应用概率统计　第二十四卷
第四期　2008年8月

Chinese Journal of Applied Probability
and Statistics Vol.24 No.4 Aug. 2008

Empirical Likelihood-Based Inference with Missing and Censored Data

ZHENG MING　　DU WEI
(*Department of Statistics, Fudan University, Shanghai, 200433*)

Abstract

In this paper, we investigate how to apply the empirical likelihood method to the mean in the presence of censoring and missing. We show that an adjusted empirical likelihood statistic follows a chi-square distribution. Some simulation studies are presented to compare the empirical likelihood method with the normal method. These results indicate that the empirical likelihood method works better than or equally to the normal method.

Keywords: Empirical likelihood, censoring, linear regression imputation.

AMS Subject Classification: 62N02, 62N03.

§1. Introduction

Many statistical experiments result in incomplete sample, even under well-controlled conditions. This is because individuals will experience some other competing events which cause them to be removed. In such cases, the event of interest is not observable. In this paper, we investigate how to estimate the mean in the presence of missing and censoring.

A common method for handling missing data is to impute a value for each missing response and then apply standard methods to the complete data set as if they were true observations. Commonly used imputation methods include ratio and linear regression imputation, nearest neighbor imputation and kernel regression imputation. The idea of filling in the least squares estimates of all missing values in the analysis of variance and covariance dates back to Yates (1933)[1] and Bartlett (1937)[2]. We refer the reader to Little & Rubin (1987, ch.4)[3] for an excellent account of imputation methods for missing responses.

Censored data often arise in the study of medical follow-up, survival analysis, biometry and reliability study. In the last two decades, statistical inference with censored data has been paid considerable attention and studied extensively.

Received January 26, 2005. Revised January 4, 2007.

Let $Y_1, Y_2, \cdots, Y_n$ be nonnegative independent and identically distributed (i.i.d.) random variables (r.v.) with unknown distribution function F, the mean of which is what we are interested in and we denote as θ. Let $C_1, C_2, \cdots, C_n$ be nonnegative i.i.d. censoring r.v. with the distribution function G which is assumed known here. It is also assumed that Y_i's and C_i's are independent. In the random censoring model, the true survival values $Y_1, Y_2, \cdots, Y_n$ are not observable. Instead, one observes only $Z_i = \min(Y_i, C_i)$ and $\delta_i = I(Y_i \le C_i)$, where $I(\cdot)$ denotes the indicator function.

In practice, we may find that Y is missing at random (MAR), which means that whether Y is missing or not is independent of the value of Y. We denote η_i as the indicator of whether Y_i is missing or not. $\eta_i = 0$ if Y_i is missing, otherwise $\eta_i = 1$.

The paper is structured as follows. In Section 2, we construct confidence intervals for the mean by both normal method and empirical likelihood method. The empirical likelihood method was first suggested by Art. Owen (1988, 1990)[4, 5]. According to him, confidence intervals are constructed by empirical likelihood ratio. Compared with the traditional parametric methods, the empirical likelihood-based intervals are more accurate under small-sample circumstances and are range preserving[6–9]. In section 3, we do some simulations to compare the performance of the two methods under various conditions. In section 4, we give some proofs of the main results.

§2. Methodology and Main Results

First of all, we use the linear regression imputation method to impute the missing values. We assume the following regression model:

$$Y_i = X_i^\tau \beta + \varepsilon_i, \qquad i = 1, 2, \cdots, n, \tag{2.1}$$

where β is a $p \times 1$ vector of regression parameter. ε_i's are i.i.d. random errors with the mean 0 and independent of X_i's.

By the method of Koul, Susarla and Van Ryzin (1981) (hereafter abbreviated as K-S-V), we get the estimator for β:

$$\widehat{\beta} = \Big(\sum_{i=1}^n \eta_i X_i X_i^\tau \Big)^{-1} \sum_{i=1}^n \eta_i X_i Y_i^*,$$

where $Y_i^* = Z_i \delta_i / [1 - G(Z_i)]$, $i = 1, 2, \cdots, n$.

Hence, we can use the regression imputation to impute Y_i by $X_i^\tau \widehat{\beta}$ if Y_i is missing. Let

$$Z_{in} = \eta_i Y_i^* + (1 - \eta_i) X_i^\tau \widehat{\beta}, \qquad i = 1, 2, \cdots, n.$$

We then use $\widehat{\theta}=(1/n)\cdot\sum\limits_{i=1}^{n}Z_{in}$ as the estimator of θ.

Theorem 2.1 Assume $\mathsf{E}\|X\|<\infty$, $\mathsf{E}\varepsilon^2<\infty$, and if θ_0 is the true value of θ, then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_{in}-\theta_0)\longrightarrow N(0,V(\theta_0)),$$

where

$$\begin{aligned}
&V(\theta_0)=\sigma_1^2+\sigma_2^2+2\sigma_{12},\\
&\sigma_1^2=\mathsf{E}[\eta(S_1S_2^{-1}X+1)^2(Y^*-X^\tau\beta)^2],\qquad \sigma_2^2=\mathsf{E}(X^\tau\beta-\theta_0)^2,\\
&\sigma_{12}=\mathsf{E}[\eta(S_1S_2^{-1}X+1)(Y^*-X^\tau\beta)(X^\tau\beta-\theta_0)],\\
&S_1=\mathsf{E}[(1-\eta)X^\tau],\qquad S_2=\mathsf{E}(\eta XX^\tau).
\end{aligned}$$

Since $V(\theta_0)$ is unknown, we can use its moment estimator $\widehat{V}(\theta_0)$ instead. $\widehat{V}(\theta_0)=\widehat{\sigma}_1^2+\widehat{\sigma}_2^2+2\widehat{\sigma}_{12}$, where

$$\begin{aligned}
&\widehat{\sigma}_1^2=n^{-1}\sum_{i=1}^{n}\eta_i(\widehat{S}_1\widehat{S}_2^{-1}X_i+1)^2(Y_i^*-X_i^\tau\widehat{\beta})^2,\qquad \widehat{\sigma}_2^2=n^{-1}\sum_{i=1}^{n}(X_i^\tau\widehat{\beta}-\theta_0)^2,\\
&\widehat{\sigma}_{12}=n^{-1}\sum_{i=1}^{n}\eta_i(\widehat{S}_1\widehat{S}_2^{-1}X_i+1)(Y_i^*-X_i^\tau\widehat{\beta})(X_i^\tau\widehat{\beta}-\theta_0),\\
&\widehat{S}_1=n^{-1}\sum_{i=1}^{n}(1-\eta_i)X_i^\tau,\qquad \widehat{S}_2=n^{-1}\sum_{i=1}^{n}\eta_iX_iX_i^\tau.
\end{aligned}$$

We can obtain normal approximation $1-\alpha$ confidence interval $(\widehat{\theta}-u_{1-\alpha/2}\sqrt{\widehat{V}(\widehat{\theta})/n}$, $\widehat{\theta}+u_{1-\alpha/2}\sqrt{\widehat{V}(\widehat{\theta})/n})$, where $u_{1-\alpha/2}$ is $1-\alpha/2$ quartile of the standard normal distribution.

But when the sample size is small, the above symmetric interval based on the central limit theorem has poor performance. So, we can also consider the confidence interval based on the empirical likelihood method, which makes very weak distributional assumptions and is justified by having asymptotically correct coverage levels.

Let F_p be the distribution function which assigns probability p_i at the point Z_{in} for $i=1,2,\cdots,n$. Then, we have $\theta(F_p)=\sum\limits_{i=1}^{n}p_iZ_{in}$. An empirical log-likelihood ratio, evaluated at $\theta=\theta_0$, is then defined as

$$\widehat{l}_n(\theta_0)=-2\max_{\theta(F_p)=\theta_0,\sum\limits_{i=1}^{n}p_i=1}\sum_{i=1}^{n}\log(np_i). \tag{2.2}$$

By considering a Lagrange multiplier argument, we get

$$\widehat{l}_n(\theta_0)=2\sum_{i=1}^{n}\log(1+\lambda(Z_{in}-\theta_0)), \tag{2.3}$$

where λ is the solution of the following equation:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{Z_{in}-\theta_0}{1+\lambda(Z_{in}-\theta_0)}=0. \tag{2.4}$$

Compared to the standard empirical log-likelihood function, the main difference is that Z_{in}'s in $\widehat{l}_n(\theta_0)$ are not i.i.d.. As a result, the asymptotic distribution of $\widehat{l}_n(\theta_0)$ is a non-standard chi-square distribution in which case confidence intervals are not easily contributed as in the case of standard chi-square distribution. So $\widehat{l}_n(\theta_0)$ should be adjusted to follow the standard chi-square distribution.

Let

$$\widehat{l}_{ad}(\theta_0)=r(\theta_0)\widehat{l}_n(\theta_0), \tag{2.5}$$

where

$$r(\theta_0)=\frac{V_n(\theta_0)}{\widehat{V}(\theta_0)} \qquad \text{and} \qquad V_n(\theta_0)=\frac{1}{n}\sum_{i=1}^{n}(Z_{in}-\theta_0)^2.$$

Then, $\widehat{l}_{ad}(\theta_0)$ has asymptotically a standard chi-square distribution with 1 degree of freedom as stated below.

Theorem 2.2 Suppose $\mathsf{E}\|X\|<\infty$, $\mathsf{E}\varepsilon^2<\infty$, and if θ_0 is the true value of θ, then $\widehat{l}_{ad}(\theta_0)$ is asymptotically χ_1^2. That is, $\mathsf{P}(\widehat{l}_{ad}(\theta_0)\le c_\alpha)=1-\alpha+o(1)$, where c_α satisfies $\mathsf{P}(\chi_1^2\le c_\alpha)=1-\alpha$.

By theorem 2.2, an asymptotically correct $1-\alpha$ level confidence interval can be obtained as $I_\alpha=\{\theta:\widehat{l}_{ad}(\theta)\le c_\alpha\}$.

Corollary 2.1 In particular, when there is no censoring, $Z_{in}=\eta_iY_i+(1-\eta_i)X_i^\tau\widehat{\beta}$, where $\widehat{\beta}=\Big(\sum\limits_{i=1}^{n}\eta_iX_iX_i^\tau\Big)^{-1}\sum\limits_{i=1}^{n}\eta_iX_iY_i$. Suppose $\mathsf{E}\|X\|<\infty$, $\mathsf{E}\varepsilon^2<\infty$, and if θ_0 is the true value of θ, then $\widehat{l}_{ad}(\theta_0)$ is also asymptotically χ_1^2.

This is Theorem 2.1 by Qihua Wang & J.N.K. Rao (2002)[10].

§3. Simulation Results

In section 2, we considered two methods for constructing confidence intervals in the presence of both missing and censoring. Following are the simulation results to compare these methods in terms of coverage accuracies based on them.

We use simple regression model $Y=\alpha+\beta X+\varepsilon$, where ε is a standard normal r.v. and X follows the exponential distribution with mean 1. The censoring r.v. C follows the exponential distribution with mean c. $N(\theta)$, $EL(\theta)$ denote the coverage probabilities for the confidence intervals on θ by normal method and empirical method respectively. We generated 1000 random samples of size $n = 20, 50, 100$. The mean of censoring r.v., $c = 10, 20, 50$. The missing probability is 0.05, 0.1 and 0.2 (see Table 1, 2 and 3).

From the results, we can find that in any situation, empirical likelihood method is better than normal method. As the missing probability increases, the coverage probability decreases.

Table 1 Coverage probabilities for the mean (missing probability is 0.05)

		$\alpha = 0.1$		$\alpha = 0.05$	
c	n	$N_\alpha(\theta)$	$EL_\alpha(\theta)$	$N_\alpha(\theta)$	$EL_\alpha(\theta)$
10	20	0.806	0.914	0.852	0.925
	50	0.831	0.886	0.896	0.917
	100	0.858	0.901	0.912	0.932
15	20	0.850	0.918	0.874	0.928
	50	0.865	0.899	0.921	0.940
	100	0.873	0.906	0.935	0.940
20	20	0.848	0.905	0.897	0.943
	50	0.875	0.906	0.902	0.929
	100	0.898	0.909	0.923	0.940

Table 2 Coverage probabilities for the mean (missing probability is 0.1)

		$\alpha = 0.1$		$\alpha = 0.05$	
c	n	$N_\alpha(\theta)$	$EL_\alpha(\theta)$	$N_\alpha(\theta)$	$EL_\alpha(\theta)$
10	20	0.789	0.902	0.830	0.918
	50	0.827	0.891	0.884	0.908
	100	0.860	0.899	0.896	0.932
15	20	0.803	0.881	0.867	0.913
	50	0.834	0.875	0.902	0.933
	100	0.862	0.891	0.919	0.937
20	20	0.827	0.896	0.898	0.933
	50	0.867	0.886	0.924	0.932
	100	0.886	0.905	0.924	0.941

Table 3 Coverage probabilities for the mean (missing probability is 0.2)

		$\alpha=0.1$		$\alpha=0.05$	
c	n	$N_\alpha(\theta)$	$EL_\alpha(\theta)$	$N_\alpha(\theta)$	$EL_\alpha(\theta)$
10	20	0.763	0.857	0.830	0.896
	50	0.818	0.877	0.856	0.901
	100	0.831	0.872	0.890	0.915
15	20	0.795	0.867	0.824	0.872
	50	0.814	0.854	0.855	0.897
	100	0.835	0.881	0.899	0.931
20	20	0.819	0.883	0.862	0.892
	50	0.809	0.843	0.876	0.902
	100	0.822	0.873	0.889	0.917

§4. Proofs

Proof of Theorem 2.1:

First, from the K-S-V method, we know that $\mathsf{E}(Z_i\delta_i/[1-G(Z_i)])=\mathsf{E}(Y_i)$. Thus, we can conclude that $\mathsf{E}(Z_{in})=\mathsf{E}(Y_i)$.

Then, recalling the definition of $\widehat{\beta}$, we have

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^n(1-\eta_i)X_i^\tau(\widehat{\beta}-\beta)\\
=\ &n^{-1/2}\sum_{i=1}^n(1-\eta_i)X_i^\tau\Big[\Big(\sum_{j=1}^n\eta_jX_jX_j^\tau\Big)^{-1}\sum_{k=1}^n\eta_kX_kY_k^*-\beta\Big]\\
=\ &n^{-1/2}\Big(\sum_{i=1}^n(1-\eta_i)X_i^\tau\Big)\Big(\sum_{j=1}^n\eta_jX_jX_j^\tau\Big)^{-1}\sum_{k=1}^n\eta_kX_k(Y_k^*-X_k^\tau\beta)\\
=\ &n^{-1/2}\mathsf{E}[(1-\eta)X^\tau][\mathsf{E}(\eta XX^\tau)]^{-1}\sum_{k=1}^n\eta_kX_k(Y_k^*-X_k^\tau\beta)+o_p(1)\\
=\ &n^{-1/2}S_1S_2^{-1}\sum_{k=1}^n\eta_kX_k(Y_k^*-X_k^\tau\beta)+o_p(1).
\end{aligned}\tag{4.1}$$

Reorganize (4.1), we get

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^n(1-\eta_i)X_i^\tau\widehat{\beta}\\
=\ &n^{-1/2}S_1S_2^{-1}\sum_{k=1}^n\eta_kX_k(Y_k^*-X_k^\tau\beta)+n^{-1/2}\sum_{i=1}^n(1-\eta_i)X_i^\tau\beta+o_p(1).
\end{aligned}\tag{4.2}$$

By applying (4.2), we have

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}(Z_{in}-\theta_0) \\
= \ & n^{-1/2}\sum_{i=1}^{n}(\eta_i Y_i^* + (1-\eta_i)X_i^\tau\widehat{\beta}-\theta_0) \\
= \ & n^{-1/2}\sum_{i=1}^{n}(\eta_i Y_i^*-\theta_0) + n^{-1/2}\sum_{i=1}^{n}((1-\eta_i)X_i^\tau\widehat{\beta}) \\
= \ & n^{-1/2}\sum_{i=1}^{n}(\eta_i Y_i^*-\theta_0) + n^{-1/2}S_1S_2^{-1}\sum_{i=1}^{n}\eta_i X_i(Y_i^*-X_i^\tau\beta) \\
& + n^{-1/2}\sum_{i=1}^{n}((1-\eta_i)X_i^\tau\beta) + o_p(1) \\
= \ & n^{-1/2}\sum_{i=1}^{n}(\eta_i Y_i^*-\theta_0 + S_1S_2^{-1}\eta_i X_i(Y_i^*-X_i^\tau\beta) + (1-\eta_i)X_i^\tau\beta) + o_p(1) \\
= \ & n^{-1/2}\sum_{i=1}^{n}(\eta_i(S_1S_2^{-1}X_i+1)(Y_i^*-X_i^\tau\beta) + (X_i^\tau\beta-\theta_0)) + o_p(1).
\end{aligned} \tag{4.3}
$$

By the central limit theorem, (4.3) $\longrightarrow N(0,V(\theta_0))$, where

$$V(\theta_0)=\mathsf{E}(\eta_i(S_1S_2^{-1}X_i+1)(Y_i^*-X_i^\tau\beta)+(X_i^\tau\beta-\theta_0))^2=\sigma_1^2+\sigma_2^2+2\sigma_{12},$$

and

$$
\begin{aligned}
& \sigma_1^2=\mathsf{E}[\eta(S_1S_2^{-1}X+1)^2(Y^*-X^\tau\beta)^2], \qquad \sigma_2^2=\mathsf{E}(X^\tau\beta-\theta_0)^2, \\
& \sigma_{12}=\mathsf{E}[\eta(S_1S_2^{-1}X+1)(Y^*-X^\tau\beta)(X^\tau\beta-\theta_0)].
\end{aligned}
$$

□

Lemma 4.1　Under the condition of theorem 2.1, we have

$$\frac{1}{n}\sum_{i=1}^{n}(Z_{in}-\theta_0)^2=\sigma_2^2+\sigma_3^2+2\sigma_{23}+o_p(1),$$

where

$$
\begin{aligned}
& \sigma_2^2=\mathsf{E}(X^\tau\beta-\theta_0)^2, \qquad \sigma_3^2=\mathsf{E}(\eta(Y^*-X^\tau\beta)^2, \\
& \sigma_{23}=\mathsf{E}[\eta(X^\tau\beta-\theta_0)(Y^*-X^\tau\beta)].
\end{aligned}
$$

Proof　By the law of large numbers and the fact that $\widehat{\beta}\xrightarrow{p}\beta$, it is easy to see that

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(Z_{in}-\theta_0)^2 &= \frac{1}{n}\sum_{i=1}^{n}(\eta_i Y_i^*+(1-\eta_i)X_i^\tau\widehat{\beta}-\theta_0)^2 \\
&= R_{n1}+R_{n2}+R_{n3}+o_p(1),
\end{aligned} \tag{4.4}
$$

where

$$R_{n1}=\frac{1}{n}\sum_{i=1}^{n}(X_i^\tau\beta-\theta_0)^2,\qquad R_{n2}=\frac{1}{n}\sum_{i=1}^{n}\eta_i(Y_i^*-X_i^\tau\beta)^2,$$
$$R_{n3}=\frac{2}{n}\sum_{i=1}^{n}\eta_i(X_i^\tau\beta-\theta_0)(Y_i^*-X_i^\tau\beta).$$

By the law of large numbers, we have

$$R_{n1}\xrightarrow{p}\sigma_2^2, \tag{4.5}$$
$$R_{n2}\xrightarrow{p}\sigma_3^2, \tag{4.6}$$
$$R_{n3}\xrightarrow{p}2\sigma_{23}. \tag{4.7}$$

From (4.5)–(4.7), lemma 4.1 is then proved. □

Lemma 4.2 Let $\widehat{Z}_{(n)}=\max\limits_{1\le i\le n}|Z_{in}|$. If the conditions of theorem 2.1 are satisfied, we have

$$Z_{(n)}=o_p(n^{1/2}).$$

Proof Notice that

$$Z_{(n)}\le\max_{1\le i\le n}|Y_i^*|+\max_{1\le i\le n}\|X_i\|\|\widehat{\beta}\|. \tag{4.8}$$

By lem.3 of Owen (1988), it follows that

$$\max_{1\le i\le n}|Y_i^*|=o_p(n^{1/2}),\qquad\max_{1\le i\le n}\|X_i\|=o_p(n^{1/2}). \tag{4.9}$$

(4.8) and (4.9) together with the fact $\widehat{\beta}=O_p(1)$ prove lemma 4.2. □

Lemma 4.3 Under assumption of theorem 2.1, we have

$$\lambda=O_p(n^{1/2}).$$

Proof By theorem 2.1, it follows that

$$\frac{1}{n}\sum_{i=1}^{n}(Z_{in}-\theta_0)=o_p(n^{-1/2}). \tag{4.10}$$

This together with lemmas 4.1 and 4.2 proves lemma 4.3 by the same arguments as used in Owen (1988). □

Proof of Theorem 2.2:

Applying Taylor's expansion to $\widehat{l}_n(\theta_0)$, we have

$$\widehat{l}_n(\theta_0)=2\sum_{i=1}^{n}\{\lambda(Z_{in}-\theta_0)-\frac{1}{2}[\lambda(Z_{in}-\theta_0)]^2\}+\gamma_n \tag{4.11}$$

with $|\gamma_n| \le C\sum_{i=1}^n |\lambda(Z_{in}-\theta_0)|^3$ in probability.

By lemmas 4.1, 4.2 and 4.3, it follows that

$$|\gamma_n| \le C|\lambda|^3 \max_{1\le i\le n}|Z_{in}-\theta_0|\sum_{i=1}^n (Z_{in}-\theta_0)^2 = o_p(1). \tag{4.12}$$

Notice that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \frac{Z_{in}-\theta_0}{1+\lambda(Z_{in}-\theta_0)} &= \frac{1}{n}\sum_{i=1}^n (Z_{in}-\theta_0) - \Big[\frac{1}{n}\sum_{i=1}^n (Z_{in}-\theta_0)^2\Big]\lambda \\ &\quad + \frac{1}{n}\sum_{i=1}^n \frac{\lambda^2(Z_{in}-\theta_0)^3}{1+\lambda(Z_{in}-\theta_0)}.\end{aligned} \tag{4.13}$$

By (2.4), (4.13) and lemma 4.1–4.3, we get

$$\lambda = \Big(\sum_{i=1}^n (Z_{in}-\theta_0)^2\Big)^{-1}\sum_{i=1}^n (Z_{in}-\theta_0) + o_p(n^{1/2}). \tag{4.14}$$

Again using (2.4), we get

$$\begin{aligned}0 &= \sum_{i=1}^n \frac{\lambda(Z_{in}-\theta_0)}{1+(Z_{in}-\theta_0)} \\ &= \sum_{i=1}^n [\lambda(Z_{in}-\theta_0)] - \sum_{i=1}^n [\lambda(Z_{in}-\theta_0)]^2 + \sum_{i=1}^n \frac{[\lambda(Z_{in}-\theta_0)]^3}{1+\lambda(Z_{in}-\theta_0)}.\end{aligned} \tag{4.15}$$

By lemmas 4.1–4.3, it follows that

$$\sum_{i=1}^n \frac{[\lambda(Z_{in}-\theta_0)]^3}{1+\lambda(Z_{in}-\theta_0)} = o_p(1). \tag{4.16}$$

From (4.15) and (4.16), we get

$$\sum_{i=1}^n [\lambda(Z_{in}-\theta_0)] = \sum_{i=1}^n [\lambda(Z_{in}-\theta_0)]^2 + o_p(1). \tag{4.17}$$

By (4.11), (4.12), (4.14) and (4.17), we have

$$\widehat{l}_n(\theta_0) = \Big[\frac{1}{n}\sum_{i=1}^n (Z_{in}-\theta_0)^2\Big]^{-1}\Big[\frac{1}{\sqrt{n}}\sum_{i=1}^n (Z_{in}-\theta_0)\Big]^2 + o_p(1).$$

Hence,

$$\widehat{l}_{ad}(\theta_0) = \Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Z_{in}-\theta_0}{[\widehat{V}(\theta_0)]^{1/2}}\Big\}^2 + o_p(1). \tag{4.18}$$

Standard arguments can be used to prove that

$$\widehat{V}(\theta_0) \longrightarrow V(\theta_0).$$

This together with theorem 2.1 and (4.18) proves theorem 2.2. □

References

[1] Yates, F., The analysis of replicated experiments when the field results are incomplete, *Emp. J. Exp. Agric.*, **1**(1933), 129–142.

[2] Bartlett, M.S., Some examples of statistical methods of research in agriculture and applied botany, *J. Roy. Statist. Soc.*, Ser. B: **4**(1937), 137–170.

[3] Little, R.J.A. & Rubin, D.B., *Statistical Analysis with Missing Data*, Wiley, New York, 1987.

[4] Owen, A., Empirical likelihood ratio confidence intervals for a single functional, *Biomctrika*, **75**(1988), 237–249.

[5] Owen, A., Empirical likelihood ratio confidence regions, *Ann. Statist.*, **18**(1990), 90–120.

[6] Chen, S.X. and Hall, P., Smoothed empirical likelihood confidence intervals for quintiles, *Ann. Statist.*, **21**(1993), 1161–1181.

[7] Wang, Q.H. and Jing, B.Y., Empirical likelihood for a class of functionals of survival distribution with censored data, *Ann. Inst. Statist. Math.*, **53(3)**(2001), 517–527.

[8] Zheng, M. & Li, S.H., Empirical likelihood in partial linear error-in-covariable model with censored data, *Communication in Statistics: Theory and Methods*, **34(2)**(2005), 389–404.

[9] Zheng, M., Li, S.H. & Yang, Y., Empirical likelihood for a class of functionals of survival distribution with censored data, *Appl. Math. J. Chinese Univ.*, Ser. A: **19(3)**(2004), 315–322.

[10] Wang, Q.H. & Rao, J.N.K., Empirical likelihood-based inference in linear models with missing data, *Scand. J. Statist.*, **29**(2002), 563–576.

含有截断和缺失数据的经验似然推断

郑 明 杜 玮

(复旦大学统计系, 上海, 200433)

本文将经验似然的方法应用到同时包含截断和缺失数据的情况. 通过定义调整后的经验似然比, 证明它服从χ^2分布. 利用随机模拟, 比较经验似然和正态方法的优劣. 结果发现经验似然方法在很多情况下都优于正态方法.

关键词: 经验似然, 截断, 线性回归估值.

学科分类号: O211.4, O212.7.