Homework 2.90 (p.103). 1. Compute the KME and Nelson-Aalen estimator of the survival function and the estimators of their standard deviations using Group 0 leukemia data.

Sol. Survival times: X_1, ..., X_n from $S_X(t)$.
If X_i's are all observed, the NPMLE of S_X is $\hat{S}_X(t) = \sum_i I(X_i > t)/n$.
If X_i's are right censored, that is, X_1, ..., X_n may not be observed, but one observed

RC data: (Z_1, δ_1), ..., (Z_n, δ_n).

the NPMLE of S_X is $\hat{S}_{pl}(t) = \prod_{t_k \le t}(1 - \frac{d_k}{r_k})$, where $t_1 < \cdots < t_m$ are distinct values of Z_i's with $\delta_i = 1$,

d_k is the number of person died at time t_k, and

r_k is the number of person at risk at time t_k $(= \sum_{i=1}^n I(Z_i \ge t_k))$.

An estimator of $\sigma^2_{\hat{S}_{pl}(t)}$ is

$$\hat{\sigma}^2_{\hat{S}_{pl}(t)} = \frac{1}{n}(\hat{S}_{pl}(t))^2 \sum_{k:\ t_k \le t} \frac{\hat{f}_{pl}(t_k)}{\hat{S}_Z(t_k-)\hat{S}_{pl}(t_k)}. \tag{1}$$

(1) The data are 6+, 6, 6, 6, 7, 9+, 10+, 10, 11+, 13, 16, 17+, 19+, 20+, 22, 23, 25+, 32+, 32+, 34+, 35+ (m=21),

Table 1. Calculation of KME

Exact remission Time	$(1 - \frac{d_k}{r_k})$	$\hat{S}_{pl}(a_k)$
6	18/21	18/21
7	16/17	$\frac{18}{21} \cdot \frac{16}{17}$
10	14/15	$\frac{18}{21} \cdot \frac{16}{17} \cdot \frac{14}{15}$
13	11/12	$\frac{18}{21} \cdot \frac{16}{17} \cdot \frac{14}{15} \cdot \frac{11}{12}$
16	10/11	$\frac{18}{21} \cdot \frac{16}{17} \cdot \frac{14}{15} \cdot \frac{10}{12}$
22	6/7	$\frac{6 \cdot 10 \cdot 14 \cdot 16 \cdot 18}{7 \cdot 12 \cdot 15 \cdot 17 \cdot 21}$
23	5/6	$\frac{5 \cdot 10 \cdot 14 \cdot 16 \cdot 18}{7 \cdot 12 \cdot 15 \cdot 17 \cdot 21}$

By (1), in order to compute $\hat{\sigma}_{\hat{S}_{pl}}$, we need

$$\hat{S}_{pl}(t) = \begin{cases} 1 & \text{if } t < 6, \\ 18/21 & \text{if } t \in [6, 7) \\ 96/119 & \text{if } t \in [7, 10) \\ \ldots & \ldots \end{cases}, \quad \hat{f}_{pl}(t) = \begin{cases} 1 - 18/21 & \text{if } t = 6 \\ 18/21 - 96/119 & \text{if } t = 7, \\ \ldots & \ldots \end{cases}$$

$$\hat{S}_Z(t) = \begin{cases} \frac{1}{n}\sum_i I(Z_i > t) & \\ 1 & \text{if } t < 6, \\ 17/21 & \text{if } t \in [6, 7) \\ 16/21 & \text{if } t \in [7, 9) \\ 15/21 & \text{if } t \in [9, 10) \\ \ldots & \ldots \end{cases} = \begin{cases} \prod_{t_k \le t}(1 - \frac{d_k}{r_k}) & \\ 1 & \text{if } t < 6, \\ \frac{17}{21} & \text{if } t \in [6, 7) \\ \frac{17}{21}\frac{16}{17} & \text{if } t \in [7, 9) \\ \frac{17}{21}\frac{16}{17}\frac{15}{16} & \text{if } t \in [9, 10) \\ \ldots & \ldots \end{cases}$$

$$\hat\sigma^2_{\hat S_{pl}(t)} = \frac{1}{n}(\hat S_{pl}(t))^2 \sum_{k:\ t_k\le t} \frac{\hat f_{pl}(t_k)}{\hat S_Z(t_k-)\hat S_{pl}(t_k)}$$
$$= \begin{cases} 0 & \text{if } t<6, \\ \frac{1}{21}(18/21)^2\frac{3/21}{1\times 18/21} & \text{if } t\in[6,7) \\ \frac{1}{21}(96/119)^2[\frac{3/21}{1\times 18/21}+\frac{18/21-96/119}{(17/21)\times 96/119}] & \text{if } t\in[7,10) \\ ... & ... \end{cases}$$

The other estimator is the **Nelson-Aalen estimator**: $\tilde S_{NA}(t) = e^{-\sum_{t_k\le t}\frac{d_k}{r_k}}$. Its variance can be estimated by $\hat\sigma^2_{\tilde S_{NA}(t)} = (\tilde S_{NA}(t))^2\hat\sigma^2_{H(t)}$, where $\hat\sigma^2_{H(t)} = \sum_{t_j\le t}\frac{(r_j-d_j)d_j}{(r_j-1)r_j^2}$.

$$\tilde S_{NA}(t) = \begin{cases} 1 & \text{if } t<6, \\ \exp(-3/21) & \text{if } t\in[6,7) \\ \exp(-3/21-1/17) & \text{if } t\in[7,10) \\ ... & ... \end{cases}$$

$$\hat\sigma^2_{\tilde S_{NA}(t)} = \begin{cases} 0 & \text{if } t<6, \\ \exp(-6/21)\frac{3*18}{20(21)^2} & \text{if } t\in[6,7) \\ \exp(-3/21-1/17)[\frac{3*18}{20(21)^2}+\frac{16*1}{16*17^2}] & \text{if } t\in[7,10) \\ ... & ... \end{cases}$$

2. Compute the MLE assuming $X\sim S_X(x)=e^{-x/\theta}I(x>0)$, the NPMLE and $\check F$ (see Eq. (1)) of the cdf using Group 1 leukemia data. Compute the estimators of their standard deviations.

x=c(1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23)

$\overline{X} = (1+1+\cdots+23)/21 = \frac{26}{3}$

The MLE of $S(t)$ is $\tilde S(t) = \exp(-t/\overline{X}) = \exp(-\frac{3t}{26})$ for $t>0$.

The NPMLE of $F(t)$ and $\check F$ are given in the textbook.

The NPMLE of $S(t)$ and $\check S$ are $\hat S(t) = 1-\hat F(t) = \begin{cases} 1 & \text{if } t<1 \\ & \cdots \end{cases}$ and

$\check S(t) = 1-\check F(t) = \begin{cases} 0 & \text{if } t<1 \\ & \cdots \end{cases}$.

$Var(\tilde S(t)|\theta) = E((\tilde S(t))^2) - (E(\tilde S(t)))^2$.

$E(\tilde S(t)|\theta) = E(\exp(-\frac{nt}{\sum_i X_i})) = \int_0^\infty \frac{\exp(-\frac{nt}{x})x^{n-1}\exp(-x/\theta)}{\Gamma(n)\theta^n}dx$.

$E((\tilde S(t))^2|\theta) = \int_0^\infty \frac{\exp(-\frac{2nt}{x})x^{n-1}\exp(-x/\theta)}{\Gamma(n)\theta^n}dx$

The estimator of the standard deviation of $\tilde{S}(t)$ is

$$\tilde{\sigma}_{\tilde{S}(t)} = \sqrt{\int_0^\infty \frac{x^{21-1}\exp(-x/(26/3) - \frac{42t}{x})}{\Gamma(21)(26/3)^{21}}dx - (\int_0^\infty \frac{x^{21-1}\exp(-x/(26/3) - \frac{21t}{x})}{\Gamma(21)(26/3)^{21}}dx)^2}$$

Since $\hat{S}(t) = \frac{1}{n}\sum_{i=1}^n I(X_i > t)$,
$\sigma^2_{\hat{S}(t)} = \sigma^2_{I(X_i>t)}/n = F(t)S(t)/n$.

$$\hat{\sigma}_{\hat{S}(t)} = \sqrt{\hat{F}(t)\hat{S}(t)/n} = \begin{cases} 0 & \text{if } t < 1 \\ \cdots & \text{if } t \in [1,2) \\ \cdots \end{cases}.$$

Notice that $\hat{\sigma}_{\hat{S}(t)}$ is a step functiona in t. $\hat{\sigma}_{\tilde{S}(t)}$ can be obtained by linear interpolation through $\hat{\sigma}_{\hat{S}(t)}$. .

About correction on midterm exam.

2. An actuary models the future lifetime of (30) as follows. $T(30)$ has force of mortality μ, where μ has pdf $f_\mu(u) = 400ue^{-20u}$, $u \geq 0$. Calculate ${}_tp_{30}$.
Modify the problem. An actuary models the future lifetime of (30) as follows. $T(30)$ has a uniform distribution $U(0, W)$, where W has pdf $f_W(t) = 1/40$ if $t \in (40, 80)$. Calculate ${}_tp_{30}$.
Sol. Let $t > 0$. Conditional on W, $S_{T(x)|W}(t) = 1 - t/W$, where $t \in (0, W)$. ${}_tp_{30} = E(1 - t/W)$

$$E(1-t/W) = \int_{40}^{80} \frac{1-t/t}{40}dt = \int_{40}^{80} \frac{0}{40}dt = 0...$$

$$E(1-t/W) = \int_{40}^{80} \frac{1-t/w}{40}dw = ... = 1 - \frac{t}{40}\ln 2, \quad t \in (0,40) \textbf{ (Is it correct ?)}$$

$$\begin{aligned}
E(1-t/W) = &\int_{40}^{80} \frac{1-t/w}{40}I(t<w)dw \quad (as \;\; f_W(w) = \frac{I(w \in (40,80))}{40}) \\
=&I(t \in (0,40])\int_{40}^{80} \frac{1-t/w}{40}dw + I(t \in (40,80])\int_{t}^{80} \frac{1-t/w}{40}dw \\
=&I(t \in (0,40])[1 - \frac{t}{40}\ln w\Big|_{40}^{80}] + I(t \in (40,80])[\frac{w}{40} - \frac{t}{40}\ln w]\Big|_t^{80} \\
=&I(t \in (0,40])(1 - \frac{t}{40}(\ln 80 - \ln 40)) + I(t \in (40,80])[\frac{80-t}{40} - \frac{t}{40}(\ln 80 - \ln t)] \\
=&I(t \in (0,40])(1 - \frac{t\ln 2}{40}) + I(t \in (40,80])[\frac{80-t}{40} - \frac{t}{40}\ln(80/t)]; \\
&Or \\
=&\begin{cases} (1 - \frac{t\ln 2}{40}) & \text{if } t \in (0,40] \\ [\frac{80-t}{40} - \frac{t}{40}\ln(80/t)] & \text{if } t \in (40,80] \end{cases}
\end{aligned}$$

3. Suppose that

x	90	91	92	93	94
p_x	0.2	0.1	0.05	0.01	0

. $E(K_{90} \wedge 2)$= ?

Solution: I first explained how to find $E(K(90))$, $E(K_{90})$, $e_{90:\overline{1}|}$, $E(K(90)\wedge 1)$ and $E(K_{90}\wedge 1)$. By [9], $E(K(90) \wedge n) = e_{x:\overline{n}|} = \sum_{k=1}^{n} {}_kp_x$.

$E(K(90)) = \sum_{k=1}^{\infty} {}_kp_x = 0.2 + 0.2 * 0.1 + 0.2 * 0.1 * 0.05 + 0.2 * 0.1 * 0.05 * 0.01$ (by [9])

$E(K_{90}) = E(K(90) + 1) = E(K(90)) + 1$ (by [8]).

$e_{90:\overline{1}|} = p_{90} = 0.2 = E(K(90) \wedge 1)$ (by [9]) Or by 447

$E(K(90) \wedge 1) = \sum_x f_X(x) = 0 \cdot P(K(90) = 0) + 1 \cdot P(K(90) \geq 1) = P(K(90) \geq 1) = p_{90} = 0.2.$

$$E(K_{90} \wedge 1) = E(K(90) \wedge 1) + 1 = 1.2 \ \ ? \ \ \textbf{Wrong !} \tag{1}$$

$K_{90} \in \{1, 2, 3, ...\}$ and $K(90) \in \{0, 1, 2, \cdots, \}$

$K_{90} \wedge 1 = (K(90) + 1) \wedge 1 =?$

$E(K_{90} \wedge 1) = \sum_{k=1}^{\infty} k f_X(k) = 1 \cdot 1 = 1$ (by 447)

$E(K_{90} \wedge 1) = \sum_{k=1}^{\infty} P(K_{90} \wedge 1 \geq k)$ (by [2] $E(X) = \sum_{k=1}^{\infty} P(X \geq k)$)

$= P(K_{90} \wedge 1 \geq 1) + P(K_{90} \wedge 1 \geq 2) + \cdots = P(K_{90} \geq 1) = 1.$

You need to redo it without copying it if your correction is wrong !

Now consider $E(K_{90} \wedge 2)$.

$K_{90} = \lceil T(90) \rceil \in \{1, 2, 3, ...\}$,

$K_{90} \wedge 2 \in \{1, 2\}$,

$P(K_{90} \geq 1) = P(T(90) > 0) = {}_0p_{90} = 1$,

$P(K_{90} \geq 2) = P(T(90) > 1) = {}_1p_{90} = p_{90}$

${}_tp_x = P(T(x) > t) \neq P(T(x) \geq t)$ **!!**

$E(K_{90} \wedge 2)$

$= \sum_{k=1}^{\infty} P(K_{90} \wedge 2 \geq k)$ (by [2]) $E(X) = \sum_{k\geq 1} P(X \geq k)$, $X =?$

$= P(K_{90} \wedge 2 \geq 1) + P(K_{90} \wedge 2 \geq 2)$

$= P(K_{90} \geq 1) + P(K_{90} \geq 2) = 1 + P(T(90) > 1) = 1 + p_{90} = 1.2.$

$E(K_{90} \wedge 2)$

$= 0P(K_{90} \wedge 2 = 0) + 1P(K_{90} \wedge 2 = 1) + 2P(K_{90} \wedge 2 = 2)$ (by 447)

$= 0 + P(K_{90} = 1) + 2P(K_{90} \geq 2)$

$= (1 - p_{90}) + 2 \cdot p_{90} = (1 - 0.2) + 2 * 0.2 = 1.2$