Analysis of Frailty-Based Competing Risk Data from Repairable Systems

Anupap Somboonsavatdee
Department of Statics, Faculty of Commerce and Accountancy,
Chulalongkorn University, Bangkok, Thailand anupap@cbs.chula.ac.th

Abstracts

The focus of study is failure history on repairable systems for which the relevant data comprise successive event times for a recurrent phenomenon along with an event-count indicator. Situations, in which individuals or systems in some populations experience recurrent events, are common in areas such as manufacturing, software debugging, risk analysis, and clinical trials. In this study, we undertake an investigation for analyzing failures from repairable systems that are subjected to multiple failure modes. Recurrent cluster of failures from multiple systems is studied both with the assumption of independent failure modes and specifically with the dependence formulation under frailty structure. Extensive simulation has been carried out that supplements the theoretical findings.

Keywords: Competing risks, frailty model, power law process, recurrent events.

1. Introduction

In this article, we undertake the investigation of a specific parametric model, known as a Power Law Process (PLP) that has found considerable attention in the repairable systems literature (Bain, 1978; Rigdon and Basu, 2000). PLP describes the repeated failure history by assuming an NHPP for the underlying counting process $N(t), t \geq 0$, with an intensity

$$\lambda(t) = \mu\beta t^{\beta-1}, \;\; t \geq 0, \mu > 0, \beta > 0,$$

matching the hazard rate of a Weibull distribution. The model, comprehensively studied first by Crow (1974), has been used widely in various application areas including, most notably, the defense industry. PLP has been popular primarily due to its attractive physical interpretation in the way of depicting the failure history of a repairable system and elegance of its statistical properties. Despite its popularity, however, little has been studied for PLP model under competing risks, which is the focus of the current investigation.

A simple approach in dealing with competing risks s to assume the multiple failure modes to act independently on the system, a framework that could be construed as overly simplistic. The premise we investigate here assumes a K-component series system where the failures of the k-th component is governed by a PLP $N_k(t)$ with intensity function

$$\lambda_k(t) = Z\mu_k\beta_k t^{\beta_k-1}, \; k = 1, \ldots, K \tag{1}$$

conditionally given a latent random variable Z. While given the frailty Z, the N_k's are assumed to be independent, the dependence among N_k's follows in an unconditional sense, upon integrating over the distribution of the shared frailty Z. The system level counting process is the superposition of all component processes, and along with the component indicators responsible for the failure, it constitutes a marked point process.

Due to limitation of number of pages allowed in this article, only selected results from the study are presented and organized as follows. Section 2 presents theoretical results for the PLP-based inference for analyzing multiple repairable systems subject to an arbitrary number of failure modes with inter-related failure processes. Section 3 details findings from extensive finite-sample simulation that supplements the results of Section 2. The article is concluded in Section 4 with a summary and discussion of some possible extensions.

2. Theoretical Results

In our framework, we shall assume the existence of m identical systems experiencing repeated failures that result from the failure of any of the K components connected in series. Let T_{ij} denote the j-th cumulative failure time of the i-th system, $i = 1, 2, \ldots, m$. Along with the failure times, we also have information on δ_{ij}, index of the mode responsible for the corresponding failure. The i-th system is observed until a censoring time τ_i, which is assumed to be noninformative of the failure process. Let n_i be the observed number of failures for the i-th system, with $n = \sum_{i=1}^m n_i$ denoting the total number of observed system failures. We shall further assume the existence of i.i.d. copies Z_i drawn from the population characterized by Z, associated with each of the m systems. Then, conditionally given Z_i, the likelihood contribution from the i-th system with (1) governing the underlying failure process is

$$L_i\left(\mu_1, \mu_2, \ldots, \mu_K, \beta_1, \beta_2, \ldots, \beta_K \mid Z_i\right) = \prod_{j=1}^{n_i} \prod_{k=1}^{K} \left[Z_i \mu_k \beta_k t_{ij}^{\beta_k - 1}\right]^{I(\delta_{ij}=k)} \times \exp\left[-Z_i \sum_{k=1}^{K} \mu_k \tau_i^{\beta_k}\right], \quad (2)$$

$I(\cdot)$ denoting the indicator function. The overall likelihood L is obtained as the product of the system specific likelihood functions. $Z_1 \ldots, Z_m$ are assumed to be i.i.d. copies of Z conforming to a continuous distribution with a probability density function g that has support on the positive real axis. In order to avoid non-identifiability, it is a common practice to assume the frailty random variable to have mean 1. A popular choice for g is the gamma pdf with parameters η^{-1}, η, so that

$$g(z) = \frac{1}{\Gamma(\eta^{-1})\eta^{\eta^{-1}}} z^{\eta^{-1}-1} \exp(-z/\eta),$$

yielding $E(Z) = 1, Var(Z) = \eta$. Let $\overline{N}_k(t)$ denotes the number of component k failures from m systems until time t. By integrating out the distribution of Z, the results in Lemma 1 follows.

Lemma 1. *With the nonnegative frailty variable $Z \sim Gamma(\eta^{-1}, \eta)$, then the followings are true.*

(a) $\overline{N}_k(t) \sim NegativeBinomial\left(\eta^{-1}, \eta\mu_k\beta_k(1 + \eta\mu_k t^{\beta_k})^{-1}\right)$,

(b) $E\left(\overline{N}_k(t)\right) = \mu_k t^{\beta_k}$, $Var\left(\overline{N}_k(t)\right) = \mu_k t^{\beta_k}(1 + \eta\mu_k t^{\beta_k})$, *and*

(c) $Cov\left(\overline{N}_k(t), \overline{N}_{k'}(t)\right) = \eta\mu_k\mu_{k'} t^{\beta_k + \beta_{k'}}$.

We shall provide the details for the case of single censoring, namely, $\tau_i = \tau$ for simplicity of exposition. With differing τ_i's, the maximum likelihood estimators no longer have closed form expressions. In practice, when τ_i's do not differ much, one can still apply the estimators from the single censoring case with $\tau = m^{-1} \sum_{i=1}^m \tau_i$, which have simple closed-form expressions, and serve as reasonable approximations.

By integrating (2) over the distribution of Z_i, the unconditional likelihood contribution from the i-th system is
$L_i(\mu_1, \ldots, \mu_K; \beta_1, \ldots, \beta_K)$

$$= \prod_{k=1}^{K} \left\{ (\mu_k \beta_k)^{n_{ik}} \prod_{j=1}^{n_i} (t_{ij}^{\beta_k - 1})^{I(\delta_{ij}=k)} \right\} \times \frac{\Gamma(n_i + \eta^{-1})}{\Gamma(1/\eta)\eta^{1/\eta}[\sum_{k=1}^{K} \mu_k \tau^{\beta_k} + \eta^{-1}]^{n_i + \eta^{-1}}},$$

where n_{ik} denotes the number of k-component failures obtained for the i-th system. Combined likelihood function based on all m systems is calculated as $L = \prod_{i=1}^{m} L_i$, which upon maximizing with respect to μ_k, β_k, η, yield the maximum likelihood estimators (MLE) for β_k, μ_k, η as shown in Lemma 2 below.

Lemma 2. *With noninformative frailty variable $Z \sim Gamma(\eta^{-1}, \eta)$, MLEs for β_k's, μ_k's, η are*

$$\hat{\beta}_k = \frac{\sum_{i=1}^{m} n_{ik}}{\sum_{i=1}^{m} \sum_{j=1}^{n_i} I(\delta_{ij} = k) \log(\tau / t_{ij})}, \quad \hat{\mu}_k = \frac{\sum_{i=1}^{m} n_{ik}}{m \tau^{\hat{\beta}_k}}, \tag{3}$$

$$\hat{\eta} = h^{-1}(0)$$

where

$$h(\eta) = -\sum_{i=1}^{m} \psi(n_i + \eta^{-1}) + m\psi(\eta^{-1}) + m \log\left(1 + \frac{m\eta}{n}\right). \tag{4}$$

for $k = 1, 2, \ldots, K$ and with ψ denoting the digamma *function.*

Note that the estimators in (3) are identical to the case when the component failure processes are independent. Further it is evident that for $\hat{\mu}_k, \hat{\beta}_k$ to be defined, at least one failure needs to be caused due to the k-th mode. $\hat{\eta}$ is obtained by solving $\partial \log L / \partial \eta = 0$ and plugging in the MLE's from (3).

Exact inference for the MLE's does not yield any tractable distributional results. Large-sample inference, however, follows in a straightforward manner. Here the vectors $\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}, \hat{\boldsymbol{\beta}}, \boldsymbol{\beta}$ denote the respective K-dimensional ensemble collectively. We present below the final result for $(\hat{\mu}_k, \hat{\beta}_k, \hat{\mu}_{k'}, \hat{\beta}_{k'})$ for any pair $k, k' \in \{1, 2, \ldots, K\}, \ \ k \neq k'$.

Theorem 1. *Let us define $Q_{1m} = \sqrt{m}(\hat{\mu}_k - \mu_k),\ Q_{2m} = \sqrt{m}(\hat{\beta}_k - \beta_k),\ Q_{3m} = \sqrt{m}(\hat{\mu}_{k'} - \mu_{k'}),\ Q_{4m} = \sqrt{m}(\hat{\beta}_{k'} - \beta_{k'})$ and $Q_m = (Q_{1m}, Q_{2m}, Q_{3m}, Q_{4m})'$. Then $Q_m \xrightarrow{d} N(0, \Omega)$ where*

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12}' & \omega_{22} \end{pmatrix}$$

with

$$\omega_{11} = \frac{1}{\tau^{\beta_k}} \begin{pmatrix} \mu_k(\beta_k \log \tau)^2 + \mu_k(1 + \eta \mu_k \tau^{\beta_k}) & -\beta_k^2 \log \tau \\ -\beta_k^2 \log \tau & \beta_k^2 / \mu_k \end{pmatrix},$$

$$\omega_{12} = \begin{pmatrix} \mu_k \mu_{k'} \eta & 0 \\ 0 & 0 \end{pmatrix}$$

and ω_{22} is identical to ω_{11} with k replaced by k'.

Theorem 1 indicates some remarkable features of the MLE's. Firstly, we note that the large-sample distribution of $\hat{\beta}_k$ is identical to that in the case where the components work independently of each other. The shared frailty induced dependence thus does not have any effect on the asymptotic distribution of $\hat{\beta}_k$. Further, $\hat{\beta}_k$ is asymptotically independent of the MLE's of the parameters specific to any other component. The large-sample distribution of $\hat{\mu}_k$, on the other hand, involves η. The difference between the asymptotic variances in the cases of independence and frailty-induced dependence increases linearly with η, and so does the large-sample covariance between $\hat{\mu}_k$ and $\hat{\mu}_{k'}$. The parametric model based estimator of the mean number of system failures $\Lambda(t)$ is $\hat{\Lambda}(t) = \sum_{k=1}^{K} \hat{\mu}_k t^{\hat{\beta}_k}$, which, at the censoring time τ, simplifies to the nonparametric estimator n/m. A $100(1-\alpha)\%$ approximate confidence interval for $\Lambda(t)$ is given by

$$\hat{\Lambda}(t) \pm z_{1-\alpha/2}\sqrt{\hat{\Lambda}(t)\left(1+\eta\hat{\Lambda}(t)\right)}.$$

The standard error explicitly indicates the extra-Poisson variation induced by dependence.

Remarks

1. The above is presented for the single censoring case. The extension to the multiple censoring case with differing τ_i is straightforward in principle. The expressions, however, are substantially more cumbersome. In particular, the MLE's no longer possess the attractive closed-form structures.

2. The contrast between the large-sample results in the single system and the multiple system case is worth a mention. Somboonsavatdee and Sen (2013) has demonstrated some peculiarities of the inference of the MLE's in the single sample case where the large-sample results are applied as the number of repeated failures from the system grows large. In that case, the MLE's of μ_k's suffer from a pretty slow rate of convergence. Further, the asymptotic variance-covariance matrix turns out to be singular, highlighting the contrast with the multiple system case which is devoid of such pathologies.

3. Simulation Results

We carried out extensive simulation in order to investigate finite-sample properties of the MLE's, especially for a small number of systems. In order to keep the exposition simple, we kept our investigation confined to a 2-mode system. Among the many parameter combinations tested, we report the findings for three sets of $(\mu_1, \beta_1, \mu_2, \beta_2)$ namely (4, 0.25, 0.25, 2), (1, 0.75, 0.3, 1.25), and (0.5, 1.5, 5, 0.8). A single-censoring scheme was simulated with censoring time τ fixed at 20. The three scenarios yield the mean number of failures to be (8.5, 100), (9.5, 12.7), (44.7, 55) by time τ, respectively, for the two modes. While the first of these corresponds to the situation with one dominating mode, the remaining two cases represent a 2-mode system where the modes contribute similarly, with the two cases differing in terms of the propensity of failures. The number of systems (m) tested ranged between 10 and 50, while the frailty variance η was varied between a small value of 0.1 to a large value of 5 representing a low to a high degree of dependence.

Table 1 and Table 2 exhibit the bias and relative efficiencies of the MLE's for the three sets of parameter values, $m = 10, 50$, and $\eta = 1, 5$ with simulation size of 5000.

The relative efficiency (RE) is defined as the ratio of the large-sample (theoretical) variance from Theorem 1, and the variance calculated from the simulated samples. For the μ parameters a second relative efficiency (RE_{Ind}) measure is calculated for comparative purposes, where the numerator of RE is replaced by the large-sample expression under the assumption of independence of the failure modes. This provides an assessment of mis-specification arising from ignoring the underlying dependence between the failure modes. As is evident from Table 1 and Table 2, the bias in the MLE's for both the scale and the shape parameters are quite small, even for small m, rarely going beyond 2%. Further, the bias does not seem to be affected by the extent of dependence. The relative efficiency (RE) of the scale parameter MLE's are close to 100% regardless of the values of m and η, demonstrating close agreement between asymptotic and finite-sample variability. It seems, however, that the large-sample variance of $\hat{\beta}_1, \hat{\beta}_2$ severely underestimate the true variability for small to moderate m coupled with a large degree of dependence. This is indicative of a hidden role played by the frailty in the variability of the shape parameters. As expected, ignoring the dependence between the failure modes results in severe underestimation of $Var(\hat{\mu}_1), Var(\hat{\mu}_2)$.

Based on our findings, we recommend estimating the dependence parameter η first for analyzing a dataset under the frailty induced PLP framework. If the estimate $\hat{\eta}$ exceeds 1, or the number of systems tested is small (say $m \leq 30$), we recommend using bootstrap methods.

Table 1: Bias and Relative Efficiencies of the MLE's for different sets of parameters, $T = 20, \eta = 1$

No. of Systems	$\mu_1 = 4,\ \beta_1 = 0.25,\ \mu_2 = 0.25,\ \beta_2 = 2$									
(m)	Bias($\hat{\mu}_1$)	RE($\hat{\mu}_1$)	$\text{RE}_{\text{Ind}}(\hat{\mu}_1)$	Bias($\hat{\beta}_1$)	RE($\hat{\beta}_1$)	Bias($\hat{\mu}_2$)	RE($\hat{\mu}_2$)	$\text{RE}_{\text{Ind}}(\hat{\mu}_2)$	Bias($\hat{\beta}_2$)	RE($\hat{\beta}_2$)
10	-0.024	1.008	0.157	0.003	0.842	0.003	0.980	0.264	0.002	0.879
50	-0.007	0.984	0.153	0.001	0.937	0.001	0.981	0.264	0.001	1.000
	$\mu_1 = 1,\ \beta_1 = 0.75,\ \mu_2 = 0.3,\ \beta_2 = 1.25$									
10	0.012	0.983	0.383	0.008	0.857	0.007	0.985	0.534	0.012	0.879
50	0.000	1.007	0.393	0.001	0.974	0.001	1.000	0.542	0.003	0.949
	$\mu_1 = 0.5,\ \beta_1 = 1.5,\ \mu_2 = 5,\ \beta_2 = .8$									
10	0.000	1.028	0.330	0.004	0.921	-0.024	1.012	0.111	0.001	0.868
50	0.002	1.007	0.324	0.001	0.998	0.009	1.016	0.111	0.000	1.004

Table 2: Bias and Relative Efficiencies of the MLE's for different sets of parameters, $T = 20, \eta = 5$

No. of Systems	$\mu_1 = 4,\ \beta_1 = 0.25,\ \mu_2 = 0.25,\ \beta_2 = 2$									
(m)	Bias($\hat{\mu}_1$)	RE($\hat{\mu}_1$)	$\text{RE}_{\text{Ind}}(\hat{\mu}_1)$	Bias($\hat{\beta}_1$)	RE($\hat{\beta}_1$)	Bias($\hat{\mu}_2$)	RE($\hat{\mu}_2$)	$\text{RE}_{\text{Ind}}(\hat{\mu}_2)$	Bias($\hat{\beta}_2$)	RE($\hat{\beta}_2$)
10	-0.051	1.024	0.036	0.007	0.309	0.001	1.025	0.070	0.004	0.544
50	-0.032	0.985	0.035	0.001	0.892	-0.001	0.979	0.067	0.001	0.863
	$\mu_1 = 1,\ \beta_1 = 0.75,\ \mu_2 = 0.3,\ \beta_2 = 1.25$									
10	0.005	0.994	0.113	0.020	0.252	0.009	1.005	0.192	0.021	0.192
50	0.007	1.000	0.113	0.001	0.888	0.003	0.980	0.188	0.003	0.881
	$\mu_1 = 0.5,\ \beta_1 = 1.5,\ \mu_2 = 5,\ \beta_2 = .8$									
10	0.010	0.964	0.083	0.006	0.452	0.023	0.957	0.023	0.004	0.374
50	-0.003	1.003	0.087	0.001	0.888	-0.042	1.007	0.024	0.001	0.869

4. Discussions

In this article we have provided an investigation of statistical inference for failure data arising from multiple repairable systems that are subject to competing risks. The setup is pertinent to a Failure Mode Effect Analysis (FMEA) program often pursued for complex multi-component systems with components connected in series. The main focus of the article is on carrying out the inference under a shared frailty structure that induces dependence across the component processes. The discourse has been confined to PLP with a Gamma frailty, that can be construed as an extension to a popular parametric framework in this context. The Poisson process framework is not critical in this context. As long as the mean model is specified correctly for the component recurrent processes, the general theory of estimating equations can be applied to ensure the large-sample normality of the solutions to the estimating equations under some mild conditions. The estimators would result from a set of equations that are slight generalizations of those provided in Lawless and Nadeau (1995). Asymptotic normality of the estimators would follow from steps similar to those outlined in the Appendix of Lawless and Nadeau (1995).

In our setup, the censoring is considered to be noninformative of the recurrent processes, an assumption that may be deemed inappropriate in contexts where recurrent events such as hospitalization are terminated by a related censoring event such as death. Although not pursued in this article, it is conceivable that dependent censoring can be incorporated by joint modeling of the recurrence and censoring time, in a manner analogous to Liu et al (2004).

The shared frailty formulation studied in this article is equivalent to a generalized linear mixed effects (GLIMMIX) model for the component specific counting processes. Specifically, one can express the model for $\Lambda_{ik}(t)$ until time t for the k-th failure mode and the i-th system as

$$\log \Lambda_{ik}(t) = \alpha_{ik} + \beta_k \log t, \quad i = 1, 2, \ldots, m;\ k = 1, \ldots, K,$$

with α_{ik} denoting the random intercept while β_k is the fixed slope. In contrast to the usual GLIMMIX, however, α_{ik} is modeled as a log-gamma random variable.

The general approach pursued in this article is valid for other parametric framework as well. Note that methodologies exist for carrying out inference in a non-parametric setting in the context of analyzing recurrent events under competing risk. The investigation in that case are restricted to mean or rate of cause-specific recurrence and are not valid for time beyond the largest censoring time. The inference in a parametric setting, on the other hand, is more comprehensive and can be extrapolated beyond the observation point. The trade-off, of course, lies in the robustness of model assumptions. We strongly believe, nonetheless, that the investigation presented herein provides an important extension to perhaps one of the most widely used parametric models in analyzing failure data from repairable systems.

References

Bain, L. J. (1978), *Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods*, Volume 24, New York: Marcel Dekker.

Crow, L. H. (1974), “Reliability analysis for complex, repairable systems,” *Reliability and Biometry*, F. Proschan and R.J. Serfling, eds., SIAM, Philadelphia, 379- 410.

Lawless, J. F. and Nadeau, C. (1995), “Some simple robust methods for the analysis of recurrent events,” *Technometrics*, 37, 158-168.

Liu, L., Wolf, R. A., and Huang, X. (2004), “Shared frailty model for recurrent events and a terminal event,” *Biometrics*, 60, 747–756.

Rigdon, S. E. and Basu, A. P. (2000), *Statistical Methods for the Reliability of Repairable Systems*, New York: Wiley.

Somboonsavatdee, A. and Sen, A.(2013) “Statistical Inference for Power Law Process with Competing Risks,” under review.