Solution to Homework 1

Date: March 2, 2023 *Scribe: Jianliang He*

***Warning**: This note is only used as a reference solution for the homework, and the solution to each question is not unique. The solution may contain factual and/or typographic errors and comments and criticism are kindly welcomed.*

Problem 1 *A manufacturer wants to market a new brand of heat-resistant tiles which may be used on the space shuttle. A random sample of m of these tiles is put on a test and the heat resistance capacities of the tiles are measured. Let $X_{(1)}$ denote the smallest of these measurements. The manufacturer is interested in finding the probability that in a future test (performed by, say, an independent agency) of a random sample of n of these tiles, at least k, $k = 1, \dots n$ will have a heat resistance capacity exceeding $X_{(1)}$ units. Assume that the heat resistance capacities of these tiles follows a continuous distribution with cdf F. Show that the probability of interest is given by $\sum_{r=k}^{n} P(r)$, where $P(r) = \frac{mn!(r+m-1)!}{r!(m+n)!}$.*

Proof: As $\{X_i\}_{i=1}^m$ and $\{Y_i\}_{i=1}^n$ are iid samples from distribution F, the probability of any ordering for the grouped samples $\{X_1, \dots, X_m, Y_1, \dots, Y_n\}$ are equal. Denote $P(r)$ to be the probability that there exists exactly r samples from $\{Y_i\}_{i=1}^n$ larger than $X_{(1)}$ in the ordered samples,

$$P(r) = \frac{\binom{m+r-1}{r}}{\binom{m+n}{m}} = \frac{m \cdot n!(r+m-1)!}{r!(m+n)!},$$

as in the ordered sequence the first $n-r$ samples must be from $\{Y_i\}_{i=1}^n$ and the $(n-r+1)^{th}$ sample is $X_{(1)}$. As the manufacturer is interested in funding when $r \geq k$, sum up $P(r)$ from k to n and we will get the probability of interest as $\sum_{r=k}^{n} P(r)$. $\square$

Remark: The solution is not unique and it's also plausible to prove the result via calculating $\mathbb{P}(Y_{(n-k+1)} > X_{(1)})$ and $f_{X_{(1)}}$ using the distribution of order statistics and independence.

Problem 2 *Let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ be order statistics for a random sample from the exponential distribution $f_X(x) = e^{-x}$ for $x \geq 0$. 1. Show that $X_{(r)}$ and $X_{(s)} - X_{(r)}$ are independent for any $s > r$; 2. Find the distribution of $X_{(r+1)} - X_{(r)}$; 3. Show that $E(X_{(i)}) = \sum_{j=1}^{i} \frac{1}{n+1-j}$; 4. Interpret the significance of these results if the sample arose from a life test on n light bulbs with exponential lifetimes.*

Solution: Based on the property of order statistics,

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!}(1-e^{-x})^{r-1}(e^{-x})^{n-r+1}\mathbb{I}(x \geq 0),$$

and the joint pdf of $(X_{(r)}, X_{(s)})$ follows

$$f_{X_{(r)},X_{(s)}}(x,y) = \frac{n!}{(r-1)!(s-r-1)!(n-r)!}(1-e^{-x})^{r-1}(e^{-x}-e^{-y})^{s-r-1}(e^{-y})^{n-s+1}e^{-x},$$

where $x \leq y \leq 0$.

1. Let $U = X_{(s)} - X_{(r)}$ and $V = X_{(r)}$, the corresponding Jocobian determinant is $|J| = 1$.

$$f_{U,V}(u,v) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}(1-e^{-v})^{r-1}(e^{-v})^{n-r+1}(1-e^{-u})^{s-r-1}(e^{-u})^{n-s+1}, \tag{1.1}$$

where $u, v \geq 0$. Marginalize the joint pdf, the marginal pdf of $X_{(s)} - X_{(r)}$ is

$$f_{X_{(s)}-X_{(r)}}(u) = \frac{(n-r)!}{(s-r-1)!(n-s)!}(1-e^{-u})^{s-r-1}(e^{-u})^{n-s+1}\mathbb{I}(u \geq 0).$$

As $f_{X_{(s)}-X_{(r)},X_{(r)}}(u,v) = f_{X_{(s)}-X_{(r)}}(u)f_{X_{(r)}}(v)$, we have $X_{(s)} - X_{(r)} \perp\!\!\!\perp X_{(r)}$.

2. Based on the result in (1.1) and take $s = r+1$, we have

$$f_{X_{(r+1)}-X_{(r)}}(u) = (n-r)e^{-u(n-r)}\mathbb{I}(u \geq 0),$$

i.e $X_{(r+1)} - X_{(r)}$ follows exponential distribution with $\lambda = n - r$.

3. $\mathbb{E}(X_{(i)}) = \sum_{r=1}^{i-1} \mathbb{E}(X_{(r+1)} - X_{(r)}) + \mathbb{E}(X_{(1)}) = \sum_{r=1}^{i-1} \frac{1}{n-r} + \frac{1}{n} = \sum_{j=1}^{i} \frac{1}{n+1-j}$, where we use the fact that $X_{(r+1)} - X_{(r)} \sim \exp(n-r)$ and $X_{(1)} \sim \exp(n)$.

4. If the sample arose from a life test on n light bulbs with exponential lifetimes, the time interval between two consecutive ending of bulbs $X_{(r+1)} - X_{(r)}$ follows exponential distribution $\exp(n-r)$ and it's independent of its own life span $X_{(r)}$. The exponential distribution is memoryless in the way that "the past has no bearing on its future behavior" and the property will be discussed detailedly in the Stochastic Process course (especially *Poissoin Process*) this semester. □

Problem 3 *If X is a continuous random variable with pdf* $f_X(x) = 2(1-x) \quad 0 \le x \le 1$*, find the transformation* $Y = g(X)$ *such that* $Y \sim U[0,2]$.

Solution: Known that $X \sim F_X = (2x - x^2)\mathbb{I}_{[0,1]}(x)$ and $F_X(X) \sim [0,1]$ as

$$\mathbb{P}(F_X(X) \le y) = \mathbb{P}(X \le F_X^{-1}(y)) = y,$$

apparently $g(x) = 2F_X(x) = (4x - 2x^2)\mathbb{I}_{[0,1]}(x)$ is exactly what we want. □