A New Way to Derive Locally Most Powerful Rank Tests

Anthony Y. C. Kuk

The standard way to derive locally most powerful rank tests in nonparametrics involves differentiation of the rank likelihood, which is a complicated integral, and the proof is typically quite long and technical. By treating the ranks as the observed data and the underlying continuous random variables as the "complete" data, we are able to give a quick derivation of locally most powerful rank tests by using a well-known relationship between the derivatives of the log-likelihoods of the observed and "complete" data, which can be either assumed known or derived from first principles in a few lines. Along with shortening the proof considerably, this proposed approach has two other advantages. First, it allows students to see that the locally most powerful rank test is just a special case of the locally most powerful test based on the efficient score (i.e., the derivative of the log-likelihood), a topic often covered in a first-year graduate course in mathematical statistics. Second, it allows students to gain the perspective of viewing a rank test as an case of inference based on incomplete data. As a novel application, the proposed approach is used to derive a new rank test based on the difference between the Savage score and the Wilcoxon score. The new test is useful in testing for a difference in two survival distributions with a time-varying hazard ratio.

KEY WORDS: Complete data; Hazard ratio; Locally most powerful test; Savage test; Score function; Wilcoxon test.

1. INTRODUCTION

To be concrete, we illustrate our point using the two-sample location problem and mention in passing that essentially the same approach can be used for the two-sample scale problem or in a regression setting. Let $X_1, \ldots, X_m$ be independent and identically distributed (iid) according to $f(x-\theta)$ and let $X_{m+1}, \ldots, X_{m+n}$ be iid according to $f(x)$, so that $N = m+n$ is the combined sample size. The two samples are assumed to be independent. A standard result in the theory of rank tests is that the locally most powerful rank test of $H_0 : \theta = 0$ against $H_1 : \theta > 0$ is based on the test statistic

$$S = -\sum_{i=1}^{m} E_{H_0}\left\{\frac{f'(X_{(R_i)})}{f(X_{(R_i)})}\right\}, \tag{1}$$

where $R_1, \ldots, R_N$ denote the ranks of $X_1, \ldots, X_N$ and $X_{(1)} < X_{(2)} < \cdots < X_{(N)}$ are the order statistics (Hajek, Sidak, and Sen 1999, p. 68). Let $r = (r_1, \ldots, r_N)$ be the realization of the rank vector $R = (R_1, \ldots, R_N)$. The standard way to derive (1) is to begin with the rank likelihood

$$\Pr(R = r) = \int \underset{x_{d_1} < \cdots < x_{d_N}}{\cdots} \int \prod_{i=1}^{m} f(x_i - \theta) \times \prod_{i=m+1}^{N} f(x_i)\, dx_1 \cdots dx_N, \tag{2}$$

where $d_1, \ldots, d_N$ denote the antiranks. This is followed by differentiating $\Pr(R = r)$ with respect to θ at $\theta = 0$. Randles and Wolfe (1979, pp. 296–298) devoted three pages to this proof, whereas Hajek, Sidak, and Sen (1999, pp. 70–73) used four pages and cited three theorems proven earlier in their book to provide a rigorous and general treatment of locally most powerful rank tests.

2. A NEW PROOF

A well-known result in mathematical statistics (Cox and Hinkley 1974, p. 113) is that if the data $X_1, \ldots, X_N$ follow a parametric family of distributions, $f(x_1, \ldots, x_N; \theta)$, indexed by a scalar parameter θ, then the locally most powerful test for testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$ is based on the efficient score, defined as the derivative of the log-likelihood function $\ell(\theta; x_1, \ldots, x_N) = \log f(x_1, \ldots, x_N; \theta)$ with respect to θ evaluated at $\theta = 0$. By the same token, the locally most powerful rank test is based on the derivative of the logarithm of the rank likelihood. Note that

$$\left.\frac{\partial \Pr(R = r)}{\partial \theta}\right|_{\theta=0} = \Pr_{\theta=0}(R = r) \left.\frac{\partial \log \Pr(R=r)}{\partial \theta}\right|_{\theta=0} = \frac{1}{N!} \left.\frac{\partial \log \Pr(R = r)}{\partial \theta}\right|_{\theta=0},$$

and so it does not matter whether the rank likelihood (2) or its logarithm is differentiated. We prefer to work with the log-likelihood, to remain in line with the standard treatment of locally most powerful tests and to take advantage of a known general result, which we elaborate later. Rather than differentiating the logarithm of (2) directly, we apply the following result. Let Y denote the observed data and $X = (Y, U)$ denote the "complete" data obtained by augmenting some missing or unobserved data U to the observed data. The log-likelihood of the observed data, $\ell(\theta; y) = \log f(y; \theta)$, and the log-likelihood of the "complete" data, $\ell(\theta; x) = \log f(x; \theta) = \log f(y, u; \theta)$, are related by

$$\ell'(\theta; y) = E\{\ell'(\theta; X) | Y = y; \theta\}. \tag{3}$$

Anthony Y. C. Kuk is Professor, Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546 (E-mail: *stakuka@nus.edu.sg*). The author thanks the editor, an associate editor, and two referees for their helpful comments, and Alex Cook for his help in polishing the language.

  DOI: 10.1198/tast.2009.08244

We can either take (3) as a known result (McLachlan and Krishnan 1997, p. 100) or quickly derive it from first principles as

$$\begin{aligned}\ell'(\theta; y) &= \frac{\partial}{\partial\theta}\log f(y;\theta) = \frac{\partial}{\partial\theta}\log\int f(y,u;\theta)\,du \\ &= \frac{\int f'(y,u;\theta)\,du}{f(y;\theta)} = \int \frac{f'(y,u;\theta)}{f(y,u;\theta)}\frac{f(y,u;\theta)}{f(y;\theta)}\,du \\ &= \int \ell'(\theta; y,u) f(u|y;\theta)\,du = E\{\ell'(\theta; X)|Y=y;\theta\}.\end{aligned}$$

Note that here we have interchanged the order of integration and differentiation. According to Wiener (2001), a sufficient condition for doing this is the continuity of $f(y,u;\theta)$ and $f'(y,u;\theta)$ as functions of u and θ.

In the context of the two-sample location problem, the complete data are, $X = (X_1,\ldots,X_N)$, and the observed data Y are in the form of the rank vector $R = (R_1,\ldots,R_N)$. The log-likelihood of the "complete" data is given by $\ell(\theta; x) = \sum_{i=1}^{m}\log f(x_i-\theta) + \sum_{i=m+1}^{N}\log f(x_i)$, so that

$$\ell'(\theta; x) = -\sum_{i=1}^{m}\frac{f'(x_i-\theta)}{f(x_i-\theta)}.$$

Substituting into (3), we obtain

$$\begin{aligned}\left.\frac{\partial\log\Pr(R=r)}{\partial\theta}\right|_{\theta=0} &= -\sum_{i=1}^{m}E_{H_0}\left\{\frac{f'(X_i)}{f(X_i)}\middle|R=r\right\} \\ &= -\sum_{i=1}^{m}E_{H_0}\left\{\frac{f'(X_{(r_i)})}{f(X_{(r_i)})}\middle|R=r\right\} \\ &= -\sum_{i=1}^{m}E_{H_0}\left\{\frac{f'(X_{(r_i)})}{f(X_{(r_i)})}\right\},\end{aligned}$$

where the last equality follows from the well-known result that the order statistics and the ranks are independent under the null hypothesis (Hajek, Sidak, and Sen 1999, p. 37). This provides more proof that (1) is the locally most powerful rank test of $H_0:\theta=0$ against $H_1:\theta>0$.

As alluded to earlier, here we use the two-sample location problem as a concrete example to illustrate how the new proof works. We can use the same technique to provide an easier proof of the most general theorem 3.4.8.1 of Hajek, Sidak, and Sen (1999, p. 73). The steps are similar to those specified earlier and are omitted to save space.

3. A TEST BASED ON DIFFERENCE OF SAVAGE AND WILCOXON SCORES

Most of the rank tests proposed in the literature are designed with location shifts or scale shifts in mind. To do something different, we use the proposed approach to derive a new rank test that is locally most powerful against an alternative hypothesis specifying that the hazard ratio varies with time in a certain way. To motivate the problem, we begin with the Savage test, which is known to be the locally most powerful rank test of $H_0:\theta=1$ against $H_1:\theta<1$ when $T_1,\ldots,T_m$ are iid $\exp(\theta)$ and $T_{m+1},\ldots,T_{m+n}$ are standard exponential (Hajek, Sidak, and Sen 1999, p. 106). This exponential model postulates that the hazard ratio between the treatment sample 1 and the control sample 2 is given by the constant θ. Although the assumption of a constant hazard ratio, as implied by the popular proportional hazards model, has dominated the survival analysis literature, people are beginning to realize that in some applications it may be unreasonable to expect the hazard ratio to remain constant over time. To address this problem, Yang and Prentice (2005) have proposed a model that allows the hazard ratio to vary over time. The following model is a special case of their model. As before, we assume that $T_{m+1},\ldots,T_{m+n}$ are standard exponential but that the data in the first sample, $T_1,\ldots,T_m$, are iid with hazard function

$$h(t) = \frac{\theta}{1+(\theta-1)e^{-t}}. \tag{4}$$

Note that $h(t)$ is also the hazard ratio, because sample 2 is standard exponential, with constant hazard 1. Because $h(t)\to 1$ as $t\to 0$ and $h(t)\to\theta$ as $t\to\infty$, we can see that this model describes a case with no treatment effect initially but with a hazard ratio that changes over time to reach the limit θ as the long-term hazard ratio. Now it is straightforward to show that the hazard function (4) corresponds to the density function

$$f(t;\theta) = e^t\left(\frac{\theta}{\theta-1+e^t}\right)^{\theta+1}$$

for $t\ge 0$ and 0 otherwise. Under $H_0:\theta=1$, the observations in both samples are iid standard exponential. We again use (3) to derive the locally most powerful rank test of $H_0:\theta=1$ against $H_1:\theta<1$. Note that $\theta<1$ corresponds to a treatment effect of decreased hazard and thus longer survival time. Because rank tests are invariant to increasing transformation of the data, we require that the data follow the foregoing model only up to an increasing transformation. The log-likelihood function is given by $\ell(\theta; T_1,\ldots,T_N) = m(\theta+1)\log\theta - (\theta+1)\sum_{i=1}^{m}\log(\theta-1+e^{T_i}) + c$, where c is a constant that does not depend on θ. Differentiating ℓ with respect to θ at $\theta=1$, we obtain the efficient score $\ell'(1; T_1,\ldots,T_N) = 2\sum_{i=1}^{m}(1-e^{-T_i}) - \sum_{i=1}^{m}T_i$. Treating $T_1,\ldots,T_N$ as the "complete" data and the ranks as the observed data and applying (3) at $\theta=1$, we obtain

$$\begin{aligned}&\left.\frac{\partial\log\Pr(R=r)}{\partial\theta}\right|_{\theta=1} \\ &\quad= E_{H_0}\left\{2\sum_{i=1}^{m}\left(1-e^{-T_{(r_i)}}\right) - \sum_{i=1}^{m}T_{(r_i)}\middle|R=r\right\} \\ &\quad= E_{H_0}\left\{2\sum_{i=1}^{m}\left(1-e^{-T_{(r_i)}}\right) - \sum_{i=1}^{m}T_{(r_i)}\right\},\end{aligned}$$

and the locally most powerful rank test of $H_0:\theta=1$ against $H_1:\theta<1$ is to reject H_0 if $\frac{\partial\log\Pr(R=r)}{\partial\theta}|_{\theta=1}$ is small (because the alternative is $\theta<1$ rather than $\theta>1$) or, equivalently, if

$\frac{-\partial \log \Pr(R=r)}{\partial \theta}|_{\theta=1}$ is large. Now under $H_0: \theta = 1, T_1, \ldots, T_N$ are iid standard exponential and $U_i = 1 - e^{-T_i}, i = 1, \ldots, N$, are iid uniform. It follows that the locally most powerful rank test of $H_0: \theta = 1$ against $H_1: \theta < 1$ is to reject H_0 if

$$S = \sum_{i=1}^{m} E\big(T_{(R_i)}\big) - 2\sum_{i=1}^{m} E\big(U_{(R_i)}\big)$$
$$= \sum_{i=1}^{m}\left(\frac{1}{N} + \frac{1}{N-1} + \cdots + \frac{1}{N-R_i+1}\right) - 2\sum_{i=1}^{m}\frac{R_i}{N+1}$$

is large. The first part of S can be recognized as the Savage score, and the second part is twice the Wilcoxon score. Taking the difference of the scores, the resulting rank test is a moderated version of the Savage test. This is reasonable because the alternative hypothesis being considered is one under which there is no treatment effect initially, which is less dramatic than the scenario of a constant hazard ratio of θ from time 0 to ∞, for which the Savage test is optimal.

4. CONCLUSIONS

We have proposed a neat way to derive locally most powerful rank tests by casting the problem in a framework in which the general result (3) can be invoked, which is something that we have not seen before. By invoking (3), one can avoid dealing with the complicated rank likelihood altogether. This new way to derive locally most powerful rank tests is recommended for a graduate course in nonparametrics. It not only shortens the proof considerably, but also demonstrates links between the theory of rank tests and statistical inference in general (i.e., that the locally most powerful rank test is just a special case of the locally most powerful test or the score test presumably taught in a first-year graduate course in mathematical statistics). Our proposed approach also reinforces the perspective that a rank test can be viewed as a case of inference based on incomplete data. In particular, the new proof provides insight into why locally most powerful rank tests take their particular form. The new rank test derived in Section 3 is of interest in its own right in the situation where we want to test for a difference between two survival distributions with a time-varying hazard ratio.

[Received November 2008. Revised March 2009.]

REFERENCES

Cox, D. R., and Hinkley, D. V. (1974), *Theoretical Statistics*, London: Chapman & Hall.

Hajek, J., Sidak, Z., and Sen, P. K. (1999), *Theory of Rank Tests* (2nd ed.), San Diego: Academic Press.

McLachlan, G. J., and Krishnan, T. (1997), *The EM Algorithm and Extensions*, New York: Wiley.

Randles, R. H., and Wolfe, D. A. (1979), *Introduction to the Theory of Nonparametric Statistics*, New York: Wiley.

Wiener, E. (2001), "Differentiating With Respect to a Parameter," *The College Mathematics Journal*, 32, 180–184.

Yang, S., and Prentice, R. (2005), "Semiparametric Analysis of Short-Term and Long-Term Hazard Ratios With Two-Sample Survival Data," *Biometrika*, 92, 1–17.