Johns Hopkins Engineering

625.464 Computational Statistics

Jackknife Bias Correction

Module 8 Lecture 8D

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Jackknife Review

- samples $\underline{Y} = \{Y_1, \ldots, Y_n\}$ from dist w/ para θ.
- T is an estimator of θ
- we wish to gain insight into the dist of T so we look at the collection of new est. $T_{(-j)}$ formed by deleting k samples from $\underline{Y}$

"pseudovalues"

$$T_j^* = rT - (r-1)T_{(-j)}$$

Jackknifed T

$$J(T) = \frac{1}{r}\sum_{j=1}^{r} T_j^* = \bar{T}^*$$

Jackknife Bias Correction

$K = 1$
$r = n$

$$\text{Bias}(T) = E(T) - \theta = \sum_{q=1}^{\infty} \frac{a_q}{n^q}$$

where a_q do not depend on n

- If all $a_q = 0$, T is unbiased
- If $a_1 \neq 0$, the order of the bias is $1/n$

Bias of the Jackknife J(T)

$$\text{Bias}(J(T)) = E(J(T)) - \theta$$

$$= E\left[nT - (n-1)\bar{T}_{(\cdot)}\right] - \theta$$

$$= nE[T] - (n-1)E\left[\bar{T}_{(\cdot)}\right] - \theta$$

$$= nE[T] - (n-1)E\left[\frac{1}{n}\sum_{j=1}^{n} T_{(-j)}\right] - \theta$$

$$= nE[T] - (n-1)\frac{1}{n}\sum_{j=1}^{n} E\left[T_{(-j)}\right] - \theta$$

$$= n\Big(\underbrace{\boxed{E(T) - \theta}}_{\text{Bias}(T)}\Big) - \frac{n-1}{n}\sum_{j=1}^{n}\Big[\underbrace{\boxed{E(T_{(-j)}) - \theta}}_{\text{Bias}(T_{(-j)})}\Big]$$

Bias of the Jackknife J(T)

$$\text{Bias}(J(T)) = n\left(\sum_{q=1}^{\infty}\frac{a_q}{n^q}\right) - (n-1)\left(\sum_{q=1}^{\infty}\frac{a_q}{(n-1)^q}\right)$$

$$= a_1 + n\left(\sum_{q=2}^{\infty}\frac{a_q}{n^q}\right) - a_1 - \left(\sum_{q=2}^{\infty}\frac{a_q}{(n-1)^q}\right)(n-1)$$

$$= a_2\left(\frac{1}{n}-\frac{1}{(n-1)}\right) + a_3\left(\frac{1}{n^2}-\frac{1}{(n-1)^2}\right) + \cdots$$

$$= a_2\left(\frac{1}{n(n-1)}\right) + a_3\left(\frac{1}{n^2}-\frac{1}{(n-1)^2}\right) + \cdots$$

if $a_q = 0$ $\forall q \geq 2$ $\Rightarrow J(T)$ is unbiased

order of $B(J(T))$ is at most $1/n^2$

The Jackknife Bias Correction

$$\underbrace{E[J(T)] - \theta}_{=0} = E[T] - \theta + (n-1)\left(E(T) - \underbrace{\frac{1}{n}\sum_{j=1}^{n} E(T_{(-j)})}_{\bar{T}_{(\cdot)}}\right)$$

$$\overbrace{\text{Bias}(T)_J} = (n-1)(\bar{T}_{(\cdot)} - T)$$

$$= (n-1)(J(T) - T)$$

$$T_J = nT - (n-1)\bar{T}_{(\cdot)}$$

Higher Order Bias Correction

$$T_J^{*(2)} = n(J(T)) - (n-1)J(T_{(-j)})$$

$J^2(T)$ is unbiased to order $O(n^{-3})$

(1) reducing bias, ↑ variance

(2) even if $g \geq 3$ has $a_g = 0$

$J^2(T)$ may still have bias $O(1/n^3)$