Lecture 11

September 27, 2018

Permutation Tests

- We have studied hypothesis testing in the linear model via the F-test so far.

Permutation Tests

- We have studied hypothesis testing in the linear model via the F-test so far.
- Suppose we want to test a linear hypothesis about $\beta = (\beta_0, \ldots, \beta_p)^T$ in the full linear model (denoted by M).

Permutation Tests

- We have studied hypothesis testing in the linear model via the F-test so far.
- Suppose we want to test a linear hypothesis about $\beta = (\beta_0, \dots, \beta_p)^T$ in the full linear model (denoted by M).
- We first construct a reduced model which incorporates the hypothesis in the full model M. Call this reduced model m. We then look at the quantity:

$$T := \frac{(RSS(m) - RSS(M))/(p - q)}{RSS(M)/(n - p - 1)}.$$

Permutation Tests

- We have studied hypothesis testing in the linear model via the F-test so far.
- Suppose we want to test a linear hypothesis about $\beta = (\beta_0, \ldots, \beta_p)^T$ in the full linear model (denoted by M).
- We first construct a reduced model which incorporates the hypothesis in the full model M. Call this reduced model m. We then look at the quantity:

$$T := \frac{(RSS(m) - RSS(M))/(p-q)}{RSS(M)/(n-p-1)}.$$

- It makes sense to reject the null hypothesis if T is large. To answer the question: *how large is large?*

Permutation Tests

- We have studied hypothesis testing in the linear model via the F-test so far.
- Suppose we want to test a linear hypothesis about $\beta = (\beta_0, \ldots, \beta_p)^T$ in the full linear model (denoted by M).
- We first construct a reduced model which incorporates the hypothesis in the full model M. Call this reduced model m. We then look at the quantity:

$$T := \frac{(RSS(m) - RSS(M))/(p-q)}{RSS(M)/(n-p-1)}.$$

- It makes sense to reject the null hypothesis if T is large. To answer the question: *how large is large?*
- We rely on the assumption of normality of the errors i.e., $e \sim N(0, \sigma^2 I)$ to assert that $T \sim F_{p-q,n-p-1}$ under H_0

Permutation Tests

- As a result, a p-value can be obtained as $\mathbb{P}\{F_{p-q,n-p-1} > T\}$.

Permutation Tests

- As a result, a p-value can be obtained as $\mathbb{P}\{F_{p-q,n-p-1} > T\}$.
- Suppose we do not want to assume normality of errors. Is there any way to obtain a p-value? This is possible in some cases via permutation tests. We provide two examples below.

Testing for all explanatory variables

- We want to test the null hypothesis that all explanatory variables can be thrown away without assuming that $e \sim N(0, \sigma^2 I)$.

Testing for all explanatory variables

- We want to test the null hypothesis that all explanatory variables can be thrown away without assuming that $e \sim N(0, \sigma^2 I)$.
- Under the null hypothesis, we assume that if the response variable y has no relation to the explanatory variables.

Testing for all explanatory variables

- We want to test the null hypothesis that all explanatory variables can be thrown away without assuming that $e \sim N(0, \sigma^2 I)$.
- Under the null hypothesis, we assume that if the response variable y has no relation to the explanatory variables.
- Therefore, it is plausible to assume that under the null hypothesis, the values of the response variable $y_1, \ldots, y_n$ are randomly distributed between the n subjects without relation to the predictors.

Testing for all explanatory variables

- This motivates the following test:

Testing for all explanatory variables

- This motivates the following test:
  1. Randomly permute the response values: $y_1, \ldots, y_n$.

Testing for all explanatory variables

- This motivates the following test:
  1. Randomly permute the response values: $y_1, \ldots, y_n$.
  2. Calculate the quantity

$$\frac{(RSS(m) - RSS(M))/p}{RSS(M)/(n-p-1)}.$$

     with the response values being the permuted values in the pervious step.

Testing for all explanatory variables

- This motivates the following test:
  1. Randomly permute the response values: $y_1, \ldots, y_n$.
  2. Calculate the quantity

     $$\frac{(RSS(m) - RSS(M))/p}{RSS(M)/(n-p-1)}.$$

     with the response values being the permuted values in the pervious step.
  3. Repeat the above pair of steps a large number of times.

Testing for all explanatory variables

- This motivates the following test:
  1. Randomly permute the response values: $y_1, \ldots, y_n$.
  2. Calculate the quantity
  $$\frac{(RSS(m) - RSS(M))/p}{RSS(M)/(n-p-1)}.$$
  with the response values being the permuted values in the pervious step.
  3. Repeat the above pair of steps a large number of times.
  4. This results in a large number of values of the test statistic (one for each permutation of the response values). Let us call them $T_1, \ldots, T_N$. The p-value is calculated as the proportion of $T_1, \ldots, T_N$ that exceed the original test statistic value T (T is calculated with the actual unpermuted response values $y_1, \ldots, y_n$).

Permutation tests

The idea behind this test is as follows:

- From the given data, we calcuate the value of

$$\frac{(RSS(m) - RSS(M))/p}{RSS(M)/(n-p-1)}.$$

Permutation tests

The idea behind this test is as follows:

- From the given data, we calcuate the value of

$$\frac{(RSS(m) - RSS(M))/p}{RSS(M)/(n-p-1)}.$$

- We need to know how extreme this value is under the null hypothesis.

Permutation tests

The idea behind this test is as follows:

- From the given data, we calcuate the value of

$$\frac{(RSS(m) - RSS(M))/p}{RSS(M)/(n-p-1)}.$$

- We need to know how extreme this value is under the null hypothesis.
- Under the assumption of normality, we can assess this by the F-distribution.

Permutation tests

The idea behind this test is as follows:

- From the given data, we calcuate the value of

$$\frac{(RSS(m) - RSS(M))/p}{RSS(M)/(n-p-1)}.$$

- We need to know how extreme this value is under the null hypothesis.
- Under the assumption of normality, we can assess this by the F-distribution.
- But we need to do this without assuming normality. For this, we try to generate values of this quantity under the null hypothesis.

Permutation tests

The idea behind this test is as follows:

- From the given data, we calcuate the value of

$$\frac{(RSS(m) - RSS(M))/p}{RSS(M)/(n-p-1)}.$$

- We need to know how extreme this value is under the null hypothesis.
- Under the assumption of normality, we can assess this by the F-distribution.
- But we need to do this without assuming normality. For this, we try to generate values of this quantity under the null hypothesis.
- The idea is to do this by calculating the statistic after permuting the response values.

Permutation tests

- Because once the response values are permuted, all association between the response and explanatory variables breaks down so that the values of

$$\frac{(RSS(m) - RSS(M))/p}{RSS(M)/(n-p-1)}$$

  for the permuted response values resembles values generated under the null hypothesis.

Permutation tests

- Because once the response values are permuted, all association between the response and explanatory variables breaks down so that the values of

$$\frac{(RSS(m) - RSS(M))/p}{RSS(M)/(n-p-1)}$$

  for the permuted response values resembles values generated under the null hypothesis.
- The p-value is then calculated as the proportion of these values larger than the observed value.

Testing for a single explanatory variable

Testing for a single explanatory variable

How do we test if, say, the first explanatory variable is useful?

Testing for a single explanatory variable

How do we test if, say, the first explanatory variable is useful?

- We calcuate the t-statistic:

$$\frac{\hat{\beta}_1}{s.e(\hat{\beta}_1)}$$

  and calculate p-value by comparing it with the t_{n-p-1} distribution (which requires normality).

Testing for a single explanatory variable

How do we test if, say, the first explanatory variable is useful?

- We calcuate the t-statistic:

$$\frac{\hat{\beta}_1}{s.e(\hat{\beta}_1)}$$

  and calculate p-value by comparing it with the t_{n-p-1} distribution (which requires normality).
- How to do this without normality?

Testing for a single explanatory variable

How do we test if, say, the first explanatory variable is useful?

- We calcuate the t-statistic:

$$\frac{\hat{\beta}_1}{s.e(\hat{\beta}_1)}$$

  and calculate p-value by comparing it with the t_{n-p-1} distribution (which requires normality).
- How to do this without normality?
- We can follow the permutation test by permuting the values of x_1.

Testing for a single explanatory variable

How do we test if, say, the first explanatory variable is useful?

- We calcuate the t-statistic:

$$\frac{\hat{\beta}_1}{s.e(\hat{\beta}_1)}$$

  and calculate p-value by comparing it with the t_{n-p-1} distribution (which requires normality).
- How to do this without normality?
- We can follow the permutation test by permuting the values of x_1.
- For each permutation, we calculate the t-statistic and the p-value is the proportion of these t-values which are larger than the observed t-value in absolute value.

The Bootstrap

The Bootstrap

- The problem with permutation tests is that they cannot be used to test arbitrary hypotheses about β (for example, there is no natural permuation test for the hypothesis $H_0 : \beta_1 = \beta_2 + \beta_5, \beta_2 - 4\beta_3 = 2$.

The Bootstrap

- The problem with permutation tests is that they cannot be used to test arbitrary hypotheses about β (for example, there is no natural permuation test for the hypothesis $H_0 : \beta_1 = \beta_2 + \beta_5, \beta_2 - 4\beta_3 = 2$.
- Moreover, there is no way of obtaining confidence intervals for components of β by permutation methods.

The Bootstrap

- The problem with permutation tests is that they cannot be used to test arbitrary hypotheses about β (for example, there is no natural permuation test for the hypothesis $H_0 : \beta_1 = \beta_2 + \beta_5, \beta_2 - 4\beta_3 = 2$.
- Moreover, there is no way of obtaining confidence intervals for components of β by permutation methods.
- The bootstrap is a general technique that can be used to obtain confidence intervals and to carry out hypothesis tests without reliance on the normal regression theory.

The Bootstrap

- Before explaining how bootstrap works in regression, let me first explain its workings for the simpler setting of i.i.d data.

The Bootstrap

- Before explaining how bootstrap works in regression, let me first explain its workings for the simpler setting of i.i.d data.
- Suppose $X_1, \dots, X_n$ are i.i.d observations having common distribution F.

The Bootstrap

- Before explaining how bootstrap works in regression, let me first explain its workings for the simpler setting of i.i.d data.
- Suppose $X_1, \ldots, X_n$ are i.i.d observations having common distribution F.
- Assume that each X_i has mean μ and finite variance σ^2.

The Bootstrap

- Before explaining how bootstrap works in regression, let me first explain its workings for the simpler setting of i.i.d data.
- Suppose $X_1, \dots, X_n$ are i.i.d observations having common distribution F.
- Assume that each X_i has mean μ and finite variance σ^2.
- We would like to estimate μ. Three natural estimators for the mean μ are:

The Bootstrap

- Before explaining how bootstrap works in regression, let me first explain its workings for the simpler setting of i.i.d data.
- Suppose $X_1, \ldots, X_n$ are i.i.d observations having common distribution F.
- Assume that each X_i has mean μ and finite variance σ^2.
- We would like to estimate μ. Three natural estimators for the mean μ are:
  1. The sample mean: $\hat{\mu}_1 := \bar{X}$.

The Bootstrap

- Before explaining how bootstrap works in regression, let me first explain its workings for the simpler setting of i.i.d data.
- Suppose $X_1, \dots, X_n$ are i.i.d observations having common distribution F.
- Assume that each X_i has mean μ and finite variance σ^2.
- We would like to estimate μ. Three natural estimators for the mean μ are:
  1. The sample mean: $\hat{\mu}_1 := \bar{X}$.
  2. The trimmed sample mean $\hat{\mu}_2$ where we drop the 5% largest and fewest observations and then take the average of the remaining observations.

The Bootstrap

- Before explaining how bootstrap works in regression, let me first explain its workings for the simpler setting of i.i.d data.
- Suppose $X_1, \dots, X_n$ are i.i.d observations having common distribution F.
- Assume that each X_i has mean μ and finite variance σ^2.
- We would like to estimate μ. Three natural estimators for the mean μ are:
  1. The sample mean: $\hat{\mu}_1 := \bar{X}$.
  2. The trimmed sample mean $\hat{\mu}_2$ where we drop the 5% largest and fewest observations and then take the average of the remaining observations.
  3. The sample median $\hat{\mu}_3$.

The Bootstrap

- I would like to determine the distribution of $\hat{\mu}_i - \mu$ for each $i = 1, \ldots, 3$.

The Bootstrap

- I would like to determine the distribution of $\hat{\mu}_i - \mu$ for each $i = 1, \ldots, 3$.
- Knowledge of this distribution will tell me all about the accuracy of each of these estimators for estimating μ. Observe that the distribution of $\hat{\mu}_i - \mu$ will depend on F in general.

The Bootstrap

- I would like to determine the distribution of $\hat{\mu}_i - \mu$ for each $i = 1, \ldots, 3$.
- Knowledge of this distribution will tell me all about the accuracy of each of these estimators for estimating μ. Observe that the distribution of $\hat{\mu}_i - \mu$ will depend on F in general.
- For some distributions F, the median may be a better estimator for μ than the mean and vice versa.

The Bootstrap

- I would like to determine the distribution of $\hat{\mu}_i - \mu$ for each $i = 1, \ldots, 3$.
- Knowledge of this distribution will tell me all about the accuracy of each of these estimators for estimating μ. Observe that the distribution of $\hat{\mu}_i - \mu$ will depend on F in general.
- For some distributions F, the median may be a better estimator for μ than the mean and vice versa.
- How does one determine the distribution of $\hat{\mu}_i - \mu$? For $i = 1$ i.e., the case of the mean, we can use the central limit theorem assuming that n is large enough.

The Bootstrap

- I would like to determine the distribution of $\hat{\mu}_i - \mu$ for each $i = 1, \ldots, 3$.
- Knowledge of this distribution will tell me all about the accuracy of each of these estimators for estimating μ. Observe that the distribution of $\hat{\mu}_i - \mu$ will depend on F in general.
- For some distributions F, the median may be a better estimator for μ than the mean and vice versa.
- How does one determine the distribution of $\hat{\mu}_i - \mu$? For $i = 1$ i.e., the case of the mean, we can use the central limit theorem assuming that n is large enough.
- This gives us that $\hat{\mu}_1 - \mu$ is approximately distributed according to the normal distribution with mean 0 and variance σ^2/n.

The Bootstrap

- The standard error of $\hat{\mu}_1$ can then be approximated by $\hat{\sigma} n^{-1/2}$ where

$$\hat{\sigma} := \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}.$$

The Bootstrap

- The standard error of $\hat{\mu}_1$ can then be approximated by $\hat{\sigma} n^{-1/2}$ where

$$\hat{\sigma} := \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}.$$

- For the other two estimators however, it is not easy to write down the asymptotic distribution of $\hat{\mu}_i - \mu$.

The Bootstrap

- The standard error of $\hat{\mu}_1$ can then be approximated by $\hat{\sigma} n^{-1/2}$ where

$$\hat{\sigma} := \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}.$$

- For the other two estimators however, it is not easy to write down the asymptotic distribution of $\hat{\mu}_i - \mu$.
- Even if F is completely known, it is not easy to write down a formula for the asymptotic distribution of $\hat{\mu}_i - \mu$.

The Bootstrap

- The standard error of $\hat{\mu}_1$ can then be approximated by $\hat{\sigma} n^{-1/2}$ where

$$\hat{\sigma} := \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}.$$

- For the other two estimators however, it is not easy to write down the asymptotic distribution of $\hat{\mu}_i - \mu$.
- Even if F is completely known, it is not easy to write down a formula for the asymptotic distribution of $\hat{\mu}_i - \mu$.
- But if F is known, one can use simulation on the computer to exactly determine the distribution of $\hat{\mu}_i - \mu$.

The Bootstrap

The algorithm to do this is the following:

The Bootstrap

The algorithm to do this is the following:

1. Suppose, for example, that F is the normal distribution with mean 5 and variance 10.

The Bootstrap

The algorithm to do this is the following:

1. Suppose, for example, that F is the normal distribution with mean 5 and variance 10.
2. We need to determine the distributon of $\hat{\mu}_i - \mu$. Suppose $i = 2$ to fix ideas. Because F is $N(5, 10)$, we have $\mu = 5$. We therefore need to find the distribution of $\hat{\mu} - 5$.

The Bootstrap

The algorithm to do this is the following:

1. Suppose, for example, that F is the normal distribution with mean 5 and variance 10.
2. We need to determine the distributon of $\hat{\mu}_i - \mu$. Suppose $i = 2$ to fix ideas. Because F is $N(5, 10)$, we have $\mu = 5$. We therefore need to find the distribution of $\hat{\mu} - 5$.
3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \ldots, N$.

The Bootstrap

The algorithm to do this is the following:

1. Suppose, for example, that F is the normal distribution with mean 5 and variance 10.
2. We need to determine the distributon of $\hat{\mu}_i - \mu$. Suppose $i = 2$ to fix ideas. Because F is $N(5, 10)$, we have $\mu = 5$. We therefore need to find the distribution of $\hat{\mu} - 5$.
3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \ldots, N$.
   3.1 Generate n observations $X_1^{(j)}, \ldots, X_n^{(j)}$ from F.

The Bootstrap

The algorithm to do this is the following:

1. Suppose, for example, that F is the normal distribution with mean 5 and variance 10.
2. We need to determine the distributon of $\hat{\mu}_i - \mu$. Suppose $i = 2$ to fix ideas. Because F is $N(5, 10)$, we have $\mu = 5$. We therefore need to find the distribution of $\hat{\mu} - 5$.
3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \ldots, N$.
   3.1 Generate n observations $X_1^{(j)}, \ldots, X_n^{(j)}$ from F.
   3.2 Compute the statistic $\hat{\mu}_2$ from the generated observations $X_1^{(j)}, \ldots, X_n^{(j)}$. Call the computed value $\hat{\mu}_2^{(j)}$.

The Bootstrap

The algorithm to do this is the following:

1. Suppose, for example, that F is the normal distribution with mean 5 and variance 10.
2. We need to determine the distributon of $\hat{\mu}_i - \mu$. Suppose $i = 2$ to fix ideas. Because F is $N(5, 10)$, we have $\mu = 5$. We therefore need to find the distribution of $\hat{\mu} - 5$.
3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \ldots, N$.
   3.1 Generate n observations $X_1^{(j)}, \ldots, X_n^{(j)}$ from F.
   3.2 Compute the statistic $\hat{\mu}_2$ from the generated observations $X_1^{(j)}, \ldots, X_n^{(j)}$. Call the computed value $\hat{\mu}_2^{(j)}$.
4. The values $\hat{\mu}_2^{(j)} - 5$ for $j = 1, \ldots, N$ form a (very large) sample from the distribution of $\hat{\mu}_2 - \mu$. These determine the distribution of $\hat{\mu}_2 - \mu$.

The Bootstrap

- How does one find the distribution of $\hat{\mu}_2 - \mu$ when F is unknown?

The Bootstrap

- How does one find the distribution of $\hat{\mu}_2 - \mu$ when F is unknown?
- Because we know how to solve the problem when F is known, a natural idea is to estimate F by some known distribution $\hat{F}$ and implement the above algorithm with $\hat{F}$. This is bootstrap.

The Bootstrap

- How does one find the distribution of $\hat{\mu}_2 - \mu$ when F is unknown?
- Because we know how to solve the problem when F is known, a natural idea is to estimate F by some known distribution $\hat{F}$ and implement the above algorithm with $\hat{F}$. This is bootstrap.
- Different versions of the bootstrap correspond to different choices of the estimate $\hat{F}$. There are two broad choices for $\hat{F}$:

The Bootstrap

- How does one find the distribution of $\hat{\mu}_2 - \mu$ when F is unknown?
- Because we know how to solve the problem when F is known, a natural idea is to estimate F by some known distribution $\hat{F}$ and implement the above algorithm with $\hat{F}$. This is bootstrap.
- Different versions of the bootstrap correspond to different choices of the estimate $\hat{F}$. There are two broad choices for $\hat{F}$:
    1. One can assume that $\hat{F}$ has a parametric form such as a normal distribution and then estimate the parameters from the observed data $X_1, \ldots, X_n$. This is called *parametric bootstrap*.

The Bootstrap

- How does one find the distribution of $\hat{\mu}_2 - \mu$ when F is unknown?
- Because we know how to solve the problem when F is known, a natural idea is to estimate F by some known distribution $\hat{F}$ and implement the above algorithm with $\hat{F}$. This is bootstrap.
- Different versions of the bootstrap correspond to different choices of the estimate $\hat{F}$. There are two broad choices for $\hat{F}$:
    1. One can assume that $\hat{F}$ has a parametric form such as a normal distribution and then estimate the parameters from the observed data $X_1, \ldots, X_n$. This is called *parametric bootstrap*.
    2. One can try to estimate F nonparametrically. The most common way of doing this is to take $\hat{F}$ to be the *empirical distribution* of the data $X_1, \ldots, X_n$. This is called *nonparametric bootstrap*. Usually people mean this when referring to the bootstrap.

The Bootstrap

- The empirical distribution of the data $X_1, \ldots, X_n$ is a discrete probability measure that gives the mass $1/n$ to each data point $X_1, \ldots, X_n$. Its distribution function is given by

$$\hat{F}(t) := \frac{1}{n} \left(\text{number of points } X_1, \ldots, X_n \text{ that are } \leq t \right).$$

The Bootstrap Algorithm for i.i.d data

- Suppose $X_1, \ldots, X_n$ are i.i.d observations having common distribution F.

The Bootstrap Algorithm for i.i.d data

- Suppose $X_1, \dots, X_n$ are i.i.d observations having common distribution F.
- Let $\mu = \mu(F)$ be a parameter that depends on F. For example, μ could be the mean of F or μ could be the standard deviation of F.

The Bootstrap Algorithm for i.i.d data

- Suppose $X_1, \dots, X_n$ are i.i.d observations having common distribution F.
- Let $\mu = \mu(F)$ be a parameter that depends on F. For example, μ could be the mean of F or μ could be the standard deviation of F.
- Let T be a statistic that is computed from the data $X_1, \dots, X_n$. For example, T could be the sample mean or the sample median or the sample standard deviation etc.

The Bootstrap Algorithm for i.i.d data

- Suppose $X_1, \ldots, X_n$ are i.i.d observations having common distribution F.
- Let $\mu = \mu(F)$ be a parameter that depends on F. For example, μ could be the mean of F or μ could be the standard deviation of F.
- Let T be a statistic that is computed from the data $X_1, \ldots, X_n$. For example, T could be the sample mean or the sample median or the sample standard deviation etc.
- We want to determine the distribution of $T - \mu$. The bootstrap algorithm for doing this is given below.

The Bootstrap Algorithm for i.i.d data

- Suppose $X_1, \dots, X_n$ are i.i.d observations having common distribution F.
- Let $\mu = \mu(F)$ be a parameter that depends on F. For example, μ could be the mean of F or μ could be the standard deviation of F.
- Let T be a statistic that is computed from the data $X_1, \dots, X_n$. For example, T could be the sample mean or the sample median or the sample standard deviation etc.
- We want to determine the distribution of $T - \mu$. The bootstrap algorithm for doing this is given below.
- Remember the notion of empirical distribution. The empirical distribution of the data $X_1, \dots, X_n$ is a discrete probability measure that gives the mass $1/n$ to each data point $X_1, \dots, X_n$.

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

$$\hat{F}(t) := \frac{1}{n}\left(\text{number of points } X_1, \ldots, X_n \text{ that are } \leq t\right).$$

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

$$\hat{F}(t) := \frac{1}{n}\left(\text{number of points } X_1, \ldots, X_n \text{ that are } \leq t\right).$$

- The following is the bootstrap algorithm for approximating the distribution of $T - \mu$:

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

$$\hat{F}(t) := \frac{1}{n}\left(\text{number of points } X_1, \ldots, X_n \text{ that are } \leq t\right).$$

- The following is the bootstrap algorithm for approximating the distribution of $T - \mu$:
    1. Estimate the distribution of the data F by the empirical distribution $\hat{F}$.

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

$$\hat{F}(t) := \frac{1}{n}\left(\text{number of points } X_1, \ldots, X_n \text{ that are } \le t\right).$$

- The following is the bootstrap algorithm for approximating the distribution of $T - \mu$:
  1. Estimate the distribution of the data F by the empirical distribution $\hat{F}$.
  2. Calculate the value of the parameter μ for the empirical distribution $\hat{F}$. Let us call this value $\mu(\hat{F})$.

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

$$\hat{F}(t) := \frac{1}{n} \left(\text{number of points } X_1, \ldots, X_n \text{ that are } \leq t\right).$$

- The following is the bootstrap algorithm for approximating the distribution of $T - \mu$:
  1. Estimate the distribution of the data F by the empirical distribution $\hat{F}$.
  2. Calculate the value of the parameter μ for the empirical distribution $\hat{F}$. Let us call this value $\mu(\hat{F})$.
  3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \ldots, N$.

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

$$\hat{F}(t) := \frac{1}{n}\,(\text{number of points } X_1, \ldots, X_n \text{ that are } \le t)\,.$$

- The following is the bootstrap algorithm for approximating the distribution of $T - \mu$:
    1. Estimate the distribution of the data F by the empirical distribution $\hat{F}$.
    2. Calculate the value of the parameter μ for the empirical distribution $\hat{F}$. Let us call this value $\mu(\hat{F})$.
    3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \ldots, N$.
        3.1 Generate n observations $X_1^{(j)}, \ldots, X_n^{(j)}$ from $\hat{F}$.

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

$$\hat{F}(t) := \frac{1}{n} (\text{number of points } X_1, \ldots, X_n \text{ that are } \leq t) .$$

- The following is the bootstrap algorithm for approximating the distribution of $T - \mu$:
  1. Estimate the distribution of the data F by the empirical distribution $\hat{F}$.
  2. Calculate the value of the parameter μ for the empirical distribution $\hat{F}$. Let us call this value $\mu(\hat{F})$.
  3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \ldots, N$.
     - 3.1 Generate n observations $X_1^{(j)}, \ldots, X_n^{(j)}$ from $\hat{F}$.
     - 3.2 Compute the statistic T from the generated observations $X_1^{(j)}, \ldots, X_n^{(j)}$. Call the computed value $T^{(j)}$.

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

$$\hat{F}(t) := \frac{1}{n}\left(\text{number of points } X_1, \ldots, X_n \text{ that are } \leq t\right).$$

- The following is the bootstrap algorithm for approximating the distribution of $T - \mu$:
  1. Estimate the distribution of the data F by the empirical distribution $\hat{F}$.
  2. Calculate the value of the parameter μ for the empirical distribution $\hat{F}$. Let us call this value $\mu(\hat{F})$.
  3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \ldots, N$.
     3.1 Generate n observations $X_1^{(j)}, \ldots, X_n^{(j)}$ from $\hat{F}$.
     3.2 Compute the statistic T from the generated observations $X_1^{(j)}, \ldots, X_n^{(j)}$. Call the computed value $T^{(j)}$.
  4. The empirical distribution of the values $T^{(j)} - \mu(\hat{F})$ for $j = 1, \ldots, N$ is an estimate of the distribution of $T - \mu$.