Probab. Th. Fields 81, 247–273 (1989)

Probability Theory and Related Fields

Unusual Properties of Bootstrap Confidence Intervals in Regression Problems

Peter Hall

Department of Statistics, Australian National University, G.P.O. Box 4, Canberra ACT 2001, Australia

Summary. We show that the percentile-t method, and one of the two percentile methods, have unusually good performance when employed to construct bootstrap confidence intervals in a regression setting. In the case of slope parameters, percentile-t produces two-sided intervals with coverage error n^{-2}, and one-sided intervals with coverage error $n^{-\frac{3}{2}}$, where n is sample size. The errors are only n^{-1} in most other problems. One of the percentile methods produces critical points which are third-order correct for Efron's [11] relatively complex accelerated bias-corrected points.

1. Introduction

1.1. Aims and principal results

In this paper we show that when the bootstrap is used to construct confidence intervals in a regression problem, it has several remarkable properties which are not seen in other applications of the bootstrap. To explain these properties, let us focus attention on the simple linear regression model,

$$Y_i = \alpha + x_i \beta + \varepsilon_i, \qquad 1 \leqq i \leqq n, \tag{1.1}$$

where the ε_i's are independent and identically distributed errors with zero mean and variance σ^2. Only the pairs (x_i, Y_i) are observable, and we wish to construct "nonparametric" confidence intervals for the slope parameter β, or for the mean $\alpha + x_0 \beta$ of Y given that $x = x_0$.

Our main conclusions are listed below. They hold true under moment conditions on the design sequence $\{x_i\}$, and moment and smoothness conditions on the error distribution. In results (i)–(vi) below we assume that slope, β, is the parameter of interest, and that the design variables x_i are fixed (or at least, conditioned upon). When considering these points the reader might like to bear in mind that asymptotic confidence bounds based on the central limit theorem have coverage error $O(n^{-\frac{1}{2}})$ in the one-sided case and $O(n^{-1})$ in the case of two-sided intervals.

(i) The percentile-t method for constructing two-sided confidence intervals for β results in confidence intervals whose coverage error is $O(n^{-2})$ as $n \to \infty$. This is most unusual, since coverage error is generally of size n^{-1} when the bootstrap (percentile-t or otherwise) is used to set two-sided intervals.

(ii) Unusually high coverage accuracy is also available for one-sided intervals. There, percentile-t intervals have coverage error $O(n^{-\frac{3}{2}})$, compared to an error of size n^{-1} in most other statistical problems.

(iii) This exemplary performance of percentile-t is not achieved at the expense of inaccuracy in position of critical points. Those points retain their property of second-order correctness noted in other statistical problems – see e.g. [11]. [A critical point is second-order correct if it is correct to order $(n^{-\frac{1}{2}})^2 = n^{-1}$; see Subsection 1.2 for a more detailed definition.]

(iv) In most statistical problems, both versions of the bootstrap percentile method (different from percentile-t) fail to be second-order correct. They usually have coverage error of size $n^{-\frac{1}{2}}$ when used to construct one-sided confidence intervals, and in this sense do little better than the simple normal approximation. However in a regression setting, when β is the parameter of interest, one of the two percentile methods is always second-order correct, and yields coverage errors of $O(n^{-1})$, for both one- and two-sided confidence intervals. The other percentile method is second-order correct when design points are chosen symmetrically – for example it is second-order correct in the case of regularly spaced designs. These results are particularly important in multiple or multivariate regression, for there the percentile method can be considerably less numerically expensive than percentile-t.

(v) That percentile method which is always second-order correct (see (iv) above) is third-order equivalent to Efron's [11] accelerated bias corrected method. This phenomenon is hardly ever observed in other statistical problems. The accelerated bias-corrected method yields one-sided and two-sided confidence intervals with coverage error $O(n^{-1})$, combared with $O(n^{-\frac{3}{2}})$ and $O(n^{-2})$ respectively in the case of percentile-t. See (i) and (ii) above.

(vi) Properties (i)–(iv) continue to hold for slope parameters in multiple regression and multivariate regression. (We exclude property (v) from this statement because accelerated bias correction is difficult to use in a general multivariate setting.) They also hold in the so-called random design model, but not in the correlation model.

(vii) These unusual properties evaporate if the intercept, α, or the conditional mean, $\alpha + x_0\beta$, is the object of interest. In that circumstance the bootstrap behaves very much as it would in other statistical problems. For example, percentile-t confidence intervals have coverage error of size n^{-1}.

All our conclusions apply to both "parametric" and "nonparametric" forms of the bootstrap, although we shall concentrate all our discussion on the "nonparametric" case, which we regard as the more important of the two. The "nonparametric" bootstrap makes no assumptions about the distribution of the errors ε_i, except that they have zero mean, sufficiently many finite moments, and a nonsingular distribution.

The reason for outstanding performance of the percentile-t bootstrap when estimating slope, is the high degree of symmetry conferred by presence of the design variables x_i. In the equivalent model $Y_i=(\alpha+\bar{x}\beta)+(x_i-\bar{x})\beta+\varepsilon_i$, slope is multiplied by the factor $x_i-\bar{x}$ which adds to zero. This is just enough to eliminate several crucial error terms which render the bootstrap relatively inaccurate in other problems.

Section 2 describes bootstrap confidence intervals for slope, in the case of the simple linear model. These results are given under explicit regularity conditions. Section 3 gives a similar treatment of bootstrap confidence intervals for intercepts and conditional means. The multivariate, multiparameter case is treated very briefly in Sect. 4, reaching the same conclusions as in the simple linear case. Finally, Sect. 5 gives a detailed proof of one of the results from Sect. 2.

Notable recent work on the bootstrap in regression includes Freedman [12], Bickel and Freedman [8] and Freedman and Peters [13]. In Freedman's [12] seminal paper, the focus is on consistency of bootstrap estimates under a wide variety of regression models. See also Hinkley [16] and Bickel and Freedman [8]. However, our emphasis on explicit calculation of coverage error, and on error in position of critical points, rather than simply consistency, makes our contributions closer to those of Bickel and Freedman [7], Singh [18], Beran [2] and Hall [15] in non-regression contexts. These authors discuss issues such as second-order correctness.

Beran [1] gives a very accessible introduction to theory for bootstrap methods, Hinkley and Wei [17] describe advantages of Studentizing (although from a viewpoint different from our own), Beran and Miller [3] discuss general confidence regions for a vector parameter, and Wu [20] gives a detailed treatment of the percentile method (not percentile-t) in regression problems.

In the remainder of the present section we introduce commonly-used versions of the bootstrap. By way of notation, π denotes a probability level, so that $0<\pi<1$. Standard normal distribution and density functions are denoted by Φ and ϕ, respectively, and z_π is the solution of $\Phi(z_\pi)=\pi$.

1.2. Nonparametric Bootstrap

Here we describe bootstrap methods for constructing confidence intervals in the case of the simple linear regression model (1.1). The errors ε_i are assumed to be independent and identically distributed with zero mean and variance σ^2.

Put $\bar{x}=n^{-1}\sum x_i$, $\bar{Y}=n^{-1}\sum Y_i$ and $\sigma_x^2=n^{-1}\sum(x_i-\bar{x})^2$. Least-squares estimates of α and β based on the sample $\mathscr{X}=\{(x_i, Y_i), 1\leqq i\leqq n\}$ are

$$\hat{\alpha}=\bar{Y}-\bar{x}\hat{\beta}, \qquad \hat{\beta}=\sigma_x^{-2}n^{-1}\sum_{i=1}^{n}(x_i-\bar{x})(Y_i-\bar{Y}).$$

The residuals are $\hat{\varepsilon}_i=Y_i-(\hat{\alpha}+x_i\hat{\beta})$, $1\leqq i\leqq n$, and the residual-based estimate of σ^2 is

$$\hat{\sigma}^2=n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i^2.$$

Observe that $E(\hat{\beta})=\beta$ and $\operatorname{var}(\hat{\beta})=\sigma^2/n\sigma_x^2$. Therefore two standardized versions of $\hat{\beta}$ are

$$S_{n1}=n^{\frac{1}{2}}\sigma_x(\hat{\beta}-\beta)/\sigma, \qquad S_{n2}=n^{\frac{1}{2}}\sigma_x(\hat{\beta}-\beta)/\hat{\sigma},$$

the second being "Studentized".

Let $\{\varepsilon_1^*, \ldots, \varepsilon_n^*\}$ denote a random n-sample of the residuals $\{\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n\}$, drawn with replacement. Define

$$Y_i^*=\hat{\alpha}+x_i\hat{\beta}+\hat{\varepsilon}_i^*, \qquad 1\leqq i\leqq n,$$

and let $\hat{\alpha}^*$, $\hat{\beta}^*$ be least-squares estimates of $\hat{\alpha}$, $\hat{\beta}$ computed from the data set $\mathscr{X}^*=\{(x_i, Y_i^*),\ 1\leqq i\leqq n\}$. The two percentile methods (called percentile(I) and percentile(II) below) of constructing a π-level confidence region $\hat{\mathscr{R}}_\pi$ $(0<\pi<1)$ for β, are as follows.

(I) Carry out B independent resampling experiments of the type just described, and define $\hat{\mathscr{R}}_\pi$ to be a regular region within which precisely πB of the B values of $\hat{\beta}^*$ lie.

(II) Let $\hat{\mathscr{S}}_\pi$ denote a regular region containing just πB of the B values of $\hat{\beta}^*-\hat{\beta}$, and put $\hat{\mathscr{R}}_\pi=\hat{\beta}-\hat{\mathscr{S}}_\pi=\{\hat{\beta}-x\colon x\in\hat{\mathscr{S}}_\pi\}$. The latter method is based on the supposition that $\hat{\beta}^*-\hat{\beta}$ has a distribution which closely approximates that of $\hat{\beta}-\beta$.

The percentile-t method runs as follows. Put $\hat{\varepsilon}_i^*=Y_i^*-(\hat{\alpha}^*+x_i\hat{\beta}^*)$, $\hat{\sigma}^{*2}=n^{-1}\sum\hat{\varepsilon}_i^{*2}$ and $S_{n2}^*=n^{\frac{1}{2}}\sigma_x(\hat{\beta}^*-\hat{\beta})/\hat{\sigma}^*$. Let $\hat{\mathscr{S}}_\pi$ denote a regular region containing just πB of the B values of S_{n2}^*, and put

$$\hat{\mathscr{R}}_\pi=\hat{\beta}-n^{-\frac{1}{2}}\hat{\sigma}\hat{\mathscr{S}}_\pi=\{\hat{\beta}-n^{-\frac{1}{2}}\hat{\sigma}x\colon x\in\hat{\mathscr{S}}_\pi\}.$$

In practice B is of course finite, but as B increases the endpoints of confidence intervals converge to the values they would assume in the case $B=\infty$. The latter, "ideal" circumstance is perhaps most easily treated in terms of distribution functions, as follows. In addition to S_{n1}, S_{n2} and S_{n2}^* defined above, put $S_{n1}^*=n^{\frac{1}{2}}\sigma_x(\hat{\beta}^*-\hat{\beta})/\hat{\sigma}$. Let

$$H(x)=P(S_{n1}\leqq x), \qquad \hat{H}(x)=P(S_{n1}^*\leqq x\,|\,\mathscr{X}),$$
$$K(x)=P(S_{n2}\leqq x), \qquad \hat{K}(x)=P(S_{n2}^*\leqq x\,|\,\mathscr{X}).$$

Define inverses of distribution functions in the usual way; for example,

$$\hat{H}^{-1}(\pi)=\sup\{x\colon \hat{H}(x)\leqq\pi\}, \qquad 0<\pi<1.$$

Put

$$\hat{\beta}_{\mathrm{PERC(I)}}(\pi)=\hat{\beta}+n^{-\frac{1}{2}}\hat{\sigma}\hat{H}^{-1}(\pi), \qquad \hat{\beta}_{\mathrm{PERC(II)}}(\pi)=\hat{\beta}-n^{-\frac{1}{2}}\hat{\sigma}\hat{H}^{-1}(1-\pi),$$
$$\hat{\beta}_{\mathrm{PERC\text{-}T}}(\pi)=\hat{\beta}-n^{-\frac{1}{2}}\hat{\sigma}\hat{K}^{-1}(1-\pi). \tag{1.2}$$

In the case $B=\infty$, one-sided bootstrap confidence intervals constructed by the two percentile methods and the percentile-t method are respectively

$$(-\infty, \hat{\beta}_{\mathrm{PERC(I)}}(\pi)),\ (-\infty, \hat{\beta}_{\mathrm{PERC(II)}}(\pi)),\ (-\infty, \hat{\beta}_{\mathrm{PERC\text{-}T}}(\pi)).$$

All have nominal coverage π, in the sense that the true coverage converges to π as $n \to \infty$.

This approach leads easily to definitions of Efron's [10, 11] bias corrected confidence intervals, as we now show. Define

$$\hat{G}(x) = P(\hat{\Theta}^* \leq x \mid \mathscr{X}) = \hat{H}\{(x-\hat{\theta})/\hat{\sigma}\}, \qquad \hat{m} = \Phi^{-1}\{\hat{G}(\hat{\theta})\},$$

$$\hat{\rho}(\pi) = \Phi(z_\pi + 2\hat{m}) \quad \text{and} \quad \hat{\rho}_a(\pi) = \Phi[\hat{m} + (\hat{m}+z_\pi)\{1 - \hat{a}(\hat{m}+z_\pi)\}^{-1}], \tag{1.3}$$

where $\hat{a}$ is the bootstrap estimate of the acceleration constant and is defined below. Put

$$\hat{\beta}_{BC}(\pi) = \hat{\beta} + n^{-\frac{1}{2}} \hat{\sigma} \hat{H}^{-1}(\hat{\rho}), \qquad \hat{\beta}_{ABC}(\pi) = \hat{\beta} + n^{-\frac{1}{2}} \hat{\sigma} \hat{H}^{-1}(\hat{\rho}_a). \tag{1.4}$$

Then one-sided bias corrected and accelerated bias corrected confidence intervals are

$$(-\infty, \hat{\beta}_{BC}(\pi)), \qquad (-\infty, \hat{\beta}_{ABC}(\pi))$$

respectively, with nominal coverage π.

The acceleration constant may be defined in terms of Edgeworth expansions of H and K, which are obtainable from Theorem 2.1 in Sect. 2:

$$H(x) = \Phi(x) + n^{-\frac{1}{2}} p_1(x)\, \phi(x) + O(n^{-1}),$$

$$K(x) = \Phi(x) + n^{-\frac{1}{2}} q_1(x)\, \phi(x) + O(n^{-1}),$$

where p_1, q_1 are polynomials. The acceleration constant is

$$a = n^{-\frac{1}{2}} x^{-2} \{p_1(x) + q_1(x) - 2p_1(0)\} \tag{1.5}$$

[15], which does not depend on x. When unknowns in the formula for a are replaced by their bootstrap estimates, we obtain the bootstrap estimate $\hat{a}$ of a.

The exact critical point, $\hat{\beta}_{\text{exact}}(\pi) = \hat{\beta} - n^{-\frac{1}{2}} \hat{\sigma} K^{-1}(1-\pi)$, satisfies

$$P\{\beta \leq \hat{\beta}_{\text{exact}}(\pi)\} = \pi.$$

We say that a confidence interval $(-\infty, \hat{\beta}(\pi))$ is *second-order correct* [relative to the exact confidence interval $(-\infty, \hat{\beta}_{\text{exact}}(\pi))$] if the endpoints of these intervals agree to second order in $n^{-\frac{1}{2}}$, that is if

$$\hat{\beta}(\pi) - \hat{\beta}_{\text{exact}}(\pi) = O_p(n^{-\frac{3}{2}}). \tag{1.6}$$

Similarly the two-sided interval $(\hat{\beta}(\pi), \hat{\beta}(1-\pi))$, which has nominal coverage $1-2\pi$, is second-order correct if (1.6) holds for π and for $1-\pi$.

In [15], the two percentile methods percentile(I) and percentile(II) were called "backwards" and "hybrid" methods, respectively.

2. Simple Linear Regression: Slope parameter

2.1. Introduction

In this section we develop properties of bootstrap confidence intervals for the slope parameter β in the simple linear regression model (1.1). Our purpose is to compare the different methods – the two percentile methods percentile(I) and percentile(II), percentile-t, bias correction and accelerated bias correction, all of which were introduced in Subsection 1.2 of Sect. 1.

The section is structured as follows. Subsection 2.2 introduces notation and regularity conditions, and Subsection 2.3 describes Edgeworth expansions and Cornish-Fisher expansions. Main results are contained in Subsections 2.4 (where we give a formula for Efron's [11] accelerated bias correction), 2.5 (where we elucidate coverage properties of one-sided confidence intervals), 2.6 (where we describe the unusually virtuous features of two-sided percentile-t confidence intervals) and 2.7 (where we treat the case of random design). Each of Subsections 2.4–2.7 concludes with a list of points which briefly summarize the main conclusions reached there.

For the sake of brevity we do not state each of our results as a formal theorem. Nevertheless, all the formulae and expansions given in Subsections 2.3–2.6 are valid under the assumptions of Theorem 2.1. The most difficult of them to derive is stated as Theorem 2.2, and given a detailed proof in Sect. 5.

2.2. Notation and Regularity Conditions

In addition to notation introduced in Subsection 1.2, let $\gamma = E(\varepsilon_i/\sigma)^3$ and $\kappa = E(\varepsilon_i/\sigma)^4 - 3$ denote standardized skewness and kurtosis, let $\hat{\gamma} = n^{-1}\sum_i (\hat{\varepsilon}_i/\hat{\sigma})^3$ and $\hat{\kappa} = n^{-1}\sum_i (\hat{\varepsilon}_i/\hat{\sigma})^4 - 3$ be the sample versions (bootstrap estimates) of γ and κ, and put $\gamma_x = n^{-1}\sum_i (x_i/\sigma_x)^3$, $\lambda_x = \sum_i (x_i/\sigma_x)^4$ and $\kappa_x = \lambda_x - 3$. We refer to γ_x and κ_x as the standardized skewness and kurtosis, respectively, of the design points.

Next we describe our regularity conditions. To keep the discussion reasonably simple we assume that the common distribution of the errors ε_i is essentially bounded. A more elaborate proof shows that the assumption $E(|\varepsilon_i|^{48+\delta}) < \infty$ for any $\delta > 0$ is sufficient for all our main results, such as (2.6)–(2.11) and (2.14)–(2.19). A still more complex analysis using techniques developed by Bhattacharya and Ghosh [5] allows us to relax that condition to $E(|\varepsilon_i|^{40+\delta}) < \infty$. There is every likelihood that the latter condition is considerably more stringent than necessary, but we cannot see how to relax it to anything like the "minimal" assumption $E(\varepsilon_i^6) < \infty$. (The term of order n^{-2} in an Edgeworth expansion of coverage error involves $E(\varepsilon_i^6)$, and so the condition $E(\varepsilon_i^6) < \infty$ can be regarded as essential.)

We further assume that the error distribution has a nontrivial absolutely continuous component. In other words, it is nonsingular.

We allow the design points x_i to depend on n. That is, we permit a completely new set of points $x_1(n), \ldots, x_n(n)$ to be chosen for each n. To control these points we assume that

$$\begin{aligned}
&\sigma_x^2 \equiv n^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \text{ is bounded away from zero as } n \to \infty, \\
&|x_i - \bar{x}| \text{ is bounded uniformly in } 1 \leqq i \leqq n < \infty, \text{ and for some} \qquad (2.1) \\
&0 < \delta < 1 \text{ the numbers } m_\pm \text{ of indices } i \leqq n \text{ such that } \pm(x_i - \bar{x}) > \delta, \\
&\text{satisfy } m_+ > n^\delta \text{ and } m_- > n^\delta \text{ for all sufficiently large } n.
\end{aligned}$$

The restriction that $|x_i - \bar{x}|$ be bounded is roughly analogous to our boundedness assumption about the error distribution, and for our main results it may be relaxed to $n^{-1} \sum_i (x_i - \bar{x})^{10} < \infty$ by using longer proofs.

2.3. *Edgeworth and Cornish-Fisher Expansions*

We require Edgeworth expansions of the distribution functions H and K, and of their bootstrap estimates $\hat{H}$ and $\hat{K}$. In the former case these expansions may be established exactly as in Bhattacharya and Ghosh [5], in the latter case exactly as in Hall [14]. We do not give proofs here, only stating the results.

Theorem 2.1. *Assume that the common error distribution is essentially bounded and nonsingular, with zero mean. Suppose the design points x_i satisfy* (2.1). *Then there exist polynomials p_j, q_j of degree $3j-1$, odd for even j and even for odd j, such that for each $m \geqq 1$,*

$$\sup_{-\infty < z < \infty} \left| H(z) - \Phi(z) - \sum_{j=1}^{m} n^{-j/2} p_j(z)\, \phi(z) \right| = O(n^{-(m+1)/2}),$$

$$\sup_{-\infty < z < \infty} \left| K(z) - \Phi(z) - \sum_{j=1}^{m} n^{-j/2} q_j(z)\, \phi(z) \right| = O(n^{-(m+1)/2}).$$

Coefficients of p_j, q_j are uniformly bounded, and are polynomials in the first $j+2$ standardized cumulants of the error distribution and of the design. In particular,

$$\begin{aligned}
-p_1(z) &= -q_1(z) = \tfrac{1}{6} \gamma \gamma_x (z^2 - 1), \\
-p_2(z) &= z\{\tfrac{1}{24} \kappa \lambda_x (z^2 - 3) + \tfrac{1}{72} \gamma^2 \gamma_x^2 (z^4 - 10 z^2 + 15)\}, \\
-q_2(z) &= z\{2 + \tfrac{1}{24} (\kappa \kappa_x + 6)(z^2 - 3) + \tfrac{1}{72} \gamma^2 \gamma_x^2 (z^4 - 10 z^2 + 15)\}.
\end{aligned}$$

Define $\hat{p}_j$, $\hat{q}_j$ by replacing cumulants of the error distribution by their sample versions (bootstrap estimates) in formulae for p_j, q_j respectively. Then for each $m \geqq 1$,

$$\sup_{-\infty<z<\infty} \left| \hat{H}(z) - \Phi(z) - \sum_{j=1}^{m} n^{-j/2} \hat{p}_j(z)\, \phi(z) \right| = O_p(n^{-(m+1)/2}),$$

$$\sup_{-\infty<z<\infty} \left| \hat{K}(z) - \Phi(z) - \sum_{j=1}^{m} n^{-j/2} \hat{q}_j(z)\, \phi(z) \right| = O_p(n^{-(m+1)/2}).$$

Examples of $\hat{p}_j$, $\hat{q}_j$ include

$$-\hat{p}_1(z) = -\hat{q}_1(z) = \tfrac{1}{6}\hat{\gamma}\gamma_x(z^2-1),$$

$$-\hat{p}_2(z) = z\{\tfrac{1}{24}\hat{\kappa}\lambda_x(z^2-3) + \tfrac{1}{72}\hat{\gamma}^2\gamma_x^2(z^4-10z^2+15)\},$$

$$-\hat{q}_2(z) = z\{2 + \tfrac{1}{24}(\hat{\kappa}\kappa_x+6)(z^2-3) + \tfrac{1}{72}\hat{\gamma}^2\gamma_x^2(z^4-10z^2+15)\}.$$

These Edgeworth expansions are readily inverted, yielding the following Cornish-Fisher expansions, valid for each $m \geqq 1$:

$$x_\pi = H^{-1}(\pi) = z_\pi + \sum_{j=1}^{m} n^{-j/2} p_{j1}(z_\pi) + O(n^{-(m+1)/2}),$$

$$y_\pi = K^{-1}(\pi) = z_\pi + \sum_{j=1}^{m} n^{-j/2} q_{j1}(z_\pi) + O(n^{-(m+1)/2}),$$

$$\hat{x}_\pi = \hat{H}^{-1}(\pi) = z_\pi + \sum_{j=1}^{m} n^{-j/2} \hat{p}_{j1}(z_\pi) + O_p(n^{-(m+1)/2}), \tag{2.2}$$

$$\hat{y}_\pi = \hat{K}^{-1}(\pi) = z_\pi + \sum_{j=1}^{m} n^{-j/2} \hat{q}_{j1}(z_\pi) + O_p(n^{-(m+1)/2}). \tag{2.3}$$

Here, p_{j1} and q_{j1} are polynomials given simply in terms of the p_j's and q_j's, respectively. They are even function for odd j and odd functions for even j. For example,

$$p_1(y) = q_1(y) = \tfrac{1}{6}\gamma\gamma_x(y^2-1),$$

$$p_{21}(y) = y\{\tfrac{1}{24}\kappa\lambda_x(y^2-3) - \tfrac{1}{36}\gamma^2\gamma_x^2(2y^2-5)\},$$

$$q_{21}(y) = y\{2 + \tfrac{1}{24}(\kappa\kappa_y+6)(y^2-3) - \tfrac{1}{36}\gamma^2\gamma_x^2(2y^2-5)\}.$$

A little algebra, starting from the Edgeworth expansions in Theorem 2.1, shows that under the conditions of Theorem 2.1 the above Cornish-Fisher expansions (with remainders of the stated orders of magnitude) are valid uniformly in $\pi \in (\delta, 1-\delta)$ for each $0<\delta<\frac{1}{2}$.

2.4. *Bootstrap critical points*

We now derive expansions for the various bootstrap critical points introduced in Sect. 1, Subsection 1.2. The expansions developed here are all valid under the conditions in Theorem 2.1 – indeed under weaker conditions, as discussed in Subsection 2.2. For the sake of brevity we shall not state the expansions as formal theorems.

We begin with formulae for quantiles in bias corrected critical points. The Edgeworth expansion of $\hat{H}$ is identical to that of H, except of course that $\hat{\gamma}$ and $\hat{\kappa}$ replace γ and κ in polynomials. Therefore

$$\begin{aligned}\hat{m} &= \Phi^{-1}\{\hat{G}(\hat{\beta})\} = \Phi^{-1}\{\hat{H}(0)\} = \Phi^{-1}\{\tfrac{1}{2} + n^{-\frac{1}{2}}\tfrac{1}{6}\hat{\gamma}\gamma_x \phi(0) + O_p(n^{-\frac{3}{2}})\} \\ &= n^{-\frac{1}{2}}\tfrac{1}{6}\hat{\gamma}\gamma_x + O_p(n^{-\frac{3}{2}}).\end{aligned}$$

Defining $\hat{p} = \Phi(z_\pi + 2\hat{m})$ we see that $z_{\hat{p}} = z_\pi + 2\hat{m}$, and so by (2.2),

$$\hat{x}_{\hat{p}} = z_\pi + n^{-\frac{1}{2}}\{\hat{p}_{11}(z_\pi) + \tfrac{1}{3}\hat{\gamma}\gamma_x\} + n^{-1}\{\hat{p}_{21}(z_\pi) + \tfrac{1}{3}\hat{\gamma}\gamma_x \hat{p}'_{11}(z_\pi)\} + O_p(n^{-\frac{3}{2}}). \quad (2.4)$$

To derive a similar formula for the accelerated bias corrected quantile $\hat{x}_{\hat{p}_a}$, note that the polynomials p_1 and q_1 in the order $n^{-\frac{1}{2}}$ terms of Edgeworth expansions of H and K are $p_1(y) = q_1(y) = -\frac{1}{6}\gamma\gamma_x(y^2-1)$; see Theorem 2.1. According to Definition (1.5) in Subsection 1.2, this means that the acceleration constant is

$$a = n^{-\frac{1}{2}} y^{-2}\{p_1(y) + q_1(y) - 2p_1(0)\} = -n^{-\frac{1}{2}}\tfrac{1}{6}\gamma\gamma_x.$$

Its bootstrap estimate is $\hat{a} = -n^{-\frac{1}{2}}\frac{1}{3}\hat{\gamma}\gamma_x$, and we define

$$\begin{aligned}\hat{\zeta}_\pi &= \hat{m} + (\hat{m} + z_\pi)\{1 - \hat{a}(\hat{m} + z_\pi)\}^{-1} \\ &= z_\pi + n^{-\frac{1}{2}}\tfrac{1}{3}\hat{\gamma}\gamma_x(1 - z_\pi^2) + n^{-1}\tfrac{1}{9}\hat{\gamma}^2\gamma_x^2 z_\pi(z_\pi^2 - 1) + O_p(n^{-\frac{3}{2}}).\end{aligned}$$

The adjusted probability level for the accelerated bias corrected critical point is $\hat{p}_a = \Phi(\hat{\zeta}_\pi)$; see Definition (1.3). Therefore $z_{\hat{p}_a} = \hat{\zeta}_\pi$, and by (2.2),

$$\begin{aligned}\hat{x}_{\hat{p}_a} = \hat{z}_\pi &+ n^{-\frac{1}{2}}\{\hat{p}_{11}(z_\pi) + \tfrac{1}{3}\hat{\gamma}\gamma_x(1 - z_\pi^2)\} \\ &+ n^{-1}\{\hat{p}_{21}(z_\pi) + \tfrac{1}{3}\hat{\gamma}\gamma_x(1 - z_\pi^2)\hat{p}'_{11}(z_\pi) + \tfrac{1}{9}\hat{\gamma}^2\gamma_x^2 z_\pi(z_\pi^2 - 1)\} + O_p(n^{-\frac{3}{2}}). \quad (2.5)\end{aligned}$$

Define the π-level percentile critical points $\hat{\beta}_{\text{PERC(I)}}(\pi)$ and $\hat{\beta}_{\text{PERC(II)}}(\pi)$, percentile-$t$ point $\hat{\beta}_{\text{PERC-T}}(\pi)$, bias corrected point $\hat{\beta}_{BC}(\pi)$ and accelerated bias corrected point $\hat{\beta}_{ABC}(\pi)$ as in Subsection 1.2; see Definitions (1.2) and (1.4). Then by (2.2)–(2.5),

$$\begin{aligned}\hat{\beta}_{\mathrm{PERC(I)}}(\pi)&=\hat{\beta}+n^{-\frac{1}{2}}\sigma_x^{-1}\hat{\sigma}\hat{x}_\pi\\&=\hat{\beta}+n^{-\frac{1}{2}}\sigma_x^{-1}\hat{\sigma}[z_\pi+n^{-\frac{1}{2}}\tfrac{1}{6}\hat{\gamma}\gamma_x(z_\pi^2-1)\\&\quad+n^{-1}z_\pi\{\tfrac{1}{24}\hat{\kappa}\delta_x(z_\pi^2-3)-\tfrac{1}{36}\hat{\gamma}^2\gamma_x^2(2z_\pi^2-5)\}]+O_p(n^{-2}),\end{aligned}\tag{2.6}$$

$$\begin{aligned}\hat{\beta}_{\mathrm{PERC(II)}}(\pi)&=\hat{\beta}-n^{-\frac{1}{2}}\sigma_x^{-1}\hat{\sigma}\hat{x}_{1-\pi}\\&=\hat{\beta}+n^{-\frac{1}{2}}\sigma_x^{-1}\hat{\sigma}[z_\pi-n^{-\frac{1}{2}}\tfrac{1}{6}\hat{\gamma}\gamma_x(z_\pi^2-1)\\&\quad+n^{-1}z_\pi\{\tfrac{1}{24}\hat{\kappa}\delta_x(z_\pi^2-3)-\tfrac{1}{36}\hat{\gamma}^2\gamma_x^2(2z_\pi^2-5)\}]+O_p(n^{-2}),\end{aligned}\tag{2.7}$$

$$\begin{aligned}\hat{\beta}_{\mathrm{PERC-T}}(\pi)&=\hat{\beta}-n^{-\frac{1}{2}}\sigma_x^{-1}\hat{\sigma}\hat{x}_{1-\pi}\\&=\hat{\beta}+n^{-\frac{1}{2}}\sigma_x^{-1}\hat{\sigma}[z_\pi-n^{-\frac{1}{2}}\tfrac{1}{6}\hat{\gamma}\gamma_x(z_\pi^2-1)\\&\quad+n^{-1}z_\pi\{2+\tfrac{1}{24}(\hat{\kappa}\kappa_x+6)(z_\pi^2-3)-\tfrac{1}{36}\hat{\gamma}^2\gamma_x^2(2z_\pi^2-5)\}]\\&\quad+O_p(n^{-2}),\end{aligned}\tag{2.8}$$

$$\begin{aligned}\hat{\beta}_{BC}(\pi)&=\hat{\beta}+n^{-\frac{1}{2}}\sigma_x^{-1}\hat{\sigma}\hat{x}_{\hat{p}}\\&=\hat{\beta}+n^{-\frac{1}{2}}\sigma_x^{-1}\hat{\sigma}[z_\pi+n^{-\frac{1}{2}}\tfrac{1}{6}\hat{\gamma}\gamma_x(z_\pi^2+1)\\&\quad+n^{-1}z_\pi\{\tfrac{1}{24}\hat{\kappa}\delta_x(z_\pi^2-3)-\tfrac{1}{36}\hat{\gamma}^2\gamma_x^2(2z_\pi^2-9)\}]+O_p(n^{-2}),\end{aligned}\tag{2.9}$$

$$\begin{aligned}\hat{\beta}_{ABC}(\pi)&=\hat{\beta}+n^{-\frac{1}{2}}\sigma_x^{-1}\hat{\sigma}\hat{x}_{\hat{p}_a}\\&=\hat{\beta}+n^{-\frac{1}{2}}\sigma_x^{-1}\hat{\sigma}[z_\pi-n^{-\frac{1}{2}}\tfrac{1}{6}\hat{\gamma}\gamma_x(z_\pi^2-1)\\&\quad+n^{-1}z_\pi\{\tfrac{1}{24}\hat{\kappa}\delta_x(z_\pi^2-3)-\tfrac{1}{36}\hat{\gamma}^2\gamma_x^2(2z_\pi^2-5)\}]+O_p(n^{-2}).\end{aligned}\tag{2.10}$$

A much simpler argument shows that the "ideal" critical point, $\hat{\beta}_{\mathrm{exact}}(\pi)=\hat{\beta}-n^{-\frac{1}{2}}\sigma_x^{-1}\hat{\sigma}y_{1-\pi}$, satisfies

$$\begin{aligned}\hat{\beta}_{\mathrm{exact}}(\pi)&=\hat{\beta}-n^{-\frac{1}{2}}\sigma_x^{-1}\hat{\sigma}x_{1-\pi}\\&=\hat{\beta}+n^{-\frac{1}{2}}\sigma_x^{-1}\hat{\sigma}[z_\pi-n^{-\frac{1}{2}}\tfrac{1}{6}\gamma\gamma_x(z_\pi^2-1)\\&\quad+n^{-1}z_\pi\{2+\tfrac{1}{24}(\kappa\kappa_x+6)(z_\pi^2-3)-\tfrac{1}{36}\gamma^2\gamma_x^2(2z_\pi^2-5)\}]\\&\quad+O_p(n^{-2}).\end{aligned}\tag{2.11}$$

The most important conclusions to be drawn from these formulae are the following. In interpreting (ii) and (iii) below, notice that $\hat{\gamma}=\gamma+O_p(n^{-\frac{1}{2}})$.

(i) Percentile(II) and accelerated bias corrected critical points are identical up to and including terms of order $n^{-\frac{3}{2}}$. That is to say, those points are *third-order equivalent*. This is unusual; see [15] for more typical results.

(ii) $\hat{\beta}_{\mathrm{PERC(II)}}$, $\hat{\beta}_{\mathrm{PERC-T}}$ and $\hat{\beta}_{ABC}$ agree with $\hat{\beta}_{\mathrm{exact}}$ in terms of order n^{-1}. That is, $\hat{\beta}_{\mathrm{PERC(II)}}$, $\hat{\beta}_{\mathrm{PERC-T}}$ and $\hat{\beta}_{ABC}$ are *second-order correct.* This is to be expected for $\hat{\beta}_{\mathrm{PERC-T}}$ and $\hat{\beta}_{ABC}$, but is unusual for $\hat{\beta}_{\mathrm{PERC(II)}}$; compare [15].

(iii) $\hat{\beta}_{\mathrm{PERC(I)}}$ and $\hat{\beta}_{BC}$ are generally not second-order correct; this is the norm for those points. However, if $\gamma_x = O(n^{-\frac{1}{2}})$ then even $\hat{\beta}_{\mathrm{PERC(I)}}$ and $\hat{\beta}_{BC}$ are second-order correct. We often have $\gamma_x = O(n^{-1})$ for equally spaced design points; for example, that is the case if $x_i = cin^{-1} + d_n$ for fixed c and variable d_n.

2.5. *Coverage Probabilities*

The expansions developed here may all be established rigorously under the conditions of Theorem 2.1. To save space we do not state the results as theorems. The proofs are no more than variants of an argument which we shall give in Sect. 5, and so we shall only sketch the main ideas.

Any one of the bootstrap critical points $\hat{\beta}(\pi)$ whose Cornish-Fisher expansions we have just derived may be written in the form

$$\hat{\beta}(\pi) = \hat{\beta} + n^{-\frac{1}{2}} \sigma_x^{-1} \hat{\sigma} \{ z_\pi + n^{-\frac{1}{2}} \hat{\gamma} c_1 + n^{-1} c_2 + O_p(n^{-\frac{3}{2}}) \}, \tag{2.12}$$

where c_1 and c_2 are nonrandom and depend on π. A little algebra shows that $\hat{\gamma} = \gamma + n^{-\frac{1}{2}} U + O_p(n^{-1})$, where, with $\xi_i = \varepsilon_i / \sigma$,

$$U = n^{-\frac{1}{2}} \sum_{i=1}^{n} (\xi_i^3 - \gamma) - \tfrac{3}{2} \gamma n^{-\frac{1}{2}} \sum_{i=1}^{n} (\xi_i^2 - 1) - 3 n^{-\frac{1}{2}} \sum_{i=1}^{n} \xi_i.$$

Therefore the exact coverage of the interval $(-\infty, \hat{\beta}(\pi)]$ equals

$$\begin{aligned} P\{\beta \leqq \hat{\beta}(\pi)\} &= P\{n^{\frac{1}{2}}(\hat{\beta} - \beta)\sigma_x/\hat{\sigma} \geqq -(z_\pi + n^{-\frac{1}{2}} \hat{\gamma} c_1 + n^{-1} c_2) + O_p(n^{-\frac{3}{2}})\} \\ &= P(S + n^{-1} U c_1 > -y_\pi) + O(n^{-\frac{3}{2}}), \end{aligned} \tag{2.13}$$

where $S_{n2} = n^{\frac{1}{2}} \sigma_x (\hat{\beta} - \beta)/\hat{\sigma}$ and $y_\pi = z_\pi + n^{-\frac{1}{2}} \gamma c_1 + n^{-1} c_2$.

To evaluate the last-written probability we need an Edgeworth expansion of the distribution of $T = S_{n2} + n^{-1} U c_1$. We claim that this expansion is identical to that of the distribution of S_{n2}, up to a remainder of order $n^{-\frac{3}{2}}$. It is this fact which enables the unusual accuracy of percentile-t confidence intervals. To verify our claim, write $S = S_{n2}$ and notice that

$$\begin{aligned} E(T) = E(S), \quad & E(T^2) = E(S^2) + 2n^{-1} c_1 E(SU) + O(n^{-2}), \\ & E(T^3) = E(S^3) + 3n^{-1} c_1 E(S^2 U) + O(n^{-2}) = E(S^3) + O(n^{-\frac{3}{2}}), \\ & E(T^4) = E(S^4) + 4n^{-1} c_1 E(S^3 U) + O(n^{-2}) \\ & \qquad = E(S^4) + 12 n^{-1} c_1 E(S^2) E(SU) + O(n^{-2}). \end{aligned}$$

Now, $S = S' + O_p(n^{-\frac{1}{2}})$, where $S' = n^{-\frac{1}{2}} \sigma_x^{-1} \sum_i (x_i - \bar{x}) \xi_i$; and $E(S'U) = 0$. Therefore $E(SU) = O(n^{-\frac{1}{2}})$. (In fact, it equals $O(n^{-1})$.) In consequence, the first four moments of T agree with their counterparts for S up to remainders of order $n^{-\frac{3}{2}}$, implying that the first four cumulants of S and T agree in the same sense. Since the $n^{-\frac{1}{2}}$ and n^{-1} terms in Edgeworth expansions of S and T depend only on the first four cumulants, this verifies our claim about Edgeworth expansions.

Returning to formula (2.13) we now see that if $\hat{\beta}(\pi)$ is the critical point defined at (2.12) then

$$\begin{aligned}P\{\beta \leqq \hat{\beta}(\pi)\} &= P(S > -y_\pi) + O(n^{-\frac{3}{2}})\\ &= \pi + n^{-\frac{1}{2}}\{\gamma c_1 + \tfrac{1}{6}\gamma\gamma_x(z_\pi^2 - 1)\}\,\phi(z_\pi) + n^{-1}[c_2 - \tfrac{1}{2}\gamma^2 c_1^2 z_\pi\\ &\quad + \tfrac{1}{6}\gamma^2\gamma_x c_1 z_\pi(3 - z_\pi^2) - z_\pi\{2 + \tfrac{1}{24}(\kappa\kappa_x + 6)(z_\pi^2 - 3)\\ &\quad + \tfrac{1}{72}\gamma^2\gamma_x^2(z_\pi^4 - 10 z_\pi^2 + 15)\}]\,\phi(z_\pi) + O(n^{-\frac{3}{2}}),\end{aligned}$$

using the Edgeworth expansion of the distribution of S given in Theorem 2.1. Noting the versions of c_1 and c_2 for the various bootstrap critical points (see (2.6)–(2.10)), we obtain the following expansion of coverage probability:

$$\begin{aligned}P\{\beta \leqq \hat{\beta}_{\mathrm{PERC(I)}}(\pi)\} &= \pi + n^{-\frac{1}{2}}\tfrac{1}{3}\gamma\gamma_x(z_\pi^2 - 1)\,\phi(z_\pi) - n^{-1} z_\pi\{2 - \tfrac{1}{8}(\kappa - 2)(z_\pi^2 - 3)\\ &\quad + \tfrac{1}{18}\gamma^2\gamma_x^2(z_\pi^4 - 4 z_\pi^2 + 3)\}\,\phi(z_\pi) + O(n^{-\frac{3}{2}}),\end{aligned} \tag{2.14}$$

$$\begin{aligned}P\{\beta \leqq \hat{\beta}_{\mathrm{PERC(II)}}(\pi)\} &= \pi - n^{-1} z_\pi\{2 - \tfrac{1}{8}(\kappa - 2)(z_\pi^2 - 3)\}\,\phi(z_\pi) + O(n^{-\frac{3}{2}})\\ &= P\{\beta \leqq \hat{\beta}_{ABC}(\pi)\} + O(n^{-\frac{3}{2}}),\end{aligned} \tag{2.15}$$

$$P\{\beta \leqq \hat{\beta}_{\mathrm{PERC\text{-}T}}(\pi)\} = \pi + O(n^{-\frac{3}{2}}), \tag{2.16}$$

$$\begin{aligned}P\{\beta \leqq \hat{\beta}_{BC}(\pi)\} &= \pi + n^{-\frac{1}{2}}\tfrac{1}{3}\gamma\,\gamma_x z_\pi^2\,\phi(z_\pi) - n^{-1} z_\pi\{2 - \tfrac{1}{8}(\kappa - 2)(z_\pi^2 - 3)\\ &\quad + \tfrac{1}{18}\gamma^2\gamma_x^2(z_\pi^4 - 2 z_\pi^2)\}\,\phi(z_\pi) + O(n^{-\frac{3}{2}}).\end{aligned} \tag{2.17}$$

Under the condition of Theorem 2.1, all these expansions are available uniformly in $0 < \pi < 1$. Sect. 5 describes the manner of proof.

The main conclusions to be drawn from these formulae are the following.

(i) The percentile-t coverage error is only $O(n^{-\frac{3}{2}})$, for one-sided confidence intervals. At first sight this is surprising, since $\hat{\beta}_{\mathrm{PERC\text{-}T}}(\pi)$ is not third-order correct.

(ii) Percentile(II) and accelerated bias corrected intervals have coverage errors of order n^{-1}, and those errors agree up to remainders of order $n^{-\frac{3}{2}}$. In the light of results in Subsection 2.4, this is to be expected, since percentile(II) and accelerated bias corrected critical points are second-order correct and third-order equivalent. However, one-sided percentile(II) method intervals have coverage error only of order $n^{-\frac{1}{2}}$ in most non-regression problems; see [15].

(iii) Percentile(I) and ordinary bias corrected intervals have coverage errors of order $n^{-\frac{1}{2}}$. This is to be expected since those intervals are only first-order correct. However the errors are order n^{-1} if $\gamma_x = O(n^{-\frac{1}{2}})$, which is often the case when the design points are equally spaced; see remark (iii) at the end of Subsection 2.4.

2.6. *Two-sided Confidence Intervals*

Again, all the expansions studied here may be given rigorous proofs under the conditions of Theorem 2.1. To economise on space we state only one of those expansions as a formal theorem, whose proof is given in Sect. 5. The main idea behind the proofs is outlined below.

Assume $0<\pi<\frac{1}{2}$. The equal-tailed, two-sided bootstrap confidence interval with nominal coverage $1-2\pi$, based on the critical point $\hat{\beta}(\pi)$, is of course $I(1-2\pi)=[\hat{\beta}(\pi), \hat{\beta}(1-\pi)]$. Its length is $l(1-2\pi)=\hat{\beta}(1-\pi)-\hat{\beta}(\pi)$, and its exact coverage is

$$P(1-2\pi)=P\{\hat{\beta}(\pi)\leqq\beta\leqq\hat{\beta}(1-\pi)\}=P\{\beta\leqq\hat{\beta}(1-\pi)\}-P\{\beta\leqq\hat{\beta}(\pi)\}.$$

To appreciate properties of these quantities, observe that in general

$$\hat{\beta}(\pi)=\hat{\beta}+n^{-\frac{1}{2}}\sigma_x^{-1}\hat{\sigma}\left\{z_\pi+\sum_{j=1}^{3} n^{-j/2}\hat{s}_j(z_\pi)+O_p(n^{-2})\right\}$$

and

$$P\{\beta\leqq\hat{\beta}(\pi)\}=\pi+\sum_{j=1}^{3} n^{-j/2}t_j(z_\pi)\,\phi(z_\pi)+O(n^{-2}),$$

where the $\hat{s}_j$'s are polynomials with random coefficients, the t_j's are polynomials with nonrandom coefficients, and odd/even indexed polynomials are even/odd functions respectively. For $j=1$ and 2, formulae for $\hat{s}_j$ follow from (2.6)–(2.10) and formulae for t_j follow from (2.14)–(2.17). (We could have taken longer series expansions, but those above suffice for our purposes.) Therefore

$$l(1-2\pi)=2n^{-\frac{1}{2}}\sigma_x^{-1}\hat{\sigma}\{z_\pi+n^{-1}\hat{s}_2(z_\pi)+O_p(n^{-2})\} \tag{2.18}$$

and

$$P(1-2\pi)=1-2\pi+2n^{-1}t_2(z_\pi)\,\phi(z_\pi)+O(n^{-2}). \tag{2.19}$$

The polynomial t_2 is identically zero in the case of percentile-t (see (2.16)), and so we have $P(1-2\pi)=1-2\pi+O(n^{-2})$ in that circumstance. Since this case is so important we have chosen to make it the subject of our formal theorem.

Theorem 2.2. *Assume the conditions of Theorem* 2.1. *Then in the percentile-t case,* $P(1-2\pi)=1-2\pi+O(n^{-2})$ *uniformly in* $0<\pi<\frac{1}{2}$.

We may draw several conclusions from formulae (2.18) and (2.19).

(i) Owing to symmetry properties of Edgeworth expansions, coverage errors for two-sided confidence intervals are generally of order n^{-1}, or smaller, even for the normal approximation method. This means that techniques such as percentile(I) and bias corrected, which have large coverage errors (of order $n^{-\frac{1}{2}}$) in the case of one-sided intervals, do not have such serious coverage problems in the case of two-sided intervals.

(ii) Only in the case of percentile-t is the polynomial t_2 in (2.19) identically zero. Therefore percentile-t stands head and shoulders above the other methods in terms of coverage accuracy, since it guarantees a coverage error of order n^{-2}; the others give order n^{-1}.

(iii) Second-order correctness of percentile(II) and accelerated bias corrected critical points shows up in the polynomial t_2, by dictating that t_2 contain no contribution from skewness of either the error distribution of the design points. See formulae (2.14)–(2.17) for values of t_2.

(iv) Percentile-t has advantages other than reduced coverage error. We already know that $\hat\beta_{\text{PERC-T}}(\pi)$ is particularly close to the exact critical point, $\hat\beta_{\text{exact}}(\pi)$; see remark (ii) at the end of Subsection 2.4. In consequence, lengths of confidence intervals based on percentile-t critical points are particularly close to lengths of intervals based on exact critical points. Indeed, let $\hat s_2$ be the polynomial appearing in (2.18) in the case where $\hat\beta(\pi)=\hat\beta_{\text{PERC-T}}(\pi)$, and let s_2 be the version of $\hat s_2$ for the case $\hat\beta(\pi)=\hat\beta_{\text{exact}}(\pi)$. Then $\hat s_2$ is obtained from s_2 by simply replacing γ by $\hat\gamma$ and κ by $\hat\kappa$; compare (2.7) and (2.11). In consequence, $E(\hat\sigma\hat s_2)=E(\hat\sigma)s_2+O(n^{-1})$, from which it follows via (2.18) that $E\{l_{\text{PERC-T}}(1-2\pi)\}=E\{l_{\text{exact}}(1-2\pi)\}+O(n^{-\frac{5}{2}})$. This formula is not valid for intervals constructed from the second-order correct points $\hat\beta_{\text{PERC(II)}}$ and $\hat\beta_{ABC}$, where the difference between $E\{l(1-2\pi)\}$ and $E\{l_{\text{exact}}(1-2\pi)\}$ is generally of order $n^{-\frac{3}{2}}$.

(v) To take maximum advantage of the accuracy of the percentile-t approximation we may construct "shortest" or "likelihood based" bootstrap confidence intervals. The argument runs as follows. For each y such that $\hat K(y)\geqq 1-2\pi$, choose $z=z(y)$ such that $\hat K(y)-\hat K(-z)$ is as close as possible to $1-2\pi$. Take $(\hat v, \hat w)$ to be that pair (y, z) which minimizes $y+z$. Then the "shortest" percentile-t confidence interval with nominal coverage $1-2\pi$ is

$$I_{\text{SHORT}}=[\hat\beta-n^{-\frac{1}{2}}\sigma_x^{-1}\hat\sigma\hat v,\ \hat\beta+n^{-\frac{1}{2}}\sigma_x^{-1}\hat\sigma\hat w].$$

Arguments summarized in [15, Subsections 2.5 and 4.6] may be adapted to show that in the present circumstances, I_{SHORT} is shorter than the equal-tailed interval by an amount of order $n^{-\frac{3}{2}}$, and that it has coverage error of order n^{-2}.

2.7. *Random Design Points*

Here we take the design points x_i to be random, and write them as X_i to indicate this distinction. The work in this subsection is at an heuristic level.

If the assumption is only that $E(Y_i|X_i)=\alpha+\beta X_i$ and that the pairs (X_i, Y_i) are independent and identically distributed, then the pairs rather than the residuals should be resampled. Techniques developed earlier in this section may be used to analyze that case. It is found that percentile-t again performs well, particularly in the senses described in remarks (iv) and (v) at the end of the previous section. However, coverage errors of two-sided intervals are generally of order n^{-1}, not n^{-2}. To appreciate why, put $\varepsilon_i=Y_i-(\alpha+X_i\beta)$ and recall from the argument following (2.13) that our proof that coverage error is of order n^{-2} depended crucially on identities such as

$$E\left\{n^{-1}\sum_{i=1}^{n}(x_i-\bar x)\varepsilon_i^2\right\}=0,$$

or at least equals $O(n^{-\frac{1}{2}})$. On the present occasion the left-hand side should be replaced by

$$n^{-1}\sum_{i=1}^{n} E\{(X_i-\bar{X})\varepsilon_i^2\}=E(X_1\varepsilon_1^2)-E(X_1)E(\varepsilon_1^2)+O(n^{-1}),$$

whose value depends on the relationship between X_1 and Y_1.

In the remainder of this section we discuss the percentile-t interval in the case where the variables $X_1, \varepsilon_1, \ldots, X_n, \varepsilon_n$ are assumed totally independent, the X_i's have a common distribution and the ε_i's have a common distribution with zero mean. Of course, Y_i is defined by $Y_i=\alpha+X_i\beta+\varepsilon_i$. Put $\sigma^2=E(\varepsilon_i^2)$, $\sigma_X^2=n^{-1}\sum_i(X_i-\bar{X})^2$, $\hat{\alpha}=\bar{Y}-\bar{X}\hat{\beta}$, $\hat{\beta}=n^{-1}\sigma_X^{-2}\sum_i(X_i-\bar{X})(Y_i-\bar{Y})$, $\hat{\varepsilon}_i=Y_i-(\hat{\alpha}+X_i\hat{\beta})$ and $\hat{\sigma}^2=n^{-1}\sum_i\hat{\varepsilon}_i^2$. Given the sample of pairs $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, conduct two totally independent resampling operations in which a random sample $\{X_1^*, \ldots X_n^*\}$ is drawn with replacement from $\{X_1, \ldots, X_n\}$ and a random sample $\{\varepsilon_1^*, \ldots, \varepsilon_n^*\}$ is drawn with replacement from $\{\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n\}$. Define $Y_i^*=\hat{\alpha}+X_i^*\hat{\beta}+\hat{\varepsilon}_i^*$, $\bar{Y}^*=n^{-1}\sum_i Y_i^*$, $\bar{X}^*=n^{-1}\sum_i X_i^*$, $\sigma_X^{*2}=n^{-1}\sum_i(X_i^*-\bar{X}^*)^2$, $\hat{\alpha}^*=\bar{Y}^*-\bar{X}^*\hat{\beta}^*$, $\hat{\beta}^*=\sigma_X^{*-2}\sum_i(X_i^*-\bar{X}^*)(Y_i^*-\bar{Y}^*)$, $\hat{\varepsilon}_i^*=Y_i^*-(\hat{\alpha}^*+x_i^*\hat{\beta}^*)$ and $\hat{\sigma}^{*2}=n^{-1}\sum_i\hat{\varepsilon}_i^{*2}$. Let K and $\hat{K}$ denote the distribution functions of $n^{\frac{1}{2}}\sigma_X(\hat{\beta}-\beta)/\hat{\sigma}$ and $n^{\frac{1}{2}}\sigma_X^*(\hat{\beta}^*-\hat{\beta})/\hat{\sigma}^*$, respectively, and put $y_\pi=K^{-1}(\pi)$ and $\hat{y}_\pi=\hat{K}^{-1}(\pi)$. The percentile-$t$ critical point is $\hat{\beta}_{\text{PERC-T}}(\pi)=\hat{\beta}-n^{-\frac{1}{2}}\sigma_X^{-1}\hat{\sigma}\hat{y}_{1-\pi}$, and the exact critical point is $\hat{\beta}_{\text{exact}}(\pi)=\hat{\beta}-n^{-\frac{1}{2}}\sigma_X^{-1}\hat{\sigma}y_{1-\pi}$. Arguments similar to those in Subsections 2.3 and 2.5 may be employed to show that $|\hat{\beta}_{\text{PERC-T}}(\pi)-\hat{\beta}_{\text{exact}}(\pi)|=O_p(n^{-\frac{3}{2}})$ (so that $\hat{\beta}_{\text{PERC-T}}(\pi)$ is second-order correct), and for $0<\pi<\frac{1}{2}$,

$$P\{\hat{\beta}_{\text{PERC-T}}(\pi)\leqq\beta\leqq\hat{\beta}_{\text{PERC-T}}(1-\pi)\}=1-2\pi+O(n^{-2})$$

(so that the percentile-t interval has coverage error $O(n^{-2})$).

The main conclusions to be drawn from these arguments are as follows.

(i) In the random design case, where the X_i's and ε_i's are assumed totally independent, the virtues of percentile-t and percentile(II) described earlier for the fixed design case, continue to hold.

(ii) In the correlation model, where it is assumed only that the pairs (X_i, Y_i) are independent and identically distributed, the aforementioned virtues may evaporate. Then, the various bootstrap methods enjoy only the properties which they do in more usual circumstances.

3. Simple Linear Regression: Intercept Parameter and Means

3.1. Introduction

In this section we describe versions of results in Sect. 2 for the case of estimating the intercept parameter α or the mean $\alpha+x_0\beta$ of Y given that $x=x_0$. Of course the former is just a special case of the latter, and so we may confine attention to the problem of setting confidence intervals for $y_0 \equiv \alpha + x_0 \beta$, where x_0 is fixed. In this context the bootstrap behaves in a manner which typifies its performance in most statistical applications, where the exceptional properties noted in Sect. 2 are seldom evidenced. Therefore our account will be particularly brief. Sufficient regularity conditions for all our results are those in Theorem 2.1.

3.2. Notation

In addition to notation introduced in Subsections 1.2 and 2.2, put $\hat{y}_0 = \hat{\alpha} + x_0 \hat{\beta}$ and $\hat{y}_0^* = \hat{\alpha}^* + x_0 \hat{\beta}^*$. Let $y_i = \sigma_x^{-2}\{(x_0 - \bar{x})(x_i - \bar{x}) + \sigma_x^2\}$, $\sigma_y^2 = n^{-1} \sum_i y_i^2 = 1 + \sigma_x^{-2}(x_0 - \bar{x})^2$, $\gamma_y = n^{-1} \sum_i y_i^3$, $\lambda_y = n^{-1} \sum_i y_i^4$ and $\kappa_y = \lambda_y - 3$. Redefine H, $\hat{H}$, K and $\hat{K}$ to be distribution functions of $n^{\frac{1}{2}}(\hat{y}_0 - y_0)/(\sigma \sigma_y)$, $n^{\frac{1}{2}}(\hat{y}_0^* - \hat{y}_0)/(\hat{\sigma} \sigma_y)$, $n^{\frac{1}{2}}(\hat{y}_0 - y_0)/(\hat{\sigma} \sigma_y)$ and $n^{\frac{1}{2}}(\hat{y}_0^* - \hat{y}_0)/(\hat{\sigma}^* \sigma_y)$, respectively. (In the case of $\hat{H}$ and $\hat{K}$, condition on the sample $\mathscr{X} = \{(x_1, Y_1), \ldots, (x_n, Y_n)\}$. Incidentally, note that $n^{\frac{1}{2}}(\hat{y}_0 - y_0)/(\sigma \sigma_y)$ has zero mean and unit variance.)

3.3. Edgeworth and Cornish-Fisher Expansions

Since we have new definitions of H, K, $\hat{H}$ and $\hat{K}$, we require corresponding new definitions of the polynomials p_j, q_j, $\hat{p}_j$ and $\hat{q}_j$. Excepting this obvious change, Theorem 2.1 holds exactly as before, under the same regularity conditions. Likewise, Cornish-Fisher expansions such as (2.2) and (2.3) are valid as before. The new versions of polynomials include

$$-p_1(z) = \tfrac{1}{6}\gamma \gamma_y (z^2 - 1), \qquad -q_1(z) = \tfrac{1}{6}\gamma(\gamma_y - 3\sigma_y^{-1})z^2 - \tfrac{1}{6}\gamma \gamma_y,$$

$$-p_2(z) = z\{\tfrac{1}{24}\kappa \lambda_y (z^2 - 3) + \tfrac{1}{72}\gamma^2 \gamma_y^2 (z^4 - 10z^2 + 15)\},$$

$$-q_2(z) = z[2 + \sigma_y^{-2}\gamma^2 + \tfrac{1}{24}\{\kappa \kappa_y + 6 - 8\gamma^2 \sigma_y^{-1}(\gamma_y - 3\sigma_y^{-1})\}(z^2 - 3)$$
$$+ \tfrac{1}{72}\gamma^2(\gamma_y - 3\sigma_y^{-1})^2(z^4 - 10z^2 + 15)].$$

Polynomials $\hat{p}_1$, $\hat{q}_1$, $\hat{p}_2$, $\hat{q}_2$ are obtained from these expressions on replacing γ, κ by $\hat{\gamma}$, $\hat{\kappa}$ respectively.

3.4. Bootstrap Critical Points

Edgeworth expansions of H and K are described above. From these, and formula (1.5) for the acceleration constant, we may deduce that the bootstrap estimate of the acceleration constant is

$$\hat{a} = n^{-\frac{1}{2}} \tfrac{1}{6} \hat{\gamma}(3\sigma_y^{-1} - 2\gamma_y).$$

This is required for constructing Efron's [11] accelerated bias corrected critical point. In the present case the various critical points are $\hat{y}_{0\,\mathrm{PERC(I)}}(\pi)=\hat{y}_0+n^{-\frac{1}{2}}\sigma_y\hat{\sigma}\hat{H}^{-1}(\pi)$, $\hat{y}_{0\,\mathrm{PERC(II)}}=\hat{y}_0-n^{-\frac{1}{2}}\sigma_y\hat{\sigma}\hat{H}^{-1}(1-\pi)$, $\hat{y}_{0\,\mathrm{PERC\text{-}T}}(\pi)=\hat{y}_0-n^{-\frac{1}{2}}\sigma_y\hat{\sigma}\hat{K}^{-1}(1-\pi)$, $\hat{y}_{0BC}=\hat{y}_0+n^{-\frac{1}{2}}\sigma_y\hat{\sigma}\hat{H}^{-1}(\hat{\rho})$ and $\hat{y}_{0ABC}=\hat{y}_0+n^{-\frac{1}{2}}\sigma_y\hat{\sigma}\hat{H}^{-1}(\hat{\rho}_a)$, where with $\hat{m}=\Phi^{-1}\{\hat{H}(0)\}=n^{-\frac{1}{2}}\frac{1}{6}\hat{\gamma}\gamma_y+O_p(n^{-\frac{3}{2}})$, we have $\hat{\rho}=\Phi(z_\pi+2\hat{m})$ and

$$\hat{\rho}_a=\Phi[\hat{m}+(\hat{m}+z_\pi)\{1-\hat{a}(\hat{m}+z_\pi)\}^{-1}].$$

Since the polynomials p_1 and q_1 appearing in Edgeworth expansions of H and K are not identical (see Subsection 3.3), then the percentile(I) critical point is not second-order correct. Neither is the percentile(II) point or the bias corrected point. However the percentile-t and accelerated bias corrected points are second-order correct.

Arguments leading to Edgeworth expansions of coverage probabilities in Subsection 2.4 may be pursued as before, the main change being that versions of S and T for the argument following (2.13) no longer have identical Edgeworth expansions. To appreciate why, observe that on the present occasion

$$S=n^{\frac{1}{2}}(\hat{y}_0-y_0)/(\hat{\sigma}\sigma_y)=n^{-\frac{1}{2}}\sigma_y^{-1}\sum_{i=1}^{n}y_i\xi_i^*+O(n^{-\frac{1}{2}})$$

where $\xi_i=\varepsilon_i/\sigma$, and $T=S+n^{-1}Uc_1$ where c_1 is a constant and U is as in Subsection 2.4. This entails

$$E(SU)=\sigma_y^{-1}\left\{n^{-1}\sum_{i=1}^{n}y_iE(\xi_i^4)-\tfrac{3}{2}\gamma n^{-1}\sum_{i=1}^{n}y_iE(\xi_i^3)-3n^{-1}\sum_{i=1}^{n}y_iE(\xi_i^2)\right\}+O(n^{-1})$$
$$=\sigma_y^{-1}(\kappa-\tfrac{3}{2}\gamma^2)+O(n^{-1}), \tag{3.1}$$

whereas the right-hand side had been $O(n^{-1})$ in the earlier case. It may be deduced from (3.1) that the one-sided percentile-t interval $(-\infty,\,\hat{y}_{0\,\mathrm{PERC\text{-}T}}(\pi)]$ has coverage probability

$$P\{y_0\leqq\hat{y}_{0\,\mathrm{PERC\text{-}T}}(\pi)\}=\pi-n^{-1}\tfrac{1}{6}(\kappa-\tfrac{3}{2}\gamma^2)z_\pi\{(\gamma_y-3\sigma_y^{-1})z_\pi^2-\gamma_y\}\phi(z_\pi)+O(n^{-\frac{3}{2}}), \tag{3.2}$$

and so has coverage error of order n^{-1}. The accelerated bias corrected one-sided interval also has coverage error of order n^{-1}, whereas one-sided percentile(I), percentile(II) and bias corrected intervals have coverage error of order $n^{-\frac{1}{2}}$. Formulae such as (3.2) are available uniformly in $0<\pi<1$, under the conditions of Theorem 2.1.

4. Multivariate, Multiparameter Regression

4.1. Introduction

In this section we take a quick look at the multivariate, multiparameter case, represented by the model

$$\mathbf{Y}_i=\boldsymbol{\alpha}+\mathbf{x}_i\boldsymbol{\beta}+\boldsymbol{\varepsilon}_i,\qquad 1\leqq i\leqq n.$$

Here, $\mathbf{Y}_i$, $\boldsymbol{\alpha}$ and $\boldsymbol{\varepsilon}_i$ are $p\times 1$ vectors, $\boldsymbol{\beta}$ is a $q\times 1$ vector, $\mathbf{x}_i$ is a $p\times q$ matrix, and the errors $\boldsymbol{\varepsilon}_i$ are assumed to be independent and identically distributed with zero mean and $p\times p$ variance matrix $\boldsymbol{\Sigma}$. Our aim is to give heuristic arguments which generalize results in Sects. 2 and 3. The arguments may be made rigorous, assuming sufficient moment and smoothness conditions on the errors $\boldsymbol{\varepsilon}_i$ and the design points $\mathbf{x}_i$, but at considerable algebraic expense. The trick in handling Edgeworth expansions for multivariate nonparametric bootstrap statistics is to smooth each resampled observation by adding to it an independent $N(\mathbf{0}, n^{-c}\mathbf{I})$ random variable, where c is fixed and arbitrarily large. The resulting statistic has a proper density function. Edgeworth expansions of this density may be developed.

Subsection 4.2 introduces notation, and Subsection 4.3 describes the bootstrap in a multivariate setting. Subsection 4.4 argues that the terms of order $n^{-\frac{1}{2}}$ in Edgeworth expansions of the non-Studentized statistic $\mathbf{S}_{n1}$ and the Studentized statistic $\mathbf{S}_{n2}$ are identical. This means that the multivariate version of the percentile(II) method produces confidence regions whose boundaries are second-order correct. Subsection 4.5 summarizes our main conclusions, which are multivariate versions of those reached for simple linear regression in Sect. 2.

4.2. Notation

Define $\bar{\mathbf{x}}=n^{-1}\sum\mathbf{x}_i$, $\bar{\mathbf{Y}}_i=n^{-1}\sum\mathbf{Y}_i$, $\boldsymbol{\Sigma}_x=n^{-1}\sum(\mathbf{x}_i-\bar{\mathbf{x}})^T(\mathbf{x}_i-\bar{\mathbf{x}})$, $\hat{\boldsymbol{\alpha}}=\bar{\mathbf{Y}}-\bar{\mathbf{x}}\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}=\boldsymbol{\Sigma}_x^{-1}n^{-1}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})^T(\mathbf{y}_i-\bar{\mathbf{Y}})$, $\hat{\boldsymbol{\varepsilon}}_i=\mathbf{Y}_i-(\hat{\boldsymbol{\alpha}}+\mathbf{x}_i\hat{\boldsymbol{\beta}})$, $\hat{\boldsymbol{\Sigma}}=n^{-1}\sum_{i=1}^{n}\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i^T$, $\mathbf{V}=n^{-1}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})^T\boldsymbol{\Sigma}(\mathbf{x}_i-\bar{\mathbf{x}})$, $\hat{\mathbf{V}}=n^{-1}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})^T\hat{\boldsymbol{\Sigma}}(\mathbf{x}_i-\bar{\mathbf{x}})$. The variance estimate $\hat{\boldsymbol{\Sigma}}$ is biased for $\boldsymbol{\Sigma}$ by an amount n^{-1}. We could correct for all or part of that bias by adjusting the factor n^{-1}, but to do so would not qualitatively affect our conclusions. If we were to assume that $\boldsymbol{\Sigma}$ was a diagonal matrix then our estimate would be a little different from $\hat{\boldsymbol{\Sigma}}$ but our conclusions would remain unchanged.

Note that $E(\hat{\boldsymbol{\beta}})=\boldsymbol{\beta}$ and $\operatorname{var}(\hat{\boldsymbol{\beta}})=n^{-1}\boldsymbol{\Sigma}_x^{-1}\mathbf{V}\boldsymbol{\Sigma}_x^{-1}$. Therefore two standardized versions of $\hat{\boldsymbol{\beta}}$ are

$$\mathbf{S}_{n1}=n^{\frac{1}{2}}(\boldsymbol{\Sigma}_x^{-1}\mathbf{V}\boldsymbol{\Sigma}_x^{-1})^{-\frac{1}{2}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}),\qquad \mathbf{S}_{n2}=n^{\frac{1}{2}}(\boldsymbol{\Sigma}_x^{-1}\hat{\mathbf{V}}\boldsymbol{\Sigma}_x^{-1})^{-\frac{1}{2}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}),$$

the second being "Studentized".

4.3. Nonparametric Bootstrap

The bootstrap argument described in paragraphs 3 and 4 of Subsection 1.2 may be applied without change in the present context, provided quantities are interpreted in a vector setting. Let $\{\boldsymbol{\varepsilon}_1^*, \ldots, \boldsymbol{\varepsilon}_n^*\}$ denote a resample drawn with replacement from the residuals $\{\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n\}$, put $\mathbf{Y}_i^* = \hat{\boldsymbol{\alpha}} + \mathbf{x}_i \hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}_i^*$, define $\hat{\boldsymbol{\alpha}}^*$, $\hat{\boldsymbol{\beta}}^*$ on replacing $(\mathbf{x}_i, \mathbf{Y}_i)$ by $(\mathbf{x}_i, \mathbf{Y}_i^*)$ in formulae for $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, and let $\hat{\boldsymbol{\varepsilon}}_i^* = \mathbf{Y}_i^* - (\hat{\boldsymbol{\alpha}}^* + \mathbf{x}_i \hat{\boldsymbol{\beta}}^*)$,

$$\hat{\boldsymbol{\Sigma}}^* = n^{-1} \sum_{i=1}^{n} \hat{\boldsymbol{\varepsilon}}_i^* \hat{\boldsymbol{\varepsilon}}_i^{*T}, \qquad \hat{\mathbf{V}}^* = n^{-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})^T \hat{\boldsymbol{\Sigma}}^* (\mathbf{x}_i - \bar{\mathbf{x}}).$$

The vector version of S_{n2}^* is $\mathbf{S}_{n2}^* = n^{\frac{1}{2}} (\boldsymbol{\Sigma}_x^{-1} \hat{\mathbf{V}}^* \boldsymbol{\Sigma}_x^{-1})^{-\frac{1}{2}} (\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}})$. Percentile(II) and percentile-t regions are $\hat{\boldsymbol{\beta}} - \hat{\mathscr{S}}_\pi$, $\hat{\boldsymbol{\beta}} - n^{-\frac{1}{2}} \hat{\boldsymbol{\Sigma}}^{\frac{1}{2}} \hat{\mathscr{T}}_\pi$ respectively, where $\hat{\mathscr{S}}_\pi$, $\hat{\mathscr{T}}$ are regular regions containing just πB of B simulated values of $\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}}$, $\mathbf{S}_{n2}^*$ respectively. Both have nominal coverage equal to π. Percentile-t requires more numerical effort than either percentile method, since it involves computation of the inverse and square root of a new matrix for each resample.

Bootstrap confidence intervals for individual components of $\boldsymbol{\beta}$ may be described in terms of distribution functions, as follows. If θ, $\hat{\theta}$ denote j'th components of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$ respectively, then $E(\hat{\theta}) = \theta$ and $\operatorname{var}(\hat{\theta}) = n^{-1} \sigma^2$, where σ^2 is the (j, j)'th component of $\boldsymbol{\Sigma}_x^{-1} \mathbf{V} \boldsymbol{\Sigma}_x^{-1}$. Let $\hat{\theta}$ be the j'th component of $\hat{\boldsymbol{\beta}}$, and $\hat{\sigma}^2$, $\hat{\sigma}^{*2}$ be respectively the (j,j)'th components of $\boldsymbol{\Sigma}_x^{-1} \hat{\mathbf{V}} \boldsymbol{\Sigma}_x^{-1}$, $\boldsymbol{\Sigma}_x^{-1} \hat{\mathbf{V}}^* \boldsymbol{\Sigma}_x^{-1}$. Write G, H and K for distribution functions of $\hat{\theta}$, $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\sigma$ and $n^{\frac{1}{2}}(\hat{\theta} - \theta)/\hat{\sigma}$, respectively, and $\hat{G}$, $\hat{H}$ and $\hat{K}$ for distribution functions of $\hat{\theta}^*$, $n^{\frac{1}{2}}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}$ and $n^{\frac{1}{2}}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^*$, conditional on $\mathscr{X}$ in each of the last three cases. If we change β to θ in formulae (1.2) and (1.4), we have formulae for endpoints $\hat{\theta}(\pi)$ of one-sided bootstrap confidence intervals having nominal coverage π.

4.4. Terms of Order $n^{-\frac{1}{2}}$ in Edgeworth Expansions

We begin by outlining theory for Edgeworth expansions of multivariate densities. Let $\mathbf{S}_n$ denote either $\mathbf{S}_{n1}$ or $\mathbf{S}_{n2}$ (both vectors of length q), let $\mathbf{v} = (v^{(1)}, \ldots, v^{(q)})^T$ be a q-vector of nonnegative integers, let $\mathbf{x} = (x^{(1)}, \ldots, x^{(q)})^T$ be a general q-vector, and define $|\mathbf{v}| = \sum_j |v^{(j)}|$, $\|\mathbf{x}\| = (\sum_j x^{(j)2})^{\frac{1}{2}}$, $\mathbf{v}! = \prod_j (v^{(j)}!)$, $\mathbf{x}^{\mathbf{v}} = \prod_j (x^{(j)})^{v^{(j)}}$ and $D^{\mathbf{v}} = \prod_j (\partial / \partial x^{(j)})^{v^{(j)}}$. Moments $\mu_{\mathbf{v}} = E(\mathbf{S}_n^{\mathbf{v}})$ and cumulants $\chi_{\mathbf{v}}$ (both scalars) are determined by the formulae

$$\psi(\mathbf{t}) = E\{\exp(i\mathbf{t}^T \mathbf{S}_n)\} \approx \sum_{\mathbf{v}} \mu_{\mathbf{v}} (i\mathbf{t})^{\mathbf{v}} / \mathbf{v}!, \qquad \log \psi(\mathbf{t}) \approx \sum_{\mathbf{v}} \chi_{\mathbf{v}} (i\mathbf{t})^{\mathbf{v}} / \mathbf{v}!.$$

In the regression case, cumulants enjoy the expansions

$$\chi_{\mathbf{v}} \approx n^{|\mathbf{v}|/2} \sum_{j \geqq |\mathbf{v}| - 1} a_{\mathbf{v}j} n^{-j}, \qquad |\mathbf{v}| \geqq 1,$$

where $a_{\mathbf{v}0} = 0$ if $|\mathbf{v}| = 1$. (Compare Withers [19].) The fact that $\mathbf{S}_n$ has been standardized to have identity asymptotic variance matrix means that when $|\mathbf{v}| = 2$, $a_{\mathbf{v}1} = 1$ if one component of $\mathbf{v}$ equals 2 and $a_{\mathbf{v}1} = 0$ otherwise. Thus,

$$\psi(\mathbf{t}) = \exp(-\tfrac{1}{2} \|\mathbf{t}\|^2) [1 + n^{-\frac{3}{2}} \{ \sum_{|\mathbf{v}|=1} a_{\mathbf{v}1} (i\mathbf{t})^{\mathbf{v}} + \sum_{|\mathbf{v}|=3} a_{\mathbf{v}3} (i\mathbf{t})^{\mathbf{v}} / \mathbf{v}! \} + O(n^{-1})].$$

Inversion of this Fourier transform indicates that the density f_n of $\mathbf{S}_n$ satisfies

$$f_n(\mathbf{x}) = \phi(\mathbf{x})\{1 + n^{-\frac{1}{2}} r(\mathbf{x})\} + O(n^{-1}),$$

where the polynomial r is given by

$$r(\mathbf{x}) = \sum_{|\boldsymbol{v}|=1} a_{\boldsymbol{v}1} H_{\boldsymbol{v}}(x) + \sum_{|\boldsymbol{v}|=3} a_{\boldsymbol{v}2} H_{\boldsymbol{v}}(\mathbf{x})/\boldsymbol{v}!. \tag{4.1}$$

(The generalized Hermite polynomial $H_{\boldsymbol{v}}$, defined by $H_{\boldsymbol{v}}(\mathbf{x})\,\phi(\mathbf{x}) = (-D)^{\boldsymbol{v}}\,\phi(\mathbf{x})$, has Fourier transform $(i\mathbf{t})^{\boldsymbol{v}} \exp(-\frac{1}{2}\|\mathbf{t}\|^2)$.)

It may be proved after lengthy algebra that the *same* polynomial r emerges from formula (4.1) in both the cases $\mathbf{S}_n = \mathbf{S}_{n1}$ and $\mathbf{S}_n = \mathbf{S}_{n2}$. This is a multivariate, density version of the fact that $p_1 = q_1$ in Theorem 2.1. That result was the key to good performance of the percentile (II) method in simple linear regression, and the fact that it persists in a multivariate setting indicates that percentile(II) will perform well here, too.

4.5. *Properties of Bootstrap Confidence Intervals*

For the sake of simplicity we shall confine attention to confidence intervals for individual components of slope or of vector means.

(i) Percentile(II) confidence intervals for components of $\boldsymbol{\beta}$ are second-order correct. Furthermore, one-sided percentile(II) intervals have coverage error $O(n^{-1})$, compared to $O(n^{-\frac{1}{2}})$ in most other statistical problems. These results may be proved from the fact that Edgeworth expansions for $\mathbf{S}_{n1}$ and $\mathbf{S}_{n2}$ agree in terms of order $n^{-\frac{1}{2}}$ – see the end of the previous section. As in the case of simple linear regression discussed in Sect. 2, two-sided percentile(II) confidence interval have coverage error $O(n^{-1})$.

(ii) Percentile-t confidence intervals for components of $\boldsymbol{\beta}$ are second-order correct. They have coverage error $O(n^{-\frac{3}{2}})$ in the one-sided case, and $O(n^{-2})$ in the two-sided case. These results may be proved very much as in Subsections 2.5 and 2.6. The arguments are more tedious in the present multivariate setting, but not conceptually more difficult.

(iii) In the case of confidence interval for components of vector means or for intercepts, rather than for slope, the above virtuous properties of percentile(II) and percentile-t evaporate. Percentile(II) intervals fail to be second-order correct, and percentile-t intervals have coverage error $O(n^{-1})$.

5. Proof of Theorem 2.2

We shall need the following notation, in addition to that introduced in Sects. 1 and 2. Use superscripts to denote elements of vectors. For d-vectors $\mathbf{x}$, write $\|\mathbf{x}\| = (\sum_i x^{(i)2})^{\frac{1}{2}}$; for d-vectors $\boldsymbol{v}$ with nonnegative integer components, write $|\boldsymbol{v}|$

$=\sum_j |v^{(j)}|$ and $D^v=\prod_j (\partial/\partial t^{(j)})^{v(j)}$ (a differential operator); let $\phi_{\mu,\Sigma}$ and $\Phi_{\mu,\Sigma}$ be the density and distribution function of the bivariate $N(\mu, \Sigma)$ distribution; and let $C, C_1, C_2, \ldots$ denote positive generic constants. Put $v_j=\sigma_x^{-1}(x_j-\bar{x})$, $U_j^{(1)}=\xi_j=\varepsilon_j/\sigma$, $U_j^{(2)}=v_j\xi_i$, $U_j^{(3)}=\xi_j^2-1$, $\mathbf{U}_j=(U_j^{(1)}, U_j^{(2)}, U_j^{(3)})^T$ and $\bar{\mathbf{U}}=n^{-1}\sum_j \mathbf{U}_j$. Observe that

$$S=n^{\frac{1}{2}}(\hat{\beta}-\beta)\sigma_x/\hat{\sigma}=n^{\frac{1}{2}}\,\bar{U}^{(2)}(1+\bar{U}^{(3)}-\bar{U}^{(1)2}-\bar{U}^{(2)2})^{-\frac{1}{2}}.$$

Techniques similar to those employed by Bhattacharya and Ghosh [4], although requiring a little elaboration because the summands comprising the mean $\bar{\mathbf{U}}$ are weighted i.i.d. vectors with different weights, show that under the conditions of Theorem 2.1,

$$P(S\leqq z)=\Phi(z)+\sum_{j=1}^{3} n^{-j/2}\,q_j(z)\,\phi(z)+O(n^{-2}) \tag{5.1}$$

uniformly in $-\infty<z<\infty$, where $-q_1(z)=\frac{1}{6}\gamma\gamma_x(z^2-1)$ and q_2, q_3 are polynomials whose coefficients depend on the first five moments of the error distribution and of the sequence of design points. Polynomial q_j is of degree $3j-1$, and odd/even indexed q_j's are even/odd functions respectively.

Define $U_j^{*(1)}=\xi_j^*=\varepsilon_j^*/\hat{\sigma}$, $U_j^{*(2)}=v_j\xi_j^*$, $U_j^{*(3)}=\xi_j^{*2}-1$, $\mathbf{U}_j^*=(U_j^{*(1)}, U_j^{*(2)}\ U_j^{*(3)})^T$ and $\bar{\mathbf{U}}^*=n^{-1}\sum_j \mathbf{U}_j^*$. Observe that

$$S^*=n^{\frac{1}{2}}(\hat{\beta}^*-\hat{\beta})\sigma_x/\hat{\sigma}^*=n^{\frac{1}{2}}\,\bar{U}^{*(2)}(1+\bar{U}^{*(3)}-\bar{U}^{*(1)^2}-\bar{U}^{*(2)^2})^{-\frac{1}{2}}.$$

Let $\hat{q}_j(1\leqq j\leqq 3)$ denote the version of q_j in which moments of the form $E(\xi_1^k)$ $(3\leqq k\leqq 5)$ appearing in coefficients are replaced by $n^{-1}\sum_j(\hat{\varepsilon}_j/\hat{\sigma})^k$. Our first goal is to establish a bootstrap version of (5.1). Let $\mathscr{C}=\mathscr{C}_n$ denote the class of all possible samples $\mathscr{X}=\{(x_1, Y_1), \ldots, (x_n, Y_n)\}$, given the design points $x_1, \ldots, x_n$. If $\mathscr{E}\subseteq\mathscr{C}$, write $P(\mathscr{E})$ for $P(\mathscr{X}\in\mathscr{E})$. Put

$$\Delta(z)=P(S^*\leqq z\,|\,\mathscr{X})-\left\{\Phi(z)+\sum_{j=1}^{3} n^{-j/2}\,\hat{q}_j(z)\,\phi(z)\right\}.$$

Theorem 5.1. *Under the conditions of Theorem 2.2, there exists $\mathscr{E}\subseteq\mathscr{C}$ with $P(\mathscr{E})=1+O(n^{-2})$ and such that*

$$\sup_{\mathscr{X}\in\mathscr{E}}\ \sup_{-\infty<z<\infty} |\Delta_n(z)|\leqq Cn^{-2}.$$

We shall prove Theorem 5.1 via an Edgeworth expansion for the trivariate distribution of $n^{\frac{1}{2}}\,\bar{\mathbf{U}}^*$. The next paragraph defines that expansion.

Given a 3-vector $\mathbf{v}$ of nonnegative integers, let $\hat{\chi}_{\mathbf{v},j}$ be the $\mathbf{v}$'th cumulant and $\hat{\mathbf{V}}_j$ the variance matrix of $\mathbf{U}_j^*$, both conditional on $\mathscr{X}$. Put

$$\hat{\chi}_{\mathbf{v}}=n^{-1}\sum_{j=1}^{n}\hat{\chi}_{\mathbf{v},j} \quad\text{and}\quad \hat{\mathbf{V}}=n^{-1}\sum_{j=1}^{n}\hat{\mathbf{V}}_j=\begin{matrix}1 & 0 & \hat{\gamma}\\ 0 & 1 & 0\\ \hat{\gamma} & 0 & \hat{\kappa}+2\end{matrix}. \tag{5.2}$$

Let $P_j(-\Phi_{\mathbf{0},\hat{\mathbf{V}}}\colon\{\hat{\chi}_{\mathbf{v}}\})$ denote the signed measure whose density is $P_j(-\phi_{\mathbf{0},\hat{\mathbf{V}}}\colon\{\hat{\chi}_{\mathbf{v}}\})$, defined by Bhattacharya and Rao [6, pp. 53–54]. Write Q_n for the random measure induced on $\mathbb{R}^3$ by $n^{\frac{1}{2}}\bar{\mathbf{U}}^*$, conditional on $\mathscr{X}$. For any real number z, let $\mathscr{S}(z)$ denote that set of values $(u^{(1)}, u^{(2)}, u^{(3)})\in\mathbb{R}^3$ such that

$$u^{(2)}(1+n^{-\frac{1}{2}}u^{(3)}-n^{-1}u^{(1)2}-n^{-1}u^{(2)2})^{-\frac{1}{2}}\leqq z$$

and the left-hand side is real-valued. The major step in proving Theorem 5.1 is deriving:

Proposition 5.2. *Under the conditions of Theorem* 4.1, *there exists* $\mathscr{E}\subseteq\mathscr{C}$ *with* $P(\mathscr{E})=1+O(n^{-2})$ *and such that*

$$\sup_{\mathscr{X}\in\mathscr{E}}\sum_{-\infty<z<\infty}\left|\int_{\mathscr{S}(z)} d\left[Q_n-\Phi_{\mathbf{0},\hat{\mathbf{V}}}-\sum_{j=1}^{4}n^{-j/2}P_j(-\Phi_{\mathbf{0},\hat{\mathbf{V}}}\colon\{\hat{\chi}_{\mathbf{v}}\})\right]\right|\leqq Cn^{-2}.$$

Proof of Proposition 5.2. Let $\mu_k=E(\varepsilon_1^k)$ be the k'th moment of the error distribution, put $\hat{\mu}_k=n^{-1}\sum_j(\hat{\varepsilon}_j/\hat{\sigma})^k=n^{-1}\hat{\sigma}^{-k}\sum_j(\varepsilon_i-\bar{\varepsilon}-v_j n^{-1}\sum_i v_i\varepsilon_i)^k$, and let $\mathscr{E}_1(\eta)$ be the set of all samples $\mathscr{X}$ such that $|\hat{\mu}_k-\mu_k|\leqq\eta$ for $1\leqq k\leqq 12$. Using the boundedness of the errors ε_j we may prove that $P\{\mathscr{E}_1(\eta)\}=1+O(n^{-2})$ (in fact, $1+O(n^{-\lambda})$ for each $\lambda>0$) for each $\eta>0$. Let $\mathbf{V}$ be the 3×3 matrix with 1, 1, $\kappa+2$ down the main diagonal, γ in the top left- and bottom right-hand corners, and zeros elsewhere; compare $\hat{\mathbf{V}}$ defined at (5.2). The determinant of $\mathbf{V}$ equals $\kappa+2-\gamma^2=E(\xi_1)^2E\{(\xi_1^2-1)^2\}-[E\{\xi_1(\xi_1^2-1)\}]^2>0$, using the Cauchy-Schwarz inequality and the fact that the distribution of ξ_1 is nonsingular. Therefore all the eigenvalues of $\mathbf{V}$ lie in an interval $(2\zeta^{-1}, \frac{1}{2}\zeta)$ for some $2<\zeta<\infty$. If η is sufficiently small and $\mathscr{X}\in\mathscr{E}_1(\eta)$ then all eigenvalues of $\hat{\mathbf{V}}$ lie within (ζ^{-1}, ζ).

Put

$$\mathbf{U}_j^{\dagger\dagger}=\begin{cases}\hat{\mathbf{V}}^{-\frac{1}{2}}\mathbf{U}_j^* & \text{if } \|\hat{\mathbf{V}}^{-\frac{1}{2}}\mathbf{U}_j^*\|\leqq n^{\frac{1}{2}}\\ 0 & \text{otherwise,}\end{cases}$$

$\mathbf{U}_j^\dagger=\mathbf{U}_j^{\dagger\dagger}-E(\mathbf{U}_j^{\dagger\dagger}|\mathscr{X})$ and $\bar{\mathbf{U}}_j^\dagger=n^{-1}\sum_j\mathbf{U}_j^\dagger$. Let $\hat{\chi}_{\mathbf{v},j}^\dagger$ be the $\mathbf{v}$'th cumulant and $\hat{\mathbf{V}}_j^\dagger$ the variance matrix of $\mathbf{U}_j^\dagger$, both conditional on $\mathscr{X}$. Set $\hat{\chi}_{\mathbf{v}}^\dagger=n^{-1}\sum_j\hat{\chi}_{\mathbf{v},j}$ and $\hat{\mathbf{V}}^\dagger=n^{-1}\sum_j\hat{\mathbf{V}}_j^\dagger$. Then $\hat{\mathbf{V}}^\dagger\to\mathbf{I}$ in probability, and it may be proved that if η is small and n large then $\hat{\mathbf{V}}^\dagger$ has all its eigenvalues within $(\frac{1}{2}, 2)$ whenever $\mathscr{X}\in\mathscr{E}_1(\eta)$.

Write $Q_n^\dagger$ for the random measure induced on $\mathbb{R}^3$ by $n^{\frac{1}{2}}\bar{\mathbf{U}}^\dagger$, conditional on $\mathscr{X}$. Put

$$H_n = Q_n - \sum_{j=0}^{4} n^{-j/2} P_j(-\Phi_{\mathbf{0},\hat{\mathbf{V}}}:\{\hat{\chi}_\nu\}),\ H_n^\dagger = Q_n^\dagger - \sum_{j=0}^{4} n^{-j/2} P_j(-\Phi_{\mathbf{0},\hat{\mathbf{V}}^\dagger}:\{\hat{\chi}_\nu^\dagger\}).$$

Let $\mathscr{S}(z)$ be as in the statement of Proposition 5.2, and $\mathscr{S}^\dagger(z)$ equal the set $\{\hat{\mathbf{V}}^{-\frac{1}{2}}\mathbf{x} - n^{-\frac{1}{2}}\sum_j E(\mathbf{U}_j^{\dagger\dagger}|\mathscr{X}): \mathbf{x}\in\mathscr{S}(z)\}$. If each $\|\hat{\mathbf{V}}^{-\frac{1}{2}}\mathbf{U}_j^*\| \leqq n^{\frac{1}{2}}$ then the event $n^{\frac{1}{2}}\bar{\mathbf{U}}^*\in\mathscr{S}(z)$ is equivalent to $n^{\frac{1}{2}}\bar{\mathbf{U}}^\dagger\in\mathscr{S}^\dagger(z)$. If $\mathscr{X}\in\mathscr{E}_1(\eta)$ and η is small then conditional on $\mathscr{X}$, the chance that $\|\hat{\mathbf{V}}^{-\frac{1}{2}}\mathbf{U}_j^*\| > n^{\frac{1}{2}}$ for some $1\leqq j\leqq n$ is less than Cn^{-2}. From this observation, and arguing as in [5, pp. 208–209], we deduce that if n is large, η is small and $\mathscr{X}\in\mathscr{E}_1(\eta)$ then

$$\sup_{-\infty<z<\infty}\left|\int_{\mathscr{S}(z)} dH_n - \int_{\mathscr{S}^\dagger(z)} dH_n^\dagger\right| \leqq Cn^{-2}.$$

Therefore Proposition 5.2 will follow if we prove that for a collection $\mathscr{E}\subseteq\mathscr{C}$ with $P(\mathscr{E}) = 1 + O(n^{-2})$,

$$\sup_{\mathscr{X}\in\mathscr{E}} s_n(\mathscr{X}) \leqq Cn^{-2} \tag{5.3}$$

where

$$s_n(\mathscr{X}) \equiv \sup_{-\infty<z<\infty} \left|\int_{\mathscr{S}^\dagger(z)} dH_n^\dagger\right|.$$

Derivation of (5.3) is along lines in [6, pp. 210–214]. We give only an outline. Assuming $\mathscr{X}\in\mathscr{E}_1(\eta)$ for small η, and using Theorem 9.11 of Bhattacharya and Rao [6] in place of Theorem 9.10 in a derivation of an analogue of their (20.21), their argument is straightforward to the foot of page 211. Thus we deduce the existence of a probability measure L on $\mathbb{R}^3$, with support confined to the closed sphere of unit radius centred at the origin, and whose Fourier-Stieltjes transform l satisfies $|(D^\nu l)(\mathbf{t})| \leqq C\exp(-\|\mathbf{t}\|^{\frac{1}{2}})$ for all $\mathbf{t}\in\mathbb{R}^3$ and all 3-vectors ν with nonnegative integer components satisfying $|\nu|\leqq 10$. Given $0<\delta<1$, put $L_\delta(E) = L(\delta^{-1}E)$ for Borel sets E, and let l_δ be the Fourier-Stieltjes transform of L_δ. Let $q_n^\dagger$ be the characteristic function associated with measure $Q_n^\dagger$. Then if $\mathscr{X}\in\mathscr{E}_1(\eta)$ and η is small,

$$\begin{aligned} s_n(\mathscr{X}) \leqq & \sup_{-\infty<z<\infty} \sum_{j=0}^{7} n^{-j/2} \int_{\{\partial\mathscr{S}^\dagger(z)\}^{2\delta}} |P_j(-\phi_{\mathbf{0},\hat{\mathbf{V}}^\dagger}:\{\hat{\chi}_\nu^\dagger\})(\mathbf{y})|\,d\mathbf{y} \\ & + C_1 \max_{\mathbf{0}\leqq\boldsymbol{\alpha}\leqq\boldsymbol{\beta},\,|\boldsymbol{\beta}|\leqq 10} t_n(\boldsymbol{\alpha},\boldsymbol{\beta}) + C_1 n^{-2}, \end{aligned} \tag{5.4}$$

where C_1 does not depend on δ and

$$t_n(\boldsymbol{\alpha},\boldsymbol{\beta}) = \int_{\|\mathbf{t}\|>C_2 n^{\frac{1}{2}}} |\{D^{\boldsymbol{\beta}-\boldsymbol{\alpha}} q_n^\dagger(\mathbf{t})\}\{D^{\boldsymbol{\alpha}} h_\delta(\mathbf{t})\}|\,d\mathbf{t}.$$

Next we select δ, and bound $t_n(\boldsymbol{\alpha},\boldsymbol{\beta})$.

Lemma 5.3. *Take $\delta = n^{-2}$. There exists $\mathscr{E}_2 \subseteq \mathscr{C}$ with $P(\mathscr{E}_2) = 1 + O(n^{-2})$ and*

$$\sup_{\mathscr{X} \in \mathscr{E}_2} \sup_{\mathbf{0} \leqq \boldsymbol{\alpha} \leqq \boldsymbol{\beta}, |\boldsymbol{\beta}| \leqq 10} t_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) \leqq C n^{-2}.$$

Proof of Lemma 5.3. Put

$$q_{nj}(\mathbf{t}) = E\{\exp(n^{-\frac{1}{2}} i \mathbf{t}^T \mathbf{U}_j^\dagger) | \mathscr{X}\}.$$

Then $q_n^\dagger = \prod_j q_{nj}$, from which it follows that for large n and all $\mathbf{t}$,

$$|D^\gamma q_n^\dagger(\mathbf{t})| \leqq n^{|\gamma|} \sum^{(|\gamma|)} r(\mathbf{t}; j_1, \ldots, j_{|\gamma|}),$$

where $r(\cdot\,; j_1, \ldots, j_k)$ denotes the product of $|q_{nj}|$ over all values $1 \leqq j \leqq n$ excluding $j_1, \ldots, j_k$, and $\sum^{(k)}$ denotes summation over distinct k-tuples $j_1, \ldots, j_k$ with each $1 \leqq j_i \leqq n$. (Note that $\|\mathbf{U}_j^\dagger\| \leqq 2n^{\frac{1}{2}}$.) Therefore if $\mathbf{0} \leqq \boldsymbol{\alpha} \leqq \boldsymbol{\beta}$ and $|\boldsymbol{\beta}| \leqq 10$,

$$t_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) \leqq C_3 n^{10} \sum^{(10)} \int_{\|\mathbf{t}\| > C_2 n^{\frac{1}{2}}} r(\mathbf{t}; j_1, \ldots, j_{10}) \exp(-\|\delta \mathbf{t}\|^{\frac{1}{2}})\, d\mathbf{t}. \tag{5.5}$$

Since $P(\|\hat{\mathbf{V}}^{-\frac{1}{2}} \mathbf{U}_j^*\| > n^{\frac{1}{2}} | \mathscr{X}) \leqq C_4 n^{-1}$ uniformly in $1 \leqq j \leqq n$ and $\mathscr{X} \in \mathscr{E}_1(\eta)$, for small η, then $|q_{nj}(\mathbf{t})| \leqq |a_{nj}(n^{-\frac{1}{2}} \hat{\mathbf{V}}^{-\frac{1}{2}} \mathbf{t})| + C_4 n^{-1}$, where

$$a_{nj}(\mathbf{t}) = E\{\exp(i \mathbf{t}^T \mathbf{U}_j^*) | \mathscr{X}\}.$$

Therefore if $c_1(\mathbf{t}; j_1, \ldots, j_k)$ denotes the product of $|a_{nj}(\mathbf{t})| + C_4 n^{-1}$ over $1 \leqq j \leqq n$ excluding $j_1, \ldots, j_k$, then the right-hand side of (5.5) is not greater than

$$C_5 n^{10} \sum^{(10)} \int_{\|\mathbf{t}\| > C_6 n^{\frac{1}{2}}} c_1(n^{-\frac{1}{2}} \mathbf{t}; j_1, \ldots, j_{10}) \exp(-C_7 \|\delta \mathbf{t}\|^{\frac{1}{2}})\, d\mathbf{t},$$

provided $\mathscr{X} \in \mathscr{E}_1(\eta)$ and η is small. But

$$|a_{nj}(\mathbf{t})| = |b_n\{(t^{(1)} + v_j t^{(2)}) \hat{\sigma}^{-1}, t^{(3)} \hat{\sigma}^{-2}\}|,$$

where $b_n(t_1, t_2) = n^{-1} \sum_k \exp(i t_1 \hat{\varepsilon}_k + i t_2 \hat{\varepsilon}_k^2)$. Hence

$$t_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) \leqq u_n = C_5 n^{10} \sum^{(10)} \int_{\|\mathbf{t}\| > C_8 n^{\frac{1}{2}}} c_1(n^{-\frac{1}{2}} \mathbf{t}; j_1, \ldots, j_{10}) \exp(-C_9 \|\delta \mathbf{t}\|^{\frac{1}{2}})\, d\mathbf{t},$$

where $c(\mathbf{t}; j_1, \ldots, j_k)$ denotes the product of $|b_n(t^{(1)} + v_j t^{(2)}, t^{(3)})| + C_4 n^{-1}$ over $1 \leqq j \leqq n$ excluding $j_1, \ldots, j_k$.

Let $\mathscr{E}_3$ denote the class of all samples $\mathscr{X}$ such that $u_n \leqq n^{-2}$. We shall complete the proof of Lemma 5.3 by demonstrating that $P(\mathscr{E}_3) = O(n^{-2})$. For that it suffices, by Markov's inequality, to show that $E(u_n) = O(n^{-4})$. Now, the series $\sum^{(10)}$ has $O(n^{10})$ different summands. Let $(k_1, \ldots, k_{10})$ denote that value of $(j_1, \ldots, j_{10})$ for which the summand has largest expectation. We must prove that

$$n^{24} \int_{\|\mathbf{t}\| > C_8 n^{\frac{1}{2}}} E\{c(n^{-\frac{1}{2}} \mathbf{t}; k_1, \ldots, k_{10})\} \exp(-C_9 \|\delta \mathbf{t}\|^{\frac{1}{2}})\, d\mathbf{t} = O(1). \tag{5.6}$$

Let $b(t_1, t_2)=E\{\exp(it_1 \varepsilon_1+it_2 \varepsilon_1^2)\}$, $b_j(\mathbf{t})=b(t^{(1)}+v_j t^{(2)}, t^{(3)})$ and $b_{nj}(\mathbf{t})=b_n(t^{(1)}+v_j t^{(2)}, t^{(3)})$. Fix $\Delta\in(0, \frac{1}{2})$ and put $I_j(\mathbf{t})=1$ if both $|b_{nj}(\mathbf{t})-b_j(\mathbf{t})|\leqq\Delta$ and $|b_j(\mathbf{t})|\leqq 1-2\Delta$, $I_j(\mathbf{t})=0$ otherwise, and $N(\mathbf{t})=\sum_j I_j(\mathbf{t})$. Then

$$c(\mathbf{t};j_1,\ldots,j_{10})\leqq C_{10}\exp\{-C_{11}N(\mathbf{t})\}$$

for all $\mathbf{t}$, all $j_1,\ldots,j_{10}$ and all samples $\mathscr{X}$. Recall that for some $0<\zeta<1$ and all large n, the number of j's such that $v_j>\zeta$ is $\geqq n^{\zeta}$, as is the number of j's such that $v_j<-\zeta$. Therefore if $\|\mathbf{t}\|>C_8$ then the number of indices j such that either $|t^{(1)}+v_j t^{(2)}|>C_{12}=\zeta C_8/3^{\frac{1}{2}}$ or $|t^{(3)}|>C_8/3^{\frac{1}{2}}$, is $\geqq n^{\zeta}$. Then if $\|t\|>c_8$, the set $\mathscr{J}(\mathbf{t})$ of indices j such that $(t^{(1)}+v_j t^{(2)})^2+t^{(3)2}\geqq C_{12}^2$ has $\geqq n^{\zeta}$ elements. Since the distribution of ε_1 is nonsingular then if we define Δ by

$$1-2\Delta=\sup_{t_1^2+t_2^2\geqq C_{12}^2}|b(t_1, t_2)|,$$

we have $0<\Delta<\frac{1}{2}$. Let $K_j(\mathbf{t})=1$ if $|b_{nj}(\mathbf{t})-b_j(\mathbf{t})|\leqq\Delta$, and $K_j(\mathbf{t})=0$ otherwise. Put $M(\mathbf{t})=\sum_{j\in\mathscr{J}(\mathbf{t})}K_j(\mathbf{t})$. Then $M(\mathbf{t})\leqq N(\mathbf{t})$, so (5.6) will follow if

$$n^{26}\int_{\|\mathbf{t}\|>C_8}E[\exp\{-C_{11}M(\mathbf{t})\}]\exp(-C_9\|\delta n^{\frac{1}{2}}\mathbf{t}\|^{\frac{1}{2}})\,d\mathbf{t}=O(1). \tag{5.7}$$

It may be proved via Markov's inequality, and after lengthy algebra, that

$$\sup_{(t_1, t_2)\in\mathbb{R}^2}P\{|b_n(t_1, t_2)-b(t_1, t_2)|>\Delta\}=O(n^{-\lambda})$$

for all $\lambda>0$. This implies that

$$\begin{aligned}P\{M(\mathbf{t})<\tfrac{1}{2}n^{\zeta}\}&\leqq P[\sum_{j\in\mathscr{J}(\mathbf{t})}\{1-K_j(\mathbf{t})\}\geqq\tfrac{1}{2}n^{\zeta}]\\&\leqq 2n^{-\zeta}\sum_{j\in\mathscr{J}(\mathbf{t})}E\{1-K_j(\mathbf{t})\}=O(n^{-\lambda})\end{aligned}$$

for all $\lambda>0$, uniformly in $\|\mathbf{t}\|>C_8$. Result (5.7) is immediate. □

Lemma 5.3 takes care of the second term on the right-hand side of (5.4). To accommodate the first term, use an argument based on formulae (9.12) and (14.74) of [6, pp. 72 and 133] to deduce that the quantity is dominated by $\sum_{0\leqq j\leqq 7}w_j$, where

$$w_j=C_1\int_{\{\partial\mathscr{S}^{\dagger}(z)\}^{2\delta}}(1+\|\mathbf{x}\|^{3j})\exp\{-\tfrac{1}{2}\mathbf{x}^T(\hat{\mathbf{V}}^{\dagger})^{-1}\mathbf{x}\}\,d\mathbf{x}.$$

After considerable algebra we may deduce that if $\mathscr{X}\in\mathscr{E}(\eta)$ and η is sufficiently small then for some $C>0$,

$$\begin{aligned} w_j &\leqq C_2 \int_{\{\partial\mathscr{S}(z)\}^{c\delta}} \exp(-C_3\|\mathbf{x}\|^2)\,d\mathbf{x}+C_2 n^{-2} \\ &\leqq C_4 n^{-2}, \end{aligned} \tag{5.8}$$

remembering that $\delta=n^{-2}$. Taking $\mathscr{E}=\mathscr{E}_1(\eta)\cap\mathscr{E}_2$, with $\mathscr{E}_2$ as in Lemma 5.3, we may deduce (5.3) from (5.8) and Lemma 5.3, completing the proof of Proposition 5.2. □

The transition from Proposition 5.2 to Theorem 5.1 is relatively straightforward, being made as on pages 443–444 of Bhattacharya and Ghosh [4]. Having derived Theorem 5.1, use arguments in steps (ii)–(vi) of the proof of Theorem 2.1 of [14] to obtain the present Theorem 2.2. In outline, that argument first inverts the Edgeworth expansion in Theorem 5.1 to obtain a Cornish-Fisher expansion: for some $\mathscr{E}\subseteq\mathscr{C}$ with $P(\mathscr{E})=1+O(n^{-2})$, and some $\delta>0$,

$$\begin{aligned} &\sup_{\mathscr{X}\in\mathscr{E}} \sup_{n^{-2-\delta}\leqq\pi\leqq 1-n^{-2-\delta}} \left|\hat\beta_{\text{PERC-T}}(\pi)-\left[\hat\beta+n^{-\frac12}\sigma_x^{-1}\hat\sigma\left\{z_\pi+\sum_{j=1}^{3} n^{-j/2}\hat s_j(z_\pi)\right\}\right]\right| \\ &\quad\leqq C n^{-\frac52}. \end{aligned} \tag{5.9}$$

Here $\hat s_j(z)$ is the bootstrap version of the polynomial s_j appearing in a Cornish-Fisher expansion of the exact quantile $\hat\beta_{\text{exact}}(\pi)$; for example, $\hat s_1=\hat q_1(z)=-\frac16\hat\gamma\gamma_x(z^2-1)$, see (2.7). The cases $\pi<n^{-2-\delta}$ and $\pi>1-n^{-2-\delta}$ are easily disposed of; see step (v) of [14].

In view of (5.9),

$$\begin{aligned} &P\{\beta\leqq\hat\beta_{\text{PERC-T}}(\pi)\} \\ &\quad\leqq P\left[\beta\leqq\hat\beta+n^{-\frac12}\sigma_x^{-1}\hat\sigma\left\{z_\pi+\sum_{j=1}^{3} n^{-j/2}\hat s_j(z_\pi)+C_1 n^{-2}\right\}\right]+C_2 n^{-2} \\ &\quad= P\left\{n^{\frac12}(\hat\beta-\beta)\,\sigma_x/\hat\sigma+\sum_{j=1}^{3} n^{-j/2}\hat s_j(z_\pi)\geqq-(z_\pi+C_1 n^{-2})\right\}+C_2 n^{-2} \\ &\quad= \pi+\sum_{j=1}^{3} n^{-j/2}\, t_j(z_\pi)\,\phi(z_\pi)+O(n^{-2}), \end{aligned} \tag{5.10}$$

where t_j is a polynomial, even for odd j and odd for even j. Similarly, the last line of (5.10) is also a lower bound to the left-hand side of (5.10). Identification of the polynomials t_1 and t_2 is via the argument leading to (2.15), which shows that both are identically zero. Coverage probability of the interval $[\hat\beta_{\text{PERC-T}}(\pi),\hat\beta_{\text{PERC-T}}(1-\pi)]$ is now seen to equal $1-2\pi+2n^{-1}t_2(z_\pi)+O(n^{-2})=1-2\pi+O(n^{-2})$; compare (2.19). □

Acknowledgements. The referee, Associate Editor and Editor gave much helpful advice and constructive criticism, which considerably improved the presentation of this work.

References

1. Beran, R.: Bootstrap methods in statistics. Jahresber. Dtsch. Math.-Ver. **86**, 14–30 (1984)
2. Beran, R.: Prepivoting to reduce level error of confidence sets. Biometrika **74**, 457–468 (1987)
3. Beran, R., Millar, P.W.: Confidence sets for a multivariate distribution. Ann. Stat. **14**, 431–443 (1986)
4. Bhattacharya, R.N., Ghosh, J.K.: On the validity of the formal Edgeworth expansion. Ann. Stat. **6**, 435–451 (1978)
5. Bhattacharya, R.N., Ghosh, J.K.: On moment conditions for valid formal Edgeworth expansions. 1986 (preprint)
6. Bhattacharya, R.N., Ranga Rao, R.: Normal approximation and asymptotic expansions. New York: Wiley 1976
7. Bickel, P.J., Freedman, D.A.: On Edgeworth expansions for the bootstrap. Preprint (1980)
8. Bickel, P.J., Freedman, D.A.: Some asymptotic theory for the bootstrap. Ann. Stat. **9**, 1196–1217 (1981)
9. Bickel, P.J., Freedman, D.A.: Bootstrapping regression models with many parameters. In: Bickel, P.J., Doksum, K., Hodges, J.L. (eds.) A Festschrift for Erich Lehmann, (pp. 28–48) Belmont: Wadsworth 1983
10. Efron, B.: Nonparametric standard errors and confidence intervals (with discussion). Can. J. Stat. **9**, 139–172 (1981)
11. Efron, B.: Better bootstrap confidence intervals (with discussion). J. Am. Stat. Assoc. **82**, 171–200 (1987)
12. Freedman, D.A.: Bootstrapping regression models. Ann. Stat. **9**, 1218–1228 (1981)
13. Freedman, D.A., Peters, S.C.: Bootstrapping a regression equation: some empirical results. J. Am. Stat. Assoc. **79**, 97–106 (1984)
14. Hall, P.: On the bootstrap and confidence intervals. Ann. Stat. **14**, 1431–1452 (1986)
15. Hall, P.: Theoretical comparison of bootstrap confidence intervals. Ann. Stat., (1988) in press
16. Hinkley, D.: On jackknifing in unbalanced situations. Technometrics **19**, 285–292 (1977)
17. Hinkley, D., Wei, B-C.: Improvement of jackknife confidence limit methods. Biometrika **71**, 331–339 (1984)
18. Singh, K.: On the asymptotic accuracy of Efron's bootstrap. Ann. Stat. **9**, 1187–1195 (1981)
19. Withers, C.S.: Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric confidence intervals. Ann. Stat. **11**, 577–587 (1983)
20. Wu, C.F.: Jackknife, bootstrap, and other resampling methods in regression analysis (with discussion). Ann. Stat. **14**, 1261–1350 (1986)

Received November 4, 1986; in revised form June 28, 1988