Anytime-Valid A/B Testing of Counting Processes

Michael Lindon
Netflix

Nathan Kallus
Netflix & Cornell University

Abstract

Motivated by monitoring the arrival of incoming adverse events such as customer support calls or crash events from users exposed to an experimental product change, we consider sequential hypothesis testing of continuous-time counting processes. Specifically, we provide a multivariate confidence process on the cumulative rates $(\Lambda_t^A, \Lambda_t^B)$ giving an anytime-valid coverage guarantee $\mathbb{P}[(\Lambda_t^A, \Lambda_t^B) \in C_t^\alpha \, \forall t > 0] \geq 1 - \alpha$. This provides simultaneous confidence process on Λ_t^A, Λ_t^B and their difference $\Lambda_t^B - \Lambda_t^A$, allowing each arm of the experiment and the difference between them to be safely monitored throughout the experiment. We extend our results by constructing a closed-form e-process for testing the equality of rates with a time-uniform Type-I error guarantee at a nominal α. We characterize the asymptotic growth rate of the proposed e-process under the alternative and show that it has power 1 when the average rates of the two process differ in the limit.

1 INTRODUCTION

Counting processes are frequently encountered in industry applications that monitor the random arrival times of events. For example, monitoring the arrival times of customer calls in a customer support center (Brown et al., 2005) or the epochs of failure times in software reliability testing (Kuo and Yang, 1996). In software delivery, it is critical to assess how a new software update or rollout could impact the volume of observability events, such as crashes or errors, to prevent bugs from reaching production, which is often performed via a randomized controlled A/B test (Lindon et al., 2022; Ham et al., 2022). Typically, the total number of events, such as successful actions or harmful exceptions, observed among treatment units is compared to the total number observed among control units. It is desirable to monitor such experiments in real-time so that harmful treatments are detected and corrected as early as possible, as opposed to analyzing the experiment after a pre-specified amount of time has elapsed, which may either be too short to detect issues, or too long so that the damage has already been done.

In this paper, we study anytime-valid inference on counting processes, with guarantees that hold uniformly over (continuous) time. This allows continuous monitoring of event streams and sequential testing hypotheses concerning differences in the cumulative arrival rates so that action can be taken immediately when a difference is detected. The ability to test hypotheses repeatedly over time, while maintaining strict Type I error guarantees necessitates careful sequential analysis.

In Section 3, we review existing results for constructing a confidence process C_t^α (Darling and Robbins, 1967; Howard et al., 2021) on $\Lambda_t := \int_0^t \lambda(s|\mathcal{H}_s)ds$, which captures the notion of a cumulative arrival rate, that is guaranteed to cover the estimand for all $t \geq 0$ with probability at least $1 - \alpha$. We provide a complementary construction of these results by considering a mixture of point process likelihoods over proportional hazards alternatives. We further show that C_t^α is powered to detect deviations in average arrival, namely that if $\Lambda_t/t \to \bar{\lambda}$ then any Λ_{1t} with $\Lambda_{1t}/t \to \bar{\lambda}_1 \neq \bar{\lambda}$ is eventually excluded with probability one.

In section 4.2, we consider the case of observing two independent processes with conditional intensity functions $\lambda^A(t|\mathcal{H}_t^A), \lambda^B(t|\mathcal{H}_t^B)$. We provide a multivariate confidence process containing $(\Lambda_t^A, \Lambda_t^B)$ jointly at all times t with probability at least $1 - \alpha$. This consequently provides *simultaneous* confidence processes on the average rates $(1/t)\Lambda_t^B$, $(1/t)\Lambda_t^A$ and the difference $(1/t)(\Lambda_t^B - \Lambda_t^A)$, enabling monitoring applications to trigger logic, such as sending alerts, not just when the difference between two counting processes is large, but

Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS) 2025, Mai Khao, Thailand. PMLR: Volume 258.

also if the rate of either process exceeds some absolute threshold.

In section 5 we consider testing equality of inhomogeneous Poisson processes, that is, the null hypothesis $H_0 : \lambda^A(t) = \lambda^B(t)$ for all $t > 0$, for which we provide a continuous-time *e-process*. That is, we construct a nonnegative process E_t such that, under the null hypothesis H_0, we have $\mathbb{E}_{H_0}[E(\tau)] = 1$ for any, possibly data-dependant, random stopping time τ (Grünwald et al., 2021; Ramdas et al., 2023). Rejecting the null at time $\tau = \inf\{t > 0 : E_t > \alpha^{-1}\}$ then has the time-uniform Type-I error guarantee $\mathbb{P}_{H_0}[\tau < \infty] \leq \alpha$. Our e-processes are constructed directly from our joint confidence processes on $(\Lambda_t^A, \Lambda_t^B)$, providing a unified approach to inference. We characterize the asymptotic growth rate of the proposed e-process when $\Lambda_t^A/t \to \bar{\lambda}^A$, $\Lambda_t^B/t \to \bar{\lambda}^B$, and we show the rate depends on a particular divergence between $\bar{\lambda}^A$ and $\bar{\lambda}^B$. As a consequence, this establishes that our test is powered to detect violations of H_0 when $\bar{\lambda}^A \neq \bar{\lambda}^B$, that is, $\mathbb{P}(\tau < \infty) = 1$ in such cases. We further compare the e-process with two alternative procedures in section 5.1: one based on the Bernoulli draw of the next event coming from A or B and one based on a normal approximation of the difference in counting processes.

We demonstrate our methodology in section 6 with an application regarding detecting increases in customer call center volume among two treatment arms of the A/B test, where the desire is to abandon the experiment as quickly as possible if any increase is detected.

1.1 Related Literature

Many classical statistical tests fail to provide *time-uniform* Type-I error and coverage guarantees, providing guarantees at a fixed sample size only, because the experiments that motivated their development, such as agriculture, yielded results simultaneously (Anscombe, 1954). In modern experiments, however, outcomes are often observed sequentially. Repeated hypothesis testing with such methods on an accumulating set of data causes the cumulative Type-I error to grow (Armitage et al., 1969). To facilitate continuous monitoring of hypotheses with the ability to perform early stopping, we need sequential analysis.

Early works date back to Wald (1945) and the sequential probability ratio test. Darling and Robbins (1967); Robbins (1970) introduced confidence sequences for continuous monitoring of estimands. Howard et al. (2021) provides nonparametric nonasymptotic confidence sequences under tail conditions. Confidence sequences for quantiles are provided by Howard and Ramdas (2022). Confidence sequences to heavy tailed and robust mean estimation are available in Wang and Ramdas (2023a,b). Lindon and Malek (2022) provide confidence sequences for multinomial parameters, while Waudby-Smith and Ramdas (2020) provide confidence sequences in sampling without replacement settings. Lindon et al. (2024) provide confidence sequences for linear model coefficients and average treatment effects in randomized experiments. Time-uniform versions of the central limit theorem via strong approximations yielding asymptotic confidence sequences can be found in Waudby-Smith et al. (2021); Bibaut et al. (2022).

Confidence sequences are closely related to e-processes (Grünwald et al., 2021; Ramdas et al., 2023). The ability to stop early has been valuable in medical survival analysis (ter Schure et al., 2020) and contingency table testing (Turner et al., 2021; Turner and Grünwald, 2023), but also in online A/B testing (Johari et al., 2021, 2017). Lindon et al. (2022) use confidence sequences to monitor quantiles of performance metrics in software A/B experiments. Ham et al. (2022) construct design-based confidence sequences on sample causal estimands to de-risk experimental product changes.

2 COUNTING AND POINT PROCESSES

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_t\}_{t\geq 0}$ be a filtration satisfying the usual conditions of right continuity and completeness. A counting process N_t is an $\mathbb{N}_0$-valued stochastic process that is adapted to $\mathcal{F}_t$, is almost surely nondecreasing i.e. $N_s \leq N_t \,\forall s \leq t$, and has locally finite increments i.e. $N_b - N_a < \infty$ almost surely for any bounded interval $[a, b]$. In what follows we specify the initial condition $N(0) = 0$. A key quantity associated with N_t is the compensator Λ_t. This is the predictable, right-continuous, increasing process Λ_t such that $N_t - \Lambda_t$ is a martingale w.r.t. $\mathcal{F}_t$. If Λ_t is absolutely continuous w.r.t. Lebesgue measure, we may write $\Lambda_t = \int_0^t \lambda(s|\mathcal{H}_s)ds$ where $\mathcal{H}_t = \sigma\{N_s : 0 \leq s \leq t\}$ is the canonical filtration generated by N_t and $\lambda(s|\mathcal{H}_s)$ is the conditional intensity. Note in general that Λ_t and $\lambda(t|\mathcal{H}_t)$ are random process as they depend on the history of N_t, while $\mathbb{E}[N_t] = \mathbb{E}[\Lambda_t]$ follows from the martingale property. In a time-inhomogeneous Poisson counting process $\lambda(t|\mathcal{H}_t) = \lambda(t)$, that is, the intensity function is purely deterministic as the process is memoryless. In survival analysis, on the other hand, consider n independent and identically distributed random failure times $\{t_i\}_{i=1}^n$ each with the hazard function $h(\cdot)$. Viewed as counting process the *conditional* intensity function is $\lambda(t|\mathcal{H}_t) = A(t|\mathcal{H}_t)h(t)$, where $A(t|\mathcal{H}_t) = n - \sum_i 1[t_i < t]$ denotes the number "at risk" at time t. The compensator

$\Lambda_t = \sum_{i=1}^{n} \int_0^t 1[t_i \geq s]h(s)ds = \sum_{i=1}^{n} H(\min(t_i, t)) = -\sum_{i=1}^{n} \log S(\min(t_i, t))$ where H and S are the cumulative hazard and survival functions respectively. Lastly, $\mathbb{E}[N_t] = \mathbb{E}[\Lambda_t] = \int_0^t nS(s)h(s)ds = n(1 - S(t))$ where S is the survival function.

When it comes to hypothesis testing, a null hypothesis $H_0 : \lambda(t) = \lambda_0(t)$ in the Poisson case implies a hypothesis about the functional form of the intensity for all t. In models with conditional intensity functions, such as in the survival analysis example, the null hypothesis $H_0 : \lambda(t|\mathcal{H}_t) = \lambda_0(t|\mathcal{H}_t)$ is a hypothesis on how the conditional intensity function will evolve over time conditional on the history. In the survival analysis example, suppose the null hypothesis determines the hazard function $h(t) = h_0(t)$, then the null hypothesis specifies $\lambda(t|\mathcal{H}_t) = (n - \sum_i 1[t_i < t])h_0(t)$, which depends on the realized path of the process.

Nevertheless, the time-inhomogeneous Poisson case is of particular interest when A/B testing two independent counting processes. If the intensity function is conditional on the history, and both realized counting processes are independent, then the sharp null hypothesis $H_0 : \lambda^A(t|\mathcal{H}_t^A) = \lambda^B(t|\mathcal{H}_t^B)$ is not plausible because $\mathcal{H}_t^A \neq \mathcal{H}_t^B$. For two sets of independent failure times, for example, then H_0 is not plausible even if they share a common hazard function. Hence the Poisson case is arguably the only reasonable setup for which it makes sense to develop a hypothesis test for difference in intensities. Fortunately, when the counting process is a sum over many (not necessarily Poisson) counting processes then it can be approximated as a Poisson process by the Palm-Khintchine theorem (Heyman and Sobel, 2004). This is often the setup in randomized A/B tests where each randomized experimental unit produces it's own (often very low intensity) counting process, which are combined into a single counting process for each treatment arm.

In the following sections we will often work with the likelihood of the corresponding *point process*. Let $\mathcal{T} \subset \mathbb{R}_{\geq 0}$ denote the times at which events occur, corresponding to the jump times of the counting process N_t. Notationally we write $\mathcal{T} \sim \mathcal{P}(\lambda)$ to denote that $\mathcal{T}$ is a point process with conditional intensity function $\lambda(t|\mathcal{H}_t)$. We use both λ and $\lambda(t|\mathcal{H})$ to refer to the conditional intensity function. To stress that the point process is *Poisson* we write $\mathcal{T} \sim \mathcal{P}\text{oisson}(\lambda)$.

3 CONFIDENCE PROCESS FOR Λ_t

Anytime-valid and sequential inference must rely on martingales (Ramdas et al., 2022). A common construction is to identify a a process X_t that is a nonnegative supermartingale under all distributions of the null hypothesis and apply Ville's inequality $\mathbb{P}[\exists t \geq 0 : X_t \geq \alpha^{-1}] = \alpha$ (Ville, 1939). An α-level sequential test is then obtained by rejecting when $X_t \geq \alpha^{-1}$ (Shafer et al. (2011) calls X_t a *test supermartingale*).

The process $M_t = N_t - \Lambda_t$ is, by definition of the compensator, a martingale, but one cannot apply Ville's inequality as it is nonnegative. The exponential process $\exp(\theta M_t)$ is, however, a nonnegative submartingale by convexity of the exponential function and Jensen's inequality. Yet, for suitable $\psi(\theta)$ and V_t, the process $M_t(\theta) := \exp(\theta(N_t - \Lambda_t) - \psi(\theta)V_t)$ is a nonnegative supermartingale. For a continuous-time counting process, the appropriate terms are $\psi(\theta) = e^\theta - 1 - \theta$ and $V_t = \Lambda_t$ (Howard et al., 2020, 2021), yielding

$$M_t(\theta) := e^{\theta N_t - \Lambda_t(e^\theta - 1)}. \tag{1}$$

Let $\mathcal{S} = \{[a, b] : a, b \in \mathbb{R}_{\geq 0}, b > a\}$ denote a set of intervals in $\mathbb{R}_{\geq 0}$. Howard et al. (2020) leverage the martingale in equation (1) to construct a function $C^\alpha : \mathbb{R}_{\geq 0} \to \mathcal{S}$ adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$, such that

$$\mathbb{P}\left[\Lambda_t \in C_t^\alpha \text{ for all } t > 0\right] \geq 1 - \alpha. \tag{2}$$

We call such a function a *confidence process*, arguing that a confidence interval provides coverage at a single time, a confidence sequence provides coverage across countably infinitely many times, whereas this function provides coverage across uncountably infinitely many times. This is important because we wish to estimate the compensator at all times, not just at the times when counts occur, as sometimes no counts occur. Indeed, observing no counts at all in an interval of time is a very informative observation - it informs us that the intensity over this interval is small.

Note that C_t^α is random as it depends on the observed data, but we suppress this dependence for ease of notation. If we write $C_t^\alpha = [l(t), u(t)]$, then a second confidence process $\bar{C}^\alpha(t) := [l(t)/t, u(t)/t]$ can be defined that covers $\frac{1}{t}\Lambda_t = \frac{1}{t}\int_0^t \lambda(s)ds$ with the same time-uniform coverage guarantee. This enables the time average of the compensator, which connects to the intuitive idea of an average rate, to be continuously estimated as the counting process evolves. In the next section we demonstrate how this result can be obtained by considering a mixture of point process likelihoods with proportional hazards alternatives.

3.1 Likelihood Ratio Mixture Construction

At time t, the intersection of $\mathcal{T} \cap [0, t]$ is observed by the experimenter, where $\mathcal{T} \sim \mathcal{P}(\lambda)$. Given this observation, the likelihood function for $\lambda(t|\mathcal{H}_t)$ is

$$\mathcal{L}_t(\lambda) \propto e^{-\int_0^t \lambda(s|\mathcal{H}_s)ds} \prod_{s \in \mathcal{T} \cap [0,t]} \lambda(s|\mathcal{H}_s). \tag{3}$$

Consider the test of a simple null λ_0 vs a simple alternative λ_1. The likelihood ratio at time t is

$$\frac{\mathcal{L}_t(\lambda_1)}{\mathcal{L}_t(\lambda_0)} = e^{-(\Lambda_{1t}-\Lambda_{0t})} \prod_{s\in\mathcal{T}\cap[0,t]} \frac{\lambda_1(s|\mathcal{H}_s)}{\lambda_0(s|\mathcal{H}_s)}. \quad (4)$$

Our first result states that the likelihood ratio in equation (4) is a continuous time non-negative supermartingale under the null hypothesis.

Theorem 3.1. *Suppose $\mathcal{T} \sim \mathcal{P}(\lambda_0)$, then for any $t > 0$, $\delta > 0$, and $\lambda_1 : \mathbb{R} \to \mathbb{R}_{\geq 0}$*

$$\mathbb{E}\left[\frac{\mathcal{L}_{t+\delta}(\lambda_1)}{\mathcal{L}_{t+\delta}(\lambda_0)} \mid \mathcal{F}_t\right] = \frac{\mathcal{L}_t(\lambda_1)}{\mathcal{L}_t(\lambda_0)}. \quad (5)$$

The likelihood ratio process satisfies the following time-uniform bound

$$\mathbb{P}\left[\sup_{t\geq 0} \frac{\mathcal{L}_t(\lambda_1)}{\mathcal{L}_t(\lambda_0)} \geq \alpha^{-1}\right] = \alpha \quad (6)$$

The likelihood ratio process is a continuous time nonnegative supermartingale under the null hypothesis, regardless of the chosen alternative, according to theorem 3.1. The time-uniform bound in equation (6) follows immediately from Ville's inequality.

Consider a proportional hazards alternative $\lambda_1(t) = e^{\theta}\lambda_0(t)$. The likelihood ratio can now be written as

$$\frac{\mathcal{L}_t(\lambda_1)}{\mathcal{L}_t(\lambda_0)} = e^{\theta N_t - (e^{\theta}-1)\Lambda_{0t}}. \quad (7)$$

The motivation for working with a proportional hazards alternative is the simplification it brings to the likelihood ratio. The data now enters the likelihood ratio only through N_t, and similarly λ_0 now enters the likelihood ratio only through Λ_{0t}. In other words, the exact times in $\mathcal{T}$ are themselves not required, only the total number N_t. The alternative in equation (7) is simple, corresponding to a singleton θ. In theorem 3.3 we will also show that this choice does not in fact limit us to detecting only such alternatives.

A test for a composite alternative can be obtained by mixing the likelihood ratio with respect to a mixture distribution on θ. In practice, distributions which are conjugate to the likelihood are preferred as they yield closed-form expressions for the likelihood ratio mixture. A logGamma(ϕ,ϕ) mixture over θ, which in turn implies a Gamma(ϕ,ϕ) mixture over e^{θ}, is conjugate to the likelihood ratio in equation (7), resulting in a convenient closed form expression (Howard et al., 2021). This mixture distribution has mean 1 and variance $1/\phi$. It is therefore centered on the null hypothesis and ϕ acts as a precision parameter to concentrate the mixture toward local alternatives.

Theorem 3.2. *Assume $\mathcal{T} \sim \mathcal{P}(\lambda_0)$ and for any fixed $\phi > 0$ let Π denote a logGamma(ϕ,ϕ) distribution. Define*

$$\begin{aligned} M_t^{\Lambda_0} :=& \int \frac{\mathcal{L}_t(e^{\theta}\lambda_0)}{\mathcal{L}_t(\lambda_0)} d\Pi(\theta) \\ =& M(N_t, \Lambda_{0t}; \phi), \end{aligned} \quad (8)$$

where

$$M(n, L; \phi) = \frac{\phi^{\phi}}{(\phi+L)^{\phi+n}} \frac{\Gamma(\phi+n)}{\Gamma(\phi)} e^{L}.$$

The process $M_t^{\Lambda_0}$ is a continuous-time nonnegative supermartingale.

Moreover,

$$\mathbb{P}[\exists t > 0 : M_t^{\Lambda_0} \geq \alpha^{-1}] = \alpha. \quad (9)$$

Equation (9) again follow from Ville's inequality. The null hypothesis $H_0 : \mathcal{T} \sim \mathcal{P}(\lambda_0)$ can therefore be rejected as soon as $M_t^{\Lambda_0}$ becomes larger the rejection threshold of α^{-1}, while maintaining a time-uniform Type-I error guarantee of α.

Since the numerator of $M_t^{\Lambda_0}$ only mixes over proportional hazard alternatives, one might worry that the test described in the previous paragraph is only powered to detect such alternatives. The next theorem demonstrates that in fact this test has power 1 against *any* alternative for which the average rate of arrivals differs from the average rate under the null hypothesis.

Theorem 3.3. *Let $\lambda(t|\mathcal{H}_t), \lambda_0(t|\mathcal{H}_t)$ be given such that $\Lambda_{0t}/t \xrightarrow{a.s.} \bar{\lambda}_0$ and $\Lambda_t/t \xrightarrow{a.s.} \bar{\lambda}$. Suppose $\mathcal{T} \sim \mathcal{P}(\lambda)$. Then,*

$$\frac{\log M_t^{\Lambda_0}}{t} \xrightarrow{a.s.} \bar{\lambda}\log\frac{\bar{\lambda}}{\bar{\lambda}_0} - (\bar{\lambda} - \bar{\lambda}_0) = D_{KL}(\bar{\lambda}||\bar{\lambda}_0), \quad (10)$$

where $D_{KL}(\bar{\lambda}||\bar{\lambda}_0)$ is the Kullback-Leibler divergence of a Poisson$(\bar{\lambda})$ distribution from a Poisson$(\bar{\lambda}_0)$ distribution.

Theorem 3.3 states that whenever arrivals come from a process with a different asymptotic average arrival than the null hypothesis, then $M_t^{\Lambda_0}$ grows exponentially quickly under the alternative, providing the exact asymptotic growth rate. Consequently, it will also almost surely grow past α^{-1} leading us to reject the null eventually with probability one.

By considering the complement of equation 9, we obtain a confidence process for the compensator Λ.

Corollary 3.4. *For any fixed $\phi > 0$, the sets defined by*

$$C_t^{\alpha} = \left\{L \in \mathbb{R}_{\geq 0} : M(N_t, L; \phi) \leq \alpha^{-1}\right\} \quad (11)$$

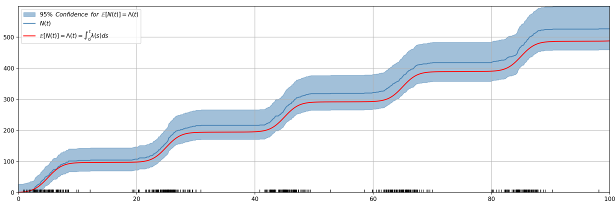

Figure 1: The intensity measure $\Lambda_t = \mathbb{E}[N_t]$ (red), the observed counting process N_t (blue), the 0.95 confidence process C_t^α (shaded blue), and the realized point process (black ticks).

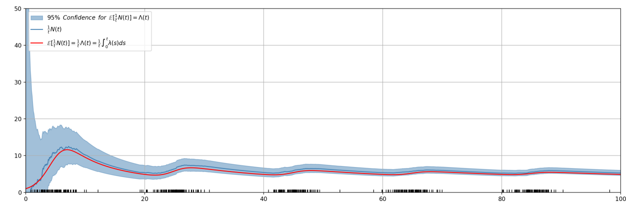

Figure 2: The time-average of intensity $\frac{1}{t}\Lambda_t$ (red), the estimated time-average of intensity (blue), the 0.95 confidence process (shaded blue), and the realized point process (black ticks).

provide a $1-\alpha$ confidence process for Λ_t. Namely, if $\mathcal{T} \sim \mathcal{P}(\lambda_0)$, then

$$\mathbb{P}[\Lambda_{0t} \in C_t^\alpha \ \forall t > 0] = 1 - \alpha.$$

Moreover, if $\lambda_0(t)$ is bounded over $t > 0$ and $\Lambda_{0t}/t \to \bar{\lambda}_0$ then

$$\mathbb{P}[\Lambda_t \notin C_t^\alpha \ \forall t > 0] = 1$$

for any Λ_t with $\Lambda_t/t \to \bar{\lambda} \neq \bar{\lambda}_0$.

Note that the confidence process C_t^α is particularly simple because $\Lambda_0(\cdot)$ only enters into $M_t^{\Lambda_0}$ as Λ_{0t}, its value at t. Had this not been the case, we would generally need to consider C_t^α to be a set of possible $\mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ *functions* $\Lambda(\cdot)$, rather than just a set in $\mathbb{R}_{\geq 0}$. This significant simplification is due to our specific choice of proportional hazard alternative. This is important because it makes C_t^α interpretable and practically usable as we can simply plot it over time (see figure 1). Moreover, since $M(n, L; \phi)$ is convex in L, the set defined in equation (11) is an interval, which can be computed numerically using a root-finding algorithm.

While benefiting greatly in terms of both computation and interpretability of C_t^α from how we chose the space of alternative arrival processes when defining $M_t^{\Lambda_0}$, we do not pay much for this choice in terms of power. We show we are able to exclude any alternative that converges at any distinct average arrival. This is made obvious by figure 2 showing how our confidence process scaled by $1/t$ concentrates at a point.

Example 3.5. *In the following simulated example $\mathcal{T} \sim \mathcal{P}(\lambda)$ with $\lambda(t) = e^{3\sin(2\pi t/20)}$. Figure 1 shows the realized point process and the confidence process C_t^α, which we remark covers $\Lambda_t = \mathbb{E}[N_t]$ at all times in this particular sample path. Figure 2 shows the confidence process on $(1/t)\Lambda_t$.*

4 CONFIDENCE PROCESS FOR $(\Lambda_t^A, \Lambda_t^B)$

In this section consider observing two independent counting processes N_t^B and N_t^A with compensators $\Lambda_t^B = \int_0^t \lambda^B(s|\mathcal{H}_s^B)ds$ and $\Lambda_t^A = \int_0^t \lambda^A(s|\mathcal{H}_s^A)ds$ respectively. We use $\mathcal{F}_t = \sigma\{N^B(s), N^A(s) : 0 \leq s \leq t\}$. To obtain a confidence process on $\Lambda_t^B - \Lambda_t^A$ it is natural to first consider replicating the logic from section 3 to the difference process $N_t^B - N_t^A$. This turns out to be unrewarding as we show in the next section, which motivates the alternative product martingale in section 4.2.

4.1 Generalized Inverse Gaussian Mixture

The process $N_t^B - N_t^A - (\Lambda_t^B - \Lambda_t^A)$ is a martingale with respect to $\mathcal{F}_t$ and by arguments similar to the preceding section

$$M_t^{\Lambda^A,\Lambda^B}(\theta) = e^{\theta(N_t^B - N_t^A) - \Lambda_t^B(e^\theta - 1) - \Lambda_t^A(e^{-\theta} - 1)}$$

is a nonnegative supermartingale. Let $\eta = e^\theta$ so that $M_t(\theta)$ becomes

$$M_t^{\Lambda^A,\Lambda^B}(\eta) = e^{\Lambda_t^B + \Lambda_t^A}\eta^{(N_t^B - N_t^A) - \eta\Lambda_t^B - \frac{\Lambda_t^A}{\eta}}$$

This matches the kernel of a Generalized inverse Gaussian $\mathrm{GIG}(a, b, p)$ with density

$$\pi(\eta) = \frac{(a/b)^{p/2}}{2K_p(2\sqrt{ab})}\eta^{p-1}e^{-a\eta - b/\eta}, \quad \eta > 0,$$

Hence, we can obtain a closed-form $GIG(a_0, b_0, p_0)$ conjugate mixture martingale via $M_t^{\Lambda^A,\Lambda^B} = \int_0^\infty M_t^{\Lambda^A,\Lambda^B}(\eta)d\Pi(\eta) = M(n^A, n^B, L^A, L^B)$ where

$$M(n^A, n^B, L^A, L^B) = e^{L^B + L^A}\frac{(a_0/b_0)^{p_0/2}}{(a/b)^{p/2}}\frac{K_p(2\sqrt{ab})}{K_{p_0}(2\sqrt{a_0 b_0})}, \tag{12}$$

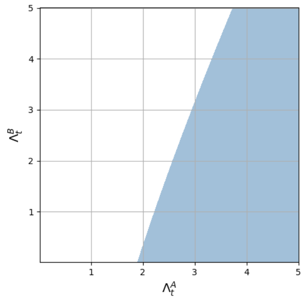

Figure 3: Confidence set $C_t^\alpha = \{(L^A, L^B) \in \mathbb{R}^2 : M(N_t^A, N_t^B, L^A, L^B) \leq \alpha\}$ for $\alpha = 0.05$, $a_0 = b_0 = 1$, $p_0 = 0$ with observations $N_t^A = 30$ and $N_t^B = 22$.

$a = a_0 + L^B$, $b = b_0 + L^A$ and $p = p_0 + n^B - n^A$, and $K_\nu(z)$ is the modified Bessel function of the second kind of order ν. The nonnegative supermartingale in equation (12) is a useful test martingale if one has a sharp null hypothesis about Λ_t^B and Λ_t^A, however inverting it to obtain a confidence process for Λ_t^A and Λ_t^B is not particularly useful as figure 3 shows. The problem is that focusing on the difference $N_t^B - N_t^A$ alone loses information about the magnitude of N_t^B and N_t^A individually. A difference of $N_t^B - N_t^A = -8$ may be surprising for a null hypothesis with relatively small Λ_t^B and Λ_t^A, but is totally in line with expectations for a null hypothesis with relatively large Λ_t^B and Λ_t^A.

4.2 Product Martingale

Next, we construct a test martingale for point nulls Λ_0^A and Λ_0^B by simply considering a product of independent test martingales.

Theorem 4.1. *Assume $\mathcal{T}^A \sim \mathcal{P}(\lambda_0^A)$ and $\mathcal{T}^B \sim \mathcal{P}(\lambda_0^B)$ independently. For any fixed $\phi > 0$ define*

$$M_t^{\Lambda_0^A, \Lambda_0^B} = M_t^{\Lambda_0^A} M_t^{\Lambda_0^B} \tag{13}$$

The process $M_t^{\Lambda_0^A, \Lambda_0^B}$ is a continuous time nonnegative supermartingale.

Theorem 4.1 follows readily from theorem 3.2, as a product of two independent nonnegative supermartingales is itself a nonnegative supermartingale.

Corollary 4.2. *For any fixed $\phi > 0$, the sets defined by*

$$\begin{aligned} C_t^\alpha = \{&(L^A, L^B) \in \mathbb{R}^2_{\geq 0} : \\ &M(N_t^A, L^A; \phi) M(N_t^B, L^B; \phi) \leq \alpha^{-1}\} \end{aligned} \tag{14}$$

form a $1-\alpha$ confidence process for $(\Lambda_t^A, \Lambda_t^B)$. Namely, if $\mathcal{T}^\mathcal{A} \sim \mathcal{P}(\lambda^A), \mathcal{T}^\mathcal{B} \sim \mathcal{P}(\lambda^B)$, then

$$\mathbb{P}[(\Lambda_t^A, \Lambda_t^B) \in C_t^\alpha \ \forall t > 0] = 1 - \alpha.$$

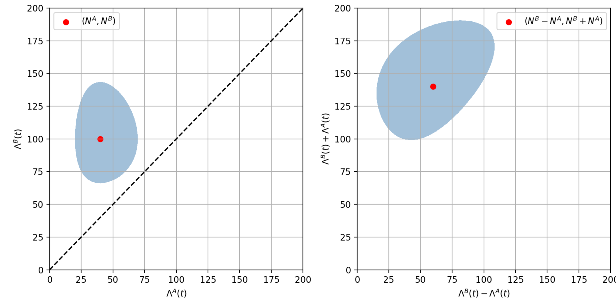

Figure 4: (Left) $1-\alpha$ Confidence set C_t^α for $(\Lambda_t^A, \Lambda_t^B)$. (Right) $1-\alpha$ confidence set $T(C_t^\alpha)$, where $T : (x, y) \to (y - x, y + x)$, for $(\Lambda_t^B - \Lambda_t^A, \Lambda_t^B + \Lambda_t^A)$. Parameters: $\alpha = 0.05$, $\phi = 1$, $N^A = 40$ and $N^B = 100$.

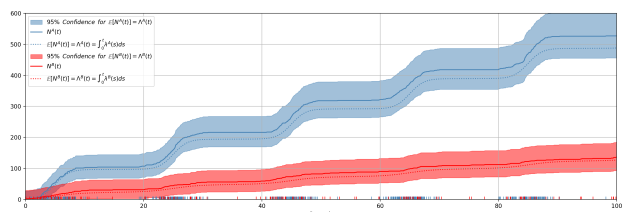

Figure 5: $1 - \alpha$ simultaneous confidence processes on Λ_t^B and Λ_t^A

The set C_t^α is convex because $M(N_t^A, L^A; \phi)M(N_t^B, L^B; \phi)$ is a product of two univariate convex functions in L^A and L^B, respectively. A joint confidence set on $(\Lambda_t^B - \Lambda_t^A, \Lambda_t^A + \Lambda_t^B)$ is obtained via the mapping $T(C_t^\alpha)$ where $T : (x, y) \to (y - x, y + x)$. These sets are visualized in figure 4.

An advantage of the joint confidence process on $(\Lambda_t^A, \Lambda_t^B)$ is that it yields confidence processes on Λ_t^B, Λ_t^A, and the difference $\Lambda_t^B - \Lambda_t^A$ simultaneously. The confidence process for $\Lambda_t^B - \Lambda_t^A$ can be obtained by simply projecting $T(C_t^\alpha)$ along the first dimension, i.e., $\{x \in \mathbb{R}_{\geq 0} : (x, y) \in T(C_t^\alpha) \text{ for some } y\}$. Confidence processes for Λ_t^B and Λ_t^A can be obtained similarly from C_t^α. The lower and upper bounds which define these intervals are trivial to find numerically, requiring only univariate root-finding algorithms, details of which are provided in the appendix.

Example 4.3. *In the following simulated example $\mathcal{T}^A \sim \mathcal{P}(\lambda^A)$ and $\mathcal{T}^B \sim \mathcal{P}(\lambda^B)$ with $\lambda^A(t) = e^{3\sin(2\pi t/20)}$ and $\lambda^B(t) = e^{2\sin(2\pi t/20)}$ respectively. Figure 5 shows the simultaneous $1-\alpha$ confidence processes on Λ_t^A and Λ_t^B from equation 14. Figure 6 shows the simultaneous $1-\alpha$ confidence processes on $(1/t)\Lambda_t^A$ and $(1/t)\Lambda_t^B$. Figure 7 visualizes the confidence process for $\Lambda_t^B/t - \Lambda_t^A/t$.*

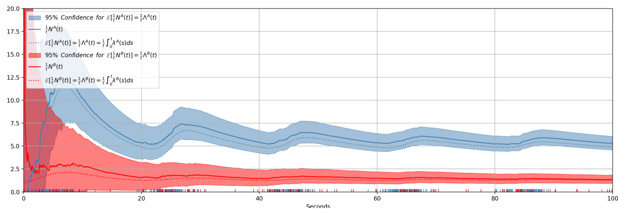

Figure 6: $1-\alpha$ confidence processes on Λ^B_t/t and Λ^A_t/t

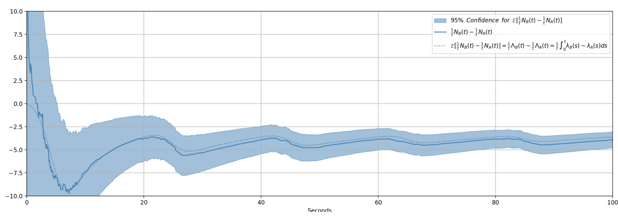

Figure 7: $1-\alpha$ simultaneous confidence processes on $\Lambda^B_t/t - \Lambda^A_t/t$

5 e-PROCESS FOR HYPOTHESIS TESTING

We now turn to the question of testing hypotheses about the A and B arrival processes and in particular equality in their inhomogeneous rates. Following the discussion in section 2, we restrict out attention to inhomogeneous Poisson processes.

Let $\mathcal{I}$ denote the set of all pairs of locally integrable nonnegative functions $\lambda : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, that is, $(\lambda^A, \lambda^B) \in \mathcal{I}$. A null hypothesis H_0 concerning λ^A and λ^B defines the subset $\mathcal{I}_0 \subset \mathcal{I}$. Let $\Theta_{0t} = \{(x, y) \in \mathbb{R}^2_{\geq 0} : x = \int_0^t f(s)ds, y = \int_0^t g(s)ds, (f, g) \in \mathcal{I}_0\}$ denote the set of values taken by the compensators at time t which are consistent with the null hypothesis. The hypothesis H_0 can be rejected at the α-level at the earliest t for which $C^\alpha_t \cap \Theta_{0t} = \emptyset$. For example, the null hypothesis of equality $\lambda^A(t) = \lambda^B(t)\ \forall t \geq$ implies $\Lambda^A_t = \Lambda^B_t\ \forall t \geq$ and can be rejected at the α level as soon as $C^\alpha_t \cap \{(x, x) : x \in \mathbb{R}_{\geq 0}\} = \emptyset$.

To measure the extent to which the null is violated, one can determine the smallest α-level sequential test that would have been rejected. This amounts to, at each t, increasing the size of the $1-\alpha$ confidence set, by decreasing α, until it intersects with Θ_0. This defines a sequential p-value process, providing the guarantee $\mathbb{P}\left[\inf_{t>0} p_t \leq \alpha\right] \leq \alpha$, where $p_t = \inf\{\alpha : C^\alpha_t \cap \Theta_0 = \emptyset\}$. Equivalently, $E_t = 1/p_t$ defines a valid e-process by proposition 12 of Ramdas et al. (2022).

Theorem 5.1. *Assume $\mathcal{T}^A \sim \mathcal{P}oisson(\lambda)$ and $\mathcal{T}^B \sim \mathcal{P}oisson(\lambda)$ independently. For any fixed $\phi > 0$*

$$E_t := \prod_{i\in\{A,B\}} \frac{\phi^\phi}{(\phi + \hat{\Lambda}_t)^{\phi+N^i_t}} \frac{\Gamma(\phi + N^i_t)}{\Gamma(\phi)} e^{\hat{\Lambda}_t}, \qquad (15)$$

where $\hat{\Lambda}_t = \frac{1}{2}\left(N^A_t + N^B_t\right)$, is an e-process.

The next theorem demonstrates that for a broad class of alternatives this test is power 1.

Theorem 5.2. *Assume $\mathcal{T}^A \sim \mathcal{P}(\lambda^A)$ and $\mathcal{T}^B \sim \mathcal{P}(\lambda^B)$ independently, with $\frac{\Lambda^A_t}{t} \xrightarrow{a.s.} \bar{\lambda}^A$ and $\frac{\Lambda^B_t}{t} \xrightarrow{a.s.} \bar{\lambda}^B$ as $t \to \infty$. Let $\bar{\lambda}^M = \frac{1}{2}\left(\bar{\lambda}^B + \bar{\lambda}^A\right)$, then for any $\phi > 0$*

$$\begin{aligned}\frac{\log E_t}{t} &\xrightarrow{a.s.} D_{KL}(\bar{\lambda}^B || \bar{\lambda}^M) + D_{KL}(\bar{\lambda}^A || \bar{\lambda}^M) \\ &= \bar{\lambda}^A \log \frac{2\bar{\lambda}^A}{\bar{\lambda}^A + \bar{\lambda}^B} + \bar{\lambda}^B \log \frac{2\bar{\lambda}^B}{\bar{\lambda}^A + \bar{\lambda}^B}\end{aligned} \qquad (16)$$

where $D_{KL}(\bar{\lambda}^B || \bar{\lambda}^M)$ is the Kullback-Leibler divergence of a Poisson($\bar{\lambda}^B$) distribution from Poisson($\bar{\lambda}^M$)

Note that theorem 5.1 requires the point process be Poisson for E_t to be an e-process, but theorem 5.2 not require the point processes to be Poisson under the alternative. Note also the term on the right hand side is zero when $\bar{\lambda}^A = \bar{\lambda}^B$ and positive otherwise. As a consequence, when the assumptions of theorem 5.2 hold with $\bar{\lambda}^A \neq \bar{\lambda}^B$ then our test will eventually reject the equality hypothesis $\lambda^A = \lambda^B$ with probability one:

$$\mathbb{P}[\exists t > 0 : E_t \geq \alpha^{-1}] = 1.$$

And, how quickly we reject depends on the divergence between $\bar{\lambda}^A$ and $\bar{\lambda}^B$ as defined in equation 16 (which looks almost like a Jenson-Shannon divergence except that Poisson($\bar{\lambda}^M$) is not the mixture of Poisson($\bar{\lambda}^A$) and Poisson($\bar{\lambda}^B$)).

5.1 Comparisons to Other Methods

We now compare to two other possible alternative methods.

5.1.1 Lindon and Malek (2022)

Lindon and Malek (2022) also proposed a sequential test for $\lambda^A(t) = \lambda^B(t)$. They observed that, at any point in time, notwithstanding inhomogeneity, if the equality hypothesis holds then the probability the next event arrives from process B is $1/2$ due to the memoryless property of the Poisson process. This simplified the problem to observing a discrete sequence of independent Bernoulli's with probability $1/2$ under the null hypothesis. They proposed an test martingale obtained by integrating the binomial likelihood ratio with respect to a conjugate $Beta(\alpha, \beta)$ mixture. We can extend their result to continuous time by writing their test martingale as follows

$$\tilde{E}_t = \frac{B(N_t^A + \beta, N_t^B + \alpha)}{B(\alpha, \beta)} 2^{N_t^A + N_t^B}, \qquad (17)$$

where $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$. We can further extend their results by characterising the asymptotic growth rate of this e-process.

Theorem 5.3. *Under the assumptions of theorem 5.2, for any $\alpha, \beta > 0$*

$$\begin{aligned} \frac{\log \tilde{E}_t}{t} &\xrightarrow{a.s.} D_{KL}(\bar{\lambda}^B || \bar{\lambda}^M) + D_{KL}(\bar{\lambda}^A || \bar{\lambda}^M) \\ &= \bar{\lambda}^A \log \frac{2\bar{\lambda}^A}{\bar{\lambda}^A + \bar{\lambda}^B} + \bar{\lambda}^B \log \frac{2\bar{\lambda}^B}{\bar{\lambda}^A + \bar{\lambda}^B} \end{aligned} \qquad (18)$$

where $D_{KL}(\bar{\lambda}^B || \bar{\lambda}^M)$ is the Kullback-Leibler divergence of a Poisson$(\bar{\lambda}^B)$ distribution from Poisson$(\bar{\lambda}^M)$

This asymptotic growth rate is identical to the rate in theorem 5.2. The advantage of our proposal, however, is that while it enjoys the same asymptotic growth rate for testing the equality hypothesis, it additionally also provides *simultaneous* confidence processes on Λ_t^A, Λ_t^B and $\Lambda_t^B - \Lambda_t^A$. Moreover, these are interpretable and visualizable.

5.1.2 Waudby-Smith et al. (2021)

If the only goal of inference is to test the null $\Lambda_t^B - \Lambda_t^A = 0$ for all $t > 0$, then the reader may wonder if our test, obtained by way of constructing a confidence process on $(\Lambda_t^A, \Lambda_t^B)$ first, may be suboptimal in comparison to a method which targets $\Lambda_t^B - \Lambda_t^A$ directly. In this section, we consider strong approximations to difference in counts $N_t^B - N_t^A$, approximating it as (scaled) Brownian motion with drift and leveraging the asymptotic confidence sequences of Waudby-Smith et al. (2021). These asymptotic confidence sequences provide a sequential analogue of fixed sample size tests based on asymptotic normality via the central limit theorem. In this section we demonstrate that this strategy actually leads to a *slower* asymptotic growth compared to our proposal. This is on top of this strategy also only being approximate. For example, in the discrete-time setting, such confidence sequences are only guaranteed to provide α-coverage in an asymptotic regime where we wait longer and longer before we even start testing (Bibaut et al., 2022).

Consider discretizing time into unit intervals and consider the sequence of Poisson random variables $z_i^B = N^B(i) - N^B(i-1)$ and $z_i^A = N^A(i) - N^A(i-1)$ with means $\Lambda_i^B := \int_{i-1}^i \lambda^B(s)ds$ and $\Lambda_i^A := \int_{i-1}^i \lambda^B(s)ds$, respectively. One could then define the sequence of differences $Y_i := z_i^B - z_i^A$, noting the partial sum process $\sum_{i=1}^t Y_i = N_t^B - N_t^A$ and cumulative variance process $\sum_{i=1}^t \Lambda_i^B + \Lambda_i^A = \Lambda_t^B + \Lambda_t^A$, and apply the asymptotic confidence sequences of Waudby-Smith et al. (2021). To apply their result, a few regularity conditions are required. For simplicity we assume that $\lambda^B(t)$ and $\lambda^A(t)$ are bounded and that $\Lambda_t^B/t \to \bar{\lambda}^B$ and $\Lambda_t^A/t \to \bar{\lambda}^A$. These assumptions ensure that the cumulative variance process diverges and that consistent variance estimation is achieved in the sense $((1/t)\sum_i^t z_i^B + z_i^A)/((1/t)\sum_i^t \Lambda_i^B + \Lambda_i^A) = \frac{N_t^B + N_t^A}{\Lambda_t^B + \Lambda_t^A} \xrightarrow{a.s.} 1$. This yields an (asymptotic) e-process, for any mixture precision $\phi > 0$, given by

$$E_t^A = \sqrt{\frac{\phi}{\phi + t}} e^{\frac{1}{2}\frac{t}{t+\phi}Z(t)^2} \qquad (19)$$

where $Z(t) = \frac{N_t^B - N_t^A}{\sqrt{N_t^B + N_t^A}}$. The expression in equation (19) takes the form of a Gaussian-mixture sequential probability ratio test statistic. Our next result shows that the asymptotic growth rate of this e-process is actually *slower* than the e-processes of the preceding sections.

Theorem 5.4. *Under the assumptions of theorem 5.2, for any $\phi > 0$*

$$\lim_{t \to \infty} \frac{\log E_t^A}{t} = \frac{1}{2} \frac{(\bar{\lambda}^B - \bar{\lambda}^A)^2}{\bar{\lambda}^B + \bar{\lambda}^A} \qquad a.s. \qquad (20)$$

which is the Kullback-Leibler divergence of a $N(\bar{\lambda}^B - \bar{\lambda}^A, \bar{\lambda}^B + \bar{\lambda}^A)$ distribution from a $N(0, \bar{\lambda}^B + \bar{\lambda}^A)$ distribution.

This limit is in fact *slower* than our rate in theorem 5.2. Figure 8 shows the relative magnitude of these two limits. When the difference between $\bar{\lambda}^A$ and $\bar{\lambda}^B$ is large, the asymptotic growth rate of E_t is considerably faster than E_t^A. Figure 9 shows 20 realizations $\frac{\log E_t^A}{t}$ and $\frac{\log E_t}{t}$ and the theoretical limits from equations (16) and (20). The slower rate should not be interpreted as a shortcoming of asymptotic confidence sequences at all, rather, it is an argument against only considering a single difference process when each individual process is available. Asymptotic procedures are, nonetheless, approximate and are therefore at risk of incorrectly rejecting the null hypothesis for small t, whereas our proposal is nonasymptotic and exactly valid over all $t > 0$.

6 APPLICATION: SOFTWARE A/B TEST

The following dataset is taken from an engineering A/B test on Android phone and tablet devices, provided with consent from a large internet streaming

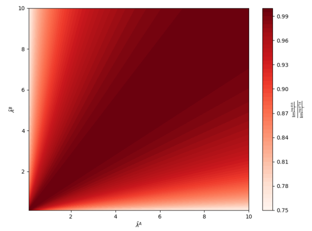

Figure 8: The value of $\lim_{t\to\infty} \frac{\log E_t^A}{t} / \lim_{t\to\infty} \frac{\log E_t}{t}$ for various $\bar{\lambda}^A$ and $\bar{\lambda}^B$

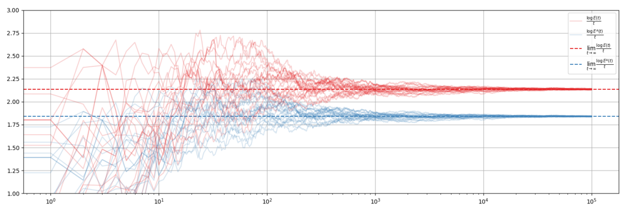

Figure 9: 20 Realizations of $\frac{\log E_t^A}{t}$ (blue) and $\frac{\log E_t}{t}$ (red) and their associated limits from theorems 5.2 and 5.4. Simulated with $\lambda^A(t) = 0.1\lambda^B(t)$ and $\lambda^B(t) = 5 + \sin(2\pi t)$.

company. It was hypothesized that a specific code rewrite could increase playback speed and reduce the number of errors produced. Unfortunately, the new version of the code deployed to devices in the treatment group contained an unintended bug that in fact caused playback to crash. Figure 10 shows a dramatic increase in customer support calls classified as streaming problems. Figure 12 shows the simultaneous 0.95 confidence process on $\Lambda_t^B - \Lambda_t^A$ in addition to the sequential p-value, where a difference is almost immediately detected at the $\alpha = 0.05$ level. The raw timestamps have been scrubbed from figures to protect sensitive information. The simultaneous confidence process on the individual Λ_t^B and Λ_t^A are given in figure 11.

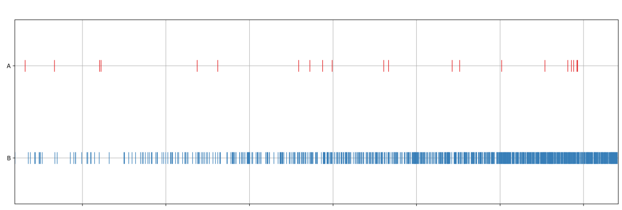

Figure 10: Arrival timestamps of customer support calls for each treatment group

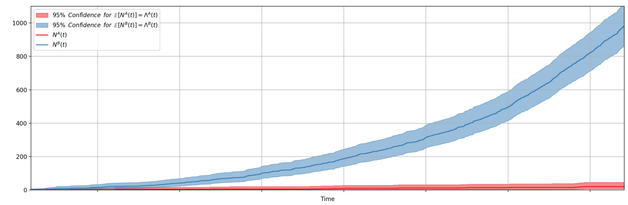

Figure 11: Counting and confidence processes on Λ_t^B and Λ_t^A.

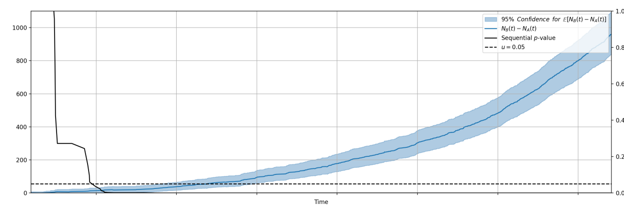

Figure 12: Confidence process on $\Lambda_t^B - \Lambda_t^A$ (left y-axis) and sequential p-value (right y-axis)

7 CONCLUSION

Motivated by the need to continuously monitor arrival processes across treatment arms of a randomized experiment, we provided an anytime-valid inference framework for counting processes. This framework provides multivariate confidence processes on $(\Lambda_t^A, \Lambda_t^B)$, which in turn provide *simultaneous* confidence sequences on Λ_t^A, Λ_t^B, and $\Lambda_t^B - \Lambda_t^A$. These confidence processes can be used to construct an e-process for sequential testing equality of intensity functions among Poisson counting processes. We demonstrate that this test is power 1 for a broad class of alternatives, and that the asymptotic growth rate of the e-process outperforms nonparametric asymptotic procedures which construct confidence sequences on the differences directly. This methodology has been successfully deployed at a large internet streaming company where it is used to monitor all active A/B tests, which we illustrate using an example where this methodology detected a increase in customer call center volume.

References

Anscombe, F. J. (1954). Fixed-sample-size analysis of sequential observations. *Biometrics*, 10(1):89–100.

Armitage, P., McPherson, C. K., and Rowe, B. C. (1969). Repeated significance tests on accumulating data. *Journal of the Royal Statistical Society. Series A (General)*, 132(2):235–244.

Bibaut, A., Kallus, N., and Lindon, M. (2022). Near-optimal non-parametric sequential tests and confidence sequences with possibly dependent observations.

Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., and Zhao, L. (2005). Statistical analysis of a telephone call center: A queueing-science perspective. *Journal of the American statistical association*, 100(469):36–50.

Daley, D. and Vere-Jones, D. (2006). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Probability and Its Applications. Springer New York.

Darling, D. A. and Robbins, H. (1967). Confidence sequences for mean, variance, and median. *Proceedings of the National Academy of Sciences*, 58(1):66–68.

Grünwald, P., de Heide, R., and Koolen, W. (2021). Safe testing.

Ham, D. W., Bojinov, I., Lindon, M., and Tingley, M. (2022). Design-based confidence sequences for anytime-valid causal inference.

Heyman, D. and Sobel, M. (2004). *Stochastic Models in Operations Research: Stochastic optimization*. Dover Books on Computer Science Series. Dover Publications.

Howard, S. R. and Ramdas, A. (2022). Sequential estimation of quantiles with applications to A/B testing and best-arm identification. *Bernoulli*, 28(3):1704 – 1728.

Howard, S. R., Ramdas, A., McAuliffe, J., and Sekhon, J. (2020). Time-uniform Chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 17(none):257 – 317.

Howard, S. R., Ramdas, A., McAuliffe, J., and Sekhon, J. (2021). Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055 – 1080.

Johari, R., Koomen, P., Pekelis, L., and Walsh, D. (2017). Peeking at a/b tests: Why it matters, and what to do about it. KDD '17, New York, NY, USA. Association for Computing Machinery.

Johari, R., Koomen, P., Pekelis, L., and Walsh, D. (2021). Always valid inference: Continuous monitoring of a/b tests. *Operations Research*.

Kuo, L. and Yang, T. Y. (1996). Bayesian computation for nonhomogeneous poisson processes in software reliability. *Journal of the American Statistical Association*, 91(434):763–773.

Lindon, M., Ham, D. W., Tingley, M., and Bojinov, I. (2024). Anytime-valid linear models and regression adjusted causal inference in randomized experiments.

Lindon, M. and Malek, A. (2022). Anytime-valid inference for multinomial count data. In *Advances in Neural Information Processing Systems*.

Lindon, M., Sanden, C., and Shirikian, V. (2022). Rapid regression detection in software deployments through sequential testing. KDD '22, New York, NY, USA. Association for Computing Machinery.

Liptser, R. and Shiryayev, A. (2012). *Theory of Martingales*. Mathematics and its Applications. Springer Netherlands.

Ramdas, A., Grünwald, P., Vovk, V., and Shafer, G. (2023). Game-theoretic statistics and safe anytime-valid inference.

Ramdas, A., Ruf, J., Larsson, M., and Koolen, W. (2022). Admissible anytime-valid sequential inference must rely on nonnegative martingales.

Robbins, H. (1970). Statistical methods related to the law of the iterated logarithm. *The Annals of Mathematical Statistics*, 41(5):1397–1409.

Shafer, G., Shen, A., Vereshchagin, N., and Vovk, V. (2011). Test martingales, bayes factors and p -values. *Statist. Sci.*, 26(1):84–101.

ter Schure, J., Pérez-Ortiz, M. F., Ly, A., and Grunwald, P. (2020). The safe logrank test: Error control under continuous monitoring with unlimited horizon. *arXiv preprint arXiv:2011.06931*.

Turner, R., Ly, A., and Grünwald, P. (2021). Generic e-variables for exact sequential k-sample tests that allow for optional stopping. *arXiv preprint arXiv:2106.02693*.

Turner, R. J. and Grünwald, P. D. (2023). Safe sequential testing and effect estimation in stratified count data.

Ville, J. (1939). *Étude critique de la notion de collectif*. PhD thesis, L'École Polytechnique.

Wald, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Statist.*, 16(2):117–186.

Wang, H. and Ramdas, A. (2023a). Catoni-style confidence sequences for heavy-tailed mean estimation. *Stochastic Processes and their Applications*, 163:168–202.

Wang, H. and Ramdas, A. (2023b). Huber-robust confidence sequences.

Waudby-Smith, I., Arbour, D., Sinha, R., Kennedy, E. H., and Ramdas, A. (2021). Time-uniform central limit theory, asymptotic confidence sequences, and anytime-valid causal inference.

Waudby-Smith, I. and Ramdas, A. (2020). Confidence sequences for sampling without replacement. In *Advances in Neural Information Processing Systems*, volume 33, pages 20204–20214. Curran Associates, Inc.

Checklist

1. For all models and algorithms presented, check if you include:
    (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
    (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes] We note that the only statistic required from the data is the number of events observed in each arm so far i.e. N_t^B and N_t^A. For online stream processing, these can be accumulated in $O(1)$
    (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [No]

2. For any theoretical claim, check if you include:
    (a) Statements of the full set of assumptions of all theoretical results. [Yes]
    (b) Complete proofs of all theoretical results. [Yes] These can be found in the supplemental material
    (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
    (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Not Applicable]
    (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]
    (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Not Applicable]
    (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
    (a) Citations of the creator If your work uses existing assets. [Not Applicable]
    (b) The license information of the assets, if applicable. [Not Applicable]
    (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
    (d) Information about consent from data providers/curators. [Yes]
    (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
    (a) The full text of instructions given to participants and screenshots. [Not Applicable]
    (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
    (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Supplemental Material

A Proofs

A.1 Useful lemmas

Lemma A.1. *Stirling's Approximation*

$$\log\Gamma(x) = \left(x - \frac{1}{2}\right)\log(x) - x - 2\pi + o(1), \tag{21}$$

as $x \to \infty$.

Lemma A.2. *Laplace Functional of Point Process*
Let $\mathcal{T} \sim \mathcal{P}(\lambda)$, f *be predictable process with respect to* $\mathcal{H}_t$ *satisfying* $\int_t^{t+\delta}(e^{f(s)} - 1)\lambda(s|\mathcal{H}_f)ds < \infty$, *then*

$$\mathbb{E}\left[e^{\sum_{s\in\mathcal{T}\cap(t,t+\delta]} f(s)}|\mathcal{H}_t\right] = \mathbb{E}\left[e^{\int_t^{t+\delta} f(s)dN(s)}|\mathcal{H}_t\right] = e^{\int_t^{t+\delta}(e^{f(s)}-1)\lambda(s|\mathcal{H}_f)ds} \tag{22}$$

Proof. See Daley and Vere-Jones (2006, Chapter 5) $\square$

Lemma A.3. *"Strong Law" for Martingales*

Let M_t *be a square integrable local martingale with*

$$\int_0^\infty \frac{1}{(1+s)^2} d\langle M\rangle_s < \infty \ a.s.$$

then $M_t/t \to 0$ *almost surely.*

Proof. This lemma is a special case of Liptser and Shiryayev (2012, p. 144, Theorem 10) $\square$

Lemma A.4. *Suppose* N_t *is a counting process with compensator* $\Lambda_t/t \xrightarrow{a.s.} \bar{\lambda}$, *then* $N_t/t \xrightarrow{a.s.} \bar{\lambda}$

Proof. Write $M_t/t = N_t/t - \Lambda_t/t$. M_t is a local square integrable martingale with predictable quadratic variation $\langle M\rangle_t = \Lambda_t$. By integration by parts

$$\int_0^\infty \frac{1}{(1+s)^2} d\langle M\rangle_s = \left[\frac{\Lambda_s}{(1+s)^2}\right]_0^\infty + 2\int_0^\infty \frac{\Lambda_s}{(1+s)^3} ds \tag{23}$$

As $\Lambda(0) = 0$ and $\Lambda_t = \bar{\lambda}t + o_{a.s.}(t)$ the first term is zero, the latter also implies that the second integral is finite, therefore $M_t/t \to 0$ by lemma A.3 $\Rightarrow N_t/t \to \bar{\lambda}$. $\square$

A.2 Proofs for section 3

A.2.1 Proof of Theorem 3.1

Proof. For any time t and $\delta > 0$, the likelihood can be decomposed into a product of two likelihoods contributed from the observations over $[0, t]$ and $(t, t + \delta]$.

$$\begin{aligned}
\frac{\mathcal{L}_{t+\delta}(\lambda)}{\mathcal{L}_{t+\delta}(\lambda_0)} &= e^{-\int_0^{t+\delta}\lambda_1(s|\mathcal{H}_s)-\lambda_0(s|\mathcal{H}_s)ds} \prod_{s\in\mathcal{T}\cap[0,t+\delta]} \frac{\lambda_1(s|\mathcal{H}_s)}{\lambda_0(s|\mathcal{H}_s)} \\
&= \left(e^{-\int_t^{t+\delta}\lambda_1(s|\mathcal{H}_s)-\lambda_0(s|\mathcal{H}_s)ds} \prod_{s\in\mathcal{T}\cap(t,t+\delta]} \frac{\lambda_1(s|\mathcal{H}_s)}{\lambda_0(s|\mathcal{H}_s)}\right)\left(e^{-\int_0^{t}\lambda_1(s|\mathcal{H}_s)-\lambda_0(s|\mathcal{H}_s)ds} \prod_{s\in\mathcal{T}\cap[0,t]} \frac{\lambda_1(s|\mathcal{H}_s)}{\lambda_0(s|\mathcal{H}_s)}\right) \\
&= \left(e^{-\int_t^{t+\delta}\lambda_1(s|\mathcal{H}_s)-\lambda_0(s|\mathcal{H}_s)ds} \prod_{s\in\mathcal{T}\cap(t,t+\delta]} \frac{\lambda_1(s|\mathcal{H}_s)}{\lambda_0(s|\mathcal{H}_s)}\right)\frac{\mathcal{L}_t(\lambda)}{\mathcal{L}_t(\lambda_0)}.
\end{aligned}$$

The likelihood ratio at time t is a constant with respect to the conditional expectation given $\mathcal{H}_t$. The goal is to show the term that multiplies the likelihood ratio at time t has expectation 1 conditional on the filtration $\mathcal{H}_t$.

$$\begin{aligned}\mathbb{E}\left[\prod_{s\in\mathcal{T}\cap(t,t+\delta]}\frac{\lambda_1(s|\mathcal{H}_s)}{\lambda_0(s|\mathcal{H}_s)}e^{-\int_t^{t+\delta}\lambda_1(s|\mathcal{H}_s)-\lambda_0(s|\mathcal{H}_s)ds}\mid\mathcal{H}_t\right] &= \mathbb{E}\left[\prod_{s\in\mathcal{T}\cap(t,t+\delta]}\frac{\lambda_1(s|\mathcal{H}_s)}{\lambda_0(s|\mathcal{H}_s)}\mid\mathcal{H}_t\right]e^{-\int_t^{t+\delta}\lambda_1(s|\mathcal{H}_s)-\lambda_0(s|\mathcal{H}_s)ds}\\ &= \mathbb{E}\left[\prod_{s\in\mathcal{T}\cap(t,t+\delta]}\frac{\lambda_1(s|\mathcal{H}_s)}{\lambda_0(s|\mathcal{H}_s)}\mid\mathcal{H}_t\right]e^{-\int_t^{t+\delta}\lambda_1(s|\mathcal{H}_s)-\lambda_0(s|\mathcal{H}_s)ds},\end{aligned} \tag{24}$$

where the second line follows from independence of the process over $[0,t]$ and $(t,t+\delta]$. The expectation in the second line can be computed using lemma A.2 of the point process as follows

$$\begin{aligned}\mathbb{E}\left[\prod_{s\in\mathcal{T}\cap(t,t+\delta]}\frac{\lambda_1(s|\mathcal{H}_s)}{\lambda_0(s|\mathcal{H}_s)}\mid\mathcal{H}_t\right] &= e^{\int_t^{t+\delta}\left(\frac{\lambda_1(s|\mathcal{H}_s)}{\lambda_0(s|\mathcal{H}_s)}-1\right)\lambda_0(s|\mathcal{H}_s)ds}\\ &= e^{\int_t^{t+\delta}\lambda_1(s|\mathcal{H}_s)-\lambda_0(s|\mathcal{H}_s)ds}\end{aligned}$$

Hence

$$\begin{aligned}\mathbb{E}\left[\frac{\mathcal{L}_{t+\delta}(\lambda)}{\mathcal{L}_{t+\delta}(\lambda_0)}\mid\mathcal{H}_t\right] &= \mathbb{E}\left[\prod_{s\in\mathcal{T}\cap(t,t+\delta]}\frac{\lambda_1(s|\mathcal{H}_s)}{\lambda_0(s|\mathcal{H}_s)}\mid\mathcal{H}_t\right]e^{-\int_t^{t+\delta}\lambda_1(s|\mathcal{H}_s)-\lambda_0(s|\mathcal{H}_s)ds}\frac{\mathcal{L}_t(\lambda)}{\mathcal{L}_t(\lambda_0)}\\ &= e^{\int_t^{t+\delta}\lambda_1(s|\mathcal{H}_s)-\lambda_0(s|\mathcal{H}_s)ds}e^{-\int_t^{t+\delta}\lambda_1(s|\mathcal{H}_s)-\lambda_0(s|\mathcal{H}_s)ds}\frac{\mathcal{L}_t(\lambda)}{\mathcal{L}_t(\lambda_0)}\\ &= \frac{\mathcal{L}_t(\lambda)}{\mathcal{L}_t(\lambda_0)}.\end{aligned}$$

□

A.2.2 Proof of Theorem 3.2

Proof.

$$\begin{aligned}M_t^{\Lambda_0} = \mathbb{E}^{\Pi}\left[\frac{\mathcal{L}_t(\theta)}{\mathcal{L}_t(\theta_0)}\right] = \int\frac{\mathcal{L}_t(\theta)}{\mathcal{L}_t(\theta_0)}d\Pi(\theta) =& \frac{\phi^\phi}{\Gamma(\phi)}e^{\Lambda_{0t}}\int x^{N_t}e^{-x\Lambda_{0t}}x^{\phi-1}e^{-x\phi}dx\\ =& \frac{\phi^\phi}{(\phi+\Lambda_{0t})^{\phi+N_t}}\frac{\Gamma(\phi+N_t)}{\Gamma(\phi)}e^{\Lambda_{0t}}\end{aligned}$$

Moreover, by Fubini's theorem

$$\begin{aligned}\mathbb{E}[M^{\Lambda_0}(t+\delta)|\mathcal{F}_t] =& \mathbb{E}\left[\int\frac{\mathcal{L}_{t+\delta}(\theta)}{\mathcal{L}_{t+\delta}(\theta_0)}d\Pi(\theta)|\mathcal{F}_t\right]\\ =& \int\mathbb{E}\left[\frac{\mathcal{L}_{t+\delta}(\theta)}{\mathcal{L}_{t+\delta}(\theta_0)}|\mathcal{F}_t\right]d\Pi(\theta)\\ =& \int\frac{\mathcal{L}_t(\theta)}{\mathcal{L}_t(\theta_0)}d\Pi(\theta)\\ =& M_t^{\Lambda_0}\end{aligned}$$

□

A.2.3 Proof of Theorem 3.3

Proof. The assumption $\Lambda_t/t \to \bar{\lambda}$ implies $N_t/t \to \bar{\lambda}$ almost surely by lemma A.4. We therefore write $N_t = \bar{\lambda}t + \epsilon(t)$ where $\epsilon(t) = o_{a.s.}(t)$. Similarly, the assumption $\Lambda_{0t}/t \to \bar{\lambda}_0$ allows us to write $\Lambda_{0t} = \bar{\lambda}_0 t + \epsilon_0(t)$ where $\epsilon_0(t) = o_{a.s.}(t)$. From equation (8)

$$\begin{aligned}\frac{\log M_t^{\Lambda_0}}{t} =& -\frac{(N_t+\phi)}{t}\log(\Lambda_{0t}+\phi) + \frac{\Lambda_{0t}}{t} + \frac{1}{t}\log\Gamma(N_t+\phi) + o(1) \\ =& -(\bar{\lambda} + \frac{\epsilon(t)+\phi}{t})\log(\bar{\lambda}_0 t + \epsilon_0(t) + \phi) + \bar{\lambda}_0 + \frac{1}{t}\log\Gamma(\bar{\lambda}t + \epsilon(t) + \phi) + o_{a.s.}(1).\end{aligned}$$

From lemma A.1, Stirling's approximation gives

$$\begin{aligned}\frac{\log M_t^{\Lambda_0}}{t} =& -\left(\bar{\lambda} + \frac{\epsilon(t)+\phi}{t}\right)\log(\bar{\lambda}_0 t + \epsilon_0(t) + \phi) + \bar{\lambda}_0 \\ &+ \left(\bar{\lambda} + \frac{\epsilon(t)+\phi-\frac{1}{2}}{t}\right)\log(\bar{\lambda}t + \epsilon(t) + \phi) - \bar{\lambda} + o_{a.s.}(1).\end{aligned}$$

Note for $c > 0$, $\log(ct + o_{a.s.}(t)) = \log(ct) + \log(1 + o_{a.s.}(t)/ct) = \log(ct) + o_{a.s.}(1)$, and so we can write

$$\begin{aligned}\frac{\log M_t^{\Lambda_0}}{t} =& -\left(\bar{\lambda} + \frac{\epsilon(t)+\phi}{t}\right)\log(\bar{\lambda}_0 t) + \bar{\lambda}_0 + \left(\bar{\lambda} + \frac{\epsilon(t)+\phi-\frac{1}{2}}{t}\right)\log(\bar{\lambda}t) - \bar{\lambda} + o_{a.s.}(1) \\ =& -\bar{\lambda}\log(\bar{\lambda}_0 t) + \bar{\lambda}\log(\bar{\lambda}t) - (\bar{\lambda} - \bar{\lambda}_0) + \frac{\epsilon(t)+\phi}{t}\log\frac{\bar{\lambda}}{\bar{\lambda}_0} - \frac{1}{2t}\log(\bar{\lambda}t) + o_{a.s.}(1) \\ =& \bar{\lambda}\log\left(\frac{\bar{\lambda}}{\bar{\lambda}_0}\right) - (\bar{\lambda} - \bar{\lambda}_0) + o_{a.s.}(1)\end{aligned}$$

□

A.3 Proofs for section 5

A.3.1 Proof of Theorem 5.1

Recall $C_t^\alpha = \{(L^A, L^B) \in \mathbb{R}^2_{\geq 0} : M(N_t^A, L^A; \phi)M(N_t^B, L^B; \phi) \leq \alpha^{-1}\}$. Let $\Theta_{0t} = \{(x, x) : x \in \mathbb{R}_{\geq 0}\}$. Claim:

$$\begin{aligned}p_t =& \inf\{\alpha : C_t^\alpha \cap \Theta_{0t}\} = \emptyset \\ =& 1/E_t\end{aligned}$$

$$\text{where} \qquad E_t = \prod_{i\in\{A,B\}} \frac{\phi^\phi}{(\phi + \hat{\Lambda}_t)^{\phi + N_t^i}} \frac{\Gamma(\phi + N_t^i)}{\Gamma(\phi)} e^{\hat{\Lambda}_t},$$

and $\hat{\Lambda}_t = \frac{1}{2}(N_t^A + N_t^B)$.

Proof. Observe that

$$E_t = \inf_{(L^A, L^B)\in\Theta_0} M(N_t^A, L^A; \phi)M(N_t^B, L^B; \phi) = M(N_t^A, \hat{\Lambda}_t; \phi)M(N_t^B, \hat{\Lambda}_t; \phi)$$

as $\hat{\Lambda}_t = \mathrm{argmin}_{L\in\mathbb{R}_{\geq 0}} M(N_t^A, L; \phi)M(N_t^B, L; \phi)$. This is easily verified by setting the first derivative to zero and solving for x. $(\hat{\Lambda}_t, \hat{\Lambda}_t)$ is therefore the "last" element in $\mathbb{R}^2_{\geq 0}$ to remain in Θ_0 as the upper bound of α^{-1} is decreased. Formally:

We first show $p_t \leq 1/E_t$:

$p_t = \inf\{\alpha : C_t^\alpha \cap \Theta_0 = \emptyset\} \Rightarrow \forall\alpha < p_t,\ C_t^\alpha \cap \Theta_{0t} \neq \emptyset \Rightarrow \forall\alpha < p_t,\ \exists(L^A, L^B) \in \Theta_0$, such that $M(N_t^A, L^A; \phi)M(N_t^B, L^B; \phi) \leq \alpha^{-1} \Rightarrow \forall\alpha < p_t,\ \inf_{(L^A,L^B)\in\Theta_0} M(N_t^A, L^A; \phi)M(N_t^B, L^B; \phi) \leq \alpha^{-1} \Rightarrow \forall\alpha < p_t$, $\alpha \leq 1/E_t \Rightarrow p_t \leq 1/E_t$.

Last, we show $p_t \geq 1/E_t$:

Conversely $p_t = \inf\{\alpha : C_t^\alpha \cap \Theta_{0t} = \emptyset\} \Rightarrow \forall \alpha > p_t,\ C_t^\alpha \cap \Theta_{0t} = \emptyset \Rightarrow \forall \alpha > p_t,\ \forall (L^A, L^B) \in \Theta_0$, $M(N_t^A, L^A; \phi) M(N_t^B, L^B; \phi) > \alpha^{-1} \Rightarrow \forall \alpha > p_t,\ \inf_{(L^A,L^B)\in\Theta_0} M(N_t^A, L^A; \phi) M(N_t^B, L^B; \phi) > \alpha^{-1}, \Rightarrow \forall \alpha > p_t$, $\alpha > 1/E_t, \Rightarrow p_t \geq 1/E_t$. □

A.3.2 Proof of Theorem 5.2

Proof. Up to terms constant in t, $\log E_t$ can be written as

$$\begin{aligned}\log E_t = &-(2\phi + N_t^A + N_t^B) \log\left(\phi + \frac{1}{2}N_t^A + \frac{1}{2}N_t^B\right)\\ &+ \log\Gamma(N_t^A + \phi)\\ &+ \log\Gamma(N_t^B + \phi)\\ &+ N_t^A + N_t^B\\ &+ const.\end{aligned}$$

Assumption $\Lambda_t^B/t \to \bar{\lambda}^B$ implies $N_t^B/t \to \bar{\lambda}^B$ almost surely by lemma A.4, which we write as $N_t^B = \bar{\lambda}^B t + \epsilon^B(t)$, where $\epsilon^B(t) = o_{a.s.}(t)$. Similarly for $N_t^A = \bar{\lambda}^A t + \epsilon^A(t)$. Substituting these expressions for N_t^A and N_t^B and applying Stirling's approximation in lemma A.1 yields

$$\begin{aligned}\log E_t = &-(\bar{\lambda}^A t + \bar{\lambda}^B t + \epsilon^A(t) + \epsilon^B(t) + 2\phi) \log\left(\frac{1}{2}\bar{\lambda}^A t + \frac{1}{2}\bar{\lambda}^B t + \frac{1}{2}\epsilon^A(t) + \frac{1}{2}\epsilon^B(t) + \phi\right)\\ &+ (\bar{\lambda}^A t + \epsilon^A(t) + \phi - \frac{1}{2}) \log(\bar{\lambda}^A t + \epsilon^A(t) + \phi) - (\bar{\lambda}^A t + \epsilon^A(t) + \phi)\\ &+ (\bar{\lambda}^B t + \epsilon^B(t) + \phi - \frac{1}{2}) \log(\bar{\lambda}^B t + \epsilon^B(t) + \phi) - (\bar{\lambda}^B t + \epsilon^B(t) + \phi)\\ &+ \bar{\lambda}^A t + \epsilon^A(t) + \bar{\lambda}^B t + \epsilon^B(t)\\ &+ const.\end{aligned}$$

Dividing by t and cancelling terms yields

$$\begin{aligned}\frac{\log E_t}{t} = &-\left(\bar{\lambda}^A + \bar{\lambda}^B + \frac{\epsilon^A(t) + \epsilon^B(t) + 2\phi}{t}\right) \log\left(\frac{1}{2}\bar{\lambda}^A t + \frac{1}{2}\bar{\lambda}^B t + \frac{1}{2}\epsilon^A(t) + \frac{1}{2}\epsilon^B(t) + \phi\right)\\ &+ \left(\bar{\lambda}^A + \frac{\epsilon^A(t) + \phi - \frac{1}{2}}{t}\right) \log(\bar{\lambda}^A t + \epsilon^A(t) + \phi)\\ &+ \left(\bar{\lambda}^B + \frac{\epsilon^B(t) + \phi - \frac{1}{2}}{t}\right) \log(\bar{\lambda}^B t + \epsilon^B(t) + \phi)\\ &+ o_{a.s.}(1)\end{aligned}$$

Replacing expressions like $\log(\bar{\lambda}^A t + \epsilon^A(t) + \phi)$ as $\log(\bar{\lambda}^A t) + o_{a.s.}(1)$ yields

$$\begin{aligned}\frac{\log E_t}{t} = &-\left(\bar{\lambda}^A + \bar{\lambda}^B + \frac{\epsilon^A(t) + \epsilon^B(t) + 2\phi}{t}\right) \log\left(\frac{1}{2}\bar{\lambda}^A t + \frac{1}{2}\bar{\lambda}^B t\right)\\ &+ \left(\bar{\lambda}^A + \frac{\epsilon^A(t) + \phi - \frac{1}{2}}{t}\right) \log(\bar{\lambda}^A t)\\ &+ \left(\bar{\lambda}^B + \frac{\epsilon^B(t) + \phi - \frac{1}{2}}{t}\right) \log(\bar{\lambda}^B t)\\ &+ o_{a.s.}(1)\end{aligned}$$

The $\frac{\epsilon^A(t)}{t}\log t$ and $\frac{\epsilon^B(t)}{t}\log t$ terms cancel, with other terms being absorbed into the $o_{a.s.}(1)$ term, yielding

$$\begin{aligned}\frac{\log E_t}{t} = &-\left(\bar{\lambda}^A + \bar{\lambda}^B\right) \log\left(\frac{1}{2}\bar{\lambda}^A + \frac{1}{2}\bar{\lambda}^B\right) + \bar{\lambda}^A \log(\bar{\lambda}^A) + \bar{\lambda}^B \log(\bar{\lambda}^B) + o_{a.s.}(1)\\ =&\bar{\lambda}^A \log\frac{2\bar{\lambda}^A}{\bar{\lambda}^A + \bar{\lambda}^B} + \bar{\lambda}^B \log\frac{2\bar{\lambda}^B}{\bar{\lambda}^A + \bar{\lambda}^B} \qquad + o_{a.s.}(1)\end{aligned}$$

□

A.3.3 Proof of Theorem 5.3

Proof. Assumption $\Lambda_t^B/t \to \bar{\lambda}^B$ implies $N_t^B/t \to \bar{\lambda}^B$ almost surely by lemma A.4, which we write as $N_t^B = \bar{\lambda}^B t + \epsilon^B(t)$, where $\epsilon^B(t) = o_{a.s.}(t)$. Similarly for $N_t^A = \bar{\lambda}^A t + \epsilon^A(t)$. Substituting these expressions for N_t^A and N_t^B and applying Stirling's approximation in lemma A.1 yields

$$\begin{aligned}\frac{\log \tilde{E}_t}{t} &= \left(\bar{\lambda}^A + \frac{\epsilon^A(t)+\beta}{t}\right)\log(\bar{\lambda}^A t + \epsilon^A(t) + \beta)\\ &+ \left(\bar{\lambda}^B + \frac{\epsilon^B(t)+\alpha}{t}\right)\log(\bar{\lambda}^B t + \epsilon^B(t) + \alpha)\\ &- \left(\bar{\lambda}^A + \bar{\lambda}^B + \frac{\alpha+\beta+\epsilon^A(t)+\epsilon^B(t)}{t}\right)\log(\alpha+\beta+\bar{\lambda}^A t + \epsilon^A(t) + \bar{\lambda}^B t + \epsilon^B(t))\\ &+ (\bar{\lambda}^A + \bar{\lambda}^B)\log 2\\ &+ o_{a.s.}(1).\end{aligned} \tag{25}$$

The first logarithmic expression can be written

$$\begin{aligned}\log(\bar{\lambda}^A t + \epsilon^A(t) + \beta) &= \log(\bar{\lambda}^A t) + \log\left(1 + \frac{\epsilon^A(t)+\beta}{\bar{\lambda}^A t}\right)\\ &= \log(\bar{\lambda}^A t) + o_{a.s.}(1),\end{aligned}$$

and similarly for the second and third, yielding

$$\begin{aligned}\frac{\log \tilde{E}_t}{t} &= \left(\bar{\lambda}^A + \frac{\epsilon^A(t)+\beta}{t}\right)\log(\bar{\lambda}^A t)\\ &+ \left(\bar{\lambda}^B + \frac{\epsilon^B(t)+\alpha}{t}\right)\log(\bar{\lambda}^B t)\\ &- \left(\bar{\lambda}^A + \bar{\lambda}^B + \frac{\alpha+\beta+\epsilon^A(t)+\epsilon^B(t)}{t}\right)\log(\bar{\lambda}^A t + \bar{\lambda}^B t)\\ &+ (\bar{\lambda}^A + \bar{\lambda}^B)\log 2\\ &+ o_{a.s.}(1).\end{aligned} \tag{26}$$

The $\frac{\epsilon^A(t)}{t}\log t$ and $\frac{\epsilon^B(t)}{t}\log t$ terms cancel

$$\begin{aligned}\frac{\log \tilde{E}_t}{t} &= \left(\bar{\lambda}^A + \frac{\epsilon^A(t)+\beta}{t}\right)\log(\bar{\lambda}^A)\\ &+ \left(\bar{\lambda}^B + \frac{\epsilon^B(t)+\alpha}{t}\right)\log(\bar{\lambda}^B)\\ &- \left(\bar{\lambda}^A + \bar{\lambda}^B + \frac{\alpha+\beta+\epsilon^A(t)+\epsilon^B(t)}{t}\right)\log(\bar{\lambda}^A + \bar{\lambda}^B)\\ &+ (\bar{\lambda}^A + \bar{\lambda}^B)\log 2\\ &+ o_{a.s.}(1)\\ &= \bar{\lambda}^A\log(\bar{\lambda}^A) + \bar{\lambda}^B\log(\bar{\lambda}^B) - \left(\bar{\lambda}^A + \bar{\lambda}^B\right)\log(\bar{\lambda}^A + \bar{\lambda}^B) + (\bar{\lambda}^A + \bar{\lambda}^B)\log 2 + o_{a.s.}(1)\\ &= \bar{\lambda}^A\log\frac{2\bar{\lambda}^A}{\bar{\lambda}^A + \bar{\lambda}^B} + \bar{\lambda}^B\log\frac{2\bar{\lambda}^B}{\bar{\lambda}^A + \bar{\lambda}^B} + o_{a.s.}(1)\end{aligned}$$

□

A.3.4 Proof of Theorem 5.4

Proof. Assumption $\Lambda_t^B/t \to \bar{\lambda}^B$ implies $N_t^B/t \to \bar{\lambda}^B$ almost surely by lemma A.4, which we write as $N_t^B = \bar{\lambda}^B t + \epsilon^B(t)$, where $\epsilon^B(t) = o_{a.s.}(t)$. Similarly for $N_t^A = \bar{\lambda}^A t + \epsilon^A(t)$, where $\epsilon^A(t) = o_{a.s.}(t)$.

$$
\begin{aligned}
\frac{\log E_t^A}{t} &= -\frac{1}{2t}\log(\phi + t) + \frac{1}{2}\frac{1}{\phi + t}\frac{(\bar{\lambda}^B t - \bar{\lambda}^A t + \epsilon^B(t) - \epsilon^A(t))^2}{\bar{\lambda}^B t + \bar{\lambda}^A t + \epsilon^B(t) + \epsilon^A(t)} + const. \\
&= o_{a.s.}(1) + \frac{1}{2}\frac{1}{1 + \phi/t}\frac{(\bar{\lambda}^B - \bar{\lambda}^A + o_{a.s.}(1))^2}{\bar{\lambda}^B + \bar{\lambda}^A + o_{a.s.}(1)} \\
&= \frac{1}{2}\frac{(\bar{\lambda}^B - \bar{\lambda}^A)^2}{\bar{\lambda}^B + \bar{\lambda}^A} + o_{a.s.}(1)
\end{aligned}
$$

□

B Numerical Computation of Confidence Processes

B.1 Numerical Computation of Confidence Processes for $\lambda^A(t)$ and $\lambda^B(t)$

We present the numerical method for computing the confidence process of Λ_t^B only. Recall from equation (14) that the set C_t^α comprises the set of elements $(x_1, x_2) \in \mathbb{R}^2_{\geq 0}$ satisfying

$$
\frac{\phi^\phi}{(\phi + x_2)^{\phi + N_t^B}}\frac{\Gamma(\phi + N_t^B)}{\Gamma(\phi)}e^{x_2} \leq \alpha^{-1}\left(\frac{\phi^\phi}{(\phi + x_1)^{\phi + N_t^A}}\frac{\Gamma(\phi + N_t^A)}{\Gamma(\phi)}e^{x_1}\right)^{-1}.
$$

The term on the right-hand side of the inequality obtains a maximum when $x_1 = N_t^A$. Define a constant

$$
k^A(t) = \frac{1}{\alpha\phi^{2\phi}}\frac{(\phi + N_t^A)^{\phi + N_t^A}}{e^{N_t^A}}\frac{\Gamma(\phi)}{\Gamma(\phi + N_t^A)}\frac{\Gamma(\phi)}{\Gamma(\phi + N_t^B)},
$$

and a function

$$
f^B(x) = \frac{e^x}{(\phi + x)^{\phi + N_t^B}} - K^A(t).
$$

Then $\{y > 0 : (x, y) \in C_t^\alpha \text{ for some } x\} = \{y : f^B(y) \leq 0\}$. The extrema of the interval defining the confidence process for Λ_t^B are simply the roots of $f^B(x)$. The result for Λ_t^A is similar.

B.2 Numerical Computation of Confidence Processes for $\Lambda_t^B - \Lambda_t^A$

The transformed confidence set is given by

$$
T(C_t^\alpha) = \{(w, v) \in \mathbb{R} \times \mathbb{R}_{\geq 0} : g(w, v) \leq -\log\alpha\},
$$

where

$$
\begin{aligned}
g(w, v) = &v - (\phi + N_t^B)\log\left(\phi + \frac{1}{2}(v + w)\right) - (\phi + N_t^A)\log\left(\phi + \frac{1}{2}(v - w)\right) \\
&+ 2\phi\log\phi - 2\log\Gamma(\phi) + \log\Gamma(\phi + N_t^A) + \log\Gamma(\phi + N_t^B).
\end{aligned}
$$

The interval for $\Lambda_t^B - \Lambda_t^A$ is the projection $\{w : (w, v) \in T(C_t^\alpha) \text{ for some } v\}$, and we seek the lower and upper bounds that define this interval. These can be obtained by maximizing and minimizing the function $f(w, v) = w$ over $T(C_t^\alpha)$. Let (w_u, v_u) and (w_l, v_l) denote the solutions to the maximization and minimization problems respectively. Now consider the two equations

$$
\begin{aligned}
g(w, v) &= \alpha^{-1}, \\
\frac{\partial g(w, v)}{\partial v} &= 0
\end{aligned}
\tag{27}
$$

The second equation allows v to be expressed in terms of w as

$$v = h(w) = \frac{1}{2}(N_t^A + N_t^B) - \phi + \sqrt{a+b},$$

where

$$a = \frac{1}{4}(N_t^A)^2 + \frac{1}{2}N_t^A N_t^B + N_t^A\phi + N_t^A w,$$
$$b = \frac{1}{4}(N_t^B)^2 + N_t^B\phi - N_t^B w + \phi^2 + w^2.$$

The solution for w can then be found by solving $g(w, h(w)) = -\log(\alpha)$.

The solutions to the optimization problem (w_l, v_l) and (w_u, v_u) must satisfy equations (27) as the solutions live on the boundary of $T(C_t^\alpha)$, satisfying the first equation, and ∇g has zero v component, satisfying the second equation. The values w_l and w_u are then the two roots of the function $g(w, h(w)) + \log\alpha$.

C Supplemental Figures

Figure 13 shows the intensity function and the realized point process from example 3.5. Figure 14 shows the two intensity functions and realized point processes from example 4.3.

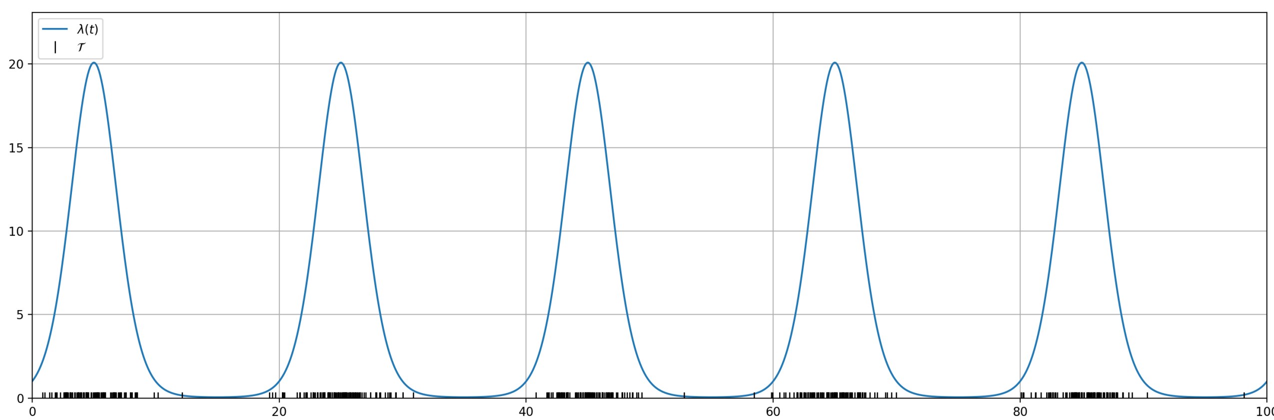

Figure 13: Intensity function $\lambda(t)$ (blue) with realized point process $\mathcal{T}$ (black ticks)

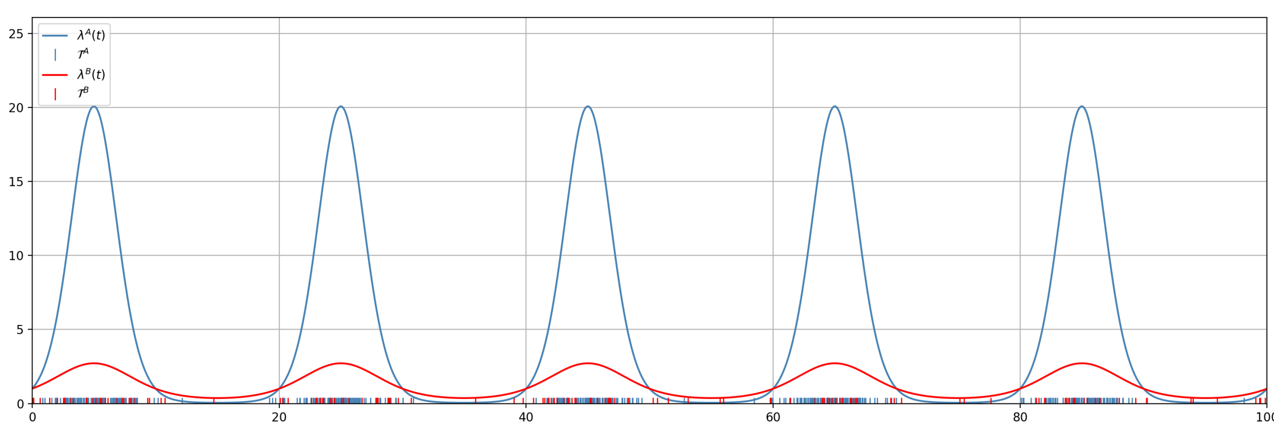

Figure 14: $\mathcal{T}^A \sim \mathcal{P}(\lambda^A)$ and $\mathcal{T}^B \sim \mathcal{P}(\lambda^B)$ with $\lambda^A(t) = e^{3\sin(2\pi t/20)}$ and $\lambda^B(t) = e^{2\sin(2\pi t/20)}$