Bootstrap for nonstandard cases

Gutti Jogesh Babu

Department of Statistics, 319 Classroom building, Pennsylvania State University, University Park, PA 16802, USA

Abstract

Scale invariant statistics like Student's t, involving mean absolute deviations instead of standard deviations, and statistics which are distributed asymptotically like linear combinations of chi-squares or F-statistics are considered. Bootstrap procedures in estimating their distributions are described. The error in the bootstrap approximation of the sampling distribution, in the latter type, is of the order $O(n^{-3/2})$. It is demonstrated that the error term cannot be improved. In such cases Bartlett correction is preferable, which is known to give an approximation with an error term of the order $O(n^{-2})$. Bootstrap procedures for errors-in-variables regression coefficients are also described. In all the cases mentioned above, the Edgeworth expansions of multivariate means, which do not satisfy standard regularity assumptions, play an important role.

Key words: Edgeworth expansions; Chi-square; F-distribution; L1-estimator; Errors-in-variable regression; Bootstrap

1. Introduction

Edgeworth expansions are known to play a crucial role in the study of bootstrap procedures. Using these expansions, Babu and Singh (1983, 1984), Babu (1984), Babu and Bose (1988), Singh and Babu (1990), Bhattacharya and Qumsiyeh (1989), Bose (1988), and Singh (1981) have established the superiority of the bootstrap methods, in approximating the sampling distribution, for a wide class of statistics. This is achieved by showing that the bootstrap procedure automatically corrects for skewness factor. This class includes statistics $\hat{\theta}_n$ which can be written as a smooth function $H$ of a sample mean $\bar{Z}_n$ of $n$ independent multivariate random variables $Z_1, \ldots, Z_n$, i.e., $\hat{\theta}_n = H(\bar{Z}_n)$. Under very general conditions the sampling distributions of the
studentized version of $H(\tilde{Z}_n) - H(E(Z_1))$ can be approximated closely by its bootstrap version.

Besides regular assumptions on the moments, several of these results are established under

(a) The distribution of $Z_1$ is strongly non-lattice, (i.e., the distribution of $Z_1$ is not concentrated on countably many parallel hyperplanes).

(b) When properly normalized, $\tilde{\theta}_n$ has a normal distribution in the limit.

(c) The random variables $Z_1, \ldots, Z_n$ are independent and identically distributed.

In general, condition (a) is violated, if a statistic is normalized using a robust $L_1$-estimate of scale rather than the traditional $L_2$-estimate like sample standard deviation. See Babu (1991) and Babu and Rao (1991). Condition (b) is not satisfied when the limiting distribution is like that of an F-statistic or like a linear combination of chi-squares. See Babu (1984) for initial work. Finally condition (c) does not hold in the context of errors-in-variables regression. See Babu and Bai (1992).

Bootstrap approximations of the sampling distributions are described in the next three sections, when one of the above three conditions is violated. The contents of Section 3 are new.

2. Mean absolute deviations

Recent research on robust statistical inference has motivated the development of statistical methodology based on the $L_1$-norm instead of the traditional methods based on the $L_2$-norm or least squares. However, from time to time attempts have been made to use the $L_1$-norm in estimation and tests of significance, but the complexity of the distributions involved stood in the way of their use in practical applications. Let $X_1, \ldots, X_n$ be independent random variables with a common continuous distribution $F$ having variance $\sigma^2$ and mean $\mu$. The mean absolute deviation

$$M_n = \frac{1}{n} \sum_{i=1}^{n} |X_i - \bar{X}_n|,$$

was considered as an estimate of scale parameter as an alternative to the root mean square deviation

$$s_n = \left( \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \right)^{1/2},$$

where $\bar{X}_n$ denotes the sample mean of $W_1, \ldots, W_n$. Herrey (1965) derived the distribution of a robust Student’s $t$-type statistic,

$$H_n = \sqrt{n(\bar{X}_n - \mu)}/M_n,$$
under normal assumptions using the independence of $\bar{X}_n$ and $M_n$. If $F$ is not Gaussian, then $\bar{X}_n$ and $M_n$ are no longer independent.

Suppose $F$ is differentiable in a neighborhood of $\mu$ and the derivative $f$ of $F$ at $\mu$ is positive. If for some $a > 0$ and $b > 0$,

$$|f(x) - f(\mu)| \leq a|x - \mu|^b,$$

for $x$ near $\mu$, then it can be shown that $H_n = S_n + R_n$, where $Y_i = \sigma^{-1}(X_i - \mu)$, $\gamma^{-1} = E[Y_1]$, $S_n = \sqrt{n\gamma} \tilde{Y}_n (1 - \tilde{U}_n + \tilde{V}_n \tilde{Y}_n + \tilde{U}_n^2)$, $U_i = \gamma(|Y_i| - \gamma^{-1} + (2F(\mu) - 1)Y_i)$, $V_i = \gamma(2F(\mu) - 2I(Y_i \leq 0) - \sigma f(\mu)Y_i)$, and for some $K > 0$ and $\varepsilon > 0$,

$$P(|R_n| > Kn^{-1-\varepsilon}) = o(n^{-1}).$$

Clearly, $S_n = \sqrt{n}(H(\bar{Z}_n) - H(E(Z_1)))$, where $Z_i = (Y_i, |Y_i| - \gamma^{-1}, I(Y_i \leq 0) - F(\mu))$, and $H$ is a smooth function given by

$$H(x, y, z) = \gamma^2(1 - xy - x\gamma(2F(\mu) - 1) - y(2z + x^2\sigma f(\mu)) + \gamma^2(y + x(2F(\mu) - 1))^2).$$

Note that $E(Z_1) = (0, 0, 0)$ and the last component of $Z_1$ is a lattice variable. The standard results on Edgeworth expansions are not applicable. The parameter $\gamma^{-1}$ is called the shape parameter and it is generally known. For example, its value for the normal family is $2\sqrt{\pi}$ and $1/\sqrt{2}$ for the double exponential family.

Using results of Babu and Singh (1989), one can establish that

$$P(H_n \leq x\gamma) = \Phi(x) + n^{-1/2} p(x) \phi(x) + o(n^{-1/2}), \quad (2.1)$$

where $p$ is a second degree even polynomial. If $X_1^* \ldots, X_n^*$ is a simple random sample from $X_1, \ldots, X_n$, the the bootstrapped version $H_n^*/\gamma_n$ of $H_n/\gamma$ can be expressed as

$$H_n^*/\gamma_n = \sqrt{n}(\bar{X}_n^* - \bar{X}_n) \left(\frac{\gamma_n}{n} \sum_{i=1}^n |X_i^* - \bar{X}_n^*|\right)^{-1},$$

where $\gamma_n = M_n/s_n$. Following Babu and Singh (1984 and 1989), it can be shown that

$$P^*(H_n^* \leq x\gamma_n) = \Phi(x) + n^{-1/2} p^*(x) \phi(x) + o(n^{-1/2}), \quad (2.2)$$

for almost all sample sequences, where $P^*$ denotes the measure induced by the bootstrap sampling. Further the coefficients of the polynomial $p^*$ converge to those of $p$. The Eqs. (2.1) and (2.2) together imply

$$\sup_x \sqrt{n}|P^*(H_n^* \leq x\gamma_n) - P(H_n \leq x\gamma)| \to 0.$$
for almost all samples. For higher order Edgeworth expansions of $H_n$ see Babu (1991). Instead of $M_1$ one can normalize using

$$M_2 = \frac{1}{n} \sum_{i=1}^{n} |X_i - \widetilde{X}|,$$

where $\widetilde{X}$ is the sample median. See Babu and Rao (1992) for details and results on other related statistics.

3. Chi-square type statistics

Let $X_i$ and $X_i^*$ be as in the previous section. If $X_1$ has the standard normal distribution, it can be shown that

$$P^*(\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \leq x s_n) = \Phi(x) - \Phi(x)(\frac{1}{6} \tau_n n^{-1/2}(x^2 - 1))$$

$$+ (1/72n) \tau_n^2(x^5 - 10x^3 + 15x)$$

$$+ (1/24n) \kappa_n(x^3 - 3x)$$

$$+ n^{-3/2} r_n(x)) + o(n^{-3/2})$$

uniformly in $x$ for almost all samples, where $r_n$ is a polynomial in $x^2$,

$$\tau_n = \frac{1}{n s^3} \sum_{i=1}^{n} (X_i - \bar{X}_n)^3 \quad \text{and} \quad \kappa_n = \frac{1}{n s_n^4} \sum_{i=1}^{n} (X_i - \bar{X}_n)^4 - 3.$$ 

As $\tau_n = O_p(n^{-1/2})$, simple algebra leads to

$$\kappa_n = \frac{1}{n s_n^4} \sum_{i=1}^{n} (X_i^4 - 3) - 6(X_i^2 - 1) + o_p(n^{-1/2}).$$

As $n\bar{X}_n^2$ has $\chi_1^2$ distribution, it follows that for almost all samples and for all $y$,

$$n^{3/2}(P(\chi_1^2 \leq y) - P^*(n(\bar{X}_n^* - \bar{X}_n)^2 \leq y s_n^2)) \quad \text{tends weakly to} \quad N_1(y), \quad (3.1)$$

where $N_1(y)$ denotes the centered normal variable with variance $\phi^2(\sqrt{y})(y - 3)^2 y/6$. Note that $s_n^{-2}E((\bar{X}_n^* - \bar{X}_n)^2) = 1$. So there is no need for Bartlett correction for the bootstrapped statistic. Another example is the Student's $t$-statistic. Using Edgeworth expansions for bootstrapped statistics as in Babu and Singh (1984), it can be shown that

$$n^{3/2}(P^*(n(\bar{X}_n^* - \bar{X}_n)^2 \leq y s_n^2)) - P(n\bar{X}_n^2 \leq y s_n^2))$$

$$= - \phi(\sqrt{y})\sqrt{n y} \left(\frac{1}{6} \tau_n^2 (y^2 + 2y - 3) - \frac{1}{6} \kappa_n (y - 3)\right) + o_p(1). \quad (3.2)$$

As $\tau_n = O_p(n^{-1/2})$ and $\sqrt{n} \kappa_n$ tends to the centered normal distribution with variance 24. It follows that the l.h.s. of (3.2) tends weakly to a centered normal distribution with
variance \( \frac{2}{3} y(y - 3)^2 \phi^2(\sqrt{y}) \). On the other hand \((n - 1) \frac{\bar{X}_n^2}{s_n^2} \) has \( F \) distribution with 1 and \( n - 1 \) degrees of freedom.

Babu (1984) obtained bootstrap approximations for statistics which are asymptotically like linear combinations of chi-squares. The results can be improved if there is additional information on the statistics as suggested by the above examples.

Suppose \( H \) is a smooth function whose first order derivatives vanish at zero. Let \( Z_1, \ldots, Z_n \) be \( m \)-variate i.i.d. random vectors with mean vector \( \mu \) and dispersion matrix \( \Sigma \). If the matrix \( L \) of second order derivatives of \( H \) at 0 is a non-null diagonal matrix, then Edgeworth expansions for \( \tilde{\theta}_n = H((\bar{Z}_n - \mu)\Sigma^{-1/2}) \) with leading term like a linear combination of chi-squares, can be established. Further the coefficient of \( n^{-1/2} \) can be shown to vanish. Similar result for the bootstrapped statistic \( \hat{\theta}_n^* = H((\bar{Z}_n^* - \bar{Z}_n)\Sigma_n^{-1/2}) \) can be established using Theorem 2 of Babu and Singh (1984), where \( Z_1^*, \ldots, Z_n^* \) are sampled with replacement from \( Z_1, \ldots, Z_n \) and \( \Sigma_n \) is the sample dispersion matrix of \( Z_1, \ldots, Z_n \). It follows from these arguments that for almost all samples and for all \( y \),

\[
n^{3/2} (P(\tilde{\theta}_n \leq y) - P(n(\hat{\theta}_n \leq y)) \text{ tends weakly to } N_2(q, y),
\]

where \( q \) is a non-null polynomial in most cases and \( N_2(q, y) \) is a centered normal with variance \( q(y)\phi^2(\sqrt{y}) \).

Using essentially the same arguments, similar results can be obtained for \( F \)-type distributions associated with analysis of one way-classified data. However, it is well known that Bartlett's correction gives a better approximation. See Bickel and Ghosh (1990).

4. Errors-in-variables regression

Consider the simple linear errors-in-variables model \((X_i, Y_i)\):

\[
X_i = u_i + \delta_i, \quad Y_i = \alpha + \beta u_i + \epsilon_i,
\]

where \((\delta_i, \epsilon_i)\) are independent mean zero random variables and \( u_i \) are unknown nuisance parameters. Let \((\delta, \epsilon)\) be independent copies of \((\delta, \epsilon)\) and \( \sigma_{\delta} \) and \( \sigma_{\epsilon} \) respectively denote the standard deviations of \( \delta \) and \( \epsilon \). Further let \( \lambda = \sigma_{\epsilon}^2 / \sigma_{\delta}^2 \). The errors-in-variables models have been studied extensively in the literature. See Fuller (1987), Gleser (1985) and Jones (1979) among others. It is well known that the least squares estimators of \( \beta \) and \( \alpha \) are given by

\[
\hat{\beta}_1 = h + \text{sign}(S_{XY})(h + h^2)^{1/2} \quad \text{and} \quad \hat{\alpha}_1 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n,
\]

when \( \lambda \) is known, where \( h = (S_{XY} - \lambda S_{XX})/2S_{XY} \).

\[
S_{XY} = \sum_{i=1}^{n} (X_i - \bar{X}_n)(Y_i - \bar{Y}_n),
\]

where \( (\delta_i, \epsilon_i) \) are independent mean zero random variables and \( u_i \) are unknown nuisance parameters.
and

\[ S_{XX} = \sum_{i=1}^{n} (X_i - \bar{X}_n)^2, \quad S_{YY} = \sum_{i=1}^{n} (Y_i - \bar{Y}_n)^2. \]  

(4.4)

The least squares method gives the same estimates as in (4.2), when both \( \sigma_\beta \) and \( \sigma_\varepsilon \) are known. Instead, if \( \sigma_\varepsilon \) alone is known, \( S_{XX} > n\sigma_\varepsilon^2 \) and \( S_{YY}(S_{XX} - n\sigma_\varepsilon^2) > S_{XY}^2 \), then the estimators of \( \beta \) and \( \varepsilon \) are given by

\[ \hat{\beta}_2 = \frac{S_{XY}}{(S_{XX} - n\sigma_\varepsilon^2)}, \quad \text{and} \quad \hat{\varepsilon}_2 = \bar{Y}_n - \hat{\beta}_2 \bar{X}_n. \]  

(4.5)

On the other hand, if \( \sigma_\beta \) alone is known, \( S_{YY} > n\sigma_\beta^2 \) and \( S_{XX}(S_{YY} - n\sigma_\beta^2) > S_{XY}^2 \), then the estimators of \( \beta \) and \( \varepsilon \) are given by

\[ \hat{\beta}_3 = \frac{(S_{YY} - n\sigma_\beta^2)}{S_{XY}}, \quad \text{and} \quad \hat{\varepsilon}_3 = \bar{Y}_n - \hat{\beta}_3 \bar{X}_n. \]  

(4.6)

It is not difficult to see that \( \hat{\beta}_r - \beta, r = 1, 2, 3 \), can be written as smooth functions of the average of

\[ \xi_{jn} = (\varepsilon_j^2 - E(\varepsilon_j^2), \delta_j^2 - E(\delta_j^2), \varepsilon_j \delta_j, \varepsilon_j, \delta_j, u_j \varepsilon_j, u_j \delta_j). \]

Edgeworth expansions for \( \xi_{jn} \) lead to those of \( \hat{\beta}_r \). Standard results on Edgeworth expansions are not applicable for two reasons. The first one being that \( \xi_{jn} \) are not identically distributed and the second one being that components of \( \xi_{jn} \) are linearly dependent. But on the average \( \xi_{jn} \) behave very well, provided for some \( v \geq 3, \)

\[ \sum u_{jn} = 0, \quad \frac{1}{n} \sum u_{jn}^2 \rightarrow \eta > 0 \quad \text{and} \quad \sup_n \frac{1}{n} \sum |u_{jn}|^v < \infty. \]  

(4.7)

Babu and Bai (1992) have shown that if \( \varepsilon \) and \( \delta \) are independent continuous centered random variables with finite sixth moments, and if (4.7) holds with \( v = 3 \), then \( \sqrt{n}(\hat{\beta}_r - \beta) \) has valid two-term Edgeworth expansion, for \( r = 1, 2, 3 \). In fact, the independence of \( \varepsilon \) and \( \delta \) is not required but very weak continuity assumptions on the conditional distributions \( \varepsilon \) and \( \delta \) are enough.

If (4.7) holds with \( v > 3 \), then higher order Edgeworth expansions can be obtained. Edgeworth expansions for studentized versions are needed to show the superiority of bootstrap approximation. Babu and Bai (1992) have obtained such Edgeworth expansions for studentized \( \sqrt{n}(\hat{\beta}_r - \beta)/\delta_r, r = 1, 2, 3 \), and their bootstrapped versions. The expressions for estimators \( \delta_r \) of the standard deviations of \( \sqrt{n}\hat{\beta}_r \) are obtained using jackknife type arguments and are given by,

\[ \delta_1^2 = n\hat{\beta}_1^2 \psi \sum_{i=1}^{n} ((Y_i - \bar{Y}_n)^2 - \lambda(X_i - \bar{X}_n)^2 - 2\xi(X_i - \bar{X}_n)(Y_i - \bar{Y}_n))^2, \]

\[ \delta_2^2 = n(\hat{\beta}_2/S_{XY})^2 \sum_{i=1}^{n} ((X_i - \bar{X}_n)(Y_i - \bar{Y}_n - \hat{\beta}_2(X_i - \bar{X}_n)) + \hat{\beta}_2 \sigma_\varepsilon^2)^2. \]
\[ \sigma^2 = nS_{XY}^2 \sum_{i=1}^{n} ((Y_i - \bar{Y}_n)(Y_i - \bar{Y}_n - \hat{\beta}_3(X_i - \bar{X}_n)) - \sigma^2)^2, \]

where \( \psi^{-1} = 4S_{XY}^2(h^2 + \lambda) \).

Bootstrap for the studentized case, corrects for the skewness of the sampling distribution. This result is established by Babu and Bai (1992) by showing that for \( r = 1, 2, 3 \),

\[ \sqrt{n} \sup_x |P(\sqrt{n}(\hat{\beta}_r - \beta) \leq x\hat{\sigma}_r) - P^*(\sqrt{n}(\hat{\beta}_r - \beta) \leq x\hat{\sigma}_r^*)| \to 0. \]

for almost all sample sequences, where \( P^* \) denotes the probability distribution induced by the bootstrap sampling, and \( \hat{\beta}_r^* \) and \( \hat{\sigma}_r^* \) denote the bootstrap estimates of the slope and the standard deviation. That is, \( \hat{\beta}_r^* \) and \( \hat{\sigma}_r^* \) are obtained by replacing \((X_i, Y_i)\) by the bootstrap samples \((X_i^*, Y_i^*)\), in the expressions for \( \hat{\beta}_r \) and \( \hat{\sigma}_r \).

References


