CENTRE FOR **STOCHASTIC GEOMETRY**
AND ADVANCED **BIOIMAGING**

Christophe A.N. Biscio, Arnaud Poinas and Rasmus Waagepetersen

A note on gaps in proofs of central limit theorems

No. 12, November 2017

A note on gaps in proofs of central limit theorems

Christophe A.N. Biscio[1], Arnaud Poinas[2] and Rasmus Waagepetersen[1]

[1]Department of Mathematical Sciences, Aalborg University, Denmark
[2]IRMAR, Campus de Beaulieu, Bat. 22/23, France

Abstract

We fill two gaps in the literature on central limit theorems. First we state and prove a generalization of the Cramér-Wold device which is useful for establishing multivariate central limit theorems without the need for assuming the existence of a limiting covariance matrix. Second we extend and provide a detailed proof of a very useful result for establishing univariate central limit theorems.

Keywords: central limit theorem, Cramér-Wold device, random field.

1 Introduction

Many applications of spatial statistics involve spatially varying covariates. One common example is universal kriging (e.g. Chilès and Delfiner, 2008), where a deterministic component of a spatial variable is modelled using a regression on spatial covariates. Another example, which we will discuss in some detail for illustrative purpose, is log-linear modelling of intensity functions for spatial point processes, see e.g. Rathbun and Cressie (1994), Rathbun (1996), Schoenberg (2005), Waagepetersen (2007), Guan and Loh (2007). Such models have for example found much use in ecology where point processes are used to model locations of plants and animals, and the covariates could describe landscape type, topography or soil properties.

Letting $Z = \{Z(u)\}_{u\in\mathbb{R}^d}$ where for each $u \in \mathbb{R}^d$, $Z(u) \in \mathbb{R}^p$ is a covariate vector, the intensity function is often assumed to be of the form

$$\rho(u;\beta) = \exp(\beta^T Z(u)) \tag{1.1}$$

where $\beta \in \mathbb{R}^p$. In the aforementioned references, the regression parameter β is inferred using an estimating function given by the score of the log likelihood function of a Poisson process. For deriving asymptotic results it is crucial to establish asymptotic normality of the estimating function. Often increasing domain asymptotics are used where a sequence $\{W_n\}_{n\geq 1}$ of bounded but increasing observation windows $W_n \subset \mathbb{R}^d$ are considered. The score estimating function is then

$$e_n(\beta) = \sum_{u\in X\cap W_n} Z(u) - \int_{W_n} Z(u)\rho(u;\beta)\mathrm{d}u \tag{1.2}$$

where X denotes the spatial point process. An estimate of β is obtained by solving $e_n(\beta) = 0$.

In spatial statistics in general, central limit theorems for α-mixing spatial processes are very popular tools for establishing asymptotic results both for random field and point process models. In case of the model (1.1), asymptotic properties of estimators of β were for example established by central limit theorems for α-mixing spatial processes in Guan and Loh (2007) and Waagepetersen and Guan (2009). Bolthausen (1982) provided a much cited central limit theorem for stationary α-mixing random fields. This result was later extended to the non-stationary case by Guyon (1995). Karácsony (2006) extended the result further to triangular arrays considering a combination of infill and increasing domain asymptotics. When taking a deeper look at the techniques employed in the proofs of the aforementioned references, a couple of issues may disturb the reader. These are addressed in Sections 2–3 below.

2 Cramér-Wold device

Bolthausen (1982), Guyon (1995) and Karácsony (2006) only gave detailed proofs of central limit theorems in the univariate case. Guyon (1995) and Karácsony (2006) also stated multivariate central limit theorems and just referred to the Cramér-Wold device for extending the univariate central limit theorems to the multivariate case. However, a closer look shows that a simple application of the Cramér-Wold device does not always suffice. Suppose that $\{X_n\}_{n\geq 1}$ is a sequence of random vectors in $\mathbb{R}^p$. The Cramér-Wold device (e.g. p. 383 in Billingsley, 1995) then says that X_n converges in distribution to a random vector X if and only if, for all $a \in \mathbb{R}^p$, $a^T X_n$ converges in distribution toward $a^T X$. In statistical applications we often want to show that $\mathrm{Var}(X_n)^{-1/2} X_n$ converges in distribution toward $\mathcal{N}(0, I_p)$ as n goes to infinity for some sequence of statistics X_n. If $\mathrm{Var}(X_n)$ converges to a fixed positive definite matrix Σ then this is equivalent to showing that X_n converges in distribution toward $\mathcal{N}(0, \Sigma)$ and the application of the Cramér-Wold device is trivial.

However, in many applications a limit for $\mathrm{Var}(X_n)$ does not exist. Returning again to the problem of inferring the intensity function (1.1) and letting $X_n = e_n(\beta)/|W_n|^{1/2}$ be given by (1.2) normalized by $|W_n|^{1/2}$, the covariance matrix of X_n is

$$\Sigma_n = \frac{1}{|W_n|}\left(\int_{W_n} Z(u)^T Z(u)\rho(u;\beta)\mathrm{d}u + \int_{W_n}\int_{W_n} Z(u)^T Z(v)[g(u,v)-1]\mathrm{d}u\mathrm{d}v\right)$$

where $g(\,\cdot\,,\,\cdot\,)$ is the so-called pair correlation function (e.g. Møller and Waagepetersen, 2004). Due to the dependency on $Z(u)$, $u \in W_n$, it is not reasonable to assume that Σ_n converges to a fixed limit. Guyon (1995) and Karácsony (2006) just refer to the simple application of the Cramér-Wold device and in particular do not discuss the issue of whether a limiting covariance matrix exists or not. To be on firm ground, a Cramér-Wold type result covering the case with no limiting covariance matrix seems missing. Therefore, we state the following generalization of the Cramér-Wold device.

Lemma 2.1. *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables in $\mathbb{R}^p$ such that*

$$0 < \liminf_{n\to\infty} \lambda_{\min}(\mathrm{Var}(X_n)) < \limsup_{n\to\infty} \lambda_{\max}(\mathrm{Var}(X_n)) < \infty,$$

where for a symmetric matrix M, $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$ denote the minimal and maximal eigenvalues of M.

Then, $\mathrm{Var}(X_n)^{-1/2}X_n \xrightarrow[n\to\infty]{distr.} \mathcal{N}(0, I_p)$ *if for all $a \in \mathbb{R}^p$,*

$$\left(a^T \mathrm{Var}(X_n)a\right)^{-\frac{1}{2}} a^T X_n \xrightarrow[n\to\infty]{distr.} \mathcal{N}(0,1).$$

The condition in this lemma regarding $\liminf_{n\to\infty} \lambda_{\min} \mathrm{Var}(X_n)$ is the kind of condition used in the central limit theorems of Guyon (1995) and Karácsony (2006) and so is not restrictive in practice. A proof of this lemma is provided in Section 4.

3 A lemma by Bolthausen

Bolthausen (1982), Guyon (1995) and Karácsony (2006) all use the following key lemma.

Lemma 3.1. *For $n \in \mathbb{N}$, let $\{X_n\}_{n\in\mathbb{N}}$ be random variables such that* $\sup_{n\in\mathbb{N}} \mathbb{E}(X_n^2) < \infty$ *and for all $t \in \mathbb{R}$,*

$$\mathbb{E}[(it - X_n)e^{itX_n}] \xrightarrow[n\to\infty]{} 0.$$

Then $X_n \xrightarrow[n\to\infty]{distr.} \mathcal{N}(0,1)$.

Karácsony (2006) somewhat misleadingly coins this “Stein’s lemma” and refers to Stein (1972) and Guyon (1995). Guyon (1995) in turn refers to Stein (1972). However, the original reference is Bolthausen (1982) who states and proves the lemma while acknowledging inspiration from Stein (1972). Stein’s lemma (Stein, 1981) says that if Z is standard normal then $\mathbb{E}[f'(Z) - Zf(Z)] = 0$ for any differentiable f with $\mathbb{E}|f'(Z)| < \infty$. This result is related to but nevertheless different from Lemma 3.1. We do not find Bolthausen (1982)’s very condensed proof easily accessible and believe it is useful to provide a more detailed proof of this crucial lemma. Moreover, the conclusion of Lemma 3.1 holds under a weaker assumption as stated below.

Lemma 3.2. *For $n \in \mathbb{N}$, let $\{X_n\}_{n\in\mathbb{N}}$ be uniformly integrable random variables such that for all $t \in \mathbb{R}$,*

$$\mathbb{E}[(it - X_n)e^{itX_n}] \xrightarrow[n\to\infty]{} 0. \tag{3.1}$$

Then $X_n \xrightarrow[n\to\infty]{distr.} \mathcal{N}(0,1)$.

A proof of the last more general lemma is presented in Section 5.

4 Proof of Lemma 2.1

Assume by contradiction that $\mathrm{Var}(X_n)^{-1/2}X_n \xrightarrow[n\to\infty]{distr.} Y \sim \mathcal{N}(0, I_p)$ does not hold. Then there exists a bounded and continuous function f such that

$$\mathbb{E}f(\mathrm{Var}(X_n)^{-1/2}X_n) - \mathbb{E}f(Y)$$

does not converge toward 0. Thus, there exists $\epsilon > 0$ and a strictly increasing function $b : \mathbb{N} \to \mathbb{N}$ such that for all $n \in \mathbb{N}$,

$$\left|\mathbb{E}f(\mathrm{Var}\,(X_{b(n)})^{-1/2}X_{b(n)}) - \mathbb{E}f(Y)\right| > \epsilon. \tag{4.1}$$

Further, since for all $a \in \mathbb{R}^p$, $(a^T \mathrm{Var}(X_n)a)^{-\frac{1}{2}}a^T X_n \xrightarrow[n\to\infty]{distr.} \mathcal{N}(0, 1)$, we also have

$$\left(a^T \mathrm{Var}(X_{b(n)})a\right)^{-\frac{1}{2}} a^T X_{b(n)} \xrightarrow[n\to\infty]{distr.} \mathcal{N}(0, 1)$$

for any $a \in \mathbb{R}^p$. Since the sequence of eigenvalues of the matrices $\mathrm{Var}(X_n)$ is bounded, there exists by the Bolzano-Weierstrass theorem a strictly increasing function $c : \mathbb{N} \to \mathbb{N}$ such that $\{c(n)\}_{n\in\mathbb{N}} \subset \{b(n)\}_{n\in\mathbb{N}}$, and a matrix Σ such that

$$\mathrm{Var}\,(X_{c(n)}) \xrightarrow[n\to\infty]{} \Sigma.$$

Since

$$\left(a^T \mathrm{Var}\,(X_{c(n)})a\right)^{-\frac{1}{2}} a^T X_{c(n)} \xrightarrow[n\to\infty]{distr.} \mathcal{N}(0, 1)$$

we obtain by multiplying with $a^T \mathrm{Var}\,(X_{c(n)})a$ and using Slutsky's lemma, that

$$a^T X_{c(n)} \xrightarrow[n\to\infty]{distr.} a^T \mathcal{N}(0, \Sigma).$$

Hence, by the Cramér-Wold device,

$$X_{c(n)} \xrightarrow[n\to\infty]{distr.} \mathcal{N}(0, \Sigma)$$

which is equivalent to

$$\Sigma^{-\frac{1}{2}} X_{c(n)} \xrightarrow[n\to\infty]{distr.} \mathcal{N}(0, I_p).$$

Therefore, by Slutsky's lemma,

$$\mathrm{Var}\,(X_{c(n)})^{-\frac{1}{2}} X_{c(n)} \xrightarrow[n\to\infty]{distr.} \mathcal{N}(0, I_p).$$

From this we can conclude

$$\left|\mathbb{E}f(\mathrm{Var}\,(X_{c(n)})^{-1/2}X_{c(n)}) - \mathbb{E}f(Y)\right| \le \epsilon$$

for n large enough which contradicts (4.1).

5 Proof of Lemma 3.2

Since the X_n are uniformly integrable, we have by (25.11) in Billingsley (1995),

$$\sup_{n\in\mathbb{N}} \mathbb{E}|X_n| < \infty. \tag{5.1}$$

By Markov's inequality, for all $\epsilon > 0$,

$$P(|X_n| > \epsilon) \leq \frac{\mathbb{E}|X_n|}{\epsilon} \leq \frac{\sup_{n\in\mathbb{N}} \mathbb{E}|X_n|}{\epsilon}. \tag{5.2}$$

By (5.2), it follows that the sequence $\{X_n\}_{n\in\mathbb{N}}$ is tight. Suppose now that $b : \mathbb{N} \to \mathbb{N}$ is an increasing function and X a random variable such that $X_{b(n)} \xrightarrow[n\to\infty]{distr.} X$. Then, by Theorem 25.11 in Billingsley (1995) and (5.1),

$$\mathbb{E}|X| \leq \liminf_{n\to\infty} \mathbb{E}|X_{b(n)}| < \infty$$

so that $E(X)$ and $E(Xe^{itX})$, for $t \in \mathbb{R}$, are well defined. Since the random variables X_n are uniformly integrable, so are $X_{b(n)}$ and $X_{b(n)}e^{itX_{b(n)}}$, for all $t \in \mathbb{R}$. Thus, by Theorem 25.12 in Billingsley (1995),

$$\lim_{n\to\infty} \mathbb{E}(X_{b(n)}) = \mathbb{E}(X) \quad \text{and} \quad \lim_{n\to\infty} \mathbb{E}(X_{b(n)}e^{itX_{b(n)}}) = \mathbb{E}(Xe^{itX}).$$

Then, by (3.1),

$$\mathbb{E}[(it - X)e^{itX}] = 0$$

which may be written

$$t\phi_X(t) + \frac{\partial}{\partial t}\phi_X(t) = 0 \tag{5.3}$$

where ϕ_X denotes the characteristic function of X. For all $t \in \mathbb{R}$, we may multiply each side of (5.3) by $e^{t^2/2}$ so that (5.3) is equivalent to

$$\frac{\partial}{\partial t}(e^{t^2/2}\phi_X(t)) = 0$$

which, since ϕ_X is a characteristic function, has the unique solution $\phi_X(t) = e^{-t^2/2}$. Thus, $X \sim \mathcal{N}(0, 1)$. Hence, we have shown that each convergent subsequence of $\{X_n\}_{n\in\mathbb{N}}$ converges in distribution towards a standard normal distribution. Moreover, since $\{X_n\}_{n\in\mathbb{N}}$ is tight, such a subsequence exists by Theorem 25.10 in Billingsley (1995). Therefore, by the Corollary, p. 337 in Billingsley (1995), $X_n \xrightarrow[n\to\infty]{distr.} \mathcal{N}(0, 1)$.

Acknowledgements

The research of Christophe A.N. Biscio and Rasmus Waagepetersen are supported by The Danish Council for Independent Research | Natural Sciences, grant DFF – 7014-00074 "Statistics for point processes in space and beyond", and by the "Centre for Stochastic Geometry and Advanced Bioimaging", funded by grant 8721 from the Villum Foundation.

References

Billingsley, P., 1995. Probability and Measure, 3rd Edition. John Wiley & Sons, New York-Chichester-Brisbane, wiley Series in Probability and Mathematical Statistics.

Bolthausen, E., 1982. On the central limit theorem for stationary mixing random fields. The Annals of Probability 10 (4), 1047–1050.

Chilès, J.-P., Delfiner, P., 2008. Geostatistics. John Wiley & Sons, Inc.

Guan, Y., Loh, J. M., 2007. A thinned block bootstrap variance estimation procedure for inhomogeneous spatial point patterns. Journal of the American Statistical Association 102 (480), 1377–1386.

Guyon, X., 1995. Random fields on a network. Probability and its Applications. Springer, New York.

Karácsony, Z., 2006. A central limit theorem for mixing random fields. Miskolc Mathematical Notes. A Publication of the University of Miskolc 7 (2), 147–160.

Møller, J., Waagepetersen, R. P., 2004. Statistical Inference and Simulation for Spatial Point Processes. Chapman and Hall/CRC, Boca Raton.

Rathbun, S. L., 1996. Estimation of Poisson intensity using partially observed concomitant variables. Biometrics 52 (1), 226–242.

Rathbun, S. L., Cressie, N., 1994. Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson point processes. Advances in Applied Probability 26 (1), 122–154.

Schoenberg, F., 2005. Consistent parametric estimation of the intensity of a spatial-temporal point process. Journal of Statistical Planning and Inference 128, 79–93.

Stein, C. M., 1972. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory. University of California Press, pp. 583–602.

Stein, C. M., 1981. Estimation of the mean of a multivariate normal distribution. The Annals of Statistics 9 (6), 1135–1151.

Waagepetersen, R., Guan, Y., 2009. Two-step estimation for inhomogeneous spatial point processes. Journal of the Royal Statistical Society. Series B. Statistical Methodology 71 (3), 685–702.

Waagepetersen, R. P., 2007. An estimating function approach to inference for inhomogeneous Neyman-scott processes. Biometrics 63, 252–258.