Sankhyā : The Indian Journal of Statistics
2004, Volume 66, Part 1, pp 63-74

Goodness-of-fit Tests When Parameters are Estimated

G. Jogesh Babu and C.R. Rao
The Pennsylvania State University, University Park, USA

Abstract

Several nonparametric goodness-of-fit tests are based on the empirical distribution function. In the presence of nuisance parameters, the tests are generally constructed by first estimating these nuisance parameters. In such a case, it is well known that critical values shift, and the asymptotic null distribution of the test statistic may depend in a complex way on the unknown parameters. In this paper we use bootstrap methods to estimate the null distribution. We shall consider both parametric and nonparametric bootstrap methods. We shall first demonstrate that, under very general conditions, the process obtained by subtracting the population distribution function with estimated parameters from the empirical distribution has the same weak limit as the corresponding bootstrap version. Of course in the nonparametric bootstrap case a bias correction is needed. This result is used to show that the bootstrap method consistently estimates the null distributions of various goodness-of-fit tests. These results hold not only in the univariate case but also in the multivariate setting.

AMS (2000) *subject classification.* 62F40, 62F12, 62G10.
Keywords and phrases. Brownian Motion, empirical process, bootstrap, Kolmogorov-Smirnov statistic, Cramér-von Mises statistic interquartile range.

1 Introduction

The paper revisits a well discussed problem of goodness-of-fit tests when parameters are estimated. It is customary in certain applied circles to use the goodness-of-fit tests suited for fixed nulls with the same critical values even when the null is composite and contains unknown parameters. It is of course well known that the cutoffs shift. However, it is less well known that unless some special structure is present, the percentiles depend in a complex way on the true value of the parameter. Furthermore, the cutoffs do not seem to have been calculated for important cases, and in most cases, would not be available in closed form. This problem is illustrated by the following example.

Example. Let $X_1, \ldots, X_n$ be i.i.d. random variables from the Uniform distribution $G = G(.;\theta)$ on $(\theta, \theta + 1)$. Let θ be estimated by $\hat{\theta}_n = \overline{X}_n - \frac{1}{2}$, where $\overline{X}_n$ denotes the sample mean of $X_1, \ldots, X_n$. Clearly,

$$F_n(x) - G(x; \hat{\theta}_n) = F_n(x) - G(x; \theta) + \overline{X}_n - 1/2 - \theta,$$

where F_n denotes the empirical distribution function of $X_1, \ldots, X_n$. The corresponding von Mises-type statistic $\hat{\omega}^2$ is given by

$$\begin{aligned}
\hat{\omega}^2 &= n \int_{-\infty}^{\infty} \big(F_n(x) - G(x; \hat{\theta}_n)\big)^2 \, dG(x; \hat{\theta}_n) \\
&= n \int_{\hat{\theta}_n}^{1+\hat{\theta}_n} \big(F_n(x) - G(x; \theta) + \overline{X}_n - 1/2 - \theta\big)^2 dx \\
&= \omega^2 + n\big(\overline{X}_n - 1/2 - \theta\big)^2 \\
&\qquad + n\big(2\overline{X}_n - 1 - 2\theta\big) \int_{\theta}^{1+\theta} \big(F_n(x) - G(x; \theta)\big) dx + \gamma_n \, ,
\end{aligned}$$

where $\mathrm{E}|\gamma_n| \to 0$ as $n \to 0$, and the distribution of von Mises statistic

$$\omega^2 = n \int_{\theta}^{1+\theta} \big(F_n(x) - G(x; \theta)\big)^2 dx$$

is free from G and θ. A simple algebra shows that

$$nE\Big(\big(\overline{X}_n - 1/2 - \theta\big)^2 + \big(2\overline{X}_n - 1 - 2\theta\big) \int_{\theta}^{1+\theta} \big(F_n(x) - G(x; \theta)\big) dx\Big) = -\frac{1}{12}.$$

As a result, the expected values differ by $\frac{1}{12}$ in the limit. Anderson and Darling (1952) obtained the explicit limiting distribution of von Mises ω^2. They also gave a table of the percentiles of the asymptotic distribution of ω^2. Consequently the use of these tables instead of the percentiles of $\hat{\omega}^2$ will lead to substantially incorrect critical values. In view of this, it is important to investigate if a procedure like the bootstrap lead to consistent estimates of the percentiles of the true limiting distribution of the test statistic when the parameters are estimated.

There has been lot of work on goodness-of-fit tests, but we shall not attempt a review here. Asymptotic distribution of test statistics based on the empirical distribution function, when parameters are estimated have been extensively studied by Darling (1955), Kac, Kiefer and Wolfowitz (1955) and others. Extending these results, Durbin (1973) has studied the weak convergence of the sample distribution function to a Gaussian process under a given sequence of contiguous alternative hypotheses when parameters are estimated from the data. However, the limiting process or the asymptotic distribution of statistics based on the empirical process depends in a complex way on the unknown parameter. In this paper, we shall show that the consistency of the bootstrap resampling scheme in estimating the limiting distribution. This will help in obtaining critical values in the testing context.

In the multivariate case the Kolmogorov-Smirnov distance is not distribution free, even when the population distribution is completely specified. Simpson (1951) gave a simple example to illustrate this in the bivariate case. In this paper we study methods of inference using bootstrap for tests based on empirical process, when some parameters are estimated.

2 Bootstrap

Let $\{F(.;\theta) : \theta \in \Theta\}$ denote a family of continuous distribution functions, where Θ is a open region in a p-dimensional Euclidean space. Let $X_1, \ldots, X_n$ be i.i.d. random variables from a distribution F. To test $F = F(.;\theta)$ for some $\theta = \theta_0$ or if θ is partially specified, we consider several tests based on empirical measures. In particular we consider Kolmogorov-Smirnov and Cramér-von Mises statistics, when θ is estimated. These statistics can be viewed as continuous functionals of the empirical process

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

where $\hat{\theta}_n = \theta_n(X_1, \ldots, X_n)$ is an estimator of the true value θ, and F_n denotes the empirical distribution function of $X_1, \ldots, X_n$. We shall develop a bootstrap procedure and show its consistency, when the parameters are (partially or completely) estimated. The procedure will also help in the computation of power under contiguous alternatives $\lambda_n = \theta_0 + n^{-1/2}\lambda$.

To describe the bootstrap procedure, let $X_1^*, \ldots, X_n^*$ be i.i.d. random variables from $\hat{F}_n$, where $\hat{F}_n$ is an estimator of the distribution function F, based on the sample $X_1, \ldots, X_n$. Let $\hat{\theta}_n^* = \theta_n(X_1^*, \ldots, X_n^*)$. For example, if $F(.,\theta)$ denotes the Gaussian distribution function with parameter vector $\theta = (\mu, \sigma^2)$, where μ denotes the mean and σ^2 denotes the variance, then $\hat{\theta}_n$ denotes the vector of sample mean and sample variance. Further, $\hat{\theta}_n^*$ denotes the vector of sample mean and sample variance based on the bootstrap sample $X_1^*, \ldots, X_n^*$. The resampling method is called nonparametric bootstrap if $\hat{F}_n = F_n$, and it is called parametric bootstrap if $\hat{F}_n = F(.;\hat{\theta}_n)$.

Under some regularity conditions both parametric and nonparametric procedures lead to correct asymptotic levels. However, in the case of nonparametric bootstrap we have to make a correction for the bias. Similar situations are encountered in the case of χ^2 type statistics (Babu, 1984) and U-statistics (Arcones and Giné, 1992).

2.1. Assumptions and their validity. Let $\Lambda \subset \Theta$ be the closure of a given neighborhood of a point $\theta_0 \in \Theta$. We use some of the assumptions listed below on the estimators $\hat{\theta}_n$ and $\hat{\theta}_n^*$, where $\ell(.;\theta)$, $\theta \in \Lambda$ is a measurable p-dimensional row vector valued function:

(E) For some $\epsilon_n = \epsilon_n(X_1, \ldots, X_n) \to 0$ in probability,

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

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

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

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

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

Assumption (E) asserts that the estimator $\hat{\theta}_n$ is locally asymptotically linear, and assumption (P) asserts the same for the bootstrapped version, given the original sample. In the case of nonparametric bootstrap, assumption (N) along with its associated centering correction is more appropriate. Assumptions (E) and (P), or (E) and (N) hold, in general, for maximum likelihood estimators, M-estimators and L-statistics.

Babu & Singh (1984) have shown that in the case of the sample median and L-statistics, condition (N) is satisfied. In fact, in the case of the sample median $\hat{\theta}_n$ and nonparametric bootstrap, assumptions (E) and (N) hold under very general conditions, with

$$\ell(x;\theta) = \frac{1}{a_\theta}\left(I_{\{x\le\theta\}} - \frac{1}{2}\right), \tag{2.1}$$

where a_θ denotes the density evaluated at the population median θ (Babu & Singh, 1984, Theorem 5). In the case of L-statistic, $\hat{\theta}_n = \int x\omega(F_n(x))\,dF_n(x)$

and $\theta = \int x\omega(F(x;\theta))dF(x;\theta)$, Babu & Singh (1984, Theorem 4 and P9) have shown that (E) and (N) hold with

$$\ell(y;\theta) = \int_{-\infty}^{\infty} \big(F(x;\theta) - I_{\{y\le x\}}\big)\omega(F(x;\theta))dx,$$

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 F^{-1} of F is continuous at $a_1, \ldots, a_k$.

We now list an additional set of assumptions on ℓ and F used in the main results.

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

(A2) For $\theta \in \Lambda$, $\int \ell(x;\theta)\, dF(x;\theta) = 0$.

(A3) For $\theta \in \Lambda$, $L(\theta) = \int \ell'(x;\theta)\ell(x;\theta)\, dF(x;\theta)$ is a finite non-negative definite matrix.

(A4) As $\gamma \to \infty$,

$$\sup_{\theta\in\Lambda} \int_{\{\|\ell(x;\theta)\|>\gamma\}} \|\ell(x;\theta)\|^2\, dF(x;\theta) \to 0.$$

(A5) For all x, the function $h(x;.)$ defined by

$$h(x;\theta) = \int_{-\infty}^{x} \ell(t;\theta)\, dF(t;\theta)$$

is continuous at θ_0.

2.2. Parametric and nonparametric bootstrap consistency. Suppose (A1) holds and (E), (P) hold for some score function ℓ satisfying (A2)-(A5). Then in the case of parametric bootstrap, it follows by Theorem 4.1, that for almost all sample sequences, the process $Y(., \hat\theta_n^*)$ given by

$$Y_n(x;\hat\theta_n^*) = \sqrt{n}\big(F_n^*(x) - F(x;\hat\theta_n^*)\big),$$

and $Y_n(.;\hat\theta_n)$ converge weakly to the same limiting centered Gaussian process Y, where F_n^* denotes the empirical distribution function of $X_1^*, \ldots, X_n^*$. The covariance function of Y is given by (4.2). Thus both

$$\sqrt{n}\sup_x |F_n(x) - F(x;\hat\theta_n)| \quad \text{and} \quad \sqrt{n}\sup_x |F_n^*(x) - F(x;\hat\theta_n^*)|$$

have the same limiting distribution for almost all sample sequences $X_1, \ldots, X_n$.

However, in the case of nonparametric bootstrap,

$$\sqrt{n}\big(F_n^*(x) - F(x;\hat\theta_n^*)\big) - B_n(x) = \sqrt{n}\big(F_n^*(x) - F_n(x)\big) + \sqrt{n}\big(F(x;\hat\theta_n) - F(x;\hat\theta_n^*)\big), \tag{2.2}$$

where

$$B_n(x) = \sqrt{n}(F_n(x) - F(x;\hat\theta_n)) \tag{2.3}$$

is the known bias term. As above, if (A1) holds, and (E) and (N) hold for some score function ℓ satisfying (A2)-(A5), then in the case of nonparametric bootstrap, we have by Theorem 4.2, that for almost all sample sequences, the bias corrected process $\sqrt{n}\big(F_n^* - F(.;\hat\theta_n^*)\big) - B_n$ and $Y_n(.;\hat\theta_n)$ converge weakly to the same limiting centered Gaussian process Y. Thus, in this case, both

$$\sqrt{n}\sup_x |F_n(x) - F(x;\hat\theta_n)| \quad \text{and} \quad \sup_x |\sqrt{n}\left(F_n^*(x) - F(x;\hat\theta_n^*)\right) - B_n(x)|$$

have the same limiting distribution for almost all sample sequences $X_1, \ldots, X_n$. To reiterate, if the bootstrap sampling is done from $F(.;\hat\theta_n)$, then the bias term disappears.

Additional complications arise for von Mises-type statistics, for example,

$$\int \big(F_n(x) - F(x;\hat\theta_n)\big)^2 dF(x;\theta_0) \quad \text{or} \quad \int \big(F_n(x) - F(x;\hat\theta_n)\big)^2 dF(x;\hat\theta_n).$$

For the second statistic above, the function with respect to which the integral is evaluated is partially estimated. Two approaches are possible here. First, we might compare the statistic based on the empirical process evaluated at the quantiles of the distribution with estimated parameters and the one based on the empirical process evaluated at the theoretical quantiles. Second, we might impose reasonable conditions on the estimators of the parameters such as a bounded variation for the score function. Details in the case of normal family are given in §3.

These ideas can be extended to product-limit estimators.

3 Applications to Location and Scale Invariant Families

Suppose $\{F(.;\theta) : \theta \in \Theta\}$ is a location family of continuous distribution functions, *i.e.*, $F(x;\theta) = F(x - \theta)$ for all x, θ, and suppose $\hat\theta_n$ satisfies

$$\hat\theta_n(X_1 + a, \ldots, X_n + a) = \hat\theta_n(X_1, \ldots, X_n) + a \tag{3.1}$$

for all a. Note that estimators $\hat\theta_n$ such as sample median and sample mean satisfy (3.1). Then clearly, for parametric bootstrap, both the processes

$Y_{(\cdot+\hat{\theta}_n)};\hat{\theta}_n^*)$ and $Y_n(.;\hat{\theta}_n)$ have the same distributions in the function space. Hence, $\sup_x|Y_n(x;\hat{\theta}_n)|$ and $\sup_x|Y_n(x;\hat{\theta}_n^*)|$ have identical distributions, Similar results hold for scale invariant families. Details for normal and Cauchy family of distributions are given below. In particular, these ideas carry over to parametric families described through a transitive group of one-to-one transformations on the sample space.

Let $\Phi_{\mu,\sigma}$ and $\phi_{\mu,\sigma}$ denote, respectively, the cumulative distribution function and the density function of the normal distribution with mean μ and variance σ. Let $\Phi=\Phi_{0,1}$. Let $V_1,\ldots,V_n$ be i.i.d. random variables from normal distribution with parameter vector $\theta=(\mu,\sigma^2)$, where μ denotes the mean and σ denotes the standard deviation. Let $\bar{V}$ and s_V denote the sample mean and standard deviation of the sample $V_1,\ldots,V_n$. Let $F_{n,V}$ and $\hat{F}_{n,V}$ denote the empirical distribution of $V_1,\ldots,V_n$ and $((V_1-\bar{V})/s_V),\ldots,V_n((V_1-\bar{V})/s_V)$ respectively. Clearly

$$\hat{F}_{n,V}(y)-\Phi(y)=F_{n,V}(\bar{V}+ys_V)-\Phi_{\bar{V},s_V}(\bar{V}+ys_V).$$

So,

$$\sup_y|\hat{F}_{n,V}(y)-\Phi(y)|=\sup_x|F_{n,V}(x)-\Phi_{\bar{V},s_V}(x)|.$$

If a further i.i.d. sample $R_1,\ldots,R_n$ is drawn from the normal distribution with mean $\bar{V}$ and variance s_V^2, then also we have

$$\sup_y|\hat{F}_{n,R}(y)-\Phi(y)|=\sup_x|F_{n,R}(x)-\Phi_{\bar{R},s_R}(x)|,$$

where $\bar{R}$, s_R, $F_{n,R}$ denote mean, standard deviation, empirical distribution of $R_1,\ldots,R_n$, and $\hat{F}_{n,R}$ denote the empirical distribution of $((R_1-\bar{R})/s_R),\ldots,(R_n-\bar{R})/s_R))$. Thus from the discussion in §2 2, it is clear that both $\sqrt{n}(F_{n,V}-\Phi_{\bar{V},s_V})$ and $\sqrt{n}(F_{n,R}-\Phi_{\bar{R},s_R})$ both converge weakly to the same Gaussian process. Thus, both $\sup_y\sqrt{n}|\hat{F}_{n,R}(y)-\Phi(y)|$ and $\sup_y\sqrt{n}|F_{n,R}(y)-\Phi(y)|$ have the same limiting distributions.

As $\sup_x\ |\phi_{\bar{R},s_R}(x)-\phi_{\bar{V},s_V}(x)|/\phi_{0,1}(x)\to 0$, it also follows that von Mises-type statistics $n\int(F_{n,V}(x)-\Phi_{\bar{V},s_V}(x))^2\,d\Phi_{\bar{V},s_V}(x)$ and $n\int(F_{n,R}(x)-\Phi_{\bar{R},s_R}(x))^2\,d\Phi_{\bar{R},s_R}(x)$ both converge in distribution to the same limit.

In practice, we can start taking samples of size n from standard normal distribution, and standardize the sample with estimated mean and variance. Then compute the statistic $\sup_y\ |\hat{F}_{n,V}(y)-\Phi(y)|$. Repeat this a large number of times. The histogram of the resultant values approximate the sampling distribution of $|\hat{F}_{n,R}(y)-\Phi(y)|$. Critical values can be computed from this histogram.

Similar analysis goes through for the Cauchy family. Let $V_1, \ldots, V_n$ be i.i.d. random variables from a Cauchy distribution with density $c(x;\theta)$ given by

$$c(x;\theta) = c(x;(\theta_1,\theta_2)) = \frac{1}{\pi}\,\frac{\theta_2}{\theta_2^2 + (y-\theta_1)^2}\,.$$

The sample median is a consistent estimator of the location θ_1. Note that the scale parameter θ_2 is given by half of the difference between the third and first quartile. Thus the sample interquartile range consistently estimates $2\theta_2$.

Let $\tilde{V}$ and $2\hat{V}$ denote, respectively, the sample median and the interquartile range of the sample $V_1, \ldots, V_n$. Let $\hat{F}_{n,V}$ and $F_{n,V}$ denote the empirical distributions of $((V_1 - \bar{V})/\hat{V}), \ldots, ((V_1 - \bar{V})/\hat{V})$ and $V_1, \ldots, V_n$ respectively. Let $C_\theta = C_{\theta_1,\theta_2}$ denote the cumulative distribution function induced by the density $c(.;(\theta_1,\theta_2))$ and $C = C_{0,1}$. That is,

$$C_{\theta_1,\theta_2}(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}((x-\theta_1)/\theta_2)).$$

Clearly

$$\hat{F}_{n,V}(y) - C(y) = F_{n,V}(\tilde{V} + y\hat{V}) - C_{\tilde{V},\hat{V}}(\tilde{V} + y\hat{V}).$$

So,

$$\sup_y |\hat{F}_{n,V}(y) - C(y)| = \sup_x |F_{n,V}(x) - C_{\tilde{V},\hat{V}}(x)|.$$

If a further i.i.d. sample $R_1, \ldots, R_n$ is drawn from the Cauchy distribution with parameters $\tilde{V}$ and $\hat{V}$, then also we have

$$\sup_y |\hat{F}_{n,R}(y) - C(y)| = \sup_x |F_{n,R}(x) - C_{\tilde{R},\hat{R}}(x)|,$$

where $\tilde{R}$, $2\hat{R}$, $F_{n,R}$ denote the sample median, the interquartile range, the empirical distribution of $R_1, \ldots, R_n$, and $\hat{F}_{n,R}$ denotes the empirical distribution of $((R_1 - \tilde{R})/\hat{R}), \ldots, ((R_n - \tilde{R})/\hat{R})$. Thus from the discussion in §2 2, it is clear that both $\sup_y |\hat{F}_{n,V}(y) - C(y)|$ and $\sup_y |\hat{F}_{n,R}(y) - C(y)|$ have the same limiting distributions. Note that the estimators involve only quantiles, as in the case of sample median (see equation (2.1)), the conditions (E), (P), and (A2)-(A5) are clearly satisfied. Thus the critical values can be computed as described in the normal case.

4 Technical Results

Recall that $\{F(.;\theta) : \theta \in \Theta\}$ denotes a family of continuous distribution functions, $\theta_0 \in \Theta$, and $\Lambda \subset \Theta$ is the closure of a given neighborhood of

θ_0, where Θ is a open region in a p-dimensional Euclidean space. Suppose $\{\theta_n\}$ is a sequence in Λ converging to θ_0 as $n \to \infty$. Let $X_{1,n}, \ldots, X_{n,n}$ be i.i.d. random variables from the distribution $F(.;\theta_n)$. Let P_{θ_n} denote the probability measure induced by $X_{1,n}, \ldots, X_{n,n}$ and let F_n denote the empirical distribution of these random variables.

Suppose $\hat{\theta}_n = \hat{\theta}_n(X_{1,n}, \ldots, X_{n,n})$ is an estimator of θ_n. In most practical situations $\hat{\theta}_n$ is a maximum likelihood estimator or an estimator obtained from an estimating equation. As mentioned in §2 2.1, the results also hold for many general estimators $\hat{\theta}_n$ including M-estimators and L-statistics.

REMARK. Even though the estimator $\hat{\theta}_n$ as a function of the observations does not depend on the parameter θ_n, it would be necessary to consider $\sqrt{n}(\hat{\theta}_n - \theta_n)$, which would in general depend on θ_n.

If $\int \|\ell(x;\theta_0)\|^2 \, dF(x;\theta_0)$ is finite, then clearly

$$\int_{\{\|\ell(x;\theta_0)\|>\epsilon\sqrt{n}\}} \|\ell(x;\theta_0)\|^2 \, dF(x;\theta_0) \to 0.$$

Consequently, if $\theta_n \equiv \theta_0$, then assumptions (A4) and (A5) are not required in the proof of Theorem 4.1. Assumption (A4) holds if for some $\delta > 0$,

$$\sup_{\theta\in\Lambda} \int \|\ell(x;\theta)\|^{2+\delta} \, dF(x;\theta) < \infty.$$

Assumption (A5) holds if (A4) holds and if

$$\{t \in (0,1) : \ell(\xi_n(t);\theta_n) \to \ell(\xi_0(t);\theta_0)\}$$

has Lebesgue measure 1 for every sequence $\theta_n \to \theta_0$, where $\xi_n(t)$ denotes the t-th quantile of $F(.;\theta_n)$, $n = 0, 1, 2, \ldots$ This can be seen easily by using results on uniform integrability (Theorem 16.14 of Billingsley 1995).

We now state the main technical result.

THEOREM 4.1 *Suppose $\theta_n \to \theta_0$, (A1) holds, and*

$$\hat{\theta}_n - \theta_n = \frac{1}{n} \sum_{i=1}^{n} \ell(X_{i,n};\theta_n) + \frac{1}{\sqrt{n}} \epsilon_n, \tag{4.1}$$

for a score function ℓ satisfying (A2)-(A5), where $\epsilon_n \to 0$ in P_{θ_n}-probability. If $L(\theta_n) \to L(\theta_0)$, then the process Y_n given by

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

converges weakly to a centered ($E(Y(x)) = 0$) Gaussian process Y. The covariance function R of Y is given by

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

PROOF. Let $Z_n(x) = \sqrt{n}\big(F_n(x) - F(x;\theta_n)\big)$. Then

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

By mean value theorem,

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

where λ is a vector lying between θ_n and $\hat{\theta}_n$. Thus by (4.1), (A1)-(A4) and the multivariate central limit theorem

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

as $\|\lambda - \theta_0\| \le \|\lambda - \theta_n\| + \|\theta_n - \theta_0\| \le \|\hat{\theta}_n - \theta_n\| + \|\theta_n - \theta_0\| \to_p 0$, where $\to_{\mathcal{D}}$ denotes convergence in distribution. Hence by (4.3) and (4.4), we have

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

Consequently by (4.1),

$$\sup_x \Big|Y_n(x) - Z_n(x) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell(X_{i,n};\theta_n) g(x;\theta_0)'\Big| \to_p 0. \tag{4.5}$$

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

$$W_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(I_{\{X_{i,n} \le x\}} - F(x;\theta_n) - \ell(X_{i,n};\theta_n)g(x;\theta_0)'\Big).$$

Measurability issues are not discussed here; see Pollard (1984, pp 155-159) for issues associated with measurability.

To prove that W_n converges to a Gaussian process, let $U_1, \ldots, U_n$ be i.i.d. uniform random variables on $(0,1)$. Let $E_n(t) = n^{-1/2} \sum_{i=1}^{n} \Big(I_{\{U_i \le t\}} - t\Big)$.

Clearly, for each n, the processes Z_n and $E_n(F(.;\theta_n))$ both have the same distribution. As the sequence $\{E_n\}$ is tight (Billingsley, 1968), $\{Z_n\}$ is a tight sequence. Since $\sqrt{n}(\hat{\theta}_n - \theta_n)$ is bounded in probability and $g(.;\theta_0)$, by (A1), is a continuous function, it follows that $\{W_n\}$ is a tight sequence. It is clear by assumptions (A2)-(A4) and 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

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

Now by (A1) we have for $x \le y$

$$\mathrm{Cov}\,(I_{\{X_{i,n}\le x\}}, I_{\{X_{i,n}\le y\}}) = F(x;\theta_n)(1 - F(y;\theta_n)) \to F(x;\theta_0)(1 - F(y;\theta_0)). \tag{4.7}$$

By (A2),

$$\begin{aligned}\mathrm{Cov}\,(W_n(x), W_n(y)) &= \mathrm{Cov}\,(I_{\{X_{i,n}\le x\}}, I_{\{X_{i,n}\le y\}}) + g(x;\theta_0)L(\theta_n)g(y;\theta_0)' \\ &\quad -h(x;\theta_n)g(y;\theta_0)' - h(y;\theta_n)g(x;\theta_0)'. \end{aligned} \tag{4.8}$$

Since $L(\theta_n) \to L(\theta_0)$, the convergence (4.6) follows from (A5), (4.7) and (4.8). This completes the proof of the theorem.

REMARK. Though Theorem 4.1 can be derived from Theorem 19.23 of Van der Vaart (1998), the direct proof given above is essentially self contained, and easily adaptable to obtain the bootstrap version to establish bootstrap consistency.

THEOREM 4.2 *Suppose* (A1) *holds, and* (E), (N) *hold for some score function ℓ satisfying* (A2)-(A5). *If L is continuous at θ_0, then for almost all sample sequences $X_1, \ldots, X_n$, the bias corrected process Y_n^b given by*

$$Y_n^b(x) = \sqrt{n}\big(F_n^*(x) - F(x;\hat{\theta}_n^*)\big) - B_n(x)$$

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 (N) $\sup_x |Y_n^b(x) - W_n^*(x)| \to 0$ in the bootstrap measure for almost all sample sequences $X_1, \ldots, X_n$, where

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

Now it is a matter of simple algebra using strong law of large numbers 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.

Acknowledgments. We are deeply grateful to the referee for valuable suggestions and comments. The work of G.J. Babu is supported in part by NSF grant DMS-0101360 and the work of C. R. Rao is supported in part by ARO grant 19-01-1-0563.

References

Anderson, T.W. and Darling, D.A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statist.*, **23**, 193-212.

Arcones, M.A. and Giné, E. (1992). On the bootstrap of U and V statistics. *Ann. Statist.*, **20**, 655–674.

Babu, G.J. (1984). Bootstrapping statistics with linear combinations of chi-squares as weak limit. *Sankhyā Ser. A*, **46**, 85-93.

Babu, G.J. and 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.

Billingsley, P. (1995). *Probability and Measure*, 3rd Edition. 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.

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

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag, New York.

Simpson, P.B. (1951). Note on the estimation of a bivariate distribution function. *Ann. Math. Statist.*, **22**, 476-478.

van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge, UK.

G. Jogesh Babu and C.R. Rao
326 Thomas Building
Department of Statistics
The Pennsylvania State University
University Park, PA 16802
E-mail: babu@stat.psu.edu
crrl@psu.edu

Paper received: February 2002; revised August 2003.