1. (a) Let $X_1, \ldots, X_n$ be a random sample from $f(x|\theta) = \theta x^{-2} I_{(\theta,\infty)}(x)$. Then we have

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} \theta x_i^{-2} I_{[\theta,\infty)}(x_i) = \theta^n \left(\prod_{i=1}^{n} x_i \right)^{-2} I_{[\theta,\infty)}(X_{(1)})$$

Therefore letting $g(T(\underline{X}), \theta) = \theta^n I_{\theta,\infty}(X_{(1)})$ and $h(x) = \prod_{i=1}^{n} x_i^{-2}$ we see that $X_{(1)}$ is a sufficent statistic for θ by the Neyman - Fisher factorization theorem.

   (b) As above we can define the likelihood of the sample as

$$\mathcal{L}(\theta) = \theta^n \left(\prod_{i=1}^{n} x_i \right)^{-2} I_{[\theta,\infty)}(X_{(1)})$$

Notice that $\prod_{i=1}^{n} x_i^{-2} > 0$ and is constant with respect to θ. Hence we only look to maximize $\theta^n I_{[\theta,\infty)}(X_{(1)})$. θ^n increases as θ increases. But notice that $\mathcal{L}(\theta) = 0$ when $\theta > X_{(1)}$. Therefore we look to maximize θ with respect to $\theta \leq X_{(1)}$. Thus

$$\hat{\theta}_{MLE} = X_{(1)}$$

   (c) Since θ is one dimensional, we need only solve $\overline{X} = \mu(f)$. But notice that

$$\mu(f) = \int_{\theta}^{\infty} x \frac{\theta}{x} dx = \theta \int_{\theta}^{\infty} \frac{1}{x} dx = \theta \log(x) \Big|_{\theta}^{\infty} = \infty$$

Therefore, for $X \sim f$, $\mathbb{E}(X^k)$ does not exist for all $k = 1, 2, \ldots$. Therefore the method of moments estiamtor does not exist.

2. Let $X_1, \ldots, X_n$ be a random sample from $f(x|\theta) = \frac{1}{\theta} I_{[0,\theta]}(x)$. Then find the likelihood as follows.

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I_{[0,\theta]}(x_i) = \frac{1}{\theta^n} I_{[0,\theta]}(X_{(n)})$$

Now, we see that $\mathcal{L}$ increases as θ decreases. But notice that $\mathcal{L} = 0$ when $\theta < X_{(n)}$. So we look to make θ as small as possible with respect to $\theta \geq X_{(n)}$. This of course corresponds to

$$\hat{\theta}_{MLE} = X_{(n)}$$

Now, deriving the mean and variance of this estimate we have $X_{(n)} \sim f(y|\theta) = n(F_X(y))^{n-1} f_X(y) = \frac{n}{\theta^n} y^{n-1}$. This leads to the mean and variance calculations

$$\mathbb{E}(\hat{\theta}_{MLE}) = \int_{0}^{\theta} \frac{n}{\theta^n} y^n dy = \frac{n}{\theta^n} \cdot \frac{1}{n+1} y^{n+1} \Big|_{0}^{\theta} = \frac{n}{n+1} \theta$$

$$\mathbb{E}(\hat{\theta}_{MLE}^2) = \int_{0}^{\theta} \frac{n}{\theta^n} y^{n+1} dy = \frac{n}{\theta^n} \cdot \frac{1}{n+2} y^{n+2} \Big|_{0}^{\theta} = \frac{n}{n+2} \theta^2$$

$$\mathbb{V}(\hat{\theta}_{MLE}) = \left[\frac{n}{n+2} \theta^2 \right] - \left[\frac{n}{n+1} \theta \right]^2 = \frac{n}{(n+2)(n+1)^2} \theta^2$$

Now looking to find the method of momements estimator notice that θ is one dimensional so we need only consider $\overline{X} = \mu(f)$. Finding this value, we see that

$$\mu(f) = \int_0^\theta \frac{x}{\theta}dx = \frac{x^2}{2\theta}\bigg|_0^\theta = \frac{\theta}{2}$$

Using this, we see that

$$\hat{\theta}_{MOM} = 2\overline{X}$$

Finding the mean and variance of this estimator we see that

$$\mathbb{E}(\hat{\theta}_{MOM}) = \mathbb{E}(2\overline{X}) = 2\mathbb{E}(\overline{X}) = 2\frac{\theta}{2} = \theta$$

$$\mathbb{V}(\hat{\theta}_{MOM}) = 4\mathbb{V}(\overline{X}) = 4\frac{\theta^2}{12n} = \frac{\theta^2}{3n}$$

Notice that the MLE is biased (slightly) but has variance that converges faster to zero than the variance of the MOM estimator. Moreover, asympotically, the MLE is unbiased. For small sample sizes MOM is preferable while for large sample sizes the MLE is preferable. In general, I would choose the MLE as it is a lower variance estimate.

3. (a) Let $X_1, \ldots, X_n \sim F(x)$ where $F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \left(\frac{x}{\beta}\right)^\alpha & \text{if} \quad 0 \le x \le \beta \\ 1 & \text{if} \quad x > \beta \end{cases}$. This corresponds to $f(x) = \alpha(x/\beta)^{\alpha-1}I_{[0,\beta]}(x)$. Therefore, we can find the joint distribution of the sample as

$$f(\underline{X}|\alpha,\beta) = \prod_{i=1}^n \alpha(x/\beta)^{\alpha-1}I_{[0,\beta]}(x) = \frac{\alpha^n}{\beta^{n(\alpha-1)}}\left(\prod_{i=1}^n x_i\right)^{\alpha-1} I_{[0,\beta]}(X_{(n)})$$

Therefore, letting $g(T(X),\alpha,\beta) = f(\underline{X}|\alpha,\beta)$ and $h(x) = 1$, we see that $\left(\prod_{i=1}^n x_i, X_{(n)}\right)$ are sufficent statistics for (α,β).

(b) We first find the MLE of $\hat{\beta}$. Notice that for $f(\underline{X}|\alpha,\beta) = \mathcal{L}(\alpha,\beta)$ as β deceases, $\mathcal{L}$ increases. However, $\mathcal{L} = 0$ when $\beta < X_{(n)}$. Therefore, we look to minimize β with respect to $\beta \ge X_{(n)}$. Thus

$$\hat{\beta}_{MLE} = X_{(n)}$$

We now look to minimize $\mathcal{L}(\alpha,\beta)$ with respect to α. This is equivalent to minimizing

$$\ell(\alpha,\beta) \equiv \log(\mathcal{L}(\alpha,\beta))$$

$$= n\log(\alpha) - n(\alpha-1)\log(\beta) + (\alpha-1)\sum_{i=1}^n \log(x_i) + \log(I_{[0,\beta]}(X_{(n)}))$$

Taking the derivative with respect to α and setting equal to zero we have

$$\frac{\partial}{\partial \alpha}\ell(\underline{X}, \alpha, \beta) = \frac{n}{\alpha} - n\log(\beta) + \sum_{i=1}^{n}\log(x_i) \overset{set}{=} 0$$

This yields

$$\frac{n}{\alpha} = n\log(\beta) - \sum_{i=1}^{n}\log(x_i) = \frac{n}{n\log(\beta) - \sum_{i=1}^{n}\log(x_i)} = \left(\log(\beta) - \frac{1}{n}\sum_{i=1}^{n}\log(x_i)\right)^{-1}$$

Now due to the invariance property of MLE's we can plug in $\hat{\beta}_{MLE}$ to achieve our MLE

$$\hat{\alpha}_{MLE} = \left(\log(X_{(n)}) - \frac{1}{n}\sum_{i=1}^{n}\log(x_i)\right)^{-1}$$

Since

$$\frac{\partial^2}{\partial \alpha^2}\ell(\alpha, \beta)\Big|_{\alpha = \hat{\alpha}_{MLE}} = -\frac{n}{\hat{\alpha}_{MLE}^2} = -\left(\frac{\sqrt{n}}{\hat{\alpha}_{MLE}}\right)^2 < 0$$

$\hat{\alpha}_{MLE}$ is a in fact an MLE.

(c) From the data, we see that $X_{(n)} = 25$ and $\sum_{i=1}^{n}\log(x_i) = 43.9527$. This yields $\hat{\beta}_{MLE} = 25$ and $\hat{\alpha}_{MLE} = 12.59487$.

4. (a) Let $X_1, X_2, \ldots, X_n \sim f(x|\theta) = \theta x^{\theta-1}I_{[0,1]}(x)$ Then we can define the likelihood as

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

We can then write the log-likelihood as

$$\ell(\theta) = n\log(\theta) + \theta\sum_{i=1}^{n}\log(x_i) - \log\left(\prod_{i=1}^{n}x_i\right)$$

Differentiating with respect to θ and setting to zero we have

$$\frac{\partial}{\partial \theta}\ell(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n}\log(x_i) \overset{set}{=} 0$$

This implies that

$$\hat{\theta}_{MLE} = \left(-\frac{1}{n}\sum_{i=1}^{n}\log(x_i)\right)^{-1}$$

Moreover we see

$$\frac{\partial^2}{\partial \theta^2}\ell(\theta)\Big|_{\theta=\hat{\theta}_{MLE}} = -\frac{n}{\theta^2}\Big|_{\theta=\hat{\theta}_{MLE}} = -\left(\frac{\sqrt{n}}{\hat{\theta}_{MLE}}\right)^2 < 0$$

so $\hat{\theta}_{MLE}$ is in fact a maximum. To calculate the variance, notice that

$$-\log(X_i) \sim f_X(e^{-y})\Big|\frac{d}{dy}e^{-y}\Big| = \theta e^{-y(\theta-1)}e^{-y} = \theta e^{-y\theta}$$

Therefore, we see that $-\log(X_i) \sim Exp(1/\theta)$ and $G := \sum_{i=1}^{n} -\log(x_i) \sim Gamma(n, 1/\theta)$. Now, defining

$$\hat{\theta}_{MLE} = \frac{n}{\sum_{i=1}^{n} -\log(X_i)} = \frac{n}{G}$$

we look to first find the density function of $\frac{n}{G}$. Using the same transformation formula we see that

$$\frac{n}{G} \sim f_G(1/y)\Big|\frac{d}{dt}1/y\Big| = \frac{n\theta^n}{\Gamma(n)}y^{n+1}e^{\theta y}$$

From here we have

$$\mathbb{E}(\frac{n}{G}) = \frac{n\theta^n}{\Gamma(n)}\int_0^\infty \frac{1}{y}y^{n+1}e^{-\theta y}dy = \frac{\theta n}{n-1}$$

$$\mathbb{E}((\frac{n}{G})^2) = \frac{n^2\theta^n}{\Gamma(n)}\int_0^\infty \frac{1}{y^2}y^{n+1}e^{-\theta y}dy = \frac{\theta^2 n^2}{(n-1)(n-2)}$$

$$Var(\frac{n}{G}) = \frac{\theta^2 n^2}{(n-1)(n-2)} - \left(\frac{\theta n}{n-1}\right)^2 = \frac{\theta^2}{(n-1)(n-2)} \longrightarrow 0$$

(b) Now to derive the method of moments estimator for θ, note that θ is one dimensional so we need only consider $\overline{X} = \mu(f)$.

$$\mathbb{E}(X) = \int_0^1 \theta x^{\theta-1}x dx = \frac{\theta}{\theta+1}x^{\theta+1}\Big|_0^1 = \frac{\theta}{1+\theta}$$

Setting equal to $\overline{X}$ and solving for θ, we see

$$\hat{\theta}_{MOM} = \frac{\overline{X}}{1 - \overline{X}}$$

5. (a) Following the model $Y_i = \beta x_i + \epsilon_i$ with $\epsilon_i \sim N(0, \sigma^2)$ we see that $Y_i \sim N(\beta x_i, \sigma^2)$. Thus for a random sample $Y_1, \ldots, Y_n$ following this model, we can derive the

joint distribution as follows

$$
\begin{aligned}
f(\underline{Y}|\beta,\sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2}(y_i - \beta x_i)^2 \right\} \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n}(y_i - \beta x_i)^2 \right\} \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{ -\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2 + \frac{1}{\sigma^2}\sum_{i=1}^{n}\beta y_i x_i - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\beta^2 x_i^2 \right\} \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{ -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\beta^2 x_i^2 \right\} \exp\left\{ -\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2 + \frac{\beta}{\sigma^2}\sum_{i=1}^{n} y_i x_i \right\}
\end{aligned}
$$

Now, letting $c(\beta,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{ -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\beta^2 x_i^2 \right\}$, $h(x) = 1$ we see we have an exponential family. Therefore we see that $T(\theta) = \left(\sum_{i=1}^{n} y_i^2, \sum_{i=1}^{n} y_i x_i\right)$ is a sufficent statistic for (σ^2, β).

(b) Using the form from above, we can write the log-likelihood function as

$$
\ell(\beta,\sigma) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2 + \frac{\beta}{\sigma^2}\sum_{i=1}^{n} y_i x_i - \frac{\beta^2}{2\sigma^2}\sum_{i=1}^{n} x_i^2
$$

Differentiating with respect to β, we have

$$
\frac{\partial}{\partial\beta}\ell(\beta,\sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^{n} y_i x_i - \frac{\beta}{\sigma^2}\sum_{i=1}^{n} x_i^2
$$

Setting eqaual to zero and solving for β, we have

$$
\hat{\beta}_{MLE} = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}
$$

To see why the estimate is unbiased consider the following

$$
\mathbb{E}(\hat{\beta}_{MLE}) = \mathbb{E}\left(\frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}\right) = \frac{\sum_{i=1}^{n} x_i \mathbb{E}(y_i)}{\sum_{i=1}^{n} x_i^2} = \beta\frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i^2} = \beta
$$

(c) Notice that $\hat{\beta}_{MLE}$ is a linear combination of normal random variables. Therefore, $\hat{\beta}$ is also normally distributed. This distribution is specified by its mean and variance. In part b we saw the mean was β. So all we must do is derive the varince.

$$
\mathbb{V}(\hat{\beta}_{MLE}) = \mathbb{V}\left(\frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}\right) = \frac{\sum_{i=1}^{n} x_i^2 \mathbb{V}(y_i)}{\left(\sum_{i=1}^{n} x_i^2\right)^2} = \sigma^2\frac{\sum_{i=1}^{n} x_i^2}{\left(\sum_{i=1}^{n} x_i^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}
$$

Therefore we see that

$$
\hat{\beta}_{MLE} \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}\right)
$$

6. (a) Consider the estimator of β, $\hat{\beta}' = \sum Y_i / \sum x_i$. Then

$$\mathbb{E}(\hat{\beta}') = \mathbb{E}\left(\frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i}\right) = \frac{\sum_{i=1}^{n} \mathbb{E}(Y_i)}{\sum_{i=1}^{n} x_i} = \beta \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i} = \beta$$

Therefore this estimator is unbiased.

(b) Now calculating the the variance to compare to the MLE

$$\mathbb{V}(\hat{\beta}') = \mathbb{V}\left(\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}\right) = \frac{\sum_{i=1}^{n} \mathbb{V}(y_i)}{(\sum_{i=1}^{n} x_i)^2} = \frac{n\sigma^2}{(\sum_{i=1}^{n} x_i)^2}$$

Now notice, in general, $(\sum x_i)^2 \leq \sum x_i^2$ so we see that $\frac{1}{(\sum x_i)^2} \geq \frac{1}{\sum x_i^2}$. This gives

$$\mathbb{V}(\hat{\beta}_{MLE}) = \frac{\sigma^2}{\sum x_i^2} \leq \frac{\sigma^2}{(\sum x_i)^2} < \frac{n\sigma^2}{(\sum x_i)^2} = \mathbb{V}(\hat{\beta}')$$

7. (a) Consider the estimator $\tilde{\beta} = \frac{1}{n} \sum Y_i / x_i$. Then we see that

$$\mathbb{E}(\tilde{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i} \mathbb{E}(Y_i) = \frac{1}{n} \sum_{i=1}^{n} \frac{\beta x_i}{x_i} = \frac{n\beta}{n} = \beta$$

Thus, $\tilde{\beta}$ is another unbiased estimate of β.

(b) We now can calculate the variance of the estimator as follows.

$$\mathbb{V}(\tilde{\beta}) = \frac{1}{n^2} \sum_{i=1}^{n} \frac{1}{x_i^2} \mathbb{V}(Y_i) = \frac{\sigma^2}{n^2} \sum_{i=1}^{n} \frac{1}{x_i^2}$$

Notice that this is just an arithmetic mean of the $1/x_i^2$ and the variance of $\tilde{\beta}$ is the harmonic mean of the $1/x_i^2$. So we see that

$$\mathbb{V}(\hat{\beta}_{MLE}) \leq \mathbb{V}(\hat{\beta}')$$

Now notice that we also have

$$\mathbb{V}(\hat{\beta}') = \frac{n\sigma^2}{(\sum x_i)^2} \leq \frac{\sigma^2}{n^2} \sum \frac{1}{x_i^2} = \mathbb{V}(\tilde{\beta})$$

This is due to the fact that by Jensen's Inequality we have $\frac{n}{(\sum x_i)^2} \leq \sum \frac{1}{n^2 x_i^2}$

8. (a) Let $a \in \mathbb{R}$. Then using the fact that $\overline{X}$ and S^2 are independent for this curved family we see

$$\mathbb{E}(a\overline{X} + (1-a)cS) = a\mathbb{E}(\overline{X}) + (1-a)\mathbb{E}(cS) = a\theta + \theta - a\theta = \theta$$

(b) Again using the fact that S^2 and $\overline{X}$ are independent we see that

$$Var(a\overline{X} + (1-a)cS) = a^2 Var(\overline{X}) + (1-a)^2 Var(cS)$$

$$= a^2 \frac{\theta^2}{n} + (1-a)^2 \left[\mathbb{E}((cS)^2) - (\mathbb{E}(cS))^2 \right]$$

$$= a^2 \frac{\theta^2}{n} + (1-a)^2 (c^2 \mathbb{E}(S^2) - \theta^2)$$

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

Now minimizing over a we see

$$\frac{\partial}{\partial a} a^2 \frac{\theta^2}{n} + (1-a)^2(c^2 - 1)\theta^2 = 2a\frac{\theta^2}{n} - 2(1-a)(c^2-1)\theta^2 \stackrel{set}{=} 0$$

$$\frac{a}{n} - (c^2 - 1) + a(c^2 - 1) = 0$$

$$a = \frac{c^2 - 1}{1/n + c^2 - 1}$$

Also notice that

$$\frac{\partial^2}{\partial^2 a^2} a^2 \frac{\theta^2}{n} + (1-a)^2(c^2 - 1)\theta^2 = 2\theta^2 (\frac{1}{n} + c^2 + 1) \geq 0$$

Therefore, $a = \frac{c^2-1}{1/n+c^2-1}$ is in fact a minimum.

(c) We've shown in class that $(\overline{X}, S^2)$ is a sufficent statistic for $N(\mu, \sigma^2)$. So letting $\mu = \theta$ and $\sigma^2 = \theta^2$ we see that $(\overline{X}, S^2)$ is certainly a sufficent statistic. But notice

$$\mathbb{E}(\overline{X} - cS) = \theta - \theta = 0$$

but $\overline{X} - cS$ need not be zero. Therefore, we see that $(\overline{X}, cS)$ is not complete.

9. (a) First note that $X \sim Pois(\lambda)$, $f_X(x) = e^{-\lambda}\lambda^x/x! = \exp\{x \log(\lambda) - \lambda - \log(x!)\}$. Therefore, the Poisson is an exponential family, with complete sufficent statistic $\sum x_i$ by Theorem 6.2.25 (take the open set to be $\mathbb{R}$). Now let $\phi(x) = x/n$. Then the statistic $\phi(\sum x_i)$ is unbiased for λ. Then by Theorem 7.3.23 is is the unique, best unbiased estimator of λ.

(b) We will show that $\mathbb{E}(S^2|\overline{X})$ and $\overline{X}$ are both unique best unbiased estimators for λ. This will imply the result. First notice that both S^2 and $\overline{X}$ are unbiased estimators for λ. Moreover since $\sum x_i$ is a complete sufficent statistic for λ, then so is $\overline{X} = \phi(\sum x_i)$ because $\phi(\cdot)$ is one to one. Therefore, by Theorem 7.3.17, we see that $\mathbb{E}(S^2|\overline{X}) = \lambda$ is a unqiue best estimator of λ. But by Theorem 7.3.23 with $\theta(\overline{X}) = \overline{X}$, we see that $\overline{X}$ is the unique best estimator of λ. Therefore, since both results are unique, we have

$$\mathbb{E}(S^2|\overline{X}) = \overline{X}$$

$$Var(S^2) = \mathbb{E}(Var(S^2|X)) + Var(\mathbb{E}(S^2|\overline{X}))$$
$$= \mathbb{E}(Var(S^2|X)) + Var(\overline{X})$$
$$> Var(\overline{X})$$

(c) We have two results from above. (1) Any statistic has greater variance than the CSS if they have the same expectation. (2) The conditional expected value of a statistic given a CSS is the CSS. Formally we have the following;

Let $T(X)$ be a CSS and $R(X)$ being another statistic, with $\mathbb{E}(T(X)) = \mathbb{E}(R(X))$ then

- $Var(T(X)) \leq Var(R(X))$
- $\mathbb{E}(R(X)|T(X)) = T(X)$