Another homework solution is on the same website "homework solution"

4.3.1.2 # 1. Derive the GMLE of F using data from Example 1. Use approaches:

(1) use formula (3.3).

The data after being ordered is: $(1,1); (2,1); (2,0); (3,1)$. By using formula (3.3)

$$\hat{F}(Y_{(j)}) = \max_{i\le j} \min_{k\ge j} \frac{\sum_{i\le h\le k} \delta_{(h)}}{k-i+1},$$

we have below table.

i/k	1	2	3	4	$\min_{k\ge1}$	$\min_{k\ge2}$	$\min_{k\ge3}$	$\min_{k\ge4}$
1	1/1	2/2	2/3	3/4	2/3	2/3	2/3	3/4
2		1/1	1/2	2/3		1/2	1/2	2/3
3			0/1	1/2			0	1/2
4				1/1				1
				$\max_{i\le j} \min_{k\ge j}$	2/3	2/3	2/3	1

Therefore,

$Y_{(j)}$	1	2	2	3
$\hat{F}(Y_{(j)})$	2/3	2/3	2/3	1

(2) Derive directly from the likelihood function in (3.1) (see Remark: 9 boundaries).

The log-likelihood is

$$\mathcal{L}(\boldsymbol{s}) = \log s_1 + \log s_2 + \log(1-s_2) + \log s_3,\ 0 \le s_1 \le s_2 \le s_3 \le 1.$$

$$\begin{cases} \frac{\partial \mathcal{L}(\boldsymbol{s})}{\partial s_1} = \frac{1}{s_1} \\ \frac{\partial \mathcal{L}(\boldsymbol{s})}{\partial s_2} = \frac{1-2s_2}{s_2(1-s_2)} \\ \frac{\partial \mathcal{L}(\boldsymbol{s})}{\partial s_3} = \frac{1}{s_3} \end{cases} \Longrightarrow (s_1, s_2, s_3) = \left(1, \frac{1}{2}, 1\right),$$

which violates the constraint. Then we need to check the boundaries.

If $s_1 = 0$, then $(s_1, s_2, s_3) = (0, \frac{1}{2}, 1)$ with $\mathcal{L}(\boldsymbol{s}) = -\infty$.

If $s_3 = 1$, then $(s_1, s_2, s_3) = (1, \frac{1}{2}, 1)$ and violates the constraint.

If $s_1 = s_2$, then $\mathcal{L}(\boldsymbol{s}) = \log s_2 + \log s_2 + \log(1-s_2) + \log s_3 = 2\log s_2 + \log(1-s_2) + \log s_3$.

$s_1 = \frac{2}{2+1}$, $s_3 = \frac{1}{1}$.

$(s_1, s_2, s_3) = (\frac{2}{3}, \frac{2}{3}, 1)$ with $\mathcal{L}(\boldsymbol{s}) = \log \frac{4}{27}$.

If $s_2 = s_3$, then $(s_1, s_2, s_3) = (1, \frac{2}{3}, \frac{2}{3})$ and violates the constraint.

Therefore, $\left(\hat{F}(1), \hat{F}(2), \hat{F}(3)\right) = \left(\frac{2}{3}, \frac{2}{3}, 1\right)$.

Another way:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{s}) &= \log s_1 + \log s_2 + \log(1-s_2) + \log s_3, \ 0 \le s_1 \le s_2 \le s_3 \le 1 \\
&= \log p_1 + \log(p_1+p_2) + \log(p_3) + \log(p_1+p_2+p_3), && p_i \ge 0, \quad p_1+p_2+p_3 \le 1 \\
&\le \log q_1 + \log q_1 + \log q_3 + \log(q_1+q_3), && q_1 = p_1+p_2, \ q_2 = 0, q_3 = p_3 \\
&\le 2\log q_1 + \log q_3 \le && 2\log(2/3) + \log(1/3).
\end{aligned}
$$

That is $\hat{q}_1 = 2/3$, $\hat{q}_2 = 0$ and $\hat{q}_3 = 1/3$.

(3) Use R codes.

```r
C1 <- matrix(c(1, 2, 2, 3, 1, 1, 0, 1), 4)
tab <- table(C1[, 1], C1[, 2]); tab <- cbind(tab, rowSums(tab))
len <- nrow(tab); mat <- matrix(0, len, len)

for (i in 1:len) mat[i, i:len] <- cumsum(tab[i:len, 2])/cumsum(tab[i:len, 3])
sapply(1:len, function(i) max(rowMin(matrix(mat[1:i, i:len], i))))
```

```
[1] 0.6666667 0.6666667 1.0000000
```

4.3.1.2 # 2. Assuming Y is discrete with finitely many values, show that the estimator $\tilde{F}$ in (3.2) is a consistent estimator of $F_o(a_j)$ and find its asymptotic variance.

$$
\begin{aligned}
\lim_{n\to\infty} \tilde{F}(a_j) &= \lim_{n\to\infty} \frac{\sum_{i=1}^n \mathbf{1}_{(X_i \le Y_i = a_j)}}{\sum_{i=1}^n \mathbf{1}_{(Y_i = a_j)}} = \lim_{n\to\infty} \frac{\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{(X_i \le Y_i = a_j)}}{\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{(Y_i = a_j)}} \\
&= \frac{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{(X_i \le Y_i = a_j)}}{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{(Y_i = a_j)}} = \frac{\mathbb{E}\left[\mathbf{1}_{(X \le Y = a_j)}\right]}{\mathbb{E}\left[\mathbf{1}_{(Y = a_j)}\right]} \\
&= \frac{\mathbb{P}\left[X \le a_j\right]\mathbb{P}\left[Y = a_j\right]}{\mathbb{P}\left[Y = a_j\right]} = \mathbb{P}\left[X \le a_j\right] \\
&= F_o(a_j).
\end{aligned}
$$

By CLT, we have

$$
\sqrt{n}\left[\begin{pmatrix} \frac{1}{n}N_-(a_j) \\ \frac{1}{n}N(a_j) \end{pmatrix} - \begin{pmatrix} F_o(a_j)f_Y(a_j) \\ f_Y(a_j) \end{pmatrix}\right] \xrightarrow{\mathscr{D}} \mathcal{N}\left(\mathbf{0}, \Sigma\right),
$$

where

$$
\begin{aligned}
\Sigma_{11} &= \mathrm{Var}\left[\mathbf{1}_{(X \le Y = a_j)}\right] = \mathbb{E}\left[\mathbf{1}^2_{(X \le Y = a_j)}\right] - \mathbb{E}^2\left[\mathbf{1}_{(X \le Y = a_j)}\right] = F_o(a_j)f_Y(a_j) - F_o^2(a_j)f_Y^2(a_j), \\
\Sigma_{22} &= \mathrm{Var}\left[\mathbf{1}_{(Y = a_j)}\right] = \mathbb{E}\left[\mathbf{1}^2_{(Y = a_j)}\right] - \mathbb{E}^2\left[\mathbf{1}_{(Y = a_j)}\right] = f_Y(a_j) - f_Y^2(a_j), \\
\Sigma_{12} &= \Sigma_{21} = \mathrm{Cov}\left[\mathbf{1}_{(X \le Y = a_j)}, \mathbf{1}_{(Y = a_j)}\right] = \mathbb{E}\left[\mathbf{1}_{(X \le Y = a_j)}\mathbf{1}_{(Y = a_j)}\right] - \mathbb{E}\left[\mathbf{1}_{(X \le Y = a_j)}\right]\mathbb{E}\left[\mathbf{1}_{(Y = a_j)}\right] \\
&= F_o(a_j)f_Y(a_j) - F_o(a_j)f_Y^2(a_j).
\end{aligned}
$$

Now since $\tilde{F}(a_j) = N_-(a_j)/N(a_j)$, take $h(x,y) = \frac{x}{y}$ and hence $(\frac{\partial h}{\partial x}, \frac{\partial h}{\partial y}) = (\frac{1}{y}, -\frac{x}{y^2}) \neq 0$ at point $(F_o(a_j)f_Y(a_j), f_Y(a_j))$. Therefore, by Delta Method,

$$\sqrt{n}\left(\tilde{F}(a_j) - F_o(a_j)\right) \xrightarrow{\mathscr{D}} \mathcal{N}\left(0, \sigma^2\right),$$

where

$$\begin{aligned}\sigma^2 &= \frac{dh}{d(x,y)}\Sigma\frac{dh}{d(x,y)^T} = \begin{pmatrix}\frac{1}{f_Y(a_j)} & -\frac{F_o(a_j)}{f_Y(a_j)}\end{pmatrix}\Sigma\begin{pmatrix}\frac{1}{f_Y(a_j)} & -\frac{F_o(a_j)}{f_Y(a_j)}\end{pmatrix}^\top \\ &= \frac{F_o(a_j)S_o(a_j)}{f_Y(a_j)}.\end{aligned}$$

Teacher's comments on the solution of §4.3.1.3. #6. The student misunderstood the problem:

"Under the assumption in Problem # 4, generate a random sample of $n = 400$ C1 data from $\text{Exp}(\rho)$. Plot the survival curves of $S(t;\rho_o)$, the MLE and the GMLE on the same graph. Now pretend the data are RC data, plot the survival curves of $S(\cdot;\rho_o)$, the PLE and MLE curves. Make comments on the plots."

The problem means "pretend the data (Y_i, δ_i)'s are RC data, not (Z_i, δ_i)". Then the plot of PLE would be different from his plot. Also his last comment should be different,