J. Appl. Prob. **50,** 889–892 (2013)
Printed in England

THE CORRECT ASYMPTOTIC VARIANCE FOR THE SAMPLE MEAN OF A HOMOGENEOUS POISSON MARKED POINT PROCESS

WILLIAM GARNER * ** AND
DIMITRIS N. POLITIS,* *** *University of California at San Diego*

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 to the variance of the sample mean. In addition, a central limit theorem in the general d-dimensional case is also established.

Keywords: Central limit theorem; marked point process

2010 Mathematics Subject Classification: Primary 60F05; 60K99

1. Introduction

Let $\{X(t) \text{ for } t \in \mathbb{R}^d\}$ be a strictly stationary random field in the continuous, d-dimensional parameter t; define $\mu = \mathbb{E}X(t)$ and $R(t) = \operatorname{cov}(X(s), X(s+t)) = \operatorname{cov}(X(0), X(t))$, which is assumed finite. Also, let $\{N(t) \text{ for } t \in \mathbb{R}^d\}$ 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 $\mathbb{R}^d$. Let $|K|$ denote the volume of K, and let $\operatorname{diam}(K)$ denote the supremum of the diameters of all l_∞ balls contained in K; so if K is a rectangle, $\operatorname{diam}(K)$ is its smallest dimension. All asymptotic results to be discussed in this paper will be taken under the condition $\operatorname{diam}(K) \to \infty$. To avoid possible pitfalls, we will also assume that the observation region K expands in a *nested* way as $\operatorname{diam}(K) \to \infty$, i.e. that $\operatorname{diam}(K) < \operatorname{diam}(K')$ 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*. Here $N(K)$ denotes the number of available observations; define $\Lambda(K) = \mathbb{E}[N(K)] = \lambda|K|$. Consider the the problem of estimation of the mean μ based on the above marked point process data. A natural estimator is

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

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. $\mathbb{E}\bar{X}_K = \mu$. The issue to be resolved in this note is a correct expression for the asymptotic variance of $\bar{X}_K$;

Received 17 May 2012; revision received 28 November 2012.
* Postal address: Department of Mathematics, University of California at San Diego, La Jolla, CA 92093, USA.
** Email address: wgarner@ucsd.edu
*** Email address: dpolitis@ucsd.edu

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; see, for example, Ballani *et al.* (2012) and the references therein. Nevertheless, the subject has been under investigation for several decades; see Masry (1983) or Kutoyants (1984a), (1984b). Going even further back, the pioneering paper of Brillinger (1973) provided an indepth 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 0 and variance given by

$$\int_{-\infty}^{\infty} R(u)\,\mathrm{d}u + \frac{R(0)}{\lambda}. \tag{2.1}$$

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) mentioned 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

$$\tilde{X}_K = \frac{1}{\Lambda(K)} \int_K X(t) N(\mathrm{d}t). \tag{2.2}$$

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

$$\int_{\mathbb{R}^d} |R(t)|\,\mathrm{d}t < \infty \tag{2.3}$$

and

$$\frac{1}{\sqrt{|K|}} \int_K (X(t) - \mu)\,\mathrm{d}t \xrightarrow{\mathcal{L}} N\biggl(0, \int_{\mathbb{R}^d} R(t)\,\mathrm{d}t\biggr). \tag{2.4}$$

In view of the $\int R(t)\,\mathrm{d}t$ term appearing in (2.1), condition (2.3) is a *sine qua non*. Furthermore, condition (2.4) follows immediately from assuming that $\mathbb{E}|X(t)|^{2+\delta} < \infty$ for some $\delta > 0$ and a corresponding mixing condition; see Ivanov and Leonenko (1986) or Lemma 2 of Politis *et al.* (1999) for examples of such mixing conditions.

Theorem 2.1. (Karr (1986).) *Assume that (2.3) and (2.4) hold. Then, as* $\operatorname{diam}(K) \to \infty$, $\sqrt{|K|}(\tilde{X}_K - \mu) \xrightarrow{\mathcal{L}} N(0, \sigma^2)$, *where* $\sigma^2 = \int_{\mathbb{R}^d} R(y)\,\mathrm{d}y + (R(0) + \mu^2)/\lambda$.

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 (2.2). Not surprisingly, this operation just leads to the sample mean $\bar{X}_K$. Based on the fact that $N(K)/\Lambda(K) \xrightarrow{\text{a.s.}} 1$, Karr (1986) further claimed the following (incorrect) lemma.

Lemma 2.1. (Karr (1986)—incorrect.) *Assume that (2.3) and (2.4) hold. Then, as* $\operatorname{diam}(K) \to \infty$*, the random variables* $\tilde{X}_K$ *and* $\bar{X}_K$ *are asymptotically equivalent, i.e.* $\sqrt{|K|}(\tilde{X}_K - \bar{X}_K) = o_{\mathbb{P}}(1)$.

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

Theorem 2.2. (Karr (1986)—incorrect.) *Assume that (2.3) and (2.4) hold. Then, as* $\operatorname{diam}(K) \to \infty$*,* $\sqrt{|K|}(\bar{X}_K - \mu) \xrightarrow{\mathcal{L}} N(0, \sigma^2)$*, where* $\sigma^2 = \int_{\mathbb{R}^d} R(y)\,\mathrm{d}y + (R(0) + \mu^2)/\lambda$.

The discrepancy between Theorem 2.2 and Brillinger's equation (2.1) 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

$$\sqrt{|K|}(\bar{X}_K - \mu) \xrightarrow{\mathcal{L}} N(0, \theta^2), \quad \text{where} \quad \theta^2 = \int_{\mathbb{R}^d} R(y)\,\mathrm{d}y + \frac{R(0)}{\lambda}, \tag{2.5}$$

attributing the mistake in Theorem 2.2 to a typo. Equation (2.5) is indeed the correct equation 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 (2.5). 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. (2.5), as well as in the bootstrap world that they also studied.

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 that (2.3) and (2.4) hold. Then, as* $\operatorname{diam}(K) \to \infty$*, it is true that*

(i) *if* $\mu = 0$ *then* $\sqrt{|K|}(\tilde{X}_K - \bar{X}_K) = o_{\mathbb{P}}(1)$*,*

(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

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

But, from Theorem 2.1, it follows that $\sqrt{|K|}\tilde{X}_K \xrightarrow{\mathcal{L}} N(0, \sigma^2)$, so that

$$\frac{\sqrt{|K|}}{\Lambda(K)} \int_K X(t) N(\mathrm{d}t) = O_{\mathbb{P}}(1).$$

Since $N(K)/\Lambda(K) \xrightarrow{\text{a.s.}} 1$, we have $(1 - \Lambda(K)/N(K)) = o_{\mathbb{P}}(1)$, and the result follows.

(ii) Let $Y(t) = X(t) - \mu$, and, hence, $\mathbb{E}Y(t) = 0$. Note that

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

and

$$\bar{X}_K = \frac{1}{N(K)} \int_K Y(t) N(\mathrm{d}t) + \mu. \tag{3.1}$$

https://doi.org/10.1239/jap/1378401244 Published online by Cambridge University Press

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

$$\sqrt{|K|}(\tilde{X}_K - \bar{X}_K) = o_{\mathbb{P}}(1) + \sqrt{|K|}\,\mu\left(\frac{N(K)}{\Lambda(K)} - 1\right). \tag{3.2}$$

Equation (3.2) together with Slutsky's theorem and the asymptotic normality of the Poisson completes the proof of part (ii).

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 (2.1) 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 $t \in \mathbb{R}^d$.

Theorem 3.1. *Assume that (2.3) and (2.4) hold. Then, as* $\operatorname{diam}(K) \to \infty$*, (2.5) holds, i.e.*

$$\sqrt{|K|}(\bar{X}_K - \mu) \stackrel{\mathcal{L}}{\Longrightarrow} N(0, \theta^2), \quad \textit{where} \quad \theta^2 = \int_{\mathbb{R}^d} R(y)\,\mathrm{d}y + \frac{R(0)}{\lambda}.$$

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

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

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

Acknowledgements

Many thanks are due to Professors Ian Abramson and Jason Schweinsberg for helpful discussions. The research of the second author was partially supported by the NSF grant DMS-11-20888.

References

BALLANI, F., KABLUCHKO, Z. AND SCHLATHER, M. (2012). Random marked sets. *Adv. Appl. Prob.* **44,** 603–616.

BRILLINGER, D. R. (1973). Estimation of the mean of a stationary time series by sampling. *J. Appl. Prob.* **10,** 419–431.

IVANOV, A. V. AND LEONENKO, N. N. (1986). *Statistical Analysis of Random Fields*. Kluwer Publishers, Dordrecht.

KARR, A. F. (1986). Inference for stationary random fields given Poisson samples. *Adv. Appl. Prob.* **18,** 406–422.

KARR, A. F. (1991). *Point Processes and Their Statistical Inference*, 2nd edn. Marcel Dekker, New York.

KUTOYANTS, Y. A. (1984a). On nonparametric estimation of intensity function of inhomogeneous Poisson process. *Problems Control Inf. Theory* **13,** 253–258.

KUTOYANTS, Y. A. (1984b). *Parameter Estimation for Stochastic Processes* (Res. Exposition Math. **6**). Heldermann, Berlin.

MASRY, E. (1983). Nonparametric covariance estimation from irregularly-spaced data. *Adv. Appl. Prob.* **15,** 113–132.

POLITIS, D. N., PAPARODITIS, E. AND ROMANO, J. P. (1999). Resampling marked point processes. In *Multivariate Analysis, Design of Experiments, and Survey Sampling* (Statist. Textbooks Monogr. **159**), ed. Subir Ghosh, Marcel Dekker, New York, pp. 163–185.