1. Suppose that $X_1, \cdots, X_n$ are independent Exponential random variables with pdf

$$f(x;\lambda) = \lambda \exp(-\lambda x) \quad \text{for } x \geq 0$$

where $\lambda > 0$ is an unknown parameter.

(a) [10 marks] Suppose that $\{k_n\}$ is a sequence of positive integers such that $k_n/n \to \tau \in (0,1)$ where $\sqrt{n}(k_n/n - \tau) \to 0$ as $n \to \infty$. Consider estimators of λ of the form

$$\widehat{\lambda}_n(\tau) = \frac{a(\tau)}{X_{(k_n)}}$$

for some $a(\tau)$.

(i) For $\tau \in (0,1)$, find $a(\tau)$ such that $\widehat{\lambda}_n(\tau) \xrightarrow{p} \lambda$ as $n \to \infty$.

Consider this. From lecture, we know that

$$x_{(k)} = x_{(k_n)} \xrightarrow{p} F^{-1}(\tau)$$

Which is equivalent to $\lim_{n\to\infty} P(|X_{(k_n)} - F^{-1}(\tau)| > \varepsilon) = 0$ for $\varepsilon > 0$

$\Rightarrow$ $\lim_{n\to\infty} P\left(X_{(k_n)}\left|1 - \frac{F^{-1}(\tau)}{X_{(k_n)}}\right| > \varepsilon\right) = 0$ taking out $X_{(k_n)}$ from absolute value

we know that $X_{(k_n)} \geq 0$ since it is an exponential family

Now, take $a(\tau) = \lambda F^{-1}(\tau)$

Then consider this, $\lim_{n\to\infty} P(|\hat{\lambda}_n - \lambda| > \varepsilon)$ for $\varepsilon > 0$

Then, $\lim_{n\to\infty} P\left(\left|\frac{a(\tau)}{X_{(k_n)}} - \lambda\right| > \varepsilon\right)$

so, $\lim_{n\to\infty} P\left(\left|\frac{\lambda F^{-1}(\tau)}{X_{(k_n)}} - \lambda\right| > \varepsilon\right)$ since $a(\tau) = \lambda F^{-1}(\tau)$

$\lim_{n\to\infty} P\left(\lambda \cdot \left|\frac{F^{-1}(\tau)}{X_{(k_n)}} - 1\right| > \varepsilon\right) = 0$

Since $\lambda > 0$, the limit evaluates to 0

Hence $\hat{\lambda}_n \xrightarrow{p} \lambda$ as $n \to \infty$ ∎

(ii) For $a(\tau)$ derived above, we will have

$$\sqrt{n}(\hat{\lambda}_n(\tau) - \lambda) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\tau))$$

Find an expression for $\sigma^2(\tau)$ and find the value of τ (to 3 decimal places) that minimizes $\sigma^2(\tau)$.

We know that $\sqrt{n}\left(\frac{k_n}{n} - \tau\right) \to 0$ as $n \to \infty$. We also know that $f(F^{-1}(\tau)) > 0$

Then from lecture, we have $\sqrt{n}\left(X_{(k_n)} - F^{-1}(\tau)\right) \xrightarrow{d} N\left(0, \frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))}\right)$

Using delta method let $g(X) = \frac{a(\tau)}{X} = \frac{\lambda F^{-1}(\tau)}{X}$

Then $g'(x) = -\frac{\lambda F^{-1}(\tau)}{x^2}$

Then $g(X_{(k_n)}) = \frac{\lambda F^{-1}(\tau)}{X_{(k_n)}} = \hat{\lambda}_n(\tau)$

$g(F^{-1}(\tau)) = \frac{\lambda F^{-1}(\tau)}{F^{-1}(\tau)} = \lambda$

$g'(F^{-1}(\tau)) = \frac{-\lambda}{F^{-1}(\tau)}$

So, $\sqrt{n}\left(g(X_{(k_n)}) - g(F^{-1}(\tau)\right) \xrightarrow{d} N\left(0, \left(\frac{-\lambda}{F^{-1}(\tau)}\right)^2 \cdot \frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))}\right)$

$\sqrt{n}\left(\hat{\lambda}_n(\tau) - \lambda\right) \xrightarrow{d} N\left(0, \left(\frac{-\lambda}{F^{-1}(\tau)}\right)^2 \cdot \frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))}\right)$

Using R to compute τ that minimizes $\sigma^2(\tau)$, $\tau_{min} = 0.797$

(b) [10 marks] The file `prob1.txt` contains the 50 smallest values of a sample of 100 observations from an Exponential distribution with parameter λ.

(i) Compute the maximum likelihood estimate of λ based on these data. (Hint: Use spacings.)

(ii) Compute an estimate of the standard error of the estimate in (i).

$Y_i \sim exp(\lambda)$

PDF of exponential distribution : $f(x) = \lambda exp(-\lambda x)$

$$P(Y;\lambda) = \prod_{i=1}^{50} \lambda e^{-\lambda y_i}$$

$$\mathcal{L}(\lambda;Y) = \ln P(y;\lambda)$$

$$= \sum_{i=1}^{50} \ln(\lambda) - \lambda y_i$$

$$\frac{\partial}{\partial\lambda}\mathcal{L}(\lambda;Y) = \sum_{i=1}^{50} \frac{1}{\lambda} - Y_i$$

$$= \frac{50}{\lambda} - \sum_{i=1}^{50} Y_i$$

Equal to 0 and solve :

$$\frac{50}{\lambda} = \sum_{i=1}^{50} Y_i$$

$$\hat{\lambda} = \frac{50}{\sum_{i=1}^{50} Y_i} = 0.405$$

Y_i is defined to be as following. $Y_n = X_{(n)} - X_{(n-1)} = D_{n-1}$

Using R $\hat{\lambda}$ is approximately :

ii) Standard error of MLE estimate is given by Fisher's information.

$$\hat{SE}(\hat{\lambda}) = \left(-\frac{\partial^2}{\partial\lambda^2}\mathcal{L}(\lambda;y)\right)^{-1/2}$$

$$= \left(-\frac{\partial}{\partial\lambda}\sum_{i=1}^{50}\frac{1}{\lambda} - y_i\right)^{-1/2}$$

$$= \left(\sum_{i=1}^{50}\frac{1}{\lambda^2}\right)^{-1/2}$$

$$= \left(\frac{50}{\lambda^2}\right)^{-1/2} = \frac{\hat{\lambda}}{\sqrt{50}}$$

Using R, $\hat{SE}(\hat{\lambda}) = 0.0573$

2. Suppose that $X_1, \cdots, X_n$ are independent positive random variables with common cdf $F(x)$. The Atkinson index of F is defined by

$$A(F) = 1 - \frac{1}{E_F(X)} \exp\left[E_F(\ln(X))\right]$$

where $E_F(\cdot)$ denotes expected value with respect to the distribution F.

(a) [10 marks] Suppose that $G(x) = F(x - b)$ for some $b > 0$.

(i) Show that $A(G) \le A(F)$. (Hint: If $X \sim F$ then $X + b \sim G$. You may find Jensen's inequality useful here.)

$$A(G) = 1 - \frac{1}{E_G(x+b)} \exp\left[E_G(\ln(x+b))\right]$$

$$A(F) = 1 - \frac{1}{E_F(X)} \exp\left[E_F(\ln(X))\right]$$

Then, $A(G) \le A(F) \Rightarrow 1 - \frac{1}{E_G(x+b)} \exp\left[E_G(\ln(x+b))\right] \le 1 - \frac{1}{E_F(X)} \exp\left[E_F(\ln(X))\right]$

Equivantley this is, $\frac{1}{E(x+b)} \exp\left[E(\ln(x+b))\right] \geq \frac{1}{E(X)} \exp\left[E(\ln(X))\right]$

Consider this.

Let $g(x) = e^x$ then, using jensen's inequality $g(E(x)) \le E(g(x))$

$$\exp(E(\ln(x+b))) \le E(\exp(\ln(x+b)))$$
$$\le E(x+b)$$
$$\exp(E(\ln(x))) \le E(\exp(\ln(x)))$$
$$\le E(x)$$

So, $\frac{1}{E(x+b)} \exp(E(\ln(x+b))) \le 1$

and, $\frac{1}{E(x)} \exp(E(\ln(x))) \le 1$

Subtracting we get $\frac{1}{E(x+b)} \exp(E(\ln(x+b))) - \frac{1}{E(x)} \exp(E(\ln(x))) \geq 0$

Then, $\frac{1}{E(x+b)} \exp(E(\ln(x+b))) \geq \frac{1}{E(x)} \exp(E(\ln(x)))$

This is what we wanted to show.

Hence $A(G) \le A(F)$ ∎

(ii) When is $A(G) = A(F)$ (for $b > 0$)?

We know $A(G) = 1 - \frac{1}{E(x+b)} \exp[E(\ln(x+b))]$

$A(F) = 1 - \frac{1}{E(X)} \exp[E(\ln(X))]$

When $A(G) = A(F)$, then

$$\frac{1}{E(x+b)} \exp[E(\ln(x+b))] = \frac{1}{E(X)} \exp[E(\ln(X))]$$

Then, $$\frac{E(X)}{E(X+b)} = \exp[E(\ln(x)) - E(\ln(X+b))]$$

$$\frac{E(X)}{E(X+b)} = \exp[E(\ln(x)) - E(\ln(X+b))]$$

$$= \exp(E(\ln(x) - \ln(x+b)))$$ Using jensen's inequality $g(x) = e^x$

$$\leq E\left(\frac{X}{X+b}\right)$$

However jensen's inequality becomes equality iff $g(x)$ is affine or X constant

If X constant, then $\frac{E(X)}{E(X+b)} = \frac{X}{X+b} = \exp(E(\ln(x) - \ln(x+b)))$

So, $A(G) = A(F)$ iff X is constant. ∎

(b) [10 marks] A sample of 200 observations from an unknown distribution F is given in a file `prob2.txt` on Quercus. Compute an estimate of $A(F)$ and an estimate of its standard error. Justify your method.

Using R code, $\hat{A}(F) = A(\hat{F}) = 0.843$

$\hat{SE}(A(F)) = 0.0305.$

Given the data, I used Jacknife method to estimate the standard error of estimate of $A(F)$.

My estimate of $\hat{A}(F)$ is about 0.843 and the average value that I got for $A(F)$ by using leave one out method came to 0.850

We see that estimator is well approximated by an average.

Hence Jacknife will perform fairly well. ∎

3. Suppose that $X_1, \cdots, X_n$ are independent continuous random variables with density

$$f(x;\theta) = \theta(\theta+1)x^{\theta-1}(1-x) \quad \text{for } 0 \le x \le 1$$

where $\theta > 0$ is an unknown parameter.

(a) [7 marks] Find the MLE of θ and give an estimator of its standard error based on the observed Fisher information.

$$P(x;\theta) = \prod_{i=1}^{n} \theta(\theta+1)x_i^{\theta-1}(1-x_i)$$

$$l(\theta;x) = \ln P(x;\theta)$$

$$= \sum_{i=1}^{n} \ln(\theta) + \ln(\theta+1) + (\theta-1)\ln(x_i) + \ln(1-x_i)$$

$$\frac{\partial}{\partial\theta} l(\theta;x) = \sum_{i=1}^{n} \frac{1}{\theta} + \frac{1}{\theta+1} + \ln(x_i)$$

$$= \frac{n}{\theta} + \frac{n}{\theta+1} + \sum_{i=1}^{n} \ln(x_i)$$

$$= n\left(\frac{2\theta+1}{\theta(\theta+1)}\right) + \sum_{i=1}^{n} \ln(x_i)$$

Equal to 0 :

$$2\theta n + n = -\theta(\theta+1)\sum_{i=1}^{n} \ln(x_i) \qquad \text{Call } \sum_{i=1}^{n} \ln(x_i) = b$$

$$b\theta^2 + b\theta + 2\theta n + n = 0$$

$$b\theta^2 + \theta(b+2n) + n = 0$$

$$\hat{\theta} = \frac{-(b+2n) \pm \sqrt{(b+2n)^2 - 4bn}}{2b}$$

SE using fisher information :

-> need 2nd derivative of likelihood

$$\frac{\partial^2}{\partial^2\theta} l(\theta;x) = \sum_{i=1}^{n} -\frac{1}{\theta^2} - \frac{1}{(\theta+1)^2}$$

$$= -\frac{n}{\theta^2} - \frac{n}{(\theta+1)^2}$$

Then, $\hat{SE}(\hat{\theta}) = \left(-\frac{\partial^2}{\partial^2\theta} l(\hat{\theta};x)\right)^{-1/2}$

$$= \left(\frac{n}{\hat{\theta}^2} + \frac{n}{(\hat{\theta}+1)^2}\right)^{-1/2}$$

(b) [13 marks] Define $\bar{X}_n$ to be the sample mean of $X_1, \cdots, X_n$.

(i) Find $\widehat{\theta}_n = g(\bar{X}_n)$ so that $\widehat{\theta}_n \xrightarrow{p} \theta$, that is, $\{\widehat{\theta}_n\}$ is a consistent sequence of estimators.

Want to find $\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0$

We know that $\bar{X} \approx E(X)$ is a consistent estimator.

So, $E(X) = \int_0^1 x \cdot f(x)\, dx$

$= \int_0^1 \theta(\theta+1) x^{\theta} (1-x)\, dx$

$= \theta(\theta+1)\int_0^1 x^{\theta}\, dx - \theta(\theta+1)\int_0^1 x^{\theta+1}\, dx$

$= \theta x^{\theta+1}\Big|_0^1 - \frac{\theta(\theta+1)}{\theta+2} x^{\theta+2}\Big|_0^1$

$= \theta - \frac{\theta(\theta+1)}{\theta+2}$

Since this is equal to $\bar{x}$,

$\frac{\theta(\theta+2) - \theta(\theta+1)}{\theta+2} = \bar{x}$

$\theta = (\theta+2)\bar{x}$

$\theta - \theta\bar{x} = 2\bar{x}$

Then, $\hat{\theta} = \frac{2\bar{x}}{1-\bar{x}}$ ∎

(ii) By the Delta Method, $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\theta))$. Find $\sigma^2(\theta)$.

By CLT we have $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \psi^2)$

where ψ^2 is the variance of X

Using delta method we get $\sqrt{n}(g(\bar{x}) - g(\mu)) \xrightarrow{d} N(0, [g'(\mu)]^2 \psi^2)$

we want $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta))$

so $g(\bar{x}) = \hat{\theta}_n$.

Then let $g(x) = \frac{2x}{1-x}$ Since $g(\bar{x}) = \frac{2\bar{x}}{1-\bar{x}} = \hat{\theta}_n$ shown previously

Now compute $\psi^2 = Var(X)$ using second moment.

$$Var(X) = E(X^2) - E(X)^2$$

$$E(X^2) = \int_0^1 x^2 f(x)\,dx$$

$$= \int_0^1 x^2 \theta(\theta+1) x^{\theta-1}(1-x)\,dx$$

$$= \theta(\theta+1)\int_0^1 x^{\theta+1}\,dx - \theta(\theta+1)\int_0^1 x^{\theta+2}\,dx$$

$$= \frac{\theta(\theta+1)}{\theta+2} x^{\theta+2}\Big|_0^1 - \frac{\theta(\theta+1)}{\theta+3} x^{\theta+3}\Big|_0^1$$

$$= \frac{\theta(\theta+1)}{\theta+2} - \frac{\theta(\theta+1)}{\theta+3}$$

$$= \frac{\theta(\theta+1)}{(\theta+2)(\theta+3)}$$

Since $E(X) = \frac{\theta}{\theta+2} \Rightarrow E(X)^2 = \frac{\theta^2}{(\theta+2)^2}$

Then $Var(X) = E(X^2) - E(X)^2$

$$= \frac{\theta(\theta+1)}{(\theta+2)(\theta+3)} - \frac{\theta^2}{(\theta+2)^2}$$

$$= \frac{2\theta}{(\theta+2)^2(\theta+3)}$$

Then $\sigma^2(\theta) = [g'(\mu)]^2 \psi^2$

$$= \left(\frac{2}{(1-\mu)^2}\right)^2 \cdot \frac{2\theta}{(\theta+2)^2(\theta+3)}$$

$$= \left(\frac{(\theta+2)^2}{2}\right)^2 \cdot \frac{2\theta}{(\theta+2)^2(\theta+3)} = \frac{\theta(\theta+2)^2}{2(\theta+3)}$$ ∎

(iii) Two estimators of the standard error of $\widehat{\theta}_n$ are

$$\widehat{\mathrm{se}}_1(\widehat{\theta}_n) = \frac{|g'(\bar{X})|S}{\sqrt{n}} \quad \text{and} \quad \widehat{\mathrm{se}}_2(\widehat{\theta}_n) = \frac{\sigma(\widehat{\theta}_n)}{\sqrt{n}}$$

where S^2 is the sample variance of $X_1, \cdots, X_n$ and $\sigma^2(\theta)$ is as defined in part (ii). When would we prefer $\widehat{\mathrm{se}}_1(\widehat{\theta}_n)$ to $\widehat{\mathrm{se}}_2(\widehat{\theta}_n)$?

$$\hat{se}_1(\hat{\theta}_n) = \frac{|g'(\bar{x})|S}{\sqrt{n}}$$

$$= \left(\left|\frac{2\bar{x}}{1-\bar{x}}\right|\right) \times \sqrt{\frac{2\theta}{n(\theta+2)^2(\theta+3)}}$$

$$\hat{se}_2(\hat{\theta}_n) = \sqrt{\frac{\theta(\theta+2)^2}{8n(\theta+3)}}$$

Essentially the only difference b/w $\hat{se}_1(\hat{\theta}_n)$ and $\hat{se}_2(\hat{\theta}_n)$ are that they are multiplied by $|g'(\bar{x})|$ and $g'(\mu)$ respectively.

This means that we prefer $\hat{se}_1(\hat{\theta}_n)$ when population mean μ is unknown and we have sufficiently large N s.t by law of large number, we can approximate μ as $\bar{x}$ with significant confidence. ∎