Department of Mathematics and Statistics

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

CSE Exercises - Week 10

1. Revisiting exercise 2 in week 9, do the following using the data in normal_data.mat:

   (a) Plot the empirical DF together with the 95% confidence band given by the DKW inequality.

   (b) Use the nonparametric bootstrap to estimate the standard error of $\hat{q}$. You should use the same number of bootstrap realizations (N) as for exercise 2(b)(ii) in week 9.

   (c) Compare and comment on your estimates of the standard error using the two bootstrap methods, i.e the nonparametric bootstrap in part (b) above and the parametric bootstrap result from week 9.

2. Revisiting exercise 3 in week 9, do the following using the data in beta_data.mat:

   (a) Plot the empirical DF together with the 95% confidence band given by the DKW inequality.

(b) Use the nonparametric bootstrap to find approximate 95% confidence intervals for α, β and m, based on ML estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{m}$.

(c) Compare and comment on the results from part (b) above with those obtained using the parametric bootstrap from exercise 3(c) in week 9.

(d) Now repeat for the larger data set with $n = 100$ given in beta_data2.mat, i.e. find approximate 95% confidence intervals for α, β and m based on ML estimates using both parametric and nonparametric bootstraps. Comment on your results vis-a-vis those for $n = 25$.

3. Here is an example where the nonparametric bootstrap performs poorly.

Let $X_1, \ldots, X_n \overset{IID}{\sim} \text{Uniform}(0, \theta)$. Recall from exercise 1 in week 9 that the ML estimator for θ is $\hat{\theta} = X_{(n)} = \max\{X_1, \ldots, X_n\}$. Suppose that $\theta = 1$ and $n = 50$.

(a) Use Monte Carlo simulation to get an idea of the true distribution of $\hat{\theta}$. To do this, generate $N = 1000$ realizations, each containing 50 sample values from Uniform $(0, 1)$. For each realization, find $\hat{\theta}$. Plot a density histogram for the 1000 $\hat{\theta}$ estimates.

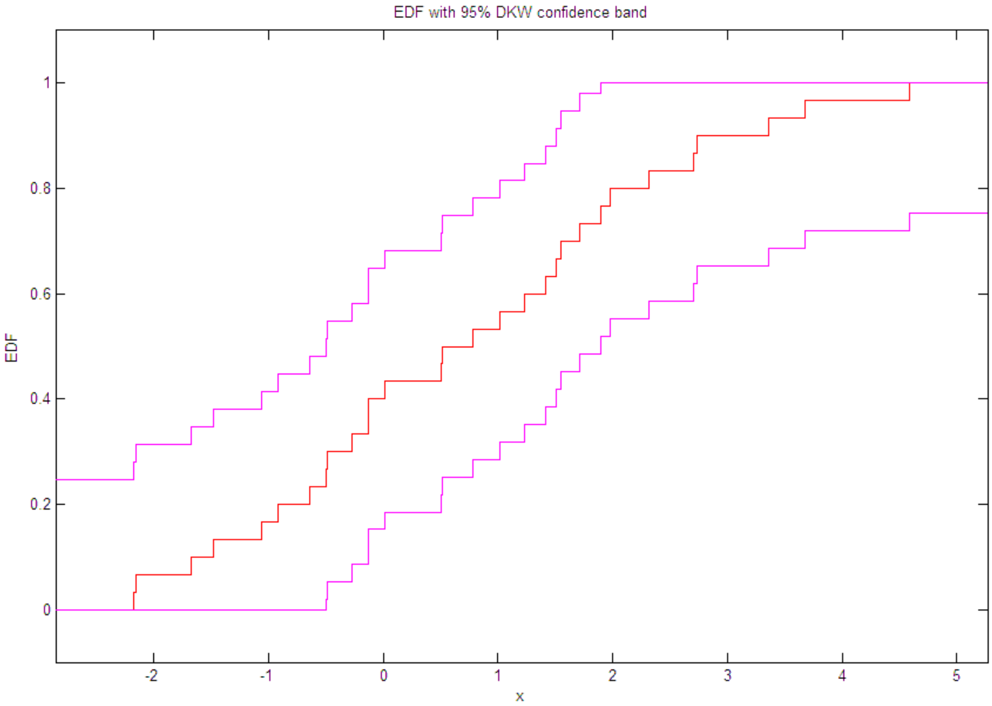
(b) Now suppose that our observed data values are contained in the file uniform_data.mat. Use the parametric bootstrap to obtain $N = 1000$ bootstrap estimates, $\hat{\theta}_1, \ldots, \hat{\theta}_N$. Plot the density histogram for these estimates.

(c) Repeat part (b) using the nonparametric bootstrap.

(d) Compare and comment on the 3 density histograms obtained.

Solutions

(1) (a)



(b) Standard error of $\hat{q}$

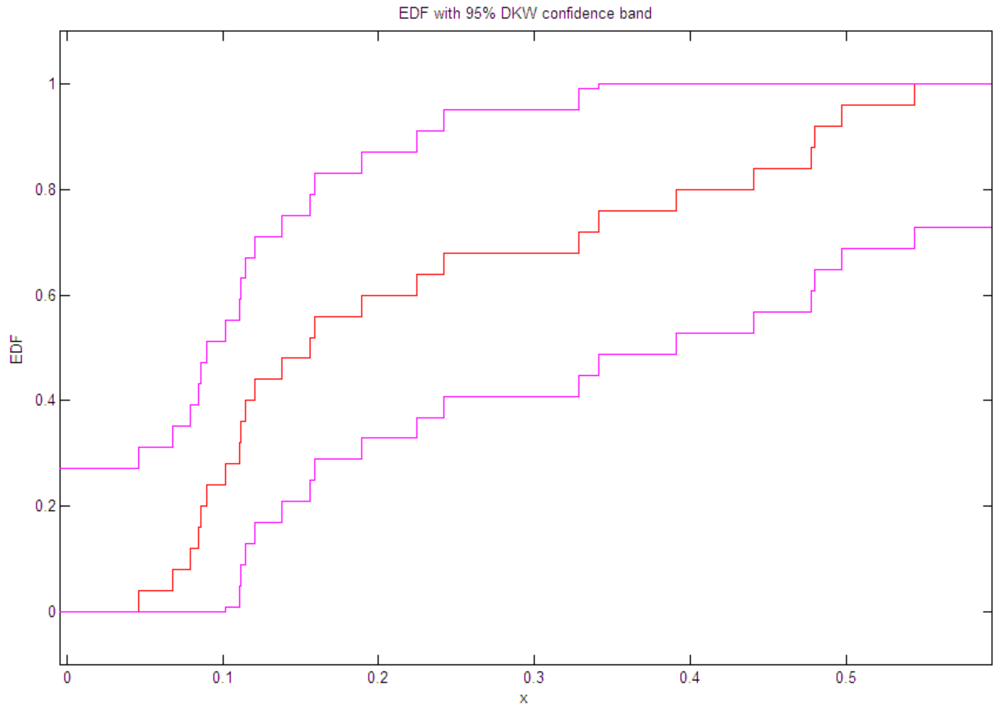
= sample standard deviation of $\hat{q}_1, \ldots, \hat{q}_N$

where $\hat{q}_j$ is the ML estimate from bootstrap realization j consisting of $n = 30$ sample values drawn randomly with replacement from the original data. Using $N = 10{,}000$ as before, I got

standard error of $\hat{q} \approx 0.4796$.

(c) For this data set, the estimates of the standard error of $\hat{q}$ from the two bootstrap procedures are about the same.

② (a)



(b) Using the same number, N = 1000, of bootstrap realizations as I did last week for this problem, I obtain the following approximate 95% confidence intervals with the nonparametric bootstrap:

for α : $(1.3424,\ 2.3849)$,

for β : $(4.1011,\ 11.1209)$,

for m : $(0.0803,\ 0.1953)$.

(c) Comparing with the confidence intervals obtained using the parametric bootstrap last week, these intervals are considerably narrower. Which set of confidence intervals is more believable? In this case, where we know that the data do indeed come from a beta distribution, I would tend to believe the parametric bootstrap confidence intervals more. Thus, for this data set, the nonparametric bootstrap confidence intervals turn out to be overly optimistic (i.e. under-estimating the estimation error, hence the narrower confidence intervals). This is caused mainly by the small sample size ($n = 25$), which does not provide a good estimate of the DF, on which the nonparametric bootstrap is based. Part (d) confirms this.

(d) Now with the data in beta_data2.mat, we have a bigger sample containing $n = 100$ values from the same beta distribution.

Approximate 95% parametric bootstrap confidence intervals :

α : $(1.5135, 2.5259)$,
β : $(5.6730, 10.2056)$,
m : $(0.0920, 0.1570)$.

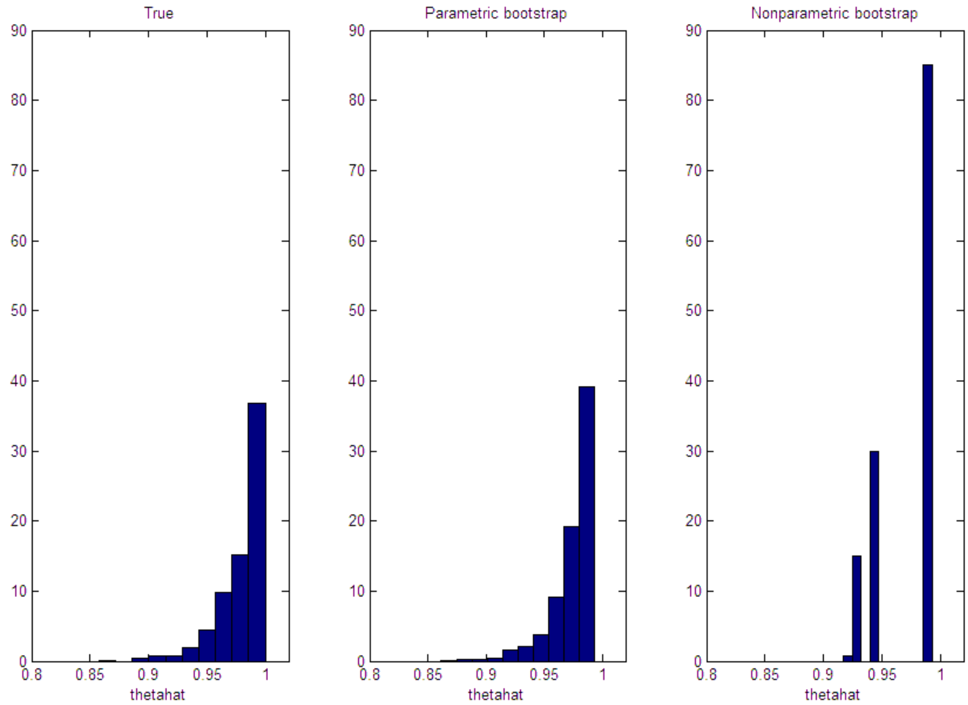
Approximate 95% nonparametric bootstrap confidence intervals :

α : $(1.4997, 2.6514)$,
β : $(5.6615, 10.7837)$,
m : $(0.0875, 0.1565)$.

We observe that the confidence intervals are all generally narrower than before because of the larger sample size. This is expected and comes as no surprise.

More importantly, the differences in the confidence intervals between the two bootstrap procedures are now less pronounced. With a larger sample size n, the empirical DF better estimates the DF. In turn, this makes the nonparametric bootstrap more reliable.

3



From the density histograms, we see that for this estimation problem, the parametric bootstrap estimates the distribution of $\hat{\theta} = \max\{X_1, \ldots, X_n\}$ fairly well. The nonparametric bootstrap, on the other hand, estimates the distribution of $\hat{\theta}$ poorly.