Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE @ DIRECT

Journal of Statistical Planning and Inference 115 (2003) 471–478

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Confidence limits to the distance of the true distribution from a misspecified family by bootstrap

G. Jogesh Babu*,[1], C.R. Rao[2]

Department of Statistics, The Pennsylvania State University, 326 Thomas Building, University Park, PA 16802, USA

Received 1 October 2001; accepted 26 March 2002

Abstract

In statistical practice, an estimated distribution function (d.f.) from a specified family is used for taking decisions. When the true d.f. from which samples are drawn does not belong to the specified family, it is of interest to know how close the true d.f. is to the specified family. In this paper, we use non-parametric bootstrap to obtain confidence limits to the difference between the true d.f. and a member of the specified family closest to it in the sense of Kullback–Leibler measure. © 2002 Elsevier Science B.V. All rights reserved.

MSC: primary 62F40; secondary 62F12; 62G10

Keywords: Brownian Motion; Empirical process; Bootstrap; L-statistics; Median; Kolmogorov–Smirnov statistic; Cramér–von Mises statistic; Kullback–Leibler measure

1. Introduction

Test statistics based on the empirical distribution function, when the family of distributions is correctly specified and the parameters are estimated have been extensively studied by Darling (1955), Kac et al. (1955), Durbin (1973) and others. However, even when the parametric model is specified, the asymptotic distribution of statistics based on the empirical process may depend on the unknown parameter.

* Corresponding author. Tel.: +1-814-863-2837; fax: +1-814-863-7114.
E-mail address: babu@stat.psu.edu (G.J. Babu).

[1] Supported in part by NSF grant DMS 0101360.
[2] Supported in part by ARO grant 19-01-1-0563.

PII: S0378-3758(02)00176-3

In Babu and Rao (2001), we provided a bootstrap approach for testing the null hypothesis that a sample comes from a specified parametric family of distributions. If the hypothesis is rejected, it would be of interest to examine how close the actual alternative distribution is to the specified family of distributions. In this paper, we define closeness in terms of Kullback–Leibler measure of separation and provide a bootstrap approach to set confidence bands to the difference of the true and the closest distribution in the specified family.

White (1982) examined the problem of detection of model misspecification when using maximum likelihood techniques for estimation and inference. Nishii (1988) studied this problem further. He considers the maximum likelihood based on a specified parametric family which provides a good approximation of the true unknown and unspecified distribution. Foutz and Srivastava (1977) obtained asymptotic results on a related likelihood ratio test when the model is incorrectly specified.

A similar problem of practical importance was considered by Konishi (1999), and Konishi and Kitagawa (1996). They obtain an asymptotically unbiased estimate of the difference between an estimated d.f. from a specified family and the true d.f. which may be outside the family, which can be used in model selection. Reference may also be made to a review paper by Rao and Wu (2001) on model selection procedures.

2. Bootstrap

Let $X_1,\ldots,X_n$ be i.i.d. random variables from a continuous distribution H. Let $\{F(\cdot;\theta)\colon \theta\in\Theta\}$ be a family of continuous distribution functions, where Θ is an open region in a p-dimensional Euclidean space. The distribution H may or may not belong to this family. Suppose H is in some sense closest to $F(\cdot;\theta_0)$ in the family $\{F(\cdot;\theta)\colon \theta\in\Theta\}$ so that an estimator, such as a maximum likelihood estimator, $\hat{\theta}_n=\theta_n(X_1,\ldots,X_n)$ converges to $\theta_0\in\Theta$. A measure of closeness is defined in Nishii (1988) through Kullback–Leibler measure

$$\int h(x)\log(h(x)/f(x;\theta))\,\mathrm{d}v(x)\geqslant 0,$$

where h and $f(\cdot;\theta)$ denote the densities of H and $F(\cdot;\theta)$ with respect to a fixed measured v and $\int|\log h(x)|h(x)\,\mathrm{d}v(x)<\infty$. Defining

$$\mathrm{LL}(h;f,\theta)=\int h(x)\log f(x;\theta)\,\mathrm{d}v(x),$$

the closest parameter θ_0 is given by

$$\mathrm{LL}(h;f,\theta_0)=\max_{\theta\in\Theta}\mathrm{LL}(h;f,\theta).$$

Nishii (1988) shows that the maximum likelihood estimator $\hat{\theta}_n\to\theta_0$ under some regularity conditions. The results of our paper hold for many types of estimators including maximum likelihood estimators.

The results are applicable to several statistics based on empirical measures. In particular, we consider Kolmogorov–Smirnov and Cramér–von Mises statistics. These statistics can be viewed as continuous functionals of the empirical process

$$Y_n^0(x;\hat{\theta}_n) = \sqrt{n}(F_n(x) - F(x;\hat{\theta}_n)),$$

where F_n denotes the empirical distribution function of $X_1,\ldots,X_n$. We shall show that under some regularity conditions,

$$Y_n(x;\hat{\theta}_n) = \sqrt{n}(F_n(x) - F(x;\hat{\theta}_n)) - \sqrt{n}(H(x) - F(x;\theta_0)) \tag{2.1}$$

converges weakly to a Gaussian process Y. The covariance function of Y is given by (4.2).

To develop a bootstrap procedure, let $X_1^*,\ldots,X_n^*$ be i.i.d. random variables from F_n. Let $\hat{\theta}_n^* = \theta_n(X_1^*,\ldots,X_n^*)$ and let F_n^* denote the empirical distribution function of $X_1^*,\ldots,X_n^*$. The bootstrap process corresponding to Y_n is given by

$$Y_n^{\mathrm{b}}(x) = \sqrt{n}(F_n^*(x) - F(x;\hat{\theta}_n^*)) - \sqrt{n}(F_n(x) - F(x;\hat{\theta}_n)). \tag{2.2}$$

We shall show that under some regularity conditions both Y_n and Y_n^{b} converge to the same limiting Gaussian process Y. Here the distribution of Y_n^{b} is induced by the bootstrap sampling P^* given the sample $X_1,\ldots,X_n$. In principle, the distribution of the bootstrap process Y_n^{b} is completely known, though it may not have any simple closed-form representation. In practice, the distribution of Y_n^{b} is approximated by the use of B bootstrap replications, where B is large (Babu and Singh, 1983).

3. Confidence bands

The weak convergence of processes (2.1) and (2.2) to the same limiting Gaussian process can be used to obtain confidence bands for the difference $H - F(\cdot;\theta_0)$. This provides an estimate of the distance between the true distribution and the family of distributions under consideration. In particular, if the distribution of the Kolmogorov–Smirnov-type distance

$$D_n(H;\theta_0) = \sqrt{n}\sup_x |F_n(x) - F(x;\hat{\theta}_n) - (H(x) - F(x;\theta_0))|$$

is known, then its critical values can be used to set up confidence bands for $H - F(\cdot;\theta_0)$. That is, if $P(D_n(H;\theta_0) \leqslant C_{\alpha,n}) - \alpha \to 0$ for some $C_{\alpha,n}$, then

$$P(F_n(x) - F(x;\hat{\theta}_n) - J_{\alpha,n} \leqslant H(x) - F(x;\theta_0) \leqslant F_n(x) - F(x;\hat{\theta}_n) + J_{\alpha,n},$$
$$\text{for all } x) \to \alpha, \tag{3.1}$$

where $0 < \alpha < 1$ and $\sqrt{n}J_{\alpha,n} = C_{\alpha,n}$. In general, $C_{\alpha,n}$ is not known. Since supremum norm is a continuous functional, Theorems 4.1 and 4.2, show that the distribution of $D_n(H;\theta_0)$ and the distribution of D_n^*, under the bootstrap sampling given the data $X_1,\ldots,X_n$, tend to the same limiting distribution, where

$$D_n^* = \sqrt{n}\sup_x |F_n^*(x) - F(x;\hat{\theta}_n^*) - (F_n(x) - F(x;\hat{\theta}_n))|.$$

That is

$$\sup_x |P(D_n(H;\theta_0) < x) - P^*(D_n^* < x \,|\, X_1,\ldots)| \to 0,$$

as $n \to \infty$. From Lemma 2.1 of Babu and Bose (1988), it follows for any $0 < \alpha < 1$ that

$$P\left(\sqrt{n}\sup_x |F_n(x) - F(x;\hat{\theta}_n) - (H(x) - F(x;\theta_0))| \leqslant C^*_{\alpha,n}\right) - \alpha \to 0,$$

where $C^*_{\alpha,n}$ is an αth quantile of the distribution of D^*_n. Thus $C^*_{\alpha,n}$ can be used in the place of $C_{\alpha,n}$ in (3.1).

In principle, $C^*_{\alpha,n}$ is completely known, but may not have a simple closed form suitable for computations. Consequently, in practice, it is estimated by using a large number of (say B) bootstrap replications. This is done by first computing D^*_n for each of the generated B bootstrap samples of size n and arranging the resulting D^*_n's in increasing order. Then $C^*_{\alpha,n}$ is chosen to be the $[B\alpha + 1]$ largest D^*_n.

Similar bootstrap analysis can be done to get one-sided confidence intervals, by using the distances

$$\sup_x (F_n(x) - F(x;\hat{\theta}_n) - (H(x) - F(x;\theta_0)))^+$$

and

$$\sup_x (F_n(x) - F(x;\hat{\theta}_n) - (H(x) - F(x;\theta_0)))^-$$

instead, where $y^+ = \max(0, y)$ and $y^- = \max(0, -y)$ denote positive and negative parts of y.

The Cramér–von Mises-type distances,

$$n\int (F_n(x) - F(x;\hat{\theta}_n) - (H(x) - F(x;\theta_0)))^2 \,\mathrm{d}F(x;\theta_0) \tag{3.2}$$

or

$$n\int (F_n(x) - F(x;\hat{\theta}_n) - (H(x) - F(x;\theta_0)))^2 \,\mathrm{d}F(x;\hat{\theta}_n) \tag{3.3}$$

can be used to treat tails with varying weights. Both (3.2) and (3.3) have the same limiting distributions as that of their bootstrap version

$$n\int (F^*_n(x) - F(x;\hat{\theta}^*_n) - (F_n(x) - F(x;\hat{\theta}_n)))^2 \,\mathrm{d}F(x;\hat{\theta}_n). \tag{3.4}$$

In practice, as in the case of Kolmogorov–Smirnov-type distance, B bootstrap replications of (3.4) can be used to estimate $B^*_{\alpha,n}$, so that

$$p\left(n\int (F_n(x) - F(x;\hat{\theta}_n) - (H(x) - F(x;\theta_0)))^2 \,\mathrm{d}F(x;\hat{\theta}_n) < B^*_{\alpha,n}\right) \to \alpha,$$

for $0 < \alpha < 1$.

4. Technical results

4.1. Assumptions

Let $\hat{\theta}_n \to \theta_0$ and let $\Lambda \subset \Theta$ be the closure of a given neighborhood of θ_0. We now list a set of assumptions used in the main results. The assumptions are also illustrated with examples below.

(A1) The row vector $g(x;\theta) = (\partial/\partial\theta)F(x;\theta)$ is uniformly continuous in x and $\theta \in \Lambda$.

(A2) For some $\varepsilon_n \to 0$ in probability,

$$\hat{\theta}_n - \theta_0 = \frac{1}{n}\sum_{i=1}^{n} \ell(X_i) + \frac{1}{\sqrt{n}}\varepsilon_n,$$

where ℓ is a measurable p-dimensional row vector valued function.

(A3) $\mathbf{E}(\ell(X_1)) = 0$.

(A4) $L = \mathbf{E}(\ell(X_1)\ell(X_1)')$ is a finite non-negative definite matrix.

(A5) For some $\varepsilon_n^* \to 0$ in probability under the bootstrap measure,

$$\hat{\theta}_n^* - \hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} \ell(X_i^*) - \frac{1}{n}\sum_{i=1}^{n} \ell(X_i) + \frac{1}{\sqrt{n}}\varepsilon_n^*$$

for almost all sample sequences $X_1, \ldots, X_n$.

Remarks on the Assumptions: Assumption (A2) asserts that $\hat{\theta}_n$ is locally asymptotically linear, and Assumption (A5) asserts the same for the bootstrapped version, given the original sample. Assumptions (A1)–(A4) in Nishii (1988) lead to our conditions (A2)–(A4) for estimators based on maximum likelihood principles. Foutz and Srivastava (1977) and White (1982) also use similar conditions in examining model misspecification when using maximum likelihood techniques. Assumption (A1) holds for many families of distributions including exponential families defined by the densities $f(x;\theta) = v_1(x)\exp\{v_2(x)\eta(\theta)^{\mathrm{T}} - \xi(\theta)\}$ with respect to a sigma finite measure, where the mean is finite and η and ξ have bounded derivatives in a neighborhood of θ_0. Assumptions (A2)–(A5) also hold, in general, for certain M-estimators and L-statistics. Explicit expressions of ℓ that appear in Assumptions (A2)–(A5) are derived in Babu and Singh (1984) for sample median and L-statistics. The details are presented below.

Sample median: Babu and Singh (1984) have shown that in the case of the sample median and L-statistics, condition (A5) is satisfied. In fact, in the case of the sample median $\hat{\theta}_n$, Assumptions (A2) and (A5) hold under very general conditions, with

$$\ell(x) = \frac{1}{a}\left(I_{\{x \leqslant \theta_0\}} - \frac{1}{2}\right), \tag{4.1}$$

where a denotes the density evaluated at the population median θ_0 of H (Babu and Singh 1984, Theorem 5).

L-statistic: In the case of L-satistic, $\hat{\theta}_n = \int x\omega(F_n(x))\mathrm{d}F_n(x)$ and $\theta_0 = \int x\omega(H(x))$ $\mathrm{d}H(x)$, Babu and Singh (1984, Theorem 4, p. 9) have shown that (A2) and (A5) hold with

$$\ell(y) = \int_{-\infty}^{\infty} (H(x) - I_{\{y \leqslant x\}})\omega(H(x))\,\mathrm{d}x,$$

when ω satisfies Lipschitz condition of order 1 on each of the intervals (a_{i-1}, a_i), $i = 1, \ldots, k+1$, $a_0 = 0 < a_1 < \cdots < a_k < a_{k+1} = 1$, and the quantile function H^{-1} of H is continuous at $a_1, \ldots, a_k$.

4.2. Theorems

Let

$$b(x) = \int_{-\infty}^{x} \ell(t)\,\mathrm{d}H(t).$$

We now state the main theorem.

Theorem 4.1. *Suppose conditions* (A1)–(A4) *hold. Then the process* Y_n *converges weakly to a centered* ($E(Y(x)) = 0$) *Gaussian process* Y. *The covariance function* R *of* Y *is given by*

$$\begin{aligned} R(x, y) &= \mathrm{Cov}(Y(x), Y(y)) \\ &= \min(H(x), H(y)) - H(x)H(y) \\ &\quad - b(x)g(y;\theta_0)' - b(y)g(x;\theta_0)' + g(x;\theta_0)Lg(y;\theta_0)'. \end{aligned} \tag{4.2}$$

Proof. Let $Z_n(x) = \sqrt{n}(F_n(x) - H(x))$. Then

$$Y_n(x) = Z_n(x) + \sqrt{n}(F(x;\theta_0) - F(x;\hat{\theta}_n)). \tag{4.3}$$

By mean value theorem,

$$\begin{aligned} F(x;\hat{\theta}_n) - F(x;\theta_0) &= (\hat{\theta}_n - \theta_0)g(x;\lambda)' \\ &= (\hat{\theta}_n - \theta_0)g(x;\theta_0)' + (\hat{\theta}_n - \theta_0)(g(x;\lambda) - g(x;\theta_0))', \end{aligned} \tag{4.4}$$

where λ is a vector lying between θ_0 and $\hat{\theta}_n$. Thus by (A1)–(A4) and the multivariate central limit theorem

$$\sup_x \|g(x;\lambda) - g(x;\theta_0)\| \to_{\mathrm{p}} 0 \quad \text{and} \quad \sqrt{n}(\hat{\theta}_n - \theta_0) \to_{\mathscr{D}} N(0, L)$$

as $\|\lambda - \theta_0\| \leqslant \|\hat{\theta}_n - \theta_0\| \to_{\mathrm{p}} 0$, where $\to_{\mathscr{D}}$ denotes convergence in distribution and $\to_{\mathrm{p}}$ denotes convergence in probability. Hence by (4.3) and (4.4), we have

$$\sup_x |Y_n(x) - Z_n(x) + \sqrt{n}(\hat{\theta}_n - \theta_0)g(x;\theta_0)'| \to_{\mathrm{p}} 0.$$

Consequently by (A2),

$$\sup_x \left| Y_n(x) - Z_n(x) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell(X_i) g(x; \theta_0)' \right| \to_{\mathrm{p}} 0. \tag{4.5}$$

By (4.5), the processes Y_n and W_n have the same weak limit, where

$$W_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (I_{\{X_i \leqslant x\}} - H(x) - \ell(X_i) g(x; \theta_0)').$$

To prove that W_n converges to a Gaussian process, let $U_1, \ldots, U_n$ be i.i.d. uniform random variables on (0,1) and $E_n(t) = (1/\sqrt{n}) \sum_{i=1}^{n} (I_{\{U_i \leqslant t\}} - t)$. As H is continuous, for each n, the processes Z_n and $E_n(H(\cdot))$ both have the same distribution. As the sequence $\{E_n\}$ is tight (Billingsley, 1968), $\{Z_n\}$ is a tight sequence. Since $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is bounded in probability and $g(\cdot; \theta_0)$ by Assumption (A1) is a continuous function, it follows that $\{W_n\}$ is a tight sequence. It is clear by Assumptions (A3), (A4) and by the multivariate central limit theorem that the finite dimensional distributions of $\{W_n\}$ converge to multivariate normal distributions. So it is enough to prove that

$$\mathrm{Cov}(W_n(x), W_n(y)) \to R(x, y). \tag{4.6}$$

We have for $x \leqslant y$

$$\mathrm{Cov}(I_{\{X_i \leqslant x\}}, I_{\{X_i \leqslant y\}}) = H(x)(1 - H(y)). \tag{4.7}$$

By (A3),

$$\begin{aligned} \mathrm{Cov}(W_n(x), W_n(y)) = {} & \mathrm{Cov}(I_{\{X_i \leqslant x\}}, I_{\{X_i \leqslant y\}}) + g(x; \theta_0) L g(y; \theta_0)' \\ & - b(x) g(y; \theta_0)' - b(y) g(x; \theta_0)'. \end{aligned} \tag{4.8}$$

Now convergence (4.6) follows from (4.7) and (4.8). This completes the proof of the theorem. □

Theorem 4.2. *Suppose* (A1) *and* (A3)–(A5) *hold. Then for almost all sample sequences $X_1, \ldots, X_n$, the process Y_n^{b} converges weakly to a centered* ($E(Y(x)) = 0$) *Gaussian process Y. The covariance function R of Y is given by* (4.2).

Proof. The proof is similar to that of Theorem 4.1. As in (4.5), we have by (A5) $\sup_x |Y_n^{\mathrm{b}}(x) - W_n^*(x)| \to 0$ in the bootstrap measure for almost all sample sequences $X_1, \ldots, X_n$, where

$$W_n^*(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(I_{\{X_i^* \leqslant x\}} - F_n(x) - \left(\ell(X_i^*) - \frac{1}{n} \sum_{i=1}^{n} \ell(X_i) \right) g(x; \theta_0)' \right).$$

Now it is a matter of simple algebra using strong law of large numbers and the central limit theorem to show that the process W_n^* converges weakly to a Gaussian process with covariance function R given by (4.2). This completes the proof. □

Acknowledgements

The authors are grateful to an associate editor and to both the referees for their useful comments.

References

Babu, G.J., Bose, A., 1988. Bootstrap confidence intervals. Statist. Probab. Lett. 7, 151–160.

Babu, G.J., Rao, C.R., 2001. Goodness-of-fit tests when parameters are estimated, submitted for publication.

Babu, G.J., Singh, K., 1983. Inference on means using the bootstrap. Ann. Statist. 11, 999–1003.

Babu, G.J., Singh, K., 1984. Asymptotic representations related to jackknifing and bootstrapping L-statistics. Sankhyā Ser. A 46, 195–206.

Billingsley, P., 1968. Convergence of Probability Measures. Wiley, New York.

Darling, D.A., 1955. The Cramér–Smirnov test in the parametric case. Ann. Math. Statist. 26, 1–20.

Durbin, J., 1973. Weak convergence of the sample distribution function when parameters are estimated. Ann. Statist. 1, 279–290.

Foutz, R.V., Srivastava, R.C., 1977. The performance of the likelihood ratio test when the model is incorrect. Ann. Statist. 5, 1183–1194.

Kac, M., Kiefer, J., Wolfowitz, J., 1955. On tests of normality and other tests of goodness of fit based on distance methods. Ann. Math. Statist. 26, 189–211.

Konishi, S., 1999. Statistical model evaluation and information criteria. In: Ghosh, S. (Ed.), Multivariate Analysis, Design of Experiments, and Survey Sampling. Marcel Dekker, New York, pp. 369–399.

Konishi, S., Kitagawa, G., 1996. Generalized information criteria in model selection. Biometrika 83, 875–890.

Nishii, R., 1988. Maximum likelihood principle and model selection when the true model is unspecified. J. Multivariate. Anal. 27, 392–403.

Rao, C.R., Wu, Y., 2001. Model selection. In: Lahiri, P. (Ed.), Model Selection. IMS Lecture Notes-Monograph Series, Vol. 38, pp. 1–18.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1–25.