METHODOLOGY AND THEORY FOR THE BOOTSTRAP
(Sixth set of two lectures)

Main topic of these lectures: Bootstrap methods in linear regression

Regression model

Assume we observe pairs $(x_1, Y_1), \ldots, (x_n, Y_n)$, generated by the model

$$Y_i = g(x_i) + \epsilon_i \,, \tag{1}$$

where g is a function that might be determined either parametrically or nonparametrically, and the errors ϵ_i have zero mean.

Regression model (cont.)

In the study of regression we take the explanatory variables x_i to be fixed, either because they are pre-determined (e.g. were regularly spaced) or are conditioned upon.

In this case, the only source of randomness in the model is the errors, ϵ_i, and so it is those that we resample, in form of residuals, when implementing the bootstrap. Our choice of lower-case notation for the explanatory variables reflects this view.

Correlation model

Alternatively, we might take the view that the explanatory variables are genuinely random and must be treated us such. For example, the data pairs (X_i, Y_i), for $1 \leq i \leq n$, might be drawn by sampling randomly from a bivariate distribution, and in our analysis of those data we might wish to preserve all the implications of this randomness, rather than condition some of it away by regarding the X_i's as fixed.

This approach to analysis might be termed the study of correlation, rather than regression. It would be addressed in bootstrap terms by resampling the pairs (X_i, Y_i), rather than resampling the residuals.

It is important to appreciate that these two different approaches to resampling — sampling the residuals or sampling the pairs (X_i, Y_i), respectively — are appropriate in different settings, for different models. They are not alternative ways of doing the same thing, and can lead to different conclusions.

Parametric regression

The good properties of percentile-t methods carry over to regression problems. However, in the setting of slope estimation those properties are significantly enhanced, and even the standard percentile method can perform unusually well.

For example, one-sided percentile-t confidence regions for slope have coverage error $O(n^{-3/2})$, not $O(n^{-1})$; and the error is only $O(n^{-2})$ in the case of two-sided intervals.

One-sided, standard percentile-method confidence intervals for slope, based on approximating the distribution of $\widehat{\theta} - \theta$ by the conditional distribution of $\widehat{\theta}^* - \widehat{\theta}$, have coverage error $O(n^{-1})$ rather than the usual $O(n^{-1/2})$.

General definition of slope

Although these exceptional coverage properties apply only to estimates of slope, not to estimates of intercept parameters or means, slope may be interpreted very generally.

For example, in the polynomial regression model

$$Y_i = c + x_i\, d_1 + \ldots + x_i^m\, d_m + \epsilon_i\,,$$

where we observe (x_i, Y_i) for $1 \le i \le n$, we regard each d_j as a slope parameter. A one-sided percentile-t interval for d_j has coverage error $O(n^{-3/2})$, although a one-sided percentile-t interval for c or for

$$E(Y \mid x = x_0) = c + x_0\, d_1 + \ldots + x_0^m\, d_m$$

has coverage error of size n^{-1}.

Why is slope favoured especially?

The reason for good performance in the case of slope parameters is the extra symmetry conferred by design points. Note that, in the polynomial regression case, we may write the model equivalently as

$$Y_i = c' + (x_i - \xi_1)\, d_1 + \ldots + (x_i^m - \xi_m)\, d_m + \epsilon_i\,,$$

where $\xi_j = n^{-1} \sum_i x_i^j$ and $c' = c + \xi_1\, d_1 + \ldots + \xi_m\, d_m$.

The extra symmetry arises from the fact that

$$\sum_{i=1}^{n} (x_i^j - \xi_j) = 0$$

for $1 \le j \le m$.

Correlation model

The results implying good coverage accuracy hold under the regression model, but not necessarily for the correlation model. For example, in the case of the linear correlation model, the symmetry discussed above will persist provided that

$$E\left\{\sum_{i=1}^{n} (X_i - \xi_1)\, \epsilon_i^k\right\} = 0 \qquad (1)$$

for sufficiently large k. Now, $\epsilon_i = Y_i - g(X_i)$, and as a result, (1) will generally not hold for $k \geq 1$. This means that, under the correlation model, the conventional properties of bootstrap confidence intervals hold; the special properties noted for regression, when estimating slope parameters, are not valid.

However, if the errors ϵ_i are independent of the explanatory variables X_i then (1) will hold for each k, and in such cases the enhanced features of the regression problem persist under the correlation model.

Simple linear regression

Consider the regression model,

$$Y_i = c + x_i\, d + \epsilon_i\,,$$

where the ϵ_i's are independent and identically distributed with zero mean and finite variance σ^2. Define $\sigma_x^2 = n^{-1} \sum_i (x_i - \bar{x})^2$,

$$\hat{d} = \sigma_x^{-2}\, \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})\, (\epsilon_i - \bar{\epsilon})$$

and $\hat{c} = \bar{Y} - \bar{x}\, \hat{d}$. Estimate $y_0 = E(Y \mid x = x_0) = c + x_0\, d$ by

$$\hat{y}_0 = \hat{c} + x_0\, \hat{d}\,,$$

and put

$$\begin{aligned} \sigma_y^2 &= 1 + \sigma_x^{-2}\, (x_0 - \bar{x})^2\,, \\ \hat{\epsilon}_i &= Y_i - \bar{Y} - (x_i - \bar{x})\, \hat{d} \end{aligned}$$

and $\hat{\sigma}^2 = n^{-1} \sum_i \hat{\epsilon}_i^2$, the latter estimating σ^2.

The asymptotic variances of $\hat{d}$ and $\hat{y}_0$ equal $\sigma^2/(n\, \sigma_x^2)$ and $\sigma^2\, \sigma_y^2/n$, respectively, and so $n^{1/2}\, (\hat{d} - d)\, \sigma_x/\hat{\sigma}$ and $n^{1/2}(\hat{y}_0 - y_0)/(\sigma\sigma_y)$ are asymptotically pivotal.

Bootstrapping the simple linear regression model

The residuals,

$$\hat{\epsilon}_i = Y_i - \bar{Y} - (x_i - \bar{x})\,\hat{d}\,,$$

are centred, in that $\sum_i \hat{\epsilon}_i = 0$. Therefore we may resample randomly, with replacement, from $\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n$, obtaining $\epsilon_1^*, \ldots, \epsilon_n^*$ say; and take as our bootstrap resample the pairs $(x_1, Y_1^*), \ldots, (x_n, Y_n^*)$, where

$$Y_i^* = \hat{c} + x_i\,\hat{d} + \epsilon_i^*\,.$$

Note particularly that, reflecting the fact that we condition upon the explanatory variables, those quantities are the same as in the original dataset.

In regression problems where the residuals are not centred, for example in nonparametric regression, we generally centre them, for example by subtracting their average value, before resampling.

Estimating quantiles of distribution of $\widehat{d}$

Let $\widehat{c}^*$, $\widehat{d}^*$ and $\widehat{\sigma}^*$ have the same formulae as $\widehat{c}$, $\widehat{d}$ and $\widehat{\sigma}$, respectively, except that we replace Y_i by Y_i^* throughout. The bootstrap versions of

$$\begin{aligned} S &= n^{1/2}\,(\widehat{d} - d)\,\sigma_x/\sigma\,, \\ T &= n^{1/2}\,(\widehat{d} - d)\,\sigma_x/\widehat{\sigma} \end{aligned}$$

are

$$\begin{aligned} S^* &= n^{1/2}\,(\widehat{d}^* - \widehat{d})\,\sigma_x/\widehat{\sigma}\,, \\ T^* &= n^{1/2}\,(\widehat{d}^* - \widehat{d})\,\sigma_x/\widehat{\sigma}^*\,, \end{aligned}$$

respectively.

We estimate the quantiles ξ_α and η_α of the distributions of S and T by $\widehat{\xi}_\alpha$ and $\widehat{\eta}_\alpha$, respectively, where

$$P(S^* \leq \widehat{\xi}_\alpha \,|\, \mathcal{X}) = \alpha\,, \quad P(T^* \leq \widehat{\eta}_\alpha \,|\, \mathcal{X}) = \alpha\,,$$

and $\mathcal{X} = \{(x_1, Y_1), \ldots, (x_n, Y_n)\}$ denotes the dataset.

Bootstrap confidence intervals for d

One-sided bootstrap confidence intervals for d, with nominal coverage α, are given by

$$
\begin{aligned}
\hat{I}_{11} &= \left(-\infty, \hat{d} - n^{-1/2}\,(\sigma/\sigma_x)\,\hat{\xi}_{1-\alpha}\right)\,, \\
\hat{I}_{12} &= \left(-\infty, \hat{d} - n^{-1/2}\,(\hat{\sigma}/\sigma_x)\,\hat{\xi}_{1-\alpha}\right)\,, \\
\hat{J}_{1} &= \left(-\infty, \hat{d} - n^{-1/2}\,(\hat{\sigma}/\sigma_x)\,\hat{\eta}_{1-\alpha}\right)\,.
\end{aligned}
$$

These are direct analogues of the intervals $\hat{I}_{11}$, $\hat{I}_{12}$ and $\hat{J}_1$ introduced earlier in non-regression problems. In particular, $\hat{I}_{12}$ and $\hat{J}_1$ are standard percentile-method and percentile-t bootstrap confidence regions.

Following the line of argument given earlier, we would expect $\hat{I}_{11}$ and $\hat{J}_1$ to have coverage error $O(n^{-1})$. In fact, they both have coverage error equal to $O(n^{-3/2})$. However, $\hat{I}_{11}$ is not of practical use, since it depends on the unknown σ, so we shall not treat it any further.

Likewise, we would expect $\hat{I}_{12}$ to have coverage error of size $n^{-1/2}$. However, we shall show that the error is actually of order n^{-1}.

Bootstrap confidence intervals for d (cont.)

Note that, although $\widehat{I}_{12}$ involves the variance estimator $\widehat{\sigma}$, it can be constructed numerically without resorting to computing $\widehat{\sigma}$.

Indeed, $\widehat{I}_{12}$ is identical to the interval

$$\widehat{I}_{12} = \left(-\infty, \widehat{d} - \widehat{w}_{1-\alpha}\right) ,$$

where $\widehat{w}_{1-\alpha}$ is the standard percentile-method estimator of $w_{1-\alpha}$, the latter defined by

$$P(\widehat{d} - d \le w_{1-\alpha}) = 1 - \alpha \,.$$

In particular, $\widehat{w}_{1-\alpha}$ is defined by

$$P(\widehat{d}^* - \widehat{d} \le \widehat{w}_{1-\alpha} \,|\, \mathcal{X}) = 1 - \alpha \,.$$

The interval $\widehat{J}_1$ is a standard percentile-t bootstrap confidence interval.

Polynomials in Edgeworth expansions

Edgeworth expansions for the non-Studentised and Studentised statistics, S and T respectively, are given by:

$$\begin{aligned} P(S \le u) &= \Phi(u) + n^{-1/2}\, P_1(u)\, \phi(u) \\ &\qquad + n^{-1}\, P_2(u)\, \phi(u) + \dots\,, \\ P(T \le u) &= \Phi(u) + n^{-1/2}\, Q_1(u)\, \phi(u) \\ &\qquad + n^{-1}\, Q_2(u)\, \phi(u) + \dots\,, \end{aligned}$$

where

$$P_1(u) = Q_1(u) = \tfrac{1}{6}\, \gamma\, \gamma_x\, (1 - u^2)\,,$$

$\gamma = E(\epsilon/\sigma)^3$, $\gamma_x = n^{-1} \sum_i \{(x_i - \bar{x})/\sigma_x\}^3$, and P_2 and Q_2 are odd, quintic polynomials, with

$$\begin{aligned} P_2(u) &= Q_2(u) + w\, \{2 + (3/24)\, (2 - \kappa) \\ &\qquad \times (w^2 - 3)\} \end{aligned}$$

and $\kappa = E(\epsilon/\sigma)^4 - 3$.

Note particularly that $P_1 = Q_1$.

Why does P_1 equal Q_1?

To understand why it is helpful to treat S as an approximation to T. Indeed, note that, by definition of $\hat{\sigma}^2$,

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^{n} \hat{\epsilon}_i^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} \left\{\epsilon_i - \bar{\epsilon} - (x_i - \bar{x})\,(\hat{d} - d)\right\}^2 \\ &= \sigma^2 + \frac{1}{n} \sum_{i=1}^{n} (\epsilon_i^2 - \sigma^2) + O_p(n^{-1})\,.\end{aligned}$$

Therefore, defining

$$\Delta = \tfrac{1}{2}\, n^{-1}\, \sigma^{-2} \sum_{i=1}^{n} (\epsilon_i^2 - \sigma^2)\,,$$

and recalling that

$$\begin{aligned}S &= n^{1/2}\,(\hat{d} - d)\,\sigma_x/\sigma\,, \\ T &= n^{1/2}\,(\hat{d} - d)\,\sigma_x/\hat{\sigma}\,,\end{aligned}$$

we deduce that

$$T = S\,(1 - \Delta) + O_p(n^{-1})\,.$$

Why does P_1 equal Q_1? (cont. 1)

Making use of the approximation

$$T = S\,(1 - \Delta) + O_p(n^{-1})\,, \qquad (1)$$

the symmetry property

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0\,, \qquad (2)$$

and the representation

$$S = n^{-1/2}\,\sigma_x^{-1}\,\sigma^{-1}\,\sum_{i=1}^{n} (x_i - \bar{x})\,\epsilon_i\,,$$

it is readily proved that

$$E\{S\,(1 - \Delta)\}^j = E(S^j) + O(n^{-1}) \qquad (3)$$

for $j = 1, 2, 3$.

Exercise: Derive (3) for $j = 1, 2, 3$.

Why does P_1 equal Q_1? (cont. 2)

Therefore, the first three cumulants of S and $S\,(1-\Delta)$ agree up to and including terms of order $n^{-1/2}$.

It follows that Edgeworth expansions of the distributions of S and $S\,(1-\Delta)$ differ only in terms of order n^{-1}. In view of (1), the same is true of the distributions of S and T:

$$P(S \le u) = P(T \le w) + O(n^{-1})\,.$$

Therefore, the $n^{-1/2}$ terms in the expansions must be identical; that is, $P_1 = Q_1$.

The chief ingredient in this argument is the symmetry property (2). It, and its analogues, guarantee that, in the problem of slope estimation for general regression problems, $P_1 = Q_1$.

Consequences of the property $P_1 = Q_1$

The identity $P_1 = Q_1$ implies that, to first order (i.e. up to and including terms of order $n^{-1/2}$), estimating the distribution of S is the same as estimating the distribution of T.

As we saw earlier in non-regression problems, the percentile method estimates the distribution of S, whereas the percentile-t method estimates the distribution of T. The fact that, in the setting of estimating slope in regression, $P_1 = Q_1$, means that these two techniques give the same results up to and including terms of order $n^{-1/2}$. They differ only in terms of order n^{-1}, and terms of higher order.

Therefore, since one-sided confidence intervals based on the percentile-t method have coverage error equal to $O(n^{-1})$, the same must be true for confidence intervals based on the percentile method.

Properties of percentile-t confidence regions

The coverage error of a one-sided percentile-t confidence interval for d is of order $n^{-3/2}$, rather than the usual n^{-1}. That is, with $\hat{J}_1$ defined by

$$\hat{J}_1 = \left(-\infty, \hat{d} - n^{-1/2}\,(\hat{\sigma}/\sigma_x)\,\hat{\eta}_{1-\alpha}\right)\,,$$

it can be shown that

$$P(d \in \hat{J}_1) = \alpha + O(n^{-3/2})\,. \tag{4}$$

Since terms of odd order in $n^{-3/2}$ cancel from formulae for coverage error of two-sided confidence intervals, then the two-sided, percentile-t bootstrap confidence interval for d has coverage error of order n^{-2}, rather than the usual n^{-1}.

Derivation of (4)

A proof of (4) can be given as follows. Recall that

$$P(T \le u) = \Phi(u) + n^{-1/2} Q_1(u)\,\phi(u) + n^{-1} Q_2(u)\,\phi(u) + \dots\,,$$

where

$$Q_1(u) = \tfrac{1}{6}\,\gamma\,\gamma_x\,(1 - u^2)\,,$$

$\gamma = E(\epsilon/\sigma)^3$ and $\gamma_x = n^{-1} \sum_i \{(x_i - \bar{x})/\sigma_x\}^3$.

Derivation of (4) (cont. 1)

The bootstrap version of the Taylor expansion is

$$P(T^* \le u \,|\, \mathcal{X}) \;=\; \Phi(u) + n^{-1/2}\,\widehat{Q}_1(u)\,\phi(u) + n^{-1}\,\widehat{Q}_2(u)\,\phi(u) + \ldots\,,$$

where

$$\widehat{Q}_1(u) = \tfrac{1}{6}\,\widehat{\gamma}\,\gamma_x\,(1-u^2)$$

and $\widehat{\gamma} = E(\widehat{\epsilon}/\widehat{\sigma})^3$. Now, the solutions η_α and $\widehat{\eta}_\alpha$, of the respective equations

$$P(T \le \eta_\alpha) = \alpha\,, \quad P(T^* \le \widehat{\eta}_\alpha \,|\, \mathcal{X}) = \alpha\,,$$

admit the Cornish-Fisher expansions

$$\eta_\alpha \;=\; z_\alpha + n^{-1/2}\,Q_1^{\mathrm{cf}}(z_\alpha) + n^{-1}\,Q_2^{\mathrm{cf}}(z_\alpha) + \ldots\,,$$
$$\widehat{\eta}_\alpha \;=\; z_\alpha + n^{-1/2}\,\widehat{Q}_1^{\mathrm{cf}}(z_\alpha) + n^{-1}\,\widehat{Q}_2^{\mathrm{cf}}(z_\alpha) + \ldots\,.$$

Derivation of (4) (cont. 2)

Cornish-Fisher expansions:

$$\eta_\alpha = z_\alpha + n^{-1/2}\, Q_1^{\mathsf{cf}}(z_\alpha) + n^{-1}\, Q_2^{\mathsf{cf}}(z_\alpha) + \ldots\,,$$
$$\widehat{\eta}_\alpha = z_\alpha + n^{-1/2}\, \widehat{Q}_1^{\mathsf{cf}}(z_\alpha) + n^{-1}\, \widehat{Q}_2^{\mathsf{cf}}(z_\alpha) + \ldots\,.$$

On subtracting these expansions we deduce that

$$\begin{aligned}\widehat{\eta}_\alpha - \eta_\alpha &= n^{-1/2}\, \{\widehat{Q}_1^{\mathsf{cf}}(z_\alpha) - Q_1^{\mathsf{cf}}(z_\alpha)\} \\ &\qquad + n^{-1}\, \{\widehat{Q}_2^{\mathsf{cf}}(z_\alpha) - Q_2^{\mathsf{cf}}(z_\alpha)\} + \ldots \\ &= n^{-1/2}\, \{Q_1(z_\alpha) - \widehat{Q}_1(z_\alpha)\} \\ &\qquad + O_p(n^{-3/2})\,,\end{aligned}$$

where we have used the fact that $Q_1^{\mathsf{cf}} = -Q_1$, $\widehat{Q}_1^{\mathsf{cf}} = -\widehat{Q}_1$ and $\widehat{Q}_2^{\mathsf{cf}} = Q_2^{\mathsf{cf}} + O_p(n^{-1/2})$.

Derivation of (4) (cont. 3)

From previous pages:

$$\widehat{Q}_1(u) - Q_1(u) = \tfrac{1}{6}\,(\widehat{\gamma} - \gamma)\,\gamma_x\,(1 - u^2)\,, \qquad (5)$$

$$\widehat{\eta}_\alpha - \eta_\alpha = n^{-1/2}\,\{Q_1(z_\alpha) - \widehat{Q}_1(z_\alpha)\} + O_p(n^{-3/2})\,. \qquad (6)$$

It may be proved by Taylor expansion that

$$\begin{aligned} \widehat{\gamma} &= \frac{n^{-1}\sum_i \{\epsilon_i - \bar{\epsilon} - (x_i - \bar{x})\,(\widehat{d} - d)\}^3}{[n^{-1}\sum_i \{\epsilon_i - \bar{\epsilon} - (x_i - \bar{x})\,(\widehat{d} - d)\}^2]^{3/2}} \\ &= \gamma + n^{-1/2}\,U + O_p(n^{-1})\,, \qquad (7) \end{aligned}$$

where

$$U = n^{-1/2} \sum_{i=1}^{n} \Big\{(\delta_i^3 - \gamma) - \tfrac{3}{2}\,\gamma\,(\delta_i^2 - 1) - 3\,\delta_i\Big\}$$

and $\delta_i = \epsilon_i/\sigma$. Combining results (5)–(7) we deduce that

$$\widehat{\eta}_\alpha - \eta_\alpha = -n^{-1}\,\tfrac{1}{6}\,U\,\gamma_x\,(1 - z_\alpha^2) + O_p(n^{-3/2})\,.$$

Derivation of (4) (cont. 4)

From the previous page:

$$\widehat{\eta}_{1-\alpha} - \eta_{1-\alpha} = -n^{-1}\, c\, U + O_p(n^{-3/2})$$

where $c = \frac{1}{6}\,\gamma_x\,(1 - z^2_{1-\alpha})$. Therefore,

$$\begin{aligned} P(d \in \widehat{J}_1) &= P\left\{d < \widehat{d} - n^{-1/2}\,(\widehat{\sigma}/\sigma_x)\,\widehat{\eta}_{1-\alpha}\right\} \\ &= P(T > \widehat{\eta}_{1-\alpha}) \\ &= P\left\{T + n^{-1}\,c\,U > \eta_{1-\alpha} \right. \\ &\qquad\qquad \left. + O_p(n^{-3/2})\right\}, \\ &= P\left(T + n^{-1}\,c\,U > \eta_{1-\alpha}\right) \\ &\qquad\qquad + O(n^{-3/2}), \end{aligned}$$

assuming we can treat the "$O_p(n^{-3/2})$" inside the probability as though it were deterministic, and take it outside.

Derivation of (4) (cont. 5)

From previous page:

$$P(d \in \hat{J}_1) = P(T + n^{-1}\,cU > \eta_{1-\alpha}) + O(n^{-3/2}). \tag{8}$$

It can be proved that, for any choice of the constant c, the first four moments (and hence also the first four cumulants) of $T + n^{-1}\,cU$ are identical to those of T, up to and including terms of order $n^{-3/2}$. Hence, recalling the way in which moments influence Edgeworth expansions,

$$\begin{aligned} P(T + n^{-1}\,cU > \eta_{1-\alpha}) & \\ &= P(T > \eta_{1-\alpha}) + O(n^{-3/2}) \\ &= \alpha + O(n^{-3/2})\,. \end{aligned}$$

Therefore, by (8),

$$P(d \in \hat{J}_1) = \alpha + O(n^{-3/2})\,,$$

as had to be proved.

The other percentile-method interval

Recall that the percentile-method interval $\hat{I}_{12}$ is based on bootstrapping $\hat{d} - d$; that is, it is based on approximating the distribution of this quantity by the conditional distribution of $\hat{d}^* - \hat{d}$.

The "other" percentile method is based on using the conditional distribution of $\hat{d}^*$ to approximate the distribution of $\hat{d}$. It leads to the interval

$$\widehat{K}_1 = \Big(-\infty, \hat{d} + n^{-1/2}\,(\hat{\sigma}/\sigma_x)\,\hat{\eta}_\alpha\Big) = (-\infty, \hat{\zeta}_\alpha)\,,$$

where $\hat{\zeta}_\alpha$ is an approximation to ζ_α, these two quantities being defined by

$$P(\hat{d}^* \le \hat{\zeta}_\alpha \,|\, \mathcal{X}) = \alpha\,, \quad P(\hat{d} \le \zeta_\alpha) = \alpha\,.$$

However, $\widehat{K}_1$ has coverage error of size $n^{-1/2}$, not n^{-1}. In particularly, $\widehat{K}_1$ does not enjoy the accuracy of the percentile-method interval $\hat{I}_{12}$. This is a consequence of it addressing the wrong tail of the distribution of $\hat{d}$.

Properties of confidence intervals for the conditional mean, $y_0 = E(Y \,|\, x = x_0)$, and the intercept, c

Recall that $y_0 = c + x_0\, d$, which in turn equals c when $x_0 = 0$. Therefore we can treat confidence intervals for c as a special case of those for y_0.

Recalling that $\hat{y}_0 = \hat{c} + x_0\, \hat{d}$, redefine

$$\begin{aligned} S &= n^{1/2}\,(\hat{y}_0 - y_0)/(\sigma\,\sigma_y)\,, \\ T &= n^{1/2}\,(\hat{y}_0 - y_0)/(\hat{\sigma}\,\sigma_y)\,; \end{aligned}$$

and taking $\hat{y}_0^* = \hat{c}^* + x_0\, \hat{d}^*$, redefine

$$\begin{aligned} S^* &= n^{1/2}\,(\hat{y}_0^* - \hat{y}_0)/(\hat{\sigma}\,\sigma_y)\,, \\ T^* &= n^{1/2}\,(\hat{y}_0^* - \hat{y}_0)/(\hat{\sigma}^*\,\sigma_y)\,. \end{aligned}$$

Properties of confidence intervals for y_0 and c (cont.)

Percentile-method and percentile-t confidence intervals for y_0 and c are given respectively by

$$\begin{aligned}\hat{I}_{12} &= \left(-\infty, \hat{y}_0 - n^{-1/2}\,(\hat{\sigma}/\sigma_y)\,\hat{\xi}_{1-\alpha}\right), \\ \hat{J}_1 &= \left(-\infty, \hat{y}_0 - n^{-1/2}\,(\hat{\sigma}/\sigma_y)\,\hat{\eta}_{1-\alpha}\right),\end{aligned}$$

where $\sigma_y^2 = 1 + \sigma_x^{-2}\,(x_0 - \bar{x})^2$ and we define $\hat{\xi}_{1-\alpha}$ and $\hat{\eta}_{1-\eta}$ by

$$P(S^* \le \hat{\xi}_\alpha \,|\, \mathcal{X}) = \alpha\,, \quad P(T^* \le \hat{\eta}_\alpha \,|\, \mathcal{X}) = \alpha\,,$$

for the new versions of S^* and T^*.

The intervals $\hat{I}_{12}$ and $\hat{J}_1$ have coverage errors $O(n^{-1/2})$ and $O(n^{-1})$, respectively. These results, unlike those for slope, are conventional.