Project 2
Computational Numerical Statistics
DM, FCT-UNL
Group: G1

FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE NOVA DE LISBOA

Deadline: 25/11/19

Bootstrap & Jackknife

In the resolution of the following problems **do not** use any of the `R` bootstrap and jackknife built-in functions.

Fix your `R` random seed to 123 in all simulations.

1. A specific electrical battery installation needs that all its batteries are regularly replaced after 1200 hours. One observed the following 20 survival times of independent batteries

1354	1354	2420	1577	1552	1299	71	3725	1766	627
2195	884	1325	695	1825	1014	2183	2586	159	965

Let

$$T = \text{ number of survival hours of a battery}$$

and $p = P(T > 1200)$. Let $\mathcal{P} = \dfrac{X}{n}$ with $X \sim Bin(n, p)$ and

$$X = \text{number of batteries that lived more than 1200 hours in } n \text{ inspected batteries}$$
$$\text{being } p = P(T > 1200) \text{ the probability of success.}$$

(a) Show that $\mathcal{P}$ is an unbiased and consistent estimator of p. Estimate p and $\mathrm{V}(\mathcal{P})$.

(b) Use the **non-parametric** bootstrap ($B = 10000$ samples) approach to estimate the variance and bias of $\mathcal{P}$. Display the histrograms of the bootstrap estimated p's. Also, report the bootstrap and bias corrected estimates of p.

(c) Using the simulated bootstrap samples from (b), provide both the non-parametric bootstrap **percentile** and **studentized** 90% CIs for p.

(d) Use the **non-parametric** bootstrap ($B = 10000$ samples) to test the hypothesis

$$H_0: \ \mu \leq 1000 \qquad \text{vs} \qquad H_1: \mu > 1000,$$

where μ is the population mean, at the 5% significance level.

2. Let $X \sim Geom(p)$, which has probability mass function (p.m.f.) given by

$$f(x;p) = p(1-p)^x \qquad x = 0, 1, \ldots \qquad p \in [0,1].$$

Let

```
0 2 3 0 0 0 0 2 3 1 0 1 0 1 0 2 5 2 5 1 2 2 1 0 0
```

be an observed sample from X.

(a) Derive the maximum likelihood estimator (MLE) of p and use it to estimate p.

(b) Use both the bootstrap and jackknife techniques to estimate the standard error and bias of the MLE of p. Compare the results.

Optimization

3. Let $X \sim Beta(\alpha, 1)$, which has p.d.f.

$$f(x;\alpha) = \frac{x^{\alpha-1}}{B(\alpha,1)}, \qquad \alpha > 0, \qquad x \in [0,1]$$

Let

```
0.5409477 0.8184872 0.7848854 0.9850439 0.8963032 0.6089008 0.9549606 0.6795304
0.8451902 0.5613979 0.4029634 0.2741569 0.3996693 0.6371445 0.7521881
```

be an observed sample from X.

(a) Derive the likelihood, log-likelihood and score functions.

(b) Display the likelihood, log-likelihood and score functions graphically in order to locate the ML estimate of α.

(c) Use the `R` function `maxLik()` from library `maxLik` to approximate the ML estimate of α.

(d) Derive the algorithms of bisection, Newton-Raphson and Fisher scoring that enable the approximation of the ML estimate of α. Implement those in `R` and use the sample above to estimate α. Justify your choice of the initial estimates.

Use the **absolute** convergence criterion as a stop rule with $\varepsilon = 0.000001$.

(e) Which of the methods in (d) is more sensitive to the initial estimates?

4. Let $X \sim Exp(\lambda)$, which has p.d.f.

$$f(x;\lambda) = \lambda e^{-\lambda x}, \qquad \lambda > 0, \qquad x \in \mathbb{R}.$$

Let

```
-1.16225338 -0.01236762  0.23144468 -2.08263805  2.64870304 -0.52868938 -1.52280636
 0.03085357 -1.54249244 -0.23164183  2.23750583 -0.64678326 -0.82817693 -1.50508448
 0.87125402 -1.67430203  1.63702338 -0.85729792 -0.67855079 -1.36315013 -1.70412209
-0.05967576 -0.88579241 -0.33469221 -0.74940615
```

be an observed sample from X.

(a) Use the observed data and the method of Newton-Raphson to approximate the ML estimate of λ. Justify your choice of the initial estimate and all the intermediate steps.

(b) Consider the following reparametrization $b = \log(\lambda)$ in the pdf above. Implement the Newton-Rapshon algorithm to approximate the ML estimate of λ using this reparametrization.

(c) Which approach, (a) or (b) is more sensitive to the initial values?

Use the **relative** convergence criterion as a stop rule with $\varepsilon = 0.000001$.