unbiased only if $d(X) = (-1)^X$
  - A ridiculous estimate!

$$e^{-2\lambda} = E_{X|\lambda} d(X) = \sum_{x=0}^{\infty} d(x) \frac{\lambda^x e^{-\lambda}}{x!}$$

$$\rightarrow e^{-\lambda} = \sum_{x=0}^{\infty} d(x) \frac{\lambda^x}{x!}$$

$$\rightarrow d(x) = (-1)^x \text{ because } e^{-\lambda} = \sum_{x=0}^{\infty} (-1)^x \frac{\lambda^x}{x!}$$

The bias-variance trade-off

- Let $y_1, \ldots, y_n$ be iid with mean μ and variance σ^2
- Consider estimates for μ of the form

$$d_B(\mathbf{y}) = B\mu_0 + (1 - B)\bar{y}$$

where $0 \leq B \leq 1$ is the "shrinkage factor".

- This estimate has variance

$$\mathrm{Var}\left[d(y)\right] = (1-B)^2 \cdot \mathrm{Var}\left(\bar{y}\right) = (1 - B)^2 \frac{\sigma^2}{n}$$

- And it has bias

$$E_{y|\mu} d_B(y) - \mu = B\mu_0 + (1-B)\mu - \mu = B(\mu_0 - \mu)$$

The bias-variance trade-off

- The frequentist risk under squared error loss for d_B is

$$
\begin{aligned}
R(\mu, d_B) &= (1 - B)^2 \frac{\sigma^2}{n} + B^2(\mu_0 - \mu)^2 \\
&= Variance + Bias^2
\end{aligned}
$$

- The "bias-variance trade-off"
- For $B = 0$, $d(\mathbf{y}) = \bar{y}$ is unbiased but maximizes variance
- For $B = 1$, $d(\mathbf{y}) = \mu_0$ has no variance but maximizes bias
- Ideal B depends on
  - σ^2 (larger $\sigma^2 \to B$ closer to 1)
  - n (larger $n \to B$ closer to 0)
  - Distance of μ from μ_0 (larger distance $\to B$ closer to 0)

The bias-variance trade-off

- Given prior $p(\mu)$ for μ with mean μ_0 and variance τ^2.
- The rule d_B that minimizes Bayes risk

$$r(p(\mu), d_B) = \int R(\mu, d_B)p(\mu)d\mu$$

has

$$B = \frac{\sigma^2}{\sigma^2 + n\tau^2}.$$

- Normal-normal model is a special case.
- Homework