1. Let X_1 and X_2 be independent and identical exponential random variables with pdf

$$f_X(x|\beta) = \frac{1}{\beta}e^{-x/\beta}, \quad 0 < x < \infty, \quad 0 < \beta < \infty,$$

and let $X_{(1)}$ and $X_{(2)}$ are order statistics.

(a) Let

$$U_1 = \frac{2X_{(1)}}{X_1 + X_2} \text{ and } U_2 = X_1 + X_2.$$

Find the joint density function $f_{U_1,U_2}(u_1, u_2)$ using the Jacobian method. [Hint: $X_1 + X_2 = X_{(1)} + X_{(2)}$.]

Solution Rewrite $U_1 = 2X_{(1)}/(X_{(1)} + X_{(2)})$ and $U_2 = X_{(1)} + X_{(2)}$. $U_{(1)}$ and $U_{(2)}$ are functions of $X_{(1)}$ and $X_{(2)}$. The inverse function is $X_{(1)} = U_1U_2/2$ and $X_{(2)} = U_2 - U_1U_2/2$ with Jocobian

$$J = \begin{vmatrix} u_2/2 & u_1/2 \\ -u_2/2 & 1 - u_1/2 \end{vmatrix} = u_2/2.$$

The joint density function of $X_{(1)}$ and $X_{(2)}$ is $f_{X_{(1)},X_{(2)}}(x_1, x_2) = 2\beta^{-2}\exp(-x_1/\beta - x_2/\beta)I(0 < x_1 < x_2 < \infty)$. Hence the joint density function of $X_{(1)}$ and $X_{(2)}$ is

$$\begin{aligned} f_{U_1,U_2}(u_1, u_2) &= \beta^{-2}\exp\{(-u_1u_2/2 - u_2 + u_1u_2/2)/\beta\}u_2 I(0 < u_1u_2/2 < u_2 - u_1u_2/2 < \infty) \\ &= \beta^{-2}u_2\exp(-u_2/\beta)I(0 < u_1 < 1)I(0 < u_2 < \infty). \end{aligned}$$

(b) Show that U_1 and U_2 are independent.

Solution Since the joint density function $f_{U_1,U_2}(u_1, u_2)$ can be factored into a product of two separate functions of u_1 and u_2, we hence can claim U_1 and U_2 are independent. Specifically,

$$f_{U_1,U_2}(u_1, u_2) = f_{U_1}(u_1)f_{U_2}(u_2),$$

where $f_{U_1}(u_1) = I(0 < u_1 < 1)$ and $f_{U_2}(u_2) = \beta^{-2}u_2\exp(-u_2)I(0 < u_2 < \infty)$.

(c) Find the marginal pdf of U_1 and U_2, and find $E(U_1)$.

Solution $f_{U_1}(u_1)$ and $f_{U_2}(u_2)$ are marginal density functions for U_1 and U_2, respectively. U_1 follows $U(0, 1)$ and $E(U_1) = 1/2$.

(d) Show that U_2 is a complete sufficient statistic and U_1 is an ancillary statistic for β.

Solution Since X_1 and X_2 are iid and $f_X(x|\beta)$ belongs to an exponential family, $U_2 = X_1 + X_2$ (keep in mid that $U_2 = \sum_{i=1}^n X_i$, $n = 2$) is a complete sufficient statistic for β. $f_{U_1}(u_1)$ is not a function of β (or say, independent of β). Hence U_1 is an ancillary statistic for β.

(e) Use Basu's Theorem to find $E(U_1)$. Is it identical to the answer in (c)?

Solution Since $E(2X_{(1)}) = E(U_1U_2) = E(U_1)E(U_2)$, we have

$$E(U_1) = 2E(X_{(1)})/E(U_2) = 2(\beta/2)/(2\beta) = 1/2.$$

Here, $f_{X_{(1)}}(x_1) = 2\beta^{-1}\exp(-x/\beta)\exp(-x/\beta) = (\beta/2)^{-1}\exp\{-x/(\beta/2)\}$. $X_{(1)}$ follows $\text{Exp}(\beta/2)$ and $E(X_{(1)}) = \beta/2$. $E(U_2) = E(X_1+X_2) = E(X_1)+E(X_2) = 2\beta$.

2. Let $X_1, \ldots, X_n$ constitute a random sample from $N(0, \sigma^2)$ with probability density function

$$f_X(x|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}, \quad -\infty < x < \infty, \quad 0 < \sigma^2 < \infty.$$

(a) Show that the distribution belongs to an exponential family by identifying $h(x)$, $c(\sigma^2)$, $w(\sigma^2)$, and $t(x)$.

Solution $h(x) = 1$, $c(\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}$, $w(\sigma^2) = -\frac{1}{2\sigma^2}$, and $t(x) = x^2$.

(b) Show that $T = \sum_{i=1}^n X_i^2$ is a sufficient statistic for unknown $\theta = \sigma^r$, where r is a known positive integer.

Solution Since the distribution belongs to an exponential, $T = \sum_{i=1}^n X_i^2$ is a sufficient statistic for σ^2, as well as for $\sigma^r = (\sigma^2)^{r/2}$.

(c) What is the exact distribution of T/σ^2? Justify your answer.

Solution Since $T/\sigma^2 = \sum_{i=1}^n (X_i/\sigma)^2$ and X_i/σ follows $N(0, 1)$, $(X_i/\sigma)^2$ follows χ_1^2 and T/σ^2 follows χ_n^2.

(d) If a random variable Y follows $\text{Gamma}(\alpha = 2, \beta = n/2)$, then

$$E(Y) = \alpha\beta = n \quad \text{and} \quad E(Y^{r/2}) = \frac{\Gamma(n/2 + r/2)}{\Gamma(n/2)}2^{r/2}.$$

Develop an explicit expression for an unbiased estimator $\hat{\theta}$ that is a function of T. You need to show that $E\{\hat{\theta}(T)\} = \sigma^r$.

Solution Since T/σ^2 follows $\chi^2_n \equiv$Gamma(2,$n/2$), we have

$$E\{(T/\sigma^2)^{r/2}\} = \frac{\Gamma(n/2 + r/2)}{\Gamma(n/2)} 2^{r/2}.$$

That gives

$$E(T^{r/2}/\sigma^r) = \frac{\Gamma(n/2 + r/2)}{\Gamma(n/2)} 2^{r/2},$$

and

$$E\left\{(T/2)^{r/2} \frac{\Gamma(n/2)}{\Gamma(n/2 + r/2)}\right\} = \sigma^r.$$

3. Let $X_1, \ldots, X_n$ be a random sample from a normal distribution $N(\mu, 1)$.

   (a) Find the limiting distribution of $U_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu)$ by the central limit theorem.

   Solution The central limit theorem gives $\sqrt{n}(\sum_{i=1}^n X_i/n - \mu) \to_d N(0, 1)$. That makes $U_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \to_d N(0, 1)$.

   (b) Show that $V_n = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \to 1$ in probability by the weak law of large numbers.

   Solution By the weak law of large numbers,

   $$V_n = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \to_p E(X_1 - \mu)^2 = 1.$$

   We may also let $Y_i = (X_i - \mu)^2$ and Y_i follows χ^2_1. The weak law of large number also gives $V_n = \bar{Y} \to_p E(Y_1) = 1$.

   (c) Find the limiting distribution of $W_n = U_n/V_n$.

   Solution By Slutsky Theorem, $W_n \to_d N(0, 1)$.

   (d) Find the limiting distribution of $\sqrt{n}(\bar{X}^2 - \mu^2)$, where $\bar{X} = n^{-1} \sum_{i=1}^n X_i$.

   Solution We know that $\sqrt{n}(\bar{X} - \mu) \to_d N(0, 1)$. Let $g(\mu) = \mu^2$. By the delta method, $\sqrt{n}(g(\bar{X}) - g(\mu)) \to_d N(0, \{g'(\mu)\}^2) \equiv N(0, 4\mu^2)$.

(e) Construct a 95% confidence interval for μ^2, either under a finite n (exact) or $n \to \infty$ (limiting).

Solution According to the limiting distribution in (d),

$$0.95 = P\left(-1.96 < \frac{\sqrt{n}(\bar{X}^2 - \mu^2)}{2\mu} < 1.96\right).$$

We can search μ in a range such that the inequality within the bracket can be satisfied. However, it is quite complicate and uneasy. What can be much easier to obtain the confidence interval is that we replace μ in the denominator by $\bar{X}$. By Slutsky Theorem,

$$\frac{\sqrt{n}(\bar{X}^2 - \mu^2)}{2\bar{X}} \to_d N(0,1).$$

Hence,

$$0.95 = P\left(-1.96 < \frac{\sqrt{n}(\bar{X}^2 - \mu^2)}{2\bar{X}} < 1.96\right),$$

and

$$0.95 = P\left(\bar{X}^2 - \frac{1}{\sqrt{n}}1.96(2\bar{X}) < \mu^2 < \bar{X}^2 + \frac{1}{\sqrt{n}}1.96(2\bar{X})\right).$$

Since μ^2 is always positive, one may consider use

$$\left(\max\left\{\bar{X}^2 - \frac{1}{\sqrt{n}}1.96(2\bar{X}), 0\right\}, \bar{X}^2 + \frac{1}{\sqrt{n}}1.96(2\bar{X})\right)$$

as the 95% confidence interval for μ^2.

4. It has been known that Pareto distribution can be used to model the distribution of family incomes in certain population, where Pareto probability density function is defined by

$$f_Y(y) = \theta\gamma^\theta y^{-(\theta+1)}, \quad 0 < \gamma < y < \infty, \quad 0 < \theta < \infty.$$

with cumulative density function

$$F_Y(y) = 1 - \left(\frac{y}{\gamma}\right)^{-\theta}.$$

Let $Y_1, \ldots, Y_n$ constitute a random sample from this density function where γ is a **known** positive constant (minimal wage, for example) and where θ is an **unknown** parameter (known as Pareto index).

(a) Consider a random variable $T_n = \theta n(Y_{(1)} - \gamma)/\gamma$. Show that the survivor function of T_n is

$$P(T_n > t) = \left\{ \left(1 + \frac{t}{\theta n}\right)^n \right\}^{-\theta}.$$

Solution

$$\begin{aligned} P(T_n > t) &= P(\theta n(Y_{(1)} - \gamma)/\gamma > t) \\ &= P(Y_{(1)} > \gamma + \gamma t/(n\theta)) \\ &= P(Y_1 > \gamma + \gamma t/(n\theta)) \times \cdots \times P(Y_n > \gamma + \gamma t/(n\theta)) \\ &= \left(1 + \frac{t}{n\theta}\right)^{-n\theta}. \end{aligned}$$

(b) Given that $\lim_{n\to\infty}(1 + x/n)^n = e^x$ for some constant x, show that the limiting distribution of T_n follows exponential distribution with mean 1, i.e., $T_n \to_d T$ where $f_T(t) = e^{-t}$, $0 < t < \infty$.

Solution

$$\begin{aligned} \lim_{n\to\infty} P(T_n > t) &= \left\{ \lim_{n\to\infty} \left(1 + \frac{t}{\theta n}\right)^n \right\}^{-\theta} \\ &= (e^{t/\theta})^{-\theta} = e^{-t}. \end{aligned}$$

The limiting cumulative density function is $1 - e^{-t}$ and the limiting density function is $f_T(t) = e^{-t}$, which is the pdf of Exp(1).

(c) Using the limiting result in (b), find a 95% confidence interval for θ given that $P(T < 0.025) = 0.025$ and $P(T < 3.689) = 0.975$.

Solution Since the limiting distribution of T_n is Exp(1), one can have

$$\begin{aligned} 0.95 &= P(0.025 < T_n < 3.689) \\ &= P(0.025 < \frac{\theta n(Y_{(1)} - \gamma)}{\gamma} < 3.689) \\ &= P\left(\frac{0.025\gamma}{n(Y_{(1)} - \gamma)} < \theta < \frac{3.689\gamma}{n(Y_{(1)} - \gamma)}\right) \end{aligned}$$

(d) Comment on if the limiting property brings any advantage for the statistical inferences for θ. Can we use the result in (a) to construct a 95% confidence interval for θ? Explain.

Solution The 95% confidence interval is much easier to construct based on the limiting distribution. One may use (a) to construct the exact confidence interval but it is much more complicate. We will learn it when we talk about "pivotal quantity" in Chapter 9. Hence it is all right to say we cannot use the result in (a) to construct the confidence interval.