Nonparametric Bootstrap

1 Sampling Distributions of Some Classes of Random Variables

In our course we discuss statistical inference statistical models for iid samples. We also some extensions to dependent r.v.s and their statitsical models, but mainly focus on iid settings.

On common type of r.v. that comes up in hypothesis testing and confidence intervals is

$$W = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sqrt{S_n^2}} \; .$$

This involves r.v.s $X_i, i = 1, \ldots, n$ which are iid from a distribution F, and

$$\mu = \mathrm{E}_F(X)$$

where X is a *generic* r.v. from distribution F. Sometimes it is helpful for us to think about this in a bit more explicit form

$$W = \frac{\sqrt{n}\left(\bar{X}_n - \mu(F)\right)}{\sqrt{S_n^2}}$$

where we need to recognize how μ is related to the population distribution F, that is μ is a function of F which is denoted as $\mu(F)$.

If the statistical model has parameter space Θ then we may estimate θ by $\hat{\theta}_n$, typically a method of moments estimator or MLE. In that case we might be interested in a r.v.

$$W = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sqrt{V(\hat{\theta}_n)}} \ .$$

In this case we have $\mu = \mathrm{E}_F(X)$ which we can think of more explicitly as

$$\mu(\theta) = \int_{-\infty}^{\infty} x f(x,\theta) dx$$

in the continuous case and as

$$\mu(\theta) = \sum_x x f(x,\theta)$$

in the discrete case, interpreting f as the pdf of pmf in the two formulae.

These are of the following form.

$$W = \frac{\sqrt{n}\left(\hat{\theta}_n - \theta\right)}{\sqrt{\hat{\tau}^2}}$$

In the case that θ is a vector, that is Θ has dimension 2 or larger, we might then pick on component, and correspondingly interpret the above. The quantity in the denominator is the expression that typically comes in from the normal approximation for the r.v. W. In the regular statistical model with an MLE, it is obtained from either Fisher's information, $I(\theta)$, or from observed Fisher's information, $I(\hat{\theta}_n)$.

We can think of this as

$$W = \frac{\sqrt{n}\left(\hat{\psi}_n - \psi\right)}{\sqrt{v_n^2}} \ .$$

ψ is the appropriate *parameter* or function of the parameter as needed in our particular application or case. Again if need to keep track of how ψ depends on the population distribution we can think of it as $\psi(F)$ or as $\psi(\theta)$ when there is a parameter.

These can be used to construct a confidence interval for θ, or a component of θ, or of μ which is a function of θ.

If we have the sampling distribution of W the we have the corresponding quantiles for this distribution. Specifically if we have the 0.025 and 0.975 quantiles, say $c_L = q_{.025}$ and $c_U = q_{.975}$ then we obtain the 95% confidence interval for ψ by solving for ψ from

$$q_{.025} \leq \frac{\sqrt{n}\left(\hat{\psi}_n - \psi\right)}{\sqrt{\hat{\tau}^2}} \leq q_{.975} \ .$$

This gives the 95% confidence for ψ as

$$\hat{\psi}_n - q_{.975}\frac{\sqrt{\hat{\tau}^2}}{\sqrt{n}} \leq \psi \leq \hat{\psi}_n - q_{.025}\frac{\sqrt{\hat{\tau}^2}}{\sqrt{n}} \ .$$

In order to make this useful in practice we need a statistic that is used to estimate v_n^2 and also to a method to obtain these quantiles. That is why we often use a sample variance when $\psi = \mu = \mathrm{E}_F(X)$ or observed Fisher's information, or another appropriate expression as needed for a method of moments estimator.

If the distribution of W is not explicitly know, we can approximate the distribution of W. This is usually done by a normal approximation. We have seen how these can be obtained Theorems in Chapter 9.5 for the regular statistical models and MLE, and by the use of the delta method in the method of moments estimation method.

Additionally we also have seen that if θ is known (generally not a realistic setting) we can also use a simulation method to approximate the distribution of W and hence obtain the needed quantiles of W.

2 Parametric Bootstrap

Earlier in the course we used the parametric bootstrap. A brief overview of this method is given below.

1. X_i, $1 = 1, \ldots, n$ are iid from $f(\cdot;\theta), \theta \in \Theta$. θ_0 is used to denote the true value of the parameter.

2. $\hat{\theta}_n = h(X_1, \ldots, X_n)$ is an estimator of θ and T_n is a centred or standardized r.v. constructed from this; for example

$$T_n = \frac{\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)}{v_n}$$

   where v_n is a known standard deviation or estimated standard deviation, say $v_n^2 = h_1(X_1, \ldots, X_n)$.

3. Consider X_i^* iid from $f(\cdot;\hat{\theta}_{\text{obs}})$. Based on these iid rv's construct (that is using the same formula as for $\hat{\theta}_n$ and v_n, that is the same h, h_1)

   - $$\hat{\theta}_n^* = h(X_1^*, \ldots, X_n^*) \ , \ (v_n^*)^2 = h_1(X_1^*, \ldots, X_n^*)$$

   - $$T_n^* = \frac{\sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_{\text{obs}}\right)}{v_n^*}$$

4. Use Monte Carlo simulation with M simulation steps of size n each to approximate the sampling distribution of T_n^* for this given $\hat{\theta}_{\text{obs}}$

We will review this further in the lecture.

3 Nonparametic Bootstrap

The nonparametric bootstrap is very similar except for the third step. This is now replaced by rv's X_i^* iid from the empirical distribution of the observed $X_1, X_2, \ldots, X_n$. Other parts of the algorithm are the same.

1. X_i, $1 = 1, \ldots, n$ are iid from $f(\cdot)$
2. $\hat{\theta}_n = h(X_1, \ldots, X_n)$ is an estimator of θ and T_n is a centred or standardized r.v. constructed from this; for example

$$T_n = \frac{\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)}{v_n}$$

   where v_n is a known standard deviation or estimated standard deviation, say $v_n^2 = h_1(X_1, \ldots, X_n)$.
3. Consider X_i^* iid from F_n, the empirical distribution function of $X_1, X_2, \ldots, X_n$. Based on these iid rv's construct (that is using the same formula as for $\hat{\theta}_n$ and v_n, that is the same h, h_1)
   - $$\hat{\theta}_n^* = h(X_1^*, \ldots, X_n^*) \ , \ (v_n^*)^2 = h_1(X_1^*, \ldots, X_n^*)$$
   - $$T_n^* = \frac{\sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_{\text{obs}}\right)}{v_n^*}$$
4. Use Monte Carlo simulation with M simulation steps of size n each to approximate the sampling distribution of T_n^* for this given $\hat{\theta}_{\text{obs}}$

Based on the bootstrap simulation we can thus obtain the *bootstrap* approximation to quantiles for the sampling distribution of

$$T_n^* = \frac{\sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_{\text{obs}}\right)}{v_n^*}$$

This allows us for example to obtain quantiles, say c_L^* and c_U^* for the central 0.95 region. The bootstrap theory then states that these quantiles are approximately the quantiles of

$$T_n = \frac{\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)}{v_n} \ .$$

Thus for example we obtain an approximate 95% confidence interval for θ_0 by solving for θ_0 from

$$c_L^* \leq \frac{\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)}{v_n} \leq c_U^* \ .$$

This idea is explained in more detail at the end of this section.

The bootstrap algorithm is easy to implement in a simulation code, for example in R. The numerical values of the appropriate quantiles c_L^* and c_U^* are then easily obtained. R also contains several packages that implement the nonparametric bootstrap.

The theory of why the nonparametric bootstrap gives good approximations to the sampling distribution of T_n and its quantiles is not discussed in this course. The method applies generally when the *true* sampling distribution of T_n converges in distribution to a normal distribution, and as a bonus usually gives a better approximation to the *true* quantiles than does many other methods of approximation.

Now we need to ask how do we implement step 3 in the nonparametric bootstrap method. For the parametric bootstrap we use a method of transformation based on methods in Chapter 2 in Stat 3657. To make this even easier R often has programs for generating random variables from some specified distributions, once we specify the parameter values.

We have seen in earlier course that there may be different functions of r.v.s that give the same resulting distribution. For example consider the Cauchy distribution, with pdf

$$f(x) = \frac{1}{\pi(1+x^2)} \ .$$

It has a nice cdf

$$F(x) = \frac{1}{2} + \frac{1}{\pi} tan^{-1}(x)$$

which is given terms of the arc tan function $\tan^{-1}$. Thus F^{-1} is very easy to obtain (the student should do this). Since F is continuous then if $U \sim \mathrm{Unif}(0,1)$ then

$$X = F^{-1}(U)$$

is a r.v. with cdf F, and thus X has a Cauchy distribution. A second method is given by Rice Example 3.6.1 Example B. Suppose that Y_1, Y_2 are iid $N(0,1)$ r.v.s. Let

$$X = \frac{Y_1}{Y_2} \ .$$

Then X has the Cauchy pdf. The student should review that example or work through the method of finding the pdf of X directly, using a quotients method or the completion of the transformation.

Something similar can be done for simulation of r.v.s which have as its cdf the empirical distribution.

Method 1

For the non parametric bootstrap we need to revisit the empirical distribution (EDF) The observed data is $x_1, x_2, \ldots, x_n$ and the EDF is

$$\hat{F}_n(y) = F_n(y) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(x_i \le y) \ .$$

This is the sample proportion of the observed data less than or equal to the argument y of the function F_n. Notice that F_n also happens to be a CDF, albeit a random CDF in the sense that if we ran the experiment again we would get a different *value* for this function.

For simplicity we consider only the case that the true CDF F is continuous. This means that with probability 1 there no ties and so each observed x_i is distinct, although they may be tied to say 2 or 4 decimal places. The *random* CDF F_n is a discrete CDF. It has jumps at arguments y that correspond to each x_i, that is n jump points. At each of these jump points the probability mass is $\frac{1}{n}$. Thus the random CDF F_n is a the CDF of a distribution that can take values $x_1, x_2, \ldots, x_n$ and each with probability $\frac{1}{n}$. A simple algorithm to simulate

$$X^* \sim F_n$$

is to simulate a uniform (0,1) random variable U and consider the mapping

$$\begin{aligned}
X^* &= x_1 \text{ if } 0 \le U < \frac{1}{n} \\
X^* &= x_2 \text{ if } \frac{1}{n} \le U < \frac{2}{n} \\
X^* &= x_3 \text{ if } \frac{2}{n} \le U < \frac{3}{n} \\
&\vdots \\
X^* &= x_{n-1} \text{ if } \frac{n-2}{n} \le U < \frac{n-1}{n} \\
X^* &= x_n \text{ if } \frac{n-1}{n} \le U \le 1
\end{aligned}$$

The student should verify that X^* has CDF F_n.

See the homework problem 44 at the end of chapter 2 for this method. There you applied this simulation method to generate a geometric random variable from U. If some the x_i are equal, then F_n will have some jumps of size $\frac{2}{n}$ (in the case of one tied pair) or possibly bigger jumps that are multiples of $\frac{1}{n}$.

Method 2

R has an even simpler method, a program called *sample*. This will sample from a vector of n a sample with replacement, by using the appropriate arguments. There are also functions which carry out the nonparametric bootstrap method, one being in an R package called *boot*.

Suppose that

$$U \sim \text{Uniform}\left(\{1, 2, \ldots, \text{n}\}\right) \ .$$

Let the n data points be $x_1, \ldots, x_n$. Define

$$X^* = x_U \ .$$

We can also think of this more explicitly as a function. Let h be a function with domain $\{1, 2, \ldots, n\}$ given by the formula

$$h(i) = x_i \ .$$

Then $X^* = h(U)$. The student should verify that X^* has cdf given by the empirical distribution function of this data.

This method does not need to be modified even if some of the x_i are tied. The r.v. X^* random chooses amongst the n objects $x_1, \ldots, x_n$.

Method 2 is in fact easier to implement or code, as long as the computing language has a random number generator, and is even easier to code if the language has a uniform discrete random number generator. R has such a program or function, called *sample*.

Further Discussion of the Nonparametric Bootstrap

The Theory to prove that the nonparametric bootstrap methods works is beyond what we discuss in this course. It depends on a study of how well the EDF F_n approximates (or converges to) the true CDF F. In Statistics 3657 we studied, for a fixed value y the sequence $F_n(y)$. Since this is an average of iid Bernoulli random variables $I(X_i \leq y)$, the Law of Large Number and Central Limit both applied since the conditions of these Theorems are satisfied. In particular we can conclude that

$$F_n(y) \to F(y) \text{ in probability, as } n \to \infty \ .$$

Thus $F_n(y)$ (as a sequence of r.v.s indexed by n) is a consistent estimator of $F(y)$.

The nonparametric bootstrap works very well for random variables of the form

$$W = \frac{\sqrt{n}\left(\hat{\theta}_n - \theta\right)}{\sqrt{\hat{\tau}^2}} \ . \tag{1}$$

Recall also that X_i are iid F. To help us interpret this first recognize that W is a function of r.v.s $X_1, \ldots, X_n$ and the "parameter" $\theta(F)$, that is we can view θ as depending on F. For example $\theta = E_F(X)$ might be the mean of F, or it might be $\theta = E_F(|X|)$ the mean absolute value. It may also be the formula used to obtain a method of moments estimator.

The bootstrap method is to consider the r.v. similar to (1) after recognizing that W is obtained from $X_1, \ldots, X_n, \theta(F)$ by a given formula. The bootstrap r.v. W^* uses this same formula and is obtained from the r.v. $X_1^*, \ldots, X_n^*, \theta(F_n)$. If we *view* W as the r.v. obtained from a *physical* experiment with cdf F and r.v.s $X_1, \ldots, X_n$ iid F we then *view* W^* as the analogous experiment based on cdf F_n and $X_1^*, \ldots, X_n^*$ iid F_n.

The bootstrap paradigm is that the quantiles of W^* is a good approximation for the quantiles of W. From this if we use W to generate a confidence interval for θ using quantiles of W, then by the bootstrap we use W to generate a confidence interval for θ but using the quantiles of W^* in place of the quantiles of W.

The nonparametric bootstrap method can be applied to other settings. In the case of simple regression, with iid random innovation or random errors ϵ_i, one can fit the regression model and calculate residuals

$$r_i = \hat{\epsilon}_i = Y_i - \hat{Y}_i \ .$$

One can then consider r.v.s $r_i^* = \epsilon_i^*$ that are iid from the empirical distribution of the residuals. Consider simple linear regression with data (x_i, Y_i) from the model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \ .$$

Fit the model. The bootstrap data is then given by

$$Y_i^* = \hat{\beta}_0 + \hat{\beta}_1 x_i + \epsilon_i^* \ .$$

From this one could then obtain the r.v. $\hat{\beta}^*$, the fitted coefficient from the bootstrap r.v. One could then for example study the bootstrap distribution of for example

$$\sqrt{n}\left(\hat{\beta}_1^* - \hat{\beta}_1\right) \ .$$

The quantiles of this bootstrap distribution then estimate the distribution of the r.v.

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right)$$

which can then be used to obtain a confidence interval for β_1 based on these bootstrap quantiles. This bootstrap method does not require that the model be a normal regression model.

Another version of a bootstrap for regression is to resample pairs (x_i^*, Y_i^*). It has slightly different properties than the method described above, but depends less on the correctness of the linear regression assumptions.

Bootstrap methods are versatile, but depend heavily on computing to implement the method.