Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

THE EASE OF FITTING BUT FUTILITY OF TESTING A NONSTATIONARY POISSON PROCESSES FROM ONE SAMPLE PATH

Barry L. Nelson

Industrial Engineering & Management Sciences
Northwestern University
Evanston, IL 60208-3119, USA

Lawrence M. Leemis

Department of Mathematics
William & Mary
Williamsburg, VA 23187-8795, USA

ABSTRACT

The nonstationary Poisson process (NSPP) is a workhorse tool for modeling and simulating arrival processes with time-dependent rates. In many applications only a single sequence of arrival times are observed. While one sample path is sufficient for estimating the arrival rate or integrated rate function of the process—as we illustrate in this paper—we show that testing for Poissonness, in the general case, is futile. In other words, when only a single sequence of arrival data are observed then one can fit an NSPP to it, but the choice of "NSPP" can only be justified by an understanding of the underlying process physics, or a leap of faith, not by testing the data. This result suggests the need for sensitivity analysis when such a model is used to generate arrivals in a simulation.

1 INTRODUCTION

Consider observing one-at-a-time "arrivals" over some finite period of time $[0,T_e]$. Let $\mathbf{T} = \{T_1, T_2, \ldots, T_{N(T_e)}\}$ denote the arrival times, where $0 < T_1 < T_2 < \cdots < T_{N(T_e)} \le T_e$, $N(t) = \sup\{n \colon T_n \le t\}$ denotes the arrival-counting process, and $\Lambda(t) = \mathrm{E}[N(t)]$ denotes the expected number of arrivals by time t (the integrated rate function), which is unknown. We have reason to believe that this process is nonstationary in the sense that there exists no constant $\lambda > 0$ such that $\Lambda(t) = \lambda t$ for all $0 \le t \le T_e$. *For simulation practitioners who want to generate relevant arrivals in their simulation, two key questions are: (a) Can we estimate $\Lambda(t)$ from this data? (b) Can we test whether the data are consistent with a nonstationary Poisson process (NSPP)?*

There are many reasons to hope that both answers are "yes:" NSPPs lead to tractable mathematical models (e.g., queueing), and variate generation as input to a stochastic simulation is easy. The answer to (a) turns out to be positive. Although not universally appreciated, there are many ways to "fit" $\Lambda(t)$ or $\lambda(t) = d\Lambda(t)/dt$ to a single sample path $\mathbf{T}$, including nonparametric methods such as Leemis (1991) and Arkin and Leemis (2000); semi-parametric methods such as Morgan et al. (2019); and fully parametric methods such as Lee et al. (1991). See Section 3 below for examples. Unfortunately, the answer to (b) is negative. In Section 4 we establish the futility of *testing* the NSPP assumption from a single sample path unless $\Lambda(t)$ is known (Section 5). The implication is that employing an NSPP is possible, but must be justified by the arrival process physics, not by testing the data. We begin with some basic background in Section 2.

2 BACKGROUND

The NSPP has received considerable attention in simulation research, particularly with respect to fitting $\Lambda(t)$ or $\lambda(t)$ to data, and simulating arrivals given $\Lambda(t)$ or $\lambda(t)$. See Nelson (2013) for the basics, and Section 3 below for examples.

Loosely speaking any test of nonstationary Poissonness will be based on some property that NSPPs have but other nonstationary arrival-counting processes do not. There are four obvious candidates: If $N(t)$

is an NSPP, then ...

Total: $N(T_e)$ has a Poisson distribution.

Transformation: $\mathscr{T}_j = \Lambda(T_j), j = 1,2,\ldots$ are equal in distribution to arrival times of a stationary, rate-1 Poisson process.

Variation: $\text{Var}[N(t)]/\Lambda(t) = 1$ for all $t \geq 0$.

Splitting: If arrivals are independently and randomly classified as type 1 with probability $0 < p < 1$, and type 2 otherwise, then the corresponding arrival times $\{T_{1j}, j = 1,2,\ldots\}$ and $\{T_{2j}, j = 1,2,\ldots\}$ are equal in distribution to arrival times from independent NSPPs with expected number of arrivals by time t of $p\Lambda(t)$ and $(1-p)\Lambda(t)$, respectively.

To the best of our knowledge the only paper to propose a test for nonstationary Poissonness in the single-sample-path case is Brown et al. (2005), which was later refined by Kim and Whitt (2014b). These papers address the special case when $\Lambda(t)$ is piecewise linear ($\lambda(t)$ is piecewise constant) and the breakpoints are known; they exploit a version of the transformation property. Such strong assumptions are unlikely to ever be true in practice, and getting the breakpoints right makes a difference (Kim and Whitt 2014a). Here we are interested in the general case.

We begin by recalling some methods for fitting an NSPP to a single sample path **T**. These are not new, but are usually presented in the context of multiple identically distributed sample paths (e.g., 52 Mondays). We then investigate testing for Poissonness when $\Lambda(t)$ is completely unknown, a topic that we have not seen addressed beyond the piecewise-linear case.

3 ESTIMATION WITH A SINGLE SAMPLE PATH

There are many applications in stochastic modeling in which only a single realization is available. This is frequently the case in time series analysis. It is also common in modeling the failure times in a repairable system in reliability and in modeling customer arrival times in queueing. We consider a reliability application as an illustration. In this section the underlying arrival process is assumed to be an NSPP.

The U.S.S. Halfbeak was launched on February 19, 1946 by the Electric Boat Company of Groton, Connecticut. There are $N = 71$ event times (failure times, read row-wise, in hours) collected on the No. 3 main propulsion diesel engine displayed in Table 1 (Crowder et al. 1994). The observation period for this single sample path is assumed to end at $T_e = 25{,}600$ hours.

A line plot of the data is given in Figure 1. A cursory visual inspection of the pattern of failure times reveals a concentration of failures after about 19,000 hours; the engine appears to be dramatically deteriorating after this point in time. This inspection leads us to conclude that a nonstationary model is appropriate. Can we infer $\Lambda(t)$ from this data?

Table 1: U.S.S. Halfbeak engine failure times (hours).

1382	2990	4124	6827	7472	7567	8845	9450
9794	10848	11993	12300	15413	16497	17352	17632
18122	19067	19172	19299	19360	19686	19940	19944
20121	20132	20431	20525	21057	21061	21309	21310
21378	21391	21456	21461	21603	21658	21688	21750
21815	21820	21822	21888	21930	21943	21946	22181
22311	22634	22635	22669	22691	22846	22947	23149
23305	23491	23526	23774	23791	23822	24006	24286
25000	25010	25048	25268	25400	25500	25518	

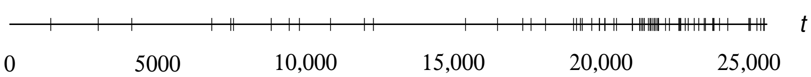

Figure 1: Unscheduled maintenance action times.

3.1 Nonparametric Estimators of $\Lambda(t)$

The standard nonparametric estimator of $\Lambda(t)$ is often known as the "step-function" estimator. For the single observed sample path of the target NSPP, let the step-function estimator $\bar{\Lambda}(t)$ denote the number of events observed in the time interval $(0,t]$, for $0 < t \le T_e$. The step-function estimator takes upward steps of height 1 at the event times from the single sample path.

Returning to the $N = 71$ U.S.S. Halfbeak unscheduled maintenance times, the nonparametric step-function estimator of $\Lambda(t)$ is shown in Figure 2. The steepness of $\bar{\Lambda}(t)$ beginning at about $t = 19{,}000$ hours is indicative of the deterioration after that time indicated by the data set.

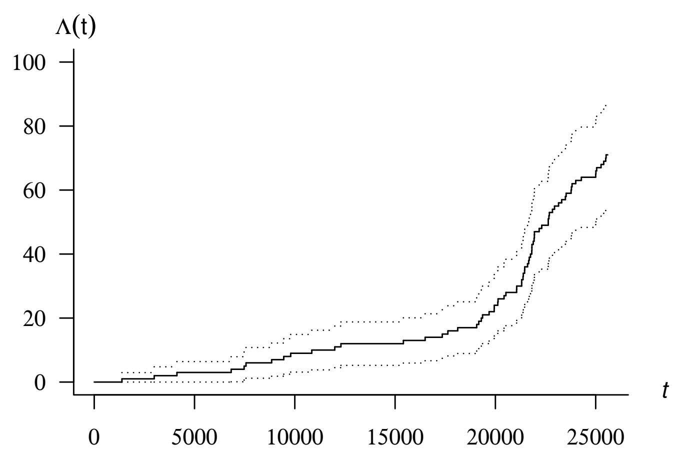

Figure 2: Step-function integrated rate function estimator $\bar{\Lambda}(t)$.

This nonparametric estimate of $\Lambda(t)$ corresponds to the intuitive interpretation of $\Lambda(t)$ as the expected number of unscheduled maintenance actions that have occurred by time t. The step-function estimator $\bar{\Lambda}(t)$ is not a consistent estimator of $\Lambda(t)$ when we are restricted to a single sample path, because extending the observation period T_e provides no additional information about any fixed time $t \le T_e$. If the arrival process is an NSPP, then we can construct the following approximate $100(1-\alpha)\%$ pointwise confidence interval for $\Lambda(t)$:

$$\bar{\Lambda}(t) - z_{\alpha/2}\sqrt{\bar{\Lambda}(t)} < \Lambda(t) < \bar{\Lambda}(t) + z_{\alpha/2}\sqrt{\bar{\Lambda}(t)},$$

which is included in Figure 2 as dotted lines. This confidence interval is obtained by observing that $\bar{\Lambda}(t)$ is an unbiased estimator of the mean and variance of $N(t)$, and when $\Lambda(t)$ is large the Poisson distribution is approximatley normal.

As expected, the width of the confidence bands increases with time. One can predict the number of failures that will occur by time 5000, for example, with greater precision than the number of failures that will occur by time 25,000. The fact that the line connecting the end points of the point estimator, $(0,0)$ and $(25{,}600, 71)$, is not contained within the confidence bands is evidence that there is a statistically significant departure from a constant-rate failure process; that is, the U.S.S. Halfbeak No. 3 main propulsion diesel engine is deteriorating over time.

If the purpose of estimating $\Lambda(t)$ is to simulate observations from the estimated NSPP, then using the step-function estimator means that only the observed event times can be generated in the simulation. Leemis (1991) developed a piecewise-linear estimator for $\Lambda(t)$ that overcomes the interpolation and extrapolation problems associated with the step-function estimator. For a single sample path of N arrivals, $\Lambda(t)$ is estimated by

$$\widehat{\Lambda}(t) = \frac{iN}{N+1} + \left[\frac{N(t-T_i)}{(N+1)(T_{i+1}-T_i)}\right], \; T_i < t \le T_{i+1}; \; i = 0, 1, \ldots, N$$

where $T_0 = 0$ and $T_{N+1} = T_e$.

Returning to the U.S.S. Halfbeak unscheduled maintenance action times, Figure 3 presents the piecewise-linear estimator, along with approximate 95% confidence bounds for $\Lambda(t)$ as dashed lines. The confidence bounds are computed using the formula associated with the step-function estimator.

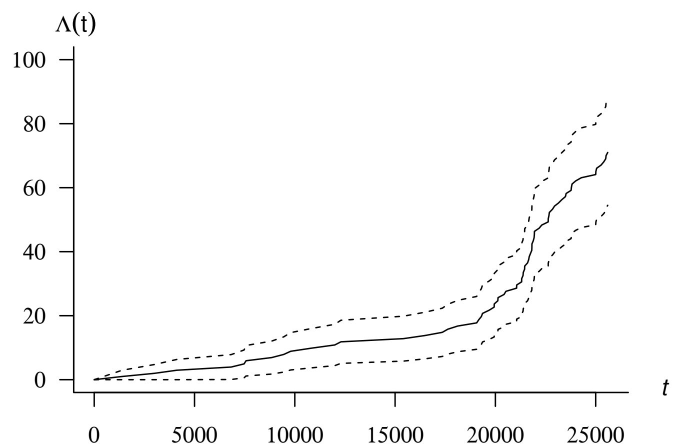

Figure 3: Piecewise linear integrated rate function estimator $\widehat{\Lambda}(t)$.

3.2 Parametric Estimators of $\Lambda(t)$

A second approach for fitting $\Lambda(t)$ or $\lambda(t)$ to a single sample path is to formulate a parametric form for $\Lambda(t)$ and estimate the parameters, say via maximum likelihood. Assume that previous experience indicates that the failure times for main propulsion diesel engines on vessels similar to the U.S.S. Halfbeak follow an integrated rate function of the form

$$\Lambda(t) = (\alpha t)^{\beta}, \ t > 0,$$

where α is a positive scale parameter and β is a positive shape parameter. This integrated rate function is known as a *power law process* and it has the same mathematical form as the integrated hazard function for the Weibull distribution.

The likelihood function for an NSPP model with rate function $\lambda(t)$ and integrated rate function $\Lambda(t)$ is

$$L = \left[\prod_{i=1}^{N} \lambda(T_i)\right] e^{-\Lambda(T_e)}$$

and the associated log-likelihood function is

$$\ln L = \left[\sum_{i=1}^{N} \ln \lambda(T_i)\right] - \Lambda(T_e).$$

In the specific case of a power law process, the log-likelihood function is

$$\ln L(\alpha, \beta) = N\beta \ln \alpha + N \ln \beta + (\beta - 1) \sum_{i=1}^{N} \ln T_i - (\alpha T_e)^{\beta}.$$

The partial derivatives of the log-likelihood function with respect to α and β are

$$\frac{\partial \ln L(\alpha, \beta)}{\partial \alpha} = \frac{N\beta}{\alpha} - \beta \alpha^{\beta - 1} T_e^{\beta}$$

and

$$\frac{\partial \ln L(\alpha, \beta)}{\partial \beta} = N \ln \alpha + \frac{N}{\beta} + \sum_{i=1}^{N} \ln T_i - (\alpha T_e)^{\beta} \ln(\alpha T_e),$$

which are the elements of the 2×1 score vector. Equating these partial derivatives to zero and solving for α and β yields the maximum likelihood estimators

$$\widehat{\beta} = \frac{N}{N \ln T_e - \sum_{i=1}^{N} \ln T_i} \text{ and } \widehat{\alpha} = \frac{N^{1/\widehat{\beta}}}{T_e}.$$

Notice that $\widehat{\Lambda}(T_e) = N$, which means that the fitted integrated rate function passes through $(0,0)$ and (T_e, N), which are the initial point and the terminal point of both the step-function and piecewise-linear nonparametric estimates of the integrated rate function. The second partial derivatives of the log-likelihood functions are

$$\frac{\partial^2 \ln L(\alpha, \beta)}{\partial \alpha^2} = -\frac{N\beta}{\alpha^2} - \beta(\beta - 1)\alpha^{\beta - 2} T_e^{\beta},$$

$$\frac{\partial^2 \ln L(\alpha, \beta)}{\partial \alpha \, \partial \beta} = \frac{N}{\alpha} - \alpha^{\beta - 1} T_e^{\beta} \left[1 + \beta \ln(\alpha T_e)\right],$$

$$\frac{\partial^2 \ln L(\alpha,\beta)}{\partial\beta^2} = -\frac{N}{\beta^2} - (\alpha T_e)^\beta \left[\ln(\alpha T_e)\right]^2.$$

The elements of the Fisher information matrix are the opposites of the expected values of these partial derivatives:

$$I(\alpha,\beta) = \begin{pmatrix} E\left[-\frac{\partial^2 \ln L(\alpha,\beta)}{\partial\alpha^2}\right] & E\left[-\frac{\partial^2 \ln L(\alpha,\beta)}{\partial\alpha\,\partial\beta}\right] \\ E\left[-\frac{\partial^2 \ln L(\alpha,\beta)}{\partial\beta\,\partial\alpha}\right] & E\left[-\frac{\partial^2 \ln L(\alpha,\beta)}{\partial\beta^2}\right] \end{pmatrix}.$$

This is the variance–covariance matrix of the score vector. This matrix can be estimated by the observed information matrix, which are these same partial derivatives evaluated at the maximum likelihood estimators:

$$O(\widehat{\alpha},\widehat{\beta}) = \begin{pmatrix} -\frac{\partial^2 \ln L(\alpha,\beta)}{\partial\alpha^2} & -\frac{\partial^2 \ln L(\alpha,\beta)}{\partial\alpha\,\partial\beta} \\ -\frac{\partial^2 \ln L(\alpha,\beta)}{\partial\beta\,\partial\alpha} & -\frac{\partial^2 \ln L(\alpha,\beta)}{\partial\beta^2} \end{pmatrix}_{\alpha=\widehat{\alpha},\beta=\widehat{\beta}}.$$

The inverse of the observed information matrix gives asymptotic estimates of the variance–covariance matrix of the maximum likelihood estimators. The diagonal elements of this matrix can be used to give asymptotically exact confidence intervals for α and β.

Returning to the U.S.S. Halfbeak data, the maximum likelihood estimates of α and β are $\widehat{\alpha} = 0.00019$ and $\widehat{\beta} = 2.7$. The step-function integrated rate function estimator and the fitted power law integrated rate function estimator are plotted in Figure 4. Both estimators end at the point $(T_e, N) = (25{,}600, 71)$. To three digits, the observed information matrix is

$$O(\widehat{\alpha},\widehat{\beta}) = \begin{pmatrix} 15400000000 & 1630000 \\ 1630000 & 182 \end{pmatrix}.$$

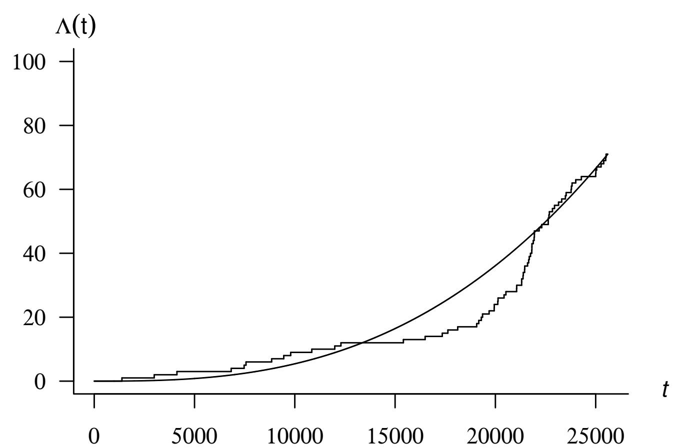

Figure 4: Step-function and fitted power law integrated rate functions.

Ignoring the off-diagonal elements of the inverse of the observed information matrix, 95% confidence intervals for these parameters are $0.00012 < \alpha < 0.00025$ and $2.1 < \beta < 3.4$. The fact that the confidence interval for β does not include $\beta = 1$ is statistical evidence that the rate function is increasing, which means that the engine is deteriorating. This is consistent with the conclusion drawn from Figure 1.

More flexible parametric models of $\lambda(t)$ have been proposed and maximum likelihood estimators obtained, most prominently the class of exponential trigonometric polynomial rate functions; see Lee et al.

(1991) and many follow-on papers. The parameters of many of these may be fit to only a single sample path.

The statistical analysis throughout this section assumed an NSPP created the observed sequence of arrivals. We next consider testing that assumption.

4 TESTING WITH A SINGLE SAMPLE PATH AND UNKNOWN RATE

Since we assume $\Lambda(t)$ is not known, direct application of the transformation property is not possible without estimating $\Lambda(t)$, but clearly there is problem if we fit $\Lambda(t)$ to $\mathbf{T}$ and then apply the fitted $\widehat{\Lambda}(t)$ to the same data $\mathbf{T}$ to test it. Further, as we have only one sample path, there is no direct way to assess the distribution of the total number of arrivals, $N(T_e)$, or to estimate $\mathrm{Var}[N(t)]/\Lambda(t)$, approaches that have been exploited when there are multiple sample paths (Gerhardt and Nelson 2009; Ross 2014). What seems to remain is splitting; we analyze the most likely approaches to exploit it in the two subsections that follow.
Confession. The analysis that follows occurred *after* the authors tried all of the ideas described below and discovered empirically that either actual NSPPs were consistently rejected as being Poisson, or non-NSPPs were consistently accepted as being Poisson. In hindsight the issues below should have been obvious.

4.1 Splitting to Facilitate Transformation

To avoid using the same data to fit $\Lambda(t)$ and then test, suppose that we randomly classify the observed arrival times in $\mathbf{T}$ as type 1 or 2 with probability p for type 1. We then use the type 2 arrivals to estimate the integrated rate function, and employ the type 1 arrivals to conduct a test of the Poisson assumption using the estimated integrated rate function.

Let N_1 and N_2 be the total number of type 1 and 2 arrivals, respectively, so that $N_1+N_2=N(T_e)$. If $N(t)$ is an NSPP, then the splitting property implies the respective integrated rate functions are $\Lambda_1(t)=p\Lambda(t)$ and $\Lambda_2(t)=(1-p)\Lambda(t)$. Thus, an unbiased step-function estimator of $\Lambda_1(t)$ is

$$\bar{\Lambda}_1(t) \equiv \frac{p}{1-p}\bar{\Lambda}_2(t) = \frac{p}{1-p}\sum_{j=1}^{N_2} I(T_{2j}\le t). \tag{1}$$

We can then apply this estimated integrated rate function to transformation the type 1 arrival times and create the data set

$$\mathscr{T}_j = \frac{p}{1-p}\bar{\Lambda}_2(T_{1j}),\ j=1,2,\ldots,N_1$$

on which to apply a test. We analyze the impact of estimating $\Lambda_1(t)$ in this way below.

Suppose $N(T_e)\gg j$. Then $\bar{\Lambda}_2(T_{1j})$ is the number of type 2 arrivals before the jth type 1 arrival; thus, based on our construction, $\bar{\Lambda}_2(T_{1j})\sim$ negative binomial$(j,1-p)$. *This result does not depend upon the original process being NSPP; it is an artifact of after-the-fact splitting.*

From this insight it is easy to show that

$$\frac{\mathrm{Var}[\mathscr{T}_j]}{\mathrm{E}[\mathscr{T}_j]} = \frac{1}{1-p} > 1.$$

Thus, even if the arrival process was NSPP, the transformed arrivals using this *estimated* $\Lambda_1(t)$ are over-dispersed, relative to a rate-1 exponential, by an amount that depends on our chosen p and not the distribution of the underlying arrival process.

Does a smoother integrated-rate function estimator than the step function solve the problem? It does not. Suppose $\Lambda(t)=\Lambda(t|\theta)$, which is known up to the value of a parameter θ with unknown true value θ_0, and is differentiable with respect to θ at θ_0. For simplicity of presentation suppose $\theta\in\Re$, although the analysis that follows does not depend on it being one-dimensional.

We use $\{T_{21}, T_{22}, \ldots, T_{2N_2}\}$ to estimate θ via $\widehat{\theta}$, and estimate $\Lambda_1(t)$ by

$$\widehat{\Lambda}_1(t) = p\Lambda(t|\widehat{\theta}) \approx p\left\{\Lambda(t|\theta_0) + (\widehat{\theta} - \theta_0)\frac{d\Lambda(t|\theta_0)}{d\theta}\right\}. \tag{2}$$

To be optimistic, suppose that the first-order Taylor approximation in (2) is exact, and that the parameter estimator $\widehat{\theta}$ is unbiased. We then apply the transformation to get $\widehat{\mathscr{T}_j} = \widehat{\Lambda}_1(T_{1j})$. Using the fact that $\widehat{\theta}$ is independent of $\{T_{1j}\}$ it is easy to show that

$$\mathrm{E}(\widehat{\mathscr{T}_j}) = \mathrm{E}[\Lambda_1(T_{1j})] \tag{3}$$

$$\mathrm{Var}(\widehat{\mathscr{T}_j}) = \mathrm{Var}[\Lambda_1(T_{1j})] + p^2\mathrm{Var}(\widehat{\theta})\mathrm{E}\left[\left(\frac{d\Lambda(T_{1j}|\theta_0)}{d\theta}\right)^2\right] > \mathrm{Var}[\Lambda_1(T_{1j})]. \tag{4}$$

Again the transformed process is over-dispersed and any test, even applied to Poisson data, should fail.

4.2 Splitting to Facilitate Testing Variability

Next consider testing variability. Randomly split the arrivals into $k \geq 2$ types, each with equal probability $1/k$. Let $N_i(t)$ and T_{ij} be the arrival-counting process and arrival times, respectively, associated with type i arrivals, and let $\mathbf{N}(t) = (N_1(t), N_2(t), \ldots, N_k(t))$. The splitting property suggests that from this data we can estimate $\mathrm{Var}[N_1(t)]/\Lambda_1(t)$ and compare it to 1.

Notice that $\mathbf{N}(T_j) \sim \mathrm{multinomial}(j, (1/k, 1/k, \ldots, 1/k))$, for $j = 1, 2, \ldots, N(T_e)$ as a consequence of after-the-fact splitting, whether the arrival process is NSPP or not. The natural estimator of the mean number of arrivals by time T_j is

$$\bar{\Lambda}_1(T_j) = \frac{1}{k}\sum_{i=1}^{k} N_i(T_j) = \frac{j}{k}.$$

Similarly, an estimator of the variance of the number of arrivals by time T_j is

$$S_1^2(T_j) = \frac{1}{k-1}\sum_{i=1}^{k}(N_i(T_j) - \bar{\Lambda}_1(T_j))^2.$$

Using properties of the multinomial distribbution it is easy to show that $\mathrm{E}[S_1^2(T_j)] = j/k$. Thus, the statistic $S_1^2(T_j)/\bar{\Lambda}_1(T_j)$ has expectation

$$\mathrm{E}\left[\frac{S_1^2(T_j)}{\bar{\Lambda}_1(T_j)}\right] = 1$$

whether or not the underlying process is an NSPP. Therefore, we obtain no traction for assessing variability, and in fact any arrival process will appear to be an NSPP based on this test.

Thus, with an unknown rate, even if known up to a parameter to be estimated, transformation cannot be exploited for testing. However, we can test if $\mathbf{T}$ is the result of an NSPP with a *given* rate, as illustrated in the next section.

5 TESTING WITH A SINGLE SAMPLE PATH AND GIVEN RATE

It is possible to conduct a test to assess the Poisson assumption when a parametric model with hypothesized parameters has been established from previous modeling. Assume again that previous testing indicates that the failure times for main propulsion diesel engines on vessels similar to the U.S.S. Halfbeak follow an integrated rate function of the form

$$\Lambda(t) = (\alpha t)^{\beta},\ t > 0,$$

where α is a positive scale parameter and β is a positive shape parameter. Assume that previous data sets concerning the sequence of failure times associated with diesel engines aboard similar vessels to the U.S.S. Halfbeak have revealed that a power law process with parameters $\alpha = 0.00011$ and $\beta = 4.2$ provide a reasonable fit to the data. Our goal is to assess whether the NSPP model with the hypothesized parameters in this power law model is appropriate for modeling the U.S.S. Halfbeak failure times.

The usual approach for testing the Poisson assumption associated with an NSPP is to transform the data values from the single sample path of the NSPP denoted by $T_1, T_2, \ldots, T_N$ to $\mathscr{T}_1 = \Lambda(T_1)$, $\mathscr{T}_2 = \Lambda(T_2), \ldots,$ $\mathscr{T}_N = \Lambda(T_N)$, which will constitute observations from a rate-1 stationary Poisson process if the hypothesized model is correct. The one-sample Kolmogorov–Smirnov test with all parameters known can be applied to the inter-event times of the unit-rate stationary Poisson process. The null hypothesis in terms of the failure times in the NSPP is that an NSPP with a hypothesized power law process governs the point process. The null hypothesis in terms of the transformed failure times is that the times between transformed failure times is unit exponential.

For the $n = 71$ U.S.S. Halfbeak unscheduled maintenance times, Figure 5 shows the nonparametric step-function estimator and the hypothesized integrated rate function for the power law process with $\alpha = 0.00011$ and $\beta = 4.2$. Figure 6 shows the empirical cumulative distribution function associated with the transformed failure times and the cumulative distribution function for the unit exponential distribution. The Kolmogorov–Smirnov test statistic is $D_{71} = 0.142$, which is the maximum vertical difference between the empirical cumulative distribution function and the cumulative distribution function for the unit exponential distribution. The p-value associated with this goodness-of-fit test using the `ks.test` function in R is $p = 0.105$. Using a threshold of $\alpha = 0.05$, we fail to reject the null hypothesis that the NSPP model governs the point process.

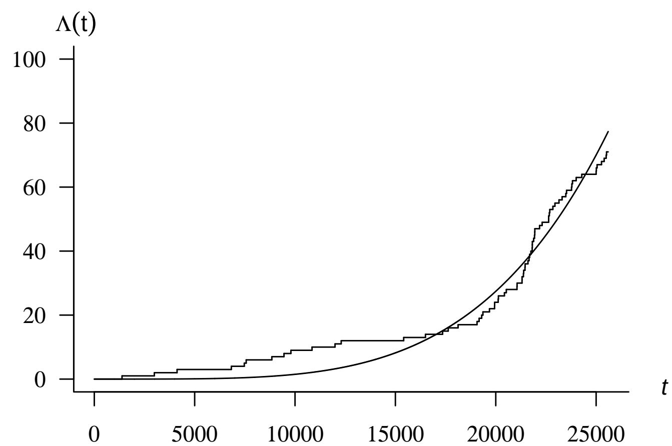

Figure 5: Step-function and power law integrated rate functions.

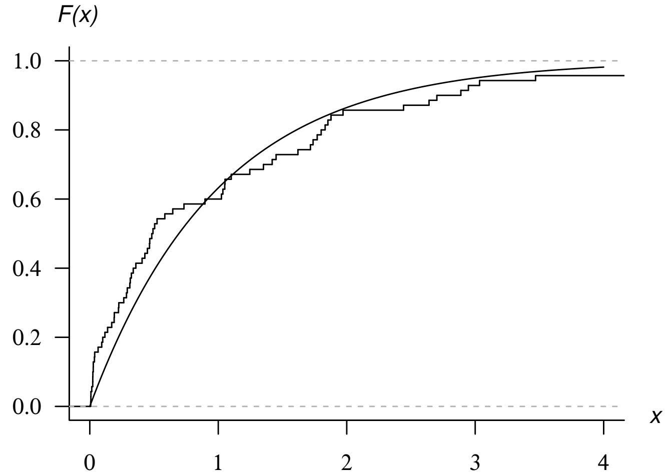

Figure 6: Kolmogorov–Smirnov test geometry.

6 CONCLUSIONS

A single sample path of a nonstationary arrival process will (typically) exhibit clusters of arrivals, as in Figure 1. However, clusters may result from $\Lambda(t)$ being steep, interarrival times being variable, or a combination of both; even stationary Poisson processes exhibit clusters. Thus, it makes intuitive sense that it is difficult to tease out Poisson and non-Poisson effects from such data. This paper shows that what might be considered obvious candidates to do the job do not work. That said, if process physics support the use of an NSPP model, then there are good methods for fitting the data and therefore providing synthetic arrivals in a stochastic simulation.

ACKNOWLEDGEMENTS

Nelson's work is partially supported by National Science Foundation Grant Number DMS-1854562. The authors thank Shane Henderson for comments on an early draft.

REFERENCES

Arkin, B. L., and L. M. Leemis. 2000. "Nonparametric Estimation of the Cumulative Intensity Function for a Nonhomogeneous Poisson Process from Overlapping Realizations". *Management Science* 46(7):989–998.

Brown, L., N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao. 2005. "Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective". *Journal of the American Statistical Association* 100(469):36–50.

Crowder, M. J., A. Kimber, T. Sweeting, and R. Smith. 1994. *Statistical Analysis of Reliability Data*. New York: CRC Press.

Gerhardt, I., and B. L. Nelson. 2009. "Transforming Renewal Processes for Simulation of Nonstationary Arrival Processes". *INFORMS Journal on Computing* 21(4):630–640.

Kim, S.-H., and W. Whitt. 2014a. "Are Call Center and Hospital Arrivals Well Modeled by Nonhomogeneous Poisson Processes?". *Manufacturing & Service Operations Management* 16(3):464–480.

Kim, S.-H., and W. Whitt. 2014b. "Choosing Arrival Process Models for Service Systems: Tests of a Nonhomogeneous Poisson Process". *Naval Research Logistics* 61(1):66–90.

Lee, S., J. R. Wilson, and M. M. Crawford. 1991. "Modeling and Simulation of a Nonhomogeneous Poisson Process Having Cyclic Behavior". *Communications in Statistics–Simulation and Computation* 20(2-3):777–809.

Leemis, L. M. 1991. "Nonparametric Estimation of the Cumulative Intensity Function for a Nonhomogeneous Poisson Process". *Management Science* 37(7):886–900.

Morgan, L. E., B. L. Nelson, A. C. Titman, and D. J. Worthington. 2019. "A Spline-based Method for Modelling and Generating a Nonhomogeneous Poisson Process". In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 356–367. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Nelson, B. L. 2013. *Foundations and Methods of Stochastic Simulation: A First Course*. New York: Springer Science & Business Media.

Ross, S. M. 2014. *Introduction to Probability Models*. 11th ed. New York: Academic Press.

AUTHOR BIOGRAPHIES

BARRY L. NELSON is the Walter P. Murphy Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University. He is a Fellow of INFORMS and IIE. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems, and he is the author of *Foundations and Methods of Stochastic Simulation: A First Course*, from Springer. His e-mail address and website are nelsonb@northwestern.edu and http://www.iems.northwestern.edu/~nelsonb/, respectively.

LAWRENCE M. LEEMIS is a Professor in the Department of Mathematics at William & Mary. He has formerly held faculty positions at The University of Oklahoma and Baylor University. He holds a B.S. in mathematics (1978), M.S. in mathematics (1980), and Ph.D. in industrial engineering (1984) from Purdue University. His email address and website are leemis@math.wm.edu and http://www.math.wm.edu/~leemis/, respectively.