Solution to Homework

§4.4.2. Homework.

1. Suppose our data consist of 5 intervals: $(-\infty, 5]$, [2,2], (4,7], (3,8], $(9, +\infty)$. Find all the innermost intervals. Compute the GMLEs of $\mathbf{s}$ and F by two methods: directly from differentiation and by the SC algorithm. Verify that it is a solution of the self-consistent equation (4.2).

Sol. The innermost intervals are $[2, 2]$, $(4, 5]$ and $(9, +\infty)$ and hence we can write down the matrix

$$\Delta = (\delta_{ij}) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The SC equation is

$$s_j = \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_{ij} s_j}{\sum_{k=1}^{m} \delta_{ik} s_k}, \text{ where } n = 5, m = 3.$$

Therefore,

$$\begin{cases} 5 = \frac{1}{s_1} + \frac{1}{s_1+s_2} \\ 5 = \frac{1}{s_1+s_2} + \frac{2}{s_2} \\ 5 = \frac{1}{s_3} \end{cases} \Longrightarrow (s_1, s_2, s_3) = \left(\frac{4}{15}, \frac{8}{15}, \frac{1}{5}\right). \tag{1}$$

```
> library(survival)
> library(MASS)
> library(rms)
> t1 = c(-Inf,2,4,3,9)
> t2 = c(5,2,7,8,Inf)
> status=c(2,3,3,3,0)
> my.surv = Surv(time=t1,time2=t2,event=status,type="interval")
> (temp = survfit(my.surv 1))
> (tt=summary(temp))
```

time	*n.risk*	*n.event*	*survival*	*std.err*	*lower* 95%*CI*	*uppe* 95%*CI*
2.0	4	1	0.6	0.219	0.2933	1
5.5	3	2	0.2	0.179	0.0346	1

Remark. It is different from EQ.(1), due to it deals with C2 data only. If one replace status as follows, it leads to the correct answer to the SCE.

```
> status=c(3,3,3,3,3)
> my.surv = Surv(time=t1,time2=t2,event=status,type="interval")
> (temp = survfit(my.surv 1))
> (tt=summary(temp))
```

time	*n.risk*	*n.event*	*survival*	*std.err*	*lower* 95%*CI*	*uppe* 95%*CI*
2.0	5.00	1.33	0.733	0.1589	0.4109	1
4.5	3.67	2.67	0.200	0.0394	0.0291	1
Inf	1.00	1.00	0.000	0.0000	*NA*	*NA*

Then the results from both approaches are identical, and the GMLE of s is

$$\hat{s} = \left(\frac{4}{15}, \frac{8}{15}, \frac{1}{5}\right).$$

Correspondingly, a GMLE of F is

$$\hat{F}(t) = \sum_{A_j \subset (-\infty, t]} \hat{s}_j = \begin{cases} 0, & \text{if } t < 2 \\ \frac{4}{15}, & \text{if } 2 \le t < 5 \\ \frac{4}{5}, & \text{if } 5 \le t < +\infty \\ 1, & \text{if } t = \infty \end{cases}$$

When plugging in (4.2), we have

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_{ij}s_j}{\sum_{k=1}^{m}\delta_{ik}s_k} = \begin{cases} \frac{1}{5}\left(1+\frac{4/15}{4/15+8/15}\right) = \frac{4}{15} \\ \frac{1}{5}\left(\frac{8/15}{4/15+8/15}+1+1\right) = \frac{8}{15} \\ \frac{1}{5}\cdot 1 = \frac{1}{5} \end{cases}$$

Therefore, this is a solution of self-constraint equation (4.2).

Another way:

```
> require(interval)
> fit1=icfit(l,r)
> summary(fit1)
```

	Interval	*Probability*
1	$[2,2]$	0.2667
2	$(4,5]$	0.5333
3	$(9,Inf)$	0.2000

§4.3.1.3. #6. Under the assumption in Problem # 4, generate a random sample of $n = 400$ C1 data from Exp(ρ). Plot the survival curves of $S(t;\rho_o)$, the MLE and the GMLE on the same graph. Now pretend the data (Y_i,δ_i)'s are RC data, plot the survival curves of $S(\cdot;\rho_o)$, the PLE and MLE curves. Make comments on the plots.

Sol. C1 data MLE: can we use survreg() ?

GMLE use R codes "c1()"

RC data MLE ? $\overline{\delta}/\overline{Z}$, or survreg()

PLE ? survfit()

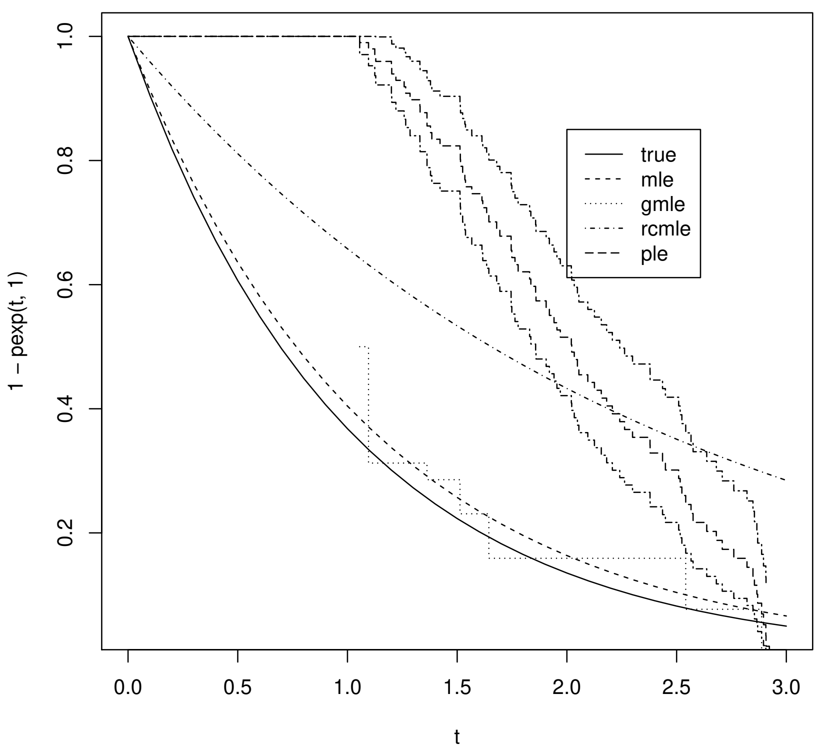
Let F_X be $Exp(\rho)$ (with mean $1/\rho$) and $Y \sim U(0,3/\rho)$.

The MLE of Exp(ρ) maximizes the likelihood

$$L(\rho) = \prod_{i=1}^{n}[(1-e^{-\rho Y_i})^{\delta_i}e^{-\rho Y_i(1-\delta_i)}]$$

```
library(survival)
n=100
x=rexp(n)
y=runif(n,1,3)
d=as.numeric(x<=y)
L=0 #Monte Carlo
M=0
N=100
for(i in 1:N) {
    gg=runif(1,0,4)
    a=prod(d*pexp(y,gg)+(1-d)*(1-pexp(y,gg)))
    if (L<a) {
    L=a
    M=gg
    }
}
x=c(y,d)
dim(x)=c(n,2)
c1=function(data) { # GMLE for C1 data
ord = order(data[, 1])
    data = data[ord, ]
    s = table(data[, 1], data[, 2])
    Nminus = s[, 2]
    N = s[, 1] + s[, 2]
```

```
    L = length(N)
    pt = unique(data[, 1])
    r = rep(0, L)
    mat = matrix(0, L, L)
    for(i in 1:L) {
    for(j in i:L)
    mat[i, j] = sum(Nminus[i:j])/sum(N[i:j]) }
    r[1] = min(mat[1, ])
    r[L] = max(mat[, L])
    for(i in 2:(L - 1))
    r[i] = max(apply(mat[1:i, i:L], 1, min))
    cbind(pt, r)
}
    zz=c1(x)
    t=(0:30)/10
    plot(t,1-pexp(t,1),type="l",lty=1,xlim=c(0,3))          # true
    lines(t,1-pexp(t,M),type="l",lty=2)          # mle
    lines(zz[,1],1-zz[,2],type="s",lty=3)          # gmle
    u=survfit(Surv(y,d)~1)          # pretend C1 data are RC data
    lines(u,lty=5)          # ple
    p=mean(d)/mean(y)
    lines(t,1-pexp(t,p),type="l",lty=4)          # rcmle
    leg.names¡-c("true","mle","gmle","rcmle","ple")
    legend(2, 0.85, leg.names, lty=c(1,2,3,4,5),cex=1.0)
```



What does the figure tell us ?