Supplementary Materials for “Estimation of Field Reliability Based on Aggregate Lifetime Data”

Piao CHEN and Zhi-Sheng YE

Department of Industrial & Systems Engineering

National University of Singapore, Singapore, 117576

1 Simulations to Assess the Robustness

The Weibull and lognormal distributions are the most common distributions used in reliability applications. However, use of these two distributions for aggregate data is not tractable because the sum of i.i.d. Weibull or lognormal random variables does not have closed-form probability density function (PDF) or cumulative density function (CDF). The gamma and IG distributions, although not as widely-used, are also popular models for lifetime data. The purpose of this section is to show that even if the underlying component lifetimes follow the Weibull/lognormal distribution, the gamma and IG distributions are robust in terms of the estimated quantiles, as long as there is not a substantial amount of extrapolation to the tails.

The procedure is as follows. We first generate the aggregate data of size n and $m_i = 3$ from the Weibull distribution $wb(k, 1)$, where k is the shape parameter of the Weibull distribution. Then we fit the data by the gamma distribution and IG distribution and afterwards we can obtain the ML estimates of $\tau_{0.1}$, $\tau_{0.5}$ and $\tau_{0.9}$, the 0.1, 0.5 and 0.9 quantiles of the random component lifetime. After 10,000 replications, we can obtain biases of $\hat{\tau}_{0.1}$, $\hat{\tau}_{0.5}$ and $\hat{\tau}_{0.9}$. Also, based on the interval estimation methods proposed in the paper, we can obtain the coverage probabilities under model-misspecification when the nominal value is set

as 95%. We consider $n = 10, 20, 40$ and the shape parameter $k = 0.5, 1, 2$. Similarly, we can generate the aggregate data of size n from the log-normal distribution $\ln\mathcal{N}(0, \sigma)$, where σ is the shape parameter. After fitting data by the gamma and IG distributions, we can also investigate the biases of the estimated quantiles as well as the coverage probabilities under model-misspecification. We also set $n = 10, 20, 40$ and $\sigma = 0.5, 1, 2$. The simulation results are shown in Table A.1 and A.2. As we can see, the gamma and IG distributions give reasonable inferences when the 0.1 and 0.5 quantiles are considered. On the other hand, the extrapolation into the tails (0.9 quantile) would lead to large biases.

Table A.1: Biases and coverage probabilities (CP) of $\hat{\tau}_{0.1}$, $\hat{\tau}_{0.5}$ and $\hat{\tau}_{0.9}$ in percentage by generating data from $wb(k, 1)$.

n	k	Gamma/ $\tau_{0.1}$		IG/ $\tau_{0.1}$		Gamma/ $\tau_{0.5}$		IG/ $\tau_{0.5}$		Gamma/ $\tau_{0.9}$		IG/ $\tau_{0.9}$	
		Bias	CP	Bias	CP	Bias	CP	Bias	CP	Bias	CP	Bias	CP
10	0.5	1.23	91.80	10.21	78.50	13.20	87.20	0.47	86.20	55.60	72.80	-117.31	73.80
	1	4.94	98.00	12.36	67.60	2.87	98.50	-5.57	87.30	-8.23	98.80	-13.51	62.80
	2	7.57	93.50	10.46	81.70	-1.70	88.90	-5.00	82.40	-3.86	85.40	-2.95	73.30
20	0.5	-0.30	95.70	7.01	69.80	4.96	96.50	-9.33	78.80	104.54	95.40	-130.47	62.30
	1	2.08	95.10	9.60	79.20	0.37	95.50	-8.95	71.80	-5.47	94.90	-10.19	58.80
	2	5.23	96.60	8.11	99.00	-2.56	95.50	-6.20	98.50	-1.94	95.10	-0.57	98.00
40	0.5	-0.64	93.50	5.60	85.70	2.25	93.50	-14.71	77.80	128.93	89.70	-146.59	63.40
	1	1.27	92.90	8.85	82.20	0.77	90.00	-9.22	71.70	-2.66	87.10	-6.20	66.10
	2	3.92	96.20	6.78	96.80	-3.10	96.10	-6.94	97.70	-0.84	97.00	0.80	98.10

Table A.2: Biases and coverage probabilities (CP) of $\hat{\tau}_{0.1}$, $\hat{\tau}_{0.5}$ and $\hat{\tau}_{0.9}$ in percentage by generating data from $\ln\mathcal{N}(0, \sigma)$.

n	σ	Gamma/ $\tau_{0.1}$		IG/ $\tau_{0.1}$		Gamma/ $\tau_{0.5}$		IG/ $\tau_{0.5}$		Gamma/ $\tau_{0.9}$		IG/ $\tau_{0.9}$	
		Bias	CP	Bias	CP	Bias	CP	Bias	CP	Bias	CP	Bias	CP
10	0.5	-0.01	96.15	4.19	92.67	4.56	94.72	1.12	97.32	-4.00	92.70	-4.18	91.80
	1	-6.19	89.01	8.50	90.40	13.35	89.64	3.64	95.35	23.75	88.10	5.69	89.80
	2	-5.30	82.62	15.22	88.05	15.01	84.52	7.90	89.36	823.36	80.00	-107.23	85.90
20	0.5	-2.94	96.46	1.61	95.19	3.62	94.16	-0.05	98.18	-1.06	92.30	-1.30	88.00
	1	-10.76	85.13	4.75	94.31	9.51	84.66	-0.51	94.79	27.19	82.20	9.11	91.80
	2	-7.06	86.37	10.88	93.99	0.47	80.73	-6.01	90.88	1290.51	63.60	-106.90	89.90
40	0.5	-3.55	92.61	1.12	92.77	4.03	93.99	0.27	96.78	0.92	92.20	0.69	90.70
	1	-13.03	78.44	3.32	96.12	8.27	76.94	-1.78	91.73	33.40	76.60	14.43	91.30
	2	-7.50	81.58	8.70	89.53	-10.76	80.70	-13.76	93.15	990.21	76.70	-76.98	89.20

2 Identifiability Issue

Identifiability is sometimes an issue in non-parametric inference. In parametric inference, the identifiability issue arises when the number of data points is less than the number of parameters, or when there are random effects (both discrete mixture and continuous mixture). In our aggregate data problem, the gamma and IG distributions were used for the aggregate data. For the gamma distribution, there will be identifiability problem if (i) there is only 1 observation, or (ii) t_i/m_i are the same for all systems. Point (i) is obvious. In the following, we would justify point (ii) in terms of the point estimation and interval estimation, respectively.

We first consider the point estimation of gamma parameters using the moment method. The moment estimators for k and θ exist if and only if the model is identifiable. The cumulative operation time T_i follows a gamma distribution with shape parameter $m_i k$ and rate parameter θ . Therefore, $E(T_i/m_i) = k/\theta$ and $E(T_i^2/m_i^2) = (k^2 + k/m_i)/\theta^2$. Let $\tilde{M} = \sum_i 1/m_i$. By solving $\sum_i (t_i/m_i) = nk/\theta$ and $\sum_i (t_i^2/m_i^2) = (nk^2 + \tilde{M}k)/\theta^2$ together, we can obtain moments estimators of k and θ as

$$\hat{k} = \frac{\tilde{M}(\sum t_i/m_i)^2}{n^2 \sum (t_i^2/m_i^2) - n(\sum t_i/m_i)^2} \quad \text{and} \quad \hat{\theta} = \frac{n\hat{k}}{\sum t_i/m_i}.$$

As can be seen, the identifiability issue occurs when $n^2 \sum (t_i^2/m_i^2) - n(\sum t_i/m_i)^2 = 0$. This happens only when $t_i/m_i = t_2/m_2 = \dots = t_n/m_n$. In fact, we can also use the contour plot of the likelihoods to illustrate this point. We take the airplane indicator light data provided in the paper as an example. The left panel of Figure A.1 shows the contour plot of the likelihoods based on the original data. As we can see, unique values of the ML estimates $\hat{k}$ and $\hat{\theta}$ can be determined. However, if we set the $t_i/m_i = 33$ for all the systems, there would be a wide range of combinations of (k, θ) such that the likelihoods are almost same, as shown in the right panel of Figure A.1. In such case, the identifiability problem arises.

In the interval estimation procedure, we use the approximation $W_0 = -2Nk \log(\tilde{y}/\bar{y}) \sim$

Figure A.1: Contour of the gamma log-likelihood functions based on the original airplane indicator light data and the modified data.

$c\chi_{(v)}^2$ in constructing the confidence interval, where $\bar{y} = \sum t_i/N$ and $\tilde{y} = [\prod (t_i/m_i)^{m_i}]^{1/N}$, and $N = \sum m_i$. The approximation fails when $\tilde{y} = \bar{y}$. Again, this is equivalent to constant t_i/m_i for all i .

The above analysis can be applied to the IG distribution, and the identifiability condition is the same. To be specific, the shape parameter λ is not identifiable when t_i/m_i is constant for all i . When the PDF of the product lifetime exists, the probability of observing constant t_i/m_i from all systems is 0. Nevertheless, due to the round-off errors in data collection, there might be a small chance to observe constant t_i/m_i . In the case that there is only one single system or t_i/m_i is constant, the exponential distribution has to be used for the data.

3 Some Technical Notes

3.1 Derivation of Equation (1)

To establish Equation (A.1) (Equation (1) in the main text)

$$E(S_0^{zN}) = \frac{N^{zN} \Gamma(kN) \prod_i \Gamma(zm_i + km_i)}{\Gamma(zN + kN) \prod_i m_i^{zm_i} \Gamma(km_i)}, \quad (\text{A.1})$$

we first show that S_0 and $\bar{y}$ are statistically independent. If we express S_0 as $S_0(t_1, \dots, t_n)$, then it is equivalent to show that $S_0(t_1, \dots, t_n)$ is independent with $\sum t_i$. In addition, because S_0 is a scale invariant statistic, without loss of generality, set $\theta = 1$ here. Let

$$\varphi(u, v) = E[e^{i(u \sum t_i + v S_0)}], \quad |u| < 1,$$

be the characteristic function of $(\sum t_i, S_0)$. Then

$$\varphi(u, v) = \frac{1}{\Gamma(m_1 k) \cdots \Gamma(m_n k)} \int_0^\infty \cdots \int_0^\infty e^{-\sum t_i} \left(\prod t_i^{m_i k - 1} \right) e^{i(u \sum t_i + v S_0)} dt_1 \cdots dt_n.$$

Substitute $q_i = t_i(1 - u)$ and together with the fact $S_0(q_1, \dots, q_n) = S_0(t_1, \dots, t_n)$, we have

$$\varphi(u, v) = (1 - iu)^{-\sum m_i k} \frac{\int_0^\infty \cdots \int_0^\infty e^{-\sum q_i} \left(\prod q_i^{m_i k - 1} \right) e^{iv S_0} dq_1 \cdots dq_n}{\Gamma(m_1 k) \cdots \Gamma(m_n k)}.$$

Note that the two terms in the above equation are the characteristic functions of $\sum t_i$ and S_0 , respectively. Therefore, S_0 is independent with $\sum t_i$ (Applebaum et al. 2005, Theorem 2.1) and thus independent with $\bar{y}$. Then we have $E(\tilde{y}^{zN}) = E[(S_0 \bar{y})^{zN}] = E(S_0^{zN}) E(\bar{y}^{zN})$ and afterwards it can be shown that when $z > -k$,

$$E(S_0^{zN}) = \frac{E(\tilde{y}^{zN})}{E(\bar{y}^{zN})} = \frac{\prod_i E(y_i^{zm_i})}{E(\bar{y}^{zN})} = \frac{N^{zN} \Gamma(kN) \prod_i \Gamma(zm_i + km_i)}{\Gamma(zN + kN) \prod_i m_i^{zm_i} \Gamma(km_i)}.$$

3.2 Proof of Theorem 1

Define $S_2 = S_0^{Nk}$ and then $W_0 = -2 \log S_2$. Let $z = rk$ in Equation (A.1), we have

$$E(S_2^r) = \frac{\prod_i E(y_i^{rk m_i})}{E(\bar{y}^{rk N})} = \frac{N^{kNr} \Gamma(kN) \prod_i \Gamma[k m_i (1+r)]}{\Gamma[kN(1+r)] \prod_i m_i^{m_i k r} \Gamma(k m_i)}, \quad (\text{A.2})$$

when $r > -1$. In addition, we can find that

$$\bar{y} = \sum_i \frac{m_i y_i}{N} \geq (\prod_i y_i^{m_i})^{1/N} = \tilde{y},$$

and thus $0 \leq S_2 \leq 1$. By substituting Stirling's approximation to the gamma function (Wrench 1968)

$$\Gamma(x) = \sqrt{2\pi} x^{x-1/2} e^{-x} [1 + o(1)],$$

to (A.2), we can show that when $k \rightarrow \infty$,

$$E(S_2^r) = (1+r)^{-kN(1+r)+1/2} \prod_i (1+r)^{k m_i (1+r)-1/2} = (1+r)^{-(n-1)/2}.$$

As $W_0 = -2 \log S_2$, by simple transformation, the moments of W_0 when $k \rightarrow \infty$ can be expressed as

$$E(W_0^r) = \frac{2^r \Gamma[r + (n-1)/2]}{\Gamma[(n-1)/2]},$$

which shows $W_0 \sim \chi_{(n-1)}^2$ when $k \rightarrow \infty$.

On the other hand, a Taylor series expansion for the gamma function (Wrench 1968) around 0 gives

$$\frac{1}{\Gamma(x)} = x + O(x^2).$$

We can then obtain that when $k \rightarrow 0$,

$$E(S_2^r) = (1+r)^{-(n-1)} N^{kNr} \prod_i m_i^{m_i k r} = (1+r)^{-(n-1)}.$$

Similarly, the moments of W_0 when $k \rightarrow 0$ can be expressed as

$$E(W_0^r) = \frac{2^r \Gamma(r + n - 1)}{\Gamma(n - 1)},$$

which shows $W_0 \sim \chi_{(2n-2)}^2$ when $k \rightarrow 0$.

3.3 Proof of Theorem 2

First, by the joint distribution of $T_1, T_2, \dots, T_n$ and the distribution of $\hat{\mu}$, the conditional joint distribution of $T_1, T_2, \dots, T_n$ given $\hat{\mu}$ can be obtained as

$$f(t_1, \dots, t_n | \hat{\mu}) = \left(\frac{\lambda}{2\pi} \right)^{(n-1)/2} \frac{\prod m_i \hat{\mu}^{3/2}}{\sqrt{N} \prod t_i^{3/2}} \exp \left(-\frac{n\lambda}{2\hat{\lambda}} \right).$$

Then, the MGF of $1/\hat{\lambda}$ given $\hat{\mu}$ can be expressed as

$$E(e^{u/\hat{\lambda}} | \hat{\mu}) = \left(1 - \frac{2u}{n\lambda} \right)^{(n-1)/2}.$$

Since the above MGF does not contain $\hat{\mu}$ and $\lambda V = \lambda n / \hat{\lambda}$, we can easily find that λV is independent of $\hat{\mu}$ and its distribution is $\chi_{(n-1)}^2$.

3.4 Proof of Theorem 3

Define $Z = \frac{\sqrt{N\lambda}(\bar{y} - \mu)}{\mu\sqrt{\bar{y}}}$ and $U = \lambda V$, and then $H = \frac{Z}{\sqrt{U/(n-1)}}$. Firstly, it is not hard to show the PDF of Z is

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \left(1 - \frac{z}{z^2 + 4N\lambda/\mu} \right) \exp(-z^2/2).$$

Then, the joint distribution of Z and U can be shown as

$$f_{Z,U}(z, u) = \frac{1}{\sqrt{2\pi} 2^{(n-1)/2} \Gamma[(n-1)/2]} \left(1 - \frac{z}{\sqrt{z^2 + 4N\lambda/\mu}} \right) u^{(n-3)/2} \exp(-u/2 - z^2/2).$$

The Jacobian of the transformation from (z, u) to (h, u) is

$$J = \frac{\partial z}{\partial h} \frac{\partial u}{\partial u} - \frac{\partial z}{\partial u} \frac{\partial u}{\partial h} = \sqrt{\frac{u}{n-1}}.$$

Therefore, the joint distribution of H and U can be expressed as

$$\begin{aligned} f_{H,U}(h, u) &= f_{Z,U}(h\sqrt{u/(n-1)}, u)|J| \\ &= \frac{u^{n/2-1}}{\sqrt{2\pi}2^{(n-1)/2}\Gamma[(n-1)/2]\sqrt{n-1}} \left[1 - \frac{h\sqrt{u}}{\sqrt{h^2u + 4N(n-1)\lambda/\mu}} \right] \exp \left[-\frac{u}{2} \left(1 + \frac{h^2}{n-1} \right) \right]. \end{aligned}$$

Integrating with respect to u , we can obtain the PDF of H

$$f_H(h) = \frac{[1 + h^2/(n-1)]^{-n/2}}{\sqrt{n-1}B(\frac{1}{2}, \frac{n-1}{2})} q(h),$$

where

$$q(h) = 1 - \frac{h}{\Gamma(n/2)} \int_0^\infty \frac{w^{(n-1)/2} e^{-w} dw}{\sqrt{h^2w + 2N\lambda(h^2 + n-1)/\mu}}.$$

This completes the proof.

4 Simulation Results

4.1 Simulation results for the shape parameter k and the rate parameter θ of the gamma distribution

We first consider the shape parameter k . A wide range of values of k , i.e., $k = 0.05, 1, 5, 10$ is considered to examine the performance of our two-moment matching method. Small sample sizes, i.e., $n = 5, 10, 20$ are used in the simulation. We are interested in the equal-tailed 90% and 95% confidence intervals of k . For each combination of k and n , the simulation was replicated 10,000 times, based on which the coverage probabilities were obtained from the

proposed method, the Wald method, and the bootstrap- t method. The results are shown in Figure A.2 and A.3. As can be seen, the bootstrap method performs poorly when the sample size is small especially for large k , and its performance improves with the sample size n . The large sample approximation tends to work well when n is large. But its performance is poor when $n = 5$, especially when the confidence level is 90%. On the other hand, our proposed two-moment matching method works for all combinations of k and n . Also, between the two methods we proposed in Section 2.1, the equation-solving method (MM1) seems to have better performance than the the plug-in method (MM2), where MM stands for moment matching.

Figure A.2: Coverage probability of k by two moment matching methods (MM1 and MM2), Wald method (WALD) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 95$.

Figure A.3: Coverage probability of k by two moment matching methods (MM1 and MM2), Wald method (WALD) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 90$.

Then we consider the generalized pivotal quantity method for the rate parameter θ .

We set $B = 10,000$ in Algorithm 1 and small sample sizes $n = 5, 10, 20$ are considered. Figure A.4-A.9 present the simulation results when $k = 0.5, 1, 3$ and $\theta = 0.5, 1, 3$ under each k , which are based on 10,000 Monte Carlo replications. The coverage probabilities of the bootstrap interval are quite sensitive to the sample size n and θ while the converge probabilities of the Wald intervals are uniformly higher than the nominal values. On the other hand, the performance of Algorithm 1 is consistent, regardless of the true parameter values θ and the sample size n . Note that the GPQ method seems not work extremely well when $k = 0.5$ and $1 - \alpha = 90\%$, as shown in Figure A.5. This is perhaps due to the error brought by the χ^2 approximation and ML estimation of k . However, the GPQ method still works reasonably well in this scenario since the difference between the coverage probabilities and nominal values is 2% in average while performance of the bootstrap- t method and Wald method is poor in certain parameter settings. We shall mention in passing that there are not good methods for interval estimation of θ based on individual lifetime data so far. Our algorithm, which includes the use of individual lifetimes as a special case, bridges this gap.

Figure A.4: Coverage probability of θ by the GPQ methods (GPQ), Wald method (WALD) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 95$ and $k = 0.5$.

Figure A.5: Coverage probability of θ by the GPQ method (GPQ), Wald method (WALD) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 90$ and $k = 0.5$.

Figure A.6: Coverage probability of θ by the GPQ method (GPQ), Wald method (WALD) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 95$ and $k = 1$.

Figure A.7: Coverage probability of θ by the GPQ method (GPQ), Wald method (WALD) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 90$ and $k = 1$.

Figure A.8: Coverage probability of θ by the GPQ method (GPQ), Wald method (WALD) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 95$ and $k = 3$.

Figure A.9: Coverage probability of θ by the GPQ method (GPQ), Wald method (WALD) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 90$ and $k = 3$.

4.2 More simulation results for the mean parameter μ of the gamma distribution

In the paper, we have shown the simulation results for the mean parameter μ of the gamma distribution when constructing the one-sided 97.5% confidence limits in the case $k = 3$. Here, the coverage probabilities of the confidence interval for μ when $k = 0.5, 1, 3$ are presented in Figure A.10-A.15. As expected, the method based on the generalized pivotal quantity has a much better performance than the Wald method and the bootstrap- t method.

Figure A.10: Coverage probability of μ by the GPQ method (GPQ), Wald method (WALD) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 95$ and $k = 0.5$.

Figure A.11: Coverage probability of μ by the GPQ method (GPQ), Wald method (WALD) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 90$ and $k = 0.5$.

Figure A.12: Coverage probability of μ by the GPQ method (GPQ), Wald method (WALD) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 95$ and $k = 1$.

Figure A.13: Coverage probability of μ by the GPQ method (GPQ), Wald method (WALD) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 90$ and $k = 1$.

Figure A.14: Coverage probability of μ by the GPQ method (GPQ), Wald method (WALD) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 95$ and $k = 3$.

Figure A.15: Coverage probability of μ by the GPQ method (GPQ), Wald method (WALD) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 90$ and $k = 3$.

4.3 More simulation results for the quantile τ_p of the gamma distribution

In the main text, we have shown the simulation results for 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 quantiles of gamma distribution in the case when the shape parameter $k = 3$. Here, we further show the results when $k = 0.5$ and 3 in Figure A.16 and A.17. As can be seen, our proposed method based on the generalized pivotal quantity performs much better than the bootstrap- t method in terms of the coverage probability.

Figure A.16: Coverage probability of τ_p by the GPQ method (GPQ) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 90$ (solid), 95 (dotted) and $k = 0.5$.

Figure A.17: Coverage probability of τ_p by the GPQ method (GPQ) and bootstrap- t method (BOOT) when $100(1 - \alpha) = 90$ (solid), 95 (dotted) and $k = 1$.

5 Normal Distribution Model

In the main text, we have used the gamma and IG distributions to fit the aggregate data. Actually, normal distribution can also be used for aggregate data because the sum of i.i.d. normal distributed random variables is also normal distributed. Though normal distribution is less commonly used in the lifetime data analysis, it is proved to be useful for certain lifetime data (Meeker and Escobar 1998, Section 4.5). Therefore, we also proposed inference method about the normal distribution based on the aggregate data here. Suppose the component lifetime X follows a normal distribution $N(\mu, \sigma^2)$ with PDF

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right].$$

Since the sum of identical normal distributed random variables is again normal distributed, the distribution of i -th aggregate lifetime is $N(m_i\mu, m_i\sigma^2)$. The log-likelihood function of θ based on the data $\mathbf{D}$, up to a constant, can be specified as

$$l(\theta; \mathbf{D}) = -n \log \sigma - \sum_i \frac{(t_i - m_i\mu)^2}{2m_i\sigma^2}.$$

The ML estimates of μ and σ can then be obtained as

$$\hat{\mu} = \bar{y} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n m_i (y_i - \bar{y})^2,$$

where $N = \sum m_i$ and $\bar{y} = \sum t_i / N$. As for the interval estimation of these two parameters, we first construct the modified unbiased sample variance as

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n m_i (y_i - \bar{y})^2,$$

where $y_i = t_i / m_i$ as defined in the article. The following two theorems show that $\frac{(n-1)s^2}{\sigma^2} \sim \chi_{(n-1)}^2$ and $\frac{\bar{y} - \mu}{s/\sqrt{N}} \sim t_{n-1}$ and then the confidence intervals of μ and σ can be constructed. The proofs are trivial and hence omitted here.

Theorem 4 Assume n aggregate lifetimes $T_1, \dots, T_n$ are observed and $T_i \sim N(m_i \mu, m_i \sigma^2)$, $i = 1, \dots, n$. Define $s^2 = \frac{1}{n-1} \sum_{i=1}^n m_i (y_i - \bar{y})^2$ and then it can be shown

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{(n-1)}^2.$$

Based on Theorem 4, an exact two-sided $1 - \alpha$ level confidence interval for σ^2 can be

$$\left[\frac{(n-1)s^2}{\chi_{(1-\alpha/2, n-1)}^2}, \frac{(n-1)s^2}{\chi_{(\alpha/2, n-1)}^2} \right].$$

Theorem 5 Assume n aggregate lifetimes $T_1, \dots, T_n$ are observed and $T_i \sim N(m_i \mu, m_i \sigma^2)$, $i = 1, \dots, n$. Define $s^2 = \frac{1}{n-1} \sum_{i=1}^n m_i (y_i - \bar{y})^2$ and then we have

$$\frac{\bar{y} - \mu}{s/\sqrt{N}} \sim t_{n-1}.$$

Based on Theorem 5, an exact two-sided $1 - \alpha$ level confidence interval for μ can be

$$\left[\bar{y} - \frac{\sqrt{N}t_{1-\alpha/2, n-1}}{s}, \bar{y} - \frac{\sqrt{N}t_{\alpha/2, n-1}}{s} \right].$$

For the interval estimation of quantiles of a normal distribution, the generalized pivotal quantity method can also be used. Define $W_4 = (n-1)s^2/\sigma^2 \sim \chi_{(n-1)}^2$. A GPQ for σ can be $\mathcal{R}_\sigma(\mathbf{D}, W_4^*) = \sqrt{(n-1)s^2/W_4^*}$, where W_4^* is treated as random. Based on Theorem 5, if we define $W_5 = \sqrt{N}(\bar{y} - \mu)/s \sim t_{n-1}$, a GPQ for μ can be $\mathcal{R}_\mu(\mathbf{D}, W_5^*) = \bar{y} - sW_5^*/\sqrt{N}$, where W_5^* is treated as random. Therefore, a GPQ for the quantile of a normal distribution $\tau_p = F^{-1}(p|\mu, \sigma)$ can be

$$\mathcal{R}_{\tau_p}(\mathbf{D}, W_4^*, W_5^*) = F^{-1}\left(p \middle| \bar{y} - sW_5^*/\sqrt{N}, \sqrt{(n-1)s^2/W_4^*}\right),$$

where W_4^*, W_5^* are treated as random and other quantities are computed based on the data $\mathbf{D}$.

Algorithm 5: Confidence interval procedure for τ_p of a normal distribution.

- (i) Based on the data $\mathbf{D}$, compute the quantities s, N and $\bar{y}$.
- (ii) Generate B realizations of W_4^* from $\chi_{(n-1)}^2$, denoted as $w_{4:1}, w_{4:2}, \dots, w_{4:B}$ and B realizations of W_5^* from t_{n-1} , denoted as $w_{5:1}, w_{5:2}, \dots, w_{5:B}$.
- (iii) Compute $r_{\tau_p:b} = F^{-1}\left(p \middle| \bar{y} - sw_{5:b}/\sqrt{N}, \sqrt{(n-1)s^2/w_{4:b}}\right)$.
- (iv) Sort the series $\{r_{\tau_p:b}; b = 1, 2, \dots, B\}$ in ascending order and use the $\alpha/2$ and $(1 - \alpha/2)$ quantiles to construct an equal-tailed $100(1 - \alpha)\%$ confidence interval for τ_p .

References

Applebaum, D., Schuermann, M., Franz, U., Bhat, B. R., Kustermans, J., and Lindsay, J. M. (2005), *Quantum Independent Increment Processes I: From Classical Probability to*

Quantum Stochastic Calculus, Springer Science & Business Media.

Meeker, W. Q. and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, vol. 314, John Wiley & Sons.

Wrench, J. W. (1968), “Concerning Two Series for the Gamma Function,” *Mathematics of Computation*, 617–626.