Statistics and inferencing
Homework 2

Name: Basil Khowaja

id: BK08432

Problem 6)

$$f(x|\alpha) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}\left[x(1-x)\right]^{\alpha-1}$$

a) as α changes, it affects the concentration and spread of the distribution.

for small values of α (close to 0), the density $f(x|\alpha)$ becomes more concentrated towards the endpoints 0 and 1, as the term $(x(1-x))^{\alpha-1}$ becomes very steep.

As α increases, the density becomes more uniform across $[0,1]$, with less concentration towards the end points and a flatter shape.

b) Given $E(x) = 1/2$ we set it equal to the sample mean $\bar{X} = 1/2$

$$\mathrm{var}(x) = \frac{1}{4(2a+1)} = s^2$$

rearranging for a

$$a = \frac{1}{8s^2} - \frac{1}{2}$$

$$\hat{a} = \frac{1}{8s^2} - \frac{1}{2}$$

where s^2 is the sample variance.

c) $$L(a) = \prod_{i=1}^{n} \frac{\Gamma(2a)}{\Gamma(a)^2} \left[x_i(1-x_i)\right]^{a-1}$$

$$\log L(a) = n\log\left(\frac{\Gamma(2a)}{\Gamma(a)^2}\right) + (a-1)\sum_{i=1}^{n}\log(x_i(1-x_i))$$

differentiating wrt to a

$$\frac{d}{da}\log L(a) = 0$$

$$\left[n\log\frac{\Gamma(2a)}{\Gamma(a)^2} + (a-1)\sum_{i=1}^{n}\log(x_i(1-x_i))\right]' = 0$$

↳ this equation satisfies the a

d) to find the asymptotic variance of the MLE for a, using fisher information

$$I(a) = -E\left(\frac{d^2}{da^2} \log L(a) \right)$$

$$\boxed{\text{asymptotic variance} = \frac{1}{I(a)}}$$

Any//

e) to find a Sufficient Statistic for a, we can use the factorization theorem which states that a statistic $T(x)$ is Sufficient for a Parameter a, if the likelihood function $L(a)$ can be factored into 2 Parts

$$f(u|a) = \frac{\Gamma(2a)}{\Gamma(a)^2} \left[u(1-u)\right]^{a-1}$$

for an i.i.d Sample, the likelihood function is $L(a) = \prod_{i=1}^{n} \frac{\Gamma(2a)}{\Gamma(a)^2} [x_i(1-x_i)]^{a-1}$

$$L(a) = \left(\frac{\Gamma(2a)}{\Gamma(a)^2} \right)^n \prod_{i=1}^{n} [x_i(1-x_i)]^{a-1}$$

taking the log likelihood we get:

$$\log l(a) = n\log\left(\frac{\Gamma(2a)}{\Gamma(a)^2}\right) + (\alpha-1)\sum_{i=1}^{n}\log(x_i(1-x_i))$$

from this expression we see that the log likelihood depends on the data only through the term.

$$T(x) = \sum_{i=1}^{n}\log(x_i(1-x_i))$$

thus this is the sufficient statistic for α.

Problem 7)

a) for a Pareto distribution the

$$E(x) = \frac{\theta x_0}{\theta - 1}$$

Setting the sample mean equal to the theoritical mean

$$\bar{X} = \frac{\theta x_0}{\theta - 1}$$

rearranging the equation to solve for θ

$$\theta = \frac{\bar{X}}{\bar{X} - x_0}$$

$$\hat{\theta} = \frac{\bar{X}}{\bar{X} - x_0} \quad \text{Ans//}$$

b) $$L(\theta) = \prod_{i=1}^{n} \theta x_0^{\theta} x_i^{-\theta-1}$$

$$\log L(\theta) = n\log\theta + n\theta\log x_0 - (\theta+1)\sum_{i=1}^{n}\log x_i$$

$$\frac{d}{d\theta}\log L(\theta) = \frac{n}{\theta} + n\log x_0 - \sum_{i=1}^{n}\log x_i = 0$$

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n}\log x_i - n\log x_0} \quad \text{Ans//}$$

c) $$I(\theta) = -E\left(\frac{d^2}{d\theta^2}\log L(\theta)\right)$$

from the log-likelihood

$$\log L(\theta) = n\log\theta + n\theta\log x_0 - (\theta+1)\sum_{i=1}^{n}\log x_i$$

$$\frac{d}{d\theta}\log L(\theta) = \frac{n}{\theta} + n\log x_0 - \sum_{i=1}^{n}\log x_i$$

$$\frac{d^2}{d\theta^2}\log L(\theta) = \frac{-n}{\theta^2}$$

$$I(\theta) = -E\left(\frac{-n}{\theta^2}\right) = \frac{n}{\theta^2}$$

$$\text{Asymptotic variance} = \frac{1}{I(\theta)} = \boxed{\frac{\theta^2}{n}}$$

Ans

d) Using the factorization theorem

$$l(\theta) = \prod_{i=1}^{n} \theta x_0^{\theta} x_i^{-\theta-1}$$

$$l(\theta) = \theta^n x_0^{n\theta} \prod_{i=1}^{n} x_i^{-\theta-1}$$

$$l(\theta) = \theta^n x_0^{n\theta} X_{(1)}^{-(\theta+1)n}$$

$T(x) = X_{(1)}$ is a sufficient statistic for θ.

Problem 8)

a) $$f(x|\theta) = \frac{x}{\theta^2} e^{-x^2(2\theta^2)}, \quad x \geq 0$$

found through internet $\longrightarrow$ $E(x) = \theta\sqrt{\frac{\pi}{2}}$

$$\bar{X} = \theta\sqrt{\frac{\pi}{2}}$$

rearranging for θ;

$$\hat{\theta} = \frac{\bar{X}}{\sqrt{\pi/2}} = \frac{\bar{X}}{\sqrt{\pi/2}}$$

$$\boxed{\hat{\theta} = \frac{\bar{X}}{\sqrt{\pi/2}}}$$ Ans

b) $$L(\theta) = \prod_{i=1}^{n} \frac{x_i}{\theta^2} e^{-x_i^2/2\theta^2}$$

$$\log L(\theta) = \sum_{i=1}^{n} \log x_i - 2n\log\theta - \frac{1}{2\theta^2} \sum_{i=1}^{n} x_i^2$$

$$\frac{d}{d\theta} \log L(\theta) = -\frac{2n}{\theta} + \frac{1}{\theta^3} \sum_{i=1}^{n} x_i^2 = 0$$

$$\theta^2 = \frac{1}{2n} \sum_{i=1}^{n} x_i^2$$

$$\hat{\theta} = \sqrt{\frac{1}{2n} \sum_{i=1}^{n} x_i^2}$$ Ans

c) as found earlier;

$$\frac{d}{d\theta} \log L(\theta) = \frac{-2n}{\theta} + \frac{1}{\theta^3} \sum_{i=1}^{n} x_i^2$$

$$\frac{d^2}{d\theta^2} \log L(\theta) = \frac{2n}{\theta^2} - \frac{3}{\theta^4} \sum_{i=1}^{n} x_i^2$$

$$I(\theta) = -E\left(\frac{d^2}{d\theta^2} \log L(\theta)\right) = \frac{2n}{\theta^2}$$

$$\text{Asymptotic variance} = \frac{1}{I(\theta)} = \frac{\theta^2}{2n} \quad \text{Ans}$$

Q9) $f(x|\theta) = (\theta+1)x^\theta, \quad 0 \leq x \leq 1$

a) $$E(x) = \int_0^1 x(\theta+1)x^\theta \, dx$$

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

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

$$\overline{X} = \frac{\theta+1}{\theta+2}$$

$$\theta+2 = \frac{\theta+1}{\overline{X}}$$

$$\hat{\theta} = \frac{\bar{x}}{1-\bar{x}} \quad \text{Ans}$$

b) $$l(\theta) = \prod_{i=1}^{n} (\theta+1) x_i^{\theta} = (\theta+1)^n \prod_{i=1}^{n} x_i^{\theta}$$

$$\log l(\theta) = n\log(\theta+1) + \theta \sum_{i=1}^{n} \log x_i$$

$$\frac{d}{d\theta} \log l(\theta) = \frac{n}{\theta+1} + \sum_{i=1}^{n} \log x_i = 0$$

$$\theta + 1 = \frac{-n}{\sum_{i=1}^{n} \log x_i}$$

$$\boxed{\hat{\theta} = -1 + \frac{n}{\sum_{i=1}^{n} \log x_i}} \quad \text{Ans}$$

c) $$\frac{d^2}{d\theta^2} \log l(\theta) = \frac{-n}{(\theta+1)^2}$$

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

$$\text{Asymptotic variance} = \frac{1}{I(\theta)}$$

Asymptotic variance $= \frac{(\theta+1)^2}{n}$ Ans

d) $L(\theta) = \prod_{i=1}^{n} (\theta+1) x_i^{\theta} = (\theta+1)^n \prod_{i=1}^{n} x_i^{\theta}$

$L(\theta) = (\theta+1)^n \left(\prod_{i=1}^{n} x_i \right)^{\theta}$

taking the Product of x_i's we get

$T(x) = \prod_{i=1}^{n} x_i$ → Ans

↳ sufficient Statistic

Q10) let $x_1 = 10$, $x_2 = 68$ and $x_3 = 112$

a) $L(\theta) = (1-\theta)^{2x_1} \cdot (2\theta(1-\theta))^{x_2} \cdot \theta^{2x_3}$

$\log L(\theta) = 2x_1 \log(1-\theta) + x_2 \log(2\theta(1-\theta))$
$+ 2x_3 \log\theta$

$\log L(\theta) = 2x_1 \log(1-\theta) + x_2 \log 2 + x_2 \log\theta + x_2 \log(1-\theta)$
$+ 2x_3 \log\theta$

$\log L(\theta) = (2x_1 + x_2)\log(1-\theta) + (x_2 + 2x_3)\log\theta$
$+ x_2 \log 2$

$$\frac{x^2 + 2x_3}{\theta} = \frac{2x_1 + x_2}{1-\theta}$$

$$\hat{\theta} = \frac{x_2 + 2x_3}{2(x_1 + x_2 + x_3)}$$

$x_1 = 10$

$x_2 = 68$

$x_3 = 112$

b) $$\frac{d}{d\theta} \log l(\theta) = \frac{-2x_1}{1-\theta} + \frac{x_2}{\theta} - \frac{x_2}{1-\theta} + \frac{2x_3}{\theta}$$

$$\frac{d}{d\theta} \log l(\theta) = -\frac{2x_1 + x_2}{1-\theta} + \frac{x_2 + 2x_3}{\theta}$$

$$\frac{d^2}{d\theta^2} \log l(\theta) = -\frac{2x_1 + x_2}{(1-\theta)^2} - \frac{x_2 + 2x_3}{\theta^2}$$

$E(x_1) = n(1-\theta)^2$

$E(x_2) = 2n\theta(1-\theta)$

$E(x_3) = n\theta^2$

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

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

Asymptotic variance $= \frac{1}{I(\theta)} = \frac{\theta(1-\theta)}{2n}$

c) $SE(\hat{\theta}) = \sqrt{\dfrac{\hat{\theta}(1-\hat{\theta})}{2n}}$

$\Rightarrow$ for 99% confidence interval, we use the critical value z from the standard normal distribution.

The critical value z for 99% confidence is approximately 2.576.

$$\hat{\theta} \pm z \times SE(\hat{\theta})$$

$$\hat{\theta} \pm 2.576 \times \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{2n}}$$