A Simple Exchangeability Argument

Charles J. Geyer

September 24, 2006

1 Introduction

These are class notes for Stat 5601 (nonparametrics) taught at the University of Minnesota, Fall 2006. This document gives a simple argument that may justify the symmetry assumption required for the Wilcoxon signed rank test and its associated confidence interval (Sections 3.1, 3.2, 3.3, and 3.7 in Hollander and Wolfe).

The required assumptions are that the one-sample data $Z_1, \ldots, Z_n$ are

- IID (independent and identically distributed)
- from a distribution that is
  - continuous
  - and symmetric.

  (page 79 in Hollander and Wolfe).

These assumptions are the same as the assumptions required for the sign test and its associated confidence interval except for symmetry. What justifies that?

In one-sample problems, there is no justification that comes from probability and statistics. Either you accept this assumption for scientific reasons or you don't (substitute business, practical or whatever for scientific). In two-sample problems, there is the following additional exchangeability argument.

Suppose the original data are paired (X_i, Y_i), $i = 1, \ldots, n$ and we have defined $Z_i = Y_i - X_i$. We assume the (X_i, Y_i) pairs are IID, which implies the Z_i are IID. We assume that the joint distribution of the (X_i, Y_i) pairs is continuous, which implies the distribution of the Z_i is continuous. What do we need to assume about the (X_i, Y_i) pairs to justify symmetry of the distribution of the Z_i? Since this issue is only about the distribution of a random vector (X, Y) and another random variable $Z = Y - X$ defined from it, we can drop the subscripts.

A random vector is said to have *exchangeable* components if permuting the components does not change its distribution. For a two-vector, (X, Y), this means that (X, Y) and (Y, X) have the same distribution. Note that this not only means that X and Y have the same marginal distribution but also that the conditional distribution of X given Y is the same as the conditional distribution of Y given X. This implies that $Z = Y - X$ and $-Z = X - Y$ have the same distribution, which says the distribution of Z is symmetric about zero.

More generally, if the pair $(X + \theta, Y)$ is exchangeable, then

$$Z - \theta = Y - (X + \theta)$$

has the same distribution as

$$\theta - Z = (X + \theta) - Y,$$

which says the distribution of Z is symmetric about θ. So this is the exchangeability assumption that justifies the Wilcoxon signed rank test and related procedures. In full, the assumptions are The random vectors (X_i, Y_i), $i = 1, \ldots, n$ are

- IID (independent and identically distributed)
- from a distribution that is
  - continuous
  - and $X_i + \theta$ and Y_i are exchangeable.

Whether one finds this exchangeability assumption any more scientifically acceptable than just assuming symmetry of the distribution of Z_i to begin with is a scientific question, not a statistical one. All probability and statistics can do is provide this additional argument, for whatever it is worth.