

**Take home exam;
Due: January 7**

Exercise 1

- (a) We know that: $\int_{-\infty}^{\infty} PDF(x)dx = 1$, as far that we know that the $0 < x < 4$, we can change integration limits, so $\int_0^4 \frac{c}{\sqrt{x}}dx = 1$. Finally we can conclude that:

$$c = \frac{1}{4}$$

- (b) $CDF(x) = \int_{-\infty}^x PDF(x)dx$

$$CDF(x) = \begin{cases} 0, & x < 0 \\ \frac{\sqrt{x}}{2}, & 0 \leq x \leq 4 \\ 1, & x > 4 \end{cases}$$

- (c) $P(x < 0.25) = CDF(0.25) = 0.25$
 $P(x > 1) = 1 - CDF(1) = 0.5$

- (d) $Y = \sqrt{X}$

$$F_Y(y) = P(Y < y) = P(\sqrt{X} < y) = P(X < y^2) = CDF(y^2)$$

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ \frac{y}{2}, & 0 \leq y \leq 2 \\ 1, & y > 2 \end{cases}$$

- (e) $E(y) = \int_{-\infty}^{\infty} y \cdot PDF(y)dy$

$$E(y) = \int_0^2 \frac{y}{2}dy = 1$$

$$Var(y) = E(y^2) - E^2(y) = \int_0^2 \frac{y^2}{2}dy - 1 = \frac{1}{3}$$

Exercise 2

- (a) We know from the CLT that, $\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X_i - Ex \right) \sim N(0, \sigma^2)$, so we can derive, that: $\bar{X} \sim N \left(Ex, \frac{\sigma^2}{n} \right)$

From the initial distribution we can find Ex and σ^2 :

$$\begin{aligned} \text{PDF}(x) &= \frac{dF(x)}{dx} = 4x^{-5} \\ E(x) &= \int_{-\infty}^{\infty} x \cdot \text{PDF}(x) dx = \int_1^{\infty} \frac{4}{x^{-4}} dx = \frac{4}{3} \\ \sigma^2 &= Ex^2 - E^2x = \int_1^{\infty} \frac{4}{x^{-3}} dx = \frac{2}{9}, \text{ and so:} \\ \bar{X} &\sim N\left(\frac{4}{3}, \frac{2}{9 \cdot n}\right) \end{aligned}$$

(b) Let $g(x) = \ln(x)$

We can use delta method, because $g'(Ex) \neq 0$ and $g(x)$ has a derivative equals to $\frac{1}{x}$

As far as we know that:

$$\sqrt{n} \left(\bar{X} - \frac{4}{3} \right) \sim N(0, \sigma^2)$$

We can conclude from delta method:

$$\sqrt{n} \left(g(\bar{X}) - g\left(\frac{4}{3}\right) \right) \sim N\left(0, \sigma^2 \left[g'\left(\frac{4}{3}\right) \right]^2\right), \text{ and so:}$$

$$\ln(x) \sim N\left(\ln\left(\frac{4}{3}\right), \frac{1}{8 \cdot n}\right)$$

(c) We know that:

$$\frac{3 \cdot \sqrt{n}}{\sqrt{2}} \left(\bar{X} - \frac{4}{3} \right) \sim N(0, 1)$$

$$\text{Thus, } \frac{9 \cdot n}{2} \left(\bar{X} - \frac{4}{3} \right)^2 \sim \chi_1^2$$

$$n \cdot \left(\bar{X} - \frac{4}{3} \right)^2 \sim \frac{2}{9} \cdot \chi_1^2$$

$$\begin{aligned} n \cdot \left(\bar{X} - \frac{4}{3} \right)^2 &= n \cdot \left(\bar{X} - 0.8 + 0.8 - \frac{4}{3} \right)^2 = n \cdot (\bar{X} - 0.8)^2 - 2 \cdot \frac{8 \cdot n}{15} \cdot (\bar{X} - 0.8) + \\ &\quad \frac{64}{15^2} \end{aligned}$$

$$2 \cdot \frac{8 \cdot n}{15} \cdot (\bar{X} - 0.8) \sim N\left(\frac{32}{15}, \frac{15^2}{16^2 \cdot n^2}\right)$$

$$\text{So, } n \cdot (\bar{X} - 0.8)^2 \sim \frac{2}{9} \cdot \chi_1^2 + N\left(\frac{416}{225}, \frac{15^2}{16^2 \cdot n^2}\right)$$

Exercise 3

(a) 1. From Chebyshev-Markov inequality we know that:

$$P(|\bar{X} - \mu| > \varepsilon) \leq \frac{\sigma^2}{n \cdot \varepsilon^2}$$

We can find that $\varepsilon = 1.4$ and so, the probability that \bar{X} is not in interval less than 25 %

2. From CLT we know that $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

To find the probability that \bar{X} is not in the interval we have to find the value of Laplace function.

$$P\left(|N(128, \frac{6.3^2}{81}) - \mu| > \varepsilon\right) = 2 \cdot \Phi\left(\frac{\mu + \varepsilon}{\sigma^2}\right) = 0.02394$$

And so, the probability that \bar{X} is not in the interval is $1 - 0.02394 = 97.6\%$

(b) $129 \pm 1.96 \cdot 6.3 = [116.652, 141.348]$

Exercise 4

$$F(x) = 1 - \frac{Q}{x}, x \geq Q$$

$$f(x|Q) = \frac{dF(x)}{dx} = \frac{Q}{x^2}, x \geq Q$$

Let us define $X_{(1)} = \min_{1 \leq i \leq n} X_i$

Joint pmf of $X_1 \dots X_n$ is :

$$f(x|Q) = \prod_{i=1}^n \frac{Q}{X_i^2} I_{(Q, \infty)}(X_i) = I_{(Q, \infty)}(X_{(1)}) \prod_{i=1}^n \frac{Q}{X_i^2} = I_{(Q, \infty)}(X_{(1)}) \frac{Q^n}{X_1^2 \cdot X_2^2 \cdot \dots \cdot X_n^2}$$

By the Factorization Theorem we can show defining:

$$h(x) = \frac{1}{X_1^2 \cdot X_2^2 \cdot \dots \cdot X_n^2}$$

$$g(t|Q) = Q^n I_{(Q, \infty)}(t)$$

that $f(x|Q) = h(x) \cdot g(T(x)|Q)$.

And so, $T(x)$ is sufficient, QED

Exercise 5

(a) Let $T = X_1 + X_2 + \dots + X_n$ and let $f_t(x_1, x_2, \dots, x_n|\theta)$ be the joint density of X_1, X_2, \dots, X_n .

$$f_t(x_1, x_2, \dots, x_n|\theta) = \begin{cases} \prod_{i=1}^n \theta e^{-\theta x_i}, & \text{if } x_i > 0, \\ 0, & \text{otherwise} \end{cases} = \begin{cases} \theta^n e^{-\theta \sum_{i=1}^n x_i}, & \text{if } x_i > 0, \\ 0, & \text{otherwise} \end{cases} =$$

$$= \theta^n e^{-\theta t} \cdot h(x_1, x_2, \dots, x_n) = g(t, \theta) h(x_1, x_2, \dots, x_n) = \theta^n e^{-\theta t}$$

Where

$$g(t, \theta) = \theta^n e^{-\theta t}$$

$$h(x_1, \dots, x_n) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{otherwise} \end{cases}$$

Hence, T is a sufficient statistic for θ .

(b)

$$f(x; \theta) = \theta e^{-\theta x}$$

$$l(x; \theta) = \log(\theta e^{-\theta x})$$

Fisher information is:

$$I(\theta) = -E\left(\frac{\partial^2 l(x; \theta)}{\partial \theta^2}\right) = -E\left(\frac{1}{\theta^2}\right) = \frac{1}{\theta^2}$$

Hence, Cramer-Rao lower bound:

$$\text{Var}(\hat{\theta}) = \frac{1}{I(\theta)}$$

(c) The likelihood function for θ , given an iid sample $X = (x_1, x_2, \dots, x_n)$:

$$L(\theta) = \prod_{i=1}^n \theta e^{-\theta x_i} = \theta^n e^{-\theta n \bar{x}}, \text{ where } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

The derivative of the likelihood function's logarithm is

$$\frac{\partial}{\partial \theta} \ln(L(\theta)) = \frac{\partial}{\partial \theta} (n \ln \theta - \theta n \bar{x}) = \begin{cases} > 0, & 0 < \theta < \frac{1}{\bar{x}}, \\ = 0, & \theta = \frac{1}{\bar{x}}, \\ < 0, & \theta > \frac{1}{\bar{x}} \end{cases}$$

Consequently the MLE

$$\hat{\theta} = \frac{1}{\bar{x}}$$

The asymptotic distribution of the estimator is $N(\theta, \frac{1}{I(\theta)})$.

(d)

$$L(\theta) = \theta^n e^{-\theta n \bar{x}}$$

$$H_0 : \theta = 1$$

$$H_1 : \theta > 1, \theta = \theta_a$$

Ratio of the likelihood functions is:

$$\frac{L(1)}{L(\theta_a)} = \frac{e^{-n\bar{x}}}{\theta_a^n e^{-\theta_a n \bar{x}}} \leq k$$

$$e^{n\bar{x}(\theta_a - 1)} \theta_a^{-n} \leq k$$

$$(\theta_a - 1) n \bar{x} - n \ln(\theta_a) \leq \ln(k)$$

$$(\theta_a - 1) \sum x_i \leq \ln(k) + n \ln(\theta_a)$$

$$\sum x_i \leq \frac{\ln(k) + n \ln(\theta_a)}{(\theta_a - 1)} = k^*$$

Therefore, the best critical region of size α for test $H_0 : \theta = 1$ against each simple alternative $H_1 : \theta = \theta_a$, where $\theta_a > 1$ if given by:

$$C = \left\{ (x_1, \dots, x_n), \sum_{i=1}^n x_i \leq k^* \right\}$$

Where k^* is selected

$$\alpha = P \left(\sum_{i=1}^n x_i \leq k^*, \text{ when } \theta = 1 \right)$$

$$0.05 = P \left(\sum_{i=1}^n x_i \leq k^*, \text{ when } \theta = 1 \right)$$

$$\sum_{i=1}^n x_i = \eta \sim \Gamma(100, 1)$$

$$F_\eta(k^*) = P(\eta \leq k^*) = 0.05$$

$$\int_0^{k^*} f_{\eta}(x) dx = \int_0^{k^*} x^0 \frac{e^{-\frac{x}{100}}}{100\Gamma(1)} dx = 0.05$$

$$\int_0^{k^*} e^{-\frac{x}{100}} dx = 5, \quad t = -\frac{x}{100}, \quad dt = -\frac{dx}{100}, \quad dx = -100dt, \quad x = k^* \rightarrow t = -\frac{k^*}{100}$$

$$\int_0^{-\frac{k^*}{100}} e^t dt = 5$$

$$\int_0^{-\frac{k^*}{100}} e^t dt = -0.05$$

$$e^{-\frac{k^*}{100}} - 1 = -0.05$$

$$k^* = -100 \ln(0.95)$$

Hence, our test is

$$\sum_{i=1}^n x_i \leq -100 \ln(0.95)$$

(e) The Wald Statistic is:

$$Z_w = \frac{\left(\frac{1}{\bar{x}} - 1\right)^2}{\theta} = (\bar{x} - 1)^2$$

In our task

$$Z_w \leq \chi^2(99) = 81.4$$

Hence

$$(\bar{x} - 1)^2 \leq 81.4 \text{ if } H_0, H_1 \text{ otherwise}$$

Exercise 6

Note that:

$L(\theta) \propto \theta^{\sum_{i=1}^n x_i} \cdot (1 - \theta)^{1 - \sum_{i=1}^n x_i}$, so we can find:

$$\frac{L(\theta_0)}{L(\theta_1)} = \left(\frac{\theta_0}{\theta_1}\right)^{\sum_{i=1}^n x_i} \cdot \left(\frac{1 - \theta_0}{1 - \theta_1}\right)^{1 - \sum_{i=1}^n x_i}$$

Since this is decreasing function of $\sum_{i=1}^n x_i$, we accept H_1 if $\sum_{i=1}^n x_i > c$

We have that $\sum_{i=1}^n x_i \sim \text{Bin}(n, \theta)$ and from CLT $\text{Bin}(n, \theta) \approx N(n\theta, n\theta(1 - \theta))$, and

hence $\frac{\sum_{i=1}^n x_i - n\theta}{\sqrt{n\theta(1 - \theta)}} \sim N(0, 1)$

We now wish to find n so that:

$$P_{\theta=0.5}(\sum_{i=1}^n x_i < c) = 0.9$$

$$P_{\theta=0.6}(\sum_{i=1}^n x_i < c) = 0.1$$

This amounts to solving the equations

$$1.28 = \frac{c - 0.5n}{\sqrt{0.25n}}$$

$$-1.28 = \frac{c - 0.6n}{\sqrt{0.24n}} \text{ and so we got, that } n > 160.54, \text{ it means, that } n = 161$$