On A Rank-Based Likelihood

Joshua D. Naranjo*

Abstract

Properties are derived for an estimator (CRL) which maximizes a pseudolikelihood function based on ranks. Similar to the MLE, the CRL gives progressively less weight to extreme observations when the tails of the underlying distribution get thicker. The CRL is more robust than the regular MLE, in the sense of a magnitude bound on each observation's contribution to the defining equation.
KEY WORDS: Composite likelihood; Ranks; Robustness.

1 Introduction

Recently, there has been some interest in a form of likelihood estimation called *pseudolikelihood* (Besag, 1975), or *composite likelihood* (Lindsay, 1988). These methods have been used because they are flexible enough to allow simplifications when maximum likelihood methods are too complicated or too difficult to calculate. Another reason for using composite likelihood is that it allows the user to focus on that part of the likelihood which is of particular interest. In this paper, for instance, we choose to focus on the information contained in the *ranks* of the observations, given the actual values.

*Dept. Of Mathematics and Statistics, Western Michigan University, Kalamazoo, MI, USA 49008

Let $X_1, \ldots, X_n$ be an i.i.d. sample from a continuous distribution with cdf $F_\theta(x)$ and pdf $f\theta(x)$. Let $R_1, \ldots, R_n$ be the ranks of $X_1, \ldots, X_n$, respectively. Recall that the MLE maximizes the log-likelihood equation

$$L(\theta) = \sum_{i=1}^{n} \log f_\theta(x_i). \tag{1}$$

Replace each component $\log f_i$ by the conditional event $P[R_i = r_i | X_i = x_i]$, i.e. the conditional probability that the ith observation has its observed rank $R_i = r_i$ given its observed value $X_i = x_i$. The resulting pseudolikelihood function

$$CRL(\theta) = \sum_{i=1}^{n} \log P[R_i = r_i | X_i = x_i] \tag{2}$$

which we will call the composite rank likelihood, was first proposed by Lindsay (1988) to illustrate a method of information assessment for composite likelihoods. The value of θ that maximizes (2) will be called the CRL estimator.

We will consider applications of (2) in estimation of the center of symmetry for continuous symmetric distributions. It turns out that the resulting CRL estimator is robust in the sense that each observation's contribution to the defining equation is bounded. Furthermore, the CRL shares the MLE property of giving extreme observations progressively less weight as the tails of the underlying distribution get thicker. Finally, it is shown that the CRL coincides with the MLE at the logistic distribution.

2 The Estimate

Let $X_1, \cdots, X_n$ be a random sample from a continuous distribution with cdf $F(x-\theta)$, where F is symmetric about 0. Hence θ is the unique median and mean (when it exists) of the underlying distribution. Let $R_i = R(X_i)$ denote the rank of X_i among $X_1, \cdots, X_n$. Let $P[R_i = r_i | X_i = x_i]$ denote the probability that the ith case has its observed rank given its observed value. Using standard probability methods, we have

$$\begin{aligned} CRL(\theta) &= \sum logP[R_i = r_i | X_i = x_i] \\ &= \sum \log \binom{n-1}{r_i - 1} F(x_i - \theta)^{r_i - 1}[1 - F(x_i - \theta)]^{n - r_i} \\ &= \sum \log \binom{n-1}{i-1} F(y_i - \theta)^{i-1}[1 - F(y_i - \theta)]^{n-i} \end{aligned}$$

where $y_1 \leq \ldots \leq y_n$ are the ordered x_i's. The maximizing value of $CRL(\theta)$ is a solution to the defining equation (or normal equation) $\frac{d}{d\theta} CRL(\theta) = 0$. Assuming F has density f, we get

$$\sum (i-1) \frac{f(y_i - \hat{\theta})}{F(y_i - \hat{\theta})} - (n-i) \frac{f(y_i - \hat{\theta})}{1 - F(y_i - \hat{\theta})} = 0$$

or equivalently

$$\sum \frac{f(y_i - \hat{\theta})}{F(y_i - \hat{\theta})[1 - F(y_i - \hat{\theta})]} \left[F(y_i - \hat{\theta}) - \frac{i-1}{n-1} \right] = 0. \tag{3}$$

Thus the gradient $\bigtriangledown CRL$ is a *weighted* and *centered* version of the cdf. This is not a surprising result since R_i/n is a nonparametric version of $F(x_i - \theta)$.

3 The CDF Defining Equation

Ignoring the weights in (3) for now, consider the estimator $\hat{\theta}_F$ which solves the defining equation

$$\begin{aligned} 0 &= \sum \left[F(y_i - \hat{\theta}_F) - \frac{i-1}{n-1} \right] \\ &= \sum \left[F(y_i - \hat{\theta}_F) - \frac{1}{2} \right]. \end{aligned}$$

This is a special case of an M-estimator defining equation $\sum \psi(y_i - \theta) = 0$ where $\psi(t) = F(t) - 1/2$. (Note that $\psi(t) = f'(t)/f(t)$ defines the MLE.) For a rigorous development of the asymptotic distribution of M-estimators, see for example Hampel et. al (1986). Here, we give a heuristic proof of asymptotic normality of $\hat{\theta}_F$.

Under regularity conditions, a first order Taylors expansion of the right side about the true value θ yields

$$0 \doteq \sum [F(y_i - \theta) - 1/2] + (\hat{\theta}_F - \theta) \sum f(y_i - \theta) \tag{4}$$

so that

$$\sqrt{n}(\hat{\theta}_F - \theta) \doteq -\frac{(1/\sqrt{n} \sum [F(y_i - \theta) - 1/2]}{(1/n) \sum f(y_i - \theta)}. \tag{5}$$

It is well known that $F(X - \theta)$ has uniform distribution on the unit interval. By the Central Limit Theorem, $(1/\sqrt{n}) \sum [F(y_i - \theta) - 1/2] = (1/\sqrt{n}) \sum [F(x_i - \theta) - 1/2]$ has asymptotic normal distribution with mean 0 and variance 1/12. Furthermore, $(1/n) \sum f(y_i - \theta)$ converges in probability to $E_\theta[f(X - \theta)] = \int f^2(x - \theta)dx = \int f^2(x)\, dx$ so that

$$\sqrt{n}(\hat{\theta}_F - \theta) \xrightarrow{d} N\left(0, \frac{1}{12[\int f^2]^2}\right). \tag{6}$$

This is the asymptotic distribution of the Hodges-Lehmann estimator based on the Wilcoxon signed-rank statistic (see e.g. Hettmansperger, 1994). When $f(x)$ is the density of $N(0, \sigma^2)$, it can be shown that

$$\int f^2(x)dx = \frac{1}{2\sigma\sqrt{\pi}}. \tag{7}$$

Since the MLE at the normal distribution has asymptotic variance σ^2, then the asymptotic relative efficiency of $\hat{\theta}_F$ with respect to the MLE at the normal is

$$ARE = \frac{\sigma^2}{[2\sigma\sqrt{\pi}]^2/12} = \frac{3}{\pi} \doteq .95. \tag{8}$$

We see that $\hat{\theta}_F$, or the CRL with constant weights, has relatively high efficiency at the normal distribution.

4 On the weight function

Now we focus on the weight function

$$w_F(y_i - \theta) = \frac{f(y_i - \theta)}{F(y_i - \theta)[1 - F(y_i - \theta)]}. \tag{9}$$

One property of the regular MLE is that extreme sample values are emphasized less as the tail-thickness of the underlying distribution increases. This is illustrated by the fact that the MLE for the thin tailed normal is the sample mean (sensitive to extreme observations), while the MLE for the thicker tailed double exponential is the sample median (insensitive to extreme observations). This sensitivity of the estimator to extreme observations reflects sensitivity of the defining equation itself to extreme observations. The normal MLE, for instance, has defining equation $\sum(x_i - \theta) = 0$. Note that each term in the sum is unbounded in magnitude. The double exponential MLE, in comparison, has defining equation $\sum sgn(x_i - \theta) = 0$, where $sgn(t)$ is $1, 0$, or -1 depending on the sign of t. Each term in the sum is obviously bounded in magnitude by 1. This shows that effects of extreme observations on the MLE defining equation (and hence, the estimate) at the double exponential is bounded in magnitude.

The following discussion will show that the weight function $w_F(y - \theta)$ serves the purpose of regulating sensitivity of the CRL estimator to extreme sample values, depending on the tail-thickness of the underlying distribution. The gold standard of tail thickness for the CRL seems to be the logistic density. We will show that for extreme values of t, $w_F(t) = f(t)/[F(t)\overline{F}(t)]$ gets larger for f lighter tailed than the logistic, and tends to 0 for heavier tailed distributions than the logistic. At the logistic distribution itself, $w_F(t)$ is a constant.

For a formalization of ordering by tail weight, we refer to a definition by van Zwet (1970).

Definition 1 *Suppose G and H are cdf's of continuous distributions that are symmetric about 0. G is said to have lighter tails than H (or H has heavier tails than G) if $H^{-1}(G(x))$ is convex for $x \geq 0$.*

Note that if F is lighter than G and G is lighter than H, then F is lighter than H, so the definition provides a weak ordering. If G and H are lighter than each other, then we say they are equivalent. Now if G and H differ only in scale ($G = H(x/a)$), it can be shown that they are equivalent, so that tail ordering is a property between families of distributions. Between frequently used distributions, it can be shown that the uniform, normal, logistic, double exponential, and Cauchy distributions are ordered from light to heavy (see Hettmansperger, 1984, Sec. 2.9).

Theorem 1 *Let F be continuous and symmetric about 0. Let $L(x) = 1/[1+exp(-x)]$ be the cdf of the logistic distribution. Then for $x \geq 0$,*
i. $w_L(x) = 1$.
ii. If F is lighter tailed than L, then $w_F(x)$ is increasing in x.
iii. If F is heavier tailed than L, then $w_F(x)$ is decreasing in x.

Noting that $L(x)$ has density function $L'(x) = \exp(-x)/[1 + \exp(-x)]^2$, result (i) is straightforward. To prove (ii), suppose that F is lighter tailed than L. Then $L^{-1}(F(x)) = log(F(x)) - log(1 - F(x))$ is convex, so that the first derivative of the right hand side is increasing. But this derivative is $w_F(x)$, so we are done. Result (iii) is proved similarly, except that $L^{-1}(F(x))$ is concave.

5 Efficiency at the logistic distribution

Though the composite rank likelihood (2) looks very different from the regular likelihood (1), the CRL behaves more like the MLE than is immediately apparent. In fact, the CRL at the logistic distribution is also the MLE.

Theorem 2 *At the logistic distribution, the CRL and MLE are the same estimator.*

Proof. We only need to show that the CRL and MLE defining equations are the same. This follows from the fact that $f/[F(1-F)] = 1$ and $F(x) - 1/2 = f'(x)/f(x)$.

Thus, the CRL at the logistic is fully efficient.

6 Robustness

In this section, we consider an estimate to be robust if it's defining equation

$$\sum_{i=1}^{n} \psi_i(X_i, \theta) = 0 \tag{10}$$

has bounded components $|\psi_i(x)| < B$ for all i and for some $B < \infty$. This reflects an upper bound on sensitivity of the defining equation to extreme values. Note that the MLE is not robust at the normal. The defining equation at the normal is

$$\sum_{i=1}^{n} (X_i - \theta) = 0$$

which has $|\psi_i(x)| = |x|$ unbounded. The MLE does have bounded defining equation at the double exponential:

$$\sum_{i=1}^{n} sgn(X_i - \theta) = 0$$

where $sgn(t)$ is $1, 0$, or -1 depending on whethe t is positive, zero, or negative.

Boundedness of the defining equation is equivalent to the boundedness of the influence function, which is typically proportional to the function ψ in (10). See Hampel et al. (1986) for details. Heuristically, the influence function measures the effect of small data contamination on the estimator.

We now show that the CRL has bounded defining equation for all distributions. Recall the CRL defining equation

$$\sum \frac{f(y_i - \theta)}{F(y_i - \theta)[1 - F(y_i - \theta)]} \left[F(y_i - \theta) - \frac{i-1}{n-1} \right] = 0. \tag{11}$$

Since $|F(y_i - \theta) - \frac{i-1}{n-1}| \leq 1$, then the CRL is robust whenever the weight function $\frac{f(y_i-\theta)}{F(y_i-\theta)[1-F(y_i-\theta)]}$ is a bounded function. This is true for all distributions with tails thicker than the logistic. It remains to show boundedness of the defining equation for thinner-tailed distributions. Without loss of generality, consider the normal distribution, for which the weight function approaches infinity for extremely small or extremely large values of the argument. If the maximum order statistic y_n is taken to approach infinity, then its contribution to the defining equation (11) becomes

$$\frac{f(y_n - \theta)}{F(y_n - \theta)[1 - F(y_n - \theta)]} \left[F(y_n - \theta) - \frac{n-1}{n-1} \right] = -\frac{f(y_n - \theta)}{F(y_n - \theta)}$$

which is bounded in magnitude. On the other hand, if the minimum y_1 approaches negative infinity, its contribution is

$$\frac{f(y_1 - \theta)}{F(y_1 - \theta)[1 - F(y_1 - \theta)]} \left[F(y_1 - \theta) - \frac{1-1}{n-1} \right] = \frac{f(y_i - \theta)}{1 - F(y_i - \theta)}$$

which is also bounded in magnitude.

The robustness of the CRL is interesting particularly by the way it is achieved, i.e. via the centering procedure for its defining equation. If shown a defining equation of the form (11) with the centering term $(i-1)/(n-1)$ covered up, the more likely guesses for it would be $1/2$ or $i/(n+1)$. Neither of these alternatives would achieve robustness of the estimate.

7 Asymptotic Normality

The following theorem follows from standard asymptotic results on order statistics. See, for example, Serfling (1980) or Chernoff, Gastwirth, and Johns (1967).

Theorem 3 *Under regularity conditions,*

$$\sqrt{n}\left(\hat{\theta}-\theta\right) \xrightarrow{d} N\left(0, \frac{\int_0^1 \frac{1}{(1-u)^2}\left[\int_u^1 \frac{1}{t} f\left(F^{-1}(t)\right) dt\right]^2 du}{\left(\int \frac{f^3(u)}{F(u)[1-F(u)]} du\right)^2}\right). \tag{12}$$

An outline of the proof follows. Let $T_n(\hat{\theta})$ denote the left hand side of (3). The first order expansion about the true value θ gives

$$0 = T_n(\hat{\theta}) \doteq T_n(\theta) + (\hat{\theta}-\theta)T_n'(\theta).$$

Thus, we have

$$\sqrt{n}\left(\hat{\theta}-\theta\right) \doteq -\frac{(1/\sqrt{n})T_n(\theta)}{(1/n)T_n'(\theta)}.$$

It can be shown that the numerator has asymptotic distribution

$$\frac{1}{\sqrt{n}}T_n(\theta) \xrightarrow{d} N\left(0, \int_0^1 \frac{1}{(1-u)^2}\left[\int_u^1 \frac{1}{t} f\left(F^{-1}(t)\right)\right]^2 du\right).$$

The denominator converges in probability to $\int \frac{f^3(u)}{F(u)[1-F(u)]} du$.

For the logistic distribution $F(x) = 1/(1+exp(-x))$, it is easy to show that $f(F^{-1}(x)) = x(1-x)$. Thus the numerator of the asymptotic variance in (12) is

$$\int_0^1 \frac{1}{(1-u)^2}\left[\int_u^1 (1-t)dt\right]^2 du = 1/12.$$

Furthermore, $f(x)/F(x)[1-F(x)] = 1$, so that the denominator of the asymptotic variance may be written $[\int f^2(u)du]^2$. Thus the CRL at the logistic has the familiar asymptotic variance $1/12(\int f^2)^2$, as it should.

8 Summary

The CRL estimator maximizes the likelihood of the observed ranks, rather then the likelihood of the observed data. This rank likelihood looks very different from the regular likelihood, but they are equivalent when the underlying distribution is logistic. Similar to regular likelihood, the defining equation for the estimator downweights extreme observations when drawing from a thick-tailed distribution, while upweighting extreme observations when drawing from thin tailed distributions like the normal.

Despite the fact that extreme observations are given more weight when the underlying distribution is thin tailed, the CRL remains robust, with the defining equation remaining bounded over all distributions. This boundedness of the defining equation is achieved, surprisingly, not by a redescending weight function but by the centering method.

REFERENCES

BESAG, J.E. (1975), "Statistical Analysis of non-lattice data," *The Statistician*, 24, 179-95.

CHERNOFF, H., GASTWIRTH, J.L., and JOHNS, M.V., Jr. (1967), "Asymptotic distribution of linear combinations of order statistics, with applications to estimation," *Annals of Mathematical Statistics*, 38, 52-72.

HAMPEL,F.R., RONCHETTI,E.M., ROUSSEEUW,P.J. and STAHEL,W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* John Wiley & Sons, New York.

HETTMANSPERGER, T. P. (1984). *Statistical Inference Based on Ranks.* John Wiley and Sons, New York.

LINDSAY, B.G. (1988), "Composite Likelihood Methods," *Contemporary Mathematics*, 80, 221-39.

SERFLING, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.

VAN ZWET, W.R. (1970). Convex transformations of random variables. Mathematical Centre Tracts No. 7, Mathematesch Centrum Amsterdam.