Data Science and Statistical Computing

Tutorial 3, Week 6

Q1

A telephone switchboard receives the following number of calls in a 4-minute period:

$$\mathbf{x} = (0, 3, 1, 2)$$

We assume that the number of calls are independent and identically distributed in any 1-minute period and want to estimate the probability that there are no calls for the next 2 minutes. We propose estimating this non-parametrically by the following statistic:

$$S(\mathbf{x}) = \left(\frac{\sum_{i=1}^{4} \mathbb{1}\{x_i = 0\}}{4} \right)^2$$

(a) What is the estimate for the probability of no calls in the next 2 minutes for this data?

You decide to use bootstrap methodology to study this estimator, and notice that you don't even need to do any simulation in this case! Can you see why? Discuss in your group before proceeding.

Hint for discussion if stuck How many possible values can $S(\mathbf{x})$ take here?

Could you exactly compute the probability of observing those values when you resample $\mathbf{x}$?

Without doing any simulation:

(b) Using the bootstrap percentile confidence interval method for your estimator, $S(\mathbf{x})$, find the largest value $\eta \in [0, 1]$ for which you are at most 95% confident that $\mathbb{P}(\text{no calls in next 2 mins}) \leq \eta$.

(c) Determine $\mathbb{E}[\bar{S}^\star]$ and hence determine the bootstrap estimate of bias (if any) in the estimator.

You read that the Poisson distribution, with probability mass function,

$$p(k \mid \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

is often used to model the number of arrivals in a queue.

(d) List the detailed steps to perform a parametric bootstrap estimate of the uncertainty in the probability that there are no calls for the next 2 minutes under the assumption of a Poisson number of calls each minute.

(e) Without performing any simulation, again determine $\mathbb{E}[\bar{S}^\star]$ and hence determine the parametric bootstrap estimate of bias (if any) in the estimator.

Q2

In the lecture we saw a method of estimating π by looking at the proportion of randomly sampled uniform values on a square falling inside the circle contained within. We will use similar ideas to estimate $\sqrt{2}$ in this question.

We certainly know $0 \leq \sqrt{2} \leq 2$, so:

(a) Write down the probability density function for the random variable X having uniform distribution between 0 and 2, and calculate the cumulative distribution function.

(b) Write down $\mathbb{P}(X \leq \sqrt{2})$ and hence propose an algorithm to estimate $\sqrt{2}$ *without* having to know how to take square roots.

(c) What is the distribution of $\mathbb{1}\{X \leq \sqrt{2}\}$? Given that we know this probability (we can in fact compute $\sqrt{2}$), write down a 95% confidence interval for the value of your Monte Carlo estimate of $\sqrt{2}$ based on your algorithm in (b), when the number of Monte Carlo samples n is large.

From a pilot run of simulations, you notice that $1.5^2 > 2$.

(d) What is the effect on your estimator if you now construct a new estimator based on simulations from a uniform distribution between 0 and 1.5?

(e) If I take 1000 samples using the Uniform [0,2] approach, how many samples must I take under the Uniform [0,1.5] approach to be equally accurate in terms of the confidence interval for the estimators?