The bootstrap is a powerful tool for inference in statistical models. This note describes briefly what the bootstrap is and how to use it.

Suppose \( X_1, \ldots, X_n \) is an i.i.d. sample drawn from an unknown distribution \( P \). Consider testing the null hypothesis \( H_0 : \theta = \theta_0 \) versus the alternative \( H_1 : \theta \neq \theta_0 \) where \( \theta \) is a parameter related to the distribution \( P \). The first step is to construct a test statistic \( T_n \), which is some function of the data. Let \( F_n(t) = \Pr(T_n \leq t) \) denote the distribution of the test statistic under the null hypothesis. Often we do not know \( F_n \) because we do not know \( P \). The exception to this rule is if \( T_n \) is what is called a pivotal statistic.

If \( F_n \) is unknown we instead must approximate it. The asymptotic approach relies on the limiting distribution, \( F_\infty(t) = \lim_{n \to \infty} F_n(t) \) which can usually be derived analytically. The limiting distribution is used to obtain a critical value, \( c_n \), such that \( F_\infty(c_n) = 1 - \alpha/2 \). The size of the test is then \( \Pr(|T_n| > c_n) = 2(1 - F_n(c_n)) \approx \alpha \). If \( F_\infty \) is a good approximation of \( F_n \), then the critical value \( c_n \) provides a test procedure with accurate size control.

The bootstrap is an alternative way of approximating \( F_n(t) \) for the purposes of constructing a critical value. There are two reasons why we might need a different approximation. First, in some cases the bootstrap will provide a better approximation. Second, the asymptotic approach may be difficult computationally.

The bootstrap uses the actual sample to estimate the distribution of the test
statistic. Let $P_n$ denote the empirical distribution based on the sample $X_1, \ldots, X_n$. For example, if the sample consists of $n$ distinct values $P_n$ consists of the distribution that assigns probability $1/n$ to each of these values. This is often a good estimate of $P$. The bootstrap distribution is the distribution of $T_n$ that would result if $P_n$ was the actual distribution of the sample.

In some cases we can figure out analytically what the bootstrap distribution is. Usually however the bootstrap distribution is obtained using a Monte Carlo simulation. First, sample the data with replacement. This means that we randomly select one of the values $X_1, \ldots, X_n$ and call this $X^*_1$ and then we repeat this $n$ times obtaining a new sample $X^*_1, \ldots, X^*_n$. The replacement part is important because otherwise we obtain the original sample, just reordered. We then compute the test statistic on this sample and call this $T^*_n$. Then we perform both of these steps repeatedly to get $T^*_n, \ldots, T^*_n,S$. Then we can use this sample of $S$ values to compute an approximation of the distribution, $F_n(t)$. Take $F_n^*(t) = \sum_{i=1}^S 1(T^*_n,i \leq t)$. Then we get a critical value $c^*_n$ from $F_n^*(t)$. Note that $S$ is limited only by computational power and our patience; it is not related to the sample size.

**Example 1. Testing the population mean** Suppose $\theta$ is the mean of the distribution $P$ and that $\theta_0 = 0$. In this case our test statistic will be the t-statistic, $T_n = \bar{X}_n/s_n$ where $s^2_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. If $P$ is the normal distribution, $N(\theta, \sigma^2)$, then $T_n$ is pivotal because it is distributed according to a student’s t distribution with $n - 1$ degrees of freedom. If we do not assume that $P$ is the normal distribution then we can use the asymptotic approximation provided by the central limit theorem. According to the CLT, $T_n \rightarrow_d N(0, 1)$. This means that $F_\infty(t) = \Phi(t)$, the standard normal cdf. Thus the critical value is the usual z-score. The bootstrap approach is only slightly more difficult to compute. We draw $S$ bootstrap samples. For each bootstrap sample we compute $T^*_{n,s} = \bar{X}^*_{n,s} - \bar{X}_n$. The critical value is obtained as the $1 - \alpha/2$ quantile of the distribution of these $T^*_{n,i}$.

The bootstrap, as described, is valid under two main conditions. First, the problem must be a “regular” one. This condition is best understood by considering some
counterexamples. Two simple counterexamples are (i) testing a composite null hypothesis with a kink or (ii) when $\theta$ is the maximum value of the sample. An example of (i) is when the null hypothesis is that two parameters are both positive and the alternative is that one or both are negative. The second condition is that the sample must be i.i.d. The basic idea of the bootstrap is more general than the bootstrap described above, though. Generally we can try to find a distribution $P_n$ by resampling the data in a way that mimics the way the data was original sampled. If we believe the sample was obtained through i.i.d. random sampling then it is natural for $P_n$ to be the empirical cdf.

**Example 2. Inference in a regression** Suppose we estimate a regression model $Y_i = \beta'X_i + \varepsilon_i$. Let $\theta = \beta_l$ be the coefficient on one of the regressors. The test statistic is the t statistic $T_n = (\hat{\beta}_l - \theta_0)/s_{n,l}$, where $s_{n,l}^2$ is a consistent estimator of the variance of $\hat{\beta}_l$. If the errors are i.i.d. normal then $T_n$ is pivotal. If the errors are i.i.d. with an unknown distribution the limiting distribution is a standard normal. This leads to the usual asymptotic critical value.

There are several ways to compute a bootstrap distribution. I consider two here. The first method resamples $(Y_i, X_i)$ $S$ times to obtain bootstrap samples. For each bootstrap sample the test statistic $T_{n,i}^*$ is computed. This is sometimes called the pairs bootstrap. The pairs bootstrap can be used in many situations other than a regression.

The second method resamples the residuals. The estimated residuals are $\hat{\varepsilon}_i = Y_i - \hat{\beta}'X_i$. In this approach you draw $S$ samples of residuals, $\varepsilon_{s,1}^*, \ldots, \varepsilon_{s,n}^*$ for $s = 1, \ldots, S$. For each bootstrap sample compute dependent variables $Y_{s,i}^* = Y_i + \varepsilon_{s,i}^*$ and estimate the test statistic, $T_{n,s}^*$ on this sample. Various scaling factors for the residuals are also generally advised.

Non-iid data can be handled by different types of resampling. Suppose in the regression example that the errors are heteroskedastic. In this case the residual bootstrap is not consistent but the pairs bootstrap is. However, the pairs bootstrap can be improved upon using something called the wild bootstrap. The wild bootstrap
can be used in many different settings for non-iid data.

In a regression framework, the wild bootstrap is performed by computing \( S \) bootstrap samples where the \( s^{th} \) sample is obtained by calculating \( Y_{s,i}^* = Y_i + \varepsilon_{s,i}^* \) with \( \varepsilon_{s,i}^* = \hat{\varepsilon}_i \nu_i \) where \( \nu_i \) is a special random variable. For example we can take \( \nu_i \) equal to 1 or \(-1\) with equal probability.

The bootstrap described so far is sometimes called a *nonparametric* or *naive* bootstrap. If one wishes to do hypothesis testing in the context of, for example, maximum likelihood then there is a version of the bootstrap called the *parametric bootstrap* that provides a better approximation.

In maximum likelihood estimation we assume that \( X_1, \ldots, X_n \) is distributed iid according to a distribution with density function \( f(x \mid \theta) \). The maximum likelihood estimator, \( \hat{\theta} \), is the value of \( \theta \) that maximizes \( \sum_{i=1}^{n} \ln(f(X_i \mid \theta)) \). The parametric bootstrap obtains samples by taking \( n \) independent draws from the distribution \( f(x \mid \hat{\theta}) \).

This note has discussed bootstrap methods for the purpose of hypothesis testing. I chose to focus on obtaining an approximation to the sampling distribution of test statistics and the computation of critical values because it is sometimes only in this case that the bootstrap provides what is called an *asymptotic refinement*. This means that the hypothesis test based on the bootstrap is more accurate than one based on the usual asymptotic critical values.

However, the bootstrap can be used to obtain an approximation to the distribution of a parameter estimate as well. This can then be used for bias correction, construction of confidence intervals, or computation of standard errors. The idea is essentially the same. Suppose the parameter of interest is \( \theta \) and the estimate is \( \hat{\theta}_n \). You can draw \( S \) bootstrap samples and compute the estimator in each sample, \( \theta_{n,1}^*, \ldots, \theta_{n,S}^* \). Then \( B_n = \theta_n - S^{-1} \sum_{s=1}^{S} \theta_{n,s}^* \) provides an approximation of the bias of the estimator. The sample standard deviation of the \( \theta_{n,s}^* \) can be used as the standard error for the estimator. And we can take the \( \alpha/2 \) quantile and the \( 1 - \alpha/2 \) quantile as the lower and upper limits for a \( 1 - \alpha \% \) confidence interval.

A useful theoretical result on the validity of the bootstrap is provided by Beran and Ducharme (1991). To state the result I need to emphasize in the notation that
the distribution of $T_n$ depends on the distribution of the sample. So $F_n(t, P)$ denotes the distribution of $T_n$ assuming that each $X_i$ has the distribution $P$. With this notation we can write the bootstrap estimate of the distribution as $F_n(t, P_n)$.

**Theorem 1.** The bootstrap provides a consistent estimator of the distribution of $T_n$ if for any $\varepsilon > 0$ and any possible distribution $P$

(i) $\lim_{n \to \infty} \Pr(\rho(P_n, P) > \varepsilon) = 0$

(ii) $F_\infty(t, P)$ is continuous in $t$ for all $P$

(iii) for any $\tilde{P}_n$ such that $\lim_{n \to \infty} \Pr(\rho(\tilde{P}_n, P) > \varepsilon) = 0$, $F_n(t, P_n) \to F_\infty(t, P)$

Further reading on the bootstrap: Joel Horowitz has a chapter in the Handbook of Econometrics in 2000 called “The Bootstrap” that is a bit long but very helpful. Also, Chapter 11 in Cameron and Trivedi provides a brief summary of bootstrap methods and there is some practical advice on the use of the bootstrap in regression in Mostly Harmless Econometrics.