Breakdown Point Theory Notes

Charles J. Geyer

February 2, 2006

1 Introduction

These are class notes for Stat 5601 (nonparametrics) taught at the University of Minnesota, Spring 2006. This not a theory course, so the bit of theory we do here is very simple, but very important.

Without *some* notion of robustness we have no way to say why nonparametrics is a good idea, or, at least, no *quantitative* way. We could say (and sometimes do say) that the assumptions for the sign test are weaker than the assumptions for the signed rank test, which in turn are weaker than the assumptions for Student's t test.

But what does that say? We know that no data are *exactly* normal, so either

> no one should ever use Student's t test

or

> one should use Student's t test so long as the population distribution isn't *too* non-normal

but what does the latter statement mean? What is *too* non-normal when we have no way to quantitate non-normality?

2 Breakdown Point

It turns out that quantitating non-normality is the wrong idea. There are just too many ways to be non-normal. A much better idea is to quantitate properties of estimators and their associated procedures. Here we look at one such idea.

The *finite sample breakdown point* of an estimator is the fraction of data that can be given arbitrary values without making the estimator arbitrarily bad. Typically this is some function of the sample size n. To get a single number, we use the *asymptotic breakdown point* which is the limit of the finite sample breakdown point as n goes to infinity.

The distinction between finite sample and asymptotic breakdown points is only worth bothering about during calculations. After we are done with these notes, we will always just say "breakdown point" to mean asymptotic breakdown point.

2.1 The Sample Mean

It is obvious from the formula from the mean

$$\frac{x_1 + \ldots + x_n}{n}$$

that if we hold $x_1, \ldots, x_{n-1}$ fixed and let x_n go to infinity, the sample mean also goes to infinity. In short even one gross outlier ruins the sample mean. The finite sample breakdown point is $1/n$. The asymptotic breakdown point is zero.

2.2 The Sample Median

If we have n data points and we let a minority of them $\lfloor (n-1)/2 \rfloor$ go to infinity leaving the rest fixed, where $\lfloor \cdot \rfloor$ denotes the "floor" operation (largest integer less than or equal to), then the median stays with the majority. The median changes, but does not become arbitrarily bad. The finite sample breakdown point is $\lfloor (n-1)/(2n) \rfloor$. The asymptotic breakdown point is one-half.

2.3 The Sample Pseudomedian

The estimator under discussion here is the median of the Walsh averages, which is the Hodges-Lehmann estimator associated with the Wilcoxon signed rank test. The Walsh averages are the $n(n+1)/2$ numbers of the form

$$\frac{x_i + x_j}{2}, \qquad i \le j$$

where $x_1, \ldots, x_n$ are the data.

If we let x_k go to infinity, leaving the rest fixed, we ruin n Walsh averages, all those involving x_k. If we let $n - m$ data points go to infinity leaving m fixed, then we *don't* ruin the Walsh averages involving only the m fixed data points, and there are $m(m+1)/2$ of them.

The median of the Walsh averages, stays with the majority, so we do not have finite sample breakdown so long as

$$\frac{m(m+1)}{2} > \frac{n(n+1) - m(m+1)}{2}$$

or

$$m(m+1) > \frac{n(n+1)}{2}$$

At this point we see that obtaining the finite sample breakdown point will be really messy. If we go directly to asymptotic breakdown point, everything is much simpler. Writing $x = m/n$ we also have (for large n) $x \approx (m+1)/(n+1)$, so we need to solve

$$x^2 = \frac{1}{2}$$

giving

$$x = \frac{1}{\sqrt{2}}$$

We're almost done. Our x is not the breakdown point but the fraction of good data we need to avoid breakdown. So $1 - x$ is the (asymptotic) breakdown point, and this is

$$1 - \frac{1}{\sqrt{2}} \approx 0.29289$$

3 Conclusions

3.1 Summary

estimator	breakdown point
sample mean	0
sample pseudomedian	0.293
sample median	0.5

3.2 Analysis

From the breakdown point characterization of robustness, the sample mean is the worst estimator ever invented. It is suitable only for perfect data. No outliers at all. Both estimators associated with nonparametric tests are good. The median is the better of the two.

It is not clear how much robustness you need. Do you really expect half your data to be junk? Or, more precisely, do you really need to protect against the possibility that half your data are junk? Perhaps it is enough to protect only against the possibility of up to 29% of your data being junk?

On the other hand, is it not just silly to have no protection whatsoever against problematic data?

This breakdown point notion of robustness must be very different from whatever people have in mind when they say (as many people and textbooks do) that Student's t test and related procedures (including the sample mean as point estimate) are "robust" against departures from normality. I have never found any mathematical arguments backing up such statements and have come to the conclusion that they are essentially tautological.

We know that any outliers at all wreck the sample mean and related procedures. We don't prove here, but it is a fact, that skewness also wrecks them. So does multimodality. Hence we conclude that the procedures that assume the normal distribution are robust against departures from normality that

1. keep the light tails of the normal distribution (no outliers),

2. keep the symmetry of the normal distribution, and

3. keep the unimodality of the normal distribution.

In short

> *So long as the "departures" from normality are so slight that a human can't tell the difference between the population distribution and a normal distribution, then Student's t and related procedures work well.*

Or, tautologically,

> *If departures from normality don't hurt Student's t, then they don't hurt Student's t.*

The tautology being that "if *blah*, then *blah*" is true whatever *blah* may be.

4 Addendum on Asymptotics

4.1 Student T

There is one notion of robustness that Student's t and related procedures do have. Perhaps we should mention it to be fair to their fans.

We do know that for large sample sizes n there is almost no difference between so-called z and t procedures. For large n, Student t critical values are almost the same as normal z critical values. That's why the bottom row of a t table (for infinite degrees of freedom) are the z values. Mathematically, we say t distributions converge to normal distributions as n goes to infinity.

We also know that the condition for the central limit theorem is finite variance, which is one notion of "light tails" but a much weaker condition than normality. The standard normal density function

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

has very very light tails. But in order to have finite variance, it is enough that a density f satisfy

$$f(x) \le \frac{C}{x^{3+\epsilon}}, \qquad |x| \ge M$$

for some constants C, M, and $\epsilon > 0$. Tails going down like $1/x^{3+\epsilon}$ are very much heavier than tails going down like $\exp(-x^2/2)$.

So *finite variance* is a much larger family than *normal*. In fact *finite variance* is a nonparametric model. The conclusion that we reach from all of this is that the conventional z (large sample normal theory) procedures are *asymptotically robust* in that they work for *sufficiently large* n for any population distribution in this large nonparametric class.

Moreover t has the same properties as z. The Student t procedures are asymptotically equivalent to the z procedures. So they are *asymptotically robust* too.

But that is not the aspect of Student t that interests us. When we are thinking of Student t as a competitor of the sign test and the signed rank test, we are think of it and them as *exact small sample procedures.* And whatever *large sample* robustness Student t may have is irrelevant to that issue.

4.2 Two-Sample T and Variance Estimation

The reader may well be wondering why we made such a big deal of the asymptotic robustness of the Student t family of procedures when it pales in comparison to the *exact small-sample robustness* of the sign test and the signed rank test and related procedures (not to mention other rank-based procedures that we haven't gotten to yet but also have similar small-sample robustness).

The reason is that asymptotic (large-sample) robustness is not to be sneezed at. There are procedures, much studied and taught in theory of statistics classes that don't even have that. The most well known are the following.

- The two-sample (non-paired) so-called "exact" Student t procedures that assume *equality of variances* of the two populations and consequently use a "pooled" estimator of the common variance. These procedures are non-robust to any violation of the equality of variance assumption. If the population variances are not exactly equal, they give incorrect inference no matter how large the sample sizes are.

  Their z competitors that do not assume equality of variances are better for sufficiently large sample sizes (they are asymptotically robust).

- The one-sample tests and confidence intervals for the population variance based on the chi-squared distribution and the analogous two-sample procedures based on the F distribution assume that the ratio of population second and forth moments (the so-called *kurtosis*) is exactly that of the normal distribution, that is,

  $$E\{(X-\mu)^4\} = 3E\{(X-\mu)^2\} \tag{1}$$

  where μ is the mean. If (1) does not hold *exactly*, then the inference is incorrect no matter how large the sample size may be. There is no way to check (1) in any application since fourth moments are *very* sensitive to tail behavior.

4.3 Conclusions for this Addendum

Really good procedures are robust (exact small-sample robust). They are not wrecked by bad data. The sign test and the signed rank test and their related procedures fall in this category.

The next best thing is asymptotically robust. Such procedures at least work well for sufficiently large sample sizes. The one-sample t procedures and the two-sample t procedure that does not assume equality of variances and is not exact either (using Welch's approximation for the distribution of the test statistic) fall in this category.

The worst thing is having no robustness whatsoever. The other procedures we mentioned, the two-sample "exact" t procedures that assume equality of variance, and the one- and two-sample variance estimation procedures based on chi-square and F, fall in this category. There are other procedures in this category but we shall be merciful and not mention them.