Homework 6

Saurabh Maheshwari - 100%

June 6, 2018

Answer 1

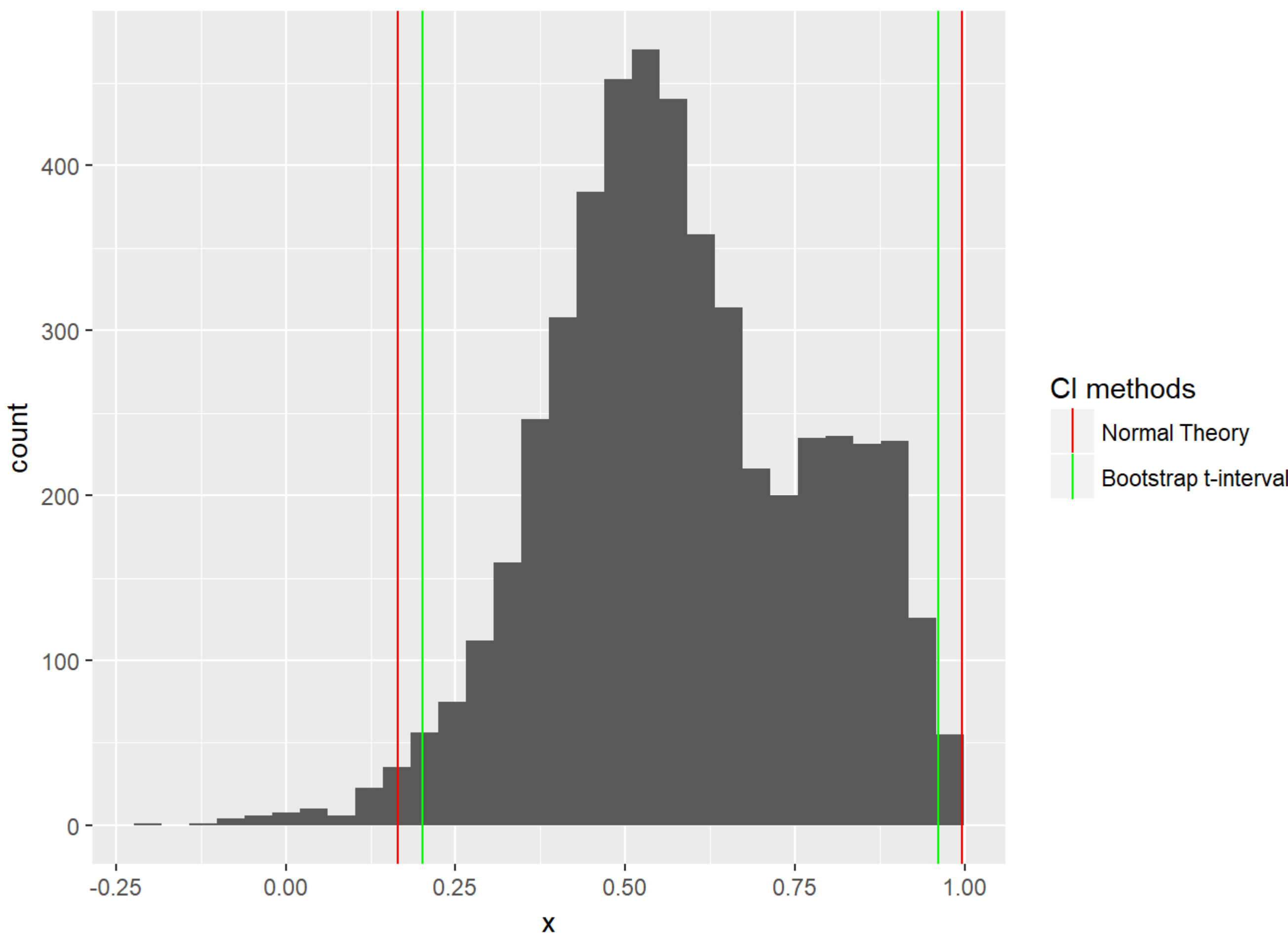
part a) Estimate of correlation coefficient comes out to be 0.5459189

part b) Standard Error using

```
1) Jackknife: 0.2547034
2) Bootstrap: 0.1939639
```

part c) 95% confidence intervals using

```
1) normal theory: (0.9611696, 0.2008449)
2) bootstrap t-interval: (0.9970185, 0.164996)
```



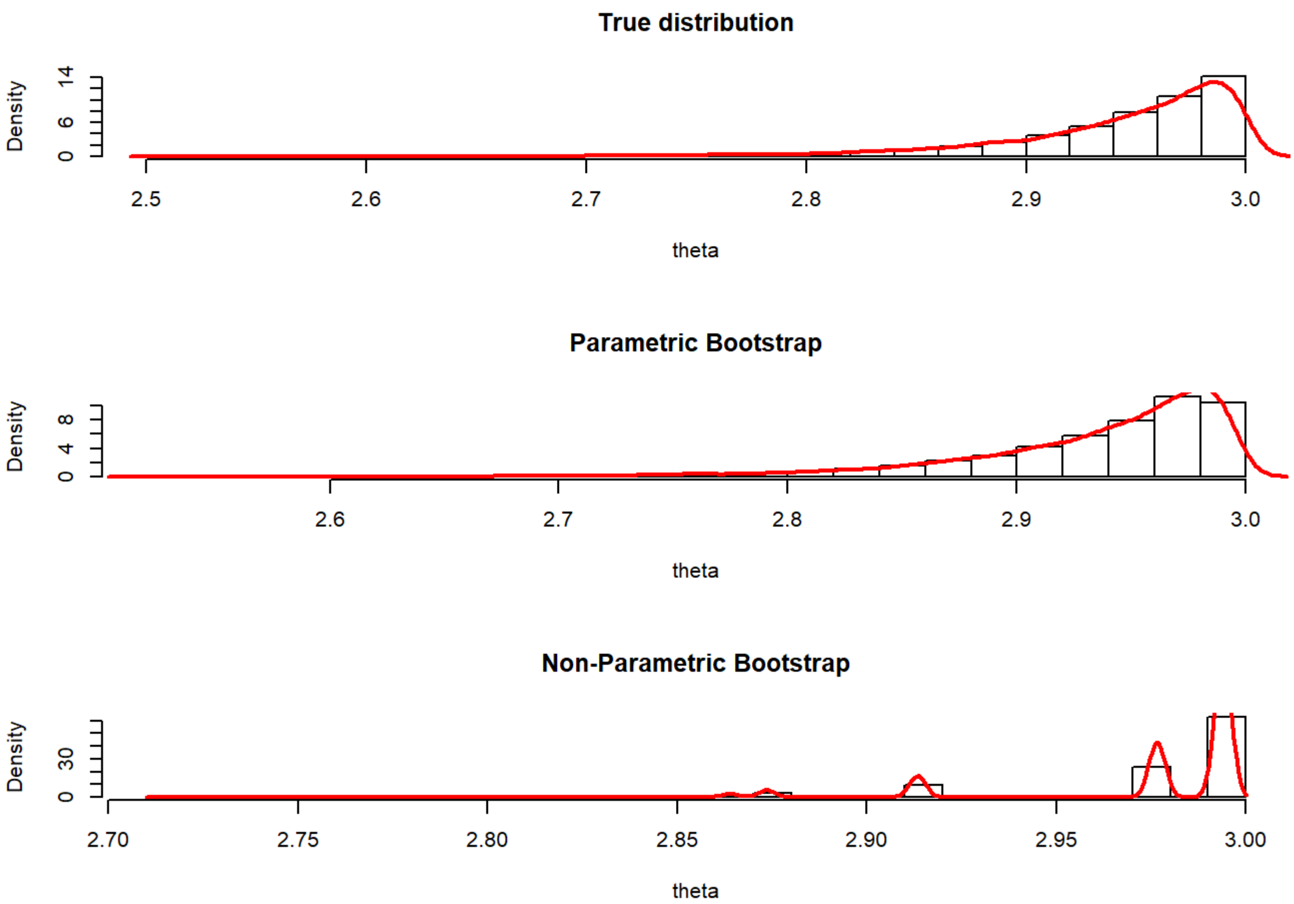
Answer2

part a, b) Derived. Handwritten document appended.

part c) Using parametric bootstrap, $Var_F(\theta)$ = 0.0033111. Variance calculated using analytic expression in part b comes out to be 0.0033271. Thus, parametric bootstrap is a good approximation to the true variance.

part d) Using non-parametric bootstrap, $Var_F(\theta)$ = 0.0010793

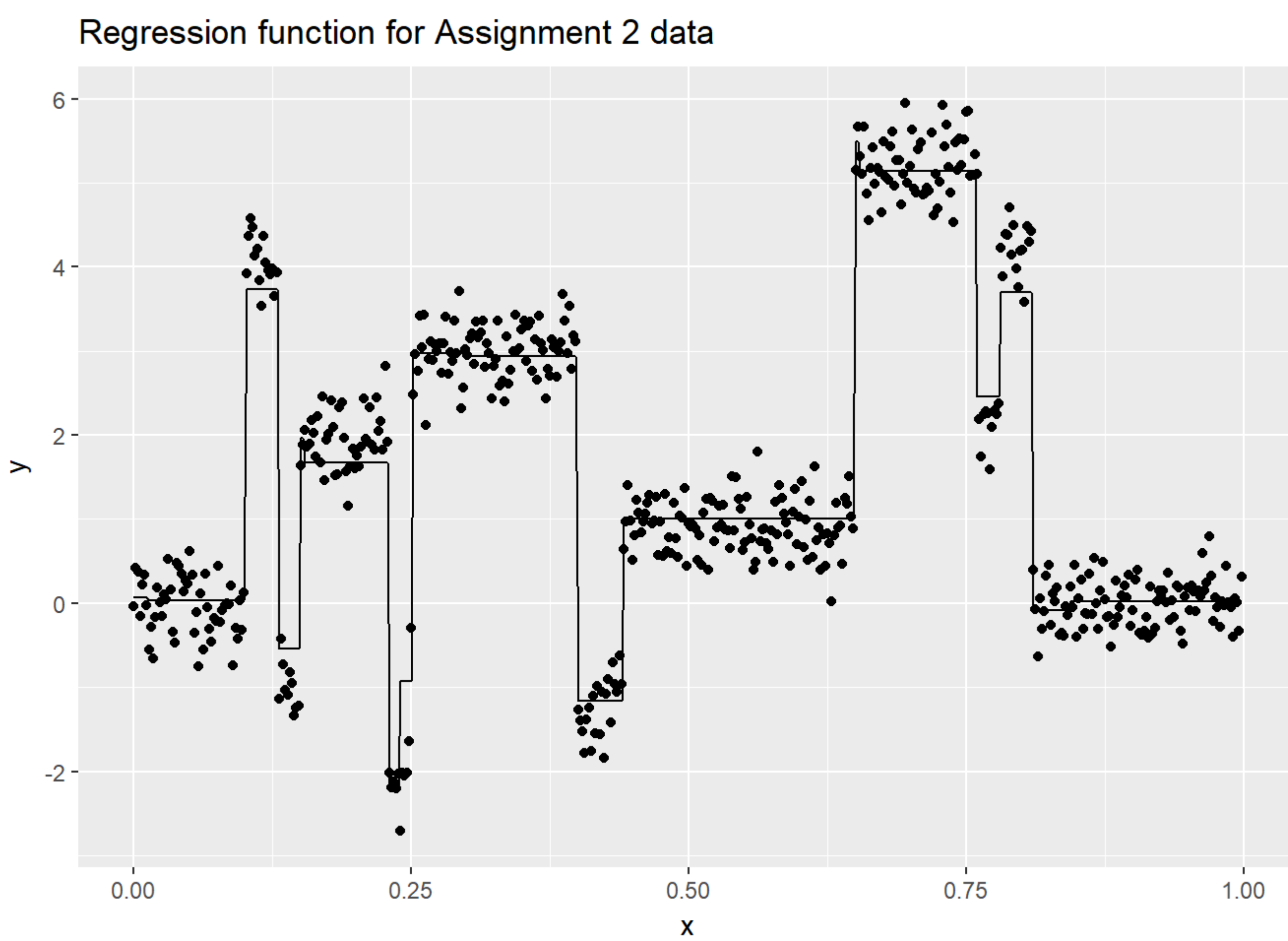
part e) Comparing parametric, non parametric bootstrap results and the actual distribution, the following histograms are plotted.



part f) Parametric bootstrap distribution is a good approximation of the true distribution as compared to the non parametric distribution. The non parametric distribution fails here probably because of small sample size that is not sufficient to capture enough information from the true distribution.

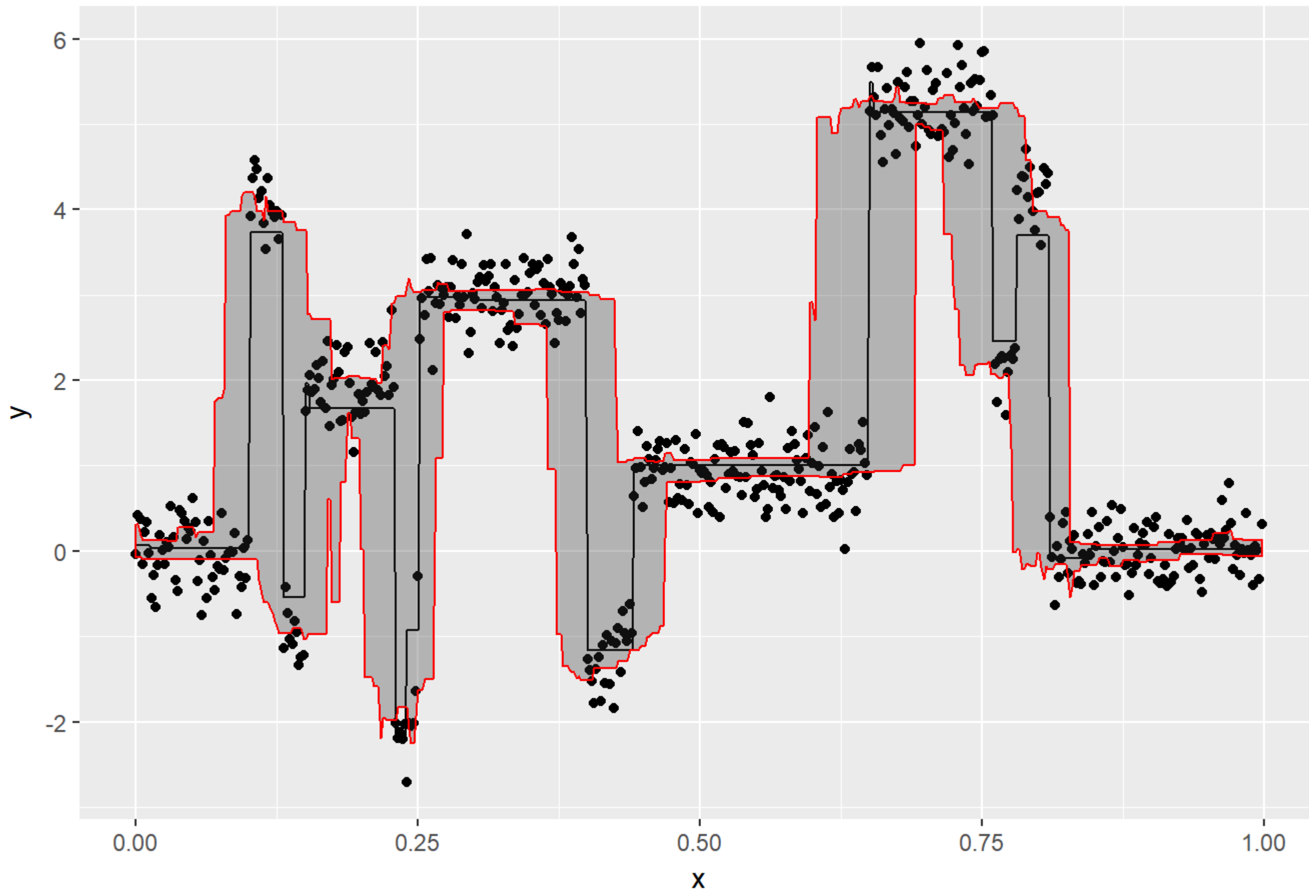
Answer 3

Fitted regression function for data from assignment 2:

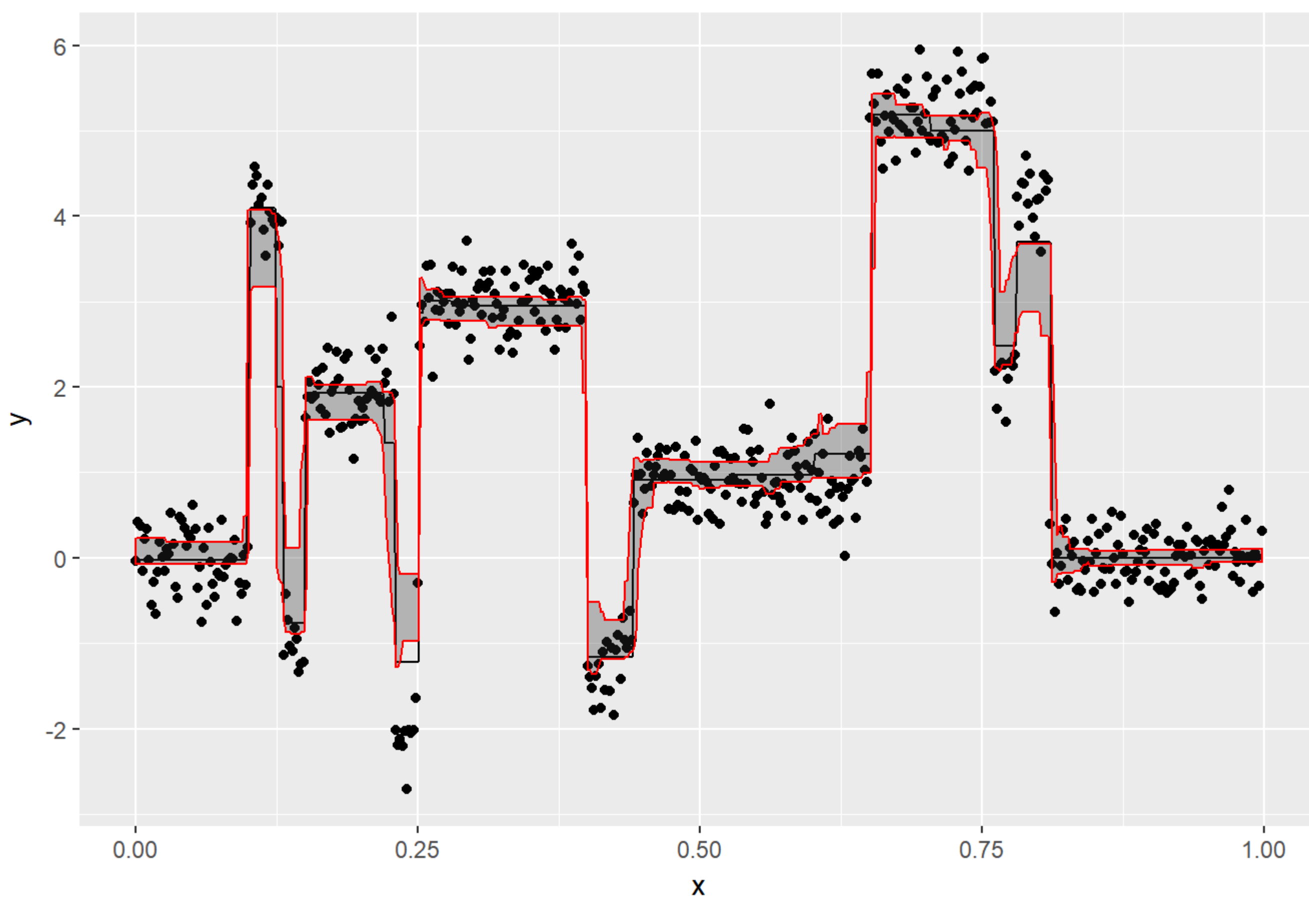


95% confidence interval are shown below:

Using bootstrap pairs

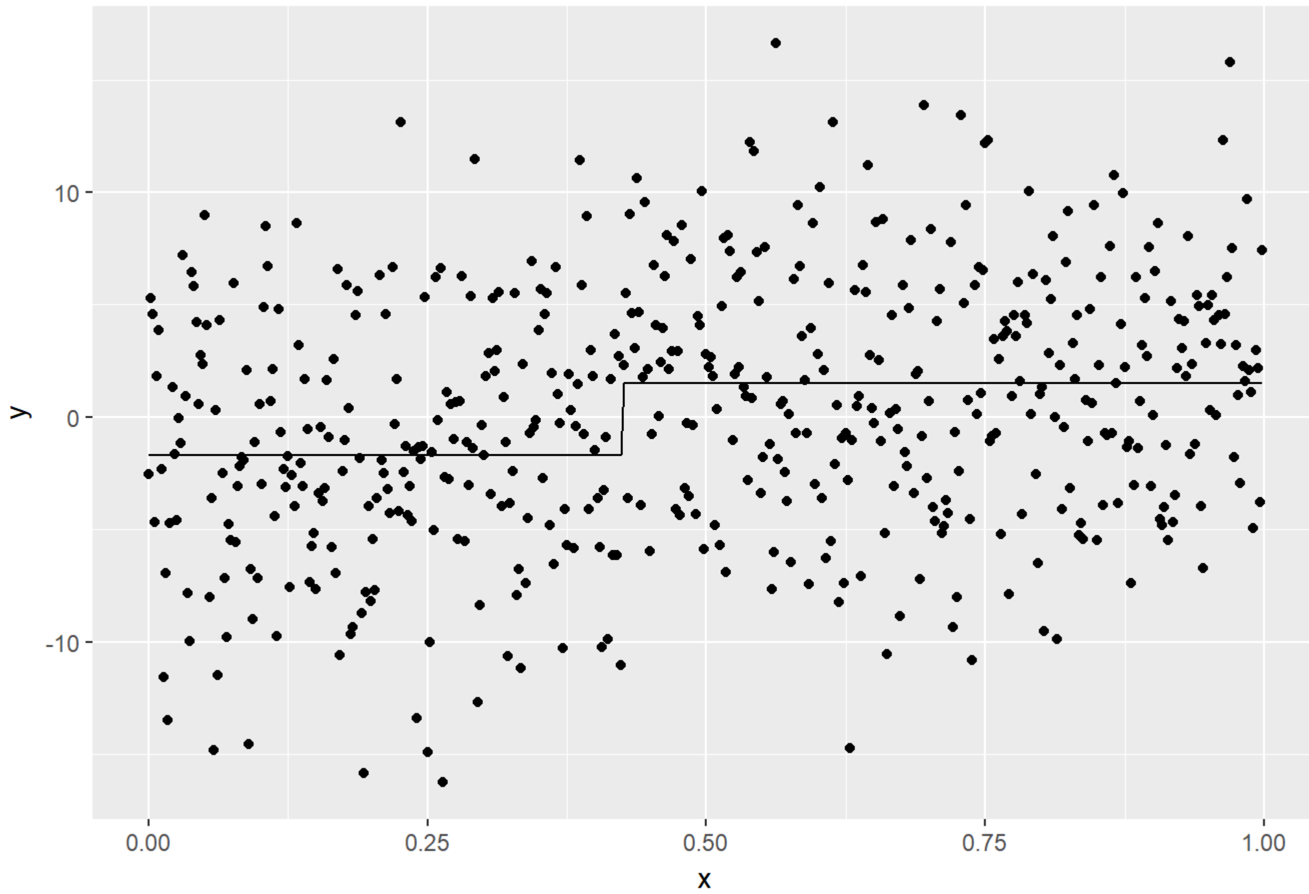


Using bootstrap residuals



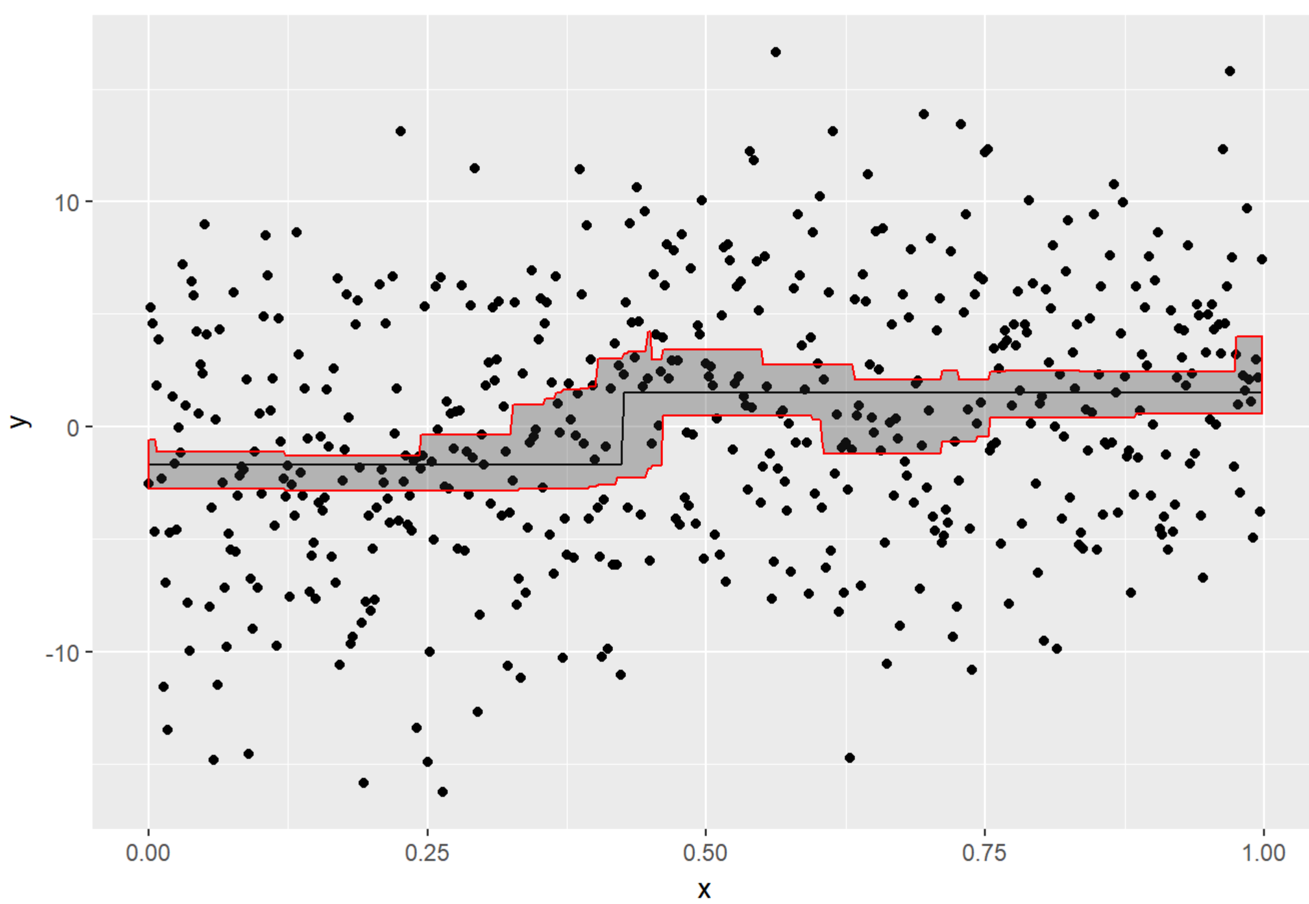
Fitted regression function for $f(x)$:

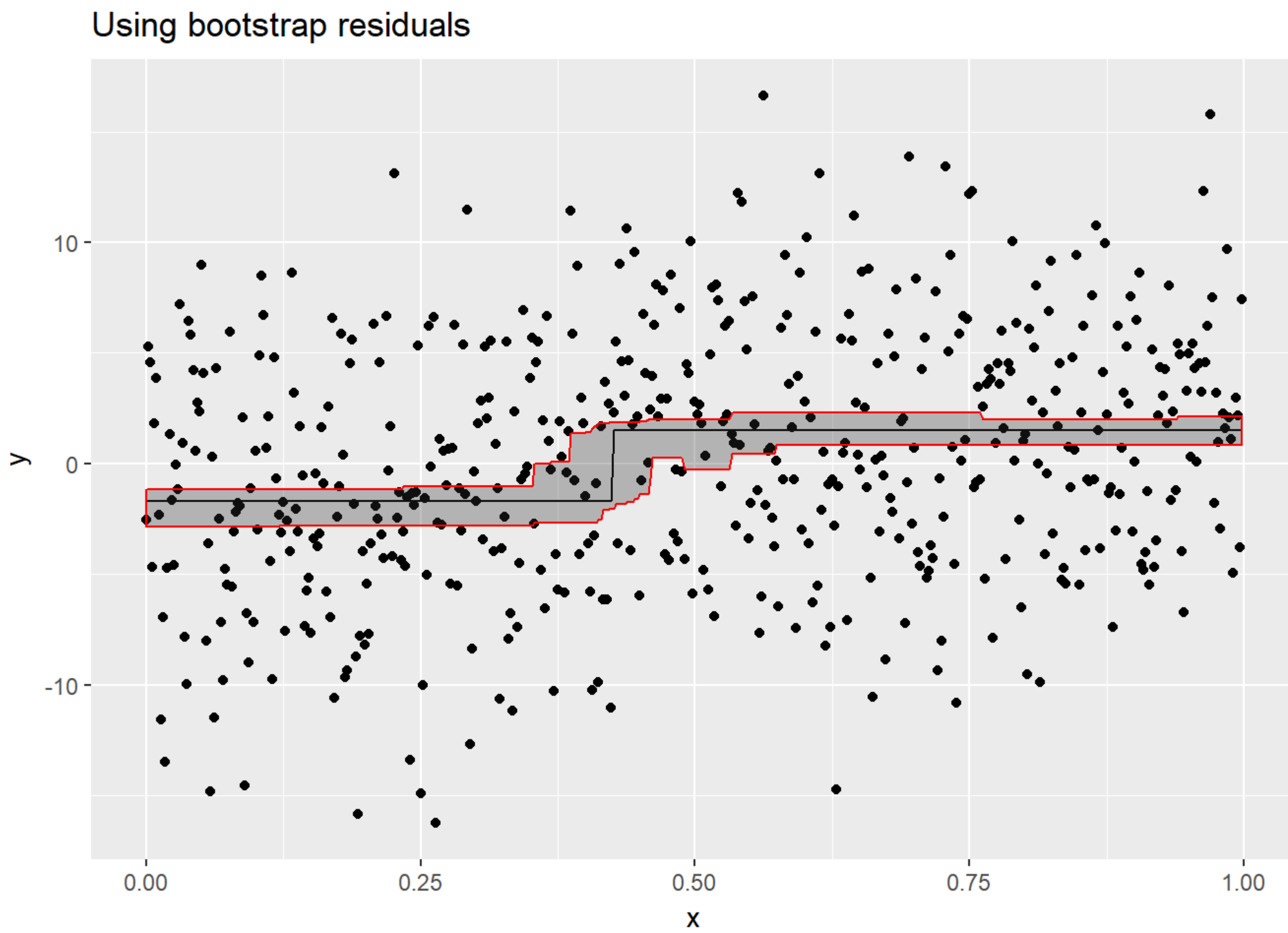
Regression function for f(x)



95% confidence intervals are shown below:

Using bootstrap pairs





Bootstrapping residuals give a better confidence interval as compared to bootstrapping pairs in both the cases. For the data of assignment 2, the confidence intervals at the jumps are very wide, especially in the case of bootstrapping pairs. Thus, they are not reasonable to use.

One way to find the confidence interval at the jumps is to do non-parametric bootstrapping on the function values obtained after genetic algorithm. Using the sampled bootstrp values, calculate an estimate for the confidence intervals using bootstrap t-interval method.

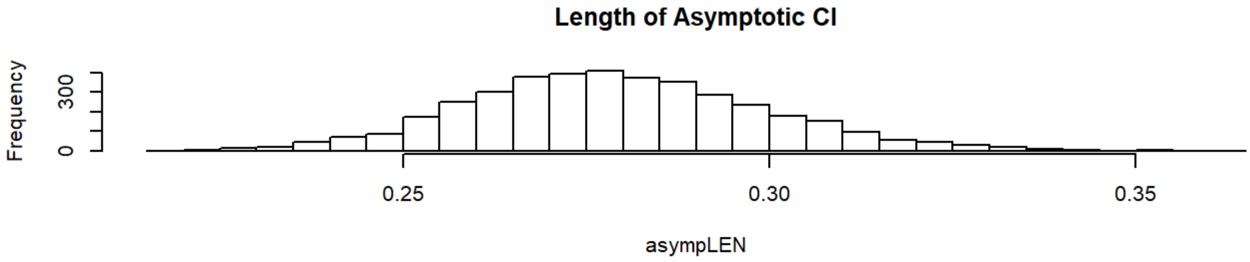
Answer 4

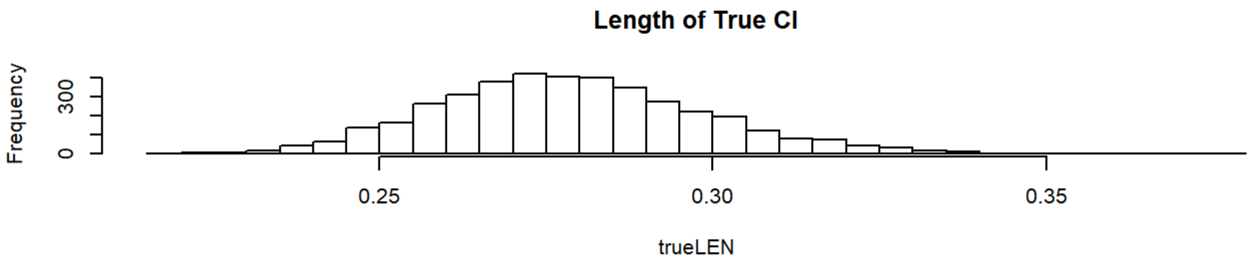
part a, b, c, d) Derived. Handwritten document appended.

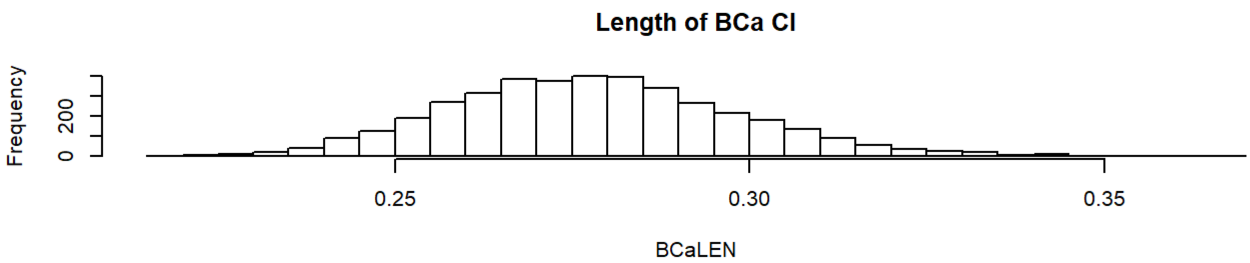
part e) A simulation study was conducted where total 4000 estimates of confidence intervals for each method were simulated. The percent coverage of each method are listed below:

```
1) Asymptotic CI: 0.94
2) exact CI: 0.952
3) BCa CI: 0.94825
```

Thus, BCa method coverage is better than the asymptotic method coverage. Though, the computational expense for BCa method is greater than the asymptotic method, the trade off should be kept in mind before selecting any of these methods. Histograms for the length of CI for each method is plotted below.







mean and standard deviation of lengths of these methods are as follows:

```
1) Asymptotic CI: mean = 0.2792105; std = 0.0201781
2) exact CI: mean = 0.2784366; std = 0.0194999
3) BCa CI: mean = 0.2779432; std = 0.0200307
```

Thus, mean and standard deviation of the 3 methods are approximately the same.

Solution 2)

$\hat{\theta} = X_{max} = max(x_1, \ldots x_n)$

a) $P(\hat{\theta} < t) = P(X_{max} < t)$

$$= P(X_1 < t) \cdot P(X_2 < t) \ldots P(X_n < t)$$

$$= \frac{t^n}{\theta^n} \; ; \forall \; t \in [0, \theta]$$

$$= 1 \; ; \forall \; t > \theta$$

$$= 0 \; ; \forall \; t < 0$$

$$\Rightarrow P(\hat{\theta} < t) = (\min(\theta, t) / \theta)^n \cdot I[t > 0]$$

b) $Var(\hat{\theta}) = E(\hat{\theta}^2) - E(\hat{\theta})^2$

$$= \int_0^\theta t^2 \cdot \frac{n t^{n-1}}{\theta^n} dt - E(\hat{\theta})^2$$

$$= \frac{n}{\theta^n} \cdot \frac{t^{n+2}}{n+2} \Big|_0^\theta - E(\hat{\theta})^2$$

$$= \frac{\theta^2 \cdot n}{n+2} - E(\hat{\theta})^2$$

$$E(\hat{\theta}) = \int_0^\theta t \cdot \frac{n t^{n-1}}{\theta^n} dt$$

$$= \frac{n \cdot \theta}{n+1}$$

thus, $var(\hat{\theta}) = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} = \left[\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right] \theta^2$

Solution 4)

a) $f(x) = \lambda e^{-\lambda x}$

log likelihood:

$\log(L) = n\log\lambda - \lambda \Sigma x_i$

$\log' L(L) = \frac{n}{\lambda} - \Sigma x_i = 0$

$\Rightarrow \lambda = n / \Sigma x_i \leftarrow$ MLE

b) By assymptotic normality of MLE

$\sqrt{n}\,(\hat{\theta} - \theta) \xrightarrow{D} N(0, \frac{1}{I_F})$; where I_F is the fisher information

$I_F = -E\left(\frac{d^2 \log L(\theta)}{d(\theta)^2} \,\Big|\, \theta\right)$

$\Rightarrow \log'(L) = n/\lambda - \Sigma x_i$

$\Rightarrow \log''(L) = -1/\lambda^2$

Thus, $I_F = 1/\lambda^2$

Hence, $\sqrt{n}\,(\hat{\lambda}_n - \lambda) \xrightarrow{D} N(0, \lambda^2)$

c) Using delta method:

$\sqrt{n}\,[g(X_n) - g(\theta)] \xrightarrow{D} N(0, \sigma^2 [g'(\theta)]^2)$

Here, $g(\theta) = \log(\theta) \Rightarrow g'(\theta) = 1/\theta$

$\Rightarrow \sqrt{n}\,[\log(\hat{\lambda}_n) - \log(\lambda)] \xrightarrow{D} N(0, \lambda^2 \cdot 1/\lambda^2) = N(0,1)$

(c) Thus, as $\sqrt{n}\,[\log(\hat{\lambda}_n) - \log(\lambda)] \xrightarrow{D} N(0,1)$

$\Rightarrow \log(\hat{\lambda}_n) - \log(\lambda) \xrightarrow{D} N(0, 1/\sqrt{n})$

$$\Rightarrow \quad \frac{\log(\hat{\lambda}_n) - \log(\lambda)}{1/\sqrt{n}} \in (-z_{\alpha/2}, z_{\alpha/2})$$

$$\Rightarrow \quad \frac{-z_{\alpha/2}}{\sqrt{n}} \leq \log(\hat{\lambda}_n) - \log(\lambda) \leq \frac{z_{\alpha/2}}{\sqrt{n}}$$

$$\Rightarrow \quad \left(\log \hat{\lambda}_n - \frac{z_{\alpha/2}}{\sqrt{n}}\right) \leq \log(\lambda) \leq \log(\hat{\lambda}_n) + \frac{z_{\alpha/2}}{\sqrt{n}}$$

$$\Rightarrow \quad e^{(\log \hat{\lambda}_n - z_{\alpha/2}/\sqrt{n})} \leq \lambda \leq e^{(\log \hat{\lambda}_n + z_{\alpha/2}/\sqrt{n})}$$

$\Rightarrow$ CI for λ at α level $\equiv \left(\hat{\lambda}_n e^{-z_{\alpha/2}/\sqrt{n}}, \ \hat{\lambda}_n e^{z_{\alpha/2}/\sqrt{n}}\right)$

d) $\lambda(X_1 + X_2 + \cdots + X_n) \sim \text{gamma}(N, 1)$

$$\Rightarrow G^{-1}\left(\frac{\alpha}{2}\right) \leq \lambda\left(\Sigma x_i\right) \leq G^{-1}(1-\alpha/2)$$

$$\Rightarrow \quad G^{-1}(\alpha/2) \cdot \frac{1}{n} \leq \lambda \frac{(\Sigma x_i)}{n} \leq G^{-1}\frac{(1-\alpha/2)}{n}$$

$$\Rightarrow \quad \hat{\lambda}_n G^{-1}(\alpha/2)/n \leq \lambda \leq \hat{\lambda}_n G^{-1}(1-\alpha/2)/n$$

Thus, CI for λ at α level $\equiv \left(\hat{\lambda}_n G^{-1}(\alpha/2)/n, \ \hat{\lambda}_n G^{-1}(1-\alpha/2)/n\right)$