Statistica Sinica **17**(2007), 873-893

A MODIFIED EM-ALGORITHM FOR ESTIMATING THE PARAMETERS OF INVERSE GAUSSIAN DISTRIBUTION BASED ON TIME-CENSORED WIENER DEGRADATION DATA

Ming-Yung Lee and Jen Tang

Chung Shan Medical University and National Tsing-hua University

Abstract: Being the solution to the stochastic linear growth model, the Wiener process has recently been used to model the degradation (or cumulative decay) of certain characteristics of test units in lifetime data analyses. When the failure threshold is constant or linear in time, the failure time, which is defined as the first-passage time of the Wiener process over the failure threshold, will follow an inverse Gaussian (IG) distribution. In this paper we consider a time-censored degradation test where, in addition to the failure times of the failed units, we assume that the degradation values at the censor time of the censored units are also available. Then, based on these degradation values, we use a modified EM-algorithm to predict the failure times of the censored units. The resulting estimator of the mean failure time is shown to be a consistent estimator, and is also an estimator that maximizes the (modified) likelihood function of the available failure times and degradation values. For the scale parameter of the IG distribution, however, the algorithm produces an inconsistent estimator. We introduce two modified estimators to reduce bias. Analytical and numerical comparisons show that our proposed estimators perform well, as compared to the traditional MLEs and the modified MLEs, for both IG parameters. An example is used to illustrate the proposed methodology.

Key words and phrases: Bias, consistency, degradation test, EM-algorithm, inverse Gaussian distribution, maximum likelihood estimator, reliability, Wiener process.

1. Introduction

Assume that a product has a critical quality characteristic (QC) whose degradation sample path, $\{W(t) \mid t \geq 0\}$, follows a Wiener process

$$W(t) = \eta t + \sigma B(t),\ \ t \geq 0, \tag{1}$$

where η is the drift parameter, $\sigma > 0$ is the diffusion coefficient, and $B(\cdot)$ is a standard Brownian motion. Being a continuous-time version of the discrete-time cumulative sum (CUSUM) process, $W(t)$ is the solution of the stochastic linear growth model $dW_t = \eta dt + \sigma dB(t)$. The product's failure time (or lifetime), denoted T, is then defined as the first-passage time of $W(t)$(with $\eta > 0$) over a constant failure threshold $a(> 0)$, i.e.,

$$T = \inf\{t \geq 0 \mid W(0) = 0, W(t) \geq a\}. \tag{2}$$

It is well-known that T follows an inverse Gaussian distribution, denoted by $IG(\mu, \lambda)$, with the location and scale parameters

$$\mu = \frac{a}{\eta}, \text{ and } \lambda = \frac{a^2}{\sigma^2}, \tag{3}$$

respectively. The p.d.f. and c.d.f. of T are given, respectively, by

$$f(t) = \sqrt{\frac{\lambda}{2\pi t^3}} \exp\Big\{ -\frac{\lambda(t-\mu)^2}{2\mu^2 t}\Big\}, \quad t > 0, \tag{4}$$

$$F(t) = \Phi\Big(\sqrt{\frac{\lambda}{t}}\Big(\frac{t}{\mu} - 1\Big)\Big) + \exp\Big(\frac{2\lambda}{\mu}\Big)\Phi\Big\{ -\sqrt{\frac{\lambda}{t}}\Big(\frac{t}{\mu} + 1\Big)\Big\}, \quad t > 0. \tag{5}$$

Chhikara and Folks (1989) and Seshadri (1999) provide systematic overviews on the IG distribution.

The Wiener/IG model has found applications in certain studies. For example, Sherif and Smith (1980) and Bhattacharyya and Fries (1982) consider a fatigue failure model in which accumulated decay is governed by a Wiener process. Whitmore and Schenkelberg (1997) also use a time-transformed Wiener process to model resistance of self-regulating heating cables. Doksum and Normand (1995) assume a Wiener process for the level of a biomarker process such as calibrated log CD4 blood cell counts in their HIV study. Singpurwalla (1995) also proposes a Wiener process for cracks caused by fatigue (with healing). Tseng, Tang and Ku (2003) suggest a transformed Wiener process for light intensity of light emitting diode (LED) lamps for contact image scanners (CISs). Doksum and Hoyland (1992) use a time-transformed Wiener process to model an accelerated degradation sample path.

If n independent units are tested until failure to obtain the complete failure times $t_1, \ldots, t_n$ on T, then the maximum likelihood estimators (MLEs) of μ and λ are

$$\hat{\mu} = \overline{T}, \tag{6}$$

$$\hat{\lambda} = \frac{n}{V}, \tag{7}$$

where $\overline{T} = \sum_{i=1}^{n} T_i/n$ and $V = \sum_{i=1}^{n}(1/T_i - 1/\overline{T})$. Since IG distributions form an exponential family, $(\overline{T}, \sum_{i=1}^{n} 1/T_i)$ is a minimal complete sufficient set of statistics for (μ, λ), and hence the uniform minimum variance unbiased estimators (UMVUEs) of μ and λ are $\overline{T}$ and $(n-3)/V$, respectively (Seshadri (1999, Chap.2)).

However, censoring is often implemented in practice, especially when testing highly reliable products. Furthermore, for estimating the IG parameter, one often assumes that only failure times of the failed units and the censor times of the censored units are available.

In this paper we consider a time-censored degradation test where, in addition to the common censor time, we assume the degradation values of the censored units at the censor time are also available. Then, based on these degradation values, a modified EM algorithm is proposed in Section 2 to predict the corresponding (expected) failure times; a closed-form estimate of the IG mean failure time is obtained in Section 2.1, denoted MEME. This estimator, which is also a modified maximum likelihood estimator (MMLE), is shown in Section 2.2 to be a consistent estimator of the true mean failure time. For the scale parameter of the IG distribution, however, the algorithm produces an inconsistent estimator. We introduce two methods for reducing the bias. We also discuss the MMLE of the IG scale parameter, and the results are given in Section 3. In Section 4 we first demonstrate that, by incorporating the degradation values of the censored units, the resulting MMLEs indeed reduce the asymptotic variances of the traditional MLEs for both IG parameters. We then numerically evaluate the performances of the MEMEs, MMLEs, and MLEs for both IG parameters for small-sample cases. We also study the performances of the MEMEs under the two bias reduction techniques given in Section 3. An example on LED lamps is used in Section 5 to illustrate the proposed methodology. Concluding remarks are given in Section 6.

2. Modified EM-Algorithm and Estimate of the Mean Failure Time

Assume that n independent units were tested and, by fixed time τ, m of them have failed, with failure times $t_1, \ldots, t_m$, and the remaining units were censored with degradation values, $w_{m+1}(\tau), \ldots, w_n(\tau)$, at time τ. Here m is an observed value of the random variable M for the member of uncensored units in this time-censored test.

The following procedure is used to estimate the IG parameters.

Step 1. Predict the failure times, denoted $t^{(1)}_{m+1}, \ldots, t^{(1)}_n$, for the censored units based on their degradation values at τ. The prediction is made using a modified expectation maximization (EM) alothrithm.

Step 2. Use the complete pseudo failure times, $t_1, \ldots, t_m$, $t^{(1)}_{m+1}, \ldots, t^{(1)}_n$ to obtain the pseudo UMVUEs, $\hat{\Theta}^{(1)}$, of the IG parameters.

Step 3. At iteration k, use the estimates of the IG parameters from the previous step to update the predicted failure times, denoted by $t_{m+1}^{(k)}, \ldots, t_n^{(k)}$, for the censored units.

Step 4. Use the complete pseudo failure times, $t_1, \ldots, t_m$, $t_{m+1}^{(k)}, \ldots, t_n^{(k)}$, to find the pseudo UMVUE, $\hat{\Theta}^{(k)}$, of the IG parameters.

Steps 3 and 4 are repeated until the limit of $\hat{\Theta}^{(k)}$, as $k \to \infty$, can be obtained, or the elements of $|\hat{\Theta}_{k+1} - \hat{\Theta}_k|$ are all smaller than a pre-determined tolerance. The estimates thus obtained will be called the modified EM estimates (MEMEs).

2.1. The MEME of μ

We apply the above procedure to obtain a closed-form estimate for the IG mean. For Step 1, since the expected degradation path is linear in time with slope η, the line through $(0, w_i(0))$ and $(\tau, w_i(\tau))$ will pass the failure threshold a at

$$t_i^{(1)} = \frac{a\tau}{w_i(\tau)}, \quad \text{for } i = m+1, \ldots, n. \tag{8}$$

The complete pseudo failure times, $t_1, \ldots, t_m$, $t_{m+1}^{(1)}, \ldots, t_n^{(1)}$, are then used in Step 2 to obtain the first pseudo UMVU estimate of μ, using (6),

$$\begin{aligned} \hat{\mu}^{(1)} = \bar{t}^{(1)} &= \frac{1}{n}\Big(\sum_{i=1}^{m} t_i + \sum_{i=m+1}^{n} t_i^{(1)}\Big) \\ &= \frac{1}{n}\sum_{i=1}^{m} t_i + \frac{a\tau}{n}\sum_{i=m+1}^{n} \frac{1}{w_i(\tau)}. \end{aligned} \tag{9}$$

Note that the estimated slope is $\hat{\eta}^{(1)} = a/\hat{\mu}^{(1)}$. For Step 3, with $k = 2$, we again use a linear degradation path after τ to obtain the predicted failure times for censored units as

$$\begin{aligned} t_i^{(2)} &= \tau + \frac{a - w_i(\tau)}{\hat{\eta}^{(1)}} \\ &= \tau + \frac{a - w_i(\tau)}{na}\Bigg(\sum_{j=1}^{m} t_j + \sum_{j=m+1}^{n} \frac{a\tau}{w_j(\tau)}\Bigg), \quad \text{for } i = m+1, \ldots, n. \end{aligned}$$

Hence for Step 4 we have, using (6),

$$\hat{\mu}^{(2)} = \frac{1}{n}\Bigg(\sum_{i=1}^{m} t_i + \sum_{i=m+1}^{n} t_i^{(2)}\Bigg)$$

$$= \frac{1}{n}\sum_{i=1}^{m} t_i + \frac{1}{n}\sum_{i=m+1}^{n}\Big(\tau + \Big(1 - \frac{w_i(\tau)}{a}\Big)\hat{\mu}^{(1)}\Big)$$
$$= C + D \cdot E, \quad \text{say,}$$

where $C = \sum_{i=1}^{m} t_i/n + (1 - m/n)\tau$, and $D = (1 - m/n) - \sum_{i=m+1}^{n} w_i(\tau)/an$, and $E = (\sum_{i=1}^{m} t_i + a\tau \sum_{i=m+1}^{n} 1/w_i(\tau))/n$.

If we iterate the algorithm k times, we obtain

$$\hat{\mu}^{(k)} = C + D(C + D(C + D(C + D(\cdots + D(C + D \cdot E))))), \tag{10}$$

where there are $(k-1)C's$, $(k-1)D's$, and one E. Because $0 < D < 1$, with probability $1, \hat{\mu}^{(k)}$ in (10) will converge with probability 1 as $k \to \infty$ to a limit denoted by $\hat{\mu}_n$ (subscript n is added to denote the fixed sample size here), where

$$\hat{\mu}_n = \frac{C}{1-D} = \frac{\frac{1}{n}\sum_{i=1}^{m} t_i + \frac{1}{n}(n-m)\tau}{\frac{1}{n}(m + \sum_{i=m+1}^{n} \frac{w_i(\tau)}{a})} = \frac{\sum_{i=1}^{m} t_i + (n-m)\tau}{m + \sum_{i=m+1}^{n} \frac{w_i(\tau)}{a}}. \tag{11}$$

To interpret (11) we first note that, based on the argument used to obtain (8), the average failure time of the censored units is a weighted average of their expected failure times $\tau/(w_j(\tau)/a)$, $j = m+1, \ldots, n$, weighted by $(w_j(\tau)/a)/[\sum_{i=m+1}^{n} (w_i(\tau)/a)]$ for the jth unit. This average turns out to be $(n-m)\tau/\sum_{i=m+1}^{n} (w_i(\tau)/a)$. The average failure time of the uncensored units is $\sum_{i=1}^{m} t_i/m$. The estimate in (11) is the weighted average of these two sub-averages.

Modified Maximum Likelihood Estimate (MMLE) of Mean Failure Time

For the mean failure time, the estimate (11) happens to be an MMLE. To see this and to compare the MMLE with the traditional MLE, we take $\delta_i = 1$ if $T_i \le \tau$, $\delta_i = 0$ if $T_i > \tau$. Then the conditional distribution of T_i, given that $\delta_i = 1$, is $P(T_i = t_i \mid T_i \le \tau) = f(t_i)/F(\tau)$, for $t_i \le \tau$. Thus the joint distribution of T_i at t_i and $\delta_i = 1$ is

$$P(T_i = t_i, T_i \le \tau) = P(T_i = t_i \mid T_i \le \tau)P(T_i \le \tau)$$
$$= \frac{f(t_i)}{F(\tau)}F(\tau) = f(t_i), \quad t_i < \tau,$$

and the joint likelihood for the uncensored observations is $\prod_{i=1}^{m} f(t_i)$. For the censored observations, we have

$$P(W_i(\tau) = w_i, \delta_i = 0) = P(W_i(\tau) = w_i, T_i > \tau)$$

$$= P(W_i(\tau) = w_i, \sup_{0\le t\le\tau} W_i(t) < a)$$
$$= \frac{1}{\sqrt{\frac{2\pi\tau a^2}{\lambda}}} \exp\Big\{ -\frac{\lambda(w_i - \frac{a\tau}{\mu})^2}{2\tau a^2}\Big\}\Big(1 - \exp\Big\{ -\frac{2\lambda(a - w_i)}{a\tau}\Big\}\Big)$$
$$= h(w_i), \quad \text{say,} \quad \text{for } 0 \le w_i < a,\ i = m+1, \ldots, n. \tag{12}$$

A simple proof of (12) is given in the Appendix; also see (Cox and Miller (1965, Chap.5)). Then the modified likelihood function is the joint distribution of the given data:

$$\prod_{i=1}^{m} f(t_i) \prod_{i=m+1}^{n} h(w_i). \tag{13}$$

The MMLE, which maximizes (13), of the mean failure time μ can be shown to be same as (11). Padgett and Tomlinson (2004) also consider (13), but they misinterpreted $h(w_i)$ as a conditional distribution. If one has only τ for the censored units, then traditional likelihood function is (Lawless (1982, Chap.2))

$$\prod_{i=1}^{n} f(t_i)^{\delta_i}(1 - F(\tau))^{1-\delta_i} = \Big(\prod_{i=1}^{m} f(t_i)\Big)(1 - F(\tau))^{n-m}. \tag{14}$$

2.2. The Consistency of the MEME/MMLE of μ

In this section we prove the large-sample property of the MEME/MMLE, $\hat{\mu}_n$, of (11), where failure times, degradation values, and failure number are all treated as random. First recall from the last section that the conditional distribution of $T_1, \ldots, T_M$ is a truncated $IG(\mu, \lambda)$. The censor number M is binomial with failure probability $F(\tau)$. The conditional joint p.d.f. of $T_1, \ldots, T_M$ and M, given $T_i \le \tau$, $i = 1, \ldots, M$, and $M \ge 1$, is

$$f(t_1, \ldots, t_m, m \mid T_i \le \tau, i = 1, \ldots, M, M \ge 1)$$
$$= f(t_1, t_2, \ldots, t_m \mid T_i \le \tau, i = 1, \ldots, M = m \ge 1) P(M = m \mid M \ge 1)$$
$$= \binom{n}{m} \frac{(1-F(\tau))^{n-m}}{1-(1-F(\tau))^n} \prod_{i=1}^{m} f(t_i), \quad \text{for } 0 < t_1, \ldots, t_m \le \tau \text{ and } m = 1, \ldots, n,$$

where $f(\cdot)$ and $F(\cdot)$ are given in (4) and (5), respectively. Then the conditional expectation $\sum_{i=1}^{M} T_i$ is (in what follows, $T_i \le \tau$ means $T_i \le \tau$, $i = 1, \ldots, M$)

$$E(\sum_{i=1}^{M} T_i \mid T_i \le \tau)$$

$$= E(\sum_{i=1}^{M} T_i \mid T_i \le \tau, M \ge 1)P(M \ge 1) + E(\sum_{i=1}^{M} T_i \mid T_i \le \tau, M = 0)P(M = 0)$$
$$= E(\sum_{i=1}^{M} T_i \mid T_i \le \tau, M \ge 1)P(M \ge 1)$$
$$= n\int_0^{\tau} tf(t)dt$$
$$= n\mu\Big\{\Phi\Big(\sqrt{\frac{\lambda}{\mu}}(\frac{\tau}{\mu} - 1)\Big) - e^{\frac{2\lambda}{\mu}}\Big[1 - \Phi\Big(\sqrt{\frac{\lambda}{\mu}}(\frac{\tau}{\mu} + 1)\Big)\Big]\Big\}. \quad (15)$$

Similarly, by the i.i.d. property of $T_i's$, we have

$$E\Big((\sum_{i=1}^{M} T_i^2) \mid T_i \le \tau\Big) = E\Big(\sum_{i=1}^{M} T_i^2 + \sum_{i=1}^{M}\sum_{i\neq j}^{M} T_iT_j \mid T_k \le \tau, k = 1, \ldots, M\Big)$$
$$= n\int_0^{\tau} t^2 f(t)dt + n(n-1)\Big(\int_0^{\tau} tf(t)dt\Big)^2.$$

Then

$$\mathrm{Var}\Big(\sum_{i=1}^{M} T_i \mid T_i \le \tau\Big) = E\Big((\sum_{i=1}^{M} T_i \mid T_i \le \tau)^2\Big) - \Big(E\Big(\sum_{i=1}^{M} T_i \mid T_i \le \tau\Big)\Big)^2$$
$$= n\int_0^{\tau} t^2 f(t)dt + n(n-1)\Big(\int_0^{\tau} tf(t)dt\Big)^2 - n^2\Big(\int_0^{\tau} tf(t)dt\Big)^2$$
$$= n\Big[\int_0^{\tau} t^2 f(t)dt - \Big(\int_0^{\tau} tf(t)dt\Big)^2\Big].$$

So the conditional variance of $n^{-1}\sum_{i=1}^{M}(T_i \mid T_i < \tau)$ converges to 0 as $n \to \infty$, since $\int_0^{\tau} t^2 f(t)dt$ and $\int_0^{\tau} tf(t)dt$ are both finite. Thus $n^{-1}\sum_{i=1}^{M}(T_i \mid T_i < \tau)$ converges in probability to $\int_0^{\tau} tf(t)dt$.

For the censored units ($i = M+1, \ldots, n$), the conditional distribution of $W_i(\tau)$, given that $T_i > \tau$, is

$$P(W_i(\tau) = w_i \mid T_i > \tau) = P\Big(W_i(\tau) = w_i \mid \sup_{0\le t\le \tau} W_i(t) < a\Big) = \frac{h(w_i)}{1 - F(\tau)},$$

where $h(w_i)$ is given in (12). We have the conditional joint distribution of $W_{M+1}(\tau), \ldots, W_n(\tau)$ and failure number M as

$$f(w_{m+1}, \ldots, w_n, m \mid M \le n-1, T_j > \tau, j = M+1, \ldots, n)$$
$$= \frac{h(w_{m+1})}{1 - F(\tau)} \cdots \frac{h(w_n)}{1 - F(\tau)} \binom{n}{m} F(\tau)^m (1 - F(\tau))^{n-m} \frac{1}{P(M \le n-1)}$$

$$= \Big(\prod_{i=m+1}^{n} h(w_i)\Big) F(\tau)^m \binom{n}{m} \frac{1}{P(M \le n-1)}, \text{ for } 0 \le m \le n-1$$
$$\text{and } 0 < w_{m+1}, \ldots, w_n < a.$$

Now, conditioned on $T_i > \tau$ for $i = M+1, \ldots, n$, we have

$$\begin{aligned}
&E\Big(\sum_{i=M+1}^{n} W_i(\tau) \mid T_i > \tau\Big) \\
&\quad = E\Big(\sum_{i=M+1}^{n} W_i(\tau) \mid T_i > \tau, M \le n-1\Big) P(M \le n-1) \\
&\quad = \sum_{m=0}^{n-1} \binom{n}{m} (F(\tau))^m \int_{-\infty}^{a} \cdots \int_{-\infty}^{a} \Big(\sum_{i=m+1}^{n} w_i h(w_i)\Big) dw_{m+1} \cdots dw_n \\
&\quad = n\Big(F(\tau) + \int_{-\infty}^{a} h(w)dw\Big)^{n-1} \int_{-\infty}^{a} wh(w)dw, \qquad \text{(by i.i.d. of } W_i(\tau)) \\
&\quad = n \int_{-\infty}^{a} wh(w)dw \qquad \Big(\text{by } \int_{-\infty}^{a} h(w)dw = 1 - F(\tau)\Big) \\
&\quad = na\Big\{\frac{\tau}{\mu}\Big(1 - \Phi\Big(\sqrt{\frac{\lambda}{\tau}}(\frac{\tau}{\mu} - 1)\Big)\Big) - (2 + \frac{\tau}{\mu}) e^{\frac{2\lambda}{\mu}} \Phi\Big(-\sqrt{\frac{\lambda}{\tau}}\Big(\frac{\tau}{\mu} + 1\Big)\Big)\Big\}. \qquad (16)
\end{aligned}$$

Furthermore, by the i.i.d. property of the $W_i(\tau)'s$, we have

$$\begin{aligned}
&E\Big(\Big(\sum_{i=M+1}^{n} W_i(\tau)\Big)^2 \mid T_i > \tau\Big) \\
&= E\Big(\sum_{i=M+1}^{n} W_i(\tau)^2 + \sum_{\substack{i,j=M+1 \\ i \ne j}}^{n} W_i(\tau) W_j(\tau) \mid T_i > \tau, M \le n-1\Big) \cdot P(M \le n-1) \\
&= \sum_{m=0}^{n-1} \int_{-\infty}^{a} \cdots \int_{-\infty}^{a} \Big(\sum_{i=m+1}^{n} w_i^2 + \sum_{\substack{i,j=M+1 \\ i \ne j}}^{n} w_i w_j\Big) \binom{n}{m} \cdot (F(\tau))^m \Big(\prod_{i=m+1}^{n} h(w_i) dw_i\Big) \\
&= n \int_{-\infty}^{a} w^2 h(w)dw + n(n-1)\Big(\int_{-\infty}^{a} wh(w)dw\Big)^2.
\end{aligned}$$

Then

$$\begin{aligned}
&\mathrm{Var}\Big(\frac{\sum_{i=M+1}^{n} W_i(\tau)}{n} \mid T_i > \tau\Big) \\
&\qquad = \frac{1}{n} \int_{-\infty}^{a} w^2 h(w)dw - \frac{1}{n}\Big(\int_{-\infty}^{a} wh(w)dw\Big)^2 \to 0, \text{ as } n \to \infty.
\end{aligned}$$

That is, we have shown that

$$\frac{1}{n}\sum_{i=M+1}^{n} W_i(\tau) \xrightarrow{P} \int_{-\infty}^{a} wh(w)dw, \text{ as } n \to \infty.$$

Now, since $(1/n)\sum_{i=1}^{M} T_i + [(n-M)/n]\tau \xrightarrow{P} \int_{-\infty}^{a} tf(t)dt + (1-F(\tau))\tau$ and $M/n + (1/an)\sum_{i=M+1}^{n} W_i(\tau) \xrightarrow{P} F(\tau) + (1/a)\int_{-\infty}^{a} wh(w)dw$, by the Slusky Theorem we have

$$\hat{\mu}_n = \frac{\sum\limits_{i=1}^{M} T_i + (n-M)\tau}{M + \frac{1}{a}\sum\limits_{i=M+1}^{n} W_i(\tau)} \xrightarrow{P} \frac{\int_{-\infty}^{a} tf(t)dt + (1-F(\tau))\tau}{F(\tau) + \frac{1}{a}\int_{-\infty}^{a} wh(w)dw},$$

which can be shown to be μ, using (15) and (16). That is, $\hat{\mu}_n \xrightarrow{P} \mu$ as $n \to \infty$.

3. Estimates of the Scale Parameter λ of IG Distribution

3.1. The MEME of λ

Since the UMVUE of λ is $(n-3)/V$ when complete failure data are available, we have an estimate of λ for iteration k as $\hat{\lambda}^{(k)} = (n-3)/\hat{v}^k$, where

$$\begin{aligned}\hat{v}^{(1)} &= \sum_{i=1}^{m}\frac{1}{t_i} + \sum_{i=m+1}^{n}\frac{1}{t_i^{(1)}} - \frac{n}{\bar{t}^{(1)}}\\ &= \sum_{i=1}^{m}\frac{1}{t_i} + \frac{1}{a\tau}\sum_{i=m+1}^{n} w_i(\tau) - \frac{n}{\hat{\mu}^{(1)}},\end{aligned}$$

and, for $k \geq 2$,

$$\begin{aligned}\hat{v}^{(k)} &= \sum_{i=1}^{m}\frac{1}{t_i} + \sum_{i=m+1}^{n}\frac{1}{t_i^{(k)}} - \frac{n}{\bar{t}^{(k)}}\\ &= \sum_{i=1}^{m}\frac{1}{t_i} + \sum_{i=m+1}^{n}\frac{1}{\tau + \left(1 - \frac{w_i(\tau)}{a}\right)\hat{\mu}^{(k)}} - \frac{n}{\hat{\mu}^{(k)}}.\end{aligned}$$

Since $\hat{\mu}^{(k)}$ converges to $\hat{\mu}_n$ as $k \to \infty$, $t_i^{(k)}$ will converge to $\tau + (1 - w_i(\tau)/a)\hat{\mu}_n$. Thus, as $k \to \infty$, $\hat{v}^{(k)}$ will converge to $\hat{v}_n$ say, where

$$\hat{v}_n = \sum_{i=1}^{m}\frac{1}{t_i} + \sum_{i=m+1}^{n}\frac{1}{\tau + \left(1 - \frac{w_i(\tau)}{a}\right)\hat{\mu}_n} - \frac{n}{\hat{\mu}_n}. \tag{17}$$

Hence, an MEME of λ is

$$\hat{\lambda}_n = \frac{n-3}{\hat{v}_n}. \tag{18}$$

The MMLE, $\hat{\lambda}'_n$, can be shown from (13) to satisfy

$$\hat{\lambda}'_n = n\left(\sum_{i=1}^{m}\frac{(t_i-\hat{\mu}_n)^2}{2\hat{\mu}_n^2 t_i}+\sum_{j=m+1}^{n}\frac{(w_j-\frac{a\tau}{\hat{\mu}_n})^2}{a^2\tau}-\sum_{j=m+1}^{n}\frac{\frac{4(a-w_j)}{a\tau}\exp\left\{-\frac{2\hat{\lambda}'_n(a-w_j)}{a\tau}\right\}}{1-\exp\left\{-\frac{2\hat{\lambda}'_n(a-w_j)}{a\tau}\right\}}\right)^{-1}. \tag{19}$$

There is no closed-form solution for $\hat{\lambda}'_n$, so we have to rely on numerical procedures to compute it. Note that, unlike the case for μ, the MEME (in (18)) and the MMLE (in (19)) of λ are not the same.

3.2. Inconsistency of $\hat{\lambda}_n$ and bias reduction

In this section we demonstrate that the MEME $\hat{\lambda}_n(=(n-3)/\hat{V}_n)$ of (18) converges in probability to a value different from λ. First note from (17) that

$$\frac{\hat{V}_n}{n} = \frac{M}{n}\frac{1}{M}\sum_{i=1}^{M}\frac{1}{T_i}+\frac{n-M}{n}\frac{1}{n-M}\sum_{i=M+1}^{n}\frac{1}{\tau+\left(1-\frac{W_i(\tau)}{a}\right)\hat{\mu}_n}-\frac{1}{\hat{\mu}_n}. \tag{20}$$

For each fixed $W_i(\tau)$, $[\tau+(1-W_i(\tau)/a)\hat{\mu}_n]^{-1}\to[\tau+(1-W_i(\tau)/a)\mu]^{-1}$ in probability as $n\to\infty$, and hence we expect that $\sum_{i=M+1}^{n}[\tau+(1-W_i(\tau)/a)\hat{\mu}_n]^{-1}/(n-M)$ will converge in probability to $E^*_\tau([\tau+(1-W(\tau)/a)\mu]^{-1})$, where $W(\tau)$ and $W_i(\tau)$ are i.i.d., and the expectation E^*_τ denotes the joint expectation with $I(T>\tau)$, or $I(\sup_{0<t<\tau}W(t)<a)$. Also, $\sum_{i=1}^{M}(1/T_i)/M$ will converge to $E(1/T\mid T\le\tau)=\int_0^\tau t^{-1}f(t)dt/F(\tau)$, where T and T_i, $1\le i\le M$, have the same distribution. Since $E(1/T)=1/\mu+1/\lambda$, if $E^*_\tau(1/T)=E^*_\tau(1/[\tau+(1-W(\tau)/a)\mu])$ then $\hat{\lambda}_n\xrightarrow{P}\lambda$. But the last equality is not true in general, implying that the MEME is asymptotically biased with

$$\begin{aligned}
&E^*_\tau(\frac{1}{T})-E^*_\tau(\frac{1}{\tau+\left(1-\frac{W(\tau)}{a}\right)\mu})\\
&=-\int_\tau^\infty\frac{1}{y}\frac{\sqrt{\lambda}}{\sqrt{2\pi\tau\mu^2}}\left\{\exp\left(-\frac{\lambda(y-\mu)^2}{2\tau\mu^2}\right)-\exp\left(\frac{2\lambda}{\mu}\right)\exp\left(-\frac{\lambda(y+\mu)^2}{2\tau\mu^2}\right)\right\}dy\\
&\quad+\int_\tau^\infty\frac{1}{t}\frac{\sqrt{\lambda}}{\sqrt{2\pi}}t^{\frac{-3}{2}}\exp\left(-\frac{\lambda(t-\mu)^2}{2\mu^2 t}\right)dy,
\end{aligned} \tag{21}$$

which is obviously not equal to 0. The expectations in (21) are obtained using the joint distribution of $W(\tau) = w$ and $T > \tau$, which is $h(w)$ in (12).

To reduce the bias, first define $Y = \tau + \mu(1 - W(\tau)/a)$ and then, by Taylor series expansion of $1/Y$ about μ, we have

$$E_\tau^*(\frac{1}{Y}) = \frac{1}{\mu} - \frac{1}{\mu^2}E_\tau^*(Y-\mu) + \frac{1}{\mu^3}E_\tau^*(Y-\mu)^2 - \frac{1}{\mu^4}E_\tau^*(Y-\mu)^3 \cdots . \tag{22}$$

Now, from the definition of Y and the normality of $W(\tau)$, $E(Y-\mu)^k$ is zero for odd k and proportional to $(\tau\mu^2/\lambda)^k$ for even k, which may not converge as $k \to \infty$ in some special cases. However, in industrial applications, λ and μ are typically large such that $\mathrm{Var}\,(T) = \mu^3/\lambda$ is relatively small. With this and the fact that $E_\tau^*(Y-\mu)^k \le E(Y-\mu)^k$, we may truncate (22) since the higher order terms are negligible.

Define $A = \sqrt{\lambda/\tau}(\tau/\mu - 1)$ and $B = \sqrt{\lambda/\tau}(\tau/\mu + 1)$. Then with some algebras we can show (with first three terms in (22))

$$\begin{aligned} E_\tau^*(\frac{1}{Y}) - E_\tau^*(\frac{1}{T}) &\approx \frac{\tau}{\lambda\mu}\Big\{[1-\Phi(A)] + \exp\Big(\frac{2\lambda}{\mu}\Big)[1-\Phi(B)]\Big\} \\ &\quad -\frac{1}{\lambda}\Big\{1-\Phi(A) - \exp\Big(\frac{2\lambda}{\mu}\Big)\Phi(-B)\Big\} + 2\Big(\frac{\tau^2}{\lambda\mu} + \frac{1}{\sqrt{\lambda\tau}}\Big)\phi(A) \\ &\approx \frac{\tau}{\lambda\mu}\Big\{[1-\Phi(A)] + \exp\Big(\frac{2\lambda}{\mu}\Big)[1-\Phi(B)]\Big\} \\ &\quad -\frac{1}{\lambda}\Big\{1-\Phi(A) - \exp\Big(\frac{2\lambda}{\mu}\Big)\Phi(-B)\Big\}. \end{aligned} \tag{23}$$

Remark 3.2.1. (The MEME3). By (23) we modify the $\hat{V}_n$ in MEME as

$$\begin{aligned} \hat{V}_n = &\sum_{i=1}^{m}\frac{1}{T_i} + \sum_{i=m+1}^{n}\frac{1}{\tau + \Big(1 - \frac{W_i(\tau)}{a}\Big)\hat{\mu}_n} - \frac{n}{\hat{\mu}_n} - (1-\hat{F}(\tau)) \\ &\times \frac{\tau}{\hat{\lambda}_n\hat{\mu}_n}\Big\{\{[1-\Phi(\hat{A})] + \exp\frac{2\hat{\lambda}_n}{\hat{\mu}_n}[1-\Phi(\hat{B})]\} - \frac{1}{\hat{\lambda}_n}\{1-\Phi(\hat{A}) - \exp\frac{2\hat{\lambda}_n}{\hat{\mu}_n}\Phi(-\hat{B})\}\Big\}, \end{aligned} \tag{24}$$

where $\hat{A} = \sqrt{\hat{\lambda}_n/\tau}(\tau/\hat{\mu}_n - 1)$ and $\hat{B} = \sqrt{\hat{\lambda}_n/\tau}(\tau/\hat{\mu}_n + 1)$, and $\hat{\lambda}_n$ and $\hat{\mu}_n$ are the unmodified MEM estimates.

Remark 3.2.2. (The MEME2). We may correct some bias in the modified EM procedures by modifying the predictions of failure times for the censored units as follows. Let t_i' be the additional time (after τ) to failure (i.e., to reach the failure threshold a) for the censored unit $i(\ge M+1)$ if we were to continue the degradation test. Then $w_i(\tau)+\eta t_i'+\sigma B(t_i')=a$, where $B(t_i')$ is (a realization of) the standard Brownian motion. Since the random $B(t_i')/\sqrt{t_i'}$ follows the standard normal distribution, we propose to replace this quantity by $\hat{B}_i=\Phi^{-1}((i-m)/(n-m+1))$, the $[(i-m)/(n-m)]$-the quantile of the standard normal distribution. So

$$w_i(\tau)+\hat{\eta}t_i'+\sqrt{t_i'}\hat{\sigma}\hat{B}_i=a, \tag{25}$$

or

$$t_i'=\hat{\mu}_n(1-\frac{w_i(\tau)}{a})+\frac{\hat{\mu}_n^2}{2}\Big(\frac{\hat{B}_i^2}{\hat{\lambda}_n}\mp\frac{\hat{B}_i}{\sqrt{\hat{\lambda}_n}}\sqrt{\frac{\hat{B}_i^2}{\hat{\lambda}_n}+\frac{4(a-w_i(\tau))}{a\hat{\mu}_n}}\Big).$$

Thus we have the predicted failure time for the ith unit as

$$T_i'=\tau+\hat{\mu}_n(1-\frac{W_i(\tau)}{a})+\frac{\hat{\mu}_n^2}{2}\Big(\frac{\hat{B}_i^2}{\hat{\lambda}_n}\mp\frac{\hat{B}_i}{\sqrt{\hat{\lambda}_n}}\sqrt{\frac{\hat{B}_i^2}{\hat{\lambda}_n}+\frac{4(a-W_i(\tau))}{a\hat{\mu}_n}}\Big),\quad i=M+1,\ldots,n. \tag{26}$$

Since the $\hat{B}_i$'s are symmetric, one can use either "+" or "−" in (26). The proposed iterative method in Section 2 will be followed to calculate $\hat{\lambda}_n$. There are no significant changes in our comparisons and conclusions in the next section when $\Phi^{-1}((i-m-1/2)/(n-m))$, suggested by (David (1981, Chap.4)) are used to approximate the normal percentiles.

4. Comparisons

First, if we write the IG c.d.f. in (5) as $F_T(t)$, then $F_{cT}(ct)=F_T(t)$ for any constant $c>0$, i.e., the c.d.f. is invariant to positive scale transformations. This implies that c can be judiciously chosen so that it is only necessary to vary one of the IG parameters, or through (3), to vary one of the Wiener parameters in our numerical study.

Large-sample Results

The traditional MLEs are known to be consistent (under mild conditions), and we proved the consistency of the MEME/MMLE for the mean failure time in Section 2.2. Now we show that the degradation values of the censored units at the censor time are indeed useful in reducing the asymptotic variance. First it can be shown that

$$Avar(\hat{\mu}_{MMLE})=(\frac{\mu^4}{\lambda})[\tau(1-F(\tau))+E(\delta T)]^{-1}, \tag{27}$$

with $E(\delta T) = \mu\{\Phi(A) - \exp(2\lambda/\mu)\Phi(-B)\}$. Then, a sufficient condition for $Avar(\hat{\mu}_{MLE}) \geq Avar(\hat{\mu}_{MMLE})$ is

$$\frac{\lambda}{\mu^4}\Big\{\tau\Big[1-\Phi(A)-e^{\frac{2\lambda}{\mu}}\Phi(-B)\Big]-2\sqrt{\tau\lambda}e^{\frac{2\lambda}{\mu}}\Big[2\sqrt{\frac{\lambda}{\tau}}\frac{1-\Phi(A)}{1-F(\tau)}\Phi(-B)-\phi(B)\Big]\Big\}\geq 0.$$

As a function of λ and τ (with fixed $\mu = 100$), various software packages indicated the inequality is true, implying $Avar(\hat{\mu}_{MLE}) > Avar(\hat{\mu}_{MMLE})$. Similarly,

$$Avar(\hat{\lambda}_{MMLE}) = \Big\{(\frac{1}{2\lambda^2}) + (1-F(\tau))e^{\frac{2\lambda}{\mu}}D\Big\}^{-1}, \tag{28}$$

with

$$D=\int_{-\infty}^{a}\frac{\sqrt{\lambda}}{\sqrt{2\pi\tau a^2}}\Big(\frac{2(a-w)}{a\tau}\Big)^2\Big(1-\exp\Big\{-\frac{2\lambda(a-w)}{a\tau}\Big\}\Big)^{-1}\exp\Big\{-\frac{\lambda(w-\frac{a\tau}{\mu}-2a)^2}{2\tau a^2}\Big\}dw.$$

Again, a sufficient condition for $Avar(\hat{\lambda}_{MLE}) - Avar(\hat{\lambda}_{MMLE}) \geq 0$ is

$$\begin{aligned}(1-F(\tau))\Big(\frac{1}{2\lambda}+\frac{1}{2\lambda^2}\Big)&+(1-F(\tau))\Big\{\frac{4}{\mu^2}(1-\Phi(A))-\frac{4}{\tau\lambda}(B^2-1)(1-F(\tau))^2\Big\}e^{\frac{2\lambda}{\mu}}\Phi(-B)\\&+\frac{1}{1-F(\tau)}\Big\{\Big[\frac{12}{\tau^2}+\frac{4}{\mu\tau}-\frac{8}{\sqrt{\lambda\tau^3}}\Big](1-F(\tau))+\Big(\frac{B}{\mu\lambda}-\frac{1+A^2}{2\sqrt{\lambda^3\tau}}\Big)\\&+\frac{1}{\tau\lambda}e^{\frac{2\lambda}{\mu}}\Big[\phi(B)-4\frac{\sqrt{\lambda\tau}}{\mu}\Phi(-B)\Big]\Big\}\phi(A)\geq 0.\end{aligned} \tag{29}$$

Among the quantities in the last term of the left hand side of (29), $12/\tau^2 + 4/(\mu\tau) - 8/\sqrt{\lambda\tau^3}$ is likely to be positive (for large λ). Since others quantities are < 1, the last term is non-negative if τ is not too large (so $1 - F(\tau)$ is not too small). Note that $B/\mu\lambda - (1 + A^2)/2\sqrt{\lambda^3\tau}$ will be negative only when $\tau < \mu$ or $\lambda < \mu$. But when $\tau > \mu$ and $\lambda < \mu$, the IG distribution has a long tail such that $(1 - F(\tau))$ is not too small, and this makes the last term of (29) positive. The second term is negative only when τ is very small and λ is sufficiently large. But, under these conditions, $e^{2\lambda/\mu}\Phi(-B)$ will be very small and then the second term is very small. The first term of (29) is always positive. Hence, we have demonstrated that $Avar(\hat{\lambda}_{MLE}) \geq Avar(\hat{\lambda}_{MMLE})$, which has been verified by graphical methods.

Small-sample Results

Next, we evaluate the small-sample performances of the MEMEs, MMLEs, and MLEs for both IG parameters. In our simulation, we choose to fix $\eta = 0.1$ and $a = 10$, which will in turn fix $\mu = 100$. We consider $\sigma = 0.2$ and 0.1, which corresponds to $\lambda = 2,500$, and $10,000$, respectively. Instead of fixing the censor time τ for comparisons, it is more appropriate to fix the failure probabilities, which are set at $p = 0.2$, 0.5 and 0.8 (the censor probabilities are 0.8, 0.5 and 0.2, respectively). For given μ, λ and p, the corresponding censor time is computed from $p = F(\tau \mid \mu, \lambda)$, where F is given in (5). For example, when $\mu = 100$, $\lambda = 2,500$, and $p = 0.2$, $\tau = 82.84$.

The values of the Wiener process considered in Tseng, Tang and Ku (2003) are such that their λ/μ is larger than the cases considered above; but when λ/μ is large, the IG distribution resembles a normal distribution and, from our comparisons below, our proposed estimators of both IG parameters are better than the corresponding traditional estimators.

The degradation test is simulated 200 times for each case of (μ, λ, p) to obtain the simulated estimates and the corresponding standard errors (s.e.). The results are given in Tables 1a−1b.

Table 1a. Comparison of MEME, MMLE and MLE($\mu = 100, \lambda = 2,500$).

p	MEME(s.e.)		MMLE(s.e.)		MLE(s.e.)		Actual p
	μ	λ	μ	λ	μ	λ	
			$n = 16$				
0.2	100.25	2713.13	100.25	2891.17	102.59	10414.20	0.201
	(5.96)	(1142.43)	(5.96)	(1130.86)	(20.81)	(48757.45)	
0.5	100.13	2463.14	100.13	2881.24	101.26	3335.07	0.496
	(5.46)	(948.37)	(5.46)	(1096.58)	(9.42)	(2305.71)	
0.8	99.97	2437.51	99.97	2989.31	100.41	2989.60	0.795
	(5.43)	(998.46)	(5.43)	(1209.30)	(5.92)	(1412.91)	
			$n = 32$				
0.2	100.14	2828.27	100.14	2731.76	101.32	3976.13	0.196
	(3.96)	(765.01)	(3.96)	(707.36)	(11.91)	(4735.44)	
0.5	100.07	2602.49	100.07	2735.26	100.42	2895.78	0.497
	(3.61)	(715.74)	(3.61)	(749.15)	(4.86)	(1271.72)	
0.8	99.98	2525.25	99.98	2766.05	100.19	2768.56	0.801
	(3.55)	(688.48)	(3.55)	(734.52)	(3.82)	(902.19)	
			$n = 64$				
0.2	100.12	2809.56	100.12	2587.70	100.73	2866.08	0.198
	(2.78)	(503.18)	(2.78)	(450.70)	(7.65)	(1152.81)	
0.5	100.19	2572.5	100.19	2579.46	100.47	2614.25	0.494
	(2.55)	(428.22)	(2.55)	(422.78)	(3.30)	(743.97)	
0.8	100.13	2495.29	100.13	2605.52	100.35	2558.04	0.795
	(2.52)	(408.93)	(2.52)	(415.16)	(2.71)	(539.53)	
			$n = 128$				
0.2	100.06	2828.80	100.06	2537.78	100.20	2717.15	0.199
	(1.90)	(364.16)	(1.90)	(311.91)	(4.87)	(801.96)	
0.5	100.14	2592.28	100.14	2532.41	100.24	2564.76	0.497
	(1.86)	(306.64)	(1.86)	(295.56)	(2.30)	(507.11)	
0.8	100.13	2508.60	100.13	2552.24	100.24	2525.3	0.797
	(1.84)	(297.02)	(1.84)	(296.19)	(1.91)	(366.61)	

Table 1b. Comparison of MEME, MMLE and MLE($\mu = 100, \lambda = 10,000$).

p	MEME(s.e.)		MMLE(s.e.)		MLE(s.e.)		Actual p
	μ	λ	μ	λ	μ	λ	
			$n = 16$				
0.2	100.04	11057.53	100.04	12597.58	102.30	28589.33	0.197
	(2.81)	(4900.35)	(2.81)	(5526.91)	(18.45)	(42857.46)	
0.5	100.01	10465.36	100.01	12527.39	100.05	15098.72	0.512
	(2.68)	(4701.39)	(2.68)	(5716.70)	(3.57)	(11138.61)	
0.8	100.04	10201.81	100.04	12494.00	100.13	12741.94	0.800
	(2.66)	(4777.61)	(2.66)	(5837.85)	(2.81)	(6662.38)	
			$n = 32$				
0.2	100.10	10943.91	100.10	11169.92	100.61	18269.33	0.193
	(1.93)	(3158.13)	(1.93)	(3165.70)	(6.63)	(24262.41)	
0.5	100.12	10429.58	100.12	11200.39	100.13	12335.60	0.502
	(1.94)	(2959.83)	(1.94)	(3218.72)	(2.42)	(5894.81)	
0.8	100.15	10186.92	100.15	11138.21	100.17	11398.90	0.798
	(1.90)	(3099.30)	(1.90)	(3398.92)	(1.96)	(4056.05)	
			$n = 64$				
0.2	99.99	10734.75	99.99	10414.53	99.96	13426.46	0.200
	(1.34)	(1813.43)	(1.34)	(1736.72)	(3.35)	(11182.31)	
0.5	100.02	10243.33	100.02	10450.20	100.03	11058.53	0.502
	(1.34)	(1736.70)	(1.34)	(1773.43)	(1.71)	(3500.61)	
0.8	100.04	10002.27	100.04	10417.32	100.05	10526.68	0.799
	(1.31)	(1744.50)	(1.31)	(1832.72)	(1.36)	(2173.44)	
			$n = 128$				
0.2	100.03	10711.80	100.03	10144.37	100.11	11051.64	0.199
	(0.90)	(1274.31)	(0.90)	(1185.74)	(2.40)	(3605.53)	
0.5	100.06	10228.90	100.06	10192.67	100.06	10458.86	0.499
	(0.90)	(1252.56)	(0.90)	(1279.00)	(1.15)	(2169.23)	
0.8	100.08	10002.04	100.08	10178.55	100.09	10194.82	0.797
	(0.88)	(1221.92)	(0.88)	(1250.00)	(0.93)	(1468.72)	

For estimating μ, the MEME and MMLE are identical, as indicated in Section 2. The MEME/MMLE generally perform better than the corresponding MLE, and with smaller standard errors. For larger p (with little censoring), the three estimates tend to perform equally well, as expected. Other than the standard error, the values of MEME/MMLE are not greatly affected by the sample size.

For estimating λ on the other hand, the MEMEs generally perform better than the MMLEs for larger p (e.g., $p = 0.5$ and 0.8). For small p, MEME tends to overestimate λ. This is because predicting the failure times for the censored units by their expected failure times is likely to underestimate $\mathrm{Var}\,(T) = \mu^3/\lambda$, and hence overestimate λ. Therefore, focus should be on how to improve the MEME for small p. Using the bias reduction techniques in Remarks 3.2.1 and

3.2.2, MEME2 and MEME3 are computed for $\lambda = 2,500$ and $p = 0.2$. We see in Table 2 that the bias of MEME3 is smaller than that of other estimators.

Table 2. Comparisons of MEME, MMLE, MME2, and MME3($\mu = 100$, $\lambda = 2,500$, $p = 0.2$).

n	MEME(s.e.)	MMLE(s.e.)	MEME2(s.e.)	MEME3(s.e.)
16	2713.13 (1142.43)	2891.17 (1130.86)	2337.66 (1056.67)	2478.37 (952.18)
32	2828.27 (765.01)	2731.76 (707.36)	2398.27 (707.73)	2587.81 (650.76)
64	2809.56 (503.18)	2587.70 (450.70)	2361.04 (467.59)	2576.28 (438.39)
128	2828.80 (364.16)	2537.78 (311.91)	2357.10 (341.51)	2594.52 (317.83)

Censored Data versus Complete Data

It may be of interest to compare MMLEs of the IG parameters when (a) some units failed before the censor time and others were censored but with their degradation values available (i.e., the case considered in this paper), (b) failure times of all units were observed, and (c) all test units are censored with their degradation values available. First, for case (a), the asymptotic variances of the MMLEs for both IG parameters are given in (27) and (28). For case (b) to occur, its censor time τ_b must be very large ($\mu < \tau < \tau_b$). Since we assume we have complete data in case (b), the censor time is basically irrelevant, so the traditional MLEs and MMLEs of both IG parameters have

$$Avar(\hat{\mu}_b) = \mu^3\lambda^{-1}, \tag{30}$$
$$Avar(\hat{\lambda}_b) = 2\lambda^2. \tag{31}$$

Assume that all units were censored in case (c) under a very small censor time τ_c ($\tau_c < \tau < \mu$), so we will use the degradation values of all units at the censor time to estimate both IG parameters. It is straightforward to show that

$$Avar(\hat{\mu}_c) = \mu^4(\tau_c\lambda)^{-1}, \tag{32}$$
$$Avar(\hat{\lambda}_c) = 2\lambda^2. \tag{33}$$

By comparing (28) and (31) we see that, for estimating λ, the asymptotic variance of the MMLE of case (a) is smaller than that of the MMLE of case (b). For estimating μ, the difference of the two asymptotic variances in (27) and (30) is $(\tau/\mu) - ((\tau/\mu) - 1)\Phi(A) - ((\tau/\mu) + 1)e^{2\lambda/\mu}\Phi(-B) - 1$. Numerically we can verify that this difference is greater than 0, approximately, when $\tau > \mu$. We conclude that the *Avar* of the MMLE of case (b) of μ is larger than that of the corresponding MMLE of case (a) when $\tau > \mu$. And the opposite is true for estimate of λ.

Comparing (32) and (27) with $\tau = \tau_c$, we need only check if the difference $(1-\mu/\tau_c)\Phi(A)+(1+\mu/\tau_c)e^{2\lambda/\mu}\Phi(-B)$ is greater than 0 or not. First, if $\tau_c > \mu$, this term is greater than 0. On the other hand if $\tau_c < \mu$, the difference may be positive or negative. So we can conclude that the *Avar* of the estimator of μ for case (c) is larger than that in our model (case (a)) when $\tau_c > \mu$. And the *Avar* of case (c) is smaller than that of case (a) when estimating λ.

5. LED Example

Degradation measures for censored units at the censor time are available in many applications; for example, crack sizes in a fatigue-crack-growth study in Bogdanoff and Kozin (1985), and fatigue crack data and carbon-file resistors data in Padgett and Tomlinson (2004). In this section we illustrate our method using the light emitting diode (LED) lamp example from Tseng, Tang and Ku (2003), where they study the problem of determining the termination time for a burn-in test.

A contact image sensor (CIS) module is a contact type image sensing module that is composed of a line of LED lamps; the light intensity (brightness) of a LED lamp has a high correlation with its lifetime. A CIS can be used in a fax machine, document scanner, copy machine, mark reader, and other office automation equipment. Due to market competition and expected high reliability of LED lamps, manufacturers normally test their CISs using an accelerated degradation test (ADT) at higher-than-normal stress, allowing time-censoring to collect timely data for accessing the reliability of their products.

Let $L(t)$ be standardized brightness of a LED with $L(0) = 1$. A LED lamp is technically defined as failed when its brightness first decreases to 0.5, whether it is under normal or accelerated conditions. The following data were obtained from one of the leading LED manufacturers in Taiwan, using an ADT under electric current = 10 amperes and temperature = 105°C. The normal operating conditions are 10 amperes and 25°C. The censor time for the accelerated test is 6,480 hours. The sample size is $n = 24$ with $m = 18$ boundary-crossing times of $L(t)$'s below 0.5: 6274.826, 6164.547, 6144.000, 6102.000, 5430.000, 6291.087, 6259.672, 5261.236, 3963.600, 6034.026, 4866.947, 3508.613, 5008.976, 2893.333, 6172.000, 6158.170, 3494.400, and 4801.878. These times were not the true failure times (lifetimes) since they were obtained from an ADT. The brightness of the six censored units at censor time are: 0.5027, 0.5438, 0.5768, 0.5516, 0.5267, and 0.5639. These $L(t)$'s do not following a Wiener process, but Tseng, Tang and Ku (2003) and (Yu and Tseng (1999, 2002)) demonstrate that $W(t) = -\ln(L(t^{\delta}))$ can be modeled as $W(t) = \eta t + \sigma B(t)$ in (1). The boundary for $W(t)$ is $a = -\ln(0.5) = 0.6932$. The boundary-crossing time T of $W(t)$ is defined in (2). To apply our results, we transform the boundary-crossing and censor times by taking δ-th powers (with $\delta = 0.60$, see Lee (2006)) and take the negative values of natural logarithms of the brightness for the censored units. These transformed values will be our t_i, τ and $w_i(\tau)$ in Sections 2 and 3. We now apply our results to obtain various estimates of μ and λ of the IG distribution for T, see Table 3. The distribution of T can be estimated by $F(t \mid \hat{\mu}, \hat{\lambda})$, using (5).

Note that the data above were obtained under an ADT, but our interest is in estimating the failure time under normal conditions. Although it is not the focus of the current paper, we briefly describe how we transform our results; more details can be found in Lee (2006). We follow the model in Doksum and Hoyland (1992), where the unaccelerated degradation process, $W_0(t)$, is assumed to satisfy $W(t) = W_0(\beta(t))$ for some continuous, increasing, and nonnegative function $\beta(t)$. In our case $\beta(t) = \beta_0 t$ for some constant $\beta_0(\geq 1)$ is a reasonable choice, since the accelerated stress was held constant at 105°C throughout the ADT. For our example, $\beta_0 = 2.61$. So, if T_0 denotes the true failure time for $W_0(t)$ under normal operating conditions, then

$$\hat{E}(T_0) = E((\beta_0 T)^{\frac{1}{\delta}}) = \beta^{\frac{1}{\delta}} \int_0^\infty t^{\frac{1}{\delta}} dF(t \mid \hat{\mu}, \hat{\lambda}).$$

The results are given in the last column of Table 3. Note that T_0 does not follow an IG so there is no $\hat{\lambda}$ to consider. The results in Table 3 for various estimators are fairly close, which is consistent with our findings in Tables 1a-1b, since the failure probability, $p = 12/18 \approx 0.7$, is rather high. This is due to the unusally high testing time of 6,480 hours, which is about 9 months. If we were to shorten the testing time, we would expect to see fewer failures (so, p would decrease). Then, from the findings in Tables 1a-1b, our proposed estimators can be expected to perform better.

Table 3. Results of MEMEs, MMLEs, and MLEs for the LED Example.

Estimators	$\hat{\mu}$	$\hat{\lambda}$	$\hat{E}(T_0)$
MLE	471.39	17298	28982
MMLE	473.45	15306	29251
MEME	473.45	15015	29260
MEME2	473.45	16518	29215
MEME3	473.45	14521	29277

Note. $\hat{\mu} = \hat{E}(T)$. T_0 is the failure time under normal conditions.

6. Concluding Remarks

The traditional maximum likelihood estimators for the parameters of a failure time distribution assume that only failure times of the failed units and the censor times of the censored units are available. In this paper we assume that the degradation follows a Wiener process and that degradation values at the censor time for the censored units are also available. These degradation values are then used to predict the failure times of the censored units, using a modified EM algorithm. The algorithm provides closed-form estimates of both the mean and the scale parameter of the IG distribution. On the other hand, the estimate of the scale that maximizes the likelihood of all available data does not have a closed form so one must rely on numerical procedures to compute it. We then demonstrate that with the help of the degradation values of the censored units, one can reduce the asymptotic variances of the estimators for both IG parameters. Small-sample performances of the proposed estimators are also compared with the traditional estimators.

Finally, a possible direction for future research is to see whether it is better to collect first-passage times of the degradation sample paths over certain non-failure thresholds, instead of collecting degradation values at prescribed time points as considered in this paper.

Appendix. Derivation of (12)

Let $(R, \mathcal{B}, P)$ be the real space with Borel sigma-algebra $\mathcal{B}$ and Lebesque measure P, under which $B(t)$ is a Brownian motion. Define $M(\tau) = \max_{0\le t\le\tau} B(t)$, then we have (Hida (1980))

$$\begin{aligned} P(B(\tau) < b, M(\tau) > m) &= P(B(\tau) > 2m - b) \\ &= \int_{2m-b}^{\infty} \frac{1}{\sqrt{2\pi\tau}} \exp\left\{\frac{-x^2}{2\tau}\right\} dx, \quad \text{for } m \ge 0 \text{ and } m > b, \end{aligned}$$

from which the joint p.d.f of $B(\tau)$ and $M(\tau)$ with respect to P is obtained by differentiating the above probability:

$$f(b, m) = \frac{2(2m-b)}{\sqrt{2\pi\tau^3}} \exp\left\{ -\frac{(2m-b)^2}{2\tau}\right\}, \quad m \ge 0,\ m > b. \tag{A.1}$$

Let $\overline{B}(t) = \theta t + B(t)$ for $t \le T$, where $T \le \infty$, and put $M_t = \exp\{-\theta B(t) - (1/2)\theta^2 t\}$, $t \in [0, T]$. Define the measure $\overline{P}$ on $(R, \mathcal{B})$ by

$$\overline{P}(A) = \int_A M_t dP, \quad \text{for all } A \in \mathcal{B}.$$

Then by Girsanov's Theorem (Oksendal (2000, p.153)), $\overline{B}(t) = \theta t + B(t)$ is a Brownian motion w.r.t. $\overline{P}$. Here $M_t = \exp\{-\theta\overline{B}(t) + \theta^2 t/2\}$, written in terms of $\overline{B}(t)$, is called the Radon-Nikodym derivative of $\overline{P}$ w.r.t. P. Define $\overline{M}(\tau) = \max_{0\le t\le\tau} \overline{B}(t)$. Similar to (A.1), the p.d.f .of $(\overline{B}(\tau), \overline{M}(\tau))$ w.r.t. $\overline{P}$ is (A.1) with m and b replaced by $\overline{m}$ and $\overline{b}$, respectively. The joint p.d.f. of $\overline{M}(\tau)$ and $\overline{B}(\tau)$ with respect to the measure P is obtained by multiplying the p.d.f. of $(\overline{B}(\tau), \overline{M}(\tau))$ by $1/M_t$, as

$$f(\overline{b}, \overline{m}) = \frac{2(2\overline{m} - \overline{b})}{\sqrt{2\pi\tau^3}} \exp\left\{ - \frac{(2\overline{m} - \overline{b})^2}{2\tau} \right\} \cdot \exp\left\{ \theta\overline{b} - \frac{1}{2}\theta^2\tau \right\}, \quad \text{for } \overline{m} > 0, \overline{m} > \overline{b}. \tag{A.2}$$

If we let $W(t) = \sigma\overline{B}(t) = \sigma\theta t + \sigma B(t)$ and $\eta = \sigma\theta (\sigma > 0)$, then $W(t) = \eta t + \sigma B(t)$ as defined in (1), and $\widetilde{M}(\tau) = \max_{0 \le t \le \tau} W(t) = \sigma\overline{M}(\tau)$. Now, from (A.2) with the standard transformation technique, we can obtain the joint p.d.f., denoted by $\underset{\sim}{f}(w, \widetilde{m})$, of $W(\tau)$ and $\widetilde{M}(\tau)$, with respect to P. Finally, from $P(W(\tau) = w, \widetilde{M} \le a) = \int_w^a f(w, \widetilde{m}) d\widetilde{m}$, we obtain (12).

References

Bhattacharyya, G. K. and Fries, A. (1982). Fatigue failure models: Birnbaum Saunders vs. Inverse Gaussian. *IEEE Trans. Reliability* **31**, 439-441.

Bogdanoff, J. L. and Kozin, F. (1985). *Probabilistic Models of Cumulative Damage*. Wiley, New York.

Chhikara, R. S. and Folks, J. L. (1989). *The Inverse Gaussian Distribution: Theorem, Methodology, and Application.* Marcel Dekker, New York.

Cox, D. R. and Miller, H. D. (1965). *The Theory of Stochastic Processes*. Wiley, New York.

David, H. A. (1981). *Order Statistics*. 2nd edition. Wiley, New York.

Doksum, K. A. and Hoyland, A. (1992). Models for variable-stress accelerated life testing experiments based on Wiener processes and the Inverse Gaussian distribution. *Technometrics* **34**, 74-82.

Doksum, K. A. and Normand, S.-L. T. (1995). Gaussian models for degradation processes-Part I: Methods for the analysis of biomarker data. *Lifetime Data Analysis* **1**, 131-144.

Hida, T. (1980). *Brownian Motion.* Springer-Verlag, NewYork.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data.* Wiley, New York.

Lee, M. Y. (2006). Estimation of lifetime distribution and optimal design of degradation test for Wiener degradation processes. Ph.D. Dissertation, Institute of Statistics, National Tsing-Hua University, Taiwan.

Oksendal, B. (2000). *Stochastic Differential Equations: An Introduction and Applications.* 5th edition. Springer-Verlag, New York.

Padgett, W. J. and Tomlinson, M. A. (2004). Inference from accelerated degradation and failure data based on Gaussian process models. *Lifetime Data Analysis* **10**, 191-206.

Seshadri, V. (1999). *The Inverse Gaussian Distribution: Statistical Theory and Applications.* Springer-Verlag, New York.

Sherif, Y. S. and Smith, M. L. (1980). First-passage time distribution of Brownian Motion as a reliability model. *IEEE Trans. on Reliability* **29**, 425-426.

Singpurwalla, N. D. (1995). Survival in dynamic environments. *Statist. Sci.* **10**, 86-103.

Tseng, S. T., Tang, J. and Ku, I. H. (2003). Determination of optimal burn-in parameters and residual life for highly reliable products. *Naval Res. Logist.* **50**, 1-14.

Whitmore, G. A. and Schenkelberg, F. (1997). Modelling accelerated degradation data using Wiener Diffusion with a time scale transformation. *Lifetime Data Analysis* **3**, 27-45.

Yu, H. F. and Tseng, S. T. (1999). Designing a degradation experiment. *Naval Res. Logist.* **46**, 689-706.

Yu, H. F. and Tseng, S. T. (2002). Designing a screening experiment for highly reliability products. *Naval Res. Logist.* **49**, 514-526.

Institute of Medicine, Chung Shan Medical University, 110, Sec. 1, Chien-Kuo N. Road, Taichung 402, Taiwan, R.O.C.

Clinical Trial Center, Chung Shan Medical University Hospital, 110, Sec. 1, Chien-Kuo N. Road, Taichung 402, Taiwan, R.O.C.

E-mail: cshn060@csh.org.tw

Krannert Graduate School of Management, Purdue University, West Lafayette, IN 47907, U.S.A.

E-mail: jtang@purdue.edu

(Received June 2005; accepted July 2006)