PROF HONG

Probability and Statistics
Chapter # 6

1.(#6.1) Consider an independent and identically distributed random sample $\mathbf{X}^n = (X_1, X_2, X_3)$, where X_i follows binary distribution with $P(X_i = 0) = P(X_i = 1) = \frac{1}{2}$, for $i = 1, 2, 3$. Define the sample mean $\bar{X}_n = \frac{1}{3}(X_1 + X_2 + X_3)$. Find: (a) the sampling distribution of $\bar{X}_n$; (b) the mean of $\bar{X}_n$; (c) the variance of $\bar{X}_n$.

Solution:

(a) To find the sampling distribution of $\bar{X}_n$, we can enumerate all outcomes and the corresponding probabilities since X_i simply follows binary distribution. The PMF of $\bar{X}_n$ is

$$f_{\bar{X}_n}(x) = \begin{cases} 1/8, \ x = 0 \\ 3/8, \ x = \frac{1}{3} \\ 3/8, \ x = \frac{2}{3} \\ 1/8, \ x = 1 \\ 0, \ \text{otherwise} \end{cases}$$

(b) mean of $\bar{X}_n = \frac{1}{8} \times 1 + \frac{3}{8} \times \frac{2}{3} + \frac{3}{8} \times \frac{1}{3} = \frac{1}{2}$. Or since the mean of sample mean is equal to the population mean, we have $E(\bar{X}_n) = E[\frac{1}{3}(X_1 + X_2 + X_3)] = E(X_i) = \frac{1}{2}$.

(c) The variance of $\bar{X}_n$ is $\sum f_{\bar{X}_n}(x)(x - E(\bar{X}_n))^2 = \frac{1}{8} \times (\frac{1}{2})^2 + \frac{3}{8} \times (\frac{1}{6})^2 + \frac{3}{8} \times (\frac{1}{6})^2 + \frac{1}{8} \times (\frac{1}{2})^2 = \frac{1}{12}$. Or since $Var(\bar{X}_n) = \frac{Var(X_i)}{3} = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}$.

2.(#6.2) A community has five families whose annual incomes are 1, 2, 3, 4, and 5 respectively. Suppose a survey is to be made to two of the five families and the choice of the two is random. Find the sampling distribution of the sample mean of the family income. Give your reasoning clearly.

Solution:
The population has been given and we just need to enumerate all the possible outcomes of

sample mean. Since there are $C_5^2 = 10$ ways of sampling. Then the possible outcomes are

$$\bar{X}_n = \begin{cases} 1.5, & \text{if 1 and 2 are chosen} \\ 2, & \text{if 1 and 3 are chosen} \\ 2.5, & \text{if 1 and 4 are chosen} \\ 3, & \text{if 1 and 5 are chosen} \\ 2.5, & \text{if 2 and 3 are chosen} \\ 3, & \text{if 2 and 4 are chosen} \\ 3.5, & \text{if 2 and 5 are chosen} \\ 3.5, & \text{if 3 and 4 are chosen} \\ 4, & \text{if 3 and 5 are chosen} \\ 4.5, & \text{if 4 and 5 are chosen} \end{cases}$$

After we collect all the possible outcomes and since each of them occurs with equal probability, we can calculate the PMF of $\bar{X}_n$ as

$$f_{\bar{X}_n}(x) = \begin{cases} 0.1, & x = 1.5 \\ 0.1, & x = 2 \\ 0.2, & x = 2.5 \\ 0.2, & x = 3 \\ 0.2, & x = 3.5 \\ 0.1, & x = 4 \\ 0.1, & x = 4.5 \\ 0, & \text{otherwise} \end{cases}$$

3.(#6.3) Suppose the return of asset i is given by

$$R_i = \alpha + \beta_i R_m + X_i, \qquad i = 1, ..., n,$$

where R_i is the return on asset i, α is the return on the risk-free asset, R_m is the return on the market portfolio which represents the market risk, and X_i represents an individual risk peculiar to the characteristics of asset i. Assume $0 < \beta_i < \infty$.

We consider an equal-weighting portfolio that consists of n assets. The return on such an equal-weighting portfolio is then given by

$$\begin{aligned} \bar{R}_n &= \sum_{i=1}^{n} \frac{1}{n} R_i \\ &= \alpha + \left(\frac{1}{n} \sum_{i=1}^{n} \beta_i \right) R_m + \bar{X}_n \\ &= \alpha + \bar{\beta} R_m + \bar{X}_n, \end{aligned}$$

where $\bar{\beta} = n^{-1}\Sigma_{i=1}^n \beta_i$ and $\bar{X}_n = n^{-1}\Sigma_{i=1}^n X_i$ is the sample mean of the random sample $\mathbf{X}^n = (X_1, \cdots, X_n)$. Assume the random sample $\mathbf{X}^n$ is an independent and identically distributed

random sample with population mean μ and population variance σ^2. Also, assume that R_m and $\mathbf{X}^n$ are mutually independent.
The total risk of the equal-weighting portfolio is measured by its variance.

(a) Show

$$\text{var}(\bar{R}_n) = \bar{\beta}^2 \text{var}(R_m) + \text{var}(\bar{X}_n).$$

That is, the risk of the portfolio contains a market risk and a component contributed by n individual risks;

(b) Show that individual risks can be eliminated by forming a portfolio with a large number of assets, that is, by letting $n \to \infty$.

Solution:

(a)

$$\begin{aligned} var(\overline{R}_n) &= var(\alpha + \overline{\beta} R_m + \overline{X}_n) \\ &= var(\overline{\beta} R_m + \overline{X}_n) \\ &= var(\overline{\beta} R_m) + var(\overline{X}_n) \quad \text{//Since } R_m, \overline{X}_n \text{ independent} \\ &= \overline{\beta}^2 var(R_m) + var(\overline{X}_n) \end{aligned}$$

(b)

$$\begin{aligned} var(\overline{X}_n) &= var(\frac{1}{n} \sum_{i=1}^{n} X_i) \\ &= \frac{1}{n^2} \sum_{i=1}^{n} var(X_i) \\ &= \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 \\ &= \frac{\sigma^2}{n} \to 0 \text{ when } n \to \infty \end{aligned}$$

4.(#6.5) Suppose $\mathbf{X}^n = (X_1, \cdots, X_n)$ is an IID $N(\mu_1, \sigma_1^2)$ random sample, $\mathbf{Y}^m = (Y_1, \cdots, Y_m)$ is an IID $N(\mu_2, \sigma_2^2)$ random sample, and the two random samples are mutually independent. Find the distribution of $\bar{X}_n - \bar{Y}_m$, where $\bar{X}_n$ and $\bar{Y}_m$ are the sample means of the first and second random samples respectively.

Solution:
Given $\{X_1, ..., X_n\}$ is an i.i.d. $N(\mu_1, \sigma_1^2)$ random sample, we know that the sample mean $\bar{X}_n$ follows $N(\mu_1, \frac{\sigma_1^2}{n})$. The logic is that $n\bar{X}_n = \sum_{i=1}^n X_i$ is a summation of n independently

distributed normal random variables. By reproductive property of normal random variable, we know $n\bar{X}_n$ follows $N(n\mu_1, n\sigma_1^2)$, then we have $\bar{X}_n$ follows $N(\mu_1, \frac{\sigma_1^2}{n})$. Be noted that the variance should be factored by $1/n^2$ rather than $1/n$. By the same logic, we have $\bar{Y}_m$ follows $N(\mu_2, \frac{\sigma_2^2}{m})$. Since $\bar{X}_n$ and $\bar{Y}_m$ are independent of each other, using the reproductive property again, we have $\bar{X}_n - \bar{Y}_m$ follows normal distribution with

$$E(\bar{X}_n - \bar{Y}_m) = E(\bar{X}_n) - E(\bar{Y}_m) = \mu_1 - \mu_2,$$

$$Var(\bar{X}_n - \bar{Y}_m) = Var(\bar{X}_n) + Var(\bar{Y}_m) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}.$$

5.(#6.6) Suppose $\mathbf{X}^n = (X_1, \cdots, X_n)$ is an IID $N(\mu, \sigma^2)$ random sample, $\mathbf{Y}^n = (Y_1, \cdots, Y_n)$ is an IID $N(\mu, \sigma^2)$ random sample, and the two random samples are mutually independent. Let $\bar{X}_n$ and $\bar{Y}_n$ be the sample means of the first and second random samples respectively, and let S_X^2 and S_Y^2 be the sample variances of the first and second random samples repectively. Find:

(a) the distribution of $(\bar{X}_n - \bar{Y}_n)/\sqrt{2\sigma^2/n}$;

(b) the distribution of $(\bar{X}_n - \bar{Y}_n)/\sqrt{2S_X^2/n}$;

(c) the distribution of $(\bar{X}_n - \bar{Y}_n)/\sqrt{2S_Y^2/n}$;

(d) the distribution of $(\bar{X}_n - \bar{Y}_n)/\sqrt{(S_X^2 + S_Y^2)/n}$;

(e) the distribution of $(\bar{X}_n - \bar{Y}_n)/\sqrt{S_n^2/n}$, where S_n^2 is the sample variance of the differenced sample $\mathbf{Z}^n = (Z_1, ..., Z_n)$, where $Z_i = X_i - Y_i, i = 1, 2, ..., n$.

Solution:

(a) First notice that $\overline{X}_n - \overline{Y}_n \sim N(0, \frac{2\sigma^2}{n})$. Then $(\overline{X}_n - \overline{Y}_n)/\sqrt{2\sigma^2/n} \sim N(0, 1)$.

(b) $(\overline{X}_n - \overline{Y}_n)/\sqrt{2S_X^2/n} = \frac{(\overline{X}_n - \overline{Y}_n)/\sqrt{2\sigma^2/n}}{\sqrt{(n-1)S_X^2/\sigma^2 \frac{1}{n-1}}} \sim t_{n-1}$ since $\overline{X}_n - \overline{Y}_n$ and S_X^2 are independent.

(c) same as part (b).

(d) $(\overline{X}_n - \overline{Y}_n)/\sqrt{(S_X^2 + S_Y^2)/n} \sim t_{2n-2}$.

(e) We can now regard the two sample as one Z^n and $\overline{X}_n - \overline{Y}_n = \overline{Z}_n$. Then $\overline{Z}_n/\sqrt{S_n^2/n} \sim t_{n-1}$

6.(#6.7) Let $\mathbf{X}^n = (X_1, \cdots, X_n)$ be an IID $N(\mu, \sigma^2)$ random sample. Find a function of S_n^2, the sample variance, say $g(S_n^2)$, that satisfies $E[g(S_n^2)] = \sigma$. (Hint: Try $g(S_n^2) = c\sqrt{S_n^2}$, where c is a constant.)

Solution:

Notice that

$$
\begin{aligned}
E(\sqrt{S_n^2}) &= E(\sqrt{\frac{\sigma^2}{n-1}}\sqrt{\frac{(n-1)S_n^2}{\sigma^2}}) \\
&= \sqrt{\frac{\sigma^2}{n-1}}E(\sqrt{\chi_{n-1}^2}) \\
&= \sqrt{\frac{\sigma^2}{n-1}}\int_0^\infty \frac{\sqrt{x}}{\Gamma((n-1)/2)2^{(n-1)/2}}x^{\frac{n-1}{2}-1}e^{-x/2}dx \\
&= \sqrt{\frac{\sigma^2}{n-1}}\frac{\Gamma(n/2)2^{n/2}}{\Gamma((n-1)/2)2^{(n-1)/2}}
\end{aligned}
$$

Then $c = \frac{\Gamma((n-1)/2)\sqrt{n-1}}{\sqrt{2}\Gamma(n/2)}$.

7.(#6.8) Establish the following recursion relations for sample means and sample variances. Let $\bar{X}_n$ and S_n^2 be the mean and variance, respectively, of $(X_1, \cdots, X_n)$. Then suppose another observation, X_{n+1}, becomes available. Show that

(a) $\bar{X}_{n+1} = \frac{X_{n+1}+n\bar{X}_n}{n+1}$;

(b) $nS_{n+1}^2 = (n-1)S_n^2 + \frac{n}{n+1}(X_{n+1} - \bar{X}_n)^2$.

Solution:

(a) $\overline{X}_{n+1} \cdot (n+1) = X_{n+1} + n \cdot \overline{X}_n$

(b) $nS_{n+1}^2 - (n-1)S_n^2 = \sum_{i=1}^{n+1}(X_i - \overline{X}_{n+1})^2 - \sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = RHS$ using part (a).

8. Suppose $X \sim \chi_p^2, Y \sim \chi_q^2$, and X and Y are independent. Show $X + Y \sim \chi_{p+q}^2$.

Solution:
Since $X \sim \chi_p^2$ and $Y \sim \chi_q^2$ are independent, by reproductive property of chi-square distribution we have

$$
\begin{aligned}
M_{X+Y}(t) &= M_X(t)M_Y(t) \\
&= (1-2t)^{-p/2}(1-2t)^{-q/2} \\
&= (1-2t)^{-(p+q)/2}
\end{aligned}
$$

It follows that $X + Y \sim \chi_{p+q}^2$

9.(#6.9)Suppose $(X_1, ..., X_n)$ is IID $N(0, \sigma^2)$. Consider the following estimator for σ^2 :

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Find: (a) the sampling distribution of $n\hat{\sigma}^2/\sigma^2$; (b)$E(\hat{\sigma}^2)$; (c)var$(\hat{\sigma}^2)$; (d)MSE$(\hat{\sigma}^2) = E(\hat{\sigma}^2 - \sigma^2)^2$. Give your reasoning.

Solution:

(a) $\frac{n\widehat{\sigma}^2}{\sigma^2} = \sum_{i=1}^n (\frac{X_i}{\sigma})^2$, since $\sum_{i=1}^n (\frac{X_i}{\sigma})^2 \sim \chi_1^2$,

$$\frac{n\widehat{\sigma}^2}{\sigma^2} = \sum_{i=1}^n (\frac{X_i}{\sigma})^2 \sim \chi_n^2$$

(b)

$$E(\widehat{\sigma}^2) = E[\frac{1}{n}\sum_{i=1}^n X_i^2] = \frac{1}{n}\sum_{i=1}^n E[X_i^2] = \frac{1}{n}\sum_{i=1}^n \sigma^2 = \sigma^2$$

(c) Since $\frac{n\widehat{\sigma}^2}{\sigma^2} \sim \chi_n^2$, var$(\frac{n\widehat{\sigma}^2}{\sigma^2})$=2n. Thus, var$(\widehat{\sigma}^2)$=$\frac{2\sigma^4}{n}$.

(d)

$$MSE(\widehat{\sigma}^2) = E[(\widehat{\sigma}^2 - \sigma^2)^2] = E[(\widehat{\sigma}^2 - E(\widehat{\sigma}^2))^2] = Var(\widehat{\sigma}^2) = \frac{2\sigma^4}{n}$$

10.(#6.10) Let $X_i, i = 1, 2, 3$, be independent with $N(i, i^2)$ distributions. For each of the following situations, use X_1, X_2, X_3 to construct a statistic with the indicated distribution.

(a) chi squared with 3 degrees of freedom;

(b) t distribution with 2 degrees of freedom;

(c) F distribution with 1 and 2 degrees of freedom.

Solution:

(a) Since Chi-square random variable can be constructed using the sum of independent squared standard normal random variable, we have $\sum_{i=1}^3 (\frac{X_i - i}{i})^2 \sim \chi_3^2$.

(b) Since Student-t random variable is defined as the ratio of standard normal random variable and the square root of Chi-square random variable over its degree of freedom, where the degree of freedom of t random variable is determined by that of the Chi-square random variable. Note the denominator and numerator should be independent of each other. Given $X_i's$ are independent of each other, we have $\frac{(\frac{X_i - i}{i})}{\sqrt{\sum_{j\neq i}(\frac{X_j - j}{j})/2}} \sim t_2$.

(c) Since Student-t random variable is defined as the ratio of two Chi-square random variables divided by their degrees of freedom respectively. Also, the denominator and numerator should be independent of each other, we have $\frac{(\frac{X_i - i}{i})^2}{\sum_{j\neq i}(\frac{X_j - j}{j})^2/2} \sim F_{1,2}$.

11.(#6.12) Show that for a Student's t_ν random variable X, (a) $E(X) = 0$; and (b) $\text{var}(X) = \nu/(\nu-2)$ for $\nu > 2$.

Solution:

(a) $E(X) = 0$ since the pdf is symmetric about 0.

(b)

$$\begin{aligned} var(X) &= E(X^2) \\ &= 2\int_0^\infty x^2 \frac{\nu^{\nu/2}}{2^{\nu/2}\Gamma(\nu/2)}(1+\frac{x^2}{\nu})^{-(\nu+1)/2}dx \\ &= 2\frac{\nu^{\nu/2}}{2^{\nu/2}\Gamma(\nu/2)}\int_0^\infty t^{3/2-1}(1+t)^{-3/2-(\nu/2-1)}dt \\ &= 2\frac{\nu^{\nu/2}}{2^{\nu/2}\Gamma(\nu/2)}Beta(3/2,\nu/2-1) \\ &= \frac{\nu}{\nu-2} \end{aligned}$$

12.(#6.15) Let X be one observation from $N(0,\sigma^2)$ population. Is $|X|$ a sufficient statistic?

Solution:
Yes, it is. By the factorization theorem, we have

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{|x|^2}{2\sigma^2}} \\ &= g(|x|,\sigma^2)\times 1, \end{aligned}$$

where we let $h(x) = 1$. Intuitively, that means if we have the information of $|x|$, then we have the information about σ^2 because the density $f_X(x)$ is symmetric about 0.

13.(#6.16) Let $X_1, ..., X_n$ be independent random variables with common densities

$$f_i(x,\theta) = \begin{cases} e^{i\theta-x} & x \geq i\theta, \\ 0 & x < i\theta. \end{cases}$$

Prove that $T = \min_{1\leq i\leq n}(X_i/i)$ is a sufficient statistic for θ.

Solution:
By independence, we have $f(x^n,\theta) = \prod_{i=1}^n f(x_i,\theta) = \prod_{i=1}^n e^{i\theta-x_i}$ if $x_i/i \geq \theta$, and 0, otherwise. Thus, we can see that $f(x^n,\theta) = \prod_{i=1}^n e^{i\theta-x_i}$ iff all x_i/i is greater than or equal to θ, this is

equivalent to $Tn = T(x^n) \geq \theta$. Thus we have

$$\begin{aligned}
f(x^n,\theta) &= \prod_{i=1}^{n} e^{i\theta - x_i} \\
&= \begin{cases} \prod_{i=1}^{n} e^{i\theta - x_i}, & \text{if } T(x^n) \geq \theta \\ 0, & \text{otherwise} \end{cases} \\
&= (\prod_{i=1}^{n} e^{i\theta - x_i})\mathbf{1}(T(x^n) \geq \theta) \\
&= e^{\frac{n(n+1}{2}}\mathbf{1}(T(x^n) \geq \theta) \times e^{-\sum_{i=1}^{n} x_i} \\
&= g[T(x^n),\theta]h(x^n),
\end{aligned}$$

where we let $e^{\frac{n(n+1}{2}}\mathbf{1}(T(x^n) \geq \theta) = g[T(x^n),\theta]$ and $e^{-\sum_{i=1}^{n} x_i} = h(x^n)$. Thus, by factorization theorem, $T(X^n)$ is a sufficient statistic for θ.

14.(#6.17) Prove the following theorem: Let $\mathbf{X}^n = (X_1, \cdots, X_n)$ be an IID random sample from a PDF or PMF $f(x,\theta)$ that belongs to an exponential family given by

$$f(x,\theta) = h(x)c(\theta)\exp\left[\sum_{j=1}^{k} w_j(\theta)t_j(x)\right],$$

where $\theta = (\theta_1, ..., \theta_d), d \leq k$. Then

$$T(\mathbf{X}^n) = \left[\sum_{i=1}^{n} t_1(X_i), \cdots, \sum_{i=1}^{n} t_k(X_i)\right]$$

is a sufficient statistic for θ.

Solution:
By independence, we have

$$\begin{aligned}
f(x^n,\theta) &= \prod_{i=1}^{n} f(x_i,\theta) \\
&= \prod_{i=1}^{n}(h(x_i)c(\theta)exp[\sum_{j=1}^{k} w_j(\theta)t_j(x_i)]) \\
&= (\prod_{i=1}^{n} h(x_i))c(\theta)^n \times exp[\sum_{i=1}^{n}\sum_{j=1}^{k} w_j(\theta)t_j(x_i)] \\
&= c(\theta)^n exp[\sum_{j=1}^{k} w_j(\theta)\sum_{i=1}^{n} t_j(x_i)] \times (\prod_{i=1}^{n} h(x_i)) \\
&= g[T(x^n),\theta]l(x^n),
\end{aligned}$$

where we let $c(\theta)^n exp[\sum_{j=1}^{k} w_j(\theta)\sum_{i=1}^{n} t_j(x_i)] = g[T(x^n),\theta]$ and $\prod_{i=1}^{n} h(x_i) = l(x^n)$. Thus, by factorization theorem, $T(X^n)$ is a sufficient statistic for θ.

15. Let $X_1, ..., X_n$ be a random sample from a $Gamma(\alpha, \beta)$ population.

(a) Suppose α is a known number. Find a sufficient statistic for β.

(b) Suppose α and β are unknown. Find a sufficient statistic for (α, β)

Solution:
For an iid random sample from Gamma(α, β) distribution, its marginal pdf is given by

$$f_{X_i}(x_i) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x_i^{\alpha-1} e^{-x_i/\beta}, & x_i > 0 \\ 0, & \text{otherwise} \end{cases}$$

(a) When α is known, then we have

$$\begin{aligned} f_{X^n}(x^n) &= \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)\beta^\alpha} x_i^{\alpha-1} e^{-x_i/\beta} \mathbf{1}(min_i x_i > 0) \\ &= \frac{1}{\beta^{n\alpha}} e^{-\sum_{i=1}^n x_i/\beta} \times \frac{1}{\Gamma(\alpha)^n} \prod_{i=1}^{n} x_i^{\alpha-1} \mathbf{1}(min_i x_i > 0) \\ &= g[T(x^n), \beta] h(x^n), \end{aligned}$$

where $g[T(x^n), \beta] = \frac{1}{\beta^{n\alpha}} e^{-\sum_{i=1}^n x_i/\beta}$ and $h(x^n) = \frac{1}{\Gamma(\alpha)^n} \prod_{i=1}^n x_i^{\alpha-1} \mathbf{1}(min_i x_i > 0)$. Then we have $T(x^n) = \sum_{i=1}^n x_i$ is a sufficient statistic for β by factorization theorem.

(b) When α and β are unknown, we have

$$\begin{aligned} f_{X^n}(x^n) &= \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)\beta^\alpha} x_i^{\alpha-1} e^{-x_i/\beta} \mathbf{1}(min_i x_i > 0) \\ &= \frac{1}{\beta^{n\alpha}} e^{-\sum_{i=1}^n x_i/\beta} \frac{1}{\Gamma(\alpha)^n} (\prod_{i=1}^{n} x_i)^{\alpha-1} \times \mathbf{1}(min_i x_i > 0) \\ &= g[T(x^n), \alpha, \beta] h(x^n), \end{aligned}$$

for $x_i > 0$, where $g[T(x^n), \alpha, \beta] = \frac{1}{\beta^{n\alpha}} e^{-\sum_{i=1}^n x_i/\beta} \frac{1}{\Gamma(\alpha)^n} (\prod_{i=1}^n x_i)^{\alpha-1}$ and $h(x^n) = \mathbf{1}(min_i x_i > 0)$. Then we have $T(x^n) = (\prod_{i=1}^n x_i, \sum_{i=1}^n x_i)$ is a sufficient statistic for (α, β) by factorization theorem.

16.(# 6.19) Let $X_1, ..., X_n$ be a random sample from a population with PDF

$$f(x, \theta) = \theta x^{\theta - 1}, 0 < x < 1, \theta > 0.$$

Is $\sum_{i=1}^n X_i$ sufficient for θ? Give your reasoning.

Solution:
To verify if $\sum_{i=1}^n$ is a sufficient statistic for θ, let's first write down the joint *pdf*.

$$\begin{aligned} f_{X^n}(x^n) &= \theta^n (\prod_{i=1}^{n} x_i)^{\theta-1} \mathbf{1}(min_i x_i > 0) \mathbf{1}(max_i x_i < 1) \\ &= g[T(x^n), \theta] h(x^n), \end{aligned}$$

where $g[T(x^n),\theta] = \theta^n(\prod_{i=1}^n x_i)^{\theta-1}$ and $h(x^n) = \mathbf{1}(min_i x_i > 0)\mathbf{1}(max_i x_i < 1)$. Thus $\sum_{i=1}^n X_i$ is **NOT** a sufficient statistic for θ.

17.(# 6.21(2)) Prove that the statistic $\sum_{i=1}^n X_i^2$ is minimal sufficient in the $N(\theta,\theta)$ family.

Solution:

$$\begin{aligned}
f(x^n,\theta) &= \prod_{i=1}^n f(x_i,\theta) \\
&= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x_i-\theta)^2}{2\theta}} \\
&= (2\pi\theta)^{\frac{n}{2}} \exp[-\sum_{i=1}^n (x_i-\theta)^2/(2\theta)] \\
&= (2\pi\theta)^{\frac{n}{2}} \exp[-\sum_{i=1}^n {x_i}^2/(2\theta) + \sum_{i=1}^n x_i - n\theta/2] \\
&= (2\pi\theta)^{\frac{n}{2}} \exp[-\sum_{i=1}^n {x_i}^2/(2\theta) - n\theta/2]\cdot \exp[\sum_{i=1}^n x_i]
\end{aligned}$$

By facterization theorem, $\sum_{i=1}^n {x_i}^2$ is a sufficient statistic. To check whether it is a minimum sufficient statistic, let x^n and y^n denotes two sample points in the sample space, then

$$\begin{aligned}
\frac{f(x^n,\theta)}{f(y^n,\theta)} &= \frac{(2\pi\theta)^{\frac{n}{2}} \exp[-\sum_{i=1}^n {x_i}^2/(2\theta) - n\theta/2]\cdot \exp[\sum_{i=1}^n x_i]}{(2\pi\theta)^{\frac{n}{2}} \exp[-\sum_{i=1}^n {y_i}^2/(2\theta) - n\theta/2]\cdot \exp[\sum_{i=1}^n y_i]} \\
&= \frac{\exp[\sum_{i=1}^n x_i]}{\exp[\sum_{i=1}^n y_i]}
\end{aligned}$$

if and only if $\sum_{i=1}^n {x_i}^2 = \sum_{i=1}^n {y_i}^2$. Thus, $\sum_{i=1}^n {x_i}^2$ is a minimum sufficient statistic for θ.

18.(# 6.26) Let $(X_1,\cdots,X_n)$ be an IID random sample from a $N(\theta,\theta)$ population, where θ is unknown. Find a sufficient statistic for θ and check if it is a minimal sufficient statistic.

Solution:

Given $(X_1,...,X_n)$ is an IID normal random sample, we can write down the joint *pdf* as

$$\begin{aligned}
f_{X^n}(x^n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta}} exp(\frac{-(x_i-\theta)^2}{2\theta}) \\
&= \frac{1}{(\sqrt{2\pi\theta})^n} exp(\frac{-\sum_{i=1}^n (x_i-\theta)^2}{2\theta}) \\
&= \frac{1}{(\sqrt{2\pi\theta})^n} exp(\frac{-\sum_{i=1}^n x_i^2 + 2\theta\sum_{i=1}^n x_i - n\theta^2}{2\theta}) \\
&= \frac{1}{(\sqrt{2\pi\theta})^n} exp(\frac{-\sum_{i=1}^n x_i^2}{2\theta} - \frac{n\theta}{2}) \times exp(\sum_{i=1}^n x_i) \\
&= g[T(x^n),\theta]h(x^n),
\end{aligned}$$

where $g[T(x^n),\theta] = \frac{1}{(\sqrt{2\pi\theta})^n}exp(\frac{-\sum_{i=1}^n x_i^2}{2\theta} - \frac{n\theta}{2})$ and $h(x^n) = exp(\sum_{i=1}^n x_i)$. By factorization theorem we know $T(X^n) = \sum_{i=1}^n X_i^2$ is a sufficient statistic for θ.
To show $T(X^n) = \sum_{i=1}^n X_i^2$ is a minimum sufficient statistic, let x^n and y^n denote two sample points in the sample space, then

$$\begin{aligned} \frac{f_{X^n}(x^n)}{f_{X^n}(y^n)} &= \frac{\frac{1}{(\sqrt{2\pi\theta})^n}exp(\frac{-\sum_{i=1}^n x_i^2}{2\theta} - \frac{n\theta}{2}) \times exp(\sum_{i=1}^n x_i)}{\frac{1}{(\sqrt{2\pi\theta})^n}exp(\frac{-\sum_{i=1}^n y_i^2}{2\theta} - \frac{n\theta}{2}) \times exp(\sum_{i=1}^n y_i)} \\ &= \frac{exp(\sum_{i=1}^n x_i)}{exp(\sum_{i=1}^n y_i)}, \end{aligned}$$

iff $T(X^n) = \sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i^2 = T(Y^n)$. Thus $\sum_{i=1}^n X_i^2$ is a minimum sufficient statistic for θ.