§4.6.2. Homework 1. Consider the MIC model (2). Suppose that

X takes values 1, 3 and 5 w.p.1.; N takes values 0, 1 and 2 w.p.1.; $T = 7$, Y_1 takes values 2 and 4 w.p.1; $Y_2 = Y_1 + 2$.

1.a. Describe all possible observed intervals. 1.b. Derive all possible SCEs of F_X.

(**Hint**: If you properly group the observations, the estimators are similar to H_1 and H_2 in Example 1 or Homework problem 1 of §4.5.)

Sol. 1.

a. Describe all possible observed intervals.

When $N = 0$, the model is right-censored. Consider X takes values $1, 3, 5$, while $T = 7$, none of observation can be censored. Possible observations are $[1, 1], [3, 3], [5, 5]$.

When $N = 1$, Y_1 takes values $2, 4$. When $Y_1 = 2$, censored observations are $(-\infty, 2], (2, \infty)$. When $Y_1 = 4$, observations are $(-\infty, 4], (4, \infty)$.

When $N = 2$, (Y_1, Y_2) takes values $(2, 4), (4, 6)$. When $Y_1 = 2$, censored observations are $(-\infty, 2], (2, 4], (4, \infty)$. When $Y_1 = 4$, censored observations are $(-\infty, 4], (4, 6]$.

From above, possible observations are $[1, 1], [3, 3], [5, 5], (-\infty, 2], (2, \infty), (-\infty, 4], (4, \infty), (2, 4], (4, 6]$.

b. Derive all possible SCEs of F_X

The innermost intervals are $[1, 1], [3, 3], [5, 5]$. The SCEs are

$$\begin{cases} s_1 = \frac{1}{9}(\frac{3s_1}{s_1} + \frac{2s_1}{s_1+s_2}) \\ s_2 = \frac{1}{9}(\frac{2s_2}{s_2} + \frac{s_2}{s_2+s_3} + \frac{2s_2}{s_1+s_2}) \\ s_3 = \frac{1}{9}(\frac{4s_3}{s_3} + \frac{s_3}{s_2+s_3}) \end{cases}$$

Solving the equations, we have

$$\begin{cases} s_1 = \frac{43}{54} - \frac{\sqrt{337}}{54} \\ s_2 = -\frac{49}{216} + \frac{7\sqrt{337}}{216} \\ s_3 = \frac{55}{72} - \frac{\sqrt{337}}{72} \end{cases} \quad \text{or} \quad \begin{cases} s_1 = \frac{43}{54} + \frac{\sqrt{337}}{54} \\ s_2 = -\frac{49}{216} - \frac{7\sqrt{337}}{216} \\ s_3 = \frac{55}{72} + \frac{\sqrt{337}}{72} \end{cases}$$

Teacher: 1. Find the SCE of F_X, not $\mathbf{s}$. $\hat{F}_X$ can be derived from $\hat{\mathbf{s}}$.

2. The possible observations should be

N_1 [1,1] or $(-\infty, 2]$,

N_2 [5,5] or (4,6] or $(4, \infty)$,

N_3 [3,3] or (2,4]

N_4 $(-\infty, 4]$

N_5 $(4, \infty)$.

$n = N_1 + \cdots + N_5$, not just 12.

2. Suppose that $F \sim Exp(\rho), K = 2$ w.p.1., $Y_1 \sim Exp(\rho)$ and $Y_2 = Y_1 + Z$, where $Z \sim Exp(\rho)$ and $Z \perp Y_1$. Let S_n^2 be the sample variance of the GMLE $\hat{F}(2)$ based on 1000 simulations of random samples of size n. What do you expect S_{100}^2/S_{400}^2 to be? State your reasoning.

Expect the ratio to be $(4\frac{\ln 400}{\ln 100})^{\frac{1}{3}}$, which is the rate from conjecture. Since $Y_1 \sim Exp(\rho), Z \sim Exp(\rho), Y \perp Z$, we have $Y_1, Z \sim Gamma(1, \rho)$, then $Y_2 \sim Gamma(2, \rho)$. Y_1, Y_2 both have supprot $[0, \infty)$. Since $Y_2 - Y_1 \geq 0$, we can always take case where $Y_1 = Y_2$, which satisfy conjecture's assumption

$$P\{(Y_1, Y_2) \in N(x_0, x_0, \epsilon)\} > 0$$

Teacher: Expect the ratio of $\sqrt{S_{100}^2/S_{400}^2}$ ...

This case does not satisfy the assumption for Thm 2, since the assumption for Thm 2 assumes F_X is discrete distributed, while F_X is actually continuous. The case also does not satisfy Thm 3. Since $Y_2 - Y_1$ is exponential distributed, it has support $[0, \infty)$, which does not satisfy

$$P\{Y_2 - Y_1 < \epsilon_0\} = 0$$

for some ϵ_0.

3. Suppose that $F \sim Exp(\rho), K = 2$ w.p.1., Y_1 has a discrete uniform distribution on the set $\{1, 2, \cdots, 9\}$ and and $Y_2 = Y_1 + 1$. Let S_n^2 be the sample variance of the GMLE $\hat{F}(2)$ based on 1000 simulations of random samples of size n. What do you expect S_{100}^2/S_{400}^2 to be? State your reasoning.

It should be $4^{1/2}$, since X is discrete distribute, and F_X is strictly increasing on b_i's, which satisfy assumption of Thm 2.

Teacher: $\sqrt{S_{100}^2/S_{400}^2}$ should be $4^{1/2}$, since X is discrete distribute, and F_X is strictly increasing on b_i's, which satisfy assumption of Thm 2.

4. Try simulation to # 2 and # 3 with hw10.r

For 2, from the code result we can see the rate of standard errors is closest to $(4\frac{\ln 400}{\ln 100})^{\frac{1}{3}}$.

```
> (d1=s100/s400-2)
[1] -0.2519533
> (d2=s100/s400-4^(1/3))
[1] 0.1606456
> (d3=s100/s400-4^(1/3)*(log(400)/log(100))^(1/3))
[1] 0.01511102
```

For 3, from the result we can see the rate is closest to $(400/1000)^{1/2}$

```
> (d1=s100/s400-2)
[1] 0.1888394
> (d2=s100/s400-4^(1/3))
[1] 0.6014383
> (d3=s100/s400-4^(1/3)*(log(400)/log(100))^(1/3))
[1] 0.4559037
```