CRAMER-RAO BOUND FOR THE TIME-VARYING POISSON

Xinhui Rong[†]*, and Victor Solo*[⋆†]

[†]School of Electrical Engineering and Telecommunications,
UNSW, Sydney, AUSTRALIA

ABSTRACT

Point processes are finding increasing applications in neuroscience, genomics, and social media. But basic modelling properties are little studied. Here we consider a periodic time-varying Poisson model and develop the asymptotic Cramer-Rao bound. We also develop, for the first time, a maximum likelihood algorithm for parameter estimation.

***Index Terms*—** Point process, Cramer-Rao bound, asymptotic rate, periodic intensity, MLE

1. INTRODUCTION

Point processes occur in a wide range of applications, such as neural coding [1], stochastic finance [2], social media [3] and an emerging area, event triggered state estimation [4].

The simplest useful model is the Time-varying Poisson (tvP) which is the focus of this paper. The tvP occurs in optical detection and auditory electro-physiology [5] where the intensity function is periodic. More recently, generalization to an almost periodic intensity function is discussed in [6] and [7].

Asymptotic performance of point process estimators is still emerging. [6] develop a central limit theorem (CLT) for frequency domain estimators for a tvP with an aperiodic intensity. [8],[9] develop CLTs for maximum likelihood estimators (MLEs).

But none of these authors has provided either an algorithm for MLE estimation or a CRB.

We make three contributions in this paper. We calculate the asymptotic CRB for a periodic model. This confirms the super efficiency of the estimator of the frequency. We also develop an MLE for a simple periodic model. It seems to be the first such algorithm. We show the CRB is achieved by the MLE, but is not achieved by the frequency domain estimator used in [6]. We illustrate these results with a number of simulations.

⋆ is the corresponding author (email: v.solo@unsw.edu.au). This work was partly supported by an Australian Research Council grant.

The rest of the paper is organized as follows. In section 2, we review point processes and the CRB. In section 3, we derive the asymptotic CRB for periodic intensities. In section 4, we propose a new ML algorithm by cyclic ascent. In section 5, we give simulations. Conclusions are in section 6.

2. PRELIMINARIES

Here we briefly review point processes and the tvP as well as the CRB.

2.1. Point Processes and the tvP

A scalar point process is defined by the counting process $N_t = N(0,t]$ =# of events up to and including time t or by the sequence of event times $T_1, T_2, \cdots$ at which the counting process jumps.

A point process is uniquely characterized by its conditional intensity function

$$\lambda(t) = \lambda(t|\mathcal{H}_t) = \lim_{\delta \downarrow 0} \frac{1}{\delta} \Pr[N_{t+\delta} - N_t = 1|\mathcal{H}_t],$$

where $\mathcal{H}_t$ is the history up to time t: $\{N_s, 0 \le s \le t\}$.

For a tvP process, the count increments are independent of the past and so its intensity function is deterministic. We assume $0 < \lambda(t) < \infty$.

The log-likelihood of a point process over an interval $[0,T]$ is given by [5]

$$\mathcal{L}_T = \int_0^T \ln \lambda(t) dN_t - \Lambda(T)$$

where $\Lambda(T) = \int_0^T \lambda(t)dt$ is the integrated intensity.

We assume the intensity is parametrized by a parameter vector θ of dimension d. We denote the estimators $\hat{\theta}$ and assume they are (asymptotically) unbiassed.

2.2. Cramer-Rao Bound for the tvP

The Cramer-Rao Bound (CRB) gives a lower bound on the variance of any parameter estimator. For lack of space we deal only with the unbiased case.

For the tvP, the CRB has been given by [5]. The more general history dependent case can be found in [10].

Theorem 1 *tvP CRB [5]. For an unbiased estimator*

$$V_T = \text{var}[\hat{\theta}] \geq \mathcal{I}_T^{-1}.$$

where $\mathcal{I}_T$ is Fisher Information Matrix (FIM)

$$\mathcal{I}_T = \int_0^T \frac{1}{\lambda(t)} \frac{\partial \lambda(t)}{\partial \theta} \frac{\partial \lambda(t)}{\partial \theta^\top} dt,$$

where $A \geq B$ means $A - B$ is positive semi-definite.

3. ASYMPTOTIC ANALYSIS

In this section, we carry out asymptotic analysis of the CRB. We only do asymptotics since exact calculations are very messy, while asymptotics yield clear formulae and allow direct comparison with CLTs.

We focus on periodic intensities motivated by the simple model

$$\lambda(t) = b + acos(\omega t)$$

where $\theta = [b, a, \omega]^\top$ and ω is the unknown frequency, b is an unknown background rate, and a is an unknown amplitude. To ensure $\lambda(t) > 0$ we require $b > a$.

3.1. Averages of Periodic Functions

Suppose $\gamma(t) \geq 0$ has period $\frac{2\pi}{\omega}$. We can then introduce the scale free function of period 2π

$$\kappa(\tau) = \gamma(\frac{\tau}{\omega}) \Rightarrow \gamma(t) = \kappa(\omega t).$$

Asymptotic analysis is enabled by the following result.

Result I. Consider the integral

$$\mathcal{J}_T = \frac{1}{T^{p+1}} \int_0^T t^p \gamma(t)(1 - \alpha f_{q,r}(\omega t))dt, \qquad (3.1)$$

where $f_{q,r}(\tau) = sin^q(\tau)cos^r(\tau)$ and $p, q, r, \geq 0$ are integers and $|\alpha| < 1$. Then

$$\begin{aligned} \mathcal{J}_T &\rightarrow \frac{1}{p+1} L_{q,r}(\alpha) \text{ as } T \rightarrow \infty \\ L_{q,r}(\alpha) &= \int_0^{2\pi} \kappa(\xi)(1 - \alpha f_{q,r}(\xi)) \frac{d\xi}{2\pi}. \end{aligned}$$

Proof. First, find the integer m, so that

$$\begin{aligned} & 2m\pi \leq \omega T < 2m\pi + 2\pi \\ \Rightarrow \quad & 0 \leq \omega T - 2m\pi \leq 2\pi \\ \Rightarrow \quad & 0 \leq 1 - \tfrac{2m\pi}{\omega T} \leq \tfrac{2\pi}{\omega T} \\ \Rightarrow \quad & \tfrac{2m\pi}{T\omega} \rightarrow 1 \text{ as } T \rightarrow \infty \end{aligned} \qquad (3.2)$$

The integrand is ≥ 0 and changing variables gives

$$\mathcal{J}_T = \frac{1}{(\omega T)^{p+1}} \int_0^{\omega T} \tau^p \kappa(\tau)(1 - \alpha f_{q,r}(\tau))d\tau.$$

Now consider that

$$\begin{aligned} & \Sigma_1^m [(k-1)2\pi]^p \int_{(k-1)2\pi}^{k2\pi} \kappa(\tau)(1 - \alpha f_{q,r}(\tau))d\tau \\ & \leq (\omega T)^{p+1} \mathcal{J}_T \\ & \leq \Sigma_1^m [k2\pi]^p \int_{(k-1)2\pi}^{k2\pi} \kappa(\tau)(1 - \alpha f_{q,r}(\tau))d\tau. \end{aligned}$$

By periodicity,

$$\int_{(k-1)2\pi}^{k2\pi} \kappa(\tau)(1 - \alpha f_{q,r}(\tau))d\tau = 2\pi L_{q,r}(\alpha).$$

Then,

$$\begin{aligned} & L_{q,r}(\alpha) 2\pi \Sigma_1^{m-1} k^p (2\pi)^p \\ = \quad & \Sigma_1^m [(k-1)2\pi]^p L_{q,r}(\alpha) 2\pi \\ \leq \quad & (\omega T)^{p+1} \mathcal{J}_T \\ \leq \quad & \Sigma_1^{m+1} [k2\pi]^p L_{q,r}(\alpha) 2\pi. \end{aligned}$$

Now divide through by $(\omega T)^{p+1} L_{q,r}(\alpha)$ to find

$$\begin{aligned} 0 \quad \leq \quad & \frac{\mathcal{J}_T}{L_{q,r}(\alpha)} - \frac{\Sigma_1^{m-1} k^p (2\pi)^{p+1}}{(\omega T)^{p+1}} \\ \leq \quad & (2\pi)^{p+1} \frac{(m+1)^p + m^p}{(\omega T)^p} \frac{1}{\omega T} \rightarrow 0 \end{aligned}$$

by (3.2). Now we observe that $\Sigma_1^{m-1} \left(\frac{k2\pi}{\omega T}\right)^p \frac{2\pi}{\omega T}$ is a Riemann sum converging to the integral $\int_0^1 x^p dx = \frac{1}{p+1}$ and the result follows.

Remarks.

(i) Set $\alpha = 0$ to find

$$\tfrac{1}{T^{p+1}} \int_0^T t^p \gamma(t) dt \rightarrow \tfrac{1}{T^{p+1}} \int_0^{2\pi} \kappa(\xi) \tfrac{d\xi}{2\pi}.$$

(ii) Subtract (3.1) with $\alpha = 1$ from (i) to find

$$\int_0^T \tfrac{t^p}{T^{p+1}} \gamma(t) f_{q,r}(\omega t) dt \rightarrow \int_0^{2\pi} \kappa(\xi) f_{q,r}(\xi) \tfrac{d\xi}{2\pi}.$$

(iii) When q is odd, the limit in (ii) is 0 since sin is an odd function. .

3.2. Asymptotic CRB

To derive the asymptotic CRB for $\hat{\theta}$ we introduce the matrix $D_T = \text{diag}(T^{\frac{1}{2}}, T^{\frac{1}{2}}, T^{\frac{3}{2}})$.

Result II.

$$\begin{aligned} D_T V_T D_T &\geq D_T \mathcal{I}_T^{-1} D_T \\ &= (D_T^{-1} \mathcal{I}_T D_T^{-1})^{-1} \to V_* \\ V_* &= \begin{bmatrix} bG & 0 \\ 0 & \frac{3R}{a\rho} \end{bmatrix} \text{ \& } \rho = \frac{a}{b} \\ G &= \begin{bmatrix} 1 & \rho \\ \rho & R \end{bmatrix} \text{ \& } R = 1 + \sqrt{1-\rho^2}. \end{aligned}$$

Proof. Set $c_t = cos(\omega t)$ and $s_t = sin(\omega t)$. The gradient is $\frac{d\lambda(t)}{d\theta} = [1, c_t, -ats_t]^\top$. Then we find $D_T^{-1}\mathcal{I}_T D_T^{-1} =$

$$\frac{1}{T}\int_0^T \begin{bmatrix} 1 & c_t & -as_t\frac{t}{T} \\ c_t & c_t^2 & -as_t c_t \frac{t}{T} \\ -as_t\frac{t}{T} & -as_t c_t \frac{t}{T} & a^2 s_t^2 \frac{t^2}{T^2} \end{bmatrix} \frac{dt}{\lambda(t)}$$

Applying result II and noting by remark(iii) that the off diagonal limits with a sine term vanish, gives V_*^{-1}. Inverting it gives the quoted result. The limits have been evaluted using classic definite integrals such as

$$\int_0^{2\pi} \frac{1}{b + acos(\xi)} \frac{d\xi}{2\pi} = \frac{1}{\sqrt{b^2 - a^2}}$$

4. MLE VIA CYCLIC ASCENT

For MLE it turns out to be useful to reparameterize

$$\lambda(t) = b + acos(\omega t) = c + a(1 + cos(\omega t))$$

where now $\theta = [c, a, \omega]^\top$ and $c = b - a$ and $a > 0$.. If $a < 0$ we write $b - |a|cos(\omega t) = c + |a|(1 - cos(\omega t)$ where $c = b - |a|$ and again the new coefficients are > 0. Since the map between b, a and c, a is 1-1 we can recover the MLE of b as $\hat{b} = \hat{c} + \hat{a}$.

We find MLEs by cyclic ascent. In the first stage, given ω we get c, a by an EM algorithm. In the second stage, given c, a we we get ω by a Newton iteration.

4.1. Stage I: EM

The EM is a special case of a tvP EM due to [11]. The idea is to partition $\lambda(t)$ into pieces, each of which is an intensity for a simpler process. Then an EM can be assembled for each piece. In the current setting we have $\lambda(t) = c + a(1 + cos(\omega t))$. The first component is c the intensity of a time-invariant Poisson process. The second component is an intensity ≥ 0 for a tvP.

We can also motivate the EM heuristically as follows. The MLE equations obtained by setting derivatives of the log-likelihood to zero are (we set $n = N_T$)

$$\begin{aligned} 0 = \frac{\partial \mathcal{L}_T}{\partial c} &= \Sigma_1^n \frac{1}{\lambda(t_r)} - T \\ 0 = \frac{d\mathcal{L}_T}{da} &= \Sigma_1^n \frac{\mu(\omega t_r)}{\lambda(t_r)} - B_T(\omega) \end{aligned}$$

where $\mu(\omega t) = 1 + cos(\omega t)$ and $B_T(\omega) = \int_0^T \mu(\omega t) dt$. We now rewrite these as

$$\begin{aligned} 1 &= \frac{1}{T}\Sigma_1^n \frac{1}{\lambda(t_r)} \\ 1 &= \frac{1}{B_T(\omega)}\Sigma_1^n \frac{\mu(\omega t_r)}{\lambda(t_r)} \end{aligned}$$

These lead to the following EM updates

$$\begin{aligned} \hat{c}_{k+1} &= \frac{\hat{c}_k}{T}\Sigma_1^n \frac{1}{\hat{\lambda}^k(t_r)} \\ \hat{a}_{k+1} &= \frac{\hat{a}_k}{B_T(\omega)}\Sigma_1^n \frac{\mu(\omega t_r)}{\hat{\lambda}^k(t_r)} \end{aligned}$$

where $\hat{\lambda}^k(t_r) = \hat{c}_k + \hat{a}_k(1 + cos(\omega t_r))$. Note that the EM updates guarantee positivity.

4.2. Stage II: Newton-Raphson iteration

We get the derivative for ω

$$\frac{\partial \mathcal{L}_T}{\partial \omega} = \Sigma_1^n \frac{a t_r \mu'(\omega t_r)}{\lambda(t_r)} - \frac{\partial B_T(\omega)}{\partial \omega}.$$

We also have

$$I_T(\omega) = \mathrm{E}[\frac{\partial^2 \mathcal{L}_T}{\partial \omega^2}] = \int_0^T \frac{(at\mu'(\omega t))^2}{\lambda(t)} dt.$$

Then, replacing the Hessian by its expected value we get the scoring algorithm

$$\hat{\omega}_{k+1} = \hat{\omega}_k + \eta I_T^{-1}(\hat{\omega}_k) \frac{\partial \ell}{\partial \hat{\omega}_k},$$

where $0 < \eta \leq 1$ is the step size.

It would be of interest to study the convergence of the overall algorithm. But even for the EM component convergence analysis is not done in [11].

5. SIMULATIONS

In this section, we show simulations of the MLE algorithm and compare results with the asymptotics.

5.1. Simulation Setup

By using scale free measures of estimator precision we can collapse many plots onto one, which simplifies interpretation. The parameters of interest b, a, ω, T are collapsed as follows. $n_b = bT$ is the expected number of background events, $m_T = \omega T/(2\pi)$ is the number of cosine periods observed. The scale-free estimator precision measure is $\pi(\hat{\theta}) = \frac{\theta}{se(\hat{\theta})}$ where $se(\hat{\theta})$ is the asymptotic CRB standard error. From result II we find

$$[se(\hat{b}), se(\hat{a}), se(\hat{\omega})] \approx \frac{1}{\sqrt{T}}[\sqrt{b}, \sqrt{bR}, \frac{\sqrt{3R/\rho}}{T\sqrt{a}}].$$

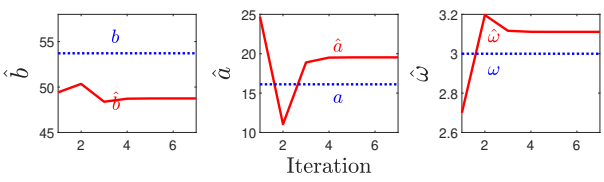

Fig. 1. Parameter iterates with initial frequency $\hat{\omega}_0 = 2.97$

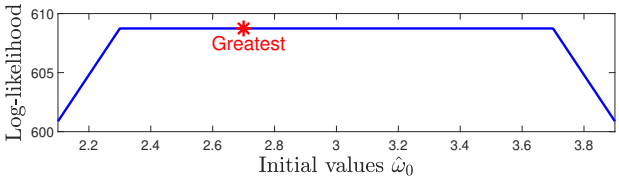

Fig. 2. Maximized Log-likelihoods from 20 initial frequencies.

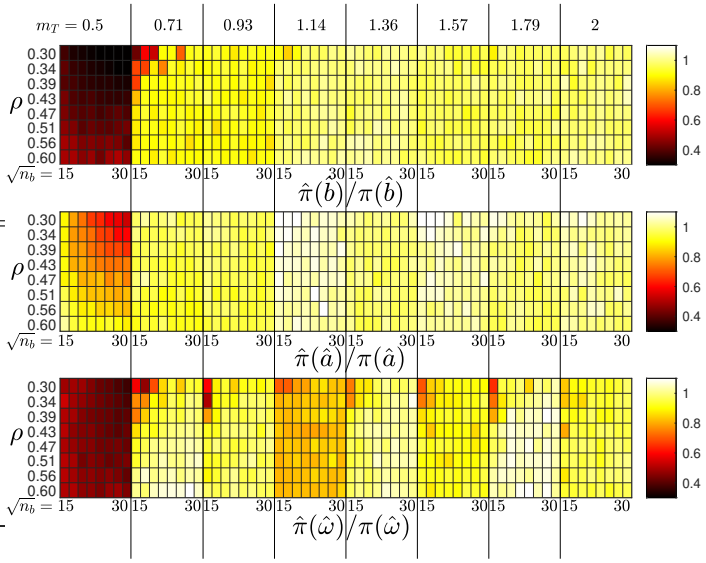

Fig. 3. Heatmaps of $\hat{\pi}/\pi$.

where $R = 1 + \sqrt{1-\rho^2}$. Set $\gamma = 1 - \sqrt{1-\rho^2} \Rightarrow \rho = \sqrt{R\gamma} \Rightarrow \frac{\rho}{\sqrt{R}} = \sqrt{\gamma}$, yielding the following precisions

$$\pi(\hat{b}) = \sqrt{n_b}, \quad \pi(\hat{a}) = \sqrt{n_b}\sqrt{\gamma}, \quad \pi(\hat{\omega}) = \frac{m_T\sqrt{n_b}\sqrt{\gamma}}{2\pi\sqrt{3}}.$$

There are two main conclusions. Firstly for all parameters precision increases only with $\sqrt{n_b}$. But for $\hat{\omega}$ precision also rises linearly with m_T. Thus we can get much greater precision for $\hat{\omega}$ than for the other parameters. We see also that if ρ is near 0 then the precision of $\hat{a}, \hat{\omega}$ can be very low.

We now illustrate these results while also investigating how 'quickly' the MLE achieves the asymptotic CRB by plotting $\hat{\pi}(\hat{\theta})/\pi(\hat{\theta})$ which should be 1 asymptotically.

We choose a grid of n_b, ρ, m_T values. We have one 'degree of freedom', so we fix $\omega = 3 \equiv$ period $= 2.1$. The grid was $8 \times 8 \times 8$ with ranges for $[n_b, m_T, \rho]$ of $[15^2 : 30^2; .5 : 2; .3 : .6]$.

We did $1,000$ repeats for each grid point. using the thinning algorithm [12] and estimated the precisions by $\hat{\pi}(\hat{\theta}) = mean(\hat{\theta})/\hat{se}(\hat{\theta})$.

We set initial values $\hat{a}_0 = \frac{N_T}{2T}, \hat{b}_0 = \frac{N_T}{T}$. However, the MLE algorithm is very sensitive to the initial frequency. By trial and error, we found choosing multiple initial frequencies $\hat{\omega}_0$ to work reliably. We do 10 MLEs with $\hat{\omega}_0 \in \{2.1, 2.3, \cdots, 3.7, 3.9\}$. We calculate the log-likelihoods of their corresponding estimates and pick the one that gives the greatest log-likelihood.

Another approach is to run the EM at each point of a grid of ωs. and find the $\hat{\omega}$ that maximizes the likelihood. This needs a much denser grid and so is much more computationally intensive.

5.2. Simulation results

As an example we first select one repeat with $[n_b, m_T, \rho] = [15^2, 2, 0.3]$. The corresponding parameters are $[b, a, \omega] = [53.7, 16.1, 3]$ and $T = 4.2$.

Fig. 1 shows parameter iterates with $\hat{\omega}_0 = 2.7$ yielding estimates $\hat{\theta} = [48.7, 19.5, 3.1]^\top$. For comparison $\hat{\omega}_o = 2.1$ or 3.9, yields $\hat{\theta} \approx [44, 4, 1.6]^\top$.

Fig. 2 shows the log-likelihoods of the 10 estimates. The greatest likelihood is marked.

We now plot heat maps of $\hat{\pi}/\pi$ in Fig. 3. Both colour extremes mean the asymptotic CRB has not been achieved. Note the powerful influence of m_T as the asymptotic CRB shows. The asymptotic/converged values are reached roughly as follows: for $\hat{b}$ at $m_T = 1.57$; for $\hat{\omega}$ at $m_T = 1.79$; for $\hat{\omega}$ at $m_T = 2$. Better convergence occurs for higher ρ and n_b as expected.

6. CONCLUSION

In this paper, we studied the widely applied time-varying Poisson model with a periodic intensity and developed two new results. Firstly we obtained an asymptotic Cramer-Rao bound. This confirmed that the frequency parameter can be estimated with super-resolution with standard error of order $\frac{1}{T^{3/2}}$. The other parameters are only estimated with the standard $\frac{1}{T^{1/2}}$ precision. Secondly we developed, apparently for the first time, an algorithm for finding the maximum likelihood estimators of the three parameters. The results were illustrated with simulations.

In future work we will tackle the much tougher problem of extensions to aperiodic models.

7. REFERENCES

[1] William Bialek, Rob de Ruyter van Steveninck, Fred Rieke, and David Warland, *Spikes: Exploring the Neural Code*, MIT press, 1997.

[2] Peter Tankov and Rama Cont, *Financial Modelling with Jump Processes*, Chapman and Hall/CRC, Boca Raton, 2004.

[3] M. Farajtabar, Y. Wang, MG. Rodriguez, S. Li, H. Zha, and L. Song, " Coevolve: A joint point process model for information diffusion and network co-evolution.," in *Proc NIPS*. ANIPS, 2015, p. 954–1962.

[4] Dawei Shi, Ling Shi, and Tongwen Chen, *Event-Based State Estimation: A Stochastic Perspective*, Studies in Systems, Decision and Control. Springer, New York, 2016.

[5] D. Snyder and M. Miller, *Random Point Processes in Time and Space*, Springer-Verlag, New York, 1991.

[6] Nan Shao and Keh-Shin Lii, "Modelling non-homogeneous poisson processes with almost periodic intensity functions," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 73, no. 1, pp. 99–122, 2011.

[7] Rodrigo Saul Gaitan and Keh-Shin Lii, "On the estimation of periodicity or almost periodicity in inhomogeneous gamma point-process data," *Journal of Time Series Analysis*, vol. n/a, no. n/a, 2021.

[8] D. Vere-Jones, "On the estimation of frequency in point-process data," *Journal of Applied Probability*, vol. 19, pp. 383–394, 1982.

[9] Yu A. Kutoyants, *Statistical Inference for Spatial Poisson Processes*, Lecture notes in statistics. Springer-Verlag, New York, 1 edition, 1998.

[10] BI. Godoy, V. Solo, and SA. Pasha, "Truncated hawkes point process modeling: System theory and system identification," *Automatica*, vol. 113, pp. 16pp, 2020.

[11] Roy L. Streit, *Poisson Point Processes: Imaging, Tracking, and Sensing*, Springer, Boston, 2010.

[12] P. A. W Lewis and G. S. Shedler, "Simulation of nonhomogeneous poisson processes by thinning," *Naval Research Logistics Quarterly*, vol. 26, no. 3, pp. 403–413, 1979.