9.1. In this problem, Y_1, Y_2, Y_3 is an iid sample of size $n = 3$ from an exponential(θ) population distribution, where the population mean $\theta > 0$ is unknown. We first want to compute

$$\text{eff}(\widehat{\theta}_1 \text{ to } \widehat{\theta}_5) = \frac{V(\widehat{\theta}_5)}{V(\widehat{\theta}_1)} = \frac{V(\overline{Y})}{V(Y_1)}.$$

We have $V(Y_1) = \theta^2$, the population variance. Also,

$$V(\overline{Y}) = \frac{\theta^2}{n} = \frac{\theta^2}{3}.$$

Therefore,

$$\text{eff}(\widehat{\theta}_1 \text{ to } \widehat{\theta}_5) = \frac{V(\overline{Y})}{V(Y_1)} = \frac{\theta^2/3}{\theta^2} = \frac{1}{3}.$$

In other words, the sample mean $\overline{Y}$ is 3 times more efficient than Y_1 as an estimator of θ. We next want to find

$$\text{eff}(\widehat{\theta}_2 \text{ to } \widehat{\theta}_5) = \frac{V(\widehat{\theta}_5)}{V(\widehat{\theta}_2)} = \frac{V(\overline{Y})}{V(\frac{1}{2}(Y_1 + Y_2))}.$$

Note that $\widehat{\theta}_2 = \frac{1}{2}(Y_1 + Y_2)$ is the sample mean based on only the first $n = 2$ observations. Therefore,

$$V\left(\frac{1}{2}(Y_1 + Y_2)\right) = \frac{\theta^2}{2}.$$

Therefore,

$$\text{eff}(\widehat{\theta}_2 \text{ to } \widehat{\theta}_5) = \frac{\theta^2/3}{\theta^2/2} = \frac{2}{3}.$$

In other words, the sample mean $\overline{Y}$ is 1.5 times more efficient than $\widehat{\theta}_2 = \frac{1}{2}(Y_1 + Y_2)$ as an estimator of θ. Finally, we want to find

$$\text{eff}(\widehat{\theta}_3 \text{ to } \widehat{\theta}_5) = \frac{V(\widehat{\theta}_5)}{V(\widehat{\theta}_3)} = \frac{V(\overline{Y})}{V(\frac{1}{3}(Y_1 + 2Y_2))}.$$

Let's find the variance of $\widehat{\theta}_3 = \frac{1}{3}(Y_1 + 2Y_2)$, a weighted average of Y_1 and Y_2. We have

$$V\left(\frac{1}{3}(Y_1 + 2Y_2)\right) = \frac{1}{9}\left[V(Y_1) + 4V(Y_2) + 2\text{Cov}(Y_1, 2Y_2)\right] = \frac{1}{9}(\theta^2 + 4\theta^2) = \frac{5\theta^2}{9}.$$

Therefore,

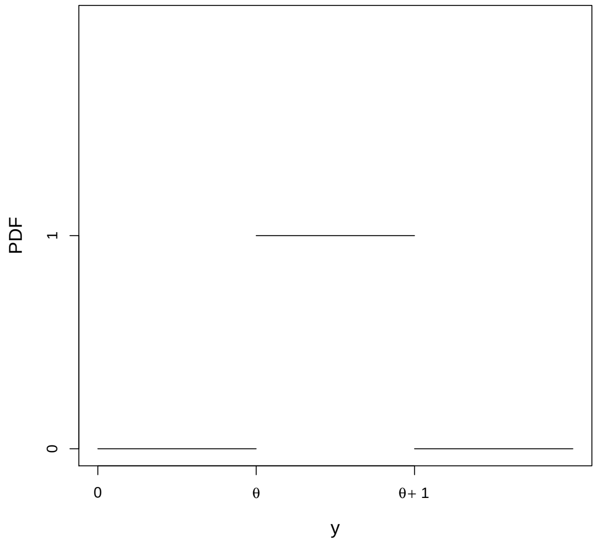
$$\text{eff}(\widehat{\theta}_3 \text{ to } \widehat{\theta}_5) = \frac{\theta^2/3}{5\theta^2/9} = \frac{3}{5}.$$

In other words, the sample mean $\overline{Y}$ is about 1.67 times more efficient than $\widehat{\theta}_3 = \frac{1}{3}(Y_1 + 2Y_2)$ as an estimator of θ.

9.3. In this problem, $Y_1, Y_2, ..., Y_n$ is an iid sample from a $\mathcal{U}(\theta, \theta + 1)$ population distribution, where the parameter θ is unknown. The population pdf is

$$f_Y(y) = \begin{cases} 1, & \theta < y < \theta + 1 \\ 0, & \text{otherwise.} \end{cases}$$

This pdf is shown at the top of the next page.



(a) We now show

$$\widehat{\theta}_1 = \overline{Y} - \frac{1}{2}$$

is an unbiased estimator of θ. The sample mean $\overline{Y}$ is always an unbiased estimator of the population mean, here,

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

Therefore,

$$E(\overline{Y}) = \theta + \frac{1}{2} \implies E\left(\overline{Y} - \frac{1}{2}\right) = \theta.$$

To show

$$\widehat{\theta}_2 = Y_{(n)} - \frac{n}{n+1}$$

is an unbiased estimator of θ, we need to derive the pdf of $Y_{(n)}$. Recall that in general,

$$f_{Y_{(n)}}(y) = n f_Y(y)[F_Y(y)]^{n-1},$$

where $F_Y(y)$ is the population cdf. We calculate

$$F_Y(y) = \begin{cases} 0, & y \leq \theta \\ y - \theta, & \theta < y < \theta + 1 \\ 1, & y \geq \theta + 1. \end{cases}$$

Therefore, for $\theta < y < \theta + 1$, the pdf of the maximum order statistic $Y_{(n)}$ is

$$f_{Y_{(n)}}(y) = n(1)(y - \theta)^{n-1} = n(y - \theta)^{n-1}.$$

Summarizing,

$$f_{Y_{(n)}}(y) = \begin{cases} n(y - \theta)^{n-1}, & \theta < y < \theta + 1 \\ 0, & \text{otherwise.} \end{cases}$$

Let's find $E(Y_{(n)})$. We have

$$E(Y_{(n)}) = \int_{\mathbb{R}} y f_{Y_{(n)}}(y)dy = \int_{\theta}^{\theta+1} ny(y-\theta)^{n-1}dy.$$

In the last integral, let

$$u = y - \theta \implies du = dy.$$

The limits change under this transformation; note $y : \theta \to \theta + 1$ implies $u : 0 \to 1$. Therefore,

$$\begin{aligned} E(Y_{(n)}) = \int_{\theta}^{\theta+1} ny(y-\theta)^{n-1}dy &= \int_0^1 n(u+\theta)u^{n-1}du \\ &= n\left[\int_0^1 (u^n + \theta u^{n-1})du\right] = n\left[\left(\frac{u^{n+1}}{n+1} + \frac{\theta u^n}{n}\right)\Bigg|_0^1\right] = \frac{n}{n+1} + \theta. \end{aligned}$$

Therefore,

$$E(Y_{(n)}) = \frac{n}{n+1} + \theta \implies E\left(Y_{(n)} - \frac{n}{n+1}\right) = \theta.$$

Therefore, both $\widehat{\theta}_1 = \overline{Y} - 1/2$ and $\widehat{\theta}_2 = Y_{(n)} - n/(n+1)$ are unbiased estimators of θ.

(b) Which point estimator is better? Let's calculate the variances of each one and form the efficiency; i.e.,

$$\text{eff}(\widehat{\theta}_1 \text{ to } \widehat{\theta}_2) = \frac{V(\widehat{\theta}_2)}{V(\widehat{\theta}_1)} = \frac{V(Y_{(n)} - \frac{n}{n+1})}{V(\overline{Y} - \frac{1}{2})}.$$

Note that

$$V\left(\overline{Y} - \frac{1}{2}\right) = V(\overline{Y}) = \frac{\sigma^2}{n} = \frac{[(\theta+1)-\theta]^2/12}{n} = \frac{1/12}{n} = \frac{1}{12n}.$$

Also,

$$V\left(Y_{(n)} - \frac{n}{n+1}\right) = V(Y_{(n)}) = E(Y_{(n)}^2) - [E(Y_{(n)})]^2 = E(Y_{(n)}^2) - \left(\frac{n}{n+1} + \theta\right)^2.$$

Note that

$$E(Y_{(n)}^2) = \int_{\mathbb{R}} y^2 f_{Y_{(n)}}(y)dy = \int_{\theta}^{\theta+1} ny^2(y-\theta)^{n-1}dy.$$

In the last integral, let

$$u = y - \theta \implies du = dy.$$

The limits change under this transformation; note $y : \theta \to \theta + 1$ implies $u : 0 \to 1$. Therefore,

$$\begin{aligned} E(Y_{(n)}^2) = \int_{\theta}^{\theta+1} ny^2(y-\theta)^{n-1}dy &= \int_0^1 n(u+\theta)^2 u^{n-1}du \\ &= n\int_0^1 (u^2 + 2\theta u + \theta^2)u^{n-1}du \\ &= n\int_0^1 (u^{n+1} + 2\theta u^n + \theta^2 u^{n-1})du \\ &= n\left[\left(\frac{u^{n+2}}{n+2} + \frac{2\theta u^{n+1}}{n+1} + \frac{\theta^2 u^n}{n}\right)\Bigg|_0^1\right] = \frac{n}{n+2} + \frac{2\theta n}{n+1} + \theta^2. \end{aligned}$$

Therefore,

$$V(Y_{(n)}) = E(Y_{(n)}^2) - [E(Y_{(n)})]^2 = \frac{n}{n+2} + \frac{2\theta n}{n+1} + \theta^2 - \left(\frac{n}{n+1} + \theta\right)^2 = \frac{n}{(n+1)^2(n+2)}.$$

Note that I did not show the algebra in the last step (about 6 lines worth). Finally, we have

$$\text{eff}(\widehat{\theta}_1 \text{ to } \widehat{\theta}_2) = \frac{V(\widehat{\theta}_2)}{V(\widehat{\theta}_1)} = \frac{V(Y_{(n)} - \frac{n}{n+1})}{V(\overline{Y} - \frac{1}{2})} = \frac{\frac{n}{(n+1)^2(n+2)}}{1/12n} = \frac{12n^2}{(n+1)^2(n+2)}.$$

I used R to calculate $\text{eff}(\widehat{\theta}_1 \text{ to } \widehat{\theta}_2)$ for $n = 2, 3, ..., 15$:

```
> n = seq(2,15,1)
> eff = 12*n^2/((n+1)^2*(n+2))
> cbind(n,eff)
       n       eff
 [1,]  2 1.3333333
 [2,]  3 1.3500000
 [3,]  4 1.2800000
 [4,]  5 1.1904762
 [5,]  6 1.1020408
 [6,]  7 1.0208333
 [7,]  8 0.9481481
 [8,]  9 0.8836364
 [9,] 10 0.8264463
[10,] 11 0.7756410
[11,] 12 0.7303466
[12,] 13 0.6897959
[13,] 14 0.6533333
[14,] 15 0.6204044
```

Therefore, $\widehat{\theta}_1 = \overline{Y} - 1/2$ is a more efficient estimator of θ when $n \leq 7$. On the other hand, $\widehat{\theta}_2 = Y_{(n)} - n/(n+1)$ is more efficient when $n \geq 8$.

9.7. In this problem, $Y_1, Y_2, ..., Y_n$ is an iid sample of size n from an exponential(θ) population distribution, where the population mean $\theta > 0$ is unknown. We want to compare the following estimators:

$$\begin{aligned} \widehat{\theta}_1 &= nY_{(1)} \\ \widehat{\theta}_2 &= \overline{Y}. \end{aligned}$$

Note that both estimators are unbiased. Recall if $Y_1, Y_2, ..., Y_n$ are iid exponential(θ), then $Y_{(1)} \sim \text{exponential}(\theta/n)$. Therefore,

$$E(\widehat{\theta}_1) = E(nY_{(1)}) = nE(Y_{(1)}) = n\left(\frac{\theta}{n}\right) = \theta.$$

Also, we know $E(\overline{Y}) = \theta$ because $\overline{Y}$ is always an unbiased estimator of the population mean.

Which point estimator is better? Let's calculate the variances of each one and form the efficiency; i.e.,

$$\text{eff}(\widehat{\theta}_1 \text{ to } \widehat{\theta}_2) = \frac{V(\widehat{\theta}_2)}{V(\widehat{\theta}_1)} = \frac{V(\overline{Y})}{V(nY_{(1)})}.$$

We know

$$V(\overline{Y}) = \frac{\theta^2}{n}.$$

Also, because $Y_{(1)} \sim \text{exponential}(\theta/n)$, we have

$$V(nY_{(1)}) = n^2 V(Y_{(1)}) = n^2 \left(\frac{\theta^2}{n^2}\right) = \theta^2.$$

Therefore,

$$\text{eff}(\widehat{\theta}_1 \text{ to } \widehat{\theta}_2) = \frac{V(\overline{Y})}{V(nY_{(1)})} = \frac{\theta^2/n}{\theta^2} = \frac{1}{n}.$$

This means the sample mean $\widehat{\theta}_2 = \overline{Y}$ is n times more efficient than $\widehat{\theta}_1 = nY_{(1)}$; i.e., $\overline{Y}$ is much better!

9.37. In this problem, $X_1, X_2, ..., X_n$ is an iid sample from a Bernoulli(p) population distribution, where $0 < p < 1$ is unknown. The population pmf is

$$p_X(x|p) = \begin{cases} p^x(1-p)^{1-x}, & x = 0, 1 \\ 0, & \text{otherwise.} \end{cases}$$

We will use the Factorization Theorem to show $T = T(X_1, X_2, ..., X_n) = \sum_{i=1}^n X_i$ is a sufficient statistic. The likelihood function is

$$\begin{aligned} L(p|\mathbf{x}) = \prod_{i=1}^n p_X(x_i|p) &= p_X(x_1|p) \times p_X(x_2|p) \times \cdots \times p_X(x_n|p) \\ &= p^{x_1}(1-p)^{1-x_1} \times p^{x_2}(1-p)^{1-x_2} \times \cdots \times p^{x_n}(1-p)^{1-x_n} \\ &= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}. \end{aligned}$$

Note that we can write the likelihood function as

$$L(p|\mathbf{x}) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} = \underbrace{\left(\frac{p}{1-p}\right)^{\sum_{i=1}^n x_i} (1-p)^n}_{g(t,p)} \times \underbrace{1}_{h(x_1,x_2,...,x_n)},$$

where $t = \sum_{i=1}^n x_i$. By the Factorization Theorem, it follows that $T = \sum_{i=1}^n X_i$ is a sufficient statistic for p.

9.38. In this problem, $Y_1, Y_2, ..., Y_n$ is an iid sample from a $\mathcal{N}(\mu, \sigma^2)$ population distribution. The likelihood function is

$$\begin{aligned} L(\mu, \sigma^2|\mathbf{y}) &= f_Y(y_1|\mu, \sigma^2) \times f_Y(y_2|\mu, \sigma^2) \times \cdots \times f_Y(y_n|\mu, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_1-\mu)^2} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_2-\mu)^2} \times \cdots \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_n-\mu)^2} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\mu)^2}. \end{aligned}$$

(a) If σ^2 is known, then there is only 1 unknown parameter, namely, the population mean μ. The likelihood function is

$$L(\mu|\mathbf{y}) = \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (y_i-\mu)^2}.$$

Note that I have adopted the notation $\sigma^2 = \sigma_0^2$ to emphasize the population variance σ^2 is known. Now, write

$$\sum_{i=1}^n (y_i - \mu)^2 = \sum_{i=1}^n (y_i^2 - 2\mu y_i + \mu^2) = \sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2.$$

Therefore,

$$\begin{aligned} L(\mu|\mathbf{y}) &= \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n e^{-\frac{1}{2\sigma_0^2}\left(\sum_{i=1}^n y_i^2 - 2\mu\sum_{i=1}^n y_i + n\mu^2\right)} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n y_i^2} e^{\frac{\mu}{\sigma_0^2}\sum_{i=1}^n y_i} e^{-\frac{n\mu^2}{2\sigma_0^2}} \\ &= \underbrace{e^{\frac{\mu}{\sigma_0^2}\sum_{i=1}^n y_i} e^{-\frac{n\mu^2}{2\sigma_0^2}}}_{g(t,\mu)} \times \underbrace{\left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n y_i^2}}_{h(y_1,y_2,\dots,y_n)}, \end{aligned}$$

where $t = \sum_{i=1}^n y_i$. By the Factorization Theorem, it follows that $T = \sum_{i=1}^n Y_i$ is a sufficient statistic for μ (when $\sigma^2 = \sigma_0^2$ is known).

Note: Because

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i = \frac{T}{n}$$

is a 1:1 function of T (i.e., it is a linear function of T), then $\overline{Y}$ is also a sufficient statistic (when $\sigma^2 = \sigma_0^2$ is known).

(b) If μ is known, then there is only 1 unknown parameter, namely, the population variance σ^2. The likelihood function is

$$L(\sigma^2|\mathbf{y}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\mu_0)^2}.$$

Note that I have adopted the notation $\mu = \mu_0$ to emphasize the population mean μ is known. Note we can write

$$L(\sigma^2|\mathbf{y}) = \underbrace{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\mu_0)^2}}_{g(t,\sigma^2)} \times \underbrace{1}_{h(y_1,y_2,\dots,y_n)},$$

where $t = \sum_{i=1}^n (y_i - \mu_0)^2$. By the Factorization Theorem, it follows that $T = \sum_{i=1}^n (Y_i - \mu_0)^2$ is a sufficient statistic for σ^2 (when $\mu = \mu_0$ is known).

(c) Now, we assume both parameters μ and σ^2 are unknown. The likelihood function is

$$\begin{aligned} L(\mu,\sigma^2|\mathbf{y}) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\mu)^2} \\ &= \underbrace{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n y_i^2} e^{\frac{\mu}{\sigma^2}\sum_{i=1}^n y_i} e^{-\frac{n\mu^2}{2\sigma^2}}}_{g(t_1,t_2,\mu,\sigma^2)} \times \underbrace{1}_{h(y_1,y_2,\ldots,y_n)}, \end{aligned}$$

where $t_1 = \sum_{i=1}^n y_i$ and $t_2 = \sum_{i=1}^n y_i^2$. By the Factorization Theorem (the multiparameter version), it follows that

$$\mathbf{T} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n Y_i^2 \end{pmatrix}$$

is a sufficient statistic for $\boldsymbol{\theta} = (\mu,\sigma^2)$.

Note: The statistics $\overline{Y}$ and S^2 (viewed together) are 1:1 functions of $\sum_{i=1}^n Y_i$ and $\sum_{i=1}^n Y_i^2$. Note that we can write

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2 = \frac{1}{n-1}\left(\sum_{i=1}^n Y_i^2 - n\overline{Y}^2\right).$$

In other words, if I know $\sum_{i=1}^n Y_i$ and $\sum_{i=1}^n Y_i^2$, then I can calculate $\overline{Y}$ and S^2 (and vice versa). Therefore,

$$\mathbf{T} = \begin{pmatrix} \overline{Y} \\ S^2 \end{pmatrix}$$

is also a sufficient statistic for $\boldsymbol{\theta} = (\mu,\sigma^2)$.

9.41. In this problem, $Y_1, Y_2, ..., Y_n$ is an iid sample from a Weibull(m,α) population distribution, where m is known and $\alpha > 0$ is unknown. The population pdf is

$$f_Y(y|\alpha) = \begin{cases} \frac{m}{\alpha} y^{m-1} e^{-y^m/\alpha}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

We will use the Factorization Theorem to show $T = T(Y_1, Y_2, ..., Y_n) = \sum_{i=1}^n Y_i^m$ is a sufficient statistic. The likelihood function is

$$\begin{aligned} L(\alpha|\mathbf{y}) = \prod_{i=1}^n f_Y(y_i|\alpha) &= f_Y(y_1|\alpha) \times f_Y(y_2|\alpha) \times \cdots \times f_Y(y_n|\alpha) \\ &= \frac{m}{\alpha} y_1^{m-1} e^{-y_1^m/\alpha} \times \frac{m}{\alpha} y_2^{m-1} e^{-y_2^m/\alpha} \times \cdots \times \frac{m}{\alpha} y_n^{m-1} e^{-y_n^m/\alpha} \\ &= \left(\frac{m}{\alpha}\right)^n \left(\prod_{i=1}^n y_i\right)^{m-1} e^{-\sum_{i=1}^n y_i^m/\alpha}. \end{aligned}$$

Note that we can write the likelihood function as

$$L(\alpha|\mathbf{y}) \;=\; \underbrace{\left(\frac{m}{\alpha}\right)^n e^{-\sum_{i=1}^n y_i^m/\alpha}}_{g(t,\alpha)} \;\times\; \underbrace{\left(\prod_{i=1}^n y_i\right)^{m-1}}_{h(y_1,y_2,...,y_n)},$$

where $t = \sum_{i=1}^n y_i^m$. By the Factorization Theorem, it follows that $T = \sum_{i=1}^n Y_i^m$ is a sufficient statistic for α. Note that this is only true when m is known (which we were told to assume). If m is unknown, then $T = \sum_{i=1}^n Y_i^m$ is not even a statistic.

9.42. In this problem, $Y_1, Y_2, ..., Y_n$ is an iid sample from a geometric(p) population distribution, where $0 < p < 1$ is unknown. The population pmf is

$$p_Y(y|p) = \begin{cases} (1-p)^{y-1}p, & y = 1,2,3,... \\ 0, & \text{otherwise.} \end{cases}$$

We will use the Factorization Theorem to show $T = T(Y_1, Y_2, ..., Y_n) = \sum_{i=1}^n Y_i$ is a sufficient statistic. The likelihood function is

$$\begin{aligned} L(p|\mathbf{y}) = \prod_{i=1}^n p_Y(y_i|p) &= p_Y(y_1|p) \times p_Y(y_2|p) \times \cdots \times p_Y(y_n|p) \\ &= (1-p)^{y_1-1}p \times (1-p)^{y_2-1}p \times \cdots \times (1-p)^{y_n-1}p \;=\; (1-p)^{\sum_{i=1}^n y_i - n}p^n. \end{aligned}$$

Note that we can write the likelihood function as

$$L(p|\mathbf{y}) \;=\; \underbrace{(1-p)^{\sum_{i=1}^n y_i - n}p^n}_{g(t,p)} \;\times\; \underbrace{1}_{h(y_1,y_2,...,y_n)},$$

where $t = \sum_{i=1}^n y_i$. By the Factorization Theorem, it follows that $T = \sum_{i=1}^n Y_i$ is a sufficient statistic for p.

9.44. In this problem, $Y_1, Y_2, ..., Y_n$ is an iid sample from a Pareto(α, β) population distribution, where β is known and α is unknown. The population pdf is

$$f_Y(y|\alpha) = \begin{cases} \dfrac{\alpha\beta^\alpha}{y^{\alpha+1}}, & y \geq \beta \\ 0, & \text{otherwise.} \end{cases}$$

Note that if β is known (an assumption), then there is only 1 population parameter; i.e., α. We will use the Factorization Theorem to show $T = T(Y_1, Y_2, ..., Y_n) = \prod_{i=1}^n Y_i$ is a sufficient statistic. The likelihood function is

$$\begin{aligned} L(\alpha|\mathbf{y}) = \prod_{i=1}^n f_Y(y_i|\alpha) &= f_Y(y_1|\alpha) \times f_Y(y_2|\alpha) \times \cdots \times f_Y(y_n|\alpha) \\ &= \frac{\alpha\beta^\alpha}{y_1^{\alpha+1}} \times \frac{\alpha\beta^\alpha}{y_2^{\alpha+1}} \times \cdots \times \frac{\alpha\beta^\alpha}{y_n^{\alpha+1}} \;=\; \frac{(\alpha\beta^\alpha)^n}{y_1^{\alpha+1}y_2^{\alpha+1}\cdots y_n^{\alpha+1}} \;=\; \frac{(\alpha\beta^\alpha)^n}{\left(\prod_{i=1}^n y_i\right)^{\alpha+1}}. \end{aligned}$$

Note that we can write the likelihood function as

$$L(\alpha|\mathbf{y}) = \frac{(\alpha\beta^{\alpha})^n}{(\prod_{i=1}^n y_i)^{\alpha+1}} = \underbrace{\frac{(\alpha\beta^{\alpha})^n}{(\prod_{i=1}^n y_i)^{\alpha}}}_{g(t,\alpha)} \times \underbrace{\frac{1}{\prod_{i=1}^n y_i}}_{h(y_1,y_2,\ldots,y_n)},$$

where $t = \prod_{i=1}^n y_i$. By the Factorization Theorem, it follows that $T = \prod_{i=1}^n Y_i$ is a sufficient statistic for α.

9.50. In this problem, $Y_1, Y_2, ..., Y_n$ is an iid sample from a $\mathcal{U}(\theta_1, \theta_2)$ population distribution, where both population parameters θ_1 and θ_2. The population pdf is

$$f_Y(y|\theta_1,\theta_2) = \begin{cases} \dfrac{1}{\theta_2-\theta_1}, & \theta_1 \le y \le \theta_2 \\ 0, & \text{otherwise.} \end{cases}$$

This pdf is shown at the top of the next page. Because the $\mathcal{U}(\theta_1, \theta_2)$ pdf is nonzero only when $\theta_1 \le y \le \theta_2$, let's write

$$f_Y(y|\theta_1,\theta_2) = \frac{1}{\theta_2-\theta_1}\, I(\theta_1 \le y \le \theta_2),$$

where $I(\cdot)$ is the **indicator function**; i.e.,

$$I(\theta_1 \le y \le \theta_2) = \begin{cases} 1, & \theta_1 \le y \le \theta_2 \\ 0, & \text{otherwise.} \end{cases}$$

The likelihood function is given by

$$\begin{aligned} L(\theta_1,\theta_2|\mathbf{y}) &= \prod_{i=1}^n f_Y(y_i|\theta_1,\theta_2) \\ &= f_Y(y_1|\theta_1,\theta_2) \times f_Y(y_2|\theta_1,\theta_2) \times \cdots \times f_Y(y_n|\theta_1,\theta_2) \\ &= \frac{1}{\theta_2-\theta_1}\, I(\theta_1 \le y_1 \le \theta_2) \times \frac{1}{\theta_2-\theta_1}\, I(\theta_1 \le y_2 \le \theta_2) \times \cdots \times \frac{1}{\theta_2-\theta_1}\, I(\theta_1 \le y_n \le \theta_2) \\ &= \left(\frac{1}{\theta_2-\theta_1}\right)^n \prod_{i=1}^n I(\theta_1 \le y_i \le \theta_2). \end{aligned}$$

A sufficient statistic is "hiding" in the

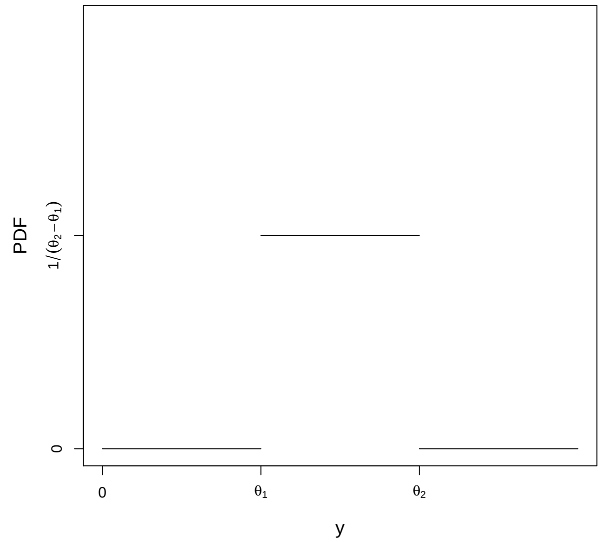
$$\prod_{i=1}^n I(\theta_1 \le y_i \le \theta_2)$$

term. To see why, note that

$$\prod_{i=1}^n I(\theta_1 \le y_i \le \theta_2) = 1 \iff I(\theta_1 \le y_{(1)} < y_{(n)} \le \theta_2) = 1.$$

Therefore, we can write the likelihood function as

$$\begin{aligned} L(\theta_1,\theta_2|\mathbf{y}) &= \left(\frac{1}{\theta_2-\theta_1}\right)^n I(\theta_1 \le y_{(1)} < y_{(n)} \le \theta_2) \\ &= \underbrace{\left(\frac{1}{\theta_2-\theta_1}\right)^n I(\theta_1 \le y_{(1)} < y_{(n)} \le \theta_2)}_{g(t_1,t_2,\theta_1,\theta_2)} \times \underbrace{1}_{h(y_1,y_2,\ldots,y_n)}, \end{aligned}$$



where $t_1 = y_{(1)}$ and $t_2 = y_{(n)}$. By the Factorization Theorem (the multiparameter version), it follows that

$$\mathbf{T} = \begin{pmatrix} Y_{(1)} \\ Y_{(n)} \end{pmatrix}$$

is a sufficient statistic for $\boldsymbol{\theta} = (\theta_1, \theta_2)$.

9.54. In this problem, $Y_1, Y_2, ..., Y_n$ is an iid sample from a power family population distribution, where both population parameters $\alpha > 0$ and $\theta > 0$ are unknown. The population pdf is

$$f_Y(y|\alpha,\theta) = \begin{cases} \dfrac{\alpha y^{\alpha-1}}{\theta^\alpha}, & 0 \le y \le \theta \\ 0, & \text{otherwise.} \end{cases}$$

Because $f_Y(y|\alpha,\theta)$ is nonzero only when $0 \le y \le \theta$, let's write

$$f_Y(y|\alpha,\theta) = \frac{\alpha y^{\alpha-1}}{\theta^\alpha}\, I(0 \le y \le \theta),$$

where

$$I(0 \le y \le \theta) = \begin{cases} 1, & 0 \le y \le \theta \\ 0, & \text{otherwise.} \end{cases}$$

The likelihood function is given by

$$\begin{aligned} L(\alpha,\theta|\mathbf{y}) &= \prod_{i=1}^n f_Y(y_i|\alpha,\theta) \\ &= f_Y(y_1|\alpha,\theta) \times f_Y(y_2|\alpha,\theta) \times \cdots \times f_Y(y_n|\alpha,\theta) \\ &= \frac{\alpha y_1^{\alpha-1}}{\theta^\alpha}\, I(0 \le y_1 \le \theta) \times \frac{\alpha y_2^{\alpha-1}}{\theta^\alpha}\, I(0 \le y_2 \le \theta) \times \cdots \times \frac{\alpha y_n^{\alpha-1}}{\theta^\alpha}\, I(0 \le y_n \le \theta) \\ &= \left(\frac{\alpha}{\theta^\alpha}\right)^n \left(\prod_{i=1}^n y_i\right)^{\alpha-1} \prod_{i=1}^n I(0 \le y_i \le \theta). \end{aligned}$$

Note that

$$\prod_{i=1}^{n} I(0 \leq y_i \leq \theta) \iff I(0 \leq y_{(n)} \leq \theta) = 1.$$

Therefore, we can write the likelihood function as

$$\begin{aligned} L(\alpha, \theta|\mathbf{y}) &= \left(\frac{\alpha}{\theta^\alpha}\right)^n \left(\prod_{i=1}^{n} y_i\right)^{\alpha-1} I(0 \leq y_{(n)} \leq \theta) \\ &= \underbrace{\left(\frac{\alpha}{\theta^\alpha}\right)^n \left(\prod_{i=1}^{n} y_i\right)^{\alpha-1} I(0 \leq y_{(n)} \leq \theta)}_{g(t_1, t_2, \alpha, \theta)} \times \underbrace{1}_{h(y_1, y_2, \ldots, y_n)}, \end{aligned}$$

where $t_1 = \prod_{i=1}^{n} y_i$ and $t_2 = y_{(n)}$. By the Factorization Theorem (the multiparameter version), it follows that

$$\mathbf{T} = \begin{pmatrix} \prod_{i=1}^{n} Y_i \\ Y_{(n)} \end{pmatrix}$$

is a sufficient statistic for $\boldsymbol{\theta} = (\alpha, \theta)$.