Computer Experiments: Statistical Inference

1- Generate N samples from an Exponential distribution.

A. Estimate its parameter. Is it unbiased?
B. What is its 93% confidence interval?
C. Repeat the process for 10000 times and see how often the confidence intervals contains the parameter?

2- Generate 100 samples from the distribution above.

A. Calculate & plot empirical distribution function.
B. Find the plug-in estimator for:
   a. Mean
   b. Variance(both reasonable estimators)
   c. Skewness

(Help Examples 8.6 to 8.8 of textbook)

3- Let $X_1, \ldots, X_n$ Normal(μ,1). Let $\theta = e^{\mu}$ and let $\hat{\theta} = e^{\bar{X}}$ be the MLE. Create a dataset(using $\mu = 5$) consisting of n=100 observations.

A. Use the bootstrap to get the se and 95 percent confidence interval for θ.
B. Plot a histogram of the bootstrap replications for the nonparametric bootstrap. These are estimates of the distribution of $\hat{\theta}$. Compare this to the true sampling distribution of $\hat{\theta}$.

4- In 1861, 10 essays appeared in the New Orleans Daily Crescent. They were signed "Quintus Curtuis Snodgrass" and some people suspected they were actually written by Mark Twain. To investigate this, we will consider the proportion of three letter words found in an author`s work. From eight Twain essays we have:
.225 .262 .217 .240 .230 .229 .235 .217
From 10 Snodgrass essays we have:
.209 .205 .196 .210 .202 .207 .224 .223 .220 .201

Perform a Wald test for equality of the means. Use the nonparametric plug-in estimator. Report the p-value and a 97% confidence interval for the difference of means. What do you conclude?

5- Let λ_0=1, n=20 and α=.05. Simulate $X_1, \ldots, X_n \sim Poisson(\lambda)$ and perform the Wald test. Repeat many times and count how often you reject the null. How close is the type I error rate to .05?

6- Let $X_1, \dots, X_n$ ~ Normal(μ,1).

A. Simulate a dataset (using $\mu = 5$) consisting of n=100 observations.
B. Take f(μ)=1 and find the posterior density. Plot the density.
C. Simulate 1000 draws from posterior. Plot a histogram of the simulated values and compare the histogram to the answer in (B).
D. Let $\theta = e^{\mu}$. Find the posterior density for θ analytically and by simulation.
E. Find a 95 percent posterior interval for θ.
F. Find a 95 percent confidence interval for θ.

7- Suppose observations $X_1, \dots, X_n$ are recorded. We assume these to be conditionally independent and exponentially distributed given a parameter θ: $X_i \sim Exponential(\theta) f$or all i = 1,...,n. The exponential distribution is controlled by one rate parameter θ > 0, and its density is $p(x;\theta) = \theta e^{-\theta x}$ for $x = R_+$.

A. Plot the graph of p(x;θ) for θ = 1 in the interval x ∈ [0,4].
B. What is the visual representation of the likelihood of individual data points? Draw it into the graph above for the samples in a toy dataset X = {1,2,4} and θ = 1. How is the likehood of this toy dataset related to that of the individual data points?
C. Would a higher rate (e.g. θ = 2) increase or decrease the likelihood for the toy data set?

8- Generate:

a. 100 samples from Gaussian($\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0.5\\0.5 & 1\end{bmatrix}$),
b. 100 samples from Gaussian($\begin{bmatrix}2\\2\end{bmatrix}, \begin{bmatrix}1 & 0.5\\0.5 & 1\end{bmatrix}$) and
c. 100 samples from Gaussian($\begin{bmatrix}3\\1\end{bmatrix}, \begin{bmatrix}1 & 0.3\\0.3 & 1\end{bmatrix}$).

Implement a Bayesian classifier where the class-conditional probabilities are estimated using histograms. Use 50 sample from each class for train (totally 150 train sample) and what remains for test (totally 150 train sample) this classifier and report the confusion matrices for both training and testing subsets.

فرمت گزارش:

گزارش بایستی به زبان فارسی و در قالب PDF باشد. در گزارش تحلیل و نتیجه‌گیری خود را در رابطه با هر تمرین به شکل مختصر بیان فرمایید.

فایل گزارش خود را به شکل «شماره دانشجویی_Report2» نام‌گذاری نمایید(مانندReport2_91234567).

فرمت کدها:

برای هر تمرین بایستی فایل کد جداگانه در محیط MATLAB تهیه شود.

هر فایل کد خود را به شکل «شماره دانشجویی_CE2_k» نام‌گذاری فرمایید که k بیانگر شماره تمرین می‌باشد. (مانند CE2_3_91234567 برای تمرین سوم).

نحوه تحویل:

فایل‌های کد و گزارش خود را که طبق فرمت‌های فوق تهیه شده‌اند، در قالب یک فایل فشرده در سایت درس بارگذاری نمایید.

فایل فشرده را به شکل «شماره دانشجویی_CE2» نام‌گذاری فرمایید(مانندCE2_91234567).

آخرین موعد تحویل: ۲۳ آذر ماه