Saddlepoint Approximations of the Two-sample Wilcoxon Statistic

BY RUNGAO JIN AND JOHN ROBINSON

University of Sydney

Froda and van Eeden [1] obtain an approximation for the two-sample Wilcoxon statistic based on the moment generating function due to van Dantzig [8]. A direct saddlepoint approximation based on this moment generating function is obtained and is shown to have uniform relative error. These approximations are compared numerically to those based on the conditional saddlepoint method and the method of [1].

1. Introduction Froda and van Eeden [1] obtain an approximation for the two-sample Wilcoxon statistic, W, the sum of the ranks of the first sample, for samples of size m and n with $m + n = N$, based on an exact moment generating function of van Dantzig [8]. This gives relative errors of order $N^{-3/2}$ for the approximation of the tail probabilities, $P((W - EW)/\sqrt{VarW} \geq w)$, in the compact region, when w is bounded, but it does not give relative errors for a region including the large deviation region, when $w = O(\sqrt{N})$.

We consider approximations for the distribution of the two-sample Wilcoxon statistic by a number of different methods and obtain numerical comparisons of them. The first method is based on the moment generating function used by Froda and van Eeden [1]. We have used the usual saddlepoint method to obtain approximations to probabilities and tail probabilities. We have obtained two distribution approximations, the indirect Edgeworth approximation and the Barndorff-Nielsen approximation, asymptotically equivalent to that of Lugananni-Rice with relative error of order N^{-1}.

The second method is based on a conditional representation of the distribution of the Wilcoxon statistic which has been given in [5] and generalised in [6]. Also the method of [7] may be applied to give a Lugananni-Rice version of the conditional method. We give indirect Edgeworth and Barndorff-Nielsen approximations based on the conditional technique. Froda and van Eeden [1] also suggest that their method should greatly improve on the conditional method, but this is shown here not to be so, particularly in the large deviation region.

Finally, we give numerical examples for $n = m = 5$ and $m = 10$, $n = 6$ comparing approximations to probabilities and to tail probabilities from the saddlepoint approximations using indirect Edgeworth and Barndorff-Nielsen forms, and based on the conditional method and the approximation of [1]. These indicate that there is effectively no difference between the new saddlepoint approximations and the approximations based on the conditional method. The approximation of [1] is very good in the center of the distribution but quite poor in the tails.

2. Direct Saddlepoint Approximation Let W be the two-sample Wilcoxon rank sum statistic. Put $U = W - m(m+1)/2$ and $N = m+n$. Then, from Question 15 of page 126 of [2], we have