Nonparametric methods for Irregularly Spaced Non-Gaussian Spatial Data Analysis

S.N. Lahiri

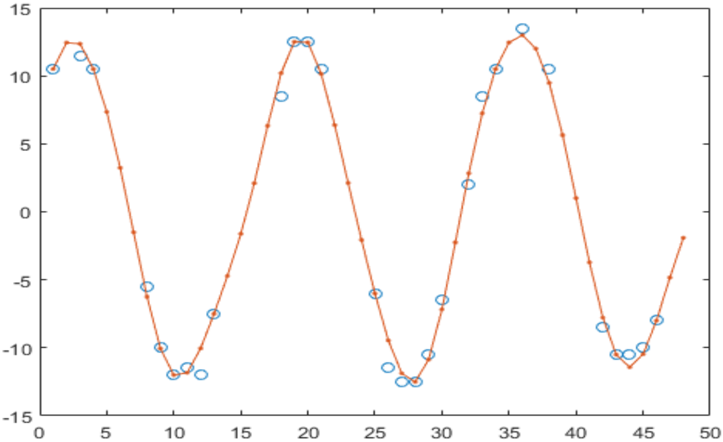
Department of Statistics

snlahiri@ncsu.edu

Aug 25, 2016

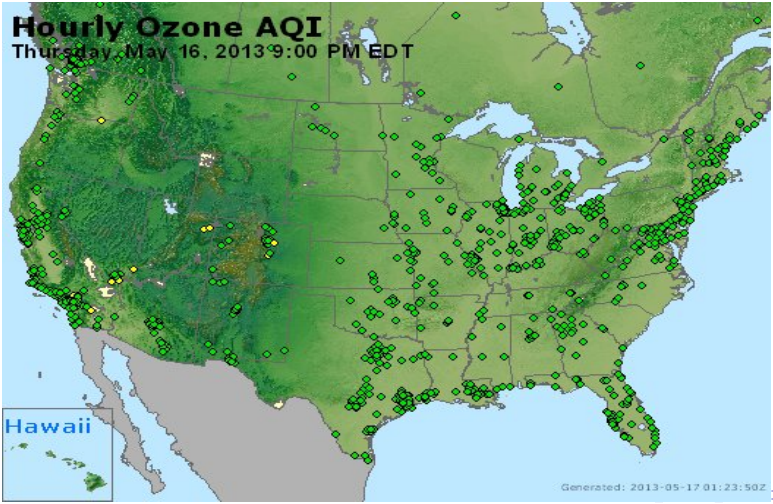
Irregularly spaced time series/spatial data

- Consider the irregularly spaced time series:
- Note that the length of the time intervals ≈ 50 and the sample size ≈ 25.



Irregularly spaced spatial data

- Locations of ground based monitoring stations in the US for the Air Quality Index (AQI).



Some important characteristics of irregularly spaced Time Series/spatial data

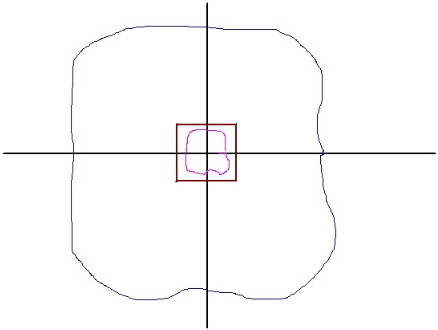
- Typically,
  S = the "size" of the domain of observations
  is **different** from
  n = the sample size.
  - In the time series case, S = the length of the time interval where the observations are taken;
  - In the spatial case, S = area/volume of the sampling region
- The **locations** of the time points OR spatial locations (in the spatial data case) **do not fall on a regular grid**.
- **Distribution of these data-locations** are also **NOT always uniform** !
- In the spatial case, the **shape** of the sampling region can be **non-convex**.

Effects

- **Each of these factors complicate sampling properties of estimators** that we know when the data are regularly spaced !
- Inference tools must be adapted/developed to deal with the complications!
- I will describe some known results and methodology that are available in a spatial framework, in $d \geq 1$-dimensions.
- The time series case will follow as a special case, with $d = 1$.

Framework -I : Sampling Region

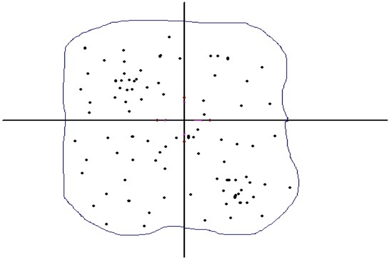
- Let $\mathcal{D}_0$ be an open connected subset of $(-1/2, 1/2]^d$, containing the origin and let $\lambda_n \to \infty$ as $n \to \infty$.
- The sampling region $\mathcal{D}_n$ is obtained by 'inflating' $\mathcal{D}_0$ by a multiplicative factor λ_n, i.e., $\mathcal{D}_n = \lambda_n \mathcal{D}_0$.



Framework -II : Sampling Design

- Let $\mathbf{X}_k \overset{iid}{\sim} f(\mathbf{x}),\ k \geq 1$, where $f(\mathbf{x})$ is a continuous, positive probability density function on $\mathcal{D}_0$.

- We assume that the sampling sites $\mathbf{s}_1, \cdots, \mathbf{s}_n$ are obtained by the relation: $\mathbf{s}_i \ \equiv \ \mathbf{s}_{in} \ = \ \lambda_n \mathbf{x}_i, \quad 1 \leq i \leq n$.



The Framework : Some Remarks

- This serves as a convenient formulation to study sampling behaviors.
- Here:
  - $S \equiv S_n =$ the size of the sampling region $= \lambda_n^d \cdot \text{vol.}(\mathcal{D}_0)$.
  - $n =$ the sample size .
- We suppose that a continuous parameter spatial process $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$ are observed at locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$.
- **Data:** $\{Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)\}$.
- Also, suppose that the $Z(\cdot)$-process is stationary and has enough finite moments.
- Let $Z(\mathbf{s}) = \mu$ and $\bar{Z}_n = n^{-1} \sum_{i=1}^{n} Z(\mathbf{s}_i)$.

Statistical Properties

- Properties of estimators like $\bar{Z}_n$ depends critically on the relative orders of n and λ_n.
- The main cases are:
  - **Case I:** $\lambda_n = O(1)$ as $n \to \infty$. (Infill)
  - **Case II:** $\lambda_n \to \infty$ as $n \to \infty$. (Increasing Domain)
- Within Case II, we have
  - **Case II.1:** $n/\lambda_n^d \to c_* \in (0, \infty)$ (Pure Increasing Domain or PID)
  - **Case II.2:** $n/\lambda_n^d \to \infty$ (Mixed Increasing Domain or MID)
- Technically, $n/\lambda_n^d \to 0$ is a possibility, but leads to uninteresting/simple results, like the independent case.

Statistical Properties : Infill case

- Under **Case I:** $\lambda_n = O(1)$ as $n \to \infty$. (Infill), estimators like $\bar{Z}_n$ are not consistent :

Theorem

If $Z(\cdot)$ is mean-square continuous and $\lambda_n \to \lambda_0 \in (0, \infty)$, then

$$\bar{Z}_n \to \int_{\mathcal{D}} Z(\lambda_0 \mathbf{s}) f(\mathbf{s}) d\mathbf{s}, \quad in \quad L^2, \quad a.s.$$

- Here $\bar{Z}_n$ has a random limit !!
- Although the estimation task is difficult, one gets consistent prediction under Case I.
- See Lahiri (1996; Sankhya), Stein (1990, 1991, ...AoS), Stein (1999; Springer) & the references therein!

- First consider the standard case of a regular grid!

Theorem

(The Regular grid case:) Suppose that $\lambda_n \to \infty$, $\mathcal{D}_0 = [-1/2, 1/2]^d$ *and the data-locations lie on the integer grid. Then, under some weak dependence condition,*

$$\sqrt{n}\Big[\bar{Z}_n - \mu\Big] \to^d N\Big(0, \sigma^2_\infty\Big)$$

where $\sigma^2_\infty = \sum_{\mathbf{i}\in\mathbb{Z}^d} Cov(Z(\mathbf{0}), Z(\mathbf{i})) = (2\pi)^d\tilde{\phi}(\mathbf{0})$, *and* $\tilde{\phi}(\cdot)$ *is the (folded) spectral density of the* $Z(\cdot)$ *process.*

- In the irregularly spaced case, we have the following (Lahiri (2003; Sankhya, Series A)):

Theorem

(The irregularly spaced case:) Suppose that $Z(\cdot)$ is SOS with ACvF $\gamma(\cdot)$ and spectral density $\phi(\cdot)$ and that some suitable weak dependence conditions hold. Let $n/\lambda_n^d \to c_* \in (0, \infty]$. Then,

$$\sqrt{\lambda_n^d}\Big[\bar{Z}_n - \mu\Big] \to^d N\Big(0, \sigma_\infty^2\Big)$$

where $\sigma_\infty^2 = c_*^{-1}\gamma(\mathbf{0}) + [\int f^2](2\pi)^d \phi(\mathbf{0})$.

- Thus, the asymptotic variance depends on the spatial sampling density f and the PID/MID constasnt c_*.

Remarks:

- The rate is determined by the volume of the sampling region - not by the sample size!
- Confidence intervals will have widths of the order $\frac{1}{\sqrt{\text{vol.}(\mathcal{D}_n)}}$, not of the usual order $\frac{1}{\sqrt{n}}$.
- Estimation of the asymptotic standard error is more difficult.
- Suitable variants of the **Spatial Block Bootstrap** are known to provide valid estimators of the (asymptotic) variance, automatically under either of the scenarios (PID/MID). (See Lahiri and Zhu (2006; AoS)).

Implications for spectrum estimation

Definition

The scaled Discrete Fourier Transform (DFT) of the sample $\{Z(\mathbf{s}_1), \ldots\ Z(\mathbf{s}_n)\}$ is given by, for $\boldsymbol{\omega} \in \mathbb{R}^d$,

$$d_n(\boldsymbol{\omega}) = \lambda_n^{d/2} n^{-1} \sum_{j=1}^{n} Z(\mathbf{s}_j) \exp\left(\iota \boldsymbol{\omega}' \mathbf{s}_j\right),$$

where $\iota = \sqrt{-1}$.

- Write $d_n(\boldsymbol{\omega}) = C_n(\boldsymbol{\omega}) + \iota S_n(\boldsymbol{\omega})$.
- Then, $C_n(\boldsymbol{\omega})$ and $S_n(\boldsymbol{\omega})$ are respectively the cosine and sine transforms of the sample.

Joint distribution of the DFTs

Theorem

Suppose that for $j = 1, \cdots, r$, $r \in \mathbb{N}$, $\{\boldsymbol{\omega}_{jn}\}$ *are sequences satisfying* $\boldsymbol{\omega}_{jn} \to \boldsymbol{\omega}_j \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ *and* $\boldsymbol{\omega}_j \pm \boldsymbol{\omega}_k \neq \mathbf{0}$ *for all* $1 \leq j \neq k \leq r$. *Then,*

$$[C_n(\boldsymbol{\omega}_{1n}), S_n(\boldsymbol{\omega}_{1n}), \cdots, C_n(\boldsymbol{\omega}_{rn}), S_n(\boldsymbol{\omega}_{rn})]'$$

$$\xrightarrow{d} N\left[\mathbf{0}, \begin{pmatrix} A_1 I_2 & \mathbf{0}_{2\times 2} & \cdots & \mathbf{0}_{2\times 2} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \cdots & A_r I_2 \end{pmatrix}\right], \quad a.s. \quad (P_{\mathbf{X}}),$$

where $2A_j = c_*^{-1}\gamma(\mathbf{0}) + \int f^2 \cdot (2\pi)^d \phi(\boldsymbol{\omega}_j)$.

Remarks

- **Irregular spacings do NOT necessarily kill the asymptotic independence property of DFTs** that is well-known in the equi-spaced time series case !
- The spectrum is now over $\mathbb{R}^d$, not over $[-\pi, \pi]^d$.
- Estimation of the spectrum using the periodogram requires slight adjustments, as implied by the theorem.
- Specifically, instead of using the raw periodogram $I_n(\boldsymbol{\omega}) \equiv \left| d_n(\boldsymbol{\omega}) \right|^2$, one must use the **bias corrected periodogram**, given by

$$\tilde{I}_n(\boldsymbol{\omega}) = I_n(\boldsymbol{\omega}) - c_*^{-1} \hat{\gamma}_n(\mathbf{0}), \quad \boldsymbol{\omega} \in \mathbb{R}^d,$$

particularly when $\lambda_n^d \asymp n$.

Estimation of the Covariance Function

- Nonparametric estimation of the covariance function of the $Z(\cdot)$-process over $\mathbb{R}^d$ (in the MID case) is addressed by
  - Hall, Fisher & Hoffman (1994; AoS) for the time series case ($d = 1$), and by
  - Hall and Patil (1994; PTRF) for the spatial case ($d \geq 2$).
- Steps include
  - Estimation of the spectral density using kernel smoothing
  - Fourier inversion to define the covariance estimator.
- The resulting estimator is **non-negative definite**!!!

Estimation of the Covariance Function

Theorem

Suppose that $n/\lambda_n^d \to \infty$ (– the MID case). Then, under the given framework, for any $a > 0$,

$$\lambda_n^d \int_{\|\mathbf{h}\| \le a} [\hat{\gamma}_n(\mathbf{h}) - \gamma(\mathbf{h})]^2 d\mathbf{h} \to^d \int_{\|\mathbf{h}\| \le a} W(\mathbf{h})^2 d\mathbf{h}$$

where $W(\cdot)$ is a zero mean Gaussian process on $\mathbb{R}^d$ with continuous sample paths.

- The most important aspect of this result is that the rate of convergence, namely $\lambda_n^{-d/2}$, is as good as that of estimating a finite dimensional parameter!!

Part 2

Methodological aspects

Inference methodology for irregularly spaced spatial data

- The Central Limit Theorem can be used to derive asymptotic distributions of asymptotically linear statistics, such as the (pseudo-) MLE, LS-estimators, etc.

- Estimation of the asymptotic variance is a difficult problem - variants of **Block Bootstrap methods** that adapt to the irregularly spaced case are available, as pointed out earlier!!

- We now describe a recent approach to nonparametric likelihood based inference, known as the

  Empirical Likelihood

  that bypasses the need for direct variance estimation.

- Consider a parametric model $\{f(\cdot;\theta) : \theta \in \Theta\}$ and let $X_1, \ldots, X_n$ be iid, $X_1 \sim f(\cdot;\theta_0)$.
  - For example, $X_1, \ldots, X_n$ be iid, $X_1 \sim N(\theta, 1)$, $\theta \in \Theta = \mathbb{R}$. Then, $f(x;\theta) = \frac{\exp(-(x-\theta)^2/2)}{\sqrt{2\pi}}$, $x \in \mathbb{R}$.
- The (parametric) *likelihood function* for θ is

$$L_n(\theta) = \prod_{i=1}^{n} f(X_i;\theta).$$

Normal Likelihood

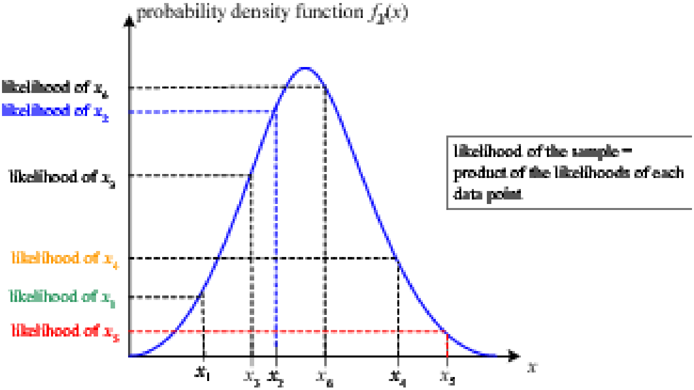


Illustration of likelihood calculation for N=6 and a normal distribution

Parametric Likelihood

- An estimator of θ is given by

$$\hat{\theta} = \text{ argmax } \log L(\theta),$$

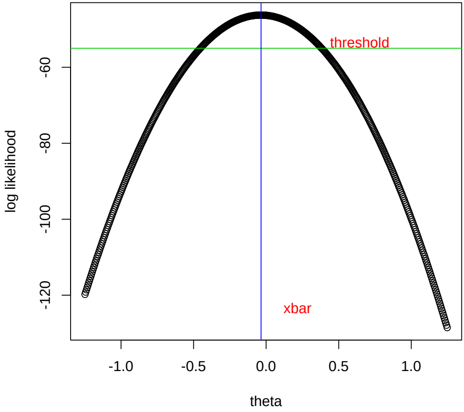
  the *maximum likelihood estimator*!
- Under some regularity conditions, Wilk's theorem asserts that

$$-2 \log R_n(\theta_0) \to^d \chi_p$$

  where $R_n(\theta_0)$ is the *likelihood ratio statistic* (LRT) for testing $H_0 : \theta = \theta_0$.

Normal log-likelihood : Based on 100 observations

- Calibration of the test $H_0 : \theta = \theta_0$ can be done using the Chi-squared limit!
- The LRT can also be inverted to get a confidence set for θ!!



Q

Can we define a likelihood without a parametric model ?

Empirical Likelihood

- Empirical Likelihood (EL) of Owen (1988) is a method that defines a likelihood for **certain** population parameters *without* requiring a parametric model.
- Let $X_1, \ldots, X_n$ be iid with mean $\mu \in \mathbb{R}$. The EL for μ is

$$L(\mu) = \sup \left\{ \prod_{i=1}^{n} \pi_i : \pi_i \geq 0, \sum \pi_i = 1, \sum \pi_i X_i = \mu \right\}$$

- i.e., $L(\mu)$ gives the max likelihood for a $\mu \in \mathbb{R}$ from discrete distributions supported on $\{X_1, \ldots, X_n\} \equiv \mathcal{X}$.

Empirical Likelihood

- The unconstrained maximum is at $\pi_i = n^{-1}$ for all i.
- Thus, the EL ratio statistic for testing $H_0 : \mu = \mu_0$ is

$$R_n(\mu_0) = \frac{L_n(\theta_0)}{n^{-n}}.$$

- Under some mild regularity conditions, Owen (1988; Biometrika) proved a version of Wilk's Theorem:

$$-2 \log R_n(\mu_0) \to^d \chi_1^2.$$

Empirical Likelihood

- EL methodology has been extended to deal with more general parameters.
- An important work by **Qin and Lawless (1994; AoS)** formulated EL for parameters defined by Estimating Equations, and proved a Chi-sq limit law.
- It allows parameters satisfying a moment condition like:

$$E\psi(X_1, \theta) = 0.$$

  for some function $\psi : \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}^p$.
- For example, with $d = p = 1$, $\psi(x, \theta) = x - \theta$ corresponds to θ =the population mean!

Some advantages of using the EL

- **It is a nonparametric method** - it does NOT require the statistician to specify a model (and hence, there is no model misspecification error)!
- **It does NOT require explicit variance estimation to construct a CI/test** !

  Contrast this with the usual approach based on large sample distribution of the M-estimator:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \ \to^d N(0, \tau^2)$$

  where $\tau^2 = [E\psi(X_1; \theta_0)^2] \cdot [E\psi'(X_1; \theta_0)]^{-2}$.
- **It allows for a distribution free calibration**, as the limit distribution is known (viz., χ^2_p.)

Literature Review

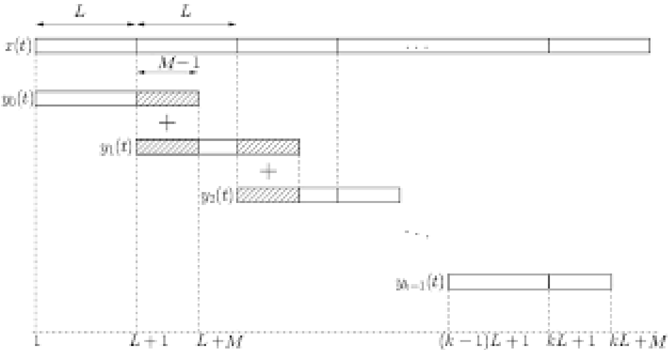
Some more references on the EL method under independence are given by

- **Owen (2001; Chapman & Hall)** - monograph
- **Chen and Hall (1993; AoS)** : Quantiles
- **Qin and Lawless (1994; AoS** : Estimating Equations
- **DiCiccio, Hall and Romano (1996; AoS)** : Bartlett Corrections
- **Bertail (2006; Bernoulli)** : Semiparametric models
- **Lahiri and Mukhopadhyay (2012; AoS)** : Penalized EL in increasing dimensions $p \gg n$,

EL under dependence : Some technical issues

- Under dependence, the standard EL fails in the sense that the limit involves population parameters.
- **Kitamura (1997; AoS)** introduced Block EL (BEL) for time series data and established Wilk's Phonomenon.
- Let $X_1, \ldots, X_n$ be a stationary time series with mean $\mu \in I\!R$.
- Let $\bar{X}_{1,L} = L^{-1} \sum_{i=1}^{L} X_t$, $\bar{X}_{2,L} = L^{-1} \sum_{i=2}^{L+1} X_t$, ..., denote the successive block averages, for some $L \approx n^{\delta}$, $\delta \in (0, 1)$.

Construction of the BEL for time series



- $M = 1$ gives the maximum overlapping version
- $M > 1$ can be used to reduce computational burden

EL under dependence : Some technical issues

- The *maximum overlapping BEL* for μ is defined as

$$L^{\mathrm{BEL}}(\mu) = \sup\left\{ \prod_{i=1}^{N} \pi_i : \pi_i \geq 0, \sum \pi_i = 1, \sum \pi_i \bar{X}_{i\ell} = \mu \right\}$$

  where $N = n - L + 1$.

- **Kitamura (1997)** established Wilk's Phonomenon for the BEL: *Under some regularity conditions,*

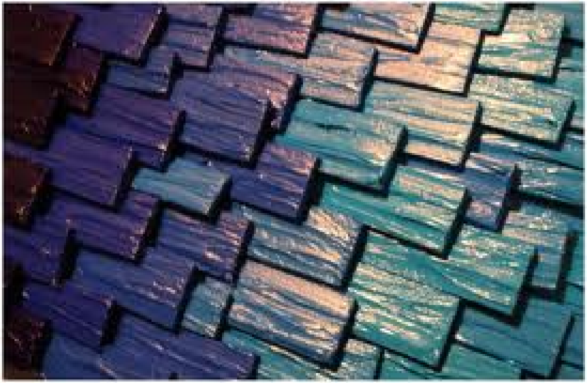
$$-2A \log R^{\mathrm{BEL}}(\mu_0) \to^{d} \chi_1,$$

  where A scale adjustment involving known quantities.

- Note that the limit is distribution free (Chi-squared).

Q: How do we extend the EL to spatial data?

Construction of the BEL for spatial data



- The idea is to use d-dimensional block averages to define the BEL for μ, as in the time series case!

EL in the frequency domain

- **For parameters related to the covariance structure** of a spatial process, a more suitable approach to formulate the **EL in the frequency domain**!!
- **Monti (1997; Biometrika)** first considered EL for time series in the frequency domain, which was refined and extended by **Nordman and Lahiri (2006; AoS)** .
- Extension to the spatial case with **irregularly spaced** data locations has been done recently by **Bandopadhyay, Nordman and Lahiri (2015; AoS)**.

EL for spatial data

- As noted before, **it does not require variance estimation** - which can be a nightmare for spatial processes (e.g, recall that for a spatial process $Z(\cdot)$ observed at n data-locations,

$$\mathrm{Var}(\bar{Z}) \approx \sigma^{*2}\lambda_n^{-d}$$

where $\sigma^{*2} = g\Big(\phi(\cdot), \frac{\mathrm{vol}(\mathcal{D}_n)}{n}, f\Big)$.

- Further, **the EL does not require the spatial process to be Gaussian** to produce valid inference !!!
- No (parametric) model formulation is needed !

Thank you!!