Lecture 13

October 2, 2018

The Bootstrap Algorithm for i.i.d data

The Bootstrap Algorithm for i.i.d data

- Suppose $X_1, \ldots, X_n$

The Bootstrap Algorithm for i.i.d data

- Suppose $X_1, \ldots, X_n$ are i.i.d observations

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 $X_1, \ldots, X_n$ are i.i.d observations having common distribution F.
- Let $\mu = \mu(F)$

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.

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,

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

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.

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

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 $X_1, \ldots, X_n$.

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,

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

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

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 $X_1, \dots, X_n$ 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 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.

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 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

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

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 Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

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

- Bootstrap algorithm for approximating the distribution of $T - \mu$:

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

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

- Bootstrap algorithm for approximating the distribution of $T - \mu$:
  1. Estimate the distribution of the data F

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

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

- 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

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

- 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

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

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

- 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 $\hat{F}$.

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

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

- 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

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

- 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 $N = 5000$)

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

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

- 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

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

- 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 $X_1^{(j)}, \ldots, X_n^{(j)}$

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

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

- 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

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

- 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 $X_1^{(j)}, \ldots, X_n^{(j)}$.

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

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

- 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

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

- 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 $j = 1, \dots, N$.
     3.1 Generate n observations $X_1^{(j)}, \dots, X_n^{(j)}$ from $\hat{F}$.
     3.2 Compute the statistic T from the generated observations $X_1^{(j)}, \dots, X_n^{(j)}$. Call the computed value $T^{(j)}$.
  4. The empirical distribution of the values $T^{(j)} - \mu(\hat{F})$ for $j = 1, \dots, N$

The Bootstrap Algorithm for i.i.d data

- Its distribution function is given by

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

- 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 $T - \mu$.

Bootstrap in Regression

Bootstrap in Regression

- We shall now study our linear model

Bootstrap in Regression

- We shall now study our linear model and see how bootstrap can be used to produce confidence intervals for $\beta_0, \ldots, \beta_p$

Bootstrap in Regression

- We shall now study our linear model and see how bootstrap can be used to produce confidence intervals for $\beta_0, \ldots, \beta_p$ without relying on the normal regression theory.

Bootstrap in Regression

- We shall now study our linear model and see how bootstrap can be used to produce confidence intervals for $\beta_0, \ldots, \beta_p$ without relying on the normal regression theory.
- Recall our linear model for Y: $Y = X\beta + e$.

Bootstrap in Regression

- We shall now study our linear model and see how bootstrap can be used to produce confidence intervals for $\beta_0, \ldots, \beta_p$ without relying on the normal regression theory.
- Recall our linear model for Y: $Y = X\beta + e$. We take X to be fixed

Bootstrap in Regression

- We shall now study our linear model and see how bootstrap can be used to produce confidence intervals for $\beta_0, \ldots, \beta_p$ without relying on the normal regression theory.
- Recall our linear model for Y: $Y = X\beta + e$. We take X to be fixed and the errors $e_1, \ldots, e_n$ to be i.i.d with mean zero and common variance σ^2 .

Bootstrap in Regression

- We shall now study our linear model and see how bootstrap can be used to produce confidence intervals for $\beta_0, \ldots, \beta_p$ without relying on the normal regression theory.
- Recall our linear model for Y: $Y = X\beta + e$. We take X to be fixed and the errors $e_1, \ldots, e_n$ to be i.i.d with mean zero and common variance σ^2 .
- We do not need to assume that they are normal.

Bootstrap in Regression

- We shall now study our linear model and see how bootstrap can be used to produce confidence intervals for $\beta_0, \ldots, \beta_p$ without relying on the normal regression theory.
- Recall our linear model for Y: $Y = X\beta + e$. We take X to be fixed and the errors $e_1, \ldots, e_n$ to be i.i.d with mean zero and common variance σ^2 .
- We do not need to assume that they are normal. Let the common distribution of $e_1, \ldots, e_n$ be denoted by G.

Bootstrap in Regression

- We shall now study our linear model and see how bootstrap can be used to produce confidence intervals for $\beta_0, \ldots, \beta_p$ without relying on the normal regression theory.
- Recall our linear model for Y: $Y = X\beta + e$. We take X to be fixed and the errors $e_1, \ldots, e_n$ to be i.i.d with mean zero and common variance σ^2 .
- We do not need to assume that they are normal. Let the common distribution of $e_1, \ldots, e_n$ be denoted by G.
- Let $\hat{\beta}$ denote the usual least squares estimate of β

Bootstrap in Regression

- We shall now study our linear model and see how bootstrap can be used to produce confidence intervals for $\beta_0, \ldots, \beta_p$ without relying on the normal regression theory.
- Recall our linear model for Y: $Y = X\beta + e$. We take X to be fixed and the errors $e_1, \ldots, e_n$ to be i.i.d with mean zero and common variance σ^2 .
- We do not need to assume that they are normal. Let the common distribution of $e_1, \ldots, e_n$ be denoted by G.
- Let $\hat{\beta}$ denote the usual least squares estimate of β (assume that X has full column rank).

Bootstrap in Regression

- We shall now study our linear model and see how bootstrap can be used to produce confidence intervals for $\beta_0, \dots, \beta_p$ without relying on the normal regression theory.
- Recall our linear model for Y: $Y = X\beta + e$. We take X to be fixed and the errors $e_1, \dots, e_n$ to be i.i.d with mean zero and common variance σ^2 .
- We do not need to assume that they are normal. Let the common distribution of $e_1, \dots, e_n$ be denoted by G.
- Let $\hat{\beta}$ denote the usual least squares estimate of β (assume that X has full column rank). We would like to determine the distribution of say $\hat{\beta}_j - \beta_j$ for some j.

Bootstrap in Regression

- We shall now study our linear model and see how bootstrap can be used to produce confidence intervals for $\beta_0, \ldots, \beta_p$ without relying on the normal regression theory.
- Recall our linear model for Y: $Y = X\beta + e$. We take X to be fixed and the errors $e_1, \ldots, e_n$ to be i.i.d with mean zero and common variance σ^2 .
- We do not need to assume that they are normal. Let the common distribution of $e_1, \ldots, e_n$ be denoted by G.
- Let $\hat{\beta}$ denote the usual least squares estimate of β (assume that X has full column rank). We would like to determine the distribution of say $\hat{\beta}_j - \beta_j$ for some j. Without loss of generality, let us take $j = 1$.

Bootstrap in Regression

- Let the joint distribution of $Y = (Y_1, \ldots, Y_n)$ be denoted by F.

Bootstrap in Regression

- Let the joint distribution of $Y = (Y_1, \ldots, Y_n)$ be denoted by F.
- Note that F depends on β and G.

Bootstrap in Regression

- Let the joint distribution of $Y = (Y_1, \ldots, Y_n)$ be denoted by F.
- Note that F depends on β and G.
- If we knew β and G,

Bootstrap in Regression

- Let the joint distribution of $Y = (Y_1, \ldots, Y_n)$ be denoted by F.
- Note that F depends on β and G.
- If we knew β and G, then we can determine the distribution of $\hat{\beta}_1 - \beta_1$ by simulation:

Bootstrap in Regression

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \ldots, -0.883)$

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \ldots, -0.883)$ and *G* is the standard *t*-distribution with 5 degrees of freedom.

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \ldots, -0.883)$ and *G* is the standard *t*-distribution with 5 degrees of freedom.
2. Since we are given that $\beta_1 = 3.65$,

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \ldots, -0.883)$ and G is the standard t-distribution with 5 degrees of freedom.
2. Since we are given that $\beta_1 = 3.65$, we need to determine the distribution of $\hat{\beta}_1 - 3.65$.

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$ and G is the standard t-distribution with 5 degrees of freedom.
2. Since we are given that $\beta_1 = 3.65$, we need to determine the distribution of $\hat{\beta}_1 - 3.65$.
3. Fix a large number N (say $N = 5000$)

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$ and G is the standard t-distribution with 5 degrees of freedom.
2. Since we are given that $\beta_1 = 3.65$, we need to determine the distribution of $\hat{\beta}_1 - 3.65$.
3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \dots, N$.

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$ and G is the standard t-distribution with 5 degrees of freedom.
2. Since we are given that $\beta_1 = 3.65$, we need to determine the distribution of $\hat{\beta}_1 - 3.65$.
3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \dots, N$.
    3.1 Generate n observations $e_1^{(j)}, \dots, e_n^{(j)}$ from the distribution G.

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$ and G is the standard t-distribution with 5 degrees of freedom.
2. Since we are given that $\beta_1 = 3.65$, we need to determine the distribution of $\hat{\beta}_1 - 3.65$.
3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \dots, N$.
   3.1 Generate n observations $e_1^{(j)}, \dots, e_n^{(j)}$ from the distribution G. Let $e^{(j)}$ denote the $n \times 1$ vector consisting of these observations.

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \ldots, -0.883)$ and G is the standard t-distribution with 5 degrees of freedom.
2. Since we are given that $\beta_1 = 3.65$, we need to determine the distribution of $\hat{\beta}_1 - 3.65$.
3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \ldots, N$.
   3.1 Generate n observations $e_1^{(j)}, \ldots, e_n^{(j)}$ from the distribution G. Let $e^{(j)}$ denote the $n \times 1$ vector consisting of these observations.
   3.2 Add the vector $X\beta$

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$ and G is the standard t-distribution with 5 degrees of freedom.
2. Since we are given that $\beta_1 = 3.65$, we need to determine the distribution of $\hat{\beta}_1 - 3.65$.
3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \dots, N$.
   - 3.1 Generate n observations $e_1^{(j)}, \dots, e_n^{(j)}$ from the distribution G. Let $e^{(j)}$ denote the $n \times 1$ vector consisting of these observations.
   - 3.2 Add the vector $X\beta$ (with $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$)

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$ and G is the standard t-distribution with 5 degrees of freedom.
2. Since we are given that $\beta_1 = 3.65$, we need to determine the distribution of $\hat{\beta}_1 - 3.65$.
3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \dots, N$.
   - 3.1 Generate n observations $e_1^{(j)}, \dots, e_n^{(j)}$ from the distribution G. Let $e^{(j)}$ denote the $n \times 1$ vector consisting of these observations.
   - 3.2 Add the vector $X\beta$ (with $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$) to $e^{(j)}$

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$ and G is the standard t-distribution with 5 degrees of freedom.
2. Since we are given that $\beta_1 = 3.65$, we need to determine the distribution of $\hat{\beta}_1 - 3.65$.
3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \dots, N$.
   - 3.1 Generate n observations $e_1^{(j)}, \dots, e_n^{(j)}$ from the distribution G. Let $e^{(j)}$ denote the $n \times 1$ vector consisting of these observations.
   - 3.2 Add the vector $X\beta$ (with $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$) to $e^{(j)}$ to create a new vector called $Y^{(j)}$.

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$ and G is the standard t-distribution with 5 degrees of freedom.
2. Since we are given that $\beta_1 = 3.65$, we need to determine the distribution of $\hat{\beta}_1 - 3.65$.
3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \dots, N$.
   - 3.1 Generate n observations $e_1^{(j)}, \dots, e_n^{(j)}$ from the distribution G. Let $e^{(j)}$ denote the $n \times 1$ vector consisting of these observations.
   - 3.2 Add the vector $X\beta$ (with $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$) to $e^{(j)}$ to create a new vector called $Y^{(j)}$.
   - 3.3 Regress $Y^{(j)}$ on X.

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$ and G is the standard t-distribution with 5 degrees of freedom.
2. Since we are given that $\beta_1 = 3.65$, we need to determine the distribution of $\hat{\beta}_1 - 3.65$.
3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \dots, N$.
   3.1 Generate n observations $e_1^{(j)}, \dots, e_n^{(j)}$ from the distribution G. Let $e^{(j)}$ denote the $n \times 1$ vector consisting of these observations.
   3.2 Add the vector $X\beta$ (with $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$) to $e^{(j)}$ to create a new vector called $Y^{(j)}$.
   3.3 Regress $Y^{(j)}$ on X. Call the resulting estimated coefficient vector $\hat{\beta}^{(j)}$.

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$ and G is the standard t-distribution with 5 degrees of freedom.
2. Since we are given that $\beta_1 = 3.65$, we need to determine the distribution of $\hat{\beta}_1 - 3.65$.
3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \dots, N$.
   - 3.1 Generate n observations $e_1^{(j)}, \dots, e_n^{(j)}$ from the distribution G. Let $e^{(j)}$ denote the $n \times 1$ vector consisting of these observations.
   - 3.2 Add the vector $X\beta$ (with $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$) to $e^{(j)}$ to create a new vector called $Y^{(j)}$.
   - 3.3 Regress $Y^{(j)}$ on X. Call the resulting estimated coefficient vector $\hat{\beta}^{(j)}$.
4. The empirical distribution of the values $\hat{\beta}_1^{(j)} - 3.65$ for $j = 1, \dots, N$

Bootstrap in Regression

1. Suppose $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$ and G is the standard t-distribution with 5 degrees of freedom.
2. Since we are given that $\beta_1 = 3.65$, we need to determine the distribution of $\hat{\beta}_1 - 3.65$.
3. Fix a large number N (say $N = 5000$) and repeat the following steps for $j = 1, \dots, N$.
   3.1 Generate n observations $e_1^{(j)}, \dots, e_n^{(j)}$ from the distribution G. Let $e^{(j)}$ denote the $n \times 1$ vector consisting of these observations.
   3.2 Add the vector $X\beta$ (with $\beta = (10.45, 3.65, -2.54, 9.94, \dots, -0.883)$) to $e^{(j)}$ to create a new vector called $Y^{(j)}$.
   3.3 Regress $Y^{(j)}$ on X. Call the resulting estimated coefficient vector $\hat{\beta}^{(j)}$.
4. The empirical distribution of the values $\hat{\beta}_1^{(j)} - 3.65$ for $j = 1, \dots, N$ gives the distribution of $\hat{\beta}_1 - \beta_1$.

Bootstrap in Regression

Bootstrap in Regression

- What can we do if we do not know β and G?

Bootstrap in Regression

- What can we do if we do not know β and G? We simply estimate them from the data.

Bootstrap in Regression

- What can we do if we do not know β and G? We simply estimate them from the data.
- β is estimated by the least squares estimate $\hat{\beta}$.

Bootstrap in Regression

- What can we do if we do not know β and G? We simply estimate them from the data.
- β is estimated by the least squares estimate $\hat{\beta}$.
- How to estimate G?

Bootstrap in Regression

- What can we do if we do not know β and G? We simply estimate them from the data.
- β is estimated by the least squares estimate $\hat{\beta}$.
- How to estimate G? For this we use residuals $\hat{e}$

Bootstrap in Regression

- What can we do if we do not know β and G? We simply estimate them from the data.
- β is estimated by the least squares estimate $\hat{\beta}$.
- How to estimate G? For this we use residuals $\hat{e}$ resulting in the method called *Residual Bootstrap*.

Bootstrap in Regression

- What can we do if we do not know β and G? We simply estimate them from the data.
- β is estimated by the least squares estimate $\hat{\beta}$.
- How to estimate G? For this we use residuals $\hat{e}$ resulting in the method called *Residual Bootstrap*.
- Recall that

Bootstrap in Regression

- What can we do if we do not know β and G? We simply estimate them from the data.
- β is estimated by the least squares estimate $\hat{\beta}$.
- How to estimate G? For this we use residuals $\hat{e}$ resulting in the method called *Residual Bootstrap*.
- Recall that

$$\hat{e} = (I - H)Y$$

- Because $Y = X\beta + e$

Bootstrap in Regression

- What can we do if we do not know β and G? We simply estimate them from the data.
- β is estimated by the least squares estimate $\hat{\beta}$.
- How to estimate G? For this we use residuals $\hat{e}$ resulting in the method called *Residual Bootstrap*.
- Recall that

$$\hat{e} = (I - H)Y$$

- Because $Y = X\beta + e$ and $HX = X$,

Bootstrap in Regression

- What can we do if we do not know β and G? We simply estimate them from the data.
- β is estimated by the least squares estimate $\hat{\beta}$.
- How to estimate G? For this we use residuals $\hat{e}$ resulting in the method called *Residual Bootstrap*.
- Recall that

$$\hat{e} = (I - H)Y$$

- Because $Y = X\beta + e$ and $HX = X$, we get

Bootstrap in Regression

- What can we do if we do not know β and G? We simply estimate them from the data.
- β is estimated by the least squares estimate $\hat{\beta}$.
- How to estimate G? For this we use residuals $\hat{e}$ resulting in the method called *Residual Bootstrap*.
- Recall that

$$\hat{e} = (I - H)Y$$

- Because $Y = X\beta + e$ and $HX = X$, we get

$$\hat{e} = (I - H)(X\beta + e)$$

Bootstrap in Regression

- What can we do if we do not know β and G? We simply estimate them from the data.
- β is estimated by the least squares estimate $\hat{\beta}$.
- How to estimate G? For this we use residuals $\hat{e}$ resulting in the method called *Residual Bootstrap*.
- Recall that

$$\hat{e} = (I - H)Y$$

- Because $Y = X\beta + e$ and $HX = X$, we get

$$\hat{e} = (I - H)(X\beta + e) = (X - HX)\beta + (I - H)e$$

Bootstrap in Regression

- What can we do if we do not know β and G? We simply estimate them from the data.
- β is estimated by the least squares estimate $\hat{\beta}$.
- How to estimate G? For this we use residuals $\hat{e}$ resulting in the method called *Residual Bootstrap*.
- Recall that

$$\hat{e} = (I - H)Y$$

- Because $Y = X\beta + e$ and $HX = X$, we get

$$\hat{e} = (I - H)(X\beta + e) = (X - HX)\beta + (I - H)e = (I - H)e.$$

Bootstrap in Regression

- What can we do if we do not know β and G? We simply estimate them from the data.
- β is estimated by the least squares estimate $\hat{\beta}$.
- How to estimate G? For this we use residuals $\hat{e}$ resulting in the method called *Residual Bootstrap*.
- Recall that

$$\hat{e} = (I - H)Y$$

- Because $Y = X\beta + e$ and $HX = X$, we get

$$\hat{e} = (I - H)(X\beta + e) = (X - HX)\beta + (I - H)e = (I - H)e.$$

- Therefore

Bootstrap in Regression

- What can we do if we do not know β and G? We simply estimate them from the data.
- β is estimated by the least squares estimate $\hat{\beta}$.
- How to estimate G? For this we use residuals $\hat{e}$ resulting in the method called *Residual Bootstrap*.
- Recall that

$$\hat{e} = (I - H)Y$$

- Because $Y = X\beta + e$ and $HX = X$, we get

$$\hat{e} = (I - H)(X\beta + e) = (X - HX)\beta + (I - H)e = (I - H)e.$$

- Therefore

$$\hat{e} = e - He.$$

- The closeness of the distribution of $\hat{e}$ to e thus depends on the term He.

- The closeness of the distribution of $\hat{e}$ to e thus depends on the term He. The ith entry of He equals

- The closeness of the distribution of $\hat{e}$ to e thus depends on the term He. The ith entry of He equals

$$(He)_i = \sum_{j=1}^{n} h_{ij} e_j$$

- The closeness of the distribution of $\hat{e}$ to e thus depends on the term He. The ith entry of He equals

$$(He)_i = \sum_{j=1}^{n} h_{ij} e_j = h_{ii} e_i + \sum_{j \neq i} h_{ij} e_j.$$

- The closeness of the distribution of $\hat{e}$ to e thus depends on the term He. The ith entry of He equals

$$(He)_i = \sum_{j=1}^{n} h_{ij} e_j = h_{ii} e_i + \sum_{j \neq i} h_{ij} e_j.$$

- Recall that h_{ii} is also known as the ith leverage.

- The closeness of the distribution of $\hat{e}$ to e thus depends on the term He. The ith entry of He equals

$$(He)_i = \sum_{j=1}^{n} h_{ij} e_j = h_{ii} e_i + \sum_{j \neq i} h_{ij} e_j .$$

- Recall that h_{ii} is also known as the ith leverage. Here are two facts about the hat matrix:

- The average of the leverages equals $(p+1)/n$

- The average of the leverages equals $(p+1)/n$ where p is the number of explanatory variables of X.

- The average of the leverages equals $(p+1)/n$ where p is the number of explanatory variables of X.
- The reason for this is the following:

- The average of the leverages equals $(p+1)/n$ where p is the number of explanatory variables of X.
- The reason for this is the following:
- The sum of the leverages equals the sum of the diagonals of the hat matrix.

- The average of the leverages equals $(p+1)/n$ where p is the number of explanatory variables of X.
- The reason for this is the following:
- The sum of the leverages equals the sum of the diagonals of the hat matrix. The sum of the diagonals of a square matrix is called its trace.

- The average of the leverages equals $(p+1)/n$ where p is the number of explanatory variables of X.
- The reason for this is the following:
- The sum of the leverages equals the sum of the diagonals of the hat matrix. The sum of the diagonals of a square matrix is called its trace. We therefore need to find the trace of the hat matrix, denoted by $tr(H)$.

- The average of the leverages equals $(p+1)/n$ where p is the number of explanatory variables of X.
- The reason for this is the following:
- The sum of the leverages equals the sum of the diagonals of the hat matrix. The sum of the diagonals of a square matrix is called its trace. We therefore need to find the trace of the hat matrix, denoted by $tr(H)$. Because $H = X(X^T X)^{-1} X^T$, we have

- The average of the leverages equals $(p+1)/n$ where p is the number of explanatory variables of X.
- The reason for this is the following:
- The sum of the leverages equals the sum of the diagonals of the hat matrix. The sum of the diagonals of a square matrix is called its trace. We therefore need to find the trace of the hat matrix, denoted by $tr(H)$. Because $H = X(X^TX)^{-1}X^T$, we have

$$tr(H) = tr\left(X(X^TX)^{-1}X^T\right)$$

- The average of the leverages equals $(p + 1)/n$ where p is the number of explanatory variables of X.
- The reason for this is the following:
- The sum of the leverages equals the sum of the diagonals of the hat matrix. The sum of the diagonals of a square matrix is called its trace. We therefore need to find the trace of the hat matrix, denoted by $tr(H)$. Because $H = X(X^T X)^{-1} X^T$, we have

$$tr(H) = tr\left(X(X^T X)^{-1} X^T\right)$$

- Here is a most important fact about trace:

$$tr(AB) = tr(BA).$$

- The average of the leverages equals $(p+1)/n$ where p is the number of explanatory variables of X.
- The reason for this is the following:
- The sum of the leverages equals the sum of the diagonals of the hat matrix. The sum of the diagonals of a square matrix is called its trace. We therefore need to find the trace of the hat matrix, denoted by $tr(H)$. Because $H = X(X^T X)^{-1} X^T$, we have

$$tr(H) = tr\left(X(X^T X)^{-1} X^T\right)$$

- Here is a most important fact about trace:

$$tr(AB) = tr(BA).$$

- This gives

- The average of the leverages equals $(p+1)/n$ where p is the number of explanatory variables of X.
- The reason for this is the following:
- The sum of the leverages equals the sum of the diagonals of the hat matrix. The sum of the diagonals of a square matrix is called its trace. We therefore need to find the trace of the hat matrix, denoted by $tr(H)$. Because $H = X(X^TX)^{-1}X^T$, we have

$$tr(H) = tr\left(X(X^TX)^{-1}X^T\right)$$

- Here is a most important fact about trace:

$$tr(AB) = tr(BA).$$

- This gives

$$tr(H) = tr\left(X(X^TX)^{-1}X^T\right) =$$

- The average of the leverages equals $(p+1)/n$ where p is the number of explanatory variables of X.
- The reason for this is the following:
- The sum of the leverages equals the sum of the diagonals of the hat matrix. The sum of the diagonals of a square matrix is called its trace. We therefore need to find the trace of the hat matrix, denoted by $tr(H)$. Because $H = X(X^TX)^{-1}X^T$, we have

$$tr(H) = tr\left(X(X^TX)^{-1}X^T\right)$$

- Here is a most important fact about trace:

$$tr(AB) = tr(BA).$$

- This gives

$$tr(H) = tr\left(X(X^TX)^{-1}X^T\right) = tr\left((X^TX)^{-1}X^TX\right)$$

- The average of the leverages equals $(p+1)/n$ where p is the number of explanatory variables of X.
- The reason for this is the following:
- The sum of the leverages equals the sum of the diagonals of the hat matrix. The sum of the diagonals of a square matrix is called its trace. We therefore need to find the trace of the hat matrix, denoted by $tr(H)$. Because $H = X(X^TX)^{-1}X^T$, we have

$$tr(H) = tr\left(X(X^TX)^{-1}X^T\right)$$

- Here is a most important fact about trace:

$$tr(AB) = tr(BA).$$

- This gives

$$tr(H) = tr\left(X(X^TX)^{-1}X^T\right) = tr\left((X^TX)^{-1}X^TX\right) = tr(I_{p+1})$$

- The average of the leverages equals $(p+1)/n$ where p is the number of explanatory variables of X.
- The reason for this is the following:
- The sum of the leverages equals the sum of the diagonals of the hat matrix. The sum of the diagonals of a square matrix is called its trace. We therefore need to find the trace of the hat matrix, denoted by $tr(H)$. Because $H = X(X^TX)^{-1}X^T$, we have

$$tr(H) = tr\left(X(X^TX)^{-1}X^T\right)$$

- Here is a most important fact about trace:

$$tr(AB) = tr(BA).$$

- This gives

$$tr(H) = tr\left(X(X^TX)^{-1}X^T\right) = tr\left((X^TX)^{-1}X^TX\right) = tr(I_{p+1}) = p+$$

- We thus have $tr(H) = p + 1$

- We thus have $tr(H) = p + 1$ which implies that the average of the leverages equals $(p + 1)/n$.

- We thus have $tr(H) = p + 1$ which implies that the average of the leverages equals $(p+1)/n$. When n is large, $(p+1)/n$ is quite small.

- We thus have $tr(H) = p + 1$ which implies that the average of the leverages equals $(p+1)/n$. When n is large, $(p+1)/n$ is quite small.
- Unless there are points with very high leverages,

- We thus have $tr(H) = p + 1$ which implies that the average of the leverages equals $(p+1)/n$. When n is large, $(p+1)/n$ is quite small.
- Unless there are points with very high leverages, all leverages will be small when n is large.

- In the ith row of H,

- In the ith row of H, the sum of the squares of the non-diagonal entries equals $h_{ii} - h_{ii}^2$.

- In the ith row of H, the sum of the squares of the non-diagonal entries equals $h_{ii} - h_{ii}^2$. This is a consequence of $H^2 = H$ as a result of which

- In the ith row of H, the sum of the squares of the non-diagonal entries equals $h_{ii} - h_{ii}^2$. This is a consequence of $H^2 = H$ as a result of which

- In the ith row of H, the sum of the squares of the non-diagonal entries equals $h_{ii} - h_{ii}^2$. This is a consequence of $H^2 = H$ as a result of which

$$h_{ii} = \sum_j h_{ij}^2$$

- In the ith row of H, the sum of the squares of the non-diagonal entries equals $h_{ii} - h_{ii}^2$. This is a consequence of $H^2 = H$ as a result of which

$$h_{ii} = \sum_j h_{ij}^2 = h_{ii}^2 + \sum_{j:j \neq i} h_{ij}^2$$

- In the ith row of H, the sum of the squares of the non-diagonal entries equals $h_{ii} - h_{ii}^2$. This is a consequence of $H^2 = H$ as a result of which

$$h_{ii} = \sum_j h_{ij}^2 = h_{ii}^2 + \sum_{j:j \neq i} h_{ij}^2$$

- Therefore

- In the ith row of H, the sum of the squares of the non-diagonal entries equals $h_{ii} - h_{ii}^2$. This is a consequence of $H^2 = H$ as a result of which

$$h_{ii} = \sum_j h_{ij}^2 = h_{ii}^2 + \sum_{j:j \neq i} h_{ij}^2$$

- Therefore

$$\sum_{j:j \neq i} h_{ij}^2 = h_{ii}\,(1 - h_{ii})\,.$$

- In the ith row of H, the sum of the squares of the non-diagonal entries equals $h_{ii} - h_{ii}^2$. This is a consequence of $H^2 = H$ as a result of which

$$h_{ii} = \sum_j h_{ij}^2 = h_{ii}^2 + \sum_{j:j \neq i} h_{ij}^2$$

- Therefore

$$\sum_{j:j \neq i} h_{ij}^2 = h_{ii} (1 - h_{ii}) .$$

- Therefore if the leverages are small,

- In the ith row of H, the sum of the squares of the non-diagonal entries equals $h_{ii} - h_{ii}^2$. This is a consequence of $H^2 = H$ as a result of which

$$h_{ii} = \sum_{j} h_{ij}^2 = h_{ii}^2 + \sum_{j:j \neq i} h_{ij}^2$$

- Therefore

$$\sum_{j:j \neq i} h_{ij}^2 = h_{ii} \left(1 - h_{ii}\right).$$

- Therefore if the leverages are small, then the non-diagonal entries of the hat matrix are also small.

- Because of the previous pair of facts,

- Because of the previous pair of facts, we can argue that *He* is negligible when *n* is large

- Because of the previous pair of facts, we can argue that *He* is negligible when *n* is large and when are there no points with high leverages.

- Because of the previous pair of facts, we can argue that He is negligible when n is large and when are there no points with high leverages. In this situation, $He \approx 0$

- Because of the previous pair of facts, we can argue that He is negligible when n is large and when are there no points with high leverages. In this situation, $He \approx 0$ which means that $\hat{e} \approx e$.

- Because of the previous pair of facts, we can argue that He is negligible when n is large and when are there no points with high leverages. In this situation, $He \approx 0$ which means that $\hat{e} \approx e$. Residual bootstrap can be proved to work in this setting.

Residual bootstrap algorithm to estimate the distribution of $\hat{\beta}_i - \beta_i$:

Residual bootstrap algorithm to estimate the distribution of $\hat{\beta}_i - \beta_i$:

1. Estimate β by the least squares estimate $\hat{\beta}$

Residual bootstrap algorithm to estimate the distribution of $\hat{\beta}_i - \beta_i$:

1. Estimate β by the least squares estimate $\hat{\beta}$ in the regression of Y on X.

Residual bootstrap algorithm to estimate the distribution of $\hat{\beta}_i - \beta_i$:

1. Estimate β by the least squares estimate $\hat{\beta}$ in the regression of Y on X. Let $\hat{e}$ denote the vector of residuals.

Residual bootstrap algorithm to estimate the distribution of $\hat{\beta}_i - \beta_i$:

1. Estimate β by the least squares estimate $\hat{\beta}$ in the regression of Y on X. Let $\hat{e}$ denote the vector of residuals.
2. Fix a large number (say $N = 5000$)

Residual bootstrap algorithm to estimate the distribution of $\hat{\beta}_i - \beta_i$:

1. Estimate β by the least squares estimate $\hat{\beta}$ in the regression of Y on X. Let $\hat{e}$ denote the vector of residuals.
2. Fix a large number (say $N = 5000$) and repeat the following steps for $j = 1, \ldots, N$:
   2.1 Resample n times with replacement from the residuals to obtain $\hat{e}_1^{(j)}, \ldots, \hat{e}_n^{(j)}$.

Residual bootstrap algorithm to estimate the distribution of $\hat{\beta}_i - \beta_i$:

1. Estimate β by the least squares estimate $\hat{\beta}$ in the regression of Y on X. Let $\hat{e}$ denote the vector of residuals.
2. Fix a large number (say $N = 5000$) and repeat the following steps for $j = 1, \ldots, N$:
   2.1 Resample n times with replacement from the residuals to obtain $\hat{e}_1^{(j)}, \ldots, \hat{e}_n^{(j)}$. Let $\hat{e}^{(j)}$ denote the $n \times 1$ vector consisting of these observations.

Residual bootstrap algorithm to estimate the distribution of $\hat{\beta}_i - \beta_i$:

1. Estimate β by the least squares estimate $\hat{\beta}$ in the regression of Y on X. Let $\hat{e}$ denote the vector of residuals.
2. Fix a large number (say $N = 5000$) and repeat the following steps for $j = 1, \dots, N$:
   2.1 Resample n times with replacement from the residuals to obtain $\hat{e}_1^{(j)}, \dots, \hat{e}_n^{(j)}$. Let $\hat{e}^{(j)}$ denote the $n \times 1$ vector consisting of these observations.
   2.2 Add the vector $X\hat{\beta}$ to $\hat{e}^{(j)}$

Residual bootstrap algorithm to estimate the distribution of $\hat{\beta}_i - \beta_i$:

1. Estimate β by the least squares estimate $\hat{\beta}$ in the regression of Y on X. Let $\hat{e}$ denote the vector of residuals.
2. Fix a large number (say $N = 5000$) and repeat the following steps for $j = 1, \ldots, N$:
   - 2.1 Resample n times with replacement from the residuals to obtain $\hat{e}_1^{(j)}, \ldots, \hat{e}_n^{(j)}$. Let $\hat{e}^{(j)}$ denote the $n \times 1$ vector consisting of these observations.
   - 2.2 Add the vector $X\hat{\beta}$ to $\hat{e}^{(j)}$ to create a new vector called $Y^{(j)}$.

Residual bootstrap algorithm to estimate the distribution of $\hat{\beta}_i - \beta_i$:

1. Estimate β by the least squares estimate $\hat{\beta}$ in the regression of Y on X. Let $\hat{e}$ denote the vector of residuals.
2. Fix a large number (say $N = 5000$) and repeat the following steps for $j = 1, \ldots, N$:
   2.1 Resample n times with replacement from the residuals to obtain $\hat{e}_1^{(j)}, \ldots, \hat{e}_n^{(j)}$. Let $\hat{e}^{(j)}$ denote the $n \times 1$ vector consisting of these observations.
   2.2 Add the vector $X\hat{\beta}$ to $\hat{e}^{(j)}$ to create a new vector called $Y^{(j)}$.
   2.3 Regress $Y^{(j)}$ on X.

Residual bootstrap algorithm to estimate the distribution of $\hat{\beta}_i - \beta_i$:

1. Estimate β by the least squares estimate $\hat{\beta}$ in the regression of Y on X. Let $\hat{e}$ denote the vector of residuals.
2. Fix a large number (say $N = 5000$) and repeat the following steps for $j = 1, \dots, N$:
   2.1 Resample n times with replacement from the residuals to obtain $\hat{e}_1^{(j)}, \dots, \hat{e}_n^{(j)}$. Let $\hat{e}^{(j)}$ denote the $n \times 1$ vector consisting of these observations.
   2.2 Add the vector $X\hat{\beta}$ to $\hat{e}^{(j)}$ to create a new vector called $Y^{(j)}$.
   2.3 Regress $Y^{(j)}$ on X. Call the resulting least squares coefficient vector $\hat{\beta}^{(j)}$.

Residual bootstrap algorithm to estimate the distribution of $\hat{\beta}_i - \beta_i$:

1. Estimate β by the least squares estimate $\hat{\beta}$ in the regression of Y on X. Let $\hat{e}$ denote the vector of residuals.
2. Fix a large number (say $N = 5000$) and repeat the following steps for $j = 1, \dots, N$:
   2.1 Resample n times with replacement from the residuals to obtain $\hat{e}_1^{(j)}, \dots, \hat{e}_n^{(j)}$. Let $\hat{e}^{(j)}$ denote the $n \times 1$ vector consisting of these observations.
   2.2 Add the vector $X\hat{\beta}$ to $\hat{e}^{(j)}$ to create a new vector called $Y^{(j)}$.
   2.3 Regress $Y^{(j)}$ on X. Call the resulting least squares coefficient vector $\hat{\beta}^{(j)}$.
3. The empirical distribution of the values $\hat{\beta}_i^{(j)} - \hat{\beta}_i$ for $j = 1, \dots, N$ provides an estimate of the distribution of $\hat{\beta}_i - \beta_i$.

- The bootstrap estimate of the standard error of $\hat{\beta}_i$ is given by the standard deviation of the values $\hat{\beta}_i^{(j)} - \hat{\beta}_i$ for $j = 1, \ldots, N$.

- The bootstrap estimate of the standard error of $\hat{\beta}_i$ is given by the standard deviation of the values $\hat{\beta}_i^{(j)} - \hat{\beta}_i$ for $j = 1, \ldots, N$.
- A bootstrap confidence interval for β_i is given by $\hat{\beta}_i$ plus/minus 1.96 times the bootstrap estimate of the standard error.

- The bootstrap estimate of the standard error of $\hat{\beta}_i$ is given by the standard deviation of the values $\hat{\beta}_i^{(j)} - \hat{\beta}_i$ for $j = 1, \ldots, N$.
- A bootstrap confidence interval for β_i is given by $\hat{\beta}_i$ plus/minus 1.96 times the bootstrap estimate of the standard error.
- An alternate bootstrap confidence interval for β_i is given by:

- The bootstrap estimate of the standard error of $\hat{\beta}_i$ is given by the standard deviation of the values $\hat{\beta}_i^{(j)} - \hat{\beta}_i$ for $j = 1, \ldots, N$.
- A bootstrap confidence interval for β_i is given by $\hat{\beta}_i$ plus/minus 1.96 times the bootstrap estimate of the standard error.
- An alternate bootstrap confidence interval for β_i is given by:

$$[\hat{\beta}_i - b_{0.975}, \hat{\beta}_i - b_{0.025}]$$

- The bootstrap estimate of the standard error of $\hat{\beta}_i$ is given by the standard deviation of the values $\hat{\beta}_i^{(j)} - \hat{\beta}_i$ for $j = 1, \ldots, N$.
- A bootstrap confidence interval for β_i is given by $\hat{\beta}_i$ plus/minus 1.96 times the bootstrap estimate of the standard error.
- An alternate bootstrap confidence interval for β_i is given by:

$$[\hat{\beta}_i - b_{0.975}, \hat{\beta}_i - b_{0.025}]$$

- The bootstrap estimate of the standard error of $\hat{\beta}_i$ is given by the standard deviation of the values $\hat{\beta}_i^{(j)} - \hat{\beta}_i$ for $j = 1, \ldots, N$.
- A bootstrap confidence interval for β_i is given by $\hat{\beta}_i$ plus/minus 1.96 times the bootstrap estimate of the standard error.
- An alternate bootstrap confidence interval for β_i is given by:

$$[\hat{\beta}_i - b_{0.975}, \hat{\beta}_i - b_{0.025}]$$

where $b_{0.025}$ and $b_{0.975}$ denote the 0.025^{th} and the 0.975^{th} quantiles of values $\hat{\beta}_i^{(j)} - \hat{\beta}_i$ for $j = 1, \ldots, N$.