

The Jackknife

Patrick Breheny

September 6

Introduction

- The last few lectures, we described influence functions as a tool for assessing the standard error of statistical functionals and obtaining nonparametric confidence intervals
- Today we will discuss a tool called the *jackknife*
- Like influence functions, the jackknife can be used to estimate standard errors in a nonparametric way
- The jackknife can also be used to obtain nonparametric estimates of bias
- Although superficially different, we will see that the jackknife is built on essentially the same idea as the influence function, although the jackknife was proposed much earlier (1949)

Jackknife setup and notation

- Suppose we have an estimator $\hat{\theta}$ which can be computed from a sample $\{x_i\}$
- Jackknife methods revolve around computing the estimates $\{\hat{\theta}_{(i)}\}$, where $\hat{\theta}_{(i)}$ denotes the estimate calculated from the data with the i^{th} observation removed (these are sometimes called the “leave one out” estimates)
- Finally, let

$$\bar{\theta} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(i)}$$

Jackknife estimation of bias

- Note that, if $\hat{\theta}$ is unbiased,

$$\begin{aligned}\mathbb{E}\bar{\theta} &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}\hat{\theta}_{(i)} \\ &= \theta\end{aligned}$$

- However, suppose that $\mathbb{E}(\hat{\theta}) = \theta + an^{-1} + bn^{-2} + O(n^{-3})$
- Then

$$\mathbb{E}(\bar{\theta} - \hat{\theta}) = \frac{a}{n(n-1)} + O(n^{-3})$$

Jackknife estimation of bias

- Thus,

$$b_{jack} = (n - 1)(\bar{\theta} - \hat{\theta}),$$

the *jackknife estimate of bias*, is correct up to second order

- Furthermore, the *bias-corrected jackknife estimate*,

$$\hat{\theta}_{jack} = \hat{\theta} - b_{jack},$$

is an unbiased estimate of θ , again up to second order

Example: Plug-in variance

- For example, consider the plug-in estimate of variance:
 $\hat{\theta} = n^{-1} \sum_i (x_i - \bar{x})^2$
- The expected value of the jackknife estimate of bias is

$$\begin{aligned}\mathbb{E}(b_{jack}) &= -\frac{\theta}{n} \\ &= \text{Bias}(\hat{\theta})\end{aligned}$$

- Furthermore, it can be shown that the bias-corrected estimate is

$$\hat{\theta}_{jack} = s^2,$$

the usual unbiased estimate of the variance

Pseudo-values

- Another way to think about the jackknife is in terms of the *pseudo-values*:

$$\tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{(i)}$$

- Note that

$$\begin{aligned} \frac{1}{n} \sum_i \tilde{\theta}_i &= n\hat{\theta} - (n-1)\bar{\theta} \\ &= \hat{\theta} - b_{jack}, \end{aligned}$$

which is the bias-corrected estimate $\hat{\theta}_{jack}$

- The idea behind pseudo-values is that it allows us to think of the bias-corrected estimate as simply the mean of n “independent” data values

Properties of the pseudo-values

- The pseudo-values $\{\tilde{\theta}_i\}$ are not, in general, independent, although note that for the special case of a linear statistic $\hat{\theta} = n^{-1} \sum_i a(x_i)$, $\tilde{\theta}_i = a(x_i)$ — *i.e.*, for the mean, $\tilde{\theta}_i = x_i$
- A reasonable idea, therefore, is to treat $\tilde{\theta}_i$ as linear approximations to iid observations and approach inference for $\hat{\theta}_{jack}$ as we would the sample mean
- Thus, letting \tilde{s}^2 denote the sample variance of the pseudo-values, consider

$$v_{jack} = \frac{\tilde{s}^2}{n}$$

as an estimate of $\mathbb{V}(\hat{\theta})$

Example: Mean

- As a somewhat trivial example, suppose $\hat{\theta} = \bar{x}$
- $b_{jack} = 0$
- $v_{jack} = s^2/n$, where s^2 is the usual, unbiased estimate of the variance

Jackknife confidence intervals?

- The pseudo-value statement of the jackknife suggests the following approach to constructing confidence intervals:

$$\hat{\theta}_{jack} \pm t_{1-\alpha/2; n-1} SE_{jack},$$

- These intervals turn out to be similar to those produced by the functional delta method, for reasons that will be discussed soon

Homework

Homework: Write an R function called `jackknife` which implements the jackknife. The function should accept two arguments: `x` (the data) and `theta` (a function which, when applied to `x`, produces the estimate). The function should return a named list with the following components:

- `bias` – the jackknife estimate of bias
- `se` – the jackknife estimate of standard error
- `values` the leave-one-out estimates $\{\hat{\theta}_{(i)}\}$

Please submit the function via Dropbox and name your file `Brehehy-jack.R`, with your last name replacing Brehehy

Is v_{jack} a good estimator?

- Obviously, v_{jack} is a good estimator of $\mathbb{V}(\bar{x})$, where it is equivalent to the usual estimator
- The same logic extends to any linear statistic
- Furthermore, if g is function continuously differentiable at μ , it can be shown that v_{jack} is a consistent estimator of $g(\bar{x})$:

$$\frac{v_{jack}}{\mathbb{V}\{g(\bar{x})\}} \xrightarrow{P} 1$$

Is v_{jack} a good estimator? (cont'd)

- However, there are also cases where v_{jack} is not a good estimator of the variance of an estimate
- In particular, v_{jack} can be shown to perform poorly when the estimator is not a smooth function of the data
- A simple example of a non-smooth estimator is the median

Jackknife applied to the median

- Suppose we observe the sample $\{1, 2, \dots, 9, 10\}$
- What will our leave-one-out estimates look like?
- We will obtain five 5's and 5 6's
- There is nothing special about these particular numbers: for any data set with an even number of observations, we will always obtain two unique values for $\hat{\theta}_{(i)}$, each with $n/2$ instances

Inconsistency

- This doesn't seem like a good way to estimate the variance of the median, and it isn't
- It can be shown that the jackknife variance estimate is inconsistent for all F and all p
- In particular, for the median,

$$\frac{v_{jack}}{\mathbb{V}(\hat{\theta})} \xrightarrow{d} \left(\frac{\chi_2^2}{2} \right)^2$$

Jackknife and influence function

- There is a close connection between the jackknife and influence functions
- Viewing the jackknife as a plug-in estimator, it calculates n estimates based on a slightly altered version of the empirical distribution function and compares these altered estimates to the plug-in estimate in order to assess variability of the estimate
- What CDF is the jackknife using?

$$\left(\frac{1}{n-1}, \dots, 0, \dots, \frac{1}{n-1} \right)$$

Jackknife and influence function (cont'd)

- This is very similar to the idea behind the influence function, with

$$1 - \epsilon = \frac{n}{n-1}$$
$$\implies \epsilon = -\frac{1}{n-1}$$

- In a sense, then, the jackknife is a numerical approximation to the functional delta method
- Indeed, an alternative name for the functional delta method is the *infinitesimal jackknife*

The positive jackknife

- There is an important difference, however, between the jackknife and the functional delta method: the delta method adds point mass to observation x_i , while the jackknife takes mass away
- Another take on the jackknife, then, is to compute n estimates $\{\hat{\theta}_{(i)}\}$ by adding an observation at x_i instead of taking one away (i.e., $\epsilon = 1/(n + 1)$)
- This method is called the *positive jackknife*; however, it is not commonly used

The delete- d jackknife

- Another variation on the jackknife that has been proposed is called the *delete- d jackknife*
- As the name suggests, instead of leaving out one observation when calculating the collection $\{\hat{\theta}_{(s)}\}$, the delete- d jackknife leaves out d observations
- This approach has merit: in particular, it can be shown that if d is appropriately chosen, then the delete- d jackknife estimate of variance is consistent for the median
- However, it has the drawback that instead of calculating n leave-one-out estimates, we now have to calculate $\binom{n}{d}$ leave- d -out estimates — a much larger number, often bordering on the computationally infeasible

Homework

Homework: What would the collection of weights $\{w_i\}_{i=1}^n$ look like for the delete- d jackknife?

Homework

Homework: The standardized test used by law schools is called the Law School Admission Test (LSAT), and it has a reasonably high correlation with undergraduate GPA. The course website contains data on the average LSAT score and average undergraduate GPA for the 1973 incoming class of 15 law schools.

- (a) Use the jackknife to obtain an estimate of the bias and standard error of the correlation coefficient between GPA and LSAT scores. Comment on whether the estimate is biased upward or downward.
- (b) If x and y are drawn from a bivariate normal distribution, then $n\mathbb{V}(\hat{\rho}) \xrightarrow{P} (1 - \rho^2)^2$. Use this to estimate the standard error of $\hat{\rho}$.

Homework (cont'd)

- (c) On page 21 of our textbook, the author gives the influence function for the correlation coefficient:

$$L(x, y) = \tilde{x}\tilde{y} - \frac{1}{2}\theta(\tilde{x}^2 + \tilde{y}^2),$$

where

$$\tilde{x} = \frac{x - \mu_x}{\sqrt{\sigma_x^2}}$$

and \tilde{y} is defined similarly. Use this to estimate the standard error of $\hat{\rho}$; compare the three estimates (a)-(c).

Homework (cont'd)

- (d) For each data point (x_i, y_i) , make a plot of $\hat{\rho}_\epsilon$ vs. the mass at point i , as the point mass at i varies from 0 to $1/3$ (and the rest of the mass is spread evenly on the rest of the observations).
- (e) Comment on why some plots slope upwards and others slope downwards.
- (f) Extra credit: Comment on the how the shape of the curve relates to the comparison between the delta method and jackknife estimates of the variance

Hints

- The R function `cov.wt` can be used to calculate weighted correlation coefficients, taking on an optional argument that consists of a vector of weights for each observation
- Please do not turn in 15 pages of plots – use `mfrow` to make a grid of plots on one page
- You will likely have to reduce the margins of your plots to eliminate whitespace; you are free to design your plot how you wish, but I will pass on the suggestion `par(mfrow=c(5,3), mar=c(2,4,2,1))`