Q1 Prove that Greenwood's formula for the estimate of the variance of the Kaplan-Meier survivor function estimator

reduces to $\frac{1}{n}S_n(t)(1-S_n(t))$, where $S_n(t)$ is the empirical survivor function, provided there are neither censored observations nor ties.

The estimate of the variance

$$Var(\hat{S}(t)) = S^2(t)Var(\log(\hat{S}(t))) = \hat{S}^2(t)\sum_{i=1}^{k}\frac{d_i}{(n_i-d_i)n_i}$$

For a survival time t, k represents the number of event happened before t, $i = 1,2,..,k-1,k$. the number of in risk $n_i = n, n-1, .., n-k+2, n-k+1$. There isn't censored observations means $d_i = 1$.

$$\hat{S}(t) = \prod_{i=1}^{k}\frac{n_i-d_i}{n_i} = \prod_{i=1}^{k}\frac{n_i-1}{n_i} = \frac{n_1-1}{n_1}\cdot\frac{n_2-1}{n_2}\cdots\frac{n_{k-1}-1}{k-1}\cdot\frac{n_k-1}{n_k} = \frac{n-1}{n}\cdot\frac{n-2}{n-1}\cdots\frac{n-k+1}{n-k+2}\cdot\frac{n-k}{n-k+1} = \frac{n-k}{n}$$

$$\sum_{i=1}^{k}\frac{d_i}{(n_i-d_i)n_i} = \sum_{i=1}^{k}\frac{1}{(n_i-1)n_i} = \sum_{i=1}^{k}(\frac{1}{n_i-1}-\frac{1}{n_i}) = \frac{1}{n_1-1}-\frac{1}{n_1}+\frac{1}{n_2-1}-\frac{1}{n_2}+..+\frac{1}{n_{k-1}-1}-\frac{1}{n_{k-1}}+\frac{1}{n_k-1}-\frac{1}{n_k}$$

$$= \frac{1}{n-1}-\frac{1}{n}+\frac{1}{n-2}-\frac{1}{n-1}+..+\frac{1}{n-k+1}-\frac{1}{n-k+2}+\frac{1}{n-k}-\frac{1}{n-k+1} = \frac{1}{n-k}-\frac{1}{n}$$

$$\hat{S}^2(t)\sum_{i=1}^{k}\frac{d_i}{(n_i-d_i)n_i} = (\frac{n-k}{n})^2(\frac{1}{n-k}-\frac{1}{n}) = \frac{1}{n}(\frac{n-k}{n})(1-\frac{n-k}{n}) = \frac{1}{n}S_n(t)(1-S_n(t))$$

Q2 Suppose that the mean residual life of a continuous survival time

T at μ is given by $mrl(u) = u+10$.

(a) Find the mean of T

$$E(T) = \mathrm{mrl}(0) = 0+10 = 10$$

(b) Find $S(t)$

$$-\int_0^t \frac{1}{mrl(u)}du = -\int_0^t \frac{1}{u+10}du = -\log(u+10)|_0^t = \log\frac{10}{(t+10)}$$

$$S(t) = \frac{mrl(0)}{mrl(t)}\cdot e^{-\int_0^t \frac{1}{mrl(u)}du} = \frac{10}{t+10}\cdot e^{\log\frac{10}{(t+10)}} = (\frac{10}{t+10})^2$$

(c) Find $h(t)$

$$h(t) = -\frac{d}{dt}\log S(t) = -\frac{d}{dt}\log(\frac{10}{t+10})^2 = -\frac{d}{dt}(2\log(10)-2\log(t+10)) = \frac{2}{t+10}$$

Q3 Derive the Greenwood's variance formula for the Kaplan-Meier survivor function estimator using Delta method.

Hint: See what happens if you take a log-transformation of $\hat{S}(t)$.

By Delta Method, $X \sim \mu, \sigma^2$, $f(x) = f(\mu) + f'(\mu)(x-\mu)$

$E[f(x)] = f(\mu)$, $Var[f(x)] \approx \sigma^2[f'(\mu)]^2$

Let $f(x) = \log(x)$, $X = \hat{S}(t)$, then $\log(\hat{S}(t)) = \log(S(t)) + \frac{1}{S(t)}(\hat{S}(t) - S(t))$

$E[\log(\hat{S}(t))] = \log(S(t))$, $Var[\log(\hat{S}(t))] \approx Var(\hat{S}(t))[\frac{1}{S(t)}]^2$

Hence

$$Var(\hat{S}(t)) \approx S^2(t) Var(\log(\hat{S}(t))) = S^2(t) \sum_{i=1}^{k} \frac{d_i}{(n_i - d_i)n_i}$$

- Proof $Var(\log(\hat{S}(t))) = \sum_{i=1}^{k} \frac{d_i}{(n_i-d_i)n_i}$

$n_i - d_i \sim Bino(n_i, p_i)$

$Var(\hat{p}_i) = \frac{1}{n_i^2} Var(n_i - d_i) = \frac{1}{n_i} p_i(1-p_i)$

Let $f(x) = \log(\hat{p}_i)$, $f(\mu) = \log(p_i)$, $f'(\mu) = \frac{1}{p_i}$

Therefore,

$$Var(\log(\hat{S}(t))) = Var(\log(\prod_{i=1}^{k} \hat{p}_i)) = \sum_{i=1}^{k} Var(\log \hat{p}_i) = \sum_{i=1}^{k} \frac{1}{p_i^2} Var(\hat{p}_i)$$

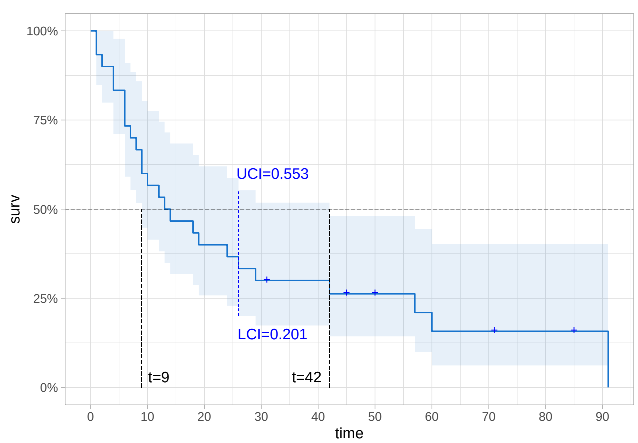
$$= \sum_{i=1}^{k} \frac{(1-p_i)}{p_i n_i} = \sum_{i=1}^{k} \frac{q_i}{p_i n_i} = \sum_{i=1}^{k} \frac{\frac{d_i}{n_i}}{(1-\frac{d_i}{n_i})n_i} = \sum_{i=1}^{k} \frac{d_i}{(n_i - d_i)n_i}$$

Q4 The data below are remission times, in weeks, for a group of 30 patients with leukemia who received similar treatment.

(a) Obtain and plot the Kaplan-Meier estimate $\hat{S}(t)$ of the survivor function for remission time.

```
me_km.fit <- survfit(Surv(time,status)~1,type="kaplan-meier")
summary(me_km.fit)
```

```
## Call: survfit(formula = Surv(time, status) ~ 1, type = "kaplan-meier")
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1     30       2    0.933  0.0455       0.8482        1.000
##     2     28       1    0.900  0.0548       0.7988        1.000
##     4     27       2    0.833  0.0680       0.7101        0.978
##     6     25       3    0.733  0.0807       0.5910        0.910
##     7     22       1    0.700  0.0837       0.5538        0.885
##     8     21       1    0.667  0.0861       0.5176        0.859
##     9     20       2    0.600  0.0894       0.4480        0.804
##    10     18       1    0.567  0.0905       0.4144        0.775
##    12     17       1    0.533  0.0911       0.3816        0.745
##    13     16       1    0.500  0.0913       0.3496        0.715
##    14     15       1    0.467  0.0911       0.3183        0.684
##    18     14       1    0.433  0.0905       0.2878        0.652
##    19     13       1    0.400  0.0894       0.2581        0.620
##    24     12       1    0.367  0.0880       0.2291        0.587
##    26     11       1    0.333  0.0861       0.2010        0.553
##    29     10       1    0.300  0.0837       0.1737        0.518
##    42      8       1    0.263  0.0812       0.1432        0.481
##    57      5       1    0.210  0.0801       0.0994        0.444
##    60      4       1    0.158  0.0754       0.0617        0.402
##    91      1       1    0.000     NaN           NA           NA
```



(b) Obtain approximate .95 confidence intervals for the median remission time and for the probability that remission lasts over 26 weeks.

```
me_km.fit
```

```
## Call: survfit(formula = Surv(time, status) ~ 1, type = "kaplan-meier")
##
##       n  events  median 0.95LCL 0.95UCL
##    30.0    25.0    13.5     9.0    42.0
```

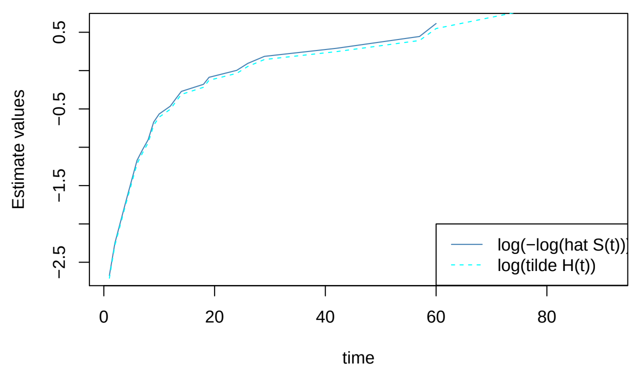
```
pander(data.frame(summary(me_km.fit)[c(2,6,7,14,15)])[15,])
```

	time	surv	std.err	lower	upper
15	26	0.3333	0.08607	0.201	0.5529

The approximate .95 confidence intervals for the median remission time ($t = 13.5$) is $[9, 42]$

The probability of remission lasts over 26 weeks is $S(26) = 0.333$. Its .95 confidence intervals is $[0.201, 0.553]$

(c) Plot $\log(-\log(\hat{S}(t)))$ and $\log(\hat{H}(t))$ on the same graph, where $\hat{H}(t)$ is the Nelson-Aalen estimate, (2.9). Is there much difference?



The two curves are almost the same. It shows that Nelson-Aalen estimate is the cumulative summation of the estimated probability.

Q5

(a) Estimate the median remission time by assuming that the underlying distribution of remission times is exponential. Compute approximate 95% confidence interval for the median remission time. Compare confidence intervals based on the nonparametric method in problem 4(b) and the confidence interval that you just computed.

	estimator	CI-L	CI-U
hat.Median	21.07	14.24	31.18

Using Weibull Distribution, the estimated median remission time is 21.0717

The confidence interval is (14.2383, 31.1845), which is narrow than CI in K-M estimate.

(b) Similarly, compare estimate of $S(26)$, the probability a remission lasts more than 26 weeks, using the nonparametric Kaplan-Meier estimate and the parametric model, respectively.

```r
mu<- weib.fit$coef # Intercept
(weib.Shat <- 1 - pweibull(26,1,exp(mu)) )
```

```
## [1] 0.4252
```

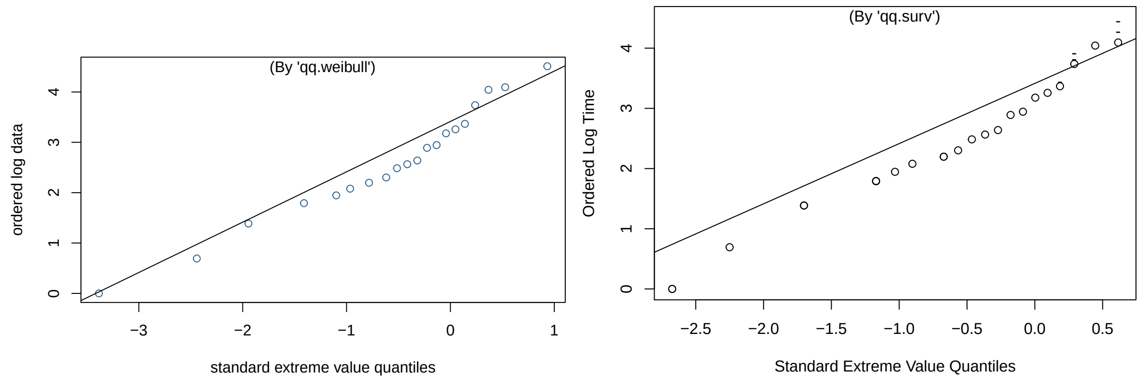
Using the parametric model, $S(26)=$0.4252. It is smaller than Kaplan-Meier estimated median remission time (0.3333)

(c) Is there any evidence against the parametric model?

```r
qq.weibull(Surv(time,status),scale = 1)
mtext("(By 'qq.weibull')",side=3,line=-1)
qq.surv(time,status,distribution = "weibull", scale =1)
```

```
##    logtime      sevq
## 1   0.0000 -2.673752
## 2   0.0000 -2.673752
## 3   0.6931 -2.250367
## 4   1.3863 -1.701983
## 5   1.3863 -1.701983
## 6   1.7918 -1.170683
## 7   1.7918 -1.170683
## 8   1.7918 -1.170683
## 9   1.9459 -1.030930
## 10  2.0794 -0.902720
## 11  2.1972 -0.671727
## 12  2.1972 -0.671727
## 13  2.3026 -0.565662
## 14  2.4849 -0.464246
## 15  2.5649 -0.366513
## 16  2.6391 -0.271625
## 17  2.8904 -0.178830
## 18  2.9444 -0.087422
## 19  3.1781  0.003297
## 20  3.2581  0.094048
## 21  3.3673  0.185627
## 22  3.4340  0.185627
## 23  3.7377  0.290805
## 24  3.8067  0.290805
## 25  3.9120  0.290805
## 26  4.0431  0.445101
## 27  4.0943  0.614282
## 28  4.2627  0.614282
## 29  4.4427  0.614282
## 30  4.5109       Inf
```

```r
mtext("(By 'qq.surv')",side=3,line=-1)
```



The QQ plot shows that some point are away from the line. If the proposed model fits the data adequately, the point should lie close to the straight line.

Q6 Suppose that the time to death T has an exponential distribution

with hazard rate λ and that the right-censoring time C is exponential with hazard rate θ.

Let $Y = \min(T, C)$ and $\delta = \begin{cases} 1 & T \le C \\ 0 & T > C \end{cases}$. Assume that T and C are independent.

(a). Find $P(\delta = 1)$. Hint: $P(\delta = 1 = P(T \le C))$

$T \sim Expo(\lambda)$, $C \sim Expo(\theta)$.$f_T(t) = \lambda e^{-\lambda t}$, $f_C(c) = \theta e^{-\theta c}$.

$T \perp C$, then $f_{T,C}(t, c) = \lambda e^{-\lambda t} \cdot \theta e^{-\theta c}$

$$\begin{aligned}
P(\delta = 1) = P(T \le C) &= \int_0^\infty \int_0^c f_{T,C}(t, c) dtdc \\
&= \int_0^\infty \int_0^c \lambda e^{-\lambda t} \cdot \theta e^{-\theta c} dtdc = \int_0^\infty \theta e^{-\theta c} (\int_0^c \lambda e^{-\lambda t}\ dt)dc = \int_0^\infty \theta e^{-\theta c}(-e^{-\lambda t}|_0^c)dc \\
&= \int_0^\infty \theta e^{-\theta c}(1 - e^{-\lambda c})dc = \int_0^\infty \theta e^{-\theta c} dc - \frac{\theta}{\theta + \lambda} \int_0^\infty (\theta + \lambda) e^{-(\theta+\lambda)c} dc \\
&= 1 - \frac{\theta}{\theta + \lambda}(-e^{-(\theta+\lambda)c}|_0^\infty) = 1 - \frac{\theta}{\theta + \lambda} = \frac{\lambda}{\theta + \lambda} \qquad \blacksquare
\end{aligned}$$

(b). Find the distribution of Y . Hint: Consider $P(Y > y)$

$$\begin{aligned}
F_Y(y) &= P(T \le y \cup C \le y) = 1 - P(T \ge y \cap C \ge y) \\
&\underset{T \perp C}{=} 1 - P(T \ge y)(C \ge y) = 1 - \int_y^\infty \lambda e^{-\lambda t} dt \cdot \int_y^\infty \theta e^{-\theta c} dc \\
&= 1 - (-e^{-\lambda t}|_y^\infty) \cdot (-e^{-\theta c}|_y^\infty) = 1 - e^{-(\lambda+\theta)y} \\
&\implies Y \sim Expo(\lambda + \theta) \qquad \blacksquare
\end{aligned}$$

(c). Show that δ and Y are independent. Hint: Consider $\lim_{\Delta y \to 0^+} \frac{P(y \le Y < y + \Delta y, \delta = 0)}{\Delta y}$ which can be written in terms of C and T and we know the joint pdf of C and T.

$$
\begin{aligned}
P(\delta=1,Y\le y) &= P(T\le C\cap Y\le y)=P(T\le C\cap T\le y)\\
&= \int_0^y\int_t^\infty f_{T,C}(t,c)dcdt=\int_0^y \lambda e^{-\lambda t}(\int_t^\infty \theta e^{-\theta c}dc)dt\\
&= \int_0^y \lambda e^{-\lambda t}(-e^{-\theta c}|_t^\infty)dt=\int_0^y \lambda e^{-(\theta+\lambda)t}dt=\frac{\lambda}{\theta+\lambda}\int_0^y(\theta+\lambda)e^{-(\theta+\lambda)t}dt\\
&= \frac{\lambda}{\theta+\lambda}(-e^{-(\theta+\lambda)t}|_0^y)=\frac{\lambda}{\theta+\lambda}(1-e^{-(\theta+\lambda)y})=P(\delta=1)\cdot P(Y\le y) \qquad \square\\
P(\delta=0,Y\le y) &= P(C\le T\cap Y\le y)=P(C\le T\cap C\le y)\\
&= \int_0^y\int_c^\infty f_{T,C}(t,c)dtdc=\int_0^y \theta e^{-\theta c}(\int_c^\infty \lambda e^{-\lambda t}dt)dc\\
&= \int_0^y \theta e^{-\theta c}(-e^{-\lambda t}|_c^\infty)dc=\int_0^y \theta e^{-(\theta+\lambda)c}dc=\frac{\theta}{\theta+\lambda}\int_0^y(\theta+\lambda)e^{-(\theta+\lambda)c}dc\\
&= \frac{\theta}{\theta+\lambda}(-e^{-(\theta+\lambda)c}|_0^y)=\frac{\theta}{\theta+\lambda}(1-e^{-(\theta+\lambda)y})=P(\delta=0)\cdot P(Y\le y) \qquad \square
\end{aligned}
$$

Therefore, $Y\perp\delta$

(d). Let $(Y_1,\delta_1),(Y_2,\delta_2),..,(Y_n,\delta_n)$ be a random sample from this model. The Maximum Likelihood Estimator of λ, $\hat\lambda$ is $\sum_{i=1}^n\delta_i/\sum_{i=1}^n Y_i$ Hint: Use (a) - (c) to write down the joint log-likelihood function for the random sample

$f(Y_i,\delta_i)=\begin{cases}\frac{\lambda}{\theta+\lambda}(\theta+\lambda)e^{-(\theta+\lambda)y_i} & T\le C\\ \frac{\theta}{\theta+\lambda}(\theta+\lambda)e^{-(\theta+\lambda)y_i} & T>C\end{cases}$

$L(Y_i,\delta_i)=\prod_{i=1}^n(\frac{\lambda}{\theta+\lambda})^{\delta_i}(\frac{\theta}{\theta+\lambda})^{1-\delta_i}(\theta+\lambda)e^{-(\theta+\lambda)y_i}$

$=(\frac{\lambda}{\theta+\lambda})^{\sum\delta_i}(\frac{\theta}{\theta+\lambda})^{n-\sum\delta_i}(\theta+\lambda)^n e^{-(\theta+\lambda)\sum y_i}=\lambda^{\sum_{i=1}^n\delta_i}\theta^{n-\sum_{i=1}^n\delta_i}e^{-(\theta+\lambda)\sum_{i=1}^n y_i}$

$l(Y_i,\delta_i)=\sum_{i=1}^n\delta_i\log\lambda+(n-\sum_{i=1}^n\delta_i)\log\theta-(\theta+\lambda)\sum_{i=1}^n y_i$

$$\frac{\partial}{\partial\lambda}l(Y_i,\delta_i)=\frac{1}{\lambda}\sum_{i=1}^n\delta_i-\sum_{i=1}^n y_i\overset{set}{=}0 \implies \hat\lambda=\frac{\sum_{i=1}^n\delta_i}{\sum_{i=1}^n Y_i} \qquad \blacksquare$$

(e). Use part (a) - (d) to find the mean and variance of $\hat\lambda$. Hint: Use the independence between δ_i and Y. Determine the distribution of $\sum_{i=1}^n\delta_i$. Remember that exponential distribution with a parameter $\lambda+\theta$ is the same as Gamma distribution with parameters 1 and $\lambda+\theta$. Then what would be the distribution of $\sum_{i=1}^n Y_i$

$\begin{cases}p(\delta_i=1)=\frac{\lambda}{\theta+\lambda} & T\le C\\ p(\delta_i=0)=\frac{\theta}{\theta+\lambda} & T>C\end{cases}$, $\sum_{i=1}^n\delta_i\sim Bino(n,\frac{\lambda}{\theta+\lambda})$

$E[\sum_{i=1}^n\delta_i]=\frac{n\lambda}{\theta+\lambda}$; $V[\sum_{i=1}^n\delta_i]=\frac{n\lambda}{\theta+\lambda}(1-\frac{\lambda}{\theta+\lambda})=\frac{n\lambda\theta}{(\theta+\lambda)^2}$

$E[(\sum_{i=1}^n\delta_i)^2]=V[\sum_{i=1}^n\delta_i]+(E[\sum_{i=1}^n\delta_i])^2=\frac{n\lambda\theta}{(\theta+\lambda)^2}+(\frac{n\lambda}{\theta+\lambda})^2=\frac{n^2\lambda^2+n\lambda\theta}{(\theta+\lambda)^2}$

$Y\sim Expo(\lambda+\theta)=Gamma(1,(\lambda+\theta)^{-1})$, Let $X=\sum_{i=1}^n Y_i\sim Gamma(n,(\lambda+\theta)^{-1})$

$$E[(\sum_{i=1}^{n} Y_i)^{-1}] = E[X^{-1}] = \int_0^\infty x^{-1}\frac{(\lambda+\theta)^n}{\Gamma(n)}x^{n-1}e^{-(\lambda+\theta)x}dx$$
$$= \frac{(\lambda+\theta)\Gamma(n-1)}{\Gamma(n)}\underbrace{\int_0^\infty \frac{(\lambda+\theta)^{n-1}}{\Gamma(n-1)}x^{n-1-1}e^{-(\lambda+\theta)x}dx}_{=1} = \frac{\lambda+\theta}{n-1}$$
$$E[(\sum_{i=1}^{n} Y_i)^{-2}] = E[X^{-2}] = \int_0^\infty x^{-2}\frac{(\lambda+\theta)^n}{\Gamma(n)}x^{n-1}e^{-(\lambda+\theta)x}dx$$
$$= \frac{(\lambda+\theta)^2\Gamma(n-2)}{\Gamma(n)}\underbrace{\int_0^\infty \frac{(\lambda+\theta)^{n-2}}{\Gamma(n-2)}x^{n-2-1}e^{-(\lambda+\theta)x}dx}_{=1} = \frac{(\lambda+\theta)^2}{(n-1)(n-2)}$$

Or by $X \sim Inverse - Gamma(n, (\lambda+\theta)^{-1})$,

$E[X^{-1}] = \frac{(\lambda+\theta)\Gamma(n-1)}{\Gamma(n)} = \frac{\lambda+\theta}{n-1}$; $E[X^{-2}] = \frac{(\lambda+\theta)^2\Gamma(n-2)}{\Gamma(n)} = \frac{(\lambda+\theta)^2}{(n-1)(n-2)}$

For $\sum_{i=1}^n \delta_i \perp \sum_{i=1}^n Y_i$

$$E[\hat\lambda] = E[\frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n Y_i}] = E[\sum_{i=1}^n \delta_i]E[(\sum_{i=1}^n Y_i)^{-1}] = \frac{n\lambda}{\theta+\lambda}\cdot\frac{\lambda+\theta}{n-1} = \frac{n\lambda}{n-1} \qquad \blacksquare$$
$$E[\hat\lambda^2] = E[(\sum_{i=1}^n \delta_i)^2]E[(\sum_{i=1}^n Y_i)^{-2}] = \frac{n^2\lambda^2+n\lambda\theta}{(\theta+\lambda)^2}\cdot\frac{(\lambda+\theta)^2}{(n-1)(n-2)} = \frac{n^2\lambda^2+n\lambda\theta}{(n-1)(n-2)}$$
$$V[\hat\lambda] = E[\hat\lambda^2] - (E[\hat\lambda])^2 = \frac{n^2\lambda^2+n\lambda\theta}{(n-1)(n-2)} - (\frac{n\lambda}{n-1})^2 = \frac{n\lambda(n\lambda+n\theta-\theta)}{(n-1)^2(n-2)} \qquad \blacksquare$$

Q7

(a) Write the self-consistency algorithm for a current status data set and create a function to compute the MLE of the survivor function. Define your notations in detail.

- Notation:

For current status data, we observe only the iid times $u_i, i = 1, ..., n$ and $\delta_i = I\{T_i \le U_i\}$.
$\delta = \begin{cases} 1 & \text{Event } T \le U \text{ Occurred} \\ 0 & \text{Event } T \le U \text{ Not occurred} \end{cases}$ but the exact time is unknown

u_i denote a grid of time points by $0 < u_1 < u_2 < .. < t_m$ at which subjects are observed. the u_i's are not all event times

d_i be the number of deaths at time u_i, d_i may be zero for some points

r_i be the number of individuals right-censored at time u_i, $\delta_i = 0$.

l_i be the number of left-censored observations at u_i, $\delta_i = 1$. The event of interest has occurred at some $t_j \le u_i$

The self-consistent estimator estimates the probability that this event occurred at each possible t_j less or more than u_i based on an initial estimate of the survival function.

Using this estimate, we compute an expected number of deaths at t_j, which is then used to update the estimate of the survival function

The procedure is repeated until the estimated survival function stabilizes.

- Algorithm:

Combine K-M estimator and self-consistent estimator based on an iterative procedure

Step 0: Produce an initial estimate of the survival function at each t_j, $\widehat{SC}(t_j)^{(0)}$

Step 1: By ignoring the left-censored observations, compute the usual K-M estimator or self-consistent estimator (By Theorem 1, K-M estimator is the unique self-consistent estimator for $u < t_n$) based on the estimated right-censored data.

$\widehat{S(t)} = \prod_{j=1}^{k}\left[\frac{n_j-d_j}{n_j}\right]$ Or

$\widehat{SC(u)} = \frac{1}{n}\left[\sum_{j=1}^{n} 1I(t_j > u, \delta_j = 0) + \sum_{j=1}^{n} I(t_j \le u, \delta_j = 1) + \sum_{j=1}^{n} \frac{SC(u)}{SC(t_j)} I(t_j \le u, \delta_j = 0)\right]$

$= \frac{1}{n}\left[N(t_j) + \sum_{j=1}^{n}(1-\delta_j)\frac{SC(u)}{SC(t_j)}\right]$

Step 2: Using the current estimate of $\widehat{SC}^{(k)}$, for $j \le i$,estimate

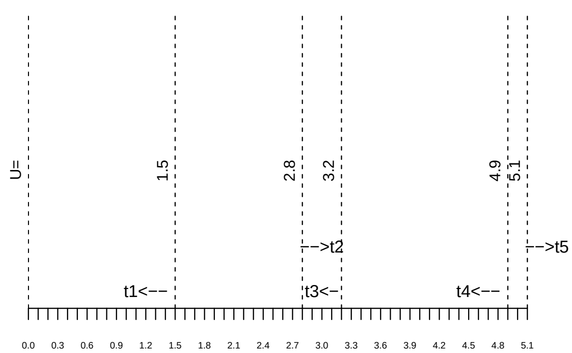
$$p_{ij} = P[t_{j-1} < X \le t_j | X \le u_i] = \frac{\widehat{SC}(t_{j-1})^{(k)} - \widehat{SC}(t_j)^{(k)}}{1 - \widehat{SC}(u_i)^{(k)}}$$

Step 3: Using the results of the previous step, estimate the number of events at time t_j by

$$\hat{d}_j = d_j + \sum_{i=j}^{n} l_i p_{ij}$$

Step 4: Compute the usual K-M estimator based on the estimated right-censored data with $\hat{d}_j$ events and r_j right-censored observations at t_j, ignoring the left-censored data

Step 5: If this estimate, $\widehat{SC}(t)^{(k+1)}$, is close to $\widehat{SC}(t)^{(k)}$ for all u_i, stop the procedure; if not, go to step 2.



(b) Using the following data, find the MLE of the survivor function: $(\mu_1 = 1.5, \delta_1 = 1), (\mu_2 = 2.8, \delta_2 = 0), (\mu_3 = 3.2, \delta_3 = 1), (\mu_4 = 4.9, \delta_4 = 1), and(\mu_5 = 5.1, \delta_5 = 0)$

```
SC_Est_CS <- function(time,status){ # For current status data
df <- data.frame(time,status)
df.distinct<- df %>% mutate(status=1-status)%>% group_by_all %>% count %>%
  rename(u_i=time,event = n)%>%as.data.frame()
df.distinct<- df.distinct %>% mutate(status=1-status)
df.distinct<- df.distinct %>% mutate(l_i= ifelse(status==1,event,0))%>%
  mutate(r_i= ifelse(status==0,event,0))%>%select(-2,-3)%>%
```

```r
  add_row(u_i=0,l_i=0,r_i=0, .before = 1)
n <- length(status)
k <- nrow(df.distinct)
n_i <- n_i.hat <- n
u_i <- df.distinct[,1]; l_i <- df.distinct[,2]; r_i <- df.distinct[,3]

d_i <- d_i.hat <- 0

for (i in 2:k) {
d_i[i] <- 0 # sum(t_j<=u_i[i])-1
n_i[i] <- n_i[i-1]-l_i[i-1]-r_i[i-1]
}

p_i <- (n_i-l_i)/n_i  # Probability of surviving through I_i | alive at beginning I_i
s_i0 <- s_i <- s_i.hat <-  1              # K-M estimator of the survivor function
for (i in 2:k) { s_i0[i]<- s_i[i]<-prod(p_i[1:i])}

conv <- 1e-3; iter <- 1
repeat{
p_ij <- matrix(rep(0,k^2),nrow = k, ncol = k,dimnames =list(c(paste("i",0:n)),c(paste("j",0:n))))
for (i in 2:k) {
  for (j in i:k) {
p_ij[i,j] <- (s_i[j-1]-s_i[j])/(1-s_i[i])
}}# d_i.hat <- l_i*rowSums(p_ij)#+l_i[i]
for (i in 2:k) {
d_i[i] <- sum(l_i[1:i])
d_i.hat[i] <- l_i[i]*sum(p_ij[i,])#+l_i[i]
n_i.hat[i] <- n_i.hat[i-1]-d_i.hat[i-1]#-r_i[i]
}
p_i.hat <- (n_i-d_i.hat)/n_i

for (i in 1:k) {s_i.hat[i]<- prod(p_i.hat[1:i])}

  if(sum(abs(s_i.hat - s_i))< conv)
break
else s_i <- s_i.hat; iter <- iter+1
}
out <- list(cbind(u_i,l_i,d_i,r_i,n_i,p_i,s_i0),p_ij,
            cbind(u_i,d_i.hat,n_i.hat,p_i.hat,s_i.hat))
names(out) <- c("Initial Values","p_ij", paste("Iteration times=",iter))
return(out)
}
```

- **Initial Values**:

u_i	l_i	d_i	r_i	n_i	p_i	s_i0
0	0	0	0	5	1	1
1.5	1	1	0	5	0.8	0.8
2.8	0	1	1	4	1	0.8
3.2	1	2	0	3	0.6667	0.5333
4.9	1	3	0	2	0.5	0.2667
5.1	0	3	1	1	1	0.2667

- **p__ij**:

	j 0	j 1	j 2	j 3	j 4	j 5
i 0	0	0	0	0	0	0
i 1	0	1	0	0.2336	0.07202	0
i 2	0	0	0	0.2336	0.07202	0
i 3	0	0	0	0.1894	0.05838	0
i 4	0	0	0	0	0.05516	0
i 5	0	0	0	0	0	0

- **Iteration times= 9**:

u_i	d__i.hat	n__i.hat	p__i.hat	s__i.hat
0	0	5	1	1
1.5	1.306	5	0.7389	0.7389
2.8	0	3.694	1	0.7389
3.2	0.2478	3.694	0.9174	0.6778
4.9	0.05516	3.447	0.9724	0.6591
5.1	0	3.391	1	0.6591