Solutions for Stat 512 — Take home exam II

1. Let $Y_1, \dots, Y_n$ be independent Poisson random variables with means $\lambda_1, \dots, \lambda_n$ respectively. Find:

   a. Probability function of $U = \sum_{i=1}^n Y_i$. (Hint: Using mgf technique.) (10 pts)

   b. Conditional probability function of Y_1, given that $U = m$. (Take a short review of conditional probability in 511). (10 pts)

Solution:

a. For $Y_1, \dots, Y_n$, the mgf is $m_{Y_i}(t) = e^{\lambda_i(e^t-1)}$. Hence,

$$\begin{aligned} m_U(t) &= \prod_{i=1}^n m_{Y_i}(t) \\ &= \prod_{i=1}^n \left[e^{\lambda_i(e^t-1)} \right]^n \\ &= e^{(e^t-1)\sum_{i=1}^n \lambda_i} \end{aligned}$$

Hence U follows a Poisson distribution with parameter $\sum_{i=1}^n \lambda_i$.

b.

$$\begin{aligned} P(Y_1 = y_1 | U = m) &= P(Y_1 = y_1 \Big| \sum_{i=1}^n Y_i = m) \\ &= \frac{P(Y_1 = y_1, \sum_{i=1}^n Y_i = m)}{P(\sum_{i=1}^n Y_i = m)} \\ &= \frac{P(Y_1 = y_1, \sum_{i=2}^n Y_i = m - y_1)}{P(\sum_{i=1}^n Y_i = m)} \\ &= \frac{P(Y_1 = y_1)P(\sum_{i=2}^n Y_i = m - y_1)}{P(\sum_{i=1}^n Y_i = m)} \qquad \text{since } Y_1 \text{ are independent with } (Y_2, Y_3, \dots, Y_n). \end{aligned}$$

Now, from part (a), we know that $\sum_{i=2}^n Y_i$ follows Poisson $(\sum_{i=2}^n \lambda_i)$ and $\sum_{i=1}^n Y_i$ follows Poisson $(\sum_{i=1}^n \lambda_i)$.

Hence,

$$\frac{P(Y_1 = y_1)P(\sum_{i=2}^n Y_i = m - y_1)}{P(\sum_{i=1}^n Y_i = m)} = \frac{\frac{e^{-\lambda_1}\lambda_1^{y_1}}{y_1!} \cdot \frac{e^{-\sum_{i=2}^n \lambda_i} \sum_{i=2}^n \lambda_i^{(m-y_1)}}{(m-y_1)!}}{\frac{e^{-\sum_{i=1}^n \lambda_i}(\sum_{i=1}^n \lambda_i)^m}{m!}}$$

$$= \frac{m!}{y_1!(m-y_1)!} \cdot \frac{\lambda_1^{y_1}(\sum_{i=2}^n \lambda_i)^{(m-y_1)}}{(\sum_{i=1}^n \lambda_i)^m}$$

$$= \frac{m!}{y_1!(m-y_1)!} \cdot \left(\frac{\lambda_1}{\sum_{i=1}^n \lambda_i}\right)^{y_1} \left(1 - \frac{\lambda_1}{\sum_{i=1}^n \lambda_i}\right)^{m-y_1}$$

$$= \binom{m}{y_1} \left(\frac{\lambda_1}{\sum_{i=1}^n \lambda_i}\right)^{y_1} \left(1 - \frac{\lambda_1}{\sum_{i=1}^n \lambda_i}\right)^{m-y_1}$$

Hence, the conditional distribution follows a Binomial distribution with m trials and $p = \frac{\lambda_1}{\sum_{i=1}^n \lambda_i}$.

****

2. If $Y_1, \ldots, Y_n$ are independent, uniformly distributed random variables on the interval $[0, \theta]$.

a. Find joint density for $(Y_{(1)}, Y_{(n)})$. (10 pts)

b. Find joint density for (U_1, U_2) where $U_1 = \frac{Y_{(1)}}{Y_{(n)}}$ and $U_2 = Y_{(n)}$. (15 pts)

c. Are U_1 and U_2 independent? Briefly discuss your reason. (10 pts)

d. If $\theta = 1$, show that $Y_{(k)}$, the kth-order statistic, has a beta distribution. Identify α and β. (15 pts)

Solution:

a. Since $Y_1, \ldots, Y_n$ follow Unif$[0, \theta]$, the pdf and the cdf are:

$$f(y) = \frac{1}{\theta}, \qquad y \in [0, \theta]$$

$$F(y) = \int_0^\theta \frac{1}{\theta} dy = \frac{y}{\theta}, \qquad y \in [0, \theta]$$

Hence,

$$\begin{aligned} f_{Y_{(1)},Y_{(n)}}(y_1,y_n) &= \frac{n!}{0!(n-2)!0!}\left(\frac{1}{\theta}\right)^2\left(\frac{y_1}{\theta}\right)^0\left(\frac{y_n}{\theta}-\frac{y_1}{\theta}\right)^{n-2}\left(1-\frac{y_n}{\theta}\right)^{n-n} \\ &= n\cdot(n-1)\cdot\frac{1}{\theta^2}\cdot\frac{(y_n-y_1)^{n-2}}{\theta^{n-2}} \\ &= \frac{n(n-1)}{\theta^n}(y_n-y_1)^{n-2}, \qquad 0\le y_1\le y_n\le\theta \end{aligned}$$

b.

$$\begin{cases} U_1 = \dfrac{Y_{(1)}}{Y_{(n)}} \\ U_2 = Y_{(n)} \end{cases} \Longrightarrow \begin{cases} Y_{(1)} = U_1U_2 \\ Y_{(n)} = U_2 \end{cases} \Longrightarrow J = U_2$$

The support:

$$\begin{cases} 0\le u_1u_2\le u_2 \\ 0\le u_2\le\theta \end{cases} \Longrightarrow \begin{cases} 0\le u_1\le 1 \\ 0\le u_2\le\theta \end{cases}$$

Hence,

$$\begin{aligned} f_{U_1,U_2}(u_1,u_2) &= \frac{n(n-1)}{\theta^n}(u_2-u_1u_2)^{n-2}\cdot u_2 \\ &= \frac{n(n-1)}{\theta^n}u_2^{n-1}(1-u_1)^{n-2}, \quad 0\le u_1\le 1,\ ,0\le u_2\le\theta \end{aligned}$$

c. Since the joint density of U_1, U_2 can be written into two pieces which only depends on U_1 and U_2 separately, U_1 and U_2 are independent.

d. If $\theta = 1$,

$$\begin{aligned} f_{Y_{(k)}}(y) &= \frac{n!}{(k-1)!(n-k)!}\cdot 1\cdot y^{k-1}(1-y)^{n-k} \\ &= \frac{\Gamma(n-k+1+k)}{\Gamma(k)\Gamma(n-k+1)}y^{k-1}(1-y)^{n-k+1-1}, \qquad y\in[0,1] \end{aligned}$$

Hence, $Y_{(k)}$ follows Beta distribution with $\alpha = k$ and $\beta = n-k+1$.

****

3. Suppose that $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$ are independent random samples, with the variables X_i normally distributed with mean μ_1 and variances σ_1^2 and the variables Y_i normally distributed with mean μ_2 and variances σ_2^2. The difference between the sample means, $\overline{X}-\overline{Y}$, is then a linear combination of $m+n$ normally distributed random variables and, is itself normally distributed.

a. Find $E(\overline{X} - \overline{Y})$. (10 pts)

b. Find var$(\overline{X} - \overline{Y})$. (10 pts)

c. Suppose that $\sigma_1^2 = 2$ and $\sigma_2^2 = 2.5$, and $m = n$. Find the sample sizes so that $(\overline{X} - \overline{Y})$ will be within 1 unit of $(\mu_1 - \mu_2)$ with probability (at least) 0.95. (10 pts)

Solution:

a. $\overline{X} - \overline{Y}$ can be written as linear combinations of $X_i's$ and $Y_j's$ as following:

$$\overline{X} - \overline{Y} = \frac{1}{m}\sum_{i=1}^{m} X_i - \frac{1}{n}\sum_{j=1}^{n} Y_i = \frac{1}{m}X_1 + \ldots + \frac{1}{m}X_m - \frac{1}{n}Y_1 - \ldots - \frac{1}{n}Y_n$$

Recall that if $L = a_1Y_1 + a_2Y_2 + \cdots + a_nY_n$ where $a_i's$ are some constants, $Y_i \sim N(\mu_i, \sigma_i^2)$ and $Y_i's$ are independent, then $L \sim N(\sum a_i\mu_i, \sum a_i^2\sigma_i^2)$. Therefore,

$$\overline{X} - \overline{Y} \sim N(\sum_{i=1}^{m} \frac{1}{m}\mu_1 - \sum_{j=1}^{n} \frac{1}{n}\mu_2, \sum_{i=1}^{m} \frac{1}{m^2}\sigma_1^2 + \sum_{j=1}^{n} \frac{1}{n^2}\sigma_2^2) = N(\mu_1 - \mu_2, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n})$$

Hence, $E(\overline{X} - \overline{Y}) = \mu_1 - \mu_2$.

b. From part (a), we know that var$(\overline{X} - \overline{Y}) = \dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}$.

c. Let $U = \overline{X} - \overline{Y}$, then $U \sim N(\mu \equiv \mu_1 - \mu_2, \frac{4.5}{N})$, where $N = m = n$. The question now is to determine the value of N such that $P(\mu - 1 < U < \mu + 1) \geq 0.95$.

$$\begin{aligned}
& P(\mu - 1 < U < \mu + 1) \geq 0.95 \\
& \Rightarrow P\left(\frac{\mu - 1 - \mu}{\sqrt{4.5/N}} < \frac{U - \mu}{\sqrt{4.5/N}} < \frac{\mu + 1 - \mu}{\sqrt{4.5/N}}\right) \geq 0.95 \\
& \Rightarrow P\left(\frac{\mu - 1 - \mu}{\sqrt{4.5/N}} < Z < \frac{\mu + 1 - \mu}{\sqrt{4.5/N}}\right) \geq 0.95 \\
& \Rightarrow 1 - 2P\left(Z \leq -\frac{1}{\sqrt{4.5/N}}\right) \geq 0.95 \\
& \Rightarrow P\left(Z \leq -\frac{1}{\sqrt{4.5/N}}\right) \leq 0.025
\end{aligned}$$

By looking at the standard normal table or through R, we know that

$$-\frac{1}{\sqrt{4.5/N}} \leq -1.96 \Rightarrow N \geq (1.96 * \sqrt{4.5})^2 \Rightarrow N \geq 17.3 \Rightarrow N = 18.$$

****

Extra credit.

4. Let Y_1 and Y_2 be independent and uniformly distributed over the interval $(0, 1)$. Find $P(2Y_{(1)} < Y_{(2)})$.

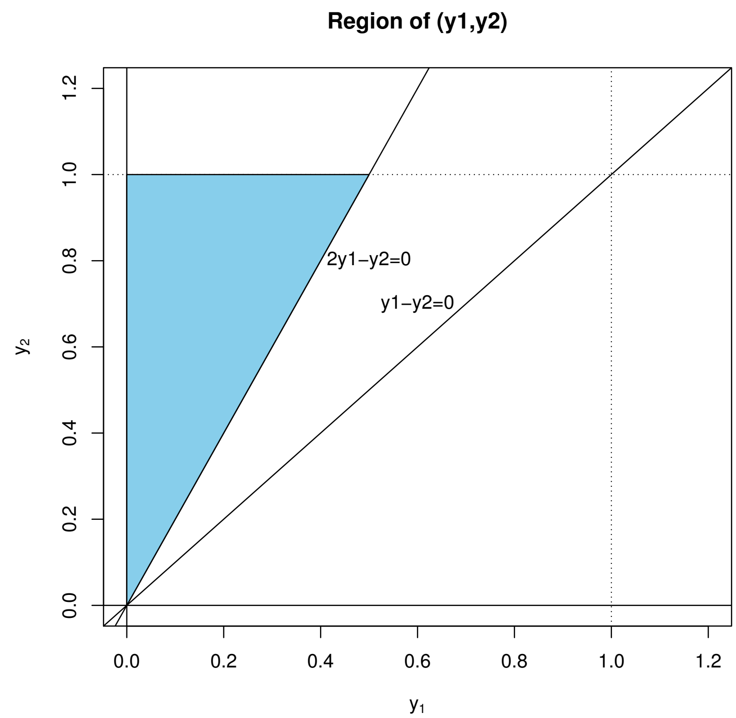
Solution:

First we find the joint distribution of $Y_{(1)}$ and $Y_{(2)}$, since $f(y) = 1$ and $F(y) = y$ for $y \in (0, 1)$, we have

$$f_{Y_{(1)},Y_{(2)}}(y_1, y_2) = \frac{2!}{0!0!0!}(y_1)^0(y_2 - y_1)^0(1 - y_2)^0 \cdot 1 \cdot 1 = 2, \qquad 0 < y_j < y_k < 1$$

Hence, for the regions in the figure below, we can find

$$P(2Y_{(1)} < Y_{(2)}) = \int_0^{0.5} \int_{2y_1}^{1} 2 \; dy_2 dy_1 = 0.5$$


Region of (y1,y2)
2y1−y2=0
y1−y2=0
y2
y1
0.0
0.2
0.4
0.6
0.8
1.0
1.2