Predictive Models for Time-to-Fail with Multiple Causes

सिद्धिमूलं प्रबन्धनम्
भा. प्र. सं. इन्दौर
IIM INDORE

A Thesis

Submitted in Partial Fulfillment of the Requirements for the Executive Doctoral Programme in Management

Indian Institute of Management Indore

by

Sandip Kumar Pal

EDPM-2020

Operations Management & Quantitative Techniques Area

January 2026

Thesis Advisory Committee:

Prof. Pritam Ranjan
[Chair]

Prof. Arnab Koley	Prof. Sutirtha Bandyopadhyay	Prof. Debasis Kundu
[Member]	[Member]	[Member]

Abstract

This thesis addresses the critical need for accurate failure time prediction for complex systems, with two real-life applications in industrial manufacturing and intensive care units (ICUs). In the first application, we developed a predictive model for analysing operational logs data from Injection Molding (IM) machine to model event durations and failure frequencies using exponential and truncated Poisson distributions respectively. Motivated by the need for accurate and interpretable diagnosis of IM machines used in plastic bottle manufacturing, a novel parametric framework for time-to-failure prediction based on operational log data. Machine operation is characterized by two latent states, 'running without alert' and 'running with alert', whose durations follow independent exponential distributions, while the number of alert events per operating epoch is modeled using a shifted Poisson distribution. The key contributions include the joint modeling of state-dependent failure dynamics and event frequency, with the proposed approach achieving performance comparable to the Cox proportional hazards model while preserving strong interpretability for predictive maintenance. The maximum likelihood estimation, with and without covariates, enables a predictive framework to optimize operational parameters and minimize the machine downtime.

In the second real application, we model ICU stay durations of several patients in Boston, MA, USA under a competing risks framework using the Medical Information Mart for Intensive Care (MIMIC-III) dataset. Each outcome, such as mortality or discharge, is treated as a latent failure time, with the observed duration determined by the earliest event. Parametric distributions including Weibull and Burr Type XII are evaluated for modeling these times. The inclusion of patient covariates (vitals, demographics, and history) allows for personalized risk estimation.

We propose a flexible piecewise linear polynomial for modeling the cumulative hazard with breakpoints estimated via a Non-Homogeneous Poisson Process. This model improves adaptability to irregular risk patterns. Rigorous validation using BIC for model selection, and MSE and Kaplan-Meier comparisons for assessing model fit, supports the appropriateness of the proposed models. Results offer actionable insights for both proac-

tive industrial maintenance and clinical decision-making, demonstrating the value of competing risks models in real-world applications.

Acknowledgment

I would like to express my sincere gratitude to everyone who supported me throughout this thesis journey. This work would not have been possible without the guidance, encouragement, and constructive feedback of many individuals who contributed at different stages of my research.

I am deeply thankful to the members of my Thesis Advisory Committee (TAC) Prof. Pritam Ranjan (Chairperson), Prof. Arnab Koley, Prof. Sutirtha Bandyopadhyay, and Prof. Debasis Kundu for their valuable insights, thoughtful suggestions, and continuous academic guidance. Their feedback helped shape the direction of my research and significantly improved the clarity, rigor, and relevance of this work.

I would also like to sincerely thank the members of my Thesis Evaluation Committee (TEC) Prof. Debanjan Mitra, Prof. Shovan Chowdhury, and Prof. Sujay K. Mukhoti for their careful evaluation of my thesis and for providing insightful and constructive comments that strengthened the final outcome of this research.

My sincere appreciation goes to my professional mentors and leaders who provided steady encouragement and support throughout this journey, particularly Mr. Guy Sears, and Prof. Arnab Laha. Their confidence in my work and their perspectives from both industry and academia were immensely motivating.

Above all, I am profoundly grateful to my family. My deepest appreciation goes to my wife, Gayatri, and my daughter, Tanirika, whose unwavering support, patience, and encouragement have been my greatest strength throughout this journey. I consider this thesis not merely an individual achievement, but a shared accomplishment made possible by their constant belief in me.

Contents

References

[1] Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. *In Proceedings of the 1993 ACM SIGMOD international conference on Management of data (pp. 207-216).*

[2] Assistance, H. C. (2003). Summary of the hipaa privacy rule. *Office for Civil Rights.*

[3] Augustin, N. H., Sauleau, E. A., & Wood, S. N. (2012). On quantile quantile plots for generalized linear models. *Computational Statistics & Data Analysis, 56(8), 2404-2409.*

[4] Balakrishnan, N., Koutras, M. V., Milienos, F. S., & Pal, S. (2016). Piecewise linear approximations for cure rate models and associated inferential issues. *Methodology and Computing in Applied Probability, 18(4), 937–966.*

[5] Bandyopadhyay, S., Wolfson, J., Vock, D. M., Vazquez-Benitez, G., Adomavicius, G., Elidrisi, M., ... & O'Connor, P. J. (2015). Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular risk from electronic health record data. *Data Mining and Knowledge Discovery, 29, 1033–1069.*

[6] Berzborn, M., & Cramer, E. (2023). Inference for Type-I and Type-II hybrid censored minimal repair and record data. *Journal of Statistical Theory and Practice, 17(4), 53.*

[7] Beyersmann, J., Allignol, A., & Schumacher, M. (2011). *Competing risks and multi-state models with R. Springer Science & Business Media.*

[8] Clifford, G. D., Scott, D. J., & Villarroel, M. (2009). User guide and documentation for the MIMIC II database. *MIMIC-II database version, 2(95).*

[9] Cox, D. R. (1959). The analysis of exponentially distributed life-times with two types of failure. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 21(2), 411–421.*

[10] Cook, R. J., & Lawless, J. F. (2007). *The statistical analysis of recurrent events.*

[11] Crowder, M. J. (2001). Classical competing risks. *Chapman and Hall/CRC.*

[12] Ding, J., Liu, Y., Zhang, L., & 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.*

[13] Eswaran, M. (2014). Why gender matters in economics. Princeton University Press.

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

[15] Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association, 94(446), 496–509.*

[16] Fox, C., Moghaddam, B., & Stone, M. (2008). Determination of the Financial Impact of Machine Downtime on the Post Large Letters Sorting Process. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 222(5), 575–584.*

[17] Ganguly, A., Mitra, D., Balakrishnan, N., & Kundu, D. (2024). A flexible model based on piecewise linear approximation for the analysis of left truncated right censored data with covariates. *Statistics in Medicine, 43(2), 233–255.*

[18] Ganguly, A., Mitra, D., & Kundu, D. (2022). Order Restricted Inference for Adaptive Progressively Censored Competing Risks Data. *arXiv preprint arXiv:2205.03550.*

[19] Glazier, S. W. (2019). Sequential Survival Analysis with Deep Learning. *Brigham Young University, Master's Thesis.*

[20] Hastie, T., Tibshirani, R., & Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Predictions, 2008. *Publishing House of Electronics Industry, Beijing.*

[21] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering, 21(9), 1263-1284.*

[22] Huang, X., & An, A. (2002, December). Discovery of interesting association rules from livelink web log data. *In 2002 IEEE International Conference on Data Mining, 2002. Proceedings. (pp. 763-766). IEEE.*

[23] Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent data analysis, 6(5), 429-449.*

[24] Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing, 20(7), 1483-1510.*

[25] Jiang, X-R., & Gruenwald, Le. (2005). Microarray gene expression data association rules mining based on BSC-tree and FIS-tree. *Data & Knowledge Engineering, 53(1), 3–29.*

[26] Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2021). *Mimic-iv (version 1.0). PhysioNet, 101, e215-e220.*

[27] Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L. W. H., Feng, M., Ghassemi, M., ... & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific data, 3(1), 1-9.*

[28] Kalantan, Z. I., Binhimd, S. M., Salem, H. N., AL-Dayian, G. R., EL-Helbawy, A. A., & Elaal, M. K. A. (2024). Chen-Burr XII Model as a Competing Risks Model with Applications to Real-Life Data Sets. *Axioms, 13(8), 531.*

[29] Kalbfleisch, J. D., & Prentice, R. L. (2002). *The statistical analysis of failure time data. John Wiley & Sons.*

[30] Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association, 53(282), 457-481.*

[31] Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC medical research methodology, 18(1), 24.

[32] Kundu, D., Mitra, D., & Ganguly, A. (2017). Analysis of left truncated and right censored competing risks data. *Computational Statistics & Data Analysis, 108, 12-26.*

[33] Kundu, D. (2024). A Stationary Proportional Hazard Class Process and its Applications. *Methodology and Computing in Applied Probability, 26(4), 45.*

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

[35] Kolmogorov, A. N. (1933). On the empirical determination of a distribution law. English version of Kolmogorov (1933a). Mathematics and Its Applications (19..), Selected Works of AN Kolmogorov, 2.

[36] Lawless, J. F. (2011). Statistical models and methods for lifetime data. *John Wiley & Sons.*

[37] Lee, S. (2006). Identification of a competing risks model with unknown transformations of latent failure times. *Biometrika, 93(4), 996–1002.*

[38] Lou, J. G., Fu, Q., Wang, Y., & Li, J. (2010). Mining dependency in distributed systems through unstructured logs analysis. *ACM SIGOPS Operating Systems Review, 44(1), 91-96.*

[39] Luo, C., Lou, J. G., Lin, Q., Fu, Q., Ding, R., Zhang, D., & 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.*

[40] Li, Z., Zhou, S., Choubey, S., & Sievenpiper, C. (2007). Failure event prediction using the Cox proportional hazard model driven by frequent failure signatures. *IIE transactions, 39(3), 303-315.*

[41] Ling, M. H., and Balakrishnan, N., Model mis-specification analyses of Weibull and gamma models based on one-shot device test data , *IEEE Transactions on Reliability, vol. 66, 641–650, 2017.*

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

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

[44] Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E. J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic bulletin & review, 23(1), 103-123.*

[45] Pal, S. K., Koley, A., Ranjan, P., & Kundu, D. (2025). Modeling time to failure using a temporal sequence of events. *Quality Engineering, 37(3), 493-512.*

[46] Pal, A., Samanta, D., & Kundu, D. (2025). A semiparametric approach for simple step-stress model. *Test, 34(1), 91-124.*

[47] Prajapati, D., Ling, M. H., Shing Chan, P., & Kundu, D. (2023). Misspecification of copula for one-shot devices under constant stress accelerated life-tests. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 237(4), 725-740.*

[48] Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V. T., & Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics, 541–554.*

[49] Putter, H., Fiocco, M., & Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine, 26(11), 2389-2430. https://doi.org/10.1002/sim.2712.*

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

[51] Ramesh, R., Muralidhar, P., Madan Kumar, K. M. V., Banothu, S., Mahto, M. K., & Kumar, R. C. (2024). *A Longitudinal And Domain Wise Survey On MIMIC III Dataset. Frontiers in Health Informatics, 13(8).*

[52] Ross, S. M. (2014). Introduction to probability models. *Academic press.*

[53] Thomas, D., Rehg, J., & Tassey, G. (2020). Economics of Manufacturing Machinery Maintenance: A Survey and Analysis of U.S. Costs and Benefits. *National Institute of Standards and Technology (NIST) Report, U.S. Department of Commerce.*

[54] Villani, F., Civico, R., Pucci, S., Pizzimenti, L., Nappi, R., & De Martini, P. M. (2018). A database of the coseismic effects following the 30 October 2016 Norcia earthquake. *Scientific Data, 5(1), 1–11.*

[55] Wunsch, H., Angus, D. C., Harrison, D. A., Collange, O., Fowler, R., Hoste, E. A., ... & Rowan, K. M. (2008). Variation in critical care services across North America and Western Europe. *Critical care medicine, 36(10), 2787-e8.*

[56] Wu, X., Sun, T., Cai, Y., Zhai, T., Liu, Y., Gu, S., ... & Zhan, Q. (2023). Clinical characteristics and outcomes of immunocompromised patients with severe community-acquired pneumonia. *Frontiers in Public Health, 11, 1070581.*

[57] Wuryandari, T., & Wilandari, Y. (2023). Survival Analysis for Recurrent Event Data Using Counting Process Approach: Application to Diabetes. *Media Statistic, 16(1), 67–75.*

[58] Zhu, J. L., Yuan, S. Q., Huang, T., Zhang, L. M., Xu, X. M., Yin, H. Y., ... & Lyu, J. (2023). Influence of systolic blood pressure trajectory on in-hospital mortality in patients with sepsis. *BMC Infectious Diseases, 23(1), 90.*