SB1.2 Computational Statistics
Hilary Term 2019

Week 8 Assessed Practical

- There are two questions in this practical but only question 2 is assessed
- It contributes 8.5% to your raw SB1 total mark.
- Write your answer to question 2 as a report.
- The deadline for submission, which is officially published in the Course Handbook (Part B synopses), is:

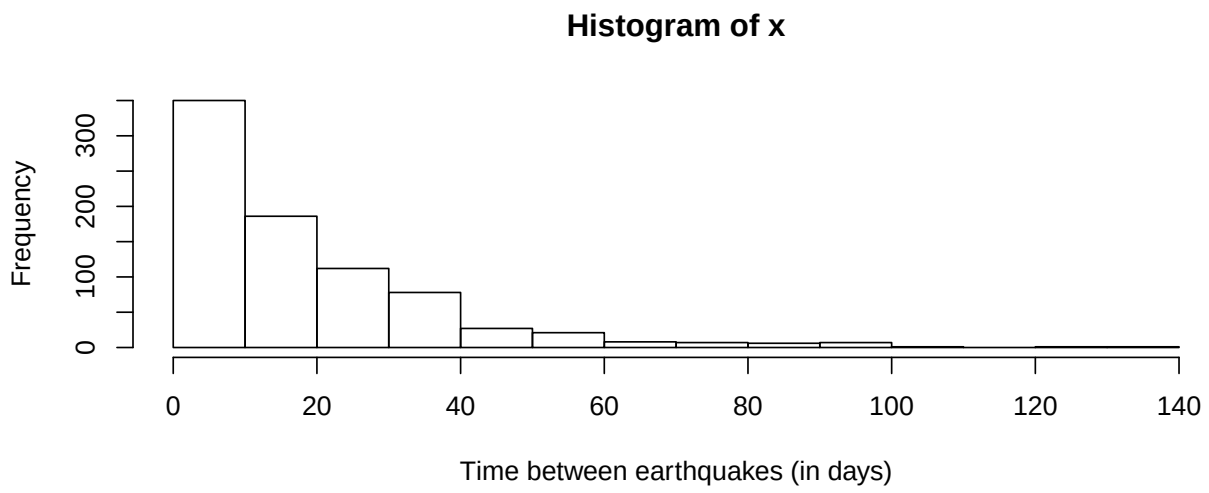
  12 noon Monday week 2, Trinity Term 2019,

  at the Statistics Department reception, 24-29 St Giles. Your report should be clearly written. There are no marks awarded for presentation but there are marks awarded for clarity. You should use captions for your tables and figures and include your commented R-code, preferably in an appendix.

1 Question 1 - non-assessed

We consider a dataset of times (in days) between $n = 174$ successive earthquakes of magnitude 6 of higher from 1970 to 2009. The dataset comes from the `R` package `resampledata`.

```
# Install and load the R package
# install.packages('resampledata') if needed
library('resampledata')
# Load Earthquakes data
x = Quakes[,2]
# Plot Histogram of the data
hist(x, xlab = "Time between earthquakes (in days)")
```



Assume that the times $X_1, \ldots, X_n$ are independent and identically distributed (iid) from some unknown cumulative distribution function F. Let F^{-1} be the associated quantile function. For a given quantile $q(\alpha) = F^{-1}(\alpha)$ with $\alpha \in (0, 1)$, we consider the nonparametric plug-in estimator

$$\hat{q}_n^{\mathtt{NP}}(\alpha) = F_n^{-1}(\alpha)$$

where F_n is the empirical cdf and F_n^{-1} the associated quantile function.

- Calculate the estimates $\hat{q}_n^{\mathtt{NP}}(\alpha)$ for $\alpha = 0.1, 0.25, 0.5, 0.75, 0.9$.
- Using the nonparametric bootstrap, provide 99% approximate confidence intervals for $q(\alpha)$ for $\alpha = 0.1, 0.25, 0.5,\ 0.75, 0.9,$.

2 Question 2 - assessed

2.1 Part A

Assume now that $F = G_\theta$, where G_θ is the cdf of an exponential random variable with unknown rate $\theta > 0$. Its probability density function g_θ is defined, for $x > 0$, as

$$g_\theta(x) = \theta e^{-\theta x}.$$

Denote G_θ^{-1} the associated quantile function. We consider the parametric estimator

$$\hat{q}_n^{\mathtt{P}}(\alpha) = G_{\hat{\theta}_n}^{-1}(\alpha) \tag{1}$$

where $\hat{\theta}_n$ is the maximum likelihood estimator of θ.

- Give a bootstrap estimate of $\mathbb{V}(\widehat{\theta}_n)$
- Calculate the estimates $\hat{q}_n^{\mathsf{P}}(\alpha)$ for $\alpha = 0.1, 0.25, 0.5, 0.75, 0.9$.
- Using the parametric bootstrap, provide 99% approximate confidence intervals for $q(\alpha)$ for $\alpha = 0.1, 0.25, 0.5,\ 0.75, 0.9,$.
- Discuss which of the parametric or nonparametric approach is more suitable here.

We now want to make prediction and find the prediction confidence region of a future observation X_{n+1} given $X_1, \ldots, X_n$.

For any $\alpha \in (0, 1)$, define

$$h(\alpha) = \mathbb{P}_F(X_{n+1} \leq \hat{q}_n^{\mathsf{P}}(\alpha)) \tag{2}$$

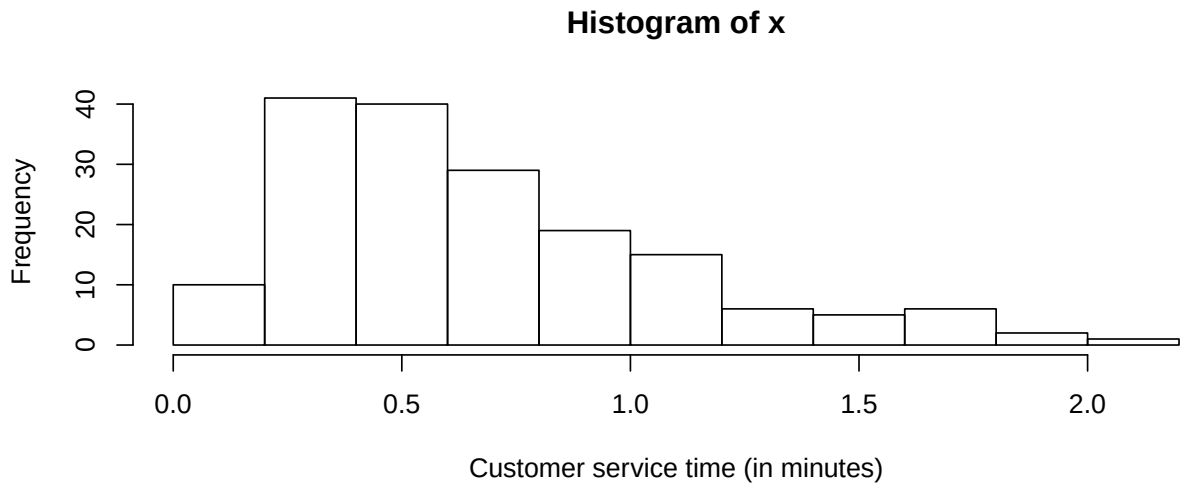
Note that we expect $h(\alpha)$ to be close to α, but its exact value cannot be computed analytically.

- Define the parametric bootstrap estimator of $h(\alpha)$, for any $\alpha \in (0, 1)$
- Calculate the bootstrap estimate of $h(\alpha)$ for a range of values α and obtain an approximate 99% prediction confidence interval for X_{n+1}.

2.2 Part B

Consider now the dataset of the service time (in minutes) for 174 customers at a college snack bar.

```
# Load Service time data
x = Service[,2]
# Plot Histogram of the data
hist(x, xlab = "Customer service time (in minutes)")
```



Assume that the service times $X_1, \ldots, X_n$ are iid from some distribution F, with quantiles $q(\alpha) = F^{-1}(\alpha)$. Propose a (parametric or nonparametric) estimator for $q(\alpha)$ and compute the estimate for this dataset for $\alpha = 0.1, 0.25, 0.5, 0.75, 0.9$. Using the (parametric or nonparametric) bootstrap, calculate 99% approximate confidence intervals for $q(\alpha)$ for $\alpha = 0.1, 0.25, 0.5, 0.75, 0.9$ and 99% approximate prediction confidence interval for a new observation X_{n+1}. Carefully justify any choice you make.