Supplementary Material – Persistent Surveillance of Events with Unknown Rate Statistics

Cenk Baykal[1], Guy Rosman[1], Kyle Kotowick[1], Mark Donahue[2], and Daniela Rus[1]

[1] Massachusetts Institute of Technology (MIT), Cambridge MA 02139, USA,
{baykal, rosman, kotowick, rus}@csail.mit.edu,
[2] MIT Lincoln Laboratory, Lexington MA 02421, USA,
mark.donahue@ll.mit.edu

1 Introduction

This document contains technical material intended to supplement the submission "Persistent Surveillance of Events with Unknown Rate Statistics." In this manuscript, we present supportive proofs of the theoretical results presented in Sect. 4.

2 Proof of Lemma 1

Lemma 1 (Satisfaction of the uncertainty constraint). *The observation time $t_{i,k}^{low}$ given by* (8) *satisfies the uncertainty constraint* (4) *for any arbitrary station $i \in [n]$ and iteration $k \in \mathbb{N}_+$.*

Proof. We consider the left-hand side of (6) from Sect. 3 and marginalize over the unknown parameter $\lambda_i \in \mathbb{R}_+$:

$$\mathbb{P}\left(N_i(t_{i,k}) \leq k(t_{i,k}) | X_i^{1:k-1}\right) = \int_0^\infty \mathbb{P}\left(N_i(t_{i,k}) \leq k(t_{i,k}) | X_i^{1:k-1}, \lambda\right) \mathbb{P}\left(\lambda | X_i^{1:k-1}\right) d\lambda$$

where the probability is with respect to the random variable $N_i(t_{i,k}) \sim \text{Pois}(\lambda t_{i,k})$ $\forall \lambda \in \mathbb{R}_+$ by definition of a Poisson process with parameter λ. Using the equal-tails credible interval constructed in Alg. 2, i.e. the interval $(\lambda_i^{\mathrm{l}}, \lambda_i^{u})$ satisfying

$$\forall i \in [n] \quad \forall \lambda_i \in \mathbb{R}_+ \; \mathbb{P}\left(\lambda_i^{\mathrm{l}} > \lambda_i | X_i^{1:k-1}\right) = \mathbb{P}\left(\lambda_i^{\mathrm{u}} < \lambda_i | X_i^{1:k-1}\right) = \frac{\epsilon}{2},$$

we establish the inequalities:

$$\begin{aligned}
\mathbb{P}\left(N_i(t_{i,k}) \leq k(t_{i,k}) | X_i^{1:k-1}\right) &> \int_0^{\lambda_i^u} \mathbb{P}\left(N_i(t_{i,k}) \leq k(t_{i,k}) | X_i^{1:k-1}, \lambda\right) \mathbb{P}\left(\lambda | X_i^{1:k-1}\right) d\lambda \\
&\geq \mathbb{P}\left(N_i(t_{i,k}) \leq k(t_{i,k}) | X_i^{1:k-1}, \lambda_i^u\right) \int_0^{\lambda_i^u} \mathbb{P}\left(\lambda | X_i^{1:k-1}\right) d\lambda \\
&= (1 - \frac{\epsilon}{2}) \, \mathbb{P}\left(N_i(t_{i,k}) \leq k(t_{i,k}) | X_i^{1:k-1}, \lambda_i^u\right). \qquad (1)
\end{aligned}$$

where we utilized the fact that $\mathbb{P}\left(N_i(t_{i,k}) \leq k(t_{i,k}) | X_i^{1:k-1}, \lambda_i^{\mathrm{u}}\right)$ is monotonically decreasing with respect to λ. By construction, $t_{i,k}^{\mathrm{low}}$ satisfies $D_{\mathrm{KL}}(\mathrm{Pois}(\lambda_i^{\mathrm{u}} t_{i,k}^{\mathrm{low}}) \,||\, \mathrm{Pois}(k(t_{i,k}^{\mathrm{low}}))) = W_\epsilon$ which yields $1 - g(W_\epsilon) = 1 - \frac{\epsilon}{2-\epsilon}$ by definition and thus by (11) we have:

$$\mathbb{P}\left(N_i(t_{i,k}^{\mathrm{low}}) \leq k(t_{i,k}^{\mathrm{low}}) | X_i^{1:k-1}, \lambda_i^{\mathrm{u}}\right) > 1 - g(W_\epsilon) = 1 - \frac{\epsilon}{2-\epsilon}.$$

Combining this inequality with the expression of (1) establishes the result. □

3 Proof of Lemma 2

Lemma 2 (Monotonicity of solutions satisfying (4)**).** *For any arbitrary station $i \in [n]$ and monitoring cycle $k \in \mathbb{N}_+$, the observation time $t_{i,k}$ satisfying $t_{i,k} \geq t_{i,k}^{low}$, where $t_{i,k}^{low}$ is given by* (8)*, satisfies the uncertainty constraint.*

Proof. Differentiation of g yields $\forall x \in \mathbb{R}_+$

$$g'(x) \geq \frac{e^{-x}}{\max\left\{2, 2\sqrt{\pi x}\right\}} > 0$$

which establishes that g is monotonically increasing in its argument. Now, consider the KL divergence function $D_{\mathrm{KL}}(\mathrm{Pois}(m) \,||\, \mathrm{Pois}(k))$ which is a function of m and k. From Sec. 3 we know that m and k are in turn functions of $t_{i,k}$, namely, $m(t_{i,k}) := \lambda_i^{\mathrm{u}} t_{i,k}$ and $k(t_{i,k}) := k(t_{i,k}) = \delta \frac{\alpha_i}{\beta_i^2}(\beta_i + t_{i,k})^2 - \alpha_i$. Taking the total derivative of D_{KL} with respect to $t_{i,k}$ yields:

$$\begin{aligned}
\frac{dD_{\mathrm{KL}}}{dt_{i,k}} &= \frac{\partial D_{\mathrm{KL}}}{\partial m}\frac{dm}{dt_{i,k}} + \frac{\partial D_{\mathrm{KL}}}{\partial k}\frac{dk}{dt_{i,k}} \\
&= \big(1 - \frac{k}{m}\big)\lambda_i^u + \ln\frac{k}{m}\frac{2\alpha_i\delta(\beta_i + t_{i,k})}{\beta_i^2} \\
&\geq \big(1 - \frac{k}{m}\big)\lambda_i^u - \big(1 - \frac{k}{m}\big)\frac{2\alpha_i\delta(\beta_i + t_{i,k})}{\beta_i^2} \\
&= \big(1 - \frac{k}{m}\big)\big(\lambda_i^u - \frac{2\alpha_i\delta(\beta_i + t_{i,k})}{\beta_i^2}\big).
\end{aligned}$$

Note that since the inequality (11) established by [1] only holds for $k \geq m > 0$, for any $\epsilon \in \big(0, 2(1+2e^{1/\pi})^{-1}\big), \delta \in (0,1)$, we have that $k(t_{i,k}^{\mathrm{low}}) \geq m(t_{i,k}^{\mathrm{low}})$ by construction of $t_{i,k}^{\mathrm{low}}$ using (8). Moreover, differentiation of the ratio $k(t_{i,k})/m(t_{i,k})$ yields

$$\frac{d(k(t_{i,k})/m(t_{i,k}))}{dt} = \frac{\alpha_i}{\lambda_i^u}\big(\frac{\delta}{\beta_i^2} + \frac{1-\delta}{t_{i,k}^2}\big) > 0$$

which proves the monotonically increasing property of this ratio. Employing the monotonicity of the ratio, we conclude that for all $t_{i,k} \geq t_{i,k}^{\mathrm{low}}$, we have $k(t_{i,k}) \geq m(t_{i,k})$.

Hence, we have the following sequence of inequalities

$$\begin{aligned} k(t_{i,k}) &\geq m(t_{i,k}) \\ \delta \frac{\alpha_i}{\beta_i^2}(\beta_i + t_{i,k})^2 - \alpha_i &\geq \lambda_i^u t_{i,k} \\ \frac{\delta \alpha_i t_{i,k}}{\beta_i^2} + \frac{2\alpha_i \delta}{\beta_i} + \frac{\alpha_i \delta}{t_{i,k}} - \frac{\alpha_i}{t_{i,k}} &\geq \lambda_i^u. \end{aligned}$$

Employing the constraint $\delta \in (0,1)$ yields $\frac{\alpha_i \delta}{t_{i,k}} - \frac{\alpha_i}{t_{i,k}} < 0$, which implies that

$$\frac{\delta \alpha_i t_{i,k}}{\beta_i^2} + \frac{2\alpha_i \delta}{\beta_i} > \lambda_i^u \quad \Rightarrow \frac{2\delta\alpha_i(\beta_i + t_{i,k})}{\beta_i^2} > \lambda_i^u.$$

Thus, we have that $\big(1 - \frac{k}{m}\big) \leq 0$ and $\big(\lambda_i^u - \frac{2\alpha_i\delta(\beta_i + t_{i,k})}{\beta_i^2}\big) \leq 0$ for all $t_{i,k} \geq t_{i,k}^{\mathrm{low}}$, which yields

$$\frac{dD_{\mathrm{KL}}}{dt_{i,k}} \geq \big(1 - \frac{k}{m}\big)\big(\lambda_i^u - \frac{2\alpha_i\delta(\beta_i + t_{i,k})}{\beta_i^2}\big) \geq 0$$

which establishes that D_{KL} is ensured to be monotonically increasing for all $t_{i,k} \geq t_{i,k}^{\mathrm{low}}$.

Putting it all together, we have that for all $t_{i,k} \geq t_{i,k}^{\mathrm{low}}$, $g\big(D_{\mathrm{KL}}(\mathrm{Pois}(m(t_{i,k}))\,||\,\mathrm{Pois}(k(t_{i,k}))\big) \geq g\big(D_{\mathrm{KL}}(\mathrm{Pois}(m(t_{i,k}^{\mathrm{low}})\,||\,\mathrm{Pois}(t_{i,k}^{\mathrm{low}}))\big)$ which implies that $t_{i,k}$ also satisfies the uncertainty constraint, i.e.,

$$\begin{aligned} \forall t_{i,k} \geq t_{i,k}^{\mathrm{low}} \quad \mathbb{P}\left(N_i(t_{i,k}) \leq k(t_{i,k}) | X_i^{1:k-1}\right) > g(t_{i,k}) &\geq g(t_{i,k}^{\mathrm{low}}) \\ &> 1 - \epsilon. \end{aligned}$$

□

4 Proof of Theorem 1

Theorem 1 (Per-cycle approximate-optimality of solutions). *For any arbitrary cycle $k \in \mathbb{N}_+$, the policy $\pi_k^* := (t_{1,k}^*, \ldots, t_{n,k}^*)$ generated by Alg. 2 is an approximately-optimal solution with respect to Problem 3.*

Proof. Recall that $\pi_k^* := (t_{1,k}^*, \ldots, t_{n,k}^*)$ is an optimal solution to Problem 3 if and only if π_k^* (i) satisfies the uncertainty constraint for all stations $i \in [n]$ and (ii) maximizes the balance of observations, i.e.

$$\begin{aligned} &\pi_k^* \in \operatorname*{argmax}_{\pi_k} \hat{f}_{\mathrm{bal}}(\pi_k) \\ &\text{s.t.} \;\; \forall i \in [n] \quad \mathbb{P}\left(\mathrm{Var}(\lambda_i | X_i^{1:k}, \pi_k) \leq \delta \mathrm{Var}(\lambda_i | X_i^{1:k-1}) \big| X_i^{1:k-1}\right) > 1 - \epsilon. \end{aligned}$$

By definition of each $t_{i,k}^*$ in π_k^*, we have that for all stations $i \in [n]$ and iterations $k \in \mathbb{N}_+$,

$$t_{i,k}^* = \frac{N_{\max}}{\hat{\lambda}_{i,k}} \geq t_{i,k}^{\mathrm{low}}.$$

Invoking Lemma 2, we conclude that $t^*_{i,k}$ satisfies the uncertainty constraint for all stations, which establishes that condition (i) holds for the constructed policy π^*_k.

Now, to establish that (ii) holds for π^*_k, note that for any arbitrary policy π_k, we have that

$$\mathbb{E}[N_1(\pi_k)] = \cdots = \mathbb{E}[N_n(\pi_k)] \Leftrightarrow \pi_k \in \underset{\pi}{\operatorname{argmax}} \hat{f}_{\text{bal}}(\pi).$$

In other words, π_k optimizes the objective function for balance if and only if the expected number of observations under π_k is equal for each station [2]. Now, note that for the constructed $\pi^*_k = (t^*_{1,k}, \ldots, t^*_{n,k})$, we have:

$$\hat{\lambda}_{1,k} t^*_{1,k} = N_{\max}, \hat{\lambda}_{2,k} t^*_{2,k} = N_{\max}, \ldots, \hat{\lambda}_{n,k} t^*_{n,k} = N_{\max}$$

which implies that

$$\mathbb{E}[N_1(\pi^*_k)] = N_{\max} = \cdots = \mathbb{E}[N_n(\pi^*_k)]$$

and thus we conclude that condition (ii) holds for π^*_k, i.e., $\pi^*_k \in \operatorname{argmax}_{\pi_k} \hat{f}_{\text{bal}}(\pi_k)$. □

5 Proof of Lemma 3

Lemma 3 (Bound on posterior variance). *After executing an arbitrary number of cycles $k \in \mathbb{N}_+$, the posterior variance $Var(\lambda_i | X_i^{1:k})$ is bounded above by $\delta^k \, Var(\lambda_i)$ with probability at least $(1-\epsilon)^k$, i.e.,*

$$\forall i \in [n] \quad \forall k \in \mathbb{N}_+ \quad \mathbb{P}\left(Var(\lambda_i | X_i^{1:k}) \le \delta^k \, Var(\lambda_i) | X_i^{1:k} \right) > (1-\epsilon)^k$$

for all stations $i \in [n]$ where $Var(\lambda_i) := \alpha_{i,0}/\beta^2_{i,0}$ is the prior variance.

Proof. From Lemma 2 we have that each $t^*_{i,k}$ is ensured to satisfy the uncertainty condition (4) $\forall i \in [n]$

$$\mathbb{P}\left(\text{Var}(\lambda_i | X_i^{1:k}) \le \delta \text{Var}(\lambda_i | X_i^{1:k-1}) | X_i^{1:k-1} \right) > 1 - \epsilon \tag{2}$$

for each iteration k regardless of the events that transpire in the other iterations. Hence, the probability of satisfying this condition for k consecutive iterations is greater than $(1-\epsilon)^k$. This implies that, with probability at least $(1-\epsilon)^k$, we have that the following chain of of inequalities holds:

$$\begin{aligned} Var\big(\lambda_i | X_i^1\big) &\le \delta Var\big(\lambda_i\big), \\ Var\big(\lambda_i | X_i^{1:2}\big) &\le \delta Var\big(\lambda_i | X_i^1\big) = \delta^2 Var\big(\lambda_i\big), \\ &\vdots \\ Var\big(\lambda_i | X_i^{1:k}\big) &\le \delta Var\big(\lambda_i | X_i^{1:k-1}\big) = \delta^k Var\big(\lambda_i\big). \end{aligned}$$

□

6 Proof of Theorem 2

Theorem 2 (ξ-bound on approximation error). *For all $\xi \in \mathbb{R}_+$ and cycles $k \in \mathbb{N}_+$, the inequality $|\hat{\lambda}_{i,k} - \lambda_i| < \xi$ holds with probability at least $(1-\epsilon)^{k-1}(1-\frac{\delta^{k-1}\,Var(\lambda_i)}{\xi^2})$, i.e.,*

$$\forall i \in [n] \quad \mathbb{P}\left(|\hat{\lambda}_{i,k} - \lambda_i| < \xi | X_i^{1:k-1}\right) > (1-\epsilon)^{k-1}\left(1 - \frac{\delta^{k-1}\,Var(\lambda_i)}{\xi^2}\right).$$

Proof. Note that by Chebyshev's inequality states the following:

$$\mathbb{P}\left(|\hat{\lambda}_{i,k} - \lambda_i| < \xi | X_i^{1:k-1}\right) > 1 - \frac{\mathrm{Var}(\hat{\lambda}_{i,k}|X_i^{1:k-1})}{\xi^2}.$$

In light of Corollary 1, we have that

$$\mathbb{P}\left(\mathrm{Var}(\hat{\lambda}_{i,k}|X_i^{1:k-1}) \leq \delta^{k-1}\mathrm{Var}(\lambda_i)|X_i^{1:k-1}\right) > (1-\epsilon)^{k-1}$$

employing this inequality and Chebyshev's inequality yields:

$$\begin{aligned}\mathbb{P}\left(|\hat{\lambda}_{i,k} - \lambda_i| < \xi \big| X_i^{1:k-1}\right) &> (1-\epsilon)^{k-1}\left(1 - \frac{\mathrm{Var}(\hat{\lambda}_{i,k}|X_i^{1:k-1})}{\xi^2}\right) \\ &> (1-\epsilon)^{k-1}\left(1 - \frac{\delta^{k-1}\mathrm{Var}(\lambda_i)}{\xi^2}\right)\end{aligned}$$

□

7 Proof of Theorem 3

Theorem 3 (Δ-bound on optimality with respect to Problem 3). *For any $\xi_i \in \mathbb{R}_+$, $i \in [n], k \in \mathbb{N}_+$, given that $|\hat{\lambda}_{i,k} - \lambda_i| \in (0, \xi_i)$ with probability as given in Theorem 2, let $\sigma_{min} := \sum_{i=1}^{n}(\lambda_i - \xi_i)^{-1}$ and $\sigma_{max} := \sum_{i=1}^{n}(\lambda_i + \xi_i)^{-1}$. Then, the objective value of the policy π_k^* at iteration k is within a factor of Δ of the ground-truth optimal solution, where $\Delta := \frac{\sigma_{min}}{\sigma_{max}}$ with probability greater than $(1-\epsilon)^{n(k-1)}\left(1 - \frac{\delta^{k-1}\,Var(\lambda_i)}{\xi^2}\right)^n$.*

Proof. Let $T = \sum_{i=1}^{n} t_{i,k}^*$ be the total observation time allocated by the generated policy. Then, by the optimality of policy $\pi_k^* = (t_{1,k}^*, \ldots, t_{n,k}^*)$ with respect to the rate approximations, we have the following equalities

$$\hat{\lambda}_{1,k}t_{1,k}^* = N_{\max}, \hat{\lambda}_{2,k}t_{2,k}^* = N_{\max}, \ldots, \hat{\lambda}_{n,k}t_{n,k}^* = N_{\max}.$$

which implies that

$$\forall i \in [n] \quad t_{i,k}^* := \frac{T}{\lambda_i \sum_{l=1}^{n} \frac{1}{\lambda_l}}.$$

Now recall that the objective function pertaining to balance (1) is given by:

$$\hat{f}_{\mathrm{bal}}(\pi_k) := \min_i \frac{\mathbb{E}[N_i(\pi_k)]}{\sum_{j=1}^n \mathbb{E}[N_j(\pi_k)]}.$$

and the optimal (maximal) value of this function is $\frac{1}{n}$. Now, using the fact that $|\hat{\lambda}_{i,k} - \lambda_i| < \xi_i$, we have the following inequalities for $\hat{f}_{\mathrm{bal}}(\pi_k^*)$

$$\begin{aligned}\hat{f}_{\mathrm{bal}}(\pi_k^*) &= \frac{\min_i \lambda_i t_{i,k}^*}{\sum_{j=1}^n \lambda_j t_{j,k}^*} = \frac{\min_i \frac{T}{\sum_{l=1}^n (\lambda_l)^{-1}}}{\sum_{j=1}^n \frac{T}{\sum_{l=1}^n (\lambda_l)^{-1}}} \\ &> \frac{\frac{T}{\sum_{l=1}^n (\lambda_i + \xi_l)^{-1}}}{\frac{nT}{\sum_{l=1}^n (\lambda_l - \xi_l)^{-1}}} = \frac{\sum_{l=1}^n (\lambda_l - \xi_l)^{-1}}{n \sum_{l=1}^n (\lambda_l + \xi_l)^{-1}} \\ &= \frac{1}{n}\Big(\frac{\sigma_{\min}}{\sigma_{\max}}\Big)\end{aligned}$$

with probability at least $(1-\epsilon)^{n(k-1)}\big(1 - \frac{\delta^{k-1}\mathrm{Var}(\lambda_i)}{\xi^2}\big)^n$. $\square$

References

1. M. Short. Improved inequalities for the Poisson and binomial distribution and upper tail quantile functions. *ISRN Probability and Statistics*, 2013.
2. J. Yu, S. Karaman, and D. Rus. Persistent monitoring of events with stochastic arrivals at multiple stations. *T-RO*, 31(3):521–535, 2015.