

Nonparametric Estimation of Mean Residual Life Function Using Scale Mixtures

Shufang Liu

PRA International, Charlottesville, VA

Sujit K. Ghosh*

Department of Statistics, North Carolina State University, Raleigh, NC 27695

*email: ghosh@stat.ncsu.edu

Institute of Statistical Mimeo Series # 2612

Abstract

The mean residual life function (mrlf) of a subject is defined as the average residual lifetime of the subject given that the subject has survived up to a given time point. A smooth nonparametric estimator of the mrlf is proposed using a scale mixtures of the empirical estimate of the mrlf. Asymptotic properties are established. The performances of the proposed estimator are studied based on simulated data sets and finally, a real data set is used to illustrate the practical relevance of the proposed estimator.

Keywords: Mean residual life function; Right-censored data; Scale mixtures; Smooth nonparametric estimate.

1 Introduction

The mean residual life function, the hazard function, and the survival function are complementary in understanding aging process, since each of them can be obtained from any one of the other two functions. The mean residual life function can serve as a more useful tool than the survival function and the hazard function to assess the remaining life expectancy. The **mean residual life function (mrlf)** of a lifetime random variable T with a finite expectation is defined as

$$m(t) \equiv E[T - t | T > t] = \begin{cases} \int_t^\infty \frac{S(u)}{S(t)} du, & S(t) > 0, \\ 0 & , \text{ otherwise,} \end{cases} \quad (1)$$

where $S(t)$ denotes the survival function associated with T . Note that $m(t)$ must satisfy the conditions of the Characterization Theorem of Hall and Weller (1981).

Since the mrlf is highly dependent on the tail behavior of the survival function. Therefore it is hard to estimate $m(t)$ with precision, especially when no parametric

form can be specified. Tsang and Jardine (1993), Agarwal and Kalla (1996), Kalla et al. (2001), Gupta and Bradley (2003), Lai et al. (2004), and Gupta and Lvin (2005) utilized parametric approaches to estimate the mrlf. However, an incorrect specification of parametric assumptions is undesirable, especially when prediction is of interest. Many researchers then focus on using nonparametric estimation procedures to study $m(t)$.

Yang (1978) employed the empirical survival function $\hat{S}(t)$ in (1) to obtain the empirical estimate of $m(t)$ for complete data. Ghorai et al. (1982) first used the estimator of Kaplan and Meier (1958) of $S(t)$ into (1) to obtain the empirical estimate of $m(t)$ for censored data. Since the empirical estimators of Yang and Ghorai et al. are discontinuous at each observation, smooth estimation of the mrlf has been developed. Essentially, all the methods derive the smooth versions of $m(t)$ by smoothing the survival functions. The kernel density method is often used. Kulasekara (1991) used the classical kernel density method. Ruiz and Guillamon (1996) used the recursive kernel density method. Abdous and Berred (2005) used the classical kernel density method with local linear fitting technique. Chaubey and Sen (1999) used the so called Hille's theorem (Hille, 1948) to smooth the survival function only for complete data, which is given by

$$\tilde{m}_n(t) = \frac{1}{\lambda_n} \frac{\sum_{k=0}^n \sum_{r=0}^k ((t\lambda_n)^{(k-r}) / (k-r)!) S_n(k/\lambda_n)}{\sum_{(k=0)} ((t\lambda_n)^k / k!) S_n(k/\lambda_n)}. \quad (2)$$

In the simulation study, we extend the method of Chaubey and Sen (1999) to right-censored data. In this study, we propose scale mixtures to smooth $\hat{m}_e(t)$ directly instead of first obtaining a smooth estimate of $\hat{S}(t)$ in Section 2. In Section 3, we present a simulation study to evaluate the performance of the proposed mrlf, the empirical mrlf, and the smooth mrlf (Chaubey and Sen, 1999). In Section 4, we compare the performance of the three mrlf estimators based on a real data set. In Section 5, we present some discussion and directions for further extensions.

2 A Scale Mixtures of the mrlf

Let T_i denote the survival time of the i -th subject for $i = 1, 2, \dots, n$ subjects. We denote the complete data set as $\mathcal{D}_0 = \{T_i : i = 1, 2, \dots, n\}$. In many applications T_i 's are often right censored by C_i 's, i.e., we only observe $X_i = \min\{T_i, C_i\}$ and the indicator of censoring $\Delta_i = I(C_i - T_i)$ for $i = 1, \dots, n$. The censored data set is

given by $\mathcal{D}_1 = \{(X_i, \Delta_i) : i = 1, 2, \dots, n\}$. We assume that (T_i, C_i) are independent and identically distributed (iid) and further we assume that T is independent of C .

In this section, we first introduce the Feller Approximation Lemma and the Scale Mixture Theory. We then develop a smooth estimator $\hat{m}_m(t)$ of $m(t)$ using the Feller Approximation Lemma. $\hat{m}_m(t)$ happens to be a scale mixtures of the empirical mrlf $\hat{m}_e(t)$. In the sense of the Scale Mixture Theory, $\hat{m}_m(t)$ is a proper mrlf which satisfies the Characterization Theorem (Hall and Weller, 1981). We then present closed calculation forms for $\hat{m}_e(t)$ and $\hat{m}_m(t)$ for complete and censored data. Finally, we provide the asymptotic properties of $\hat{m}_m(\cdot)$.

First, we restate the result due to Petrone and Veronese (2002) based on Feller (1966, p.219).

LEMMA 1. *Feller Approximation Lemma:*

Let $g(\cdot)$ be a bounded and right continuous function on \mathbb{R} for each t . Let $Z_k(t)$ be a sequence of random variables for each t such that $\mu_k(t) \equiv E[Z_k(t)] \rightarrow t$ and $\sigma_k^2(t) \equiv Var[Z_k(t)] \rightarrow 0$ as $k \rightarrow \infty$. Then

$$E[g(Z_k(t))] \rightarrow g(t) \quad \forall t.$$

The proof of the lemma is given in Appendix A. Now, we introduce the Scale Mixture Theory.

THEOREM 1. *Scale Mixture Theory:*

Let $m(t) = E[T - t | T > t]$ be an mrlf. Then

- (a) $\frac{m(t\theta)}{\theta}$ is a proper mrlf for any $\theta > 0$;
- (b) $\int_0^{+\infty} \frac{m(t\theta)}{\theta} \pi(\theta) d\theta$ is a proper mrlf for any density $\pi(\cdot)$ for $[0, \infty)$.

The proof of the theory is given in Appendix B. Now we propose a scale mixtures of the empirical mrlf $\hat{m}_e(t)$. It is known that $\hat{m}_e(t)$ is a right continuous function on $[0, T_{(n)}]$ and $\hat{m}_e(t) = 0$ for $t > T_{(n)}$ and hence $\hat{m}_e(t)$ is a bounded function. Then, we can use the Feller Approximation Lemma to approximate $\hat{m}_e(t)$ by $\hat{m}_m(t)$. Let $Z_n(t) \sim Ga(k_n, \frac{t}{k_n})$ for $t > 0$, where $Ga(k_n, \frac{t}{k_n})$ denotes a Gamma distribution with mean $\mu_n(t) = t$ and variance $\sigma_n^2(t) = \frac{t^2}{k_n}$ and the density function is given by

$$f_{k_n, t}(z) = f\left(z \middle| k_n, \frac{t}{k_n}\right) = \left(\frac{k_n}{t}\right)^{k_n} \frac{1}{\Gamma(k_n)} z^{k_n-1} e^{-\frac{k_n z}{t}}.$$

Define,

$$\hat{m}_m(t) = E[\hat{m}_e(Z_n(t))] = \int_0^\infty \hat{m}_e(u) dF_{k_n, t}(u) \quad (3)$$

$$= \int_0^\infty \frac{\hat{m}_e(t\theta)}{\theta} \pi(\theta|k_n) d\theta, \quad (4)$$

where $\pi(\theta|k_n) = \frac{k_n^{k_n}}{\Gamma(k_n)} \theta^{k_n} e^{-k_n \theta}$ is the density function of a $Ga(k_n + 1, \frac{1}{k_n})$ distribution, and hence it follows from (b) of Theorem 1 that $\hat{m}_m(t)$ is a scale mixture mrlf and it is a proper mrlf.

We give a unified closed calculation form for $\hat{m}_e(t)$ for complete and censored data. Let the ordered observed data be $0 \stackrel{def}{=} X_0 < X_1 < \dots < X_n < X_{n+1} \stackrel{def}{=} \infty$. The empirical mrlf $\hat{m}_e(t)$ can be expressed as

$$\hat{m}_e(t) = \begin{cases} \frac{\sum_{j=1}^n (X_j - t) I(X_j - t) w_j}{\sum_{j=1}^n I(X_j - t) w_j}, & \text{if } 0 < t \leq X_n, \\ 0, & \text{if } t > X_n, \end{cases} \quad (5)$$

$$= \begin{cases} \frac{\sum_{j=l+1}^n X_j w_j}{\sum_{j=l+1}^n w_j} - t, & \text{if } X_l \leq t < X_{l+1}, l = 0, 1, \dots, n-1, \\ 0, & \text{if } t > X_n, \end{cases} \quad (6)$$

where $w_j = F_n(X_j) - F_n(X_{j-})$ and $F_n(X_j)$ is the KM estimate. Notice that $w_j = 1/n$ for complete data.

Then, $\hat{m}_m(t)$ can also be calculated with the same closed form for complete and censored data. The details of the calculation are given in the Appendix C.

$$\hat{m}_m(t) = \int_0^\infty \hat{m}_e(u) f\left(u \left| k_n, \frac{t}{k_n} \right.\right) du \quad (7)$$

$$= \sum_{l=0}^{n-1} \frac{\sum_{j=l+1}^n X_j w_j}{\sum_{j=l+1}^n w_j} \left[F\left(X_{l+1} \left| k_n, \frac{t}{k_n} \right.\right) - F\left(X_l \left| k_n, \frac{t}{k_n} \right.\right) \right] - t F\left(X_n \left| k_n + 1, \frac{t}{k_n} \right.\right), \quad (8)$$

where $F(\cdot | k_n, \frac{t}{k_n})$ is the cdf of $Ga(k_n, \frac{t}{k_n})$.

REMARK 1. *If $F_n(X_j)$ is the cdf estimate for data with tied observations, $\hat{m}_e(t)$ and $\hat{m}_m(t)$ based on data with tied observations can be calculated using the same closed forms as (6,8).*

Next we show that $\hat{m}_m(t)$ is a smooth mrlf. To be a smooth estimate of $m(t)$, $\hat{m}_m(t)$ must satisfy the following conditions:

- (i) $\pi(\theta|k_n)$ is a smooth density on $[0, \infty)$ such that $\int_0^\infty \frac{1}{\theta} \pi(\theta|k_n) d\theta < \infty$ for all n ;

(ii) If $Y_n \sim \pi_n(\cdot|k_n)$ then $Y_n \xrightarrow{P} 1$ as $n \rightarrow \infty$.

Since

$$\int_0^\infty \frac{1}{\theta} \pi(\theta|k_n) d\theta = \int_0^\infty \frac{1}{\theta} \frac{k_n^{k_n}}{\Gamma(k_n)} \theta^{k_n} e^{-k_n \theta} = 1,$$

condition (i) is satisfied.

Since $Y_n \sim Ga(k_n + 1, 1/k_n)$,

$$E[Y_n] = \frac{k_n + 1}{k_n} = 1 + \frac{1}{k_n} \rightarrow 1 \text{ as } n \rightarrow \infty$$

and

$$var[Y_n] = \frac{k_n + 1}{k_n^2} = \frac{k_n + 1}{k_n} \frac{1}{k_n} \rightarrow 0 \text{ as } n \rightarrow \infty,$$

then condition (ii) is satisfied. Therefore $\hat{m}_m(t)$ is a smooth estimator.

Under conditions (i) and (ii) and the assumption of $k_n = cn^{1+\epsilon}$ where $\epsilon > 0$ and c is a constant, we claim that $\hat{m}_m(t)$ has the following properties:

THEOREM 2. $\hat{m}_m(t)$ pointwisely converges in probability to $m(t)$.

THEOREM 3. $\sqrt{n}(\hat{m}_m(t) - m(t))$ is asymptotically a Gaussian process with mean identically zero and covaraince function $\sigma(s, t)$, where

$$\sigma(s, t) = \frac{1}{S(s)S(t)} \int_t^\infty u^2 dF(u) - \frac{1}{S(t)} \left(\int_t^\infty u dF(u) \right)^2, \quad 0 \leq s \leq t < \infty$$

for complete data and

$$\sigma(s, t) = \frac{1}{S(s)S(t)} \int_t^\infty \frac{\phi^2(\nu)}{(S(\nu)G(\nu))^2} d\tilde{H}(\nu), \quad 0 \leq s \leq t < \infty,$$

where $\phi(t) = \int_t^\infty S(u) du$, $G(t) = Pr[C > t]$, and $\tilde{H}(t) = Pr[\Delta_1 = 1, X_1 \leq t]$ for censored data.

We know that $\hat{m}_e(t)$ for complete data by Yang(1978) and $\hat{m}_e(t)$ for right-censored by Ghorai et al. (1982) have the following asymptotic properties:

(a) $\hat{m}_e(t)$ pointwisely converges in probability to $m(t)$:

$$\lim_{n \rightarrow \infty} P(|\hat{m}_e(t) - m(t)| \geq \epsilon) = 0.$$

(b) $\sqrt{n}(\hat{m}_e(t) - m(t))$ is asymptotically a Gaussian process with mean identically zero and covaraince function given by $GP(0, \sigma(s, t))$.

Now we prove Theorem 2.

Proof:

Since $\hat{m}_e(t)$ pointwisely converges in probability to $m(t)$ from Yang (1978) and Ghorai et al. (1982), i.e.,

$$\hat{m}_e(t) \xrightarrow{P} m(t) \text{ as } n \rightarrow \infty$$

and

$$Y_n \xrightarrow{P} 1 \text{ as } n \rightarrow \infty,$$

then by Slutsky's Theorem

$$\frac{\hat{m}_e(tY_n)}{Y_n} \xrightarrow{P} \frac{m(t * 1)}{1} = m(t).$$

Notice that $\frac{\hat{m}_e(tY_n)}{Y_n}$ is the mrlf of $\frac{\tilde{T}_n}{Y_n}$, where \tilde{T}_n takes T_1, \dots, T_n with probability of $\frac{1}{n}$ each. Then we can see that $E[\frac{\tilde{T}_n}{Y_n} - t | \frac{\tilde{T}_n}{Y_n} > t, Y_n]$ is bounded. Thus we can obtain the following relationship by DCT

$$E \left[\frac{\hat{m}_e(tY_n)}{Y_n} \right] \xrightarrow{P} E[m(t)] \text{ wrt } \pi,$$

i.e.,

$$\hat{m}_m(t) \xrightarrow{P} m(t).$$

This completes the proof of Theorem 2.

Now we prove Theorem 3.

Proof:

$$\sqrt{n}(\hat{m}_m(t) - m(t)) = \sqrt{n}(\hat{m}_m(t) - \hat{m}_e(t)) + \sqrt{n}(\hat{m}_e(t) - m(t)),$$

where $\sqrt{n}(\hat{m}_e(t) - m(t)) \sim GP(0, \sigma(\cdot, \cdot))$.

Now we consider $\sqrt{n}(\hat{m}_m(t) - \hat{m}_e(t)) = \sqrt{n}E[\frac{\hat{m}_e(tY_n)}{Y_n} - \hat{m}_e(t)]$.

First, we take Taylor's expansion on $\frac{\hat{m}_e(tY_n)}{Y_n}$

$$\frac{\hat{m}_e(tY_n)}{Y_n} = \frac{\hat{m}_e(t)}{Y_n} + \frac{(tY_n - t)\hat{m}'_e(t_n^*)}{Y_n},$$

where $t_n^* \xrightarrow{P} t$ as $n \rightarrow \infty$. It can be seen from (6) that $\hat{m}_e(t)$ is differentiable iff $t \in (X_l, X_{l+1}), l = 1, \dots, n-1$ and $\hat{m}_e(t)$ is not differentiable if $t = X_l, l = 1, \dots, n-1$. Thus, $\hat{m}'_e(t_n^*)$ exists for $t_n^* \in [0, \infty) \setminus \{X_1, \dots, X_n\}$.

Second, we substitute the Taylor's expansion into $\sqrt{n}E[\frac{\hat{m}_e(tY_n)}{Y_n} - \hat{m}_e(t)]$

$$\begin{aligned}\sqrt{n}(\hat{m}_m(t) - \hat{m}_e(t)) &= \sqrt{n}E\left[\frac{\hat{m}_e(tY_n)}{Y_n} - \hat{m}_e(t)\right] \\ &= \sqrt{n}E\left[\frac{\hat{m}_e(t)}{Y_n} - \hat{m}_e(t) + \frac{(tY_n - t)\hat{m}'_e(t^*)}{Y_n}\right] \\ &= E\left[\sqrt{n}\left(\frac{1}{Y_n} - 1\right)\hat{m}_e(t) - \sqrt{n}\left(\frac{1}{Y_n} - 1\right)t\hat{m}'_e(t^*)\right]\end{aligned}$$

If we can prove $\sqrt{n}\left(\frac{1}{Y_n} - 1\right) \xrightarrow{P} 0$ as $n \rightarrow \infty$, then we can obtain $\sqrt{n}(\hat{m}_m(t) - \hat{m}_e(t)) \xrightarrow{P} 0$ as $n \rightarrow \infty$ and the proof is complete.

Since $Y_n \sim Ga(k_n + 1, 1/k_n)$, then

$$E\left[\sqrt{n}\left(\frac{1}{Y_n} - 1\right)\right] = \sqrt{n}\left(E\left[\frac{1}{Y_n}\right] - 1\right) = \sqrt{n}\left(\frac{k_n}{k_n + 1} - 1\right) = 0$$

and

$$Var\left[\sqrt{n}\left(\frac{1}{Y_n} - 1\right)\right] = n * Var\left[\frac{1}{Y_n}\right] = n\frac{k_n^2}{k_n^2(k_n - 1)} = \frac{n}{k_n - 1}.$$

Since we assume $k_n = cn^{1+\epsilon}$ where $\epsilon > 0$, it can be seen that $var\left[\sqrt{n}\left(\frac{1}{Y_n} - 1\right)\right] \rightarrow 0$ as $n \rightarrow \infty$. Therefore, $\sqrt{n}\left(\frac{1}{Y_n} - 1\right) \xrightarrow{P} 0$ as $n \rightarrow \infty$ is proved. Hence, this completes the proof of Theorem 3.

REMARK 2. Notice that the asymptotic properties of $\hat{m}_m(t)$ are the same as those of the estimator of Yang (1978) for complete data and the same as those of the estimator of Ghorai et al. (1982) for censored data.

3 A Simulation Study

We conducted a simulation study to explore the properties of the scale mixture mrlf. We extended the smooth estimation procedures of Chaubey and Sen (1999) to censored data. The simulation study was used to compare the performance of the empirical mrlf $\hat{m}_e(t)$ (emrlf), the scale mixture mrlf $\hat{m}_m(t)$ (mmrlf), and the smooth mrlf $\hat{m}_s(t)$ (smrlf)(Chaubey and Sen, 1999). The simulation study was conducted under these conditions:

- The data are generated from Wei(2, 2). The performance of emrlf, smrlf, and mmrlf is evaluated at the time points from quantile 0.01 to quantile 0.90 of this distribution.
- The censoring distribution is Exp(λ). The choices of λ to obtain the averaged censoring rates of 0%, 20%, 40%, and 60% are listed in the following table:

Censoring Rate	Exp
0%	$\lambda = \infty$
20%	$\lambda = 7.633$
40%	$\lambda = 3.225$
60%	$\lambda = 1.667$

- A sample size of $n = 100$ and a Monte Carlo sample size of $N = 1000$ are used.
- The parameter k_n in the Gamma distribution to calculate the scale mixture mrlf $\hat{m}_m(t)$ (mmrlf) is set to be $n^{1.01}$.
- The bias can be estimated as $\hat{B}(t) = \frac{\sum_{i=1}^N \hat{m}_i(t)}{N} - m(t)$, and then the relative bias is $\frac{\hat{B}(t)}{m(t)}$. The MSE can be estimated as $\widehat{MSE}(t) = \frac{1}{N-1} \sum_{i=1}^N (\hat{m}_i(t) - \bar{\hat{m}}(t))^2 + (\bar{\hat{m}}(t) - m(t))^2$. The relative efficiency $\frac{\widehat{MSE}_{emrlf}(t)}{\widehat{MSE}_{mmrlf}(t)}$ and $\frac{\widehat{MSE}_{emrlf}(t)}{\widehat{MSE}_{smrlf}(t)}$ is compared.

Figures 1-4 show the graphical summary of the simulation for the targeted censoring rates of 0%, 20%, 40%, and 60%, respectively. In each figure, Panels 1, 3, and 5 feature the boxplots of the 1000 replications of $\hat{m}(t)$ changing with time for emrlf, mmrlf, and smrlf, respectively. Panel 2 compares the averaged emrlf, mmrlf, and smrlf from the 1000 replications with the true mrlf. Panel 4 compares the bias of averaged emrlf, mmrlf, and smrlf with 0. Panel 6 shows the relative efficiency (RE) of emrlf to mmrlf (in dotted line), emrlf to smrlf (in dash line). Tables 2-5 list the relative bias and the relative efficiency for the experiments of the censoring rates of 0%, 20%, 40%, and 60%, respectively.

Generally, emrlf, mmrlf, and smrlf fit the data well for all the censoring rates of 0%, 20%, 40%, and 60%. With the increase of the censoring rate, the fluctuation of emrlf, mmrlf, and smrlf also increases, which indicates that the variation of emrlf,

mmrlf, and smrlf increases with the censoring rate as expected. With the increase of the censoring rate, the bias for emrlf, mmrlf, and smrlf also increase. These are reasonable results, because the information provided by the data is limited by the censored observations.

Figure 1 and Table 2 present the simulation results for data with 0% censoring rate, i.e., complete data. It can be seen from Figure 1 that emrlf, mmrlf, and smrlf have the similar performance for the small and moderate time points. While emrlf has the largest variation and the smrlf has the least variation for the extreme time points. From Table 2, it can be seen that smrlf has larger relative biases at small time points compared to emrlf and mmrlf. It turns out that these biases of smrlf are statistically significant (P-value < 0.05) for all time points smaller than 1.92 (the 33%tile of this Weibull distribution). Generally, mmrlf has smaller biases than emrlf, though none of the biases from emrlf or mmrlf are significantly different from 0. In terms of relative efficiency, mmrlf is always better than emrlf. The smrlf is less efficient for small time points and more efficient for large time points compared to emrlf and smrlf. Therefore, based on relative bias and relative efficiency, mmrlf is a better estimator compared to emrlf, and mmrlf is a better estimator for small time points compared to smrlf. The reason why smrlf does better for large time points might be that the number of the observations used to calculate mmrlf and emrlf decreases with the increase of time, while the number of the observation used to calculate smrlf stays the same. In summary, mmrlf is a definitely better estimator compared to emrlf and a competitive estimator to smrlf for complete data.

Figure 2 and Table 3 present the simulation results for data with 20% censoring rate. We get the similar results for slightly censored data as for the complete data, i.e., mmrlf is a better estimator compared to emrlf and a better estimator for small data points compared to smrlf.

Figure 3 and Table 4 present the simulation results for data with 40% censoring rate. For all the time points studied in this simulation, smrlf is significantly biased. It appears that emrlf and mmrlf are only biased for time points larger than quantile 0.875 (P-value < 0.05). Therefore, for moderately censored data, mmrlf is better than emrlf and smrlf.

Figure 4 and Table 5 present the simulation results for data with 60% censoring rate. For highly censored data, emrlf, mmrlf and smrlf are all significantly biased for all the time points evaluated (P-value < 0.05). However, the biases from emrlf

and smrlf are smaller than those from smrlf. Therefore, for highly censored data, emrlf, smrlf and mmrlf are not so good estimators.

Considering the fact that we sometimes may encounter data sets with a very large sample size, we attempted to compare the performance of emrlf, mmrlf, and smrlf for a simulation study for data with a sample size of 5000. We set the other conditions as before, except that the sample size n changed to 5000 from 100. Since the calculation of smrlf involves smoothing the survival function in the numerator and denominator (see (2)), the computation for data with a sample size of 5000 is very intensive. We failed to do the simulation for smrlf under our computing resource. Since $\hat{m}_e(t)$ and $\hat{m}_m(t)$ have closed calculation forms, it did not take much time to do the simulation for emrlf and mmrlf. The results show that mmrlf is more efficient compared to emrlf for data with a very large sample size.

4 Application to a Melanoma Study

We apply our estimation methods to the data set given in Ghorai et al. (1982). This data set has the survival times (in weeks) of 68 participants of a melanoma study conducted by the Central Oncology Group with headquarters at the University of Wisconsin-Madison. This data set has a censoring rate of 58.8%.

Figure 5 shows the results of the three methods. Panel 1 compares emrlf, mmrlf, and smrlf. Panel 2-4 show the 95% asymptotic confidence interval from case resampling for emrlf, mmrlf, and smrlf, respectively. It can be seen that emrlf, mmrlf, and smrlf have similar results for this highly censored data. However, unlike emrlf is a discontinuous estimator, mmrlf and smrlf are smooth estimators. It also can be seen that emrlf vanishes at the largest time point, while mmrlf and smrlf can take positive values. This indicates that mmrlf and smrlf can have a better prediction performance than emrlf even outside the observed time range.

Table 1 shows the estimates of the mean residual life function and their corresponding 95% confidence intervals from the empirical mrlf procedure, the scale mixture mrlf procedure, and the modified Chaubey and Sen's mrlf procedure at some time points. The time points are 23.4, 58.5, 117, 175.5, and 210.6, which are the 10th, 25th, 50th, 75th, and 90th percentiles of observed times. For example, the 50th percentile of the times is 117, then $\hat{m}_m(117) = 54.2$. This indicates that the expected number of weeks that a melanoma patient will survive is about 54,

Table 1: Estimated mrlf's with 95% CI at some time points for the Melanoma study

time	23.4	58.5	117	175.5	210.6
emrlf	120.8	89.4	53.2	31.2	10.1
	(97.1, 139.5)	(75.3, 108.1)	(38.8, 71.8)	(21.6, 50.8)	(1.9, 30.8)
mmrlf	120.8	89.1	54.2	32.2	15.5
	(105.6, 139.3)	(74.3, 107.2)	(40.6, 69.8)	(23.9, 45.0)	(8.8, 24.5)
smrlf	118.6	88.4	55.0	33.6	21.9
	(101.5, 137.1)	(71.9, 105.8)	(39.3, 70.2)	(21.8, 44.3)	(11.8, 28.3)

given that the patient has already survived at least 117 weeks after entering the study. Also a 95% confidence interval for $\hat{m}_m(117)$ is (40.6, 69.8), which indicates that it is highly unlikely that the patient will survival beyond about 70 weeks and it is also highly unlikely that the patient will survival less than about 40 weeks after entering the study for 117 weeks.

Notice that the conclusions derived from Table 1 can not be made using the hazard function directly. This illustrates one of the advantages of using the mean residual life function over the traditional hazard function.

5 Discussion

In the literature, researchers often first estimate the survival function and then calculate $\hat{m}_e(t)$. The closed form of $\hat{m}_e(t)$ from (6) provides an easy and unified calculation procedure for complete data and right-censored data. The closed form of our scale mixture mrlf $\hat{m}_m(t)$ from (8) is a nonparametric smooth estimator of the mrlf for complete and right censored data.

$\hat{m}_m(t)$ is always a better estimator compared to $\hat{m}_e(t)$. In terms of bias, $\hat{m}_m(t)$ is always a better estimator compared to the Chaubey and Sen's (1999) estimator $\hat{m}_s(t)$. In terms of MSE, $\hat{m}_m(t)$ is competitive to $\hat{m}_s(t)$.

Finally, $\hat{m}_e(t)$ and $\hat{m}_m(t)$ are computationally less demanding compared to the $\hat{m}_s(t)$.

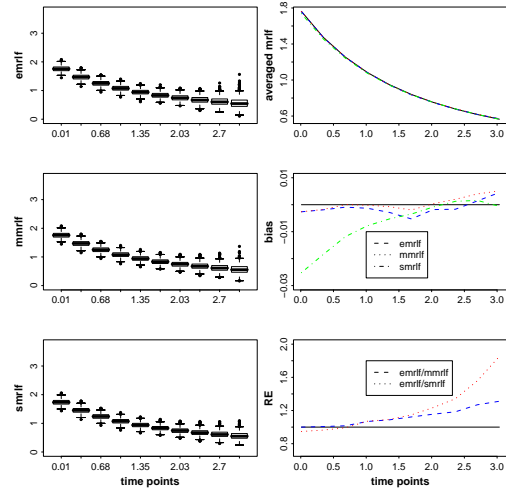


Figure 1: Estimated emrlf, mmrlf, and smrlf under 0% censoring rate

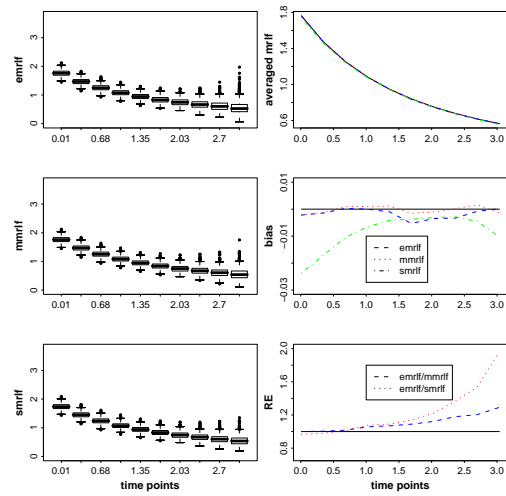


Figure 2: Estimated emrlf, mmrlf, and smrlf under 20% censoring rate

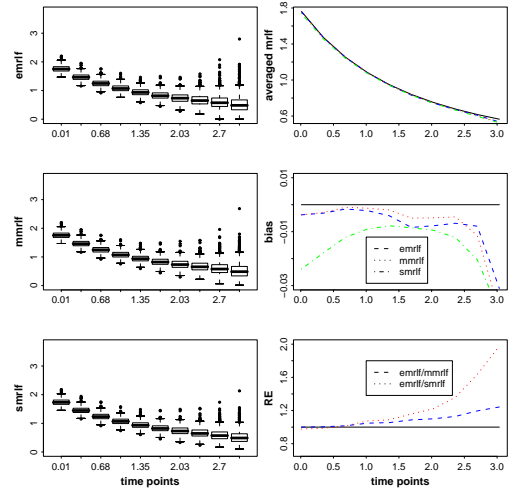


Figure 3: Estimated emrlf, mmrlf, and smrlf under 40% censoring rate

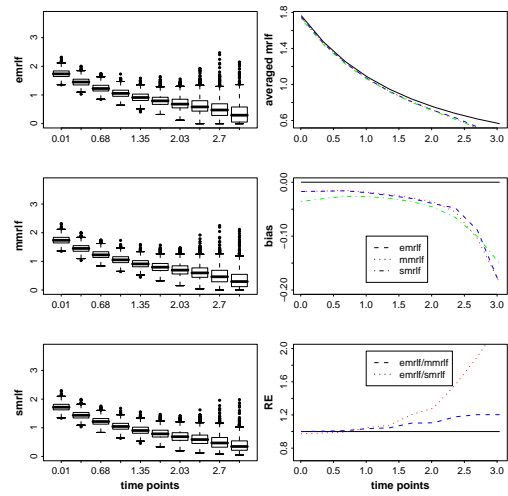


Figure 4: Estimated emrlf, mmrlf, and smrlf for under 60% censoring rate

Table 2: Relative biases and efficiency of estimated mrlf's under 0% censoring rate

time	$\frac{B_{mrlf}(t)}{m(t)}$	$\frac{B_{mrlf}(t)}{m(t)}$	$\frac{B_{mrlf}(t)}{m(t)}$	$\frac{MSE_{mrlf}(t)}{MSE_{mrlf}(t)}$	$\frac{MSE_{mrlf}(t)}{MSE_{mrlf}(t)}$	$\frac{MSE_{mrlf}(t)}{MSE_{mrlf}(t)}$
0.010	-0.001	-0.001	-0.014	1.000	0.947	1.056
0.169	-0.002	-0.001	-0.014	1.003	0.961	1.044
0.328	-0.001	-0.001	-0.013	1.014	0.970	1.045
0.488	-0.001	-0.001	-0.011	1.015	0.992	1.024
0.647	-0.001	0.000	-0.010	1.012	0.987	1.025
0.806	-0.001	0.000	-0.009	1.037	1.021	1.015
0.965	-0.001	0.000	-0.008	1.056	1.056	1.000
1.124	-0.002	0.000	-0.007	1.067	1.069	0.998
1.284	-0.001	0.000	-0.006	1.064	1.067	0.997
1.443	-0.004	-0.001	-0.005	1.096	1.099	0.998
1.602	-0.004	-0.002	-0.005	1.091	1.102	0.990
1.761	-0.004	-0.002	-0.004	1.159	1.198	0.967
1.920	-0.004	-0.001	-0.002	1.162	1.226	0.948
2.080	-0.002	0.001	-0.001	1.155	1.244	0.928
2.239	0.001	0.002	0.001	1.182	1.309	0.903
2.398	-0.003	0.003	0.002	1.186	1.354	0.876
2.557	-0.001	0.005	0.003	1.261	1.490	0.846
2.716	0.000	0.007	0.003	1.248	1.541	0.810
2.876	0.008	0.009	0.001	1.256	1.643	0.765
3.035	0.008	0.009	-0.001	1.312	1.852	0.708

On average the MCSE is about 0.004.

Table 3: Relative biases and efficiency of estimated mrlf's under 20% censoring rate

time	$\frac{B_{mrlf}(t)}{m(t)}$	$\frac{B_{mrlf}(t)}{m(t)}$	$\frac{B_{mrlf}(t)}{m(t)}$	$\frac{MSE_{mrlf}(t)}{MSE_{mrlf}(t)}$	$\frac{MSE_{mrlf}(t)}{MSE_{mrlf}(t)}$	$\frac{MSE_{mrlf}(t)}{MSE_{mrlf}(t)}$
0.010	-0.001	-0.001	-0.013	1.000	0.966	1.036
0.169	-0.001	-0.001	-0.013	1.002	0.977	1.026
0.328	-0.001	-0.001	-0.012	1.010	0.981	1.030
0.488	0.000	0.000	-0.010	1.013	1.007	1.006
0.647	0.000	0.001	-0.009	1.010	1.002	1.008
0.806	0.000	0.001	-0.007	1.035	1.038	0.998
0.965	0.000	0.001	-0.006	1.052	1.067	0.986
1.124	-0.001	0.001	-0.005	1.057	1.077	0.981
1.284	0.001	0.001	-0.005	1.063	1.084	0.981
1.443	-0.001	0.001	-0.004	1.083	1.105	0.980
1.602	-0.002	-0.001	-0.004	1.083	1.117	0.969
1.761	-0.005	-0.002	-0.004	1.124	1.187	0.947
1.920	-0.005	-0.002	-0.004	1.152	1.240	0.929
2.080	-0.005	-0.001	-0.004	1.143	1.253	0.912
2.239	-0.002	0.000	-0.004	1.160	1.308	0.887
2.398	-0.006	0.000	-0.004	1.177	1.374	0.857
2.557	-0.002	0.002	-0.005	1.222	1.486	0.822
2.716	-0.005	0.003	-0.008	1.178	1.517	0.777
2.876	0.002	0.003	-0.012	1.237	1.722	0.718
3.035	0.000	-0.002	-0.018	1.292	1.963	0.658

On average the MCSE is about 0.004.

Table 4: Relative biases and efficiency of estimated mrlf's under 40% censoring rate

time	$\frac{B_{mrlf}(t)}{m(t)}$	$\frac{B_{mrlf}(t)}{m(t)}$	$\frac{B_{mrlf}(t)}{m(t)}$	$\frac{MSE_{mrlf}(t)}{MSE_{mrlf}(t)}$	$\frac{MSE_{mrlf}(t)}{MSE_{mrlf}(t)}$	$\frac{MSE_{mrlf}(t)}{MSE_{mrlf}(t)}$
0.010	-0.002	-0.002	-0.014	1.000	0.977	1.024
0.169	-0.002	-0.002	-0.013	1.004	0.990	1.014
0.328	-0.002	-0.002	-0.012	1.006	0.986	1.020
0.488	-0.002	-0.001	-0.011	1.014	1.013	1.001
0.647	-0.002	-0.001	-0.010	1.004	1.001	1.002
0.806	-0.002	-0.001	-0.009	1.021	1.025	0.996
0.965	-0.002	-0.001	-0.008	1.040	1.061	0.980
1.124	-0.004	-0.001	-0.008	1.047	1.077	0.973
1.284	-0.002	-0.002	-0.008	1.060	1.091	0.971
1.443	-0.004	-0.003	-0.009	1.070	1.110	0.964
1.602	-0.006	-0.005	-0.010	1.069	1.126	0.949
1.761	-0.009	-0.007	-0.011	1.110	1.193	0.930
1.920	-0.010	-0.007	-0.012	1.107	1.212	0.913
2.080	-0.009	-0.006	-0.013	1.113	1.246	0.893
2.239	-0.008	-0.006	-0.015	1.161	1.346	0.862
2.398	-0.011	-0.007	-0.019	1.104	1.345	0.820
2.557	-0.009	-0.011	-0.025	1.171	1.515	0.773
2.716	-0.019	-0.020	-0.034	1.178	1.623	0.726
2.876	-0.028	-0.038	-0.049	1.160	1.712	0.678
3.035	-0.055	-0.070	-0.071	1.243	1.977	0.628

On average the MCSE is about 0.005.

Table 5: Relative biases and efficiency of estimated mrlf's under 60% censoring rate

time	$\frac{B_{mrlf}(t)}{m(t)}$	$\frac{B_{mrlf}(t)}{m(t)}$	$\frac{B_{mrlf}(t)}{m(t)}$	$\frac{MSE_{mrlf}(t)}{MSE_{mrlf}(t)}$	$\frac{MSE_{mrlf}(t)}{MSE_{mrlf}(t)}$	$\frac{MSE_{mrlf}(t)}{MSE_{mrlf}(t)}$
0.010	-0.010	-0.010	-0.020	1.000	0.974	1.027
0.169	-0.010	-0.010	-0.021	1.003	0.980	1.023
0.328	-0.011	-0.011	-0.021	1.007	0.986	1.022
0.488	-0.011	-0.011	-0.020	1.010	0.988	1.023
0.647	-0.013	-0.012	-0.021	1.005	0.987	1.018
0.806	-0.014	-0.013	-0.022	1.022	1.014	1.008
0.965	-0.017	-0.016	-0.024	1.020	1.026	0.994
1.124	-0.022	-0.019	-0.027	1.043	1.063	0.981
1.284	-0.023	-0.022	-0.030	1.046	1.073	0.975
1.443	-0.028	-0.026	-0.034	1.030	1.074	0.959
1.602	-0.033	-0.031	-0.039	1.092	1.174	0.930
1.761	-0.039	-0.037	-0.046	1.088	1.201	0.906
1.920	-0.047	-0.044	-0.054	1.095	1.242	0.882
2.080	-0.055	-0.052	-0.065	1.102	1.300	0.848
2.239	-0.058	-0.065	-0.081	1.114	1.394	0.799
2.398	-0.081	-0.087	-0.102	1.132	1.531	0.739
2.557	-0.110	-0.124	-0.131	1.162	1.711	0.680
2.716	-0.155	-0.176	-0.168	1.206	1.921	0.628
2.876	-0.233	-0.246	-0.213	1.196	2.066	0.579
3.035	-0.332	-0.329	-0.266	1.203	2.295	0.524

On average the MCSE is about 0.006.

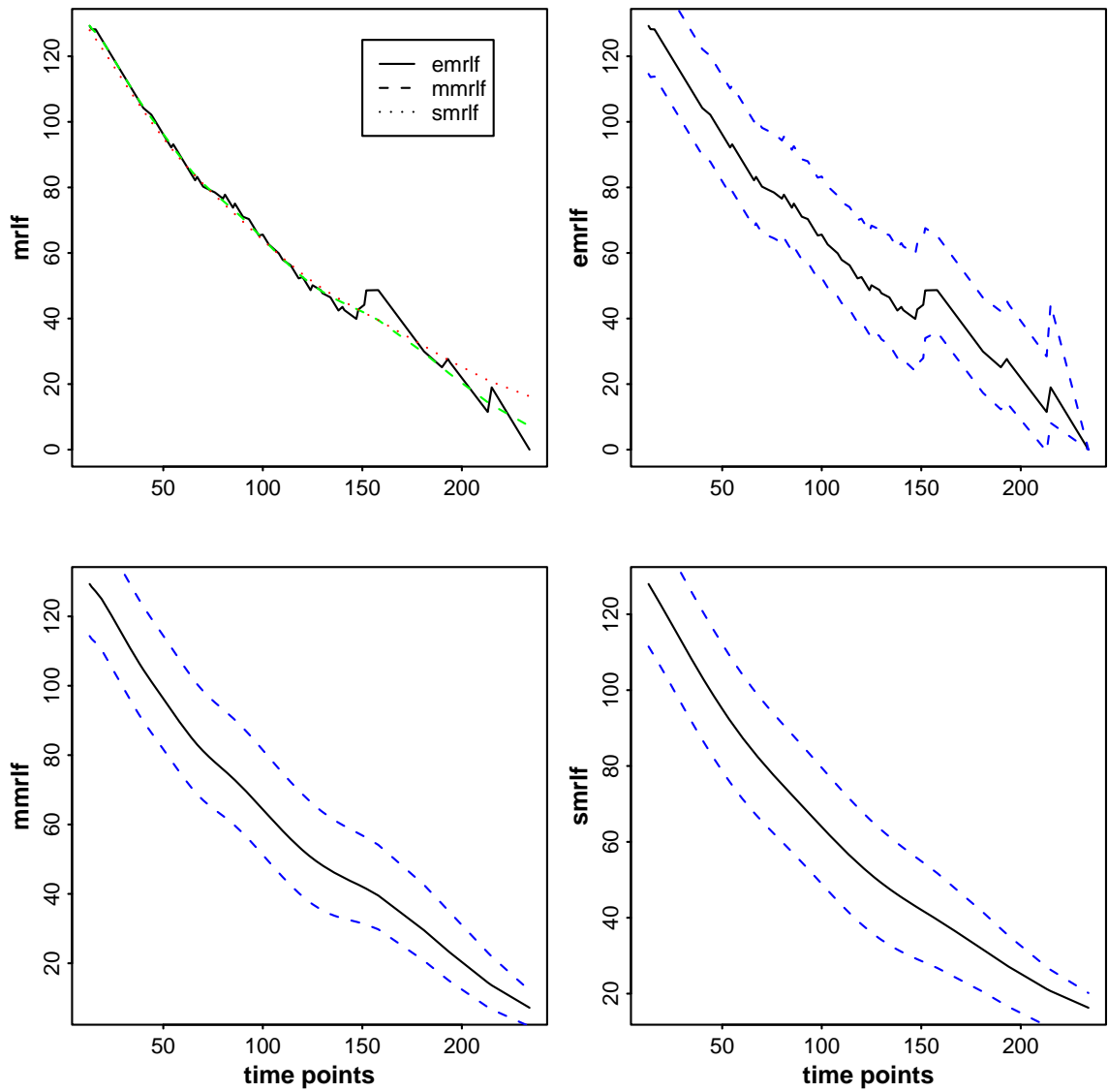


Figure 5: Estimated mrlf's and 95% CI from case resampling for the Melanoma Study

References

- Abdous, B. and Berred, A. (2005). Mean Residual Life Estimation. *Journal of Statistical Planning and Inference* 132, 3-19.
- Agarwal, S.L. and Kalla, S.L. (1996). A Generalized Gamma Distribution and its Application in Reliability, *Communications in Statistics, Theory and Methods* 25, 201-210.
- Chaubey, Y.P. and Sen, P.K. (1999). On Smooth Estimation of Mean Residual Life. *Journal of Statistical Planning and Inference* 75, 223-236.
- Feller W. (1966). *An Introduction to Probability Theory and its Applications*. Vol. II. Wiley, New York.
- Ghorai, J., Susarla, A., Susarla, V., and Van-Ryzin, J. (1982). Nonparametric Estimation of Mean Residual Life Time with Censored Data. *Nonparametric Statistical Inference*. Vol.I, *Colloquia Mathematica Societatis*, 32. North-Holland, Amsterdam-New York, 269-291.
- Gupta, R.C. and Bradley, D.M. (2003). Representing the Mean Residual Life in Terms of the Failure Rate. *Mathematical and Computer Modelling* 37, 1271-1280.
- Gupta, R.C. and Lvin, S. (2005). Monotonicity of Failure Rate and Mean Residual Life Function of a Gamma-type Model. *Applied Mathematics and Computation* 165, 623-633.
- Hall, W.J. and Wellner, J.A. (1981). Mean Residual Life. *Statistics and Related Topics*. Edited by M. Csorgo, D.A. Dawson, J.N.K. Rao , and A.K.Md.E. Saleh. North-Holland, Amsterdam. 169-184.
- Hille, E. (1948). *Functional Analysis and Semigrups*. Am.Math.Soc.Colloq. Pub., Vol.31, American Mathematical Society.
- Kalla, S.L., Al-Saqabi, H.G., Khajah, H.G. (2001). A Unified Form of Gamma-type Distributions. *Applied Mathematics and Computation* 118, 175-187.
- Kaplan, E.L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of American Statistical Association* 53, 457-481.

- Kulasekera, K.B. (1991). Smooth Nonparametric Estimation of Mean Residual Life. *Microelectron. Reliability* 31(1), 97-108.
- Lai, C.D., Zhang, L., Xie, M. (2004). Mean Residual Life and Other Properties of Weibull Related Bathtub Shape Failure Rate Distributions. *International Journal of Reliability, Quality and Safety Engineering* 11, 113-132.
- Petrone, S. and Veronese, P. (2002). Non Parametric Mixture Priors Based on an Exponential Random Scheme. *Statistical Methods and Applications* 11, 105-116.
- Ruiz, J.M. and Guillamon, A. (1996). Nonparametric Recursive Estimator of Residual Life and Vitality Functions under Mixing Dependence Conditions. *Communication in Statistics, Theory and Methods* 4, 1999-2011.
- Tsang, A.H.C. and Jardine, A.K.S. (1993). Estimation of 2-parameter Weibull Distribution from Incomplete Data with Residual Lifetimes. *IEEE Trans. Reliability* 42, 291-298.
- Yang, G.L. (1978). Estimation of a Biometric Function. *Annals of Statistics* 6, 112-116.

Appendix A

Proof of Feller Approximation Lemma:

Let $F_{k,t}$ denote the cumulative distribution function of $Z_k(t)$. Since g is right continuous at t , given $\epsilon > 0$ there exists $\delta_1 > 0$ such that $z \in (t, t + \delta_1) \Rightarrow |g(z) - g(t)| < \epsilon$ and $P[t - \delta_1 < Z_k(t) < t] < \epsilon$. Since g is bounded, we can find a δ_0 such that $|g| \leq M$ for all $z - t \geq \delta_0$. Take $\delta = \max(\delta_1, \delta_0)$, then

$$\begin{aligned} & |E(g(Z_k(t))) - g(t)| \leq E[|g(Z_k(t)) - g(t)|] \\ &= \int |g(z) - g(t)| dF_{k,t}(z) \\ &= \int_{z \in (t, t+\delta)} |g(z) - g(t)| dF_{k,t}(z) + \int_{z \notin (t, t+\delta)} |g(z) - g(t)| dF_{k,t}(z) \\ &\leq \epsilon P[t < Z_k(t) < t + \delta] + 2MP[Z_k(t) < t \text{ or } Z_k(t) > t + \delta] \\ &\leq \epsilon + 2M\{P[|Z_k(t) - t| > \delta] + P[t - \delta_1 < Z_k(t) < t]\} \\ &\leq \epsilon + 2M \left(\frac{\sigma_k^2(t)}{\delta^2} + \epsilon \right) \end{aligned}$$

Let $k \rightarrow \infty$, we get $\sigma_k^2(t) \rightarrow 0$ and

$$\lim_{k \rightarrow \infty} E[|g(Z_k(t)) - g(t)|] \leq \epsilon.$$

But $\epsilon > 0$ is arbitrary, so

$$\lim_{k \rightarrow \infty} E[|g(Z_k(t)) - g(t)|] = 0,$$

i.e.,

$$\lim_{k \rightarrow \infty} E(g(Z_k(t))) = g(t).$$

This completes the proof.

Appendix B

Proof of the Scale Mixture Theory:

Proof: Notice that, for $\theta > 0$,

$$\begin{aligned}\frac{m(t\theta)}{\theta} &= E[T - t\theta | T > t\theta] \frac{1}{\theta} \\ &= E\left[\frac{T}{\theta} - t \mid \frac{T}{\theta} > t\right].\end{aligned}$$

Hence $\frac{m(t\theta)}{\theta}$ is the mrlf of $\frac{T}{\theta}$ and (a) is proved.

Next,

$$\begin{aligned}\int_0^\infty \frac{m(t\theta)}{\theta} \pi(\theta) d\theta &= E_\pi \left[E\left[\frac{T}{\theta} - t \mid \frac{T}{\theta} > t \mid \theta\right]\right] \\ &= E\left[\frac{T}{\theta} - t \mid \frac{T}{\theta} > t\right]\end{aligned}$$

Hence, $\int_0^\infty \frac{m(t\theta)}{\theta} \pi(\theta) d\theta$ is the mrlf of $\frac{T}{\theta}$ where $\frac{T}{\theta} | \theta \sim \frac{m(t\theta)}{\theta}$ and $\theta \sim \pi(\cdot)$, and (b) is proved. Notice that the mrlf uniquely determines a distribution, thus $\frac{T}{\theta} | \theta \sim \frac{m(t\theta)}{\theta}$, which means that the conditional distribution of $\frac{T}{\theta}$ given θ is the distribution induced by $\frac{m(t\theta)}{\theta}$. This completes the proof.

Appendix C

The details of the calculation of $\hat{m}_m(t)$:

$\hat{m}_m(t)$, the smooth estimator of $m(t)$ for censored data (complete data are special case of censored data where weight function $w_j = \frac{1}{n}$), can be calculated with a closed form. The details of the calculation are given here.

$$\begin{aligned}
\hat{m}_m(t) &= \int_0^\infty \hat{m}_e(u) f\left(u \mid k_n, \frac{t}{k_n}\right) du = \sum_{l=0}^{n-1} \int_{X_l}^{X_{l+1}} \hat{m}_e(u) f\left(u \mid k_n, \frac{t}{k_n}\right) du \\
&= \sum_{l=0}^{n-1} \left[\frac{\sum_{j=l+1}^n X_j w_j}{\sum_{j=l+1}^n w_j} \int_{X_l}^{X_{l+1}} f\left(u \mid k_n, \frac{t}{k_n}\right) du - \int_{X_l}^{X_{l+1}} u f\left(u \mid k_n, \frac{t}{k_n}\right) du \right] \\
&= \sum_{l=0}^{n-1} \left[\frac{\sum_{j=l+1}^n X_j w_j}{\sum_{j=l+1}^n w_j} \int_{X_l}^{X_{l+1}} f\left(u \mid k_n, \frac{t}{k_n}\right) du - \int_{X_l}^{X_{l+1}} u \frac{u^{k_n-1} \left(\frac{t}{k_n}\right)^{k_n}}{\Gamma(k_n)} e^{-\left(\frac{t}{k_n}\right)u} du \right] \\
&= \sum_{l=0}^{n-1} \left[\frac{\sum_{j=l+1}^n X_j w_j}{\sum_{j=l+1}^n w_j} \int_{X_l}^{X_{l+1}} f\left(u \mid k_n, \frac{t}{k_n}\right) du - \left(\frac{t}{k_n}\right)^{k_n} \frac{1}{\Gamma(k_n)} \int_{X_l}^{X_{l+1}} u^{k_n} e^{-\left(\frac{t}{k_n}\right)u} du \right] \\
&= \sum_{l=0}^{n-1} \left[\frac{\sum_{j=l+1}^n X_j w_j}{\sum_{j=l+1}^n w_j} \int_{X_l}^{X_{l+1}} f\left(u \mid k_n, \frac{t}{k_n}\right) du - \frac{\Gamma(k_n+1)}{\Gamma(k_n)} \frac{t}{k_n} \int_{X_l}^{X_{l+1}} \frac{u^{k_n} \left(\frac{t}{k_n}\right)^{k_n+1}}{\Gamma(k_n+1)} e^{-\left(\frac{t}{k_n}\right)u} du \right] \\
&= \sum_{l=0}^{n-1} \left[\frac{\sum_{j=l+1}^n X_j w_j}{\sum_{j=l+1}^n w_j} \int_{X_l}^{X_{l+1}} f\left(u \mid k_n, \frac{t}{k_n}\right) du - t \int_{X_l}^{X_{l+1}} f\left(u \mid k_n+1, \frac{t}{k_n}\right) du \right] \\
&= \sum_{l=0}^{n-1} \frac{\sum_{j=l+1}^n X_j w_j}{\sum_{j=l+1}^n w_j} \left[F\left(X_{l+1} \mid k_n, \frac{t}{k_n}\right) - F\left(X_l \mid k_n, \frac{t}{k_n}\right) \right] \\
&\quad - \sum_{l=0}^{n-1} t \left[F\left(X_{l+1} \mid k_n+1, \frac{t}{k_n}\right) - F\left(X_l \mid k_n+1, \frac{t}{k_n}\right) \right] \\
&= \sum_{l=0}^{n-1} \frac{\sum_{j=l+1}^n X_j w_j}{\sum_{j=l+1}^n w_j} \left[F\left(X_{l+1} \mid k_n, \frac{t}{k_n}\right) - F\left(X_l \mid k_n, \frac{t}{k_n}\right) \right] \\
&\quad - t \left[F\left(X_n \mid k_n+1, \frac{t}{k_n}\right) - F\left(X_0 \mid k_n+1, \frac{t}{k_n}\right) \right] \\
&= \sum_{l=0}^{n-1} \frac{\sum_{j=l+1}^n X_j w_j}{\sum_{j=l+1}^n w_j} \left[F\left(X_{l+1} \mid k_n, \frac{t}{k_n}\right) - F\left(X_l \mid k_n, \frac{t}{k_n}\right) \right] - t F\left(X_n \mid k_n+1, \frac{t}{k_n}\right),
\end{aligned}$$

where $w_j = F_n(X_j) - F_n(X_{j-})$ and $F_n(X_j)$ is the KM estimate.