The correct asymptotic variance for the sample mean of a homogeneous Poisson Marked Point Process

William Garner and Dimitris N. Politis

Department of Mathematics

University of California at San Diego

La Jolla, CA 92093, USA

Emails: wgarner@ucsd.edu; dpolitis@ucsd.edu

Abstract

The asymptotic variance of the sample mean of a homogeneous Poisson marked point process has been studied in the literature but confusion has arisen as to the correct expression due to some technical intricacies. This note sets the record straight with regards the variance of the sample mean. In addition, a central limit theorem in the general d -dimensional case is also established.

Math. Subject Classification: 60F05;60K99

Keywords: Central Limit Theorem; Marked Point Process.

1 Introduction

Let $\{X(t) \text{ for } t \in \mathbf{R}^d\}$ be a strictly stationary random field in the continuous, d -dimensional parameter t ; denote $\mu = EX(t)$ and $R(t) = Cov(X(s), X(s+t)) = Cov(X(0), X(t))$ which is assumed finite. Also let $\{N(t) \text{ for } t \in \mathbf{R}^d\}$ be a homogeneous Poisson point process with rate λ . The point process $\{N(t)\}$ is assumed to be independent of the random field $\{X(t)\}$. Let $\tau_1, \tau_2, \dots, \tau_{N(K)}$ denote the points generated by $\{N(t)\}$ inside the set K .

The observation region K will be assumed to be a compact, convex subset of $\mathbf{R}^d$. Let

$|K|$ denote the volume of K , and $\text{diam}(K)$ the supremum of the diameters of all l_∞ balls contained in K ; so if K is a rectangle, $\text{diam}(K)$ is its smallest dimension. All asymptotic results to be discussed in this paper will be taken under the condition $\text{diam}(K) \rightarrow \infty$. To avoid possible pitfalls, we will also assume that the observation region K expands in a “nested” way as $\text{diam}(K) \rightarrow \infty$, i.e., that $\text{diam}(K) < \text{diam}(K')$ has as a necessary implication that $K \subset K'$.

The pairs $(X(\tau_i), N(\tau_i))$ for $i = 1, \dots, N(K)$ constitute the data from a homogeneous *Marked Point Process*. $N(K)$ denotes the number of available observations; denote $\Lambda(K) = E[N(K)] = \lambda|K|$. Consider the the problem of estimation of the mean μ based on the above marked point process data. A natural estimator is

$$\bar{X}_K = \frac{1}{N(K)} \int_K X(t) N(dt) = \frac{1}{N(K)} \sum_{i=1}^{N(K)} X(\tau_i) \quad (1)$$

which is nothing other than the sample mean of the available X data, the so-called ‘marks’. A simple conditioning (on $N(K)$) argument shows that $\bar{X}_K$ is unbiased for μ , i.e., $E\bar{X}_K = \mu$. The issue to be resolved in this note is a correct expression for the asymptotic variance of $\bar{X}_K$; such an expression is needed for the construction of confidence intervals and hypothesis tests for the parameter μ . We also establish a central limit theorem for the sample mean in the general d -dimensional case under weak assumptions.

2 A critical review of existing results

The setup of data from a Marked Point Process is still at the forefront of active research; for example, see Ballani, Kabluchko, and Schlather (2012) and the references therein. Nevertheless, the subject has been under investigation for several decades; see Masry (1983) or Kutoyants (1984a,b). Going even further back, the pioneering paper of Brillinger (1973) provided an in-depth study of the asymptotic distribution of $\bar{X}_K$ in the one-dimensional case ($d = 1$). Under a condition of (absolute) summability of all cumulants of the random field $\{X(t)\}$, Brillinger (1973) showed that $\sqrt{|K|}(\bar{X}_K - \mu)$ is asymptotically Normal with mean zero and variance given by

$$\int_{-\infty}^{\infty} R(u)du + \frac{R(0)}{\lambda}. \quad (2)$$

Nevertheless, the assumption of summability of all cumulants can be quite restrictive. For example, it excludes all random fields that have a certain degree of heavy tails, e.g., those that do not have all moments finite. Brillinger (1973) mentions that assumptions involving mixing coefficients could be used instead.

This more general approach was undertaken by Karr (1986) in the d -dimensional case. In order to prove his results, Karr (1986) introduced the auxiliary random variable

$$\tilde{X}_K = \frac{1}{\Lambda(K)} \int_K X(t)N(dt). \quad (3)$$

Recall that $\Lambda(K) = \lambda|K|$; thus, $\tilde{X}_K$ is not a *bona fide* estimator as it depends on the unknown rate λ . However, $\tilde{X}_K$ is easier to work with since it is devoid of the random denominator inherent in $\bar{X}_K$.

Karr (1986) worked under the minimal assumptions that

$$\int_{\mathbf{R}^d} |R(t)|dt < \infty. \quad (4)$$

and

$$\frac{1}{\sqrt{|K|}} \int_K (X(t) - \mu)dt \xrightarrow{\mathcal{L}} N\left(0, \int_{\mathbf{R}^d} R(t)dt\right). \quad (5)$$

In view of the $\int R(t)dt$ term appearing in (2), condition (4) is a *sine qua non*. Furthermore, condition (5) follows immediately from assuming $E|X(t)|^{2+\delta} < \infty$ for some $\delta > 0$ and a corresponding mixing condition; see Ivanov and Leonenko (1986) or Lemma 2 of Politis, Paparoditis and Romano (1999) for examples of such mixing conditions.

Theorem 2.1 [Karr (1986)] *Assume equations (4) and (5). Then, as $\text{diam}(K) \rightarrow \infty$, $\sqrt{|K|}(\tilde{X}_K - \mu) \xrightarrow{\mathcal{L}} N(0, \sigma^2)$ where $\sigma^2 = \int_{\mathbf{R}^d} R(y)dy + \frac{R(0) + \mu^2}{\lambda}$.*

In trying to render $\tilde{X}_K$ as a usable statistic, one may plug-in $N(K)$ as an estimator of the unknown $\Lambda(K)$ in expression (3). Not surprisingly, this operation just leads to the sample mean $\bar{X}_K$. Based on the fact that $N(K)/\Lambda(K) \xrightarrow{a.s.} 1$, Karr (1986) further claimed the following (incorrect) lemma:

Lemma 2.1 [Karr (1986)—incorrect.] *Assume equations (4) and (5). Then, as $\text{diam}(K) \rightarrow \infty$, the random variables $\tilde{X}_K$ and $\bar{X}_K$ are asymptotically equivalent, i.e., $\sqrt{|K|}(\tilde{X}_K - \bar{X}_K) = o_P(1)$.*

Based on the incorrect Lemma 2.1, Karr (1986) claimed the following as a corollary.

Theorem 2.2 [Karr (1986)—incorrect.] *Assume equations (4) and (5). Then, as $\text{diam}(K) \rightarrow \infty$, $\sqrt{|K|}(\bar{X}_K - \mu) \xrightarrow{\mathcal{L}} N(0, \sigma^2)$ where $\sigma^2 = \int_{\mathbf{R}^d} R(y)dy + \frac{R(0) + \mu^2}{\lambda}$.*

The discrepancy between Theorem 2.2 and Brillinger's eq. (2) went unnoticed for a long time, and the monograph by Karr (1991) did not shed any more light on the issue. Based on invariance considerations, Politis et al. (1999) suggested that the correct result is

$$\sqrt{|K|}(\bar{X}_K - \mu) \xrightarrow{\mathcal{L}} N(0, \theta^2) \quad \text{where} \quad \theta^2 = \int_{\mathbf{R}^d} R(y)dy + \frac{R(0)}{\lambda} \quad (6)$$

attributing the mistake in Theorem 2.2 to a typo. Eq. (6) is indeed the correct one as will be shown in the next section.

Note, however, that Politis et al. (1999) were not aware of the error in Lemma 2.1, and subsequently stated the incorrect fact that the auxiliary variable $\tilde{X}_K$ has the same asymptotic distribution as given in (6). Fortunately, as far as the estimator of interest is concerned, i.e., the sample mean $\bar{X}_K$, all asymptotic results of Politis et al. (1999) are correct as stated both in the real world, e.g., eq. (6), as well as in the bootstrap world that they also study.

3 A correct expression for the asymptotic variance of the sample mean and a Central Limit Theorem

A correct version of Lemma 2.1 goes as follows:

Lemma 3.1 *Assume equations (4) and (5). Then, as $\text{diam}(K) \rightarrow \infty$, it is true that:*

- (i) *If $\mu = 0$, then $\sqrt{|K|}(\tilde{X}_K - \bar{X}_K) = o_P(1)$.*
- (ii) *If $\mu \neq 0$, then $\sqrt{|K|}(\tilde{X}_K - \bar{X}_K) \xrightarrow{\mathcal{L}} N(0, \mu^2)$.*

Proof. (i) Note that

$$\sqrt{|K|}(\tilde{X}_K - \bar{X}_K) = \frac{\sqrt{|K|}}{\Lambda(K)} \int_K X(t) N(dt) \left(1 - \frac{\Lambda(K)}{N(K)}\right).$$

But from Theorem 2.1 it follows that $\sqrt{|K|}\tilde{X}_K \xrightarrow{\mathcal{L}} N(0, \sigma^2)$, so that $\frac{\sqrt{|K|}}{\Lambda(K)} \int_K X(t) N(dt) = O_P(1)$. Since $N(K)/\Lambda(K) \xrightarrow{a.s.} 1$, we have $\left(1 - \frac{\Lambda(K)}{N(K)}\right) = o_P(1)$, and the result follows.

(ii) Let $Y(t) = X(t) - \mu$, and hence $EY(t) = 0$. Notice that

$$\tilde{X}_K = \frac{1}{\Lambda(K)} \int_K Y(t) N(dt) + \mu \frac{N(K)}{\Lambda(K)} \quad (7)$$

and

$$\bar{X}_K = \frac{1}{N(K)} \int_K Y(t) N(dt) + \mu. \quad (8)$$

Using part (i) in connection with the mean zero random field $Y(t)$ gives

$$\sqrt{|K|}(\tilde{X}_K - \bar{X}_K) = o_P(1) + \sqrt{|K|} \mu \left(\frac{N(K)}{\Lambda(K)} - 1 \right). \quad (9)$$

Eq. (9) together with Slutsky's Theorem and the asymptotic normality of the Poisson complete the proof of part (ii). $\square$

We are now ready to state and prove a correct Central Limit Theorem for the sample mean $\bar{X}_K$. This is apparently novel in the literature, and represents an extension over

Brillinger's (1973) result of eq. (2) in two ways: relaxing the conditions of summability of all cumulants, and addressing the setup of a marked point process in d -dimensions, i.e., having $t \in \mathbf{R}^d$.

Theorem 3.1 *Assume equations (4) and (5). Then, as $\text{diam}(K) \rightarrow \infty$, eq. (6) holds true, i.e.,*

$$\sqrt{|K|} (\bar{X}_K - \mu) \xrightarrow{\mathcal{L}} N(0, \theta^2) \quad \text{where} \quad \theta^2 = \int_{\mathbf{R}^d} R(y) dy + \frac{R(0)}{\lambda}.$$

Proof. As in the proof of Lemma 3.1, we let $Y(t) = X(t) - \mu$, and also define

$$\tilde{Y}_K = \frac{1}{\Lambda(K)} \int_K Y(t) N(dt) \quad \text{and} \quad \bar{Y}_K = \frac{1}{N(K)} \int_K Y(t) N(dt).$$

Noting that $Y(t)$ has mean zero but the same covariance structure as $X(t)$, the (correct) Theorem 2.1 implies that $\sqrt{|K|} \tilde{Y}_K \xrightarrow{\mathcal{L}} N(0, \theta^2)$. But part (i) of Lemma 3.1 as applied to $Y(t)$ implies $\sqrt{|K|} \bar{Y}_K \xrightarrow{\mathcal{L}} N(0, \theta^2)$. Finally, eq. (8) implies that $\bar{Y}_K = \bar{X}_K - \mu$, and the proof is completed. $\square$.

Acknowledgement. Many thanks are due to Professors Ian Abramson and Jason Schweinsberg for helpful discussions.

References

- [1] Ballani, F., Kabluchko, Z., and Schlather, M. (2012). Random Marked Sets, *Adv. Appl. Prob.*, vol. 44, pp. 603–616.
- [2] Brillinger, D.R. (1973), Estimation of the mean of a stationary time series by sampling, *J. Appl. Prob.*, vol. 10, pp. 419–431.
- [3] Ivanov, A.V. and Leonenko, N.N. (1986). *Statistical Analysis of Random Fields*, Kluwer Publishers, Dordrecht.

- [4] Karr, A.F. (1986). Inference for stationary random fields given Poisson samples, *Adv. Appl. Prob.*, vol. 18, pp. 406-422.
- [5] Karr, A.F. (1991). *Point Processes and their Statistical Inference*, 2nd Ed., Marcel Dekker, New York.
- [6] Kutoyants, Yu. A. (1984a). On nonparametric estimation of intensity function of inhomogeneous Poisson process. *Problems Control Inform. Theory*, vol. 13, no. 4, 253-258.
- [7] Kutoyants, Yu. A. (1984b). *Parameter estimation for stochastic processes*, (Translated from the Russian and edited by B.L.S. Prakasa Rao), Research and Exposition in Mathematics, Vol. 6, Heldermann Verlag, Berlin.
- [8] Masry, E. (1983). Nonparametric covariance estimation from irregularly spaced data, *Adv. Appl. Prob.*, vol. 15, pp. 113-132.
- [9] Politis, D.N., Paparoditis, E. and Romano, J.P. (1999). Resampling Marked Point Processes, in *Multivariate Analysis, Design of Experiments, and Survey Sampling*, Subir Ghosh (Ed.), Statist. Textbooks Monogr., vol. 159, Marcel Dekker, New York, pp. 163-185.