Dr. David Draper
Statistics Group, School of Mathematical Sciences
University of Bath

X3: Some Notes on the Jackknife

The **jackknife** was a forerunner of the bootstrap, developed by Maurice Quenouille in 1949 to estimate the **bias** of an estimator $\hat{\theta}$ of a population parameter θ, and extended (and given its name) by John Tukey in 1958 to estimate the **standard error** of $\hat{\theta}$.

Bias Estimation

The idea behind the jackknife: Resampling from $x = (x_1, \ldots, x_n)$ in a particularly simple way—successively deleting one point at a time—can give you some idea of how stable $\hat{\theta}$ is, which in turn should provide some information about how close $\hat{\theta}$ is to θ.

Define $x_{(i)} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$, the *ith jackknife sample*. Imagine recalculating $\hat{\theta}$ based on $x_{(i)}$ for each i, rather than basing it on the whole data vector x—after all, $\hat{\theta}$ is just some function $s(x)$, so define $\hat{\theta}_{(i)} = s(x_{(i)})$, i.e., what you'd get by applying that function to the jackknife samples—and define $\hat{\theta}_{(\cdot)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{(i)}$.

Quenouille's idea was that $(\hat{\theta}_{(\cdot)} - \hat{\theta})$ should be relevant to bias assessment. He proposed the **jackknife estimate of bias**,

$$\widehat{\text{bias}}_{\text{jack}} = (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta}), \tag{1}$$

and used this to suggest a **bias-corrected** version of the estimator $\hat{\theta}$,

$$\bar{\theta} = \hat{\theta} - \widehat{\text{bias}}_{\text{jack}}. \tag{2}$$

The main mystery in this is how Quenouille came up with the multiplier $(n-1)$ in front of $(\hat{\theta}_{(\cdot)} - \hat{\theta})$. One way to figure this out is to see what the right multiplier is in a simple problem where you know the right answer—i.e., find a biased estimator whose bias adjustment you know, apply the jackknife to it and see what multiplier is needed to make $\bar{\theta}$ unbiased.

Example. Take $(X_1, \ldots, X_n) \stackrel{\text{IID}}{\sim} F$ and consider $\theta = V_F(X) = E(X - \mu)^2 = \int (x-\mu)^2 dF(x)$, where $\mu = E_F(X) = \int x \, dF(x)$. As summaries of the population cdf F, μ and θ may be thought of literally as functions of F, e.g., $\mu = t(F)$. When viewed in this way, finding a good estimate of μ amounts to finding a good estimate of F and substituting it in: $\hat{\mu} = t(\hat{F})$, which Efron and Tibshirani (1993, hereafter ET) call the **plug-in estimate** of μ. Without any other information about F the best you can do in estimating $F(x) = P(X \le x)$ is to use the **empirical cdf** $\hat{F}(x) = (\#\text{ of } X_i \le x)/n$, which (e.g.) in the case of the population mean leads to the estimate $\hat{\mu} = t(\hat{F}) = \int x d\hat{F}(x) =$ our old friend the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ because $\hat{F}$ is a step function and (Riemann-Stieltjes) integration with respect to the derivative of a step function turns out to correspond to summation of the integrand at the jump points. So by the same token a good estimate of $\theta = V_F(X)$ should be the plug-in estimate $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$ (for instance you will recognize this as the maximum likelihood estimate of θ in the Gaussian case with both μ and θ unknown). However you probably also remember that $\hat{\theta}$ is biased on the low side in this case: $E(\hat{\theta}) = \frac{n-1}{n}\theta$, so that $\text{bias}(\hat{\theta}) = -\frac{\theta}{n}$. You can show (and indeed I have encouraged you to show in the extra credit to problem 2 in Assignment 1) that the right multiplier of $(\hat{\theta}_{(\cdot)} - \hat{\theta})$ in equation (1) above to make the bias-corrected jackknife estimate of variance unbiased is $(n-1)$, as Quenouille suggested.

Standard Error Estimation

Like Quenouille, Tukey was interested in the values $\hat{\theta}_{(i)}$ of the estimator $\hat{\theta}$ applied to the jackknife samples $x_{(i)}$ formed by deleting points $i = 1, \ldots, n$ from the data set, but the use to which Tukey put the $\hat{\theta}_{(i)}$ was different: he thought that the variability of the $\hat{\theta}_{(i)}$ around their average $\hat{\theta}_{(\cdot)}$ should provide standard error information for $\hat{\theta}$. Specifically, he proposed the **jackknife estimate of standard error**

$$\widehat{\text{SE}}_{\text{jack}} = \sqrt{\frac{n-1}{n}\sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2}. \tag{3}$$

Once again the main mystery is where Tukey got the multiplier, and the solution is the same as it was with Quenouille: Tukey picked a simple problem where he knew the right answer and chose the multiplier to match. As I have asked you to examine in problem 2 of Assignment 1, instead of Quenouille's variance parameter Tukey used something even simpler, the mean: he showed that equation (3) applied to the sample mean, $\hat{\theta} = \bar{x}$, produces the correct standard error $\sqrt{\frac{1}{n(n-1)}\sum_{i=1}^n (X_i - \bar{X})^2}$.

Tukey's use of the jackknife is typically more valuable in practice than Quenouille's, because standard error assessment is usually both easier and more important than bias assessment and adjustment (see ET chapter 10).

Warning About the Jackknife

As ET note in their Chapters 10 and 11, the jackknife works only on plug-in estimates $t(\hat{F})$ in which the function $t(\cdot)$ is *smooth* (e.g., twice differentiable). This includes many standard statistics like the mean and variance but also excludes a number of standard quantities such as the median. The jackknife can be fixed for non-smooth statistics by deleting more than one point at a time in constructing the jackknife samples, but by the time you have gone that far you might as well use the bootstrap.

The Relationship Between the Jackknife and the Bootstrap

The jackknife is like a bootstrap in which sampling is done without replacement instead of with, and the samples are of size $(n-1)$ instead of n. Since there are only n such possible samples, there is no point in doing the resampling at random, which would just introduce a new layer of Monte Carlo uncertainty into the answer. People who like the bootstrap (such as ET) tend to think of the jackknife as a computationally less intensive way to approximate the answers you would get from bootstrapping (and you can show, as ET do in Section 11.5, that *when the jackknife works* [and it does not always do so, as noted above] its estimates of bias and standard error, like those of the bootstrap, are relatively unbiased but have more sampling variability than those of the bootstrap). Probably the jackknife is an idea whose time has mostly come and gone, but it has been useful in suggesting methods such as the bootstrap and **cross-validation**, which (as we will later see) is a kind of predictive jackknife.