Special Proceedings: ISBN #: 978-81-950383-8-1
27th Annual Conference, 21-23 February 2025; pp 1-11

Failure Time Prediction Model for an Injection Molding System

Pritam Ranjan[1], Arnab Koley[1], Sandip K. Pal[1] and Debasis Kundu[2]

[1] *OMQT Area, Indian Institute of Management Indore, MP, India*
[2] *Dept. of Mathematics and Statistics, Indian Institute of Technology Kanpur, UP, India*

Received: 01 May 2025; Revised: 11 May 2025; Accepted: 15 May 2025

Abstract

In industrial sectors, understanding machine behaviour in real-time is crucial for minimizing unscheduled downtime and maximizing production with expected quality. Advanced machines like injection molding machine used for manufacturing plastic bottles for soft drinks are equipped with sensors that record event log times. We adopt a hierarchical parametric model to predict machine failure time based on its current state, that are, "running with alerts," "running without alerts," and system breakdown. The model utilizes Weibull distribution for the event duration to predict failure times.

Key words: Generalized linear regression; Predictive modeling; Count data distribution; Lifetime distribution.

AMS Subject Classifications: 62N05, 90B25

1. Injection molding machine

Running manufacturing equipment involves maintenance of machines on a regular basis. Preventive maintenance is a popular and well accepted approach, however, such tasks are carried out according to a timetable, and not always done when the equipment specifically calls for them. Thus, it is crucial to predict machine failures with enough lead time.

Several predictive models have been proposed by different authors to predict the failure time using the sequence of events. Li *et al.* (2007) developed a Cox-proportional hazard (CPH) model to predict the time to failure model. Luo *et al.* (2014) proposed a framework which consists of three stages: data pre-processing, event extraction, and correlation analysis. In the data pre-processing stage. A few other works on the correlation based event prediction model are Motahari-Nezhad *et al.* (2011); Wu *et al.* (2010); Zhu and Shasha (2002); Lou *et al.* (2010). Agrawal *et al.* (1993) use association mining to learn a pattern based on a historical sequence of past events to predict the probable occurrence of next event(s). In retail sector, the market basket analysis has been recognized as a proven

Corresponding Author: Pritam Ranjan
Email: pritamr@iimidr.ac.in

and successful application of association rule mining for cross selling, product placement, promotion affinity analysis, and product promotion and targeting (Kohavi *et al.* , 2004), for mining gene sequence expression (Jiang and Gruenwald , 2005) and for web-log mining (Huang and An , 2002; Rudin *et al.* , 2011).

The main objective of this study is to propose a time to failure model of an injection molding (IM) machine for a plastic soft drink bottle (see Figure 1).

Figure 1: An injection molding (IM) machine schematic diagram (source: https://prototechasia.com/en/injection-molding/questions-injection-molding)

Industry 4.0 brings forth intelligent machines equipped with sophisticated sensors, embedded software, and robotics which gather and store data as machine logs in a semi-structured format. These data are usually collected while machine is in running condition, and primarily consists of operation events, performance counters, and alert messages, among others. This research focuses on the system logs data that are captured through various sensors mounted in the IM machine (see Figure 2).

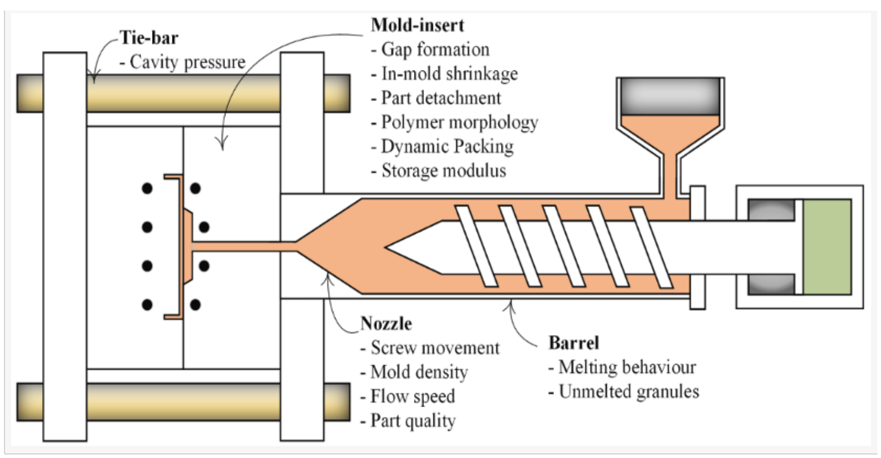

Figure 2: A schematics of IM machine with various sensors.

Figure 3 depicts different operational sequence of a few events of an IM machine which typically provide sufficient information for engineers to diagnose the working condition of equipment.

The alert messages have been clubbed into three groups. When the machine is running smoothly and does not produce any message or alert we label it as "running without alert". Alternatively, when the machine is running but generate some warning or requires human

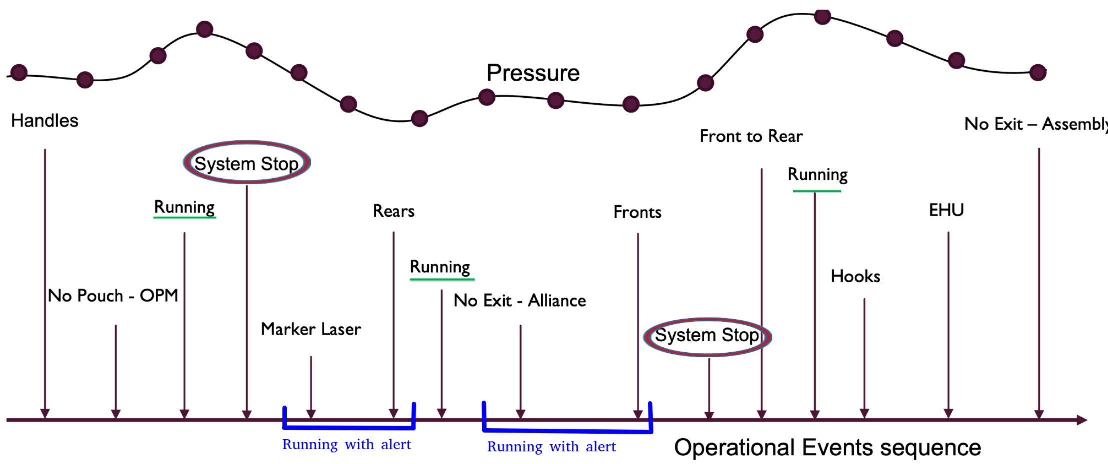

Figure 3: An example of sequence of events while machine is in operational condition.

intervention, we call it as "running with alert". Finally, "failure" refers to the state of the machine when the system is down and requires maintenance.

Furthermore, if there are two or more consecutive occurrences of the same type of events (say, "running with alert") then Pal *et al.* (2024) clubbed them together as one event. This implied that "running without alert" and "running with alert" will occur alternatively followed by "failure". It may also be possible that the machine experiences only one type of events, say "running without alert" or "running with alert" in an epoch before "failure". One sequence of events until the failure is also referred to as an epoch. In this illustrative image only two epochs are shown for the purpose of understanding. Figure 4 illustrate the sequence of events leading to failure.

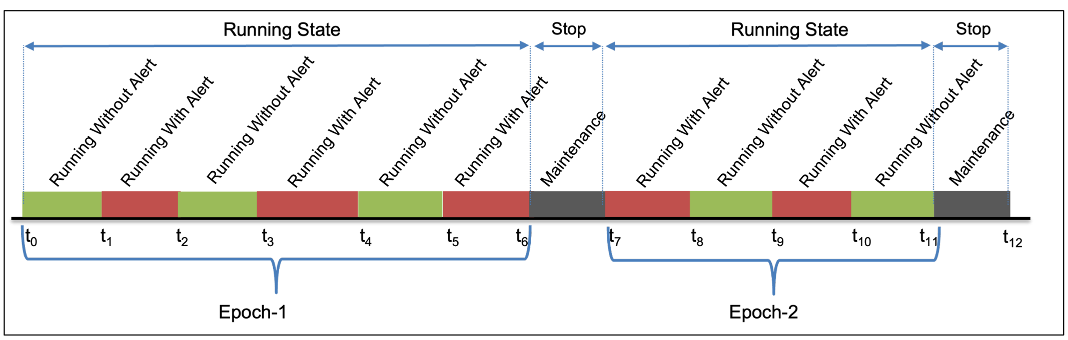

Figure 4: Illustration of a snapshot of different states of the machine (*i.e.*, "running without alert", "running with alert" and "failure").

The data considered by Pal *et al.* (2024) consists of 45 epochs, in which the total number of "running with alert" events is 1584, whereas the total number of "running without alert" events is 1606 (*i.e.*, 3190 running events and 45 failures). Moreover, the IM machine considered here consists of 72 different sensors that may explain the reasons behind the time spent on the three states. These sensors are majorly related to *mold surface temperature*, *cooling rate of cavities*, *post gate cavity pressure*, *filled area of post gate cavity*, *filled area of molding*, *injection fill time*, *screw runtime*, *etc.*

2. Model developed by Pal *et al.* (2024)

This section summarizes the key aspects of the failure time prediction methodologies developed by Pal *et al.* (2024).

1. Pre-processing of the data: the machine states are labelled into three categories: failure, running with alert, and running without alert. Furthermore, the consecutive (different) alerts with the same label (*e.g.*, running with alert) are clubbed together as the same state / event.

2. Distributional assumption of the key variables:

   (a) Number of events per epoch (R_i): Since $R_i \geq 1$ (0 is not possible), a shifted Poisson distribution is assumed. The probability mass function (PMF) is

   $$P(R_i = r_i) = e^{-\mu} \; \frac{\mu^{(r_i-1)}}{(r_i-1)!} \; ; \; r_i = 1, 2, \cdots . \tag{1}$$

   (b) Duration of a "running without alert" event j in epoch i (denoted by X_{ij}^1): exponential with rate parameter λ_1

   (c) Duration of a "running with alert" event j in epoch i (denoted by X_{ij}^2): exponential with rate parameter λ_2

3. Let N_i^1 be the number of "running without alert" events in the i-th epoch, and N_i^2 refers to the number of "running with alert" event in the i-th epoch. Grouping of epochs into four situations:

   (a) **Situation 1**: The epoch starts with the event "running without alert" and the number of events r_i is odd. Hence $N_i^1 = \frac{r_i+1}{2}$ and $N_i^2 = \frac{r_i-1}{2}$.

   (b) **Situation 2**: The epoch starts with the event "running without alert" and the number of events r_i is even. Hence $N_i^1 = \frac{r_i}{2}$ and $N_i^2 = \frac{r_i}{2}$.

   (c) **Situation 3**: The epoch starts with the event "running with alert" and the number of events r_i is odd. Hence $N_i^1 = \frac{r_i-1}{2}$ and $N_i^2 = \frac{r_i+1}{2}$.

   (d) **Situation 4**: The epoch starts with the event "running with alert" and the number of events r_i is even. Hence $N_i^1 = \frac{r_i}{2}$ and $N_i^2 = \frac{r_i}{2}$.

4. Likelihood calculation: the likelihood for Situation 1 can be written as:

   $$L_1(\theta) = c_1 \prod_{i \in S_1} \left[P(R_i = r_i) \times p \times \prod_{j=1}^{\frac{r_i+1}{2}} f^1(x_{ij}^1) \times \prod_{j=1}^{\frac{r_i-1}{2}} f^2(x_{ij}^2) \right], \tag{2}$$

   where, $f^k(.)$ is the probability density function (PDF) of exponential distribution with mean $1/\lambda_k$, for $k = 1, 2$, p is the probability of the first event being "running without alert", and c_1 is the proportionality constant independent of the parameters θ. After ignoring the constant, using appropriate PDFs and PMF in the above likelihood function, and taking natural-log we get the log-likelihood,

$$\begin{aligned}\mathcal{L}_1(\theta) = &- n_1\ \mu + ln(\mu)\sum_{i\in S_1}(r_i-1) - \sum_{i\in S_1} ln((r_i-1)!) + ln(\lambda_1)\sum_{i\in S_1}\left(\frac{r_i+1}{2}\right) \\ &- \lambda_1 \sum_{i\in S_1}\sum_{j=1}^{\frac{r_i+1}{2}} x_{ij}^1 + ln(\lambda_2)\sum_{i\in S_1}\left(\frac{r_i-1}{2}\right) - \lambda_2\sum_{i\in S_1}\sum_{j=1}^{\frac{r_i-1}{2}} x_{ij}^2 + n_1 ln(p).\end{aligned} \tag{3}$$

For other situations, the likelihood expression will be similar and the readers can refer to Appendix A1 in Pal *et al.* (2024). Subsequently, the log-likelihood of the data from all n epochs and four situations can be written as, $\mathcal{L}(\theta) = \mathcal{L}_1(\theta)+\mathcal{L}_2(\theta)+\mathcal{L}_3(\theta)+\mathcal{L}_4(\theta)$. The parameter vector $\theta = (\lambda_1, \lambda_2, p, \mu)$ is estimated by maximizing $\mathcal{L}(\theta)$. By defining $N_{il}^s = \frac{r_i+a_l^s}{2}$ and

$$a_l^s = \begin{cases}(-1)^{s+1},\ \text{if } l=1\\ 0,\ \text{if } l=2,4\\ (-1)^s,\ \text{if } l=3\end{cases}$$

for $s = 1, 2$, the closed form analytical expression of the maximum likelihood estimators (MLEs) are given by

$$\widehat{\lambda}_s = \frac{\sum_{l=1}^{4}\sum_{i\in S_l} N_{il}^s}{\sum_{l=1}^{4}\sum_{i\in S_l}\sum_{j=1}^{N_{il}^s} x_{ij}^s}\ \text{for } s=1,2,\quad \widehat{p} = \frac{n_1+n_2}{n}\ \text{and } \widehat{\mu} = \frac{1}{n}\sum_{l=1}^{4}\sum_{i\in S_l}(r_i-1). \tag{4}$$

5. Since there are 72 sensors, important ones that might influence the current state of the machine are identified via variable importance method within the random forest model framework.

6. Subsequently, these m important sensor-based covariates are introduced in the model via generalized linear regressors. That is, the generalized linear model (GLM) considered for λ_1, λ_2 and μ in i-th epoch can be written as,

$$\lambda_{1i} = \exp\left(\beta_0 + \sum_{k=1}^{m} F_{ki}\beta_k\right),\ \lambda_{2i} = \exp\left(\gamma_0 + \sum_{k=1}^{m} F_{ki}\gamma_k\right),\ \mu_i = \exp\left(\eta_0 + \sum_{k=1}^{m} F_{ki}\eta_k\right),$$

where $\beta_k, \gamma_k, \eta_k$ for $k = 0, 1, \cdots, m$ denote the unknown regression coefficients and F_{ki} denotes the k-th sensor value in the i-th epoch.

7. Next, the MLEs of these regression parameters are obtained using numerical optimization. Additionally, uncertainty bounds for these estimates are obtained through non-parametric (asymptotic and Bootstrap) confidence intervals.

8. Finally, Pal *et al.* (2024) addressed the main objective of the paper, *i.e.*, the derivation of the expected time to fail for the IM machine. Given that the epochs in four situations are based on whether the number of events is even or odd, and whether the first event

is of "running with alert" or "running without alert", the expected time to fail can be written as:

$$E[\text{Time to fail}] = \frac{(1-e^{-2\mu})(\mu+1)}{4}\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\right) + \frac{(1+e^{-2\mu})}{4}\left[\frac{\mu+2p}{\lambda_1}+\frac{\mu+2(1-p)}{\lambda_2}\right]. \qquad (5)$$

In practice, the values of $\lambda_1, \lambda_2, \mu$ and p are required, which in-turn requires the values of covariates, to compute the expected time to fail for an out-of-sample epoch. Pal *et al.* (2024) have taken the epoch-wise average value of covariates for comparison the model performance. Alternatively, one can take the average sensor values across all 45 epochs (*i.e.*, over 3190 events) to estimate the expected time to fail. Of course, if we know the true values of the sensors, one can use that instead, however these values are typically not known in advance.

Pal *et al.* (2024) implemented the methodology on the data obtained from the IM machine that manufactures softdrink bottles. The performance comparison of the actual time to fail with the expected time to fail derived in Step 8, and the popular Cox-proportional hazard (CPH) model is presented in Figure 5.

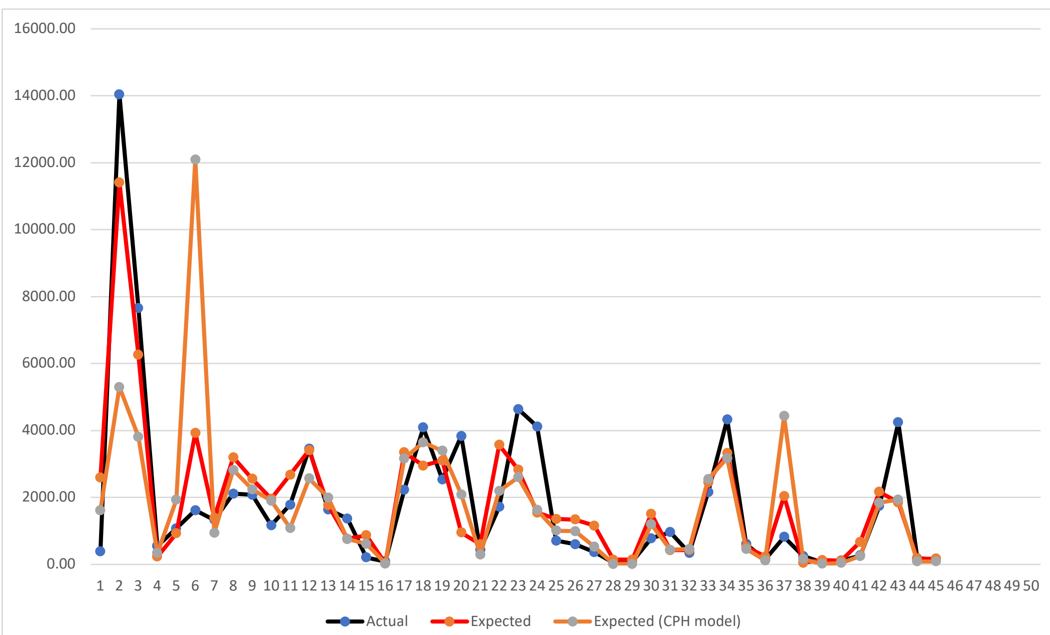

Figure 5: Plot of epoch-wise actual and expected time to fail by the proposed model and the CPH model.

The visual comparison between the three sets of values in Figure 5 clearly show the superior performance of the proposed model as compared to the CPH model. However, one can compute various goodness of fit measures for quantitative comparison as well. Table 1 presents the mean square error (MSE), mean absolute error (MAE), maximum error (MaxE) and correlation between actual data and the proposed model.

Table 1: Performance of the proposed model vs CPH model

Model	MSE	MAE	MaxE	Correlation
Proposed	1299394.36	796.38	2881.12	0.89
CPH	5389542.03	1102.77	10477.12	0.48

3. Proposed extension

Although the model proposed by Pal *et al.* (2024) demonstrates superior performance than the popular CPH model, there is a room for further investigation and possible improvement. For instance, the distribution of X_{ij}^k (j-th event of type k ($k = 1, 2$) in epoch i), the time duration spent by the machine on a given state (*i.e.*, duration of "running without alert" or "running with alert") was assumed to be exponential because of popularity and simplicity. It turns out that Weibull distribution is more general and hence a better choice than exponential for modeling X_{ij}^k. This paper discusses the key expressions of Pal *et al.* (2024) that need to be modified as per the Weibull distribution.

Let $X_{ij}^k \sim Weibull(\alpha_k, \lambda_k)$, for $k = 1, 2$. For simplicity, one can take identical shape parameters, *i.e.*, $\alpha_k = \alpha$. As a result, the PDF of X_{ij}^k is given by

$$f_k(x) = \lambda_k \alpha x^{\alpha-1} e^{-\lambda_k x^\alpha},$$

with mean $\frac{1}{\lambda^{1/k}}\Gamma(1 + \frac{1}{\alpha})$.

First, the likelihood in (2), and for other situations, will be modified as

$$L_1(\theta) = c_1 \prod_{i \in S_1} \left[P(R_i = r_i) \times p \times \prod_{j=1}^{\frac{r_i+1}{2}} \left\{ \lambda_1 \alpha (x_{ij}^1)^{\alpha-1} e^{-\lambda_1 (x_{ij}^1)^\alpha} \right\} \times \prod_{j=1}^{\frac{r_i-1}{2}} \left\{ \lambda_2 \alpha (x_{ij}^2)^{\alpha-1} e^{-\lambda_2 (x_{ij}^2)^\alpha} \right\} \right].$$

This leads to the update of the log-likelihood expression in (3) as

$$\begin{aligned} \mathcal{L}_1(\theta) \;=\; & -n_1\, \mu + ln(\mu) \sum_{i \in S_1} (r_i - 1) - \sum_{i \in S_1} ln((r_i - 1)!) + ln(\alpha) \sum_{i \in S_1} r_i + n_1 ln(p) \\ & + ln(\lambda_1) \sum_{i \in S_1} \left(\frac{r_i + 1}{2} \right) - \lambda_1 \sum_{i \in S_1} \sum_{j=1}^{\frac{r_i+1}{2}} (x_{ij}^1)^\alpha + (\alpha - 1) \sum_{i \in S_1} \sum_{j=1}^{\frac{r_i+1}{2}} ln(x_{ij}^1) \\ & + ln(\lambda_2) \sum_{i \in S_1} \left(\frac{r_i - 1}{2} \right) - \lambda_2 \sum_{i \in S_1} \sum_{j=1}^{\frac{r_i-1}{2}} (x_{ij}^2)^\alpha + (\alpha - 1) \sum_{i \in S_1} \sum_{j=1}^{\frac{r_i-1}{2}} ln(x_{ij}^2). \end{aligned}$$

AS earlier, the total log-likelihood of the data from all n epochs and four situations, $\mathcal{L}(\theta) = \mathcal{L}_1(\theta) + \mathcal{L}_2(\theta) + \mathcal{L}_3(\theta) + \mathcal{L}_4(\theta)$, can be maximized to obtain

$$\widehat{\lambda}_s = \frac{\sum_{l=1}^{4} \sum_{i \in S_l} N_{il}^s}{\sum_{l=1}^{4} \sum_{i \in S_l} \sum_{j=1}^{N_{il}^s} (x_{ij}^s)^{\widehat{\alpha}}} \text{ for } s = 1, 2, \quad \widehat{p} = \frac{n_1 + n_2}{n} \text{ and } \widehat{\mu} = \frac{1}{n} \sum_{l=1}^{4} \sum_{i \in S_l} (r_i - 1), \tag{6}$$

where, $\widehat{\alpha}$ can be obtained by maximizing the profile log-likelihood function of α, given by,

$$\begin{aligned} g(\alpha) &= ln(\alpha)\sum_{l=1}^{4}\sum_{i\in S_l} r_i - \sum_{s=1}^{2}\sum_{l=1}^{4}\sum_{i\in S_l}\left(\frac{r_i+a_l^s}{2}\right) ln\left(\sum_{l=1}^{4}\sum_{i\in S_l}\sum_{j=1}^{N_{il}^s}(x_{ij}^s)^{\alpha}\right) \\ &\quad +\alpha\sum_{s=1}^{2}\sum_{l=1}^{4}\sum_{i\in S_l}\sum_{j=1}^{N_{il}^s} ln(x_{ij}^s). \qquad (7) \end{aligned}$$

The uniqueness of $\widehat{\alpha}$ can be established with the help of the following two theorems

Theorem 1: The profile log-likelihood of α, given by, $g(\alpha)$ in (7) is a concave function.

Proof: We skip the derivation of the first derivative of $g(\alpha)$ and directly jump to the second derivative of $g(\alpha)$, *i.e.*,

$$\frac{d^2 g(\alpha)}{d\alpha^2} = -\frac{1}{\alpha^2}\sum_{l=1}^{4}\sum_{i\in S_l} r_i - \sum_{s=1}^{2}\sum_{l=1}^{4}\sum_{i\in S_l}\left(\frac{r_i+a_l^s}{2}\right)(D_s(\alpha)-E_s(\alpha))\left(\sum_{l=1}^{4}\sum_{i\in S_l}\sum_{j=1}^{N_{il}^s}(x_{ij}^s)^{\alpha}\right)^{-2},$$

where, $D_s = \sum_{l=1}^{4}\sum_{i\in S_l}\sum_{j=1}^{N_{il}^s}(x_{ij}^s)^{\alpha}\sum_{l=1}^{4}\sum_{i\in S_l}\sum_{j=1}^{N_{il}^s}(x_{ij}^s)^{\alpha}(ln(x_{ij}))^2$ and $E_s = \left(\sum_{l=1}^{4}\sum_{i\in S_l}\sum_{j=1}^{N_{il}^s}(x_{ij}^s)^{\alpha} ln(x_{ij})\right)^2$.

Using Cauchy-Schwarz inequality, we get $D_s(\alpha)-E_s(\alpha) \geq 0$ confirming $d^2g(\alpha)/d\alpha^2 \leq 0$. Hence $g(\alpha)$ is a concave function. □

Theorem 2: The profile log-likelihood of α, given by, $g(\alpha)$ in (7) has a unique maximum.

Proof: Given Theorem 1, we only need to show that the first order derivative of $g(\alpha)$ has a unique root. Note that $dg(\alpha)/d\alpha = 0$ can be written as $G(\alpha) - H(\alpha) = 0$,

where, $G(\alpha) = \dfrac{\sum_{l=1}^{4}\sum_{i\in S_l} r_i}{\alpha} + \sum_{s=1}^{2}\sum_{l=1}^{4}\sum_{i\in S_l}\sum_{j=1}^{N_{il}^s} ln(x_{ij}^s)$,

and $$H(\alpha) = \sum_{s=1}^{2}\sum_{l=1}^{4}\sum_{i\in S_l}\left(\frac{r_i+a_l^s}{2}\right)\frac{\sum_{l=1}^{4}\sum_{i\in S_l}\sum_{j=1}^{N_{il}^s}(x_{ij}^s)^{\alpha} ln(x_{ij}^s)}{\left(\sum_{l=1}^{4}\sum_{i\in S_l}\sum_{j=1}^{N_{il}^s}(x_{ij}^s)^{\alpha}\right)}. \qquad (8)$$

Clearly $G(\alpha)$ is a decreasing function of α. Also the first order derivative of $H(\alpha)$ is

$$\frac{d}{d\alpha}H(\alpha) = \sum_{s=1}^{2}\sum_{l=1}^{4}\sum_{i\in S_l}\left(\frac{r_i+a_l^s}{2}\right)\frac{D_s(\alpha)-E_s(\alpha)}{\left(\sum_{l=1}^{4}\sum_{i\in S_l}\sum_{j=1}^{N_{il}^s}(x_{ij}^s)^{\alpha}\right)^2},$$

where, $D_s(\alpha)$ and $E_s(\alpha)$ are defined in Theorem 1. By using Cauchy-Schwarz inequality, it can be noted that $dH(\alpha)/d\alpha \geq 0$, ensuring $H(\alpha)$ is an increasing function of α. Since $G(\alpha)$ is a decreasing function of α and the function $g(\alpha)$ has at-least one maximum, it is clear that $G(\alpha)$ and $H(\alpha)$ intersect at only one point ensuring unique solution of (8). Hence $g(\alpha)$ has the unique maximum value. □

Using Theorem 1 and Theorem 2, it is proved that $\widehat{\alpha}$ exists and is unique. By using invariance property of MLE, $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ are also unique.

As proposed by Pal *et al.* (2024), the sensors-based covariates are used to form generalized linear regression with the same re-parametrization as follows,

$$\lambda_{1i} = \exp\left(\beta_0 + \sum_{k=1}^{m} F_{ki}\beta_k\right), \ \lambda_{2i} = \exp\left(\gamma_0 + \sum_{k=1}^{m} F_{ki}\gamma_k\right), \ \mu_i = \exp\left(\eta_0 + \sum_{k=1}^{m} F_{ki}\eta_k\right).$$

For the sake of simplicity, the parameter α is not parametrized in terms of the covariates although one can re-parametrize it if needed as $\alpha_i = \exp\left(\zeta_0 + \sum_{k=1}^{m} F_{ki}\zeta_k\right)$, for some unknown regression coefficients ζ_k for $k = 0, 1, \cdots m$.

Following the similar approach by Pal *et al.* (2024), the expression for the expected time to fail can be written as:

$$\begin{aligned} E[\text{Time to fail}] &= \Gamma\left(1 + \frac{1}{\alpha}\right)\left[\frac{(1 - e^{-2\mu})(\mu + 1)}{4}\left(\frac{1}{\lambda_1^{1/\alpha}} + \frac{1}{\lambda_2^{1/\alpha}}\right)\right. \\ &\left. + \frac{(1 + e^{-2\mu})}{4}\left[\frac{\mu + 2p}{\lambda_1^{1/\alpha}} + \frac{\mu + 2(1 - p)}{\lambda_2^{1/\alpha}}\right]\right]. \end{aligned} \tag{9}$$

Therefore after substituting the values of $\widehat{\alpha}$, $\widehat{\lambda}_1$, $\widehat{\lambda}_2$ and $\widehat{p}$ in (9), the estimated time to fail of the machine is obtained.

4. Concluding remarks

This study extends the model proposed by Pal *et al.* (2024) to analyze sequential data from an IM machine, focusing on alternating periods of operation with alerts and without alerts, resulting in machine failure. The durations with alerts is assumed to follow Weibull distribution with scale parameter λ_1 and shape parameter α, while durations without alerts is assumed to follow Weibull distribution independently with scale parameter λ_2 and the same shape parameter α, allowing for flexible modeling. Notably setting $\alpha = 1$ recovers the earlier model by Pal *et al.* (2024) as a special case. The number of events before failure is modeled using a conditional Poisson distribution, given at least one has happened prior to failure. We have derived maximum likelihood estimators for the parameters and used these to formulate the expected time to machine failure. However we have not reported the numerical findings of the proposed model and we at this stage leave it for a future work.

Acknowledgements

Our sincere thanks to the organizers of the conference for providing a wonderful opportunity to discuss the paper with eminent researchers around the globe. We are also grateful to the Editors and reviewers for their valuable comments and suggestions.

Conflict of interest

The authors do not have any financial or non-financial conflict of interest to declare for the research work included in this article.

References

Agrawal, R., Imieliński, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, **22**, 207-216.

Ding, J., Liu, Y., Zhang, L., and Wang, J. (2015). Modeling the process of event sequence data generated for working condition diagnosis. *Mathematical Problems in Engineering*, https://doi.org/10.1155/2015/693450.

Efron, B., and Tibshirani, R.J. (1994). *An Introduction to the Bootstrap.* Chapman and Hall/CRC, New York.

Glazier, S. W. (2019). *Sequential Survival Analysis with Deep Learning.* Brigham Young University, Master of Science Thesis.

Hastie, T., Tibshirani, R., and Friedman, J. (2013). *The Elements of Statistical Learning.* Springer, New York.

Huang, X., and An, A. (2002). Discovery of interesting association rules from livelink web log data. *IEEE International Conference on Data Mining, 2002 Proceedings, Maebashi City, Japan.*

Jiang, X-R., and Gruenwald, Le. (2005). Microarray gene expression data association rules mining based on BSC-Tree and FIS-Tree. *Data & Knowledge Engineering*, **53**, 3–29.

Kohavi, R., Mason, L., Parekh, R., and Zheng, Z. (2004). Lessons and challenges from mining retail e-commerce data. *Machine Learning*, **57**, 83-113.

Li, Z., Zhou, S., Choubey, S., and Sievenpiper, C. (2007). Failure event prediction using the Cox proportional hazard model driven by frequent failure signatures. *IIE transactions*, **39**, 303-315.

Lou, J. G., Fu, Q., Wang, Y., and Li, J. (2010). Mining dependency in distributed systems through unstructured logs analysis. *ACM SIGOPS Operating Systems Review*, **44**, 91-96.

Luo, C., Lou, J. G., Lin, Q., Fu, Q., Ding, R., Zhang, D., and Wang, Z. (2014). Correlating events with time series for incident diagnosis. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining.*

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2021). *Introduction to Linear Regression Analysis.* Wiley.

Motahari-Nezhad, H. R., Saint-Paul, R., Casati, F., and Benatallah, B. (2011). Event correlation for process discovery from web service interaction logs. *The VLDB Journal*, **20**, 417-444.

Pal, S.K., Koley, A., Ranjan, P., and Kundu, D. (2024). Modeling time to failure using a temporal sequence of events, *Quality Engineering*, 1-20, https://doi.org/10.1080/08982112.2024.2441367

Rudin, C., Letham, B., Salleb-Aouissi, A., Kogan, E., and Madigan, D. (2011). Sequential event prediction with association rules. *JMLR: Workshop and Conference Proceedings* **19**, 615–634.

Wu, D., Ke, Y., Yu, J. X., Yu, P. S., and Chen, L. (2010). Detecting leaders from correlated time series. In *Kitagawa, H., Ishikawa, Y., Li, Q., Watanabe, C. (eds) Database Systems for Advanced Applications. DASFAA 2010. Lecture Notes in Computer Science, vol 5981. Springer, Berlin, Heidelberg.*

Zhu, Y., and Shasha, D. (2002). StatStream: Statistical monitoring of thousands of data. In *Proceedings 2002 VLDB Conference: 28th International Conference on Very Large Databases (VLDB). Elsevier.*