

# Statistics Based on Different Subsamples

Suppose a sample  $X_1, X_2, \dots, X_n$  is to be used to make inferences about a population.

We form a statistic  $T$  that estimates some measure of the population distribution. In order to do much with this estimator, we need to know something about its sampling distribution; e.g., is it biased? what is its variance?, and so on.

If we know the distribution from which a sample is taken, we may have complete information about the distribution of a statistic based on the sample.

We may not be able to get this information if the statistic is very “complicated”, or if we do not have complete information about the distribution from which a sample is taken

# Statistics Based on Different Subsamples: The Parametric Case

Consider the case in which a sample  $X_1, X_2, \dots, X_n$  is to be used to inferences about a population parameter,  $\theta$ .

We form a statistic  $T$  that estimates  $\theta$ .

We want to know its sampling moments (its mean, variance, etc.).

We may be able to develop information about the distribution of a statistic by studying the distribution of the statistic based on subsamples of the original sample.

There are various ways that subsampling can be used to get information about the distribution of a statistic.

## Jackknife Methods

Suppose we delete one observation, say  $X_j$ , from the sample and compute our estimator from the reduced sample.

Let  $T_{(-j)}$  denote the estimator computed from the sample with the  $j^{\text{th}}$  observation removed.

If  $T$  is the mean, for example,

$$T_{(-j)} = \sum_{i \neq j} X_i / (n - 1).$$

For the mean note that

$$X_j = nT - (n - 1)T_{(-j)}.$$

# The Jackknife Variance Estimate

The  $nT - (n - 1)T_{(-j)}$  are called “pseudovalues”, denoted by  $T_j^*$ :

$$T_j^* = nT - (n - 1)T_{(-j)}.$$

The mean of the pseudovalues,  $\bar{T}^*$ , is called the *jackknifed estimator*.

The sample variance of the jackknifed estimator can be used as an estimate of the variance of the estimator  $T$ . (The pseudovalues are *not* independent, of course.)

The sample variance of the jackknife estimator is

$$\frac{1}{n} \frac{\sum (T_j^* - \bar{T}^*)^2}{n - 1}$$

Notice what this is in the case of  $T$  being the sample mean.

## **How good is the Jackknife Variance Estimate?**

Pretty good, but Monte Carlo studies indicate that this quantity often overestimates the variance.

# The Jackknife Bias Correction

We can also use the jackknife to reduce the bias of the original estimator  $T$ .

Suppose  $T$  is biased, and suppose we can represent the bias as a power series in  $n^{-1}$ , i.e.,

$$E(T) - \theta = \sum_{r=1}^{\infty} a_r/n^r,$$

where the  $a_r$  do not involve  $n$ .

Now consider the jackknife estimator, i.e.,

$$\begin{aligned}\bar{T}^* &= nT - (n-1) \sum_{j=1}^n T_{(-j)}/n \\ &= T + (n-1)(T - \sum_{j=1}^n T_{(-j)}/n).\end{aligned}$$

So, therefore,

$$\begin{aligned} \mathbb{E}(\bar{T}^*) - \theta &= \mathbb{E}(T) - \theta + (n-1)(\mathbb{E}(T) - \sum_{j=1}^n \mathbb{E}(T_{(-j)})/n) \\ &= \sum_{r=1}^{\infty} a_r/n^r + (n-1)(\sum_{r=1}^{\infty} a_r/n^r + \theta) - \\ &\quad (n-1)(\sum_{r=1}^{\infty} a_r/(n-1)^r + \theta) \\ &= a_2 \left( \frac{1}{n} - \frac{1}{n-1} \right) + a_3 \left( \frac{1}{n^2} - \frac{1}{(n-1)^2} \right) + \dots, \end{aligned}$$

that is, the bias is only of order  $1/n^2$ .

# Bias Correction

Suppose we have two biased estimators of  $\theta$ ,  $T_1$  and  $T_2$ . Let  $b_1$  and  $b_2$  be the biases, and let

$$R = \frac{b_1}{b_2}$$

Now consider the estimator

$$T^\dagger = \frac{T_1 - RT_2}{1 - R}.$$

We have

$$\begin{aligned} \mathbb{E}(T^\dagger) &= \frac{1}{1 - R} \mathbb{E}(T_1) - \frac{R}{1 - R} \mathbb{E}(T_2) \\ &= \theta. \end{aligned}$$

If  $R$  were only known!

Notice if  $R = (n - 1)/n$ , the jackknife estimator,

$$nT - (n - 1)\bar{T}_{(-j)},$$

is unbiased.

## Higher Order Bias Corrections

Suppose we pursue the bias correction to higher orders, i.e., consider a second application of the jackknife:

$$\bar{T}^{**} = \frac{n^2 T^* - (n-1)^2 \sum_{j=1}^n T_{(-j)}^* / n}{n^2 - (n-1)^2}.$$

This estimator is unbiased to order  $O(n^{-3})$ .

## Higher Order Bias Corrections

There are two major differences between this estimator and the first-order jackknifed estimator.

For the first-order jackknife,  $\bar{T}^*$  differs from  $T$  by a quantity of order  $n^{-1}$ , hence if  $T$  has variance of order  $n^{-1}$  (as we usually hope), the variance of  $\bar{T}^*$  is asymptotically the same as that of  $T$ . That is, the bias reduction carries no penalty in increased variance. This is not the case for higher order bias corrections.

The other difference is that if in the bias expansion,

$$E(T) - \theta = \sum_{r=1}^{\infty} a_r/n^r,$$

if  $a_r = 0$  for  $r = 2, \dots$ , the first-order jackknifed estimator is unbiased.

For the second-order jackknifed estimator, even if  $a_r = 0$  for  $r = 3, \dots$ , the estimator is not unbiased; in fact, its bias is

$$\frac{a_2}{(n-1)(n-2)(2n-1)},$$

i.e., still of order  $n^{-3}$ .

# Another Way of Forming the Second-Order Jackknife

Reference: Schucany, Gray, and Owen, *JASA* (1971).

Looking at

$$E(\bar{T}^*) - \theta = -\frac{a_2}{n(n-1)} - \frac{a_3(2n-1)}{n^2(n-1)^2} - \dots$$

and

$$E(\bar{T}^{**}) - \theta = -\frac{a_2}{(n-1)(n-2)(2n-1)} + O(n^{-3}),$$

and letting

$$R = \frac{\text{bias}(\bar{T}^*)}{\text{bias}(\bar{T}^{**})}$$

suggests taking

$$R = \frac{1}{n(n-1)} / \frac{1}{(n-1)(n-2)}$$

$$= \frac{n-2}{n}.$$

## The Second-Order Jackknives

The Schucany-Gray-Owen jackknife described above will not suffer from the bias for the first second-order jackknifed estimator we described.

So for the second application of the Schucany-Gray-Owen jackknife, we use

$$T^{(2)} = \frac{n}{2}T^* - \frac{n-2}{2} \sum_{j=1}^n T_{(-j)}^*/n$$

Writing  $T_1 = T$  and  $T_2 = \bar{T}_{(-j)}$ , we “jackknife”  $T_1$  by the ratio of the determinants

$$J(T_1) = \left| \begin{array}{cc} T_1 & T_2 \\ 1/n & 1/(n-1) \end{array} \right| / \left| \begin{array}{cc} 1 & 1 \\ 1/n & 1/(n-1) \end{array} \right|$$

# The Generalized Jackknife

So, to generalize, suppose we have two biased estimators,  $T_1$  and  $T_2$ , with

$$E(T_1) - \theta = f_1(n)b(\theta)$$

and

$$E(T_2) - \theta = f_2(n)b(\theta).$$

The generalized jackknife of  $T_1$  is

$$\begin{aligned} J(T_1) &= \begin{vmatrix} T_1 & T_2 \\ f_1(n) & f_2(n) \end{vmatrix} / \begin{vmatrix} 1 & 1 \\ f_1(n) & f_2(n) \end{vmatrix} \\ &= \frac{1}{1-R}T_1 - \frac{R}{1-R}T_2, \end{aligned}$$

where

$$R = \frac{f_1(n)}{f_2(n)}.$$

# The Higher Order Generalized Jackknife

The higher order generalized jackknife estimators can be developed by writing the bias of the  $j^{\text{th}}$  estimator as

$$E(T_j) - \theta = \sum_{i=1}^{\infty} f_{ij}(n)b_i(\theta)$$

for  $j = 1, \dots, k + 1$ . Then

$$J(T_k) = \frac{\begin{vmatrix} T_1 & T_2 & \cdots & T_{k+1} \\ f_{11}(n) & f_{12}(n) & \cdots & T_{1,k+1}(n) \\ \vdots & \vdots & \ddots & \vdots \\ f_{k1}(n) & f_{k2}(n) & \cdots & T_{k,k+1}(n) \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ f_{11}(n) & f_{12}(n) & \cdots & T_{1,k+1}(n) \\ \vdots & \vdots & \ddots & \vdots \\ f_{k1}(n) & f_{k2}(n) & \cdots & T_{k,k+1}(n) \end{vmatrix}}$$

- The generalized jackknife reduces the order of the bias by  $1/n$  in each application.
- If all terms beyond the  $k^{\text{th}}$  in the expansion of the bias are zero, then  $J(T_k)$  is unbiased.

- The variance may increase, however, ...

## The Delete- $d$ Jackknife

For estimators that are not differentiable functions of the sample, such as the sample median, the jackknife is not consistent. It also does not perform well. (Note that “performing well” in real applications may not be very closely related to consistency.)

Similar ideas that lead to the development of the jackknife suggest a delete- $d$  jackknife that is formed by deleting  $d$  observations to form each pseudo-observation. This, of course could lead to a large number of pseudo-observations:  $\binom{n}{d}$ .

The delete- $d$  jackknifed estimator may be consistent. The asymptotics require  $d$  also to get large. (For the median,  $n^{1/2}/d \rightarrow 0$  and  $n - d \rightarrow \infty$ .)

The large number of pseudo observations can be accommodated by random sampling.