Resampling methods

Module DES130: Computational Statistics

Dr. Erick A. Chacón Montalván (echacon@uni.edu.pe)

Escuela de Profesional de Ingeniería Estadística
Facultad de Ingeniería Económica, Estadística y Ciencias Sociales
Universidad Nacional de Ingeniería (UNI)
Lima – Perú

Introduction

Resampling

Concept

Resampling is the procedure of generating new samples using observed data or a data generating mechanism.

Resampling is the same as:

- Given a set of observations $x_1, \ldots, x_n$; use the empirical cumulative density function $\hat{F}$ as an estimate of the cumulative density function F
- Generate random values $x_1^*, \ldots, x_n^*$ coming from $\hat{F}$.

Common types

- Jacknife: Takes one-leave-out samples.
- Bootstrap: Take random samples with repetition.
- Permutations: Relabel the data.

Resampling inference

Resampling is used with the goal of:

- Approximate the distribution of the statistics of interest.
- Estimate the standard error of an statistics.
- Compute confidence intervals.
- Perform hypothesis testing.

Resampling and monte carlo

Let consider a quantity of interest μ which can be writen as the mean of $h(X)$ where X is a random variable with density function $f(x)$,

$$\mu = \int h(x)f(x)dx \approx \int h(x)\hat{f}(x)dx.$$

Considering the random sample $X_1^*, X_2^*, \ldots, X_m^*$ of size m with density $\hat{f}(x)$, we define the resampling estimator

$$\hat{\mu}_{RE} = \frac{\sum_{i=1}^{m} h(X_i^*)}{m}.$$

The estimator improves as long as $m \to \infty$.

Motivation

- Classical statistical inference includes **asymptotic and distributional assumptions**.
- Asymptotic assumptions do not hold for **small datasets**.
- Uncertainty quantification is not obvious in **non-linear and hierarchical models**.
- Some statistics tests have **unknown analytical probability or density function.**
- Monte carlo methods **assumes a known cumulative density function**.

Jacknife

Estandar error

Let $\hat{\theta}$ be an estimator of θ. It is customary to work with the pivotal statistics

$$\frac{\hat{\theta} - \theta}{se(\hat{\theta})} \sim G(\cdot).$$

Using this pivotal statistics, we can obtain confidence interval or perform hypothesis testing, but we need a good computation of $se(\hat{\theta})$. Jacknife provides an alternative way to approximate $se(\hat{\theta})$.

Jacknife estimate

Let consider a random sample $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ with realization $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ such as $X_i \sim F$. Let $\hat{\theta}(\boldsymbol{X})$ be a statistic of and

$$\boldsymbol{x}_{(i)} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n),$$

such as

$$\hat{\theta}_{(i)} = \hat{\theta}(\boldsymbol{x}_{(i)}) \text{ for } i = 1, \ldots, n.$$

Then the Jacknife estimation of the standard error is

$$\hat{se}_{jack}(\hat{\theta}) = \left[\frac{n-1}{n} \sum_{i=1}^{n} (\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2 \right]^{1/2}, \text{ with } \hat{\theta}_{(.)} = \sum_{i=1}^{n} \hat{\theta}_{(i)}/n.$$

Jacknife estimate

For $\hat{\theta} = \bar{X}$, it can be shown that

$$\hat{se}_{jack}(\hat{\theta}) = \left[\sum_{i=1}^{n}(x_i - \bar{x})^2/(n(n-1))\right]^{1/2},$$

because $\hat{\theta}_{(i)} = (n\bar{x} - x_i)/(n-1)$.

Example for correlation

$$\hat{\theta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_{i=1}^{n}(x_i - \bar{x})\sum_{i=1}^{n}(y_i - \bar{y})\right]^{1/2}}$$

Jacknife features

- The advantage of the jacknife estimate is that $\hat{\theta}$ could be any estimator or statistic of interest, and not only the sample average.
- It is nonparametric because nothing is assumed about F.
- Automatic.

Bootstrap

Example: sample average

TABLE 9.1 Possible bootstrap pseudo-datasets from {1, 2, 6} (ignoring order), the resulting values of $\widehat{\theta}^* = T(\widehat{F}^*)$, the probability of each outcome in the bootstrapping experiment ($P^*\left[\widehat{\theta}^*\right]$), and the observed relative frequency in 1000 bootstrap iterations.

$\mathcal{X}^*$	$\widehat{\theta}^*$	$P^*\left[\widehat{\theta}^*\right]$	Observed Frequency
1 1 1	3/3	1/27	36/1000
1 1 2	4/3	3/27	101/1000
1 2 2	5/3	3/27	123/1000
2 2 2	6/3	1/27	25/1000
1 1 6	8/3	3/27	104/1000
1 2 6	9/3	6/27	227/1000
2 2 6	10/3	3/27	131/1000
1 6 6	13/3	3/27	111/1000
2 6 6	14/3	3/27	102/1000
6 6 6	18/3	1/27	40/1000

A 93% confidence interval for θ is $[4/3, 14/3]$.

Notation I

Theoretical concepts:

- F: Cumulative density function.
- X_i: A **random variable** with some associated density f and cumulative F functions.
- $\boldsymbol{X} = (X_1, \ldots, X_n)^\intercal$: A **random vector** representing a random sample such as $X_i \sim f$ with the cumulative density function F.
- $\theta = g(F)$: A parameter of interest of F. For example $\theta = \int x dF(x)$.
- $T(\boldsymbol{X}, F)$: An statistics of interest.

Notation II

Empirical concepts:

- $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\intercal$: Observed data, which is a realization of $\boldsymbol{X}$.
- $\hat{F}$: The empirical cumulative distribution function obtained using $\boldsymbol{x}$.
- $\boldsymbol{X}^* = (X_1^*, \ldots, X_n^*)^\intercal$: A **random vector** representing a bootstrap sample such as X_i^* have the cumulative density function $\hat{F}$.
- $\hat{\theta} = g(\hat{F})$: The estimator of θ using $\hat{F}$. For example, $\hat{\theta} = \sum X_i / n$.
- $T(\boldsymbol{X}^*, \hat{F})$: An statistics of interest.

Bootstrap

Bootstrap

The goal of bootstrap is to aproximate the distribution of $T(\boldsymbol{X}, F)$ using the distribution of $T(\boldsymbol{X}^*, \hat{F})$.

Once we obtain the distribution of $T(\boldsymbol{X}^*, \hat{F})$, then we can estimate the standard error, provide confidence intervals or perform hypothesis testing.

Non parametric bootstrap

In most applications, we will not be able to work with all the possible resamples. Lets define a number of resamples B such as $\boldsymbol{X}_i^* \sim \hat{F}$ for $i = 1, \ldots, B$. Then the distribution of $T(\boldsymbol{X}, F)$ is approximated by the empirical distribution of $T(\boldsymbol{X}_i^*, \hat{F})$ for $i = 1, \ldots, B$.

Parametric bootstrap

In case that data modelled comes from a parametric distribution $F(\boldsymbol{x}, \theta)$, then we can use $F(\boldsymbol{x}, \hat{\theta})$ to generate $\boldsymbol{X}_i^*$ for $i = 1, \ldots, B$.

Bootstrap bias

A quantity of interest can be

$$T(\boldsymbol{X}, F) = g(\hat{F}) - g(F),$$

which has mean $\mathbb{E}\left[\hat{\theta}\right] - \theta$ representing the bias of $g(\hat{F})$. The bootstrap bias estimate is

$$\sum_{i=1}^{B} (\hat{\theta}(\boldsymbol{x}_i^*) - \hat{\theta}(\boldsymbol{x}))/B = \bar{\theta}^* - \hat{\theta}(\boldsymbol{x}).$$

Bootstrap regression

$$z_i = (y_i, \boldsymbol{x}_i). \tag{1}$$

Percentile Method

$1 - \alpha$ confidence interval

$$\Pr\left(Q_{\alpha/2}(\hat{\theta}(\boldsymbol{X}^*)) \leq \theta \leq Q_{1-\alpha/2}(\hat{\theta}(\boldsymbol{X}^*))\right) = 1 - \alpha.$$

Justification of the percentile method

Let consider a strictly increasing transformation ϕ and a continuous and symmetric distribution function H such as:

$$\Pr\left(h_{\alpha/2} \leq \phi(\hat{\theta}) - \phi(\theta) \leq h_{1-\alpha/2}\right) = 1 - \alpha,$$
$$\Pr\left(\phi^{-1}(h_{\alpha/2} + \phi(\hat{\theta})) \leq \theta \leq \phi^{-1}(h_{1-\alpha/2} + \phi(\hat{\theta}))\right) = 1 - \alpha.$$

Similarly

$$\Pr\left(h_{\alpha/2} \leq \phi(\hat{\theta}) - \phi(\hat{\theta}(\boldsymbol{X}^*)) \leq h_{1-\alpha/2}\right) = 1 - \alpha,$$
$$\Pr\left(\phi^{-1}(h_{\alpha/2} + \phi(\hat{\theta})) \leq \hat{\theta}(\boldsymbol{X}^*) \leq \phi^{-1}(h_{1-\alpha/2} + \phi(\hat{\theta}))\right) = 1 - \alpha.$$

As an approximation we can obtain that $\phi^{-1}(h_{\alpha/2} + \phi(\hat{\theta}))$ is the $\alpha/2$ quantile of the bootstrap estimator. The approximation is good if $\phi(\hat{\theta})$ is unbiased and its vairance does not depend on θ.

Bootstrap t (studentized)

Let consider $T(\boldsymbol{X}, F) = \frac{g(\hat{F}) - g(F)}{\sqrt{\mathbb{V}\left[g(\hat{F})\right]}}$ with cumulative distribution function G, it will be roughly pivotal.

$$\Pr\left(Q_{\alpha/2,G} \leq T(\boldsymbol{X}, F) \leq Q_{1-\alpha/2,G}\right) = 1 - \alpha,$$

$$\Pr\left(\hat{\theta} - \mathbb{V}\left[g(\hat{F})\right] Q_{1-\alpha/2,G} \leq \theta \leq \hat{\theta} - \mathbb{V}\left[g(\hat{F})\right] Q_{\alpha/2,G}\right) = 1 - \alpha,$$

Using bootstrap

$$\Pr\left(\hat{\theta}^* - \mathbb{V}\left[g(\hat{F})\right] Q_{1-\alpha/2,\hat{G}} \leq \theta \leq \hat{\theta}^* - \mathbb{V}\left[g(\hat{F})\right] Q_{\alpha/2,\hat{G}}\right) = 1 - \alpha.$$

Permutation tests

Permutation tests

Hypothesis testing