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) = $\begin{cases} 0, x < 0 \\ \frac{\sqrt{x}}{2}, 0 \leqslant x \leqslant 4 \\ 1, x > 4 \end{cases}$

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

(d) $Y = \sqrt{X}$
$F_Y(y) = P(Y < y) = P(\sqrt{X} < y) = P(X < y^2) = CDF(y^2)$
$F_Y(y) = \begin{cases} 0, y < 0 \\ \frac{y}{2}, 0 \leqslant y \leqslant 2 \\ 1, y > 2 \end{cases}$

(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: $\overline{X} \sim N\left(Ex, \frac{\sigma^2}{n}\right)$

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

PDF(x) $= \dfrac{dF(x)}{dx} = 4x^{-5}$

$E(x) = \int_{-\infty}^{\infty} x \cdot PDF(x)dx = \int_{1}^{\infty} \dfrac{4}{x^{-4}}dx = \dfrac{4}{3}$

$\sigma^2 = Ex^2 - E^2x = \int_{1}^{\infty} \dfrac{4}{x^{-3}}dx = \dfrac{2}{9}$, and so:

$\overline{X} \sim N\left(\dfrac{4}{3}, \dfrac{2}{9 \cdot n}\right)$

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

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

As far as we know that:

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

We can conclude from delta method:

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

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

(c) We know that:

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

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

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

$n \cdot \left(\overline{X} - \dfrac{4}{3}\right)^2 = n \cdot \left(\overline{X} - 0.8 + 0.8 - \dfrac{4}{3}\right)^2 = n \cdot \left(\overline{X} - 0.8\right)^2 - 2 \cdot \dfrac{8 \cdot n}{15} \cdot \left(\overline{X} - 0.8\right) + \dfrac{64}{15^2}$

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

Let us define $Z = n \cdot \left(\overline{X} - \dfrac{4}{3}\right)^2 + \dfrac{16 \cdot n}{15} \cdot \left(\overline{X} - 0.8\right) - \dfrac{64}{15^2}$

As far as Z consists of two random variables, the pdf of Z is difficult to find.
We can just write that g(z), where g(z) is the PDF of Z is equal to

g(z) $= \int_{-\infty}^{\infty} f_1(x) \cdot f_2(z - x)dx$, where $f_1(x)$ is the PDF of $n \cdot \left(\overline{X} - \dfrac{4}{3}\right)^2$ and

$f_2(x)$ is the PDF of $\dfrac{16 \cdot n}{15} \cdot \left(\overline{X} - 0.8\right)$

However we can see, that σ^2 of $\dfrac{16 \cdot n}{15} \cdot \left(\overline{X} - 0.8\right)$ asymptotically goes to zero, and so, we can think that it is the constant.

So, we can think, that asymptotically $n \cdot \left(\overline{X} - 0.8\right)^2 \sim \dfrac{2}{9} \cdot \chi_1^2 + 1.848$

Exercise 3

(a)
1. From Chebyshev-Markov inequality we know that:
   $P\left(|\overline{X} - \mu| > \varepsilon\right) \leq \dfrac{\sigma^2}{n \cdot \varepsilon^2}$
   We can find that $\varepsilon = 1.4$ and so, the probability that $\overline{X}$ is not in interval less than 25 %
2. From CLT we know that $\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$
   To find the probability that $\overline{X}$ is not in the interval we have to find the value of Laplass function.
   $P\left(|N(128, \dfrac{6.3^2}{81}) - \mu| > \varepsilon\right) = 2 \cdot \Phi\left(\dfrac{\mu + \varepsilon}{\sigma^2}\right) = 0.02394$

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

Exercise 4

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

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

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

Joint pmf of $X_1..X_n$ is :

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

By the Factorization Theorem we can show defining:

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

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

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

And so, T(x) is sufficient, QED

Exercise 5

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

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

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

Where

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

$$h\left(x_1, \cdots, x_n\right) = \begin{cases} 1, if\ x > 0, \\ 0, otherwise \end{cases}$$

Hence, T is a sufficient statistic for θ.

(b)

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

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

Fisher information is:

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

Hence, Cramer-Rao lower bound:

$$Var\left(\hat{\theta}\right) = \frac{\theta^2}{n}$$

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

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

The derivative if the likelihood function's logarithm is

$$\frac{\eth}{\eth\theta} ln\left(L\left(\theta\right)\right) = \frac{\eth}{\eth\theta}\left(nln\theta - \theta n\overline{x}\right) = \begin{cases} > 0, 0 < \theta < \frac{1}{\overline{x}}, \\ = 0, \theta = \frac{1}{\overline{x}}, \\ < 0, \theta > \frac{1}{\overline{x}} \end{cases}$$

Consequently the MLE

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

We know, that $\overline{x} \sim N\left(\theta, \frac{\theta^2}{n}\right)$.

The question is how to find the distribution of $\frac{1}{\overline{x}}$
We know that:
$\sqrt{n} \cdot (\overline{x} - \theta) \sim N(0, \theta^2)$
We know from delta method, that:
$\sqrt{n} \cdot (g(\overline{x}) - g(\theta)) \sim N(0, \sigma^2 \cdot \left[g^{'}(\theta)\right]^2)$
And so, the asymptotic distribution of the estimator is:
$N(\frac{1}{\theta}, \frac{1}{\theta^2 \cdot n})$

(d)

$$L\left(\theta\right) = \theta^n e^{-\theta n\overline{x}}$$

$$H_0 : \theta = 1$$

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

Ratio of the likelihood functions is:

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

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

$$\left(\theta_a - 1\right) n\overline{x} - nln\left(\theta_a\right) \leq ln(k)$$

$$\left(\theta_a - 1\right) \sum x_i \leq ln\left(k\right) + nln\left(\theta_a\right)$$

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

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:

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

Where k^* is selected

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

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

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

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

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

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

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

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

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

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

Hence, our test is

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

(e) The Wald Statistic is:

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

In our task

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

Hence

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

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 $\sum_{i=1}^n x_i \sim \mathrm{Bin}(n,\theta)$ and from CLT $\mathrm{Bin}(n,\theta) \approx N(n\theta, n\theta(1-\theta))$, 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:
$P_{\theta=0.5}\left(\sum_{i=1}^n x_i < c\right) = 0.9$
$P_{\theta=0.6}\left(\sum_{i=1}^n x_i < c\right) = 0.1$
This amounts to solving the equations
$1.28 = \frac{c - 0.5n}{\sqrt{0.25n}}$
$-1.28 = \frac{c - 0.6n}{\sqrt{0.24n}}$ and so we got, that n $>$ 160.54, it means, that n $=$ 161