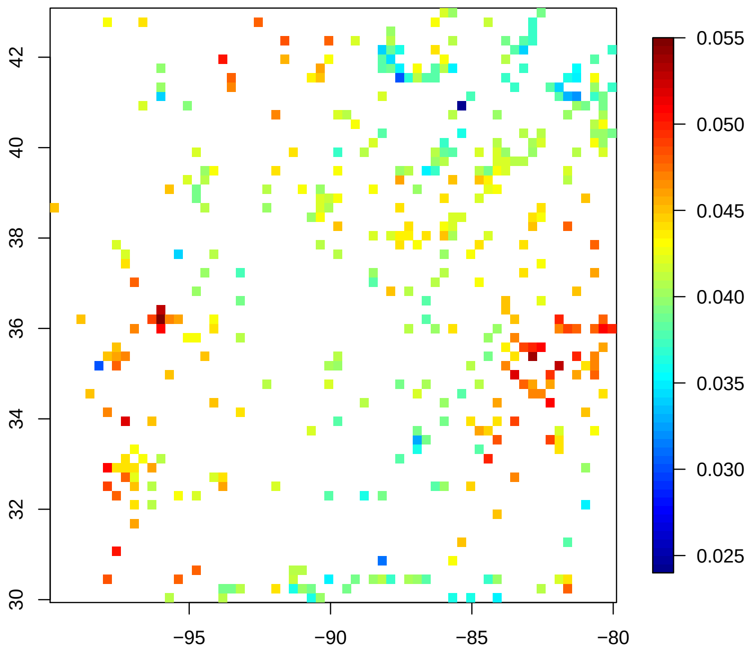
Fourier based methods for spatial data observed on irregularly spaced locations

Suhasini Subba Rao
Texas A&M University

CIRM, February, 2016

Spatial Data



- The ground ozone levels taken on 4th April in the Ohio Valley, USA.
- Each point corresponds to a measurement station.

- Central to the analysis of spatial data is understanding the underlying process (usually assumed to be stochastic) which generates the data. Typically, this means modelling the spatial covariance structure of the stochastic process.

Classical spatial procedures

- Nonparametric covariance estimation methods include:
  - Estimation of the Variogram (see Cressie (1993)).
  - Directly estimating the covariance nonparametrically and then "adjusting it" by taking Fourier transforms to ensure it is non-negative definite (Hall et. al. (1994)).

- Parametric covariance estimation methods include:
  - Likelihood approaches based on composite Gaussian likelihood (Stein et. al. (2005)).
  - Spectral/Whittle approaches (Fuentes (2007), Matsuda and Yajima (2009)).

- Often the underlying assumption is stationarity. This assumption needs to be checked (Jun and Genton (2012)).

- All the above procedures work quite well and their sampling properties are understood.

- However, there is no real coherency in the approaches eg. nonparametric covariance estimation procedure is completely different to the parametric estimation procedure.

- The purpose of this talk is to develop a unified approach to some of the problems described above, which we hope will yield statistics with "good" statistical properties and are computationally feasible.

A time series motivation

- Our motivation comes from discrete time time series where several parameters of interest can be written in terms of the functional

$$A(\phi, f) = \int_0^{2\pi} \phi(\omega) f(\omega) d\omega.$$

where f is the spectral density function. Different ϕ lead to different parameters (see Dahlhaus and Janas (1996)).

- Given the stationary time series $\{X_t\}_{t=1}^{T}$, to estimate $A(\phi, f)$ we replace f with the periodogram $I_T(\omega_k) = |J_T(\omega_k)|^2$

$$A(\phi, I_T) = \frac{1}{T}\sum_{k=1}^{T} \phi(\omega_k)|J_T(\omega_k)|^2, \qquad J_T(\omega_k) = \frac{1}{\sqrt{2\pi}}\sum_{t=1}^{T} X_t e^{it\omega_k}.$$

$$A(\phi, I_T) = \frac{1}{T}\sum_{k=1}^{T}\phi(\omega_k)|J_T(\omega_k)|^2$$

- $\omega_k = \frac{2\pi k}{T}$ are often called the Fourier frequencies.

- $A(\phi, I_T)$ is an estimator of $A(\phi, f)$ (includes the Whittle likelihood, spectral density estimator, covariance estimator etc).

- Can a similar class of statistics be defined for spatial data observed at irregular locations?

Overview

- We define frequency grid over which we define the class the statistics.

- Consider the sampling properties of this class of statistics:
  - The mean and variance.
  - There are stark differences between Gaussian and non-Gaussian.
  - Obtain a CLT.

- The variance is difficult to directly estimate and we propose a simple method for estimating the variance.

- Flavour of proofs.

The observations

- Typically, for spatial data one observes a spatial process $\{Z(\boldsymbol{s}); \boldsymbol{s} \in \mathbb{R}^d\}$ only at a finite number of locations:
  - Locations denoted as $\{\boldsymbol{s}_j; j = 1, \ldots, n\}$ on square of length λ ($[-\lambda/2, \lambda/2]^d$; one can also use a rectangle).
  - Observations: $\{(Z(\boldsymbol{s}_j), \boldsymbol{s}_j); j = 1, \ldots, n\}$.
  - We assume that the locations $\{\boldsymbol{s}_j\}$ are iid random variables that are independent of the spatial process.

- We will assume that process is covariance stationary with $\mathrm{cov}[Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2)] = c(\boldsymbol{s}_1 - \boldsymbol{s}_2)$. If the process were not nonstationary the results would be very different.

- Asymptotic formulation:
  - Mixed domain asymptotics: The spatial domain $\lambda \to \infty$ and the number of locations $n \to \infty$ at such a rate that $\lambda^d/n \to 0$. In other words the locations get denser as the field grows.
    We will work mainly under this asymptotic set-up.
  - Pure increasing domain asymptotics: when λ and $n \to \infty$ in such a way that $\lambda^d/n \to c$ $(0 < c < \infty)$.
    Essentially all the results hold in the Gaussian case (with an additional bias), however in the non-Gaussian case some differences arise.
  - A different asymptotic formulation is to keep the domain λ fixed, but let the number observations $n \to \infty$. Under this set-up the estimators will be biased and no CLT can be derived (not considered here).

The Fourier transform of irregular sampled spatial data

- Unlike discrete time time series, there is no unique way of defining the Fourier transform of in irregular locations.

- Following Matsuda and Yajima (2009) and Bandyopadhyay and Lahiri (2009) we define the Fourier transform:

$$J_n(\boldsymbol{\omega}) = \frac{\lambda^d}{n}\sum_{j=1}^{n} Z(\boldsymbol{s}_j)\exp(i\boldsymbol{s}_j'\boldsymbol{\omega}) \qquad \boldsymbol{\omega} \in \mathbb{R}^d$$

  (note Masry (1978) defined something similar for continuous time time series with Poission sampling).

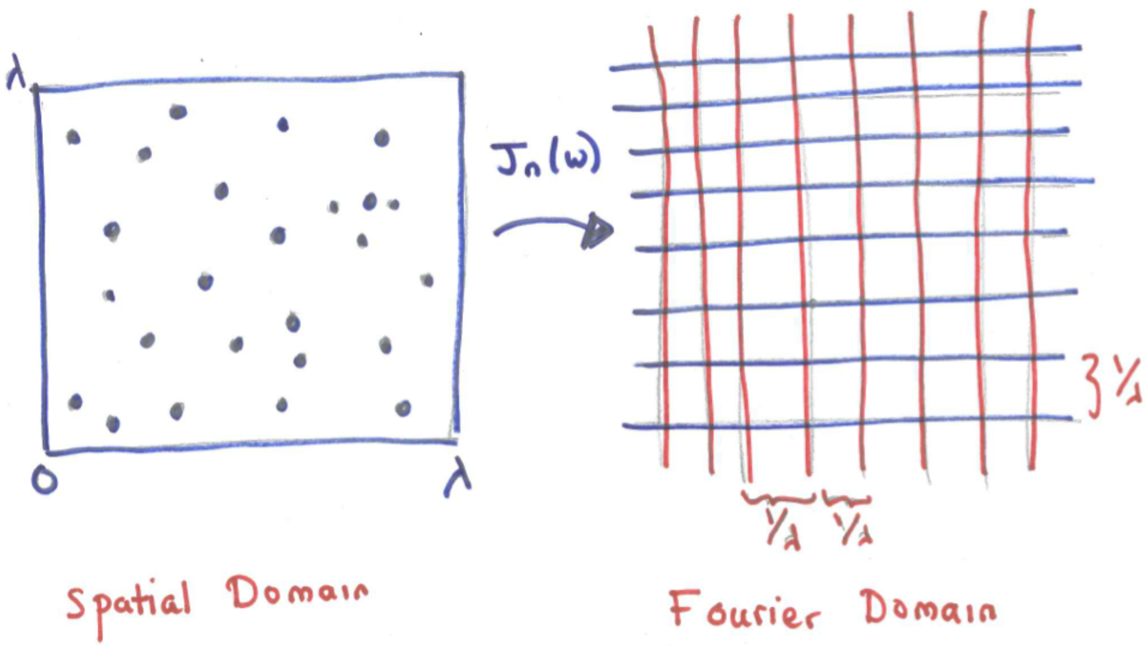
- Computing $J_n(\boldsymbol{\omega})$ at all frequencies $\boldsymbol{\omega} \in \mathbb{R}^d$ is infeasible.

$$J_n(\boldsymbol{\omega}) = \frac{\lambda^d}{n}\sum_{j=1}^{n} Z(\boldsymbol{s}_j)\exp(i\boldsymbol{s}_j'\boldsymbol{\omega}) \qquad \boldsymbol{\omega} \in \mathbb{R}^d$$

- Question Can we “gridify” $\mathbb{R}^d$ in such a way that we do not loose any information on $J_n(\boldsymbol{\omega})$?

  What are the analogous “Fourier frequencies” for the Fourier transform of observations observed at irregular locations.

- We focus on the frequencies $\boldsymbol{\omega_k} = (2\pi k_1/\lambda, \ldots, 2\pi k_d/\lambda)$ $(\boldsymbol{k} = (k_1, \ldots, k_d))$.

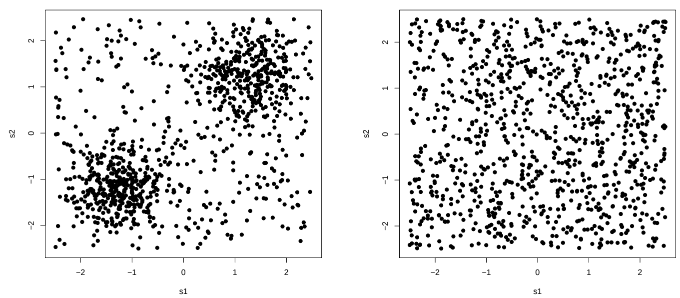


Case $d = 2$.

Properties of the Fourier transform (Assumptions)

- Spatial Process The covariance of the tails of the tails decay sufficiently fast $|c(\boldsymbol{s})| \sim C|\boldsymbol{s}|^{-(2+\delta)}$ (a type of short memory assumption).

- Location density The locations $\{\boldsymbol{s}_j\}$ are iid random variables defined on $[-\lambda/2, \lambda/2]^d$ and the sampling density is density $\lambda^{-d}h(\boldsymbol{s}/\lambda)$.



The Fourier representation of the location density $h(\boldsymbol{s}) = \sum_{\boldsymbol{j}\in\mathbb{Z}^d} \gamma_{\boldsymbol{j}} \exp(i\boldsymbol{j}'\boldsymbol{\omega})$ is such that the Fourier coefficients satisfy $\sum_{\boldsymbol{j}\in\mathbb{Z}^d} |\gamma_{\boldsymbol{j}}| < \infty$. Thus $|\gamma_{\boldsymbol{j}}| \to 0$ as $|\boldsymbol{j}|_1 \to \infty$.

Properties of the Fourier transform

- Under general random sampling of locations:

$$\mathrm{cov}\left[J_n(\boldsymbol{\omega}_{\boldsymbol{k}_1}), J_n(\boldsymbol{\omega}_{\boldsymbol{k}_2})\right] = \langle \gamma, \gamma_{(\boldsymbol{k}_2-\boldsymbol{k}_1)}\rangle f\left(\boldsymbol{\omega}_{\boldsymbol{k}_1}\right) + \frac{c(0)\gamma_{\boldsymbol{k}_2-\boldsymbol{k}_1}\lambda^d}{n} + O\left(\frac{1}{\lambda}\right),$$

  where $\langle \gamma, \gamma_{\boldsymbol{r}}\rangle = \sum_{\boldsymbol{j}\in\mathbb{Z}^d} \gamma_{\boldsymbol{j}}\overline{\gamma_{\boldsymbol{j}+\boldsymbol{r}}}$ and $f(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} c(\boldsymbol{s})\exp(i\boldsymbol{s}'\boldsymbol{\omega})d\boldsymbol{s}$ is the spatial spectral density.

- $\mathrm{var}[J_n(\boldsymbol{\omega}_{\boldsymbol{k}})] = f(\boldsymbol{\omega}_{\boldsymbol{k}})\int_{\mathbb{R}^d} h(\boldsymbol{s})^2 d\boldsymbol{s} + O(\lambda^d n^{-1} + \lambda^{-1})$.

- $\mathrm{cov}\left[J_n(\boldsymbol{\omega}_{\boldsymbol{k}_1}), J_n(\boldsymbol{\omega}_{\boldsymbol{k}_2})\right] = O(|\gamma_{\boldsymbol{k}_1-\boldsymbol{k}_2}| + \lambda^d n^{-1} + \lambda^{-1})$.

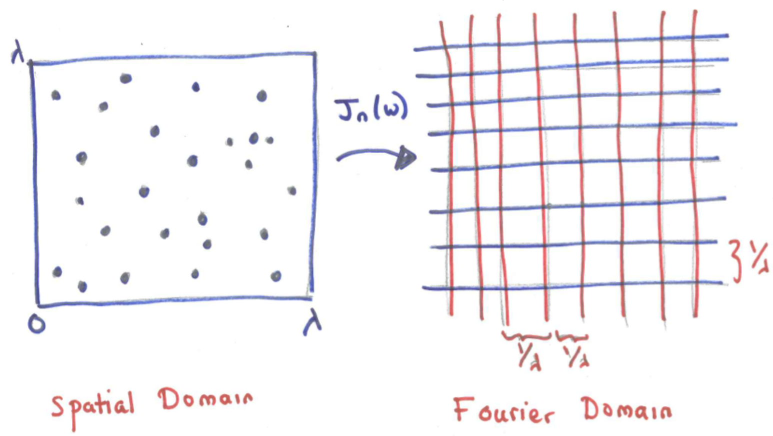
  Recall that $\gamma_{\boldsymbol{k}}$ are the Fourier coefficients of the location density.

- In discrete time time series the frequency domain is limited to $[0, 2\pi]$. Whereas when the locations are irregular, the frequency grid is not limited (since aliasing cannot arise). In theory we can identify all frequencies (see Shapiro and Silverman (1960)).

- In the case the locations come from a uniform distribution $(h(\boldsymbol{s}) = \prod_{i=1}^{d} I_{[-1/2,1/2]}(s_i))$ we have a stronger result:

$$\mathrm{cov}\left[J_n(\boldsymbol{\omega}_{\boldsymbol{k}_1}), J_n(\boldsymbol{\omega}_{\boldsymbol{k}_2})\right] = \begin{cases} f(\boldsymbol{\omega}_k) + O(\frac{1}{\lambda} + \frac{\lambda^d}{n}) & \boldsymbol{k}_1 = \boldsymbol{k}_2 (= \boldsymbol{k}) \\ O(\frac{1}{\lambda^{d-b}}) & \boldsymbol{k}_1 - \boldsymbol{k}_2 \neq 0 \end{cases}$$

where $b =$ the number of zero elements in the vector $\boldsymbol{k}_1 - \boldsymbol{k}_2$ (or number of common elements between $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$) and $f(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} c(\boldsymbol{s}) \exp(i\boldsymbol{s}'\boldsymbol{\omega}) d\boldsymbol{s}$ is the spatial spectral density.



- The results imply that the Fourier transform defined on the grid $\boldsymbol{\omega_k} = (2\pi k_1, \ldots, 2\pi k_d)/\lambda$ is not highly correlated and suggest that estimators can be defined on such a grid.

- Bandyopadhyay, Lahiri and Nordman (2015) use the Fourier transform on a wider grid in order to define the spatial empirical likelihood.

The integrated Fourier transform

Define the analogous version of the integrated periodogram [1]:

$$Q_{a,\lambda}(g;\boldsymbol{r}) = \frac{1}{\lambda^d}\sum_{k_1,\dots,k_d=-a}^{a} g(\boldsymbol{\omega_k})J_n(\boldsymbol{\omega_k})\overline{J_n(\boldsymbol{\omega_{k+r}})} - \text{bias} \qquad \boldsymbol{r}\in\mathbb{Z}^d.$$

- In the case that $\boldsymbol{r}=0$, $Q_{a,\lambda}$ is a sum over $|J_n(\boldsymbol{\omega_k})|^2$. Heuristics suggest that we can replace $|J_n(\boldsymbol{\omega_k})|^2$ with the spectral density $f(\boldsymbol{\omega_k})$, then $Q_{a,\lambda}$ is an estimator of the functional

$$I\left(g;\frac{a}{\lambda}\right) = \frac{1}{(2\pi)^d}\int_{[-a/\lambda,a/\lambda]^d} g(\boldsymbol{\omega})f(\boldsymbol{\omega})d\boldsymbol{\omega} \to \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} g(\boldsymbol{\omega})f(\boldsymbol{\omega})d\boldsymbol{\omega}$$

if we let the frequency grid a be such that $a/\lambda\to\infty$ as $\lambda\to\infty$.

[1] Requires $O(n^2a^d)$ computing operations.

Examples: Bounded frequency grid $a = O(\lambda)$

- Whittle likelihood ($g = f_\theta^{-1}(\cdot)$)

$$Q_{a,\lambda}(f_\theta^{-1};0) \propto \frac{1}{\lambda^d}\sum_{k_1,\ldots,k_d=-C\lambda}^{C\lambda}\left(\log f_\theta(\boldsymbol{\omega_k}) + \frac{|J_n(\boldsymbol{\omega_k})|^2}{f_\theta(\boldsymbol{\omega_k})}\right).$$

  We need to constrain the frequency grid since $f_\theta(\boldsymbol{\omega}) \to 0$ as $|\boldsymbol{\omega}| \to \infty$. Discretised version of the Matusda and Yajima (2009) Whittle likelihood.

- Spectral density estimator $g = W_b(\boldsymbol{\omega} - \cdot)$ (and $W_b(\boldsymbol{\omega}) = b^{-d}W(\boldsymbol{\omega}/b)$):

$$Q_{a,\lambda}(W_b(\boldsymbol{\omega}),0) = \frac{1}{\lambda^d}\sum_{k_1,\ldots,k_d=-a}^{a} W_b(\boldsymbol{\omega} - \boldsymbol{\omega_k})|J_n(\boldsymbol{\omega_k})|^2.$$

Examples: Unbounded frequency grid $a/\lambda \to \infty$

- A nonparametric estimator of the spatial (stationary) covariance function $g = e^{i\boldsymbol{v}'\cdot}$:

$$\widehat{c}_n(\boldsymbol{v}) \quad = \quad T\left(\frac{2\boldsymbol{v}}{\lambda}\right) Q_{a,\lambda}(e^{i\boldsymbol{v}'\cdot}; \mathbf{0})$$

where $T(\boldsymbol{u})$ is the triangle kernel. It can be shown that this covariance estimator is a non-negative function.

- The quadratic loss function for parameter estimation:

$$L_n(\theta) = \frac{1}{\lambda^d} \sum_{k_1,\ldots,k_d=-a}^{a} \left(|J_n(\boldsymbol{\omega_k})|^2 - f_\theta(\boldsymbol{\omega_k})\right)^2.$$

The quadratic loss $L_n(\theta)$ does not belong to the $Q_{a,\lambda}$-class but similar methods described below can be used to analysis it. Moreover, the sampling properties of θ depend on $\nabla_\theta L_n(\theta)$ which does belong to the $Q_{a,\lambda}$-class.

$$Q_{a,\lambda}(g;\boldsymbol{r}) = \frac{1}{\lambda^d}\sum_{k_1,\ldots,k_d=-a}^{a} g(\boldsymbol{\omega_k})J_n(\boldsymbol{\omega_k})\overline{J_n(\boldsymbol{\omega_{k+r}})} - \text{bias} \qquad \boldsymbol{r}\in\mathbb{Z}^d.$$

Question What happens when $\boldsymbol{r}\neq 0$?

If the sampling location density is uniform, then

- We will show that $Q_{a,\lambda}(g;\boldsymbol{r}\neq 0)$ asymptotically behaves like an ancillary variable to $Q_{a,\lambda}(g;0)$; does not contain any mean information but does contain information about the variance.

- In the case the process is nonstationary, it contains information about the nonstationarity; this it can be used to test for stationarity.

The mean

Let $I(g;\frac{a}{\lambda}) = \int_{-a/\lambda}^{a/\lambda} g(\boldsymbol{\omega}) f(\boldsymbol{\omega}) d\boldsymbol{\omega}$. Then

- Uniform sampling density ($\{m_j\}$ are non-zero elements of $\boldsymbol{r}$):

$$\mathbf{E}\left[Q_{a,\lambda}(g;\boldsymbol{r})\right] = \begin{cases} O\left(\frac{1}{\lambda^{d-b}}\prod_{j=1}^{d-b}(\log\lambda + \log|m_j|)\right) & \boldsymbol{r} \in \mathbf{Z}^d/\{0\} \\ I\left(g;\frac{a}{\lambda}\right) + O\left(\frac{\log\lambda}{\lambda} + \frac{1}{n}\right) & \boldsymbol{r} = \mathbf{0} \end{cases}.$$

- Non-uniform sampling density

$$\mathbf{E}\Big[Q_{a,\lambda}(g;\boldsymbol{r})\Big] = \langle\gamma,\gamma_{-\boldsymbol{r}}\rangle I\left(g;\frac{a}{\lambda}\right) + O\left(\frac{\log\lambda + I_{\boldsymbol{r}\neq\mathbf{0}}\log|\boldsymbol{r}|_1}{\lambda}\right).$$

[2]

[2] One can estimate $\langle\gamma,\gamma_{-\boldsymbol{r}}\rangle$ using the methods proposed in Laurent (1996) or Gine and Nickl (2008).

The covariance: Under general sampling

Under Gaussianity of the random field:

- $\lambda^d \sup_{a,\boldsymbol{r}_1,\boldsymbol{r}_2} |\text{cov}\{Q_{a,\lambda}(g;\boldsymbol{r}_1), Q_{a,\lambda}(g;\boldsymbol{r}_2)\}| < \infty$

- Additional conditions on the smoothness of the spectral density

$$\begin{aligned}
&\lambda^d \text{cov}\left[Q_{a,\lambda}(g;\boldsymbol{r}_1), Q_{a,\lambda}(g;\boldsymbol{r}_2)\right] \\
= \quad & \Gamma_{\boldsymbol{r}_1-\boldsymbol{r}_2} C_{a/\lambda} + O\left(\underbrace{\log^2(a)\left[\frac{\log a + \log\lambda}{\lambda}\right] + \frac{\lambda^d}{n}}_{=\ell_{a,\lambda,n}} + \frac{|\boldsymbol{r}_1|_1 + |\boldsymbol{r}_2|_1}{\lambda}\right)
\end{aligned}$$

- $\lambda^d \text{cov}\left[Q_{a,\lambda}(g;\boldsymbol{r}_1), Q_{a,\lambda}(g;\boldsymbol{r}_2)\right] = \Gamma_{\boldsymbol{r}_1-\boldsymbol{r}_2} C_{a/\lambda} + o(1)$.

- $\Gamma_{\boldsymbol{r}} = \sum_{\boldsymbol{j}_1+\boldsymbol{j}_2+\boldsymbol{j}_3+\boldsymbol{j}_4=\boldsymbol{r}} \gamma_{\boldsymbol{j}_1}\gamma_{\boldsymbol{j}_2}\gamma_{\boldsymbol{j}_3}\gamma_{\boldsymbol{j}_4}$; a product of the Fourier coefficients of the spatial density.

$$C_{a/\lambda} = \frac{1}{(2\pi)^d} \int_{2\pi[-a/\lambda, a/\lambda]^d} f(\boldsymbol{\omega})^2 \left[|g(\boldsymbol{\omega})|^2 + g(\boldsymbol{\omega})\overline{g(-\boldsymbol{\omega})} \right] d\boldsymbol{\omega}.$$

- The correlations decay for large lag differences $|\boldsymbol{r}_1 - \boldsymbol{r}_2|_1$.

- The expressions are unwieldy, but simplifications can be made if the sampling of locations are uniformly distributed.

The covariance: uniform sampling

Simplications can be made in the case that the sampling is uniform:

$$
\begin{aligned}
&\lambda^d \mathrm{cov}\left[\Re Q_{a,\lambda}(g;\boldsymbol{r}_1), \Re Q_{a,\lambda}(g;\boldsymbol{r}_2)\right] \\
= &\left\{ \begin{array}{cc} \frac{1}{2}C_{a/\lambda} + O\left(\ell_{a,\lambda,n} + \frac{|\boldsymbol{r}_1|}{\lambda}\right) & \boldsymbol{r}_1 = \boldsymbol{r}_2 \\ O\left(\ell_{a,\lambda,n}\right) & \boldsymbol{r}_1 \neq \boldsymbol{r}_2 \text{ or } -\boldsymbol{r}_2 \end{array} \right. ,
\end{aligned}
$$

- Similar results hold for the imaginary parts.

- Furthermore, $\lambda^d \mathrm{cov}\left[\Re\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r}_1), \Im\widetilde{Q}_{a,\lambda}(g;\boldsymbol{r}_2)\right] = O\left(\ell_{a,\lambda,n}\right)$ (if $\boldsymbol{r}_1 \neq -\boldsymbol{r}_2$).

- This means that real and imaginary parts are near uncorrelated and for $|\boldsymbol{r}|$ small, the variance of the statistics are similar.

Implications

- For any sampling scheme, the construction of the estimator $\widetilde{Q}_{a,\lambda}(g;0)$ depends on the number of frequencies used

$$Q_{a,\lambda}(g;0) = \frac{1}{\lambda^d} \sum_{k_1,\ldots,k_d=-a}^{a} g(\boldsymbol{\omega_k})|J_n(\boldsymbol{\omega_k})|^2 - \text{bias} \qquad \boldsymbol{r} \in \mathbb{Z}^d.$$

However, we have shown that the rate of convergence *does not* depend on a. In other words

$$\mathbf{E}\left|Q_{a,\lambda}(g;0) - \langle\boldsymbol{\gamma},\boldsymbol{\gamma}_0\rangle I(g;\infty)\right|^2 = O\left(\frac{1}{\lambda^d} + \left(\frac{\lambda}{a}\right)^{\beta-1}\right),$$

where $f(\boldsymbol{\omega}) \sim C|\boldsymbol{\omega}|_1^{-\beta}$ for $|\boldsymbol{\omega}|_1 \to \infty$. $O((\lambda/a)^{\beta-1})$ is the bias in approximating $I(g;\frac{a}{\lambda})$ with $I(g;\infty)$.

We can choose a as large is computationally feasible.

- However, in order to obtain an explicit expression for the variance $(\lambda^d \mathrm{var}\,[Q_{a,\lambda}(g;\boldsymbol{r})] = \Gamma_0 C_{a/\lambda} + o(1))$ we need to constrain the rate of growth such that $\log^3(a)/\lambda = o(1)$.
  - Use $a = O(\lambda^k)$ for some $k > 1$.
  - The derivation gave sufficient conditions, but I suspect the rates cannot be improved by much.

The covariance for non-Gaussian random fields

- The result is simplest when stated for uniform sampling (but similar results hold for non-uniform sampling):

$$
\begin{aligned}
&\lambda^d \mathrm{cov}\left[\Re Q_{a,\lambda}(g;\boldsymbol{r}_1), \Re Q_{a,\lambda}(g;\boldsymbol{r}_2)\right] \\
&= \begin{cases} \frac{1}{2}[C_{a/\lambda} + D_{a/\lambda}] + O\left(\underbrace{\ell_{a,\lambda,n}}_{\text{same as Gaussian}} + \underbrace{\frac{(a\lambda)^d}{n^2}}_{\text{extra}}\right) & \boldsymbol{r}_1 = \boldsymbol{r}_2 \\ O\left(\ell_{a,\lambda,n} + \frac{(a\lambda)^d}{n^2}\right) & \boldsymbol{r}_1 \neq \boldsymbol{r}_2 \end{cases},
\end{aligned}
$$

- $D_{a/\lambda} = \frac{1}{(2\pi)^{2d}} \int_{2\pi[-a/\lambda,a/\lambda]^{2d}} g(\boldsymbol{\omega}_1)\overline{g(\boldsymbol{\omega}_2)} f_4(-\boldsymbol{\omega}_1, -\boldsymbol{\omega}_2, \boldsymbol{\omega}_2) d\boldsymbol{\omega}_1 d\boldsymbol{\omega}_2$, where f_4 the fourth order spectral density of the process.

- We would expect the additional fourth order cumulant term (it is analogous to that in time series) in the case the process is non-Gaussian.

- But there is also the additional $O(\frac{(a\lambda)^d}{n^2})$. This terms is due to the fourth order cumulant and arises from an interplay between the random sampling $\boldsymbol{s}_j$ and the fourth order cumulant of the spectral density.

- This term grows if $a >> n^{2/d}/\lambda$. Therefore, unlike the Gaussian case the growth the frequency grid is limited to $a = O(n^{2/d}/\lambda)$.
  - In the non-Gaussian setting it is sensible to choose to $a = O(n^{1/d})$, in this case all the results hold for mixed asymptotics.
  - However, under pure increasing domain asymptotics where $n \sim \lambda$, this constraint means that the frequency grid cannot grow. One solution is to use a wider frequency grid (similar to the empirical likelihood approach of Bandyopadhyay, Lahiri and Nordman (2015)).

Central limit theorem

Under Gaussianity of the random field

- For $q \geq 3$: $|\text{cum}_q[Q_{a,\lambda}(g,\boldsymbol{r}_1),\ldots,Q_{a,\lambda}(g,\boldsymbol{r}_q)]| = O\left(\frac{\log^{2d(q-2)}(a)}{\lambda^{d(q-1)}}\right)$.

- For general sampling of the locations

$$\lambda^{d/2}\left[Q_{a,\lambda}(g,\mathbf{0}) - \langle\gamma,\gamma_0\rangle I\left(g;\frac{a}{\lambda}\right)\right] \stackrel{\mathcal{D}}{\to} \mathcal{N}(0,\Delta),$$

as $\log^2(a)/\lambda^{1/2} \to 0$ and $\lambda^d/n \to \infty$ as $a,\lambda,n \to \infty$[3]. Similar result holds for $\{Q_{a,\lambda}(g,\boldsymbol{r});\boldsymbol{r} \neq 0\}$.

[3] Recall

$$Q_{a,\lambda}(g;0) = \frac{1}{\lambda^d}\sum_{k_1,\ldots,k_d=-a}^{a} g(\boldsymbol{\omega}_{\boldsymbol{k}})|J_n(\boldsymbol{\omega}_{\boldsymbol{k}})|^2 - \text{bias} \qquad .$$

- It is likely that similar results hold for non-Gaussian random fields (under suitable mixing conditions); possibly by using some of the methods developed by Soumendra Lahiri and co-authors. However, in this case additional constraints are required on the growth of the unbounded frequency grid a.

- The result shows that for a given g, $Q_{a,\lambda}(g, \mathbf{0})$ consistently estimates $\langle \gamma, \gamma_0 \rangle I(g; \frac{a}{\lambda})$ and the estimator is Gaussian.

  However, for both testing and the construction of CIs, the variance Δ is unknown and needs to be estimated.

Variance estimation

- During the 1930's Fisher introduced the notion of an ancillary variable. These are statistics whose distribution does not depend on the parameter of interest. For example in the case of Gaussian random variables $X_i - \bar{X}$ are ancillary to the mean μ.

- In time series and spatial data ancillary variables are rarely used. However, under certain conditions the Fourier transform can be asymptotically treated as if it were ancillary.

- To do so, we return to the CLT of $Q_{a,\lambda}(g; \boldsymbol{r})$ in the case that the sampling of the locations are *uniformly* distributed.

- In the case that the sampling is uniform for $\boldsymbol{r} \neq 0$ the real and imaginary $\{Q_{a,\lambda}(g;\boldsymbol{r})\}$ "asymptotically" have zero mean and are uncorrelated.

- If $Q_{a,\lambda}(g;0)$ is real and $\boldsymbol{r}_1,\ldots,\boldsymbol{r}_m \in \mathbb{Z}^d/\{0\}$ where $\boldsymbol{r}_i \neq -\boldsymbol{r}_j$, then

$$\frac{1}{C_{a/\lambda}} \begin{pmatrix} \left[Q_{a,\lambda}(g;0) - I\left(g;\frac{a}{\lambda}\right)\right] \\ \sqrt{2}\Re Q_{a,\lambda}(g;\boldsymbol{r}_1) \\ \vdots \\ \sqrt{2}\Re Q_{a,\lambda}(g;\boldsymbol{r}_m) \\ \sqrt{2}\Im Q_{a,\lambda}(g;\boldsymbol{r}_1) \\ \vdots \\ \sqrt{2}\Im Q_{a,\lambda}(g;\boldsymbol{r}_m) \end{pmatrix} \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(0, I_{2m+1}\right).$$

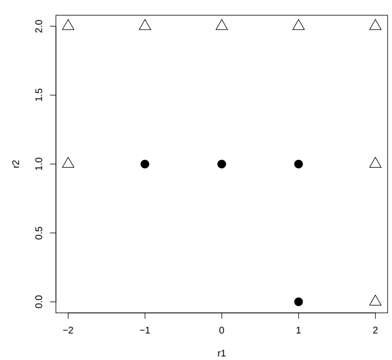
- Thus for $\boldsymbol{r}_i << \lambda$ $Q_{a,\lambda}(g;0)$, $\sqrt{2}\Re Q_{a,\lambda}(g;\boldsymbol{r}_j)$ and $\sqrt{2}\Im Q_{a,\lambda}(g;\boldsymbol{r}_j)$ asymptotically have the same variance.

Constructing asymptotically pivotal quantities

- Therefore the variables $\{\Re Q_{a,\lambda}(g;\boldsymbol{r}_j), \Im Q_{a,\lambda}(g;\boldsymbol{r}_j); j = 1,\ldots,m\}$ can be considered as ancillary to $Q_{a,\lambda}(g;0)$; they contain no information about the parameter $I(g;\frac{a}{\lambda})$ but they do contain information about the variance.

- Estimating the variance $C_{a/\lambda}$

$$\widehat{\sigma}^2_{2m} = \frac{\lambda^d}{m}\sum_{j=1}^{m}\left|Q_{a,\lambda}(g;\boldsymbol{r}_j)\right|^2.$$



The same estimator consistly estimates the variance in the case of non-Gaussian random fields.

- Using that

$$\frac{1}{C_{a/\lambda}}\begin{pmatrix}\left[Q_{a,\lambda}(g;0)-I\left(g;\frac{a}{\lambda}\right)\right]\\ \sqrt{2}\Re Q_{a,\lambda}(g;\boldsymbol{r}_1)\\ \vdots\\ \sqrt{2}\Re Q_{a,\lambda}(g;\boldsymbol{r}_m)\\ \sqrt{2}\Im Q_{a,\lambda}(g;\boldsymbol{r}_1)\\ \vdots\\ \sqrt{2}\Im Q_{a,\lambda}(g;\boldsymbol{r}_m)\end{pmatrix}\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,I_{2m+1}\right).$$

we can construct the asymptotically pivotal statistic

$$\lambda^{d/2}\left(\frac{Q_{a,\lambda}(g;0)-I(g;\frac{a}{\lambda})}{\sqrt{\widehat{\sigma}^2_{2m}}}\right)\xrightarrow{\mathcal{D}}t_{2m}.$$

Which can be used for testing and the construction of CIs.

A flavour of proofs

Focus on the case of uniform sampling and dimension $d = 1$.

- We recall

$$\mathrm{cov}\left[J_n(\omega_{k_1}), J_n(\omega_{k_2})\right] = \begin{cases} f(\omega_{k_1}) + O(\frac{1}{\lambda} + \frac{\lambda}{n}) & k_1 = k_2 \\ O(\frac{1}{\lambda}) & k_1 \neq k_2 \end{cases}$$

- Returning to the statistic

$$Q_{a,\lambda}(g; r) = \frac{1}{\lambda} \sum_{k=-a}^{a} g(\omega_k) J_n(\omega_k) \overline{J_n(\omega_{k+r})} - \mathrm{bias}.$$

Aim to show:

$$\mathbf{E}\left[\widetilde{Q}_{a,\lambda}(g; \boldsymbol{r})\right] = \begin{cases} O(\frac{\log \lambda}{\lambda}) & r \neq 0 \\ I\left(g; \frac{a}{\lambda}\right) + O\left(\frac{\log \lambda}{\lambda} + \frac{1}{n}\right) & r = 0 \end{cases}.$$

- If the frequency grid is bounded ($a = O(\lambda)$); we can immediately apply the DFT result.

- If the frequency grid is unbounded $a >> \lambda$; the errors $O(\lambda^{-1})$ accumulant.

- In this case we have to expand $J_n(\omega_k)$
  $Q_{a,\lambda}(g;r) = \frac{1}{n^2}\sum_{j_1 \neq j_2 = 1}^{n} \sum_{k=-a}^{a} g(\omega_k) Z(s_{j_1}) Z(s_{j_2}) e^{i\omega_k(s_{j_1} - s_{j_2})} e^{-i\omega_r s_{j_2}}$.

Taking expectations and exploit the randomness of locations:

$$\begin{aligned} \mathbf{E}\left[Q_{a,\lambda}(g;r)\right] &\approx \sum_{k=-a}^{a} g(\omega_k) \mathbf{E}\left[c(s_1 - s_2)\exp(i\omega_k(s_1 - s_2) - is_2\omega_r)\right] \\ &\approx \frac{1}{\lambda^2} \sum_{k=-a}^{a} g(\omega_k) \int_{-\lambda/2}^{\lambda/2} \int_{-\lambda/2}^{\lambda/2} c(s_1 - s_2) e^{i\omega_k(s_1 - s_2) - is_2\omega_r} ds_1 ds_2. \end{aligned}$$

Using the standard trick in time series, we replace $c(s_1 - s_2)$ with the Fourier representation of the covariance function

$$
\begin{aligned}
& \mathbf{E}\left[Q_{a,\lambda}(g;r)\right] \\
\approx \ & \frac{1}{\lambda^2}\sum_{k=-a}^{a} g(\omega_k)\int_{\mathbb{R}} f(\omega)\int_{-\lambda/2}^{\lambda/2}\int_{-\lambda/2}^{\lambda/2} e^{i(s_1-s_2)\omega} e^{i\omega_k(s_1-s_2)-is_2\omega_r} ds_1 ds_2 d\omega \\
= \ & \frac{1}{2\pi}\sum_{k=-a}^{a} g(\omega_k)\int_{-\infty}^{\infty} f(\omega)\mathrm{sinc}\left(\frac{\lambda\omega}{2}+k\pi\right)\mathrm{sinc}\left(\frac{\lambda\omega}{2}+(k+r)\pi\right) d\omega \\
= \ & \frac{1}{\pi}\int_{-\infty}^{\infty}\mathrm{sinc}(y)\mathrm{sinc}(y+r\pi)\int_{-2\pi a/\lambda}^{2\pi a/\lambda} g(u) f\left(\frac{2y}{\lambda}-u\right) du dy + O\left(\frac{1}{\lambda}\right)
\end{aligned}
$$

- $\mathrm{sinc}(y) = \sin(y)/y$. Using the griding of $\omega_k = \frac{2\pi k}{\lambda}$ in the construction of $Q_{a,\lambda}$ leads to the orthonormal functions $\{\frac{1}{\sqrt{\pi}}\mathrm{sinc}(y+r\pi); r \in \mathbb{Z}\}$.

- Under the assumption that f and f' are absolutely integrable we replace $f(\frac{2y}{\lambda} - u)$ with $f(-u)$. This separates the two integrals:

$$
\begin{aligned}
& \mathbf{E}\left[Q_{a,\lambda}(g;r)\right] \\
= \; & \underbrace{\frac{1}{\pi}\int_{-\infty}^{\infty} \mathrm{sinc}(y)\mathrm{sinc}(y+r\pi)dy}_{0 \text{ if } r\neq 0 \text{ else } 1} \int_{-2\pi a/\lambda}^{2\pi a/\lambda} g(u) f(-u) du \\
& \underbrace{+\frac{1}{\pi}\int_{-\infty}^{\infty} \mathrm{sinc}(y)\mathrm{sinc}(y+r\pi) \int_{-2\pi a/\lambda}^{2\pi a/\lambda} g(u)\left\{ f(\frac{2y}{\lambda} - u) - f(-u)\right\} du dy}_{O\left(\frac{\log\lambda + I_{r\neq 0}\log|r|}{\lambda}\right)} .
\end{aligned}
$$

- Essentially the same method can be applied to the covariance and higher order cumulants. The general recipe is:

(a) $\{Z(\boldsymbol{s}_j)\}$ is a composite random variable in terms the process and random location. Thus even if the process is Gaussian, $Z(\boldsymbol{s}_j)$ is not Gaussian.

    We use Brillinger (1969) to condition on the locations $\boldsymbol{s}_j$, which allows us to reduce covariance and cumulants of $\{Z(\boldsymbol{s}_j)\}$ in terms the covariance and cumulants of the underlying spatial process. Duplicate locations can cause a lot of problems, especially when the process is non-Gaussian.

(b) Represent the spatial covariance and cumulants in terms of spectral densities or higher order spectral densities; this separates out the random locations.

(c) Integrating out the density of the locations gives rise to products of sinc functions (similar to the Dirichlet kernel) which satisfy several orthogonality and self-similar properties

$$\frac{1}{\pi}\int_{\mathbb{R}} \mathrm{sinc}(u)\mathrm{sinc}(u+r\pi)du = \begin{cases} 1 & r=0 \\ 0 & r\neq 0 \end{cases}$$

$$\frac{1}{\pi}\int_{\mathbb{R}} \mathrm{sinc}(u)\mathrm{sinc}(u+v)du = \mathrm{sinc}(v).$$

- This allows us project everything onto a basis of sinc functions. Finding the approximation errors is the difficult part.

  In many ways this generalizes the results of Bochner and Kawata (1958)/Kawata (1959) who considered Fourier transforms of continuous time processes (observed over continuous time).