Statistical Validation of Non-Homogeneous Poisson Processes via Time-Rescaling

Ryan Rodrigue
Santa Fe Christian Schools, Solana Beach, CA

Abstract

Non-Homogeneous Poisson Processes (NHPPs) model events with time-varying arrival rates, making them useful across fields such as seismology, meteorology, and infrastructure risk analysis. This paper tests for NHPP behavior by estimating time-dependent intensity functions, applying the Time-Rescaling Theorem to transform event times, and using the Kolmogorov-Smirnov test to evaluate goodness of fit. Using R software, we apply this framework to real-world datasets, including tornado occurrences, earthquake events, wildfire incidents, and oil pipeline accidents, providing a practical approach to assessing the suitability of NHPPs in diverse settings.

Keywords: non-homogeneous Poisson process, time-rescaling theorem, Kolmogorov-Smirnov test, point process modeling

1 Introduction

1.1 Background

Non-Homogeneous Poisson Processes (NHPPs) are a class of stochastic processes used to model the occurrence of random events over time, where the event rate is allowed to vary with time. Formally, an NHPP is defined on a time interval $[0, \infty)$ with a time-dependent **intensity function** $\lambda(t) \geq 0$ such that the number of events $N(t)$ occurring in $[0, t]$ satisfies the following conditions: (1) $N(0) = 0$, (2) the process has independent increments, and (3) for any $s,\, t \geq 0$,

$$P\big(N(s+t) - N(s) = n\big) = \frac{1}{n!}\left[\Lambda(s+t) - \Lambda(s)\right]^n e^{-\left[\Lambda(s+t)-\Lambda(s)\right]}, \quad n = 0, 1, 2, \ldots,$$

where $\Lambda(t) = \displaystyle\int_0^t \lambda(u)\, du$ is the integrated intensity rate function.

In this paper, we investigate whether certain real-world event sequences are consistent with an NHPP model. Since the intensity function $\lambda(t)$ is typically unknown, we estimate it from data using regression techniques. We then apply the Time-Rescaling Theorem to transform the observed event times, and use the Kolmogorov-Smirnov test to assess whether the transformed data follow the expected distribution under an NHPP. Rather than proving that a process is definitively Poissonian, we adopt a practical approach: we assume the NHPP model is valid and evaluate the degree to which the data conform to that assumption. This framework allows us to test for deviations from the model and identify potential structural inconsistencies in the underlying event-generating process.

1.2 Literature Review

Non-Homogeneous Poisson Processes (NHPPs) are widely applied in diverse scientific domains. In reliability engineering, NHPPs model time-varying failure rates in repairable systems and inform maintenance strategies [1,2,3]. In neuroscience, NHPPs characterize neural spike trains by capturing time-dependent firing rates in response to stimuli [4,5]. In telecommunications and network traffic modeling, NHPPs are used to describe non-stationary arrival rates of calls or data packets [6,7]. NHPPs are employed in seismology and climatology to model earthquake occurrences or rainfall events with non-constant intensity [8,9]. In queueing theory and operations research, NHPPs help analyze service systems with variable arrival rates, such as call centers and emergency rooms [10,11]. Finance and insurance apply NHPPs to model the timing of claims, defaults, or market shocks when risk evolves over time [12].

2 Theoretical Framework

2.1 Time-Rescaling Theorem

The Time-Rescaling Theorem is a fundamental result in point process theory that provides a method for testing whether a given event sequence follows a Non-Homogeneous Poisson Process (NHPP). Suppose we are given a sequence of event times $t_1, t_2, \ldots, t_n$ and we hypothesize that the underlying process follows a Non-Homogeneous Poisson Process with intensity function $\lambda(t)$.

The Time-Rescaling Theorem states that if the event times t_i are generated by an NHPP, then the transformed (rescaled) event times $\Lambda(t_i)$, defined by the integral of the intensity function:

$$\Lambda(t_i) = \int_0^{t_i} \lambda(u)\, du$$

will correspond to the event times of a homogeneous Poisson process with rate 1, meaning that the transformed inter-arrival times should be exponentially distributed with rate parameter $\lambda = 1$. Specifically, the rescaled inter-arrival times $\Delta\Lambda(t_i) = \Lambda(t_i) - \Lambda(t_{i-1})$ are i.i.d. with distribution:

$$P(\Delta\Lambda(t_i) \le t) = 1 - e^{-t}$$

for $t \ge 0$, where $\Delta\Lambda(t_i)$ is the transformed inter-arrival time [13].

The Time-Rescaling Theorem provides a way to test the hypothesis that the original process follows a Non-Homogeneous Poisson Process. If the transformed inter-arrival times follow the expected exponential distribution with rate 1, we can conclude that the event sequence indeed follows an NHPP with the given intensity function $\lambda(t)$. Otherwise, deviations from the exponential distribution suggest that the process is not an NHPP.

In practice, the Time-Rescaling Theorem is applied by computing the integral of $\lambda(t)$ for each event time and transforming the data accordingly. This transformation allows us to test the fit of the event times to an exponential distribution using statistical methods such as the Kolmogorov-Smirnov (KS) test.

2.2 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (KS) test is a non-parametric statistical test used to compare the empirical cumulative distribution function (ECDF) of a sample to a theoretical cumulative distribution function (CDF). In the context of testing whether a sequence of rescaled event times follows an exponential distribution with rate 1 (i.e., Exp(1)), the KS test is applied as follows.

Let $\Lambda_n(t)$ denote the ECDF of the rescaled event times $\Lambda(t_1), \Lambda(t_2), \ldots, \Lambda(t_n)$, and let $F(t)$ be the theoretical CDF of an exponential distribution with rate 1, given by:

$$F(t) = 1 - e^{-t}, \quad t \geq 0.$$

The KS test statistic is defined as the maximum absolute difference between the ECDF of the data and the theoretical CDF:

$$D = \sup_{t \geq 0} |\Lambda_n(t) - F(t)|$$

where sup denotes the supremum (maximum) over all values of t [14].

To evaluate the significance of the observed value of D, we compare it to the critical value from the KS distribution for a given significance level α. If the observed value of D is larger than the critical value, we reject the null hypothesis that the rescaled event times follow the theoretical exponential distribution. Otherwise, we fail to reject the null hypothesis, implying that the data do not significantly differ from the expected exponential distribution.

The p-value associated with the KS test is computed based on the distribution of D under the null hypothesis. The KS test provides a powerful way to test the goodness of fit of a sample to a theoretical distribution without making any assumptions about the underlying data other than continuity. A small p-value indicates a significant deviation from the theoretical distribution, while a large p-value suggests that the hypothesis of an exponential distribution cannot be rejected.

Mathematically, the p-value and $D_{critical}$ for the KS test can be approximated by the following:

$$P(D_n > D) \approx 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2 D^2 n}, \tag{1}$$

$$D_{critical} \approx \frac{1.36}{\sqrt{n}} \tag{2}$$

where n is the number of events and D is the observed value of the KS statistic. The first expression arises from the limiting distribution of the Kolmogorov-Smirnov test statistic under the null hypothesis and is highly accurate. The second expression provides an accurate approximation for large n when $\alpha = 0.05$, but is not exact and may be less reliable for small sample sizes [15]. Both (1) and (2) will be used in all examples when calculating p-values and $D_{critical}$ values.

3 Applications

3.1 Earthquakes in Southern California

Earthquake occurrences in Southern California are modeled by a Non-Homogeneous Poisson Process. Using earthquake data from the IRIS Seismic Network, the dataset was filtered to include only earthquakes that occurred between October and December of 2024, and removed any aftershocks within a 2-hour window. Additionally, one extreme outlier week was removed. After filtering, a total of 90 unique earthquakes were recorded within the three-month period.

In order to approximate the intensity function $\lambda(t)$ for the hypothesized Non-Homogeneous Poisson Process, we graph the observed daily intensity, grouped by week, and attempt to find a function of good fit. Seeing the strikingly quadratic trend in Figure 1, we use polynomial regression to find the intensity function below.

$$\lambda(t) = 0.17653 \; + \; 44.27205 \cdot \left(\frac{t}{1000}\right) \; - \; 436.15277 \cdot \left(\frac{t}{1000}\right)^2.$$

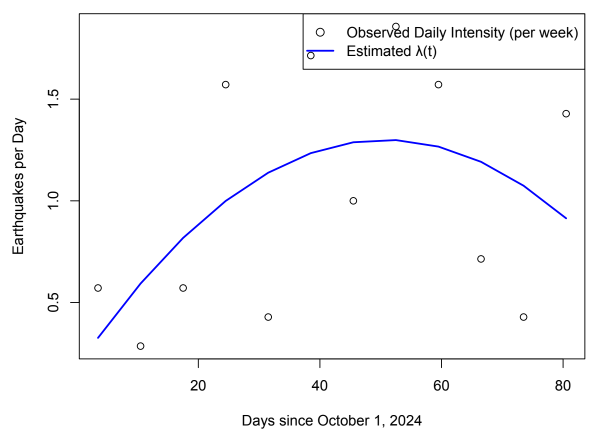

Figure 1: Earthquake Intensity Approximation $\lambda(t)$

Next, to test the validity of the Non-Homogeneous Poisson Process model with the intensity function described by $\lambda(t)$, we applied the Time-Rescaling Theorem. The Time-Rescaling Theorem converts the original event times into “rescaled” event times by integrating the intensity function $\lambda(t)$ from 0 to t, where t represents the time of each event. Taking the difference of adjacent rescaled event times transforms them into a new set of inter-arrival times, which should follow an Exp(1) distribution if the original data are from a NHPP.

To test whether the rescaled inter-arrival times follow an Exp(1) distribution, a Kolmogorov-Smirnov (KS) test between the rescaled inter-arrival times and an Exp(1) distribution was performed.

The KS test resulted in a test statistic of $D = 0.09629$ and a p-value of 0.3518. The test statistic D represents the maximum difference between the empirical cumulative distribution function (ECDF) of the rescaled inter-arrival times and the theoretical CDF of an Exp(1) distribution. Using $D_{\text{critical}} \approx \frac{1.36}{\sqrt{n}}$ (for $\alpha = 0.05$) where n is the number of observations, we find $D_{\text{critical}} \approx 0.14335$. Since $D < D_{\text{critical}}$ and the p-value is greater than the significance level of 0.05, we fail to reject the null hypothesis. This suggests that the rescaled inter-arrival times are not significantly different from an Exp(1) distribution, supporting the hypothesis that the earthquake occurrences in Southern California follow a Non-Homogeneous Poisson Process with the specified intensity function $\lambda(t)$.

This finding aligns with geophysical understanding—while earthquakes are generally unpredictable on an individual level, their broader occurrence patterns over time can be influenced by tectonic stress accumulation and fault activity, which evolve gradually. The ability to model these patterns using a Non-Homogeneous Poisson Process provides a valuable statistical foundation for earthquake risk analysis. Incorporating such models into seismic monitoring systems may help improve early warning frameworks, resource allocation for emergency preparedness, and long-term infrastructure planning in seismically active regions like Southern California.

3.2 Tornadoes in Tornado Alley

Tornado events in Tornado Alley are not modeled by a Non-Homogeneous Poisson Process. Using tornado data from the NOAA's National Weather Service Storm Prediction Center, the dataset was filtered to only include tornadoes in Tornado Alley (Texas, Oklahoma, Kansas, and Nebraska) in the year 2020. After removing duplicate tornado sightings, a total of 160 unique tornadoes occurred in this time period.

In order to approximate the intensity function $\lambda(t)$ for the hypothesized Non-Homogeneous Poisson Process, we graph the observed daily intensity, grouped by week and attempt to find a function of good fit. Seeing the trend in Figure 2, we tried many different functional forms such as Sigmoid, Log-normal, and Gamma. Ultimately, a Weibull function was the closest fit. Thus, the intensity function can be represented as:

$$\lambda(t) = a \left(\frac{t}{b}\right)^{c-1} \exp\left[-\left(\frac{t}{b}\right)^{c}\right].$$

Using non-linear squares optimization on a, b, and c, we find

$$\lambda(t) = 2.34216 \left(\frac{t}{99.29182}\right)^{0.31682} \exp\left[-\left(\frac{t}{99.29182}\right)^{1.31682}\right].$$

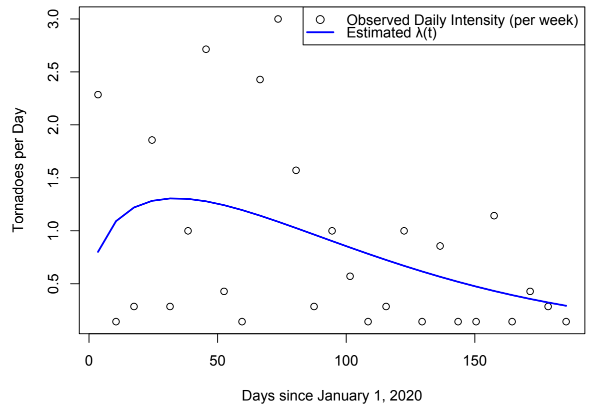

Figure 2: Tornado Intensity Approximation

Note that the x-axis of Figure 2 does not go up to 365 because the last tornado sighting of 2020 was on September 1. Next, to test the validity of the Non-Homogeneous Poisson Process model with the intensity function described by $\lambda(t)$, we applied the Time-Rescaling Theorem. The Time-Rescaling Theorem converts the original event times into "rescaled" event times by integrating the intensity function $\lambda(t)$ from 0 to t, where t represents the time of each event. Taking the difference of adjacent rescaled event times transforms them into a new set of inter-arrival times, which should follow an Exp(1) distribution if the original data are from a NHPP.

To test whether the rescaled inter-arrival times follow an Exp(1) distribution, a Kolmogorov-Smirnov (KS) test between the rescaled inter-arrival times and an Exp(1) distribution was performed.

The KS test resulted in a test statistic of $D = 0.63732$ and a p-value ≈ 0. The test statistic D represents the maximum difference between the empirical cumulative distribution function (ECDF) of the rescaled inter-arrival times and the theoretical CDF of an Exp(1) distribution. Using $D_{\text{critical}} \approx \frac{1.36}{\sqrt{n}}$ (for $\alpha = 0.05$) where n is the number of observations, we find $D_{\text{critical}} \approx 0.10752$. Since $D > D_{\text{critical}}$ and the p-value is less than the significance level of 0.05, we must reject the null hypothesis. This suggests that the rescaled inter-arrival times differ significantly from an Exp(1) distribution, and provides evidence that tornadoes in Tornado Alley do not follow a Non-Homogeneous Poisson distribution. While only the calculations for Tornado Alley were shown above, the same results are found for all 50 U.S. states.

The poor fit suggests there may be temporal clustering, such as multiple tornadoes during the same storm, which violates the assumption of conditional independence given $\lambda(t)$. Additionally, the intensity function derived from the Weibull function may not fully capture the complex, multi-scale dynamics driving tornado formation, such as atmospheric instability or seasonal weather patterns. This insight is crucial for improving tornado forecasting models by implying that alternative models (e.g., clustered point processes or Cox processes) may be more suitable to capture the complex structure of tornado events.

3.3 Oil Pipeline Accidents

Oil pipeline accidents in the United States are modeled by a Non-Homogeneous Poisson Process. Using records for oil pipeline spills and leaks from the Department of Transportation's Pipeline and Hazardous Materials Safety Administration, the dataset was filtered to include only events that occurred in the first half of the year 2016 (January to June). After filtering, a total of 210 unique accidents were recorded within the one-year period.

In order to approximate the intensity function $\lambda(t)$ for the hypothesized Non-Homogeneous Poisson Process, we graph the observed daily intensity, grouped by week, and do not notice a clear trend in Figure 3. We attempt to fit a constant intensity function by finding the mean number of events per day. Thus, the intensity function can be represented as:

$$\lambda = 1.11111.$$

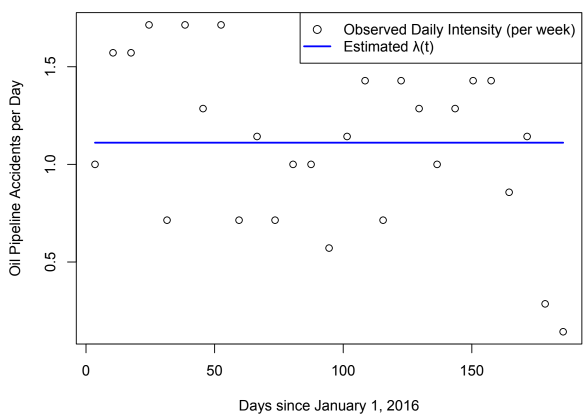

Figure 3: Oil Pipeline Accident Intensity Approximation

Next, to test the validity of the Non-Homogeneous Poisson Process model with the intensity function described by $\lambda(t)$, we applied the Time-Rescaling Theorem. The Time-Rescaling Theorem converts the original event times into "rescaled" event times by integrating the intensity function $\lambda(t)$ from 0 to t, where t represents the time of each event. Taking the difference of adjacent rescaled event times transforms them into a new set of inter-arrival times, which should follow an Exp(1) distribution if the original data are from a NHPP.

To test whether the rescaled inter-arrival times follow an Exp(1) distribution, a Kolmogorov-Smirnov (KS) test between the rescaled inter-arrival times and an Exp(1) distribution was performed.

The KS test resulted in a test statistic of $D = 0.07674$ and a p-value of 0.1685. The test statistic D represents the maximum difference between the empirical and theoretical CDFs. Using $D_{\text{critical}} \approx \frac{1.36}{\sqrt{n}}$ (for $\alpha = 0.05$) where $n = 210$, we find $D_{\text{critical}} \approx 0.0938$. Since $D < D_{\text{critical}}$ and the p-value is greater than the significance level of 0.05, we fail to reject the null hypothesis. This suggests that the rescaled inter-arrival times are not significantly different from an Exp(1) distribution, supporting the hypothesis that oil pipeline accidents follow a Non-Homogeneous Poisson Process with constant intensity $\lambda = 1.11111$.

Furthermore, because the intensity function λ is constant and independent of t, the process also satisfies the conditions of a Homogeneous Poisson Process (HPP), where:

$$P(N(t) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}.$$

While it may seem contradictory to label a point process as both Non-Homogeneous and Homogeneous, the definition of an NHPP allows $\lambda(t)$ to vary with time—it does not require it. Therefore, a HPP is simply a special case of an NHPP with constant intensity. This dual classification means that techniques developed for NHPPs in this paper also apply to HPPs, and in this case, modeling is greatly simplified due to the constant rate.

Oil pipelines are monitored under consistent operational and regulatory standards throughout the year, making accidents roughly equally likely on any given day. As such, the assumption of constant intensity is a reasonable approximation for this system in the real world. This simplification makes it significantly easier to forecast and allocate resources for oil spill response teams, allowing regulatory agencies to better prepare for and mitigate future pipeline accidents across the country.

3.4 Wildfires in California

Wildfire incidents in California are not modeled by a Non-Homogeneous Poisson Process. Using spatial wildfire data acquired from the reporting systems of federal, state, and local fire organizations, we filtered the dataset to include only wildfires that happened in California in 2018, and removed any duplicate sightings of the same fire. After filtering, a total of 7,589 unique wildfires were recorded within the one-year period.

In order to approximate the intensity function $\lambda(t)$ for the hypothesized Non-Homogeneous Poisson Process, we graph the observed daily intensity, grouped by week, and notice a trend very similar to a normal distribution in Figure 4. Thus, the intensity function can be represented as:

$$\lambda(t) = \alpha \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right).$$

Using non-linear squares optimization on σ, μ, and α, we find

$$\lambda(t) = 107.4295 \cdot \frac{1}{68.8687\sqrt{2\pi}} \exp\left(-\frac{(t-193.9674)^2}{2\cdot(69.8687)^2}\right).$$

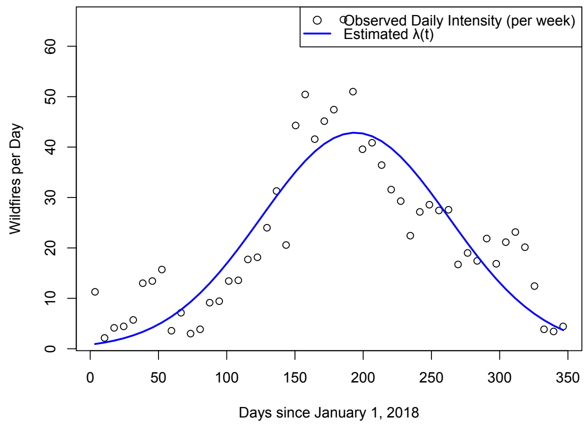

Figure 4: Wildfire Intensity Approximation

Next, to test the validity of the Non-Homogeneous Poisson Process model with the intensity function described by $\lambda(t)$, we applied the Time-Rescaling Theorem. The Time-Rescaling Theorem converts the original event times into "rescaled" event times by integrating the intensity function $\lambda(t)$ from 0 to t, where t represents the time of each event. Taking the difference of adjacent rescaled event times transforms them into a new set of inter-arrival times, which should follow an Exp(1) distribution if the original data are from a NHPP.

To test whether the rescaled inter-arrival times follow an Exp(1) distribution, a Kolmogorov-Smirnov (KS) test between the rescaled inter-arrival times and an Exp(1) distribution was performed.

The KS test resulted in a test statistic of $D = 0.2413$ and a p-value ≈ 0. The test statistic D represents the maximum difference between the empirical and theoretical CDFs. Using $D_{\text{critical}} \approx \frac{1.36}{\sqrt{n}}$ (for $\alpha = 0.05$) where $n = 7,589$, we find $D_{\text{critical}} \approx 0.0156$. Since $D > D_{\text{critical}}$ and the p-value is less than the significance level of 0.05, we can confidently reject the null hypothesis. While only the calculations for California were shown above, very similar results are found for all 50 U.S. states. This suggests that the rescaled inter-arrival times significantly differ from an Exp(1) distribution, providing evidence for the claim that wildfire incidents in California (or any other U.S. state) do not follow a Non-Homogeneous Poisson distribution.

This result is consistent with real-world dynamics—wildfires are highly seasonal and influenced by environmental factors like temperature, humidity, and wind, which fluctuate significantly over the course of a year. As a result, more sophisticated models that incorporate weather patterns, vegetation dryness, and human activity are likely needed for accurate forecasting. Understanding these variables is crucial for developing more adaptive fire management policies, improving early warning systems, and efficiently allocating firefighting resources throughout California and other U.S. states.

4 Future Work

Future research on non-homogeneous Poisson processes (NHPPs) could pursue several promising directions. Perhaps the most compelling (and time-saving) direction would be the development of a robust method for selecting the functional form of the intensity function $\lambda(t)$. While polynomial regression is effective when the underlying form is polynomial, it is often unclear whether $\lambda(t)$ is best modeled using a polynomial, exponential, logarithmic, trigonometric, or another functional family. So, creating a way to determine the underlying form of the intensity function would save a lot of guess-and-check work. Techniques such as nonparametric model selection, cross-validation across basis function classes, or adaptive model fitting using splines and kernel methods could offer optimal strategies for identifying the most appropriate functional form from the data alone.

Another research direction could be the extension of NHPPs to higher dimensions, where the intensity function varies not only over time but also across space. This spatial-temporal generalization would allow for modeling more complex stochastic behavior in dynamic systems. Using the tornado modeling from Section 3.2 as an example, we may create a more accurate intensity function $\lambda(t, l)$ that takes into account not just time but also latitude to improve results. Another promising approach to look into is the use of Bayesian methods to estimate the intensity function, allowing for formal incorporation of prior structural assumptions (e.g., smoothness or functional form) and providing a full distribution over possible intensity curves to quantify uncertainty in both the function shape and event rate estimation.

5 Supplemental Materials

To support and extend the findings in this paper, all relevant R codes and datasets are available in a public GitHub repository. This repository, created under the username ryrod724 and titled Statistical-Validation-of-Non-Homogeneous-Poisson-Processes-via-Time-Rescaling, provides readers with access to the full suite of resources used in this study.

Acknowledgments

I would like to express my gratitude to all those who supported and guided me throughout this journey. To my friends and family, thank you for your encouragement, patience, and belief in me from start to finish. I am truly appreciative of my mentor, Dr. Olga Korosteleva, Professor of Statistics at California State University, Long Beach, for her expert guidance, insightful feedback, and unwavering commitment. This project would not have been possible without you.

References

[1] Rigdon, S.E. and A.P. Basu, (2000). *Statistical Methods for the Reliability of Repairable Systems*. Wiley.

[2] Crow, L.H. (1974). Reliability analysis for complex, repairable systems. *Reliability and Biometry* (F. Proschan and R.J. Serfling, eds.) SIAM, Philadelphia, 379–410.

[3] Yanez, M., Joglar, F., and M. Modarres. (2002). Generalized renewal process for analysis of repairable systems with limited failure experience. *Reliability Engineering & System Safety*, 77(2): 167–180.

[4] Kass, R.E., Ventura, V., and E.N. Brown. (2005). Statistical issues in the analysis of neuronal data. *Journal of Neurophysiology*, 94(1): 8–25.

[5] Truccolo, W., Eden, U.T., Fellows, M.R., Donoghue, J.P., and E.N. Brown. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology*, 93(2): 1074–1089.

[6] Ross, S.M. (2023). *Introduction to Probability Models*, 13th ed. Academic Press.

[7] Gaver, D.P. and P.A.W. Lewis. (1980). The use of nonstationary Poisson processes in telephone traffic. *Bell Labs Technical Journal*, 59(10): 1991–2021.

[8] Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401): 9–27.

[9] Rodriguez-Iturbe, I. and A. Rinaldo. (1997). *Fractal River Basins: Chance and Self-Organization*. Cambridge University Press.

[10] Green, L.V., Kolesar, P.J., and W. Whitt. (2007). Coping with time-varying demand when setting staffing requirements for a service system. *Production and Operations Management*, 16(1): 13–39.

[11] Ibrahim, R. and W. Whitt. (2009). Real-time delay estimation based on delay history in many-server service systems with time-varying arrivals. *Production and Operations Management*, 18(2): 156–170.

[12] Dassios, A. and J. Jang. (2003). Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity. *Finance and Stochastics*, 7(1): 73–95.

[13] Brown, E.N., Barbieri, R., Ventura, V., Kass, R.E., and L.M. Frank. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*, 14(2), 325–346.

[14] Massey Jr., Frank J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253), 68–78.

[15] Marsaglia, G., Tsang, W.W., and J. Wang. (2003). Evaluating Kolmogorov's distribution. *Journal of Statistical Software*, 8(18), 1–4.