The most common applications of this one-sample bootstrap are:

1) Estimation of $\sqrt{\text{Var}_F(T_n)}$
   \[ \frac{1}{B-1} \sum (T_{n_i}^* - \overline{T}_n^*)^2 \]
   via $\frac{1}{B} \sum T_{n_i}^*$
   this gets me a "standard error" for $T_n$

   \[ T_n - (\overline{T}_n^* - \eta(\hat{F})) \]

   Notice that in the case that $\hat{F}$ is the empirical dsn and $T_n = \eta(\hat{F})$
   This suggests a biased corrected version of $T_n$ that is
   \[ 2T_n - \overline{T}_n^* \]

   "Examples" a) $T_n = \text{sample median}$
   $\theta = \eta(F) = \text{E}_F Y = \text{population mean}$

2) Bias of $T_n$ as an estimator of some quantity $\theta = \eta(F)$

   \[ \text{Bias}_F(T_n) = E_F T_n - \eta(F) \]

   and a standard methodology is to estimate $F$ (with $\hat{F}$) and get $B$ bootstrap versions of $T_n$ and use
   \[ \overline{T}_n^* - \eta(\hat{F}) \]

   as an estimated bias — and a possible "bias-corrected" version of $T_n$ is then
   an estimated bias of $T_n$ for $\theta$ is
   \[ \overline{T}_n^* - \eta(\hat{F}) \]
   mean of $\hat{F}$
   average of bootstrap sample
   (nonparametric)
   version is the sample mean

   and a bias-corrected version of $T_n$
   is thus
   \[ T_n - (\overline{T}_n^* - \eta(\hat{F})) \]

   b) $T_n = \text{sample median}$
   $\theta = \eta(F) = F^{-1}(0.5) = \text{population median}$
   Here $T_n = \eta(\text{empirical dsn}) = \eta(F)$
3) Confidence limits for some \( \theta \)

Again suppose \( \theta = \eta(F) \) is some characteristic of \( F \) and

\[ T_n = \eta(\text{empirical cdf/histogram}) \]

Based on \( B \) bootstrapped values

\[ T_{n1}^*, T_{n2}^*, \ldots, T_{nB}^* \]

I make ordered values

\[ T_{n(1)}^* \leq T_{n(2)}^* \leq \ldots \leq T_{n(B)}^* \]

and find lower and upper \( \frac{\alpha}{2} \) pts of the empirical cdf of \( T_n^* \)’s

as an approximate \( (1-\alpha) \) - level CI for \( \theta = \eta(F) \)

why this should work is not so clear... in fact it's not elementary and in fact this need not work...

(partial 11.22-11.24 Kochers notes/ Ch 13 of Efron and Tibshirani) - the standard "argument" as to why this might work is:

Suppose there is a function \( m(\cdot) \) s.t.

\[ \phi = m(\theta) = m(\eta(F)) \]

est'd bias of \( T_n \) for \( \theta \) is then

\[ \overline{T_n^*} - \eta(F) \]

average bootstrap median

this is \( T_n \) in nonparametric context where \( F = \text{empirical cdf} \)

So a (nonparametric) bias-corrected version of \( T_n \) is

\[ 2T_n - \overline{T_n^*} \] average bootstrap median

For

\[ k_L = \left\lfloor \frac{\alpha}{2} (B+1) \right\rfloor \]

= largest integer \( \leq \frac{\alpha}{2} (B+1) \)

and \( k_U = (B+1) - k_L \)

these \( \frac{\alpha}{2} \) pts are roughly

\( T_{n(k_L)}^* \) and \( T_{n(k_U)}^* \)

and people use the interval

\[ [T_{n(k_L)}^*, T_{n(k_U)}^*] \]

For \( \theta = \eta(F) \)
and 
\[ \hat{\phi} = m(T_n) = m(\text{empirical distribution of } Y_1, \ldots, Y_n) \]
so that for \( n \) large 
\[ \hat{\phi} \sim N(\phi, \sigma^2) \]
then a CI for \( \phi \) is 
\[ (\hat{\phi} - z \sigma, \hat{\phi} + z \sigma) \]
and a CI for \( \Theta \) is 
\[ (m^{-1}(\hat{\phi} - z \sigma), m^{-1}(\hat{\phi} + z \sigma)) \]

"Bias-corrected Bootstrap Percentile Confidence Intervals" (BCa intervals)

Get 
\[ T_n^*(1), \ldots, T_n^*(B) \]
Then use 
\[ [T_n^*(k_1(B+1)), T_n^*(k_2(B+1))] \]
(round \( k_1(B+1) \) down and \( k_2(B+1) \) up if necessary) where

\[ \hat{\xi}_0 = \Phi^{-1}(\text{fraction of } T_n^* \text{ that are smaller than } T_n) \]

and the argument is that 
\[ [T_n^*(k_1), T_n^*(k_2)] \]
approximates this CI for \( \Theta \) (without having to know or use \( m(\cdot) \))
This is called the "empirical percentile" bootstrap interval for \( \Theta \), there are variations on this/ improvements that are more complicated but seem to work better in terms of holding their nominal coverage probability 

\[ k_1 = \Phi \left( \frac{\hat{\xi}_0 + \hat{\xi}_0 - Z_{\text{upper}}}{1 - \Phi(\hat{\xi}_0 - Z_{\text{upper}})} \right) \]

\[ k_2 = \Phi \left( \frac{\hat{\xi}_0 + Z_{\text{upper}}}{1 - \Phi(\hat{\xi}_0 + Z_{\text{upper}})} \right) \]

for \( \Phi \) the standard normal cdf
\[ Z_{\text{upper}} = \text{upper } \frac{\alpha}{2} \text{ pt of std normal} \]
(for confidence level \((1 - \alpha)\))

\[ \hat{\xi}_0 = \Phi^{-1}(\text{fraction of } T_n^* \text{ that are smaller than } T_n) \]

a measure of median bias of \( T_n \) in normal units
\[ T_{n,j} = T_n \text{ computed dropping the jth observation from data set} \]
\[ = \tilde{\eta} (\text{all but the jth } \tilde{Y}_i) \]
\[ \overline{T_n} = \frac{1}{n} \sum_{j=1}^{n} T_{n,j} \]

Statistical folklore says that the BCa and ABC methods do a better job of holding their nominal coverage probability than the uncorrected version.

So what about bootstrapping in more complicated contexts? For example, what about the bootstrap in non-linear regression? There are myriad possibilities ... e.g. one could think about resampling data vectors \((\tilde{X}_1, Y_1), (\tilde{X}_2, Y_2), \ldots, (\tilde{X}_n, Y_n)\)

\[ \hat{a} = \frac{\sum_{i=1}^{n} (\overline{T_{n,i} - T_{n,j}})^3}{6 \left( \sum_{i=1}^{n} (\overline{T_{n,i} - T_{n,j}})^2 \right)^{3/2}} \]

= "an estimated acceleration factor"

Another method called the ABC (approximate bias-corrected percentile bootstrap method) that approximates BCa method analytically - a contributed R package will compute both BCa and ABC intervals -

Another approach is to bootstrap residuals.

See Kaehler's notes or Efron and Tibshirani -

"Generalized" Linear Models

(Another kind of generalization of the LM) responses are assumed to come from (not necessarily the normal dsn but) some exponential family of dsn's (e.g. normal, binomial, Poisson, Gamma, inverse Gamma)
y a response

( \rightleftharpoons an explanatory vector ) ( later )

assume that y has a pdf or pmf of the form

\[ f(y | \theta, \phi) = \exp \left[ \frac{y \theta - b(\theta)}{a(\phi)} - c(y, \phi) \right] \]

(later we'll let \( \theta \) depend on \( X \))

for \( a(\cdot) \), \( b(\cdot) \), \( c(\cdot) \) some functions, \( \theta \)

a "canonical parameter" and \( \phi \) some

\[ \text{Binomial case } \quad y \sim \text{Bi}(n, p) \]

\[ f(y|p) = \binom{n}{y} p^y (1-p)^{n-y} \]

\[ = \binom{n}{y} \left( \frac{p}{1-p} \right)^y (1-p)^n \]

\[ f(y|p) = \exp \log f(y|p) \]

\[ = \exp \left[ y \log \frac{p}{1-p} + n \log (1-p) + \log \binom{n}{y} \right] \]

Take \( \theta = \log \frac{p}{1-p} \quad ( p = \frac{e^\theta}{1 + e^\theta} ) \)

\[ \text{Normal case } \quad y \sim N(\mu, \sigma^2) \]

\[ \theta = \mu = \mathbb{E}y \quad \phi = \sigma^2 = \text{Var} y \]

\[ b(\theta) = \frac{\mu^2}{2} \]

\[ a(\phi) = \phi \]

\[ c(y, \phi) = \frac{1}{2} \left( \frac{y^2}{\phi} + \log 2\pi \phi \right) \]

\[ \text{Binomial case } \quad y \sim \text{Bi}(n, p) \]

\[ b(\theta) = -n \log (1-p) = n \log (1+e^\theta) \]

\[ \phi = 1 \]

\[ a(\phi) = 1 \]

\[ c(y, \phi) = \log \binom{n}{y} \]

\[ \text{Poisson case } \quad y \sim \text{Poisson} \lambda \]

\[ f(y|\lambda) = \frac{e^{-\lambda} \lambda^y}{y!} \]

\[ = \exp \left[ y \log \lambda - \lambda - \log y! \right] \]

\[ \theta = \log \lambda \quad (\lambda = e^\theta) \]
\[ b(\theta) = \lambda = e^{\theta} \]
\[ \phi = 1 \]
\[ a(\phi) = 1 \]
\[ c(y, \phi) = \log y! \]

Koehler's notes also show \( n \) cases and inverse \( n \) cases.

We'll proceed by finding general theory (large sample theory) for all models of the basic exponential family form.

Applying these to our basic exponential family form we get

\[ m = E(y) = b'(\theta) \quad \theta = b^{-1}(m) \]

and after a bit of work

\[ \text{Var}(y) = a(\phi) b''(\theta) \]

write

\[ \text{Var}(y) = b''(b^{-1}(m)) \]

and note that

\[ \text{Var}(y) = a(\phi) \text{Var}(m) \]

(i.e., the variance of \( y \) is a function of \( m \)

unless \( v \) is constant in \( m \))

General theory says that under regularity conditions

\[ E_{\theta_0, \phi_0} \left( \frac{\partial}{\partial \theta} \ln f(y | \theta, \phi) \right) \bigg|_{\theta_0, \phi_0} = 0 \]

\[ \text{Var}_{\theta_0, \phi_0} \left( \frac{\partial}{\partial \theta} \ln f(y | \theta, \phi) \right) \bigg|_{\theta_0, \phi_0} = -E_{\theta_0, \phi_0} \left( \frac{\partial^2}{\partial \theta \partial \theta} \ln f(y | \theta, \phi) \right) \bigg|_{\theta_0, \phi_0} \]

The heart of a generalized linear model is then the decision to model some function of \( m = Ey \) as linear in predictors \( \xi \), i.e., for some (link) function \( h(\cdot) \) we suppose

\[ h(m) = \xi' \beta \quad m = h^{-1}(\xi' \beta) \]

Then if \( y_1, y_2, \ldots, y_n \) are independent with \( E(y_i) = m_i \) and \( h(m_i) = \xi_i' \beta \)

and further suppose

\[ X = \left( \begin{array}{c} \xi_1' \\ \vdots \\ \xi_n' \end{array} \right) \]

is full rank.
the likelihood function is

\[ f(Y \mid \beta, \phi) = \prod_{i=1}^{n} \exp \left[ \frac{y_i \theta_i - b(\theta_i) - c(y_i, \phi)}{a(\phi)} \right] \]

where \( \theta_i = B^{-1}(u_i) = B^{-1}(h^{-1}(x'_i \beta)) \)

and standard inference in this model is based on the "usual" large sample theory (i.e. MVN limit for MLE's and \( \chi^2 \) limits for LRT's) —

numerically — unless your specialty is statistical computing, that's a (nasty) detail —

For a solution to the set of likelihood equations for large \( n \), some general theory suggests

\[ \hat{\beta} \sim \text{MVN}_k \left( \beta, a(\phi) \left( X' W(\beta) X \right)^{-1} \right) \]

for

\[ W(\beta) = \text{diag} \left( \frac{1}{h''(x'_i \beta)} \right)_{i=1}^{n} \]

For what it's worth an MLE of \( \beta \) must satisfy

\[ \frac{\partial}{\partial \beta_j} \log f(Y \mid \beta, \phi) = 0 \quad \forall j \]

and as it turns out this boils down to the set of equations

\[ 0 = \sum_{i=1}^{n} \left[ y_i - h'(x'_i \beta) \right] \frac{1}{\sqrt{v(h'(x'_i \beta)) h''(h'(x'_i \beta))}} \quad \forall j \]

these are the generalized equations of the normal equations and typically must be solved

so one form of inference for elements of \( \beta \) is using

\[ Q = \left( a(\phi) \left( X' W(\hat{\beta}) X \right)^{-1} \right) \]

in the event \( \lambda \neq 1 \)

and with \( q_j \) the \( j \)th diagonal entry of \( Q \) we make an approximate C.I.

for \( \beta_j \) as

\[ \hat{\beta}_j \pm z \sqrt{q_j} \]