2016 May Theory - Solution

1. You have two components, whose lifetimes are X and Y, two independent Exponentially distributed random variables with mean λ. You put the two components in series, meaning that your system works as long as both components work. Then, your system stop working at some random time T, with $T = \min(X, Y)$.

   (a) Compute $P(X > t)$, for any $t \geq 0$.

   (b) Compute the p.d.f. of T.

   (c) Your remaining component has a remaining lifetime U (that is $U = \max(X, Y) - T$). Compute the joint p.d.f. of U and T, $f_{U,T}(u,t)$ for $t, u \geq 0$.

   (d) Are U and T independent?

2. Three random points A, B, and C, are chosen independently from the uniform distribution on a unit circle.

   (a) State the distribution of the minimal distance of A and B along the circle (i.e. the shorter arc length).

   (b) Give the minimal distance of A and B along the circle is x, find the probability that the center of the circle lies in the triangle ABC.

   (c) Find the probability that the center of the circle lies in the triangle ABC.

3. A coin having probability p of coming up heads is continually flipped until both heads and tails have appeared. Let X denote the total number of flips necessary.

   (a) What is the probability that the last flip lands Head?

   (b) Conditioning on the first flip being a head, what is the distribution of $X - 1$?

   (c) Compute $\mathrm{E}(X)$.

4. Let $X_1, \cdots, X_n$ be independently and identically distributed random variables with pdf

$$f(x;\theta) = \begin{cases} \frac{2}{\pi\theta}\exp\{-\frac{x^2}{\pi\theta^2}\} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0, \end{cases}$$

   where $\theta > 0$ is an unknown parameter. Note that

$$E(X_1) = \theta,\ E(X_1^2) = \left(\frac{\pi}{2}\right)\theta^2,\ E(X_1^4) = \left(\frac{3\pi^2}{4}\right)\theta^4.$$

(a) Find the uniformly minimal variance unbiased estimator (UMVUE) for θ^2.

(b) Does the variance of the estimator in part (a) achieve the Cramér-Rao lower bound? Justify your answer.

(c) Derive the uniformly most powerful (UMP) size α test of the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta = \theta_1$, where θ_0 and θ_1 are given positive numbers with $\theta_1 > \theta_0$. Express the test in terms of $\sum_{i=1}^n X_i^2$ and a quantile of χ^2 distribution.

5. Suppose $X_1, \ldots, X_n$ are independent and identically distributed random variables with the common pdf

$$f(x|\theta) = e^{-(x-\theta)}, \quad x \geq \theta,$$

where θ is an unknown parameter. Let $X_{(1)} = \min_{1 \leq i \leq n} X_i$ and $X_{(n)} = \max_{1 \leq i \leq n} X_i$.

(a) Show that $X_{(1)}$ is sufficient for θ. Find the pdf and cdf of $X_{(1)}$.

(b) Show that $X_{(n)} - X_{(1)}$ is an ancillary statistic. Compute $E(X_{(1)})$.

(c) Consider a class of confidence intervals for θ in the form of $(X_{(1)} - a, X_{(1)} - b)$ with $a \geq b \geq 0$. Find the level $1 - \alpha$ confidence interval that has the minimum length.

6. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables with $P(X_1 = i) = p_i$ for $i = 1, 2, 3$, where $0 < p_i < 1$ and $\sum_{i=1}^3 p_i = 1$. Define

$$Y_i = \sum_{j=1}^n I(X_j = i)$$

for $i = 1, 2, 3$, where $I(X_j = i)$ is the indicator function that takes the value 1 if $X_j = i$ and the value 0 otherwise.

(a) Find the probability mass function (pmf) of Y_1. Find the joint pmf of (Y_1, Y_2).

(b) Are Y_1 and Y_2 independent? Justify your answer.

(c) Find the maximum likelihood estimator (MLE) $\widehat{p_1 p_2}$ for the product $p_1 p_2$,

(d) Find the uniformly minimum variance unbiased estimator (UMVUE) for $p_1 p_2$.

(e) Find the limiting distribution of $\sqrt{n}(\hat{p}_1 - p_1)$ as $n \to \infty$, where $\hat{p}_1 = Y_1/n$.

(f) Find the MLE's for p_1, p_2, and p_3 assuming that $p_1 = p_2$.

Solutions:

1. (a) $P(X > t) = e^{-t/\lambda}$.

   (b) By the definition of T, $P(T > t) = P(X > t, Y > t)) = e^{-2t/\lambda}$. So the the p.d.f. of T is $f_T(t) = \frac{2}{\lambda}e^{-2t/\lambda}$.

   (c) From the textbook, we can obtain the joint density function of $(X_{(1)}, X_{(2)})$ is $f_{X_{(1)},X_{(2)}}(x_1, x_2) = \frac{2}{\lambda^2}e^{-(x_1+x_2)/\lambda}$. Using the the linear transformation $U = X_{(2)} - X_{(1)}, T = X_{(1)}$, we can get the joint density function of (U, T) is $f_{U,T}(u, t) = \frac{2}{\lambda^2}e^{-u/\lambda}e^{-2t/\lambda}$.

   (d) Yes, they are independent.

2. (a) Uniform distribution on $[0, \pi]$.

   (b) Conditional on the distance between A and B is $0 \le x < \pi$, the probability that the center is in the triangle ABC is $x/2\pi$.

   (c) The (unconditional) probability of the event is $\int_0^\pi \frac{x}{2\pi}\frac{1}{\pi}dx = \frac{1}{4}$.

3. (a) $P(\text{last flip is Head}) = P(TH)+P(TTH)+P(TTTH)+\cdots = \sum_{j=1}^\infty q^j p = pq\frac{1}{1-q} = q$.

   (b) Let $H =$ first flip is Head, then $X - 1|H \sim Geometric(1 - p)$.

   (c)

$$
\begin{aligned}
\mathrm{E}(X) &= \mathrm{E}(X|H)P(H) + \mathrm{E}(X|H^c)P(H^c) \\
&= (1 + \frac{1}{1-p})p + (1 + \frac{1}{p})(1 - p) \\
&= 1 + \frac{p}{1-p} + \frac{1-p}{p}.
\end{aligned}
$$

4. (a) One-parameter full-rank exponential family. $Y = \sum_{i=1}^n X_i^2$ is complete and sufficient for θ. And $E(Y) = nE(X^2) = \frac{n\pi}{2}\theta^2$. By Rao-Blackwell theorem, the estimator $\hat{\theta}_1 = \frac{2Y}{n\pi} = \frac{2\sum_{i=1}^n X_i^2}{n\pi}$ is UMVUE for θ^2. Alternatively, one can show that $\frac{2Y}{n}$ achieves the C-R bound (as shown in part (b)), so it is UMVUE.

   (b) Calculate

$$
\mathrm{Var}(\hat{\theta}_1) = \frac{4}{n^2\pi^2}\left(nVar(X^2)\right) = \frac{4}{n^2\pi^2}\frac{\pi^2\theta^4 n}{2} = \frac{2\theta^4}{n}.
$$

   Note that

$$
s_1(x, \theta) = \frac{\partial}{\partial\theta}\log f(x; \theta) = -\frac{1}{\theta} + \frac{2}{\pi\theta^3}x^2.
$$

Then

$$I_1(\theta) = \text{Var}(s_1) = \frac{4}{\pi^2\theta^6}\text{Var}(X^2).$$

Since $Var(X^2) = E(X^4) - [E(X^2)]^2 = \left(\frac{3\pi^2}{4}\right)\theta^4 - \left(\frac{\pi}{2}\right)^2\theta^4 = \frac{\pi^2\theta^4}{2}$, we have

$$I_1(\theta) = \frac{4}{\pi^2\theta^6}\frac{\pi^2}{2}\theta^4 = \frac{2}{\theta^2}.$$

So the Cramér lower variance bound for θ^2 is

$$\frac{(2\theta)^2}{I_n(\theta)} = \frac{(2\theta)^2}{2n/\theta^2} = \frac{2\theta^4}{n}.$$

Yes, its variance achieves the Cramér-Rao lower bound. Alternatively, decompose the score function as

$$s_n(x,\theta) = \frac{\partial}{\partial\theta}\sum_{i=1}^{n}\log f(x_i;\theta) = -\frac{n}{\theta} + \frac{2}{\pi\theta^3}\sum_{i=1}^{n}x_i^2 = \frac{n}{\theta^3}\left(\frac{2\sum_{i=1}^{n}x_i^2}{n\pi} - \theta^2\right).$$

(c) Using the NP lemma, the uniformly most powerful (UMP) size α test of the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta = \theta_1$ is: reject H_0 if $\frac{f(x;\theta_1)}{f(x;\theta_0)} > c$ for some c. Since

$$\frac{f(x;\theta_1)}{f(x;\theta_0)} = \frac{\theta_0^n}{\theta_1^n}\exp\{\frac{1}{\pi}\sum_{i=1}^{n}x_i^2\}$$

is increasing in x^2, the test is equivalent to: reject H_0 if $\sum_{i=1}^{n}X_i^2 \geq c$. To get size α test, we need

$$\alpha = P(\sum_{i=1}^{n}X_i^2 \geq c|\theta=\theta_0) = P(Y \geq c|\theta=\theta_0) = P(\frac{2}{\pi\theta_0^2}Y \geq \frac{2}{\pi\theta_0^2}c) = P(\chi_n^2 \geq \frac{2}{\pi\theta_0^2}c),$$

therefore $\chi^2_{n,\alpha} = \frac{2c}{\pi\theta_0^2}$ and $c = \frac{\pi\theta_0^2\chi^2_{n,\alpha}}{2}$.

5. (a) The joint pdf of data is

$$f(\mathbf{x}|\theta) = e^{n\theta}e^{-\sum_{i=1}^{n}x_i}I(X_{(1)} \geq \theta),$$

where $X_{(1)} = \min_{1\leq i\leq n} X_i$. By factorization theorem, $X_{(1)}$ is a sufficient statistic for θ. The density of $X_{(1)}$ is $f_{X_{(1)}}(x|\theta) = ne^{-n(x-\theta)}I(x \geq \theta)$. The cdf is: if $x < \theta$, then $F_{X_{(1)}}(x) = 0$ if $x < \theta$; if $x \geq \theta$, then $F_{X_{(1)}}(x) = \int_\theta^x ne^{n(\theta-x)}dx = 1 - e^{n(\theta-x)}$.

(b) $f(x|\theta)$ is a location family. Define $Z_i = X_i - \theta$, then Z_i has the exponential distribution with mean 1. Note that $X_{(n)} - X_{(1)} = Z_{(n)} - Z_{(1)}$ whose distribution does not depend on θ, so $X_{(n)} - X_{(1)}$ is ancillary. Note $X_{(1)} - \theta \sim \exp(1/n)$, so $E(X_{(1)}) = \theta + \frac{1}{n}$.

(c) Since $P(X_{(1)} - a \le \theta \le X_{(1)} - b) = P(b + \theta \le X_{(1)} \le a + \theta) = F_{X_{(1)}}(a+\theta) - F_{X_{(1)}}(b+\theta) = e^{-nb} - e^{-na} = 1 - \alpha$, so we have $a = -\frac{1}{n}\log[e^{-nb} - (1-\alpha)]$. The length of CI is $a - b = -\frac{1}{n}\log[1 - (1-\alpha)e^{nb}]$, which is increasing in b and takes the minimum at $b = 0$. So the shortest length is given by $a = -\frac{1}{n}\log\alpha$ and $b = 0$, which leads to the CI $[X_{(1)} + \frac{1}{n}\log\alpha, X_{(1)}]$.

6. (a) The p.m.f. of Y_1 is

$$f_{Y_1}(y_1) = P(Y_1 = y_1) = \frac{n!}{y_1!(n-y_1)!} p_1^{y_1}(1-p_1)^{n-y_1}, \quad y_1 = 0, 1, \ldots, n.$$

The p.m.f. of (Y_1, Y_2) is

$$f_{Y_1,Y_2}(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2) = \frac{n!}{y_1!y_2!(n-y_1-y_2)!} p_1^{y_1} p_2^{y_2} (1-p_1-p_2)^{n-y_1-y_2}$$

for $y_1, y_2 = 0, 1, \ldots, n$ and $y_1 + y_2 \le n$.

(b) Y_1 and Y_2 are not independent, because $f_{Y_1,Y_2}(y_1,y_2) \ne f_{Y_1}(y_1) f_{Y_2}(y_2)$.

(c) The log likelihood function of (p_1, p_2) is

$$l(p_1, p_2) = Y_1 \log(p_1) + Y_2 \log(p_2) + (n - Y_1 - Y_2)\log(1 - p_1 - p_2).$$

Its first derivatives with respect to p_1 and p_2 are

$$\frac{\partial l}{\partial p_1} = \frac{Y_1}{p_1} - \frac{n - Y_1 - Y_2}{1 - p_1 - p_2},$$

$$\frac{\partial l}{\partial p_2} = \frac{Y_2}{p_2} - \frac{n - Y_1 - Y_2}{1 - p_1 - p_2}.$$

Letting these derivative equal zero, we will have

$$\frac{Y_1}{p_1} = \frac{Y_2}{p_2} = \frac{n - Y_1 - Y_2}{1 - p_1 - p_2}.$$

This will give $\hat{p}_1 = \frac{Y_1}{n}$ and $\hat{p}_2 = \frac{Y_2}{n}$. So the MLE of the product is $\frac{Y_1}{n}\frac{Y_2}{n} = \frac{Y_1 Y_2}{n^2}$.

(d) Note that

$$E(Y_1Y_2) = Cov(Y_1, Y_2) + E(Y_1)E(Y_2) = -np_1p_2 + (np_1)(np_2) = (n^2 - n)p_1p_2 = n(n-1)p_1p_2.$$

Alternatively,

$$\begin{aligned} E(Y_1Y_2) &= E\left(\sum_{j=1}^n I(X_j = 1) \sum_{k=1}^n I(X_k = 2)\right) = \sum_{j=1}^n \sum_{k=1}^n E\left(I(X_j = 1)I(X_k = 2)\right) \\ &= \sum_{j\neq k} E\left(I(X_j = 1)I(X_k = 2)\right) = n(n-1)P(X_j = 1)P(X_k = 2) = n(n-1)p_1p_2. \end{aligned}$$

So $Y_1Y_2/[n(n-1)]$ is unbiased for p_1p_2. Note that the joint p.m.f of X_i's are

$$\begin{aligned} f(x_1, \ldots, x_n) &= \prod_{i=1}^n p_1^{I(X_i=1)} p_2^{I(X_i=2)} p_3^{I(X_i=3)} \\ &= p_1^{\sum_{i=1}^n I(X_i=1)} p_2^{\sum_{i=1}^n I(X_i=2)} p_3^{\sum_{i=1}^n (X_i=3)} \\ &= p_1^{Y_1} p_2^{Y_2} p_3^{Y_3} = p_1^{Y_1} p_2^{Y_2} (1 - p_1 - p_2)^{n - Y_1 - Y - 2} \\ &= (1 - p_1 - p_2)^n \exp\{Y_1[\ln p_1 - \ln(1 - p_1 - p_2)] + Y_2[\ln p_2 - \ln(1 - p_1 - p_2)]\}, \end{aligned}$$

which is a full-rank two-parameter exponential family, so (Y_1, Y_2) are sufficient statistics. By Rao-Blackwell. the estimator $\frac{Y_1Y_2}{n(n-1)}$ is UMVUE.

(e) Note that $I(X_1 = 1), \ldots, I(X_n = 1)$ is i.i.d Bernoulli random variable with mean p_1. By CLT, we have $\sqrt{n}(\hat{p}_1 - p_1) \to N(0, p_1(1 - p_1))$ as $n \to \infty$.

(f) If $p_1 = p_2$, the log likelihood function becomes

$$l(p_1) = (Y_1 + Y_2)\log(p_1) + (n - Y_1 - Y_2)\log(1 - 2p_1).$$

Its first derivatives with respect to p_1 is $\tilde{p}_1 = \tilde{p}_2 = \frac{Y_1+Y_2}{2n}$ and $\tilde{p}_3 = 1 - 2\tilde{p}_2 = \frac{Y_3}{n}$.