A Frequency Domain Empirical Likelihood Method for Irregularly Spaced Dependent Data

Dan Nordman

Department of Statistics

Iowa State University

dnordman@iastate.edu

Joint work with Soutir Bandyopadhyay & Soumen Lahiri

(Lehigh University & North Carolina State University)

Introduction

EL: likelihood-based inference without a full distribution for data

- method formulates a likelihood function nonparametrically & has properties similar to a parametric likelihood function (e.g., chi-squared limits)
- has parallels to bootstrap, but probability profiles data rather than resampling data
- can calibrate nonparametric confidence intervals/tests for parameters

Goal: Develop an EL for dependent data with irregular/random sampling locations using a data transformation (Fourier transform)

Overview of talk on Empirical Likelihood (EL)

1. Review of EL basics
2. Recall a Frequency Domain EL for stationary (equi-spaced) time series
3. Describe a sampling design for irregularly spaced dependent data
4. Provide a Frequency Domain EL method for such data & distributional results
5. Discuss some applications

All data will be real-valued for simplicity

Empirical Likelihood (EL) for iid data - Owen (1990)

Usually, EL is a *multinomial likelihood on the sample*: $X_1, \ldots, X_n$

$$\text{consider likelihood: } \prod_{i=1}^{n} p_i, \text{ assigning } p_i \longmapsto \text{each } X_i, \ \sum_{i=1}^{n} p_i = 1$$

For inference, probability profile data under an expectation constraint

EL for mean: To assess μ as mean of $X_1, \ldots, X_n \in \mathbb{R}$, EL function is

$$L_n(\mu) = \sup \left\{ \prod_{i=1}^{n} p_i : p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i (X_i - \mu) = 0 \right\}$$

probabilities on data with mean μ

$\prod_{i=1}^{n} p_i$ maximized at $p_i = n^{-1}$ w/o expectation constraint

$\Rightarrow$ EL ratio $R_n(\mu) = L_n(\mu) / \prod_{i=1}^{n} n^{-1} \in [0, 1]$

EL for iid data: mean case $\mathrm{E}(X_i) = \mu \in \mathbb{R}$, continued

Owen (1988, 1990) established a Wilk's theorem for EL:

$$\text{at true } \mu_0 \in \mathbb{R}, \quad -2\log R_n(\mu_0) \xrightarrow{d} \chi^2_1 \quad \text{as } n \to \infty$$

$\Rightarrow$ EL Confidence Regions for μ, $\{\mu : -2\log R_n(\mu) \leq \chi^2_{1,\alpha}\}$

1. nice properties in region shapes, accuracies comparable to bootstrap
2. EL requires *no* variance estimates/studentization/pivots

But, EL formulation for independent data generally **fails** under dependence

One approach for extending EL to time series: Frequency Domain EL
i.e., use Fourier transform to whiten data (N. & Lahiri, 2006)

Frequency Domain EL (FDEL) for Time Series

For $X_1, \ldots, X_n$ from stationary $\{X_t\}$ with spectral density $\phi(\cdot)$, FDEL uses

1. approximate independence of discrete Fourier transform or

$$\text{periodogram} \;\; I_n(\omega) = \frac{1}{2\pi}\left|\sum_{t=1}^{n} X_t e^{-\imath t\omega}\right|^2, \quad \imath = \sqrt{-1}$$

   at Fourier frequencies $\omega_j = 2\pi j/n$, $j = 1, \ldots, N = \lfloor (n-1)/2 \rfloor$

2. p spectral estimating functions $G_\theta(\omega) : \mathbb{R}^p \times [-\pi, \pi] \to \mathbb{R}^p$ to provide info about $\theta \in \mathbb{R}^p$

$$\int_{-\pi}^{\pi} G_\theta(\omega)\phi(\omega)\, d\omega = 0_p$$

   e.g., Autocorrelations, Spectral Distributions, Whittle parameters, namely, (normalized) spectral mean parameters targeted by Frequency Domain Bootstrap (Dahlhaus & Janas, 1996)

Frequency Domain EL (FDEL) for Time Series

To assess $\theta \in \mathbb{R}^p$, FDEL function profiles to mimic spectral expectation:

$$L_n(\theta) = \sup\left\{ \prod_{j=-N}^{N} p_j : p_j \geq 0, \sum_{j=-N}^{N} p_j = 1, \sum_{j=-N}^{N} 2\pi p_j G_\theta(\omega_j) I_n(\omega_j) = 0_p \right\}$$

Ratio $R_n(\theta) = L_n(\theta) / \prod_{j=-N}^{N} (2N)^{-1}$

For stationary, linear $\{X_t\}$, the FDEL log-ratio satisfies Wilk's theorem:

at true $\theta_0 \in \mathbb{R}^p$, $\quad -2 \log R_n(\theta_0) \xrightarrow{d} \chi^2_p \quad$ as $n \to \infty$

Monti (1997): Whittle estimation under weak dependence

N. & Lahiri (2006): general estimating functions under weak-/long-memory

Frequency Domain EL (FDEL) for Irregularly Spaced Data

Rather equi-spaced time series, we wish to develop a FDEL

- for irregularly spaced dependent data sampled in $\mathbb{R}^d$
  (e.g., $d = 1$ for time series, $d > 1$ for spatial data)

  a general class of sampling designs/second-order stationary processes

- based on the periodogram & a framework of spectral estimating functions

- preserving the Wilk's phenomenon

For inference about such processes, this FDEL has an advantage in requiring no variance estimation steps (i.e., utterly difficult directly)

Challenges for Irregularly Spaced Data

In contrast to equi-spaced time series case, we now have

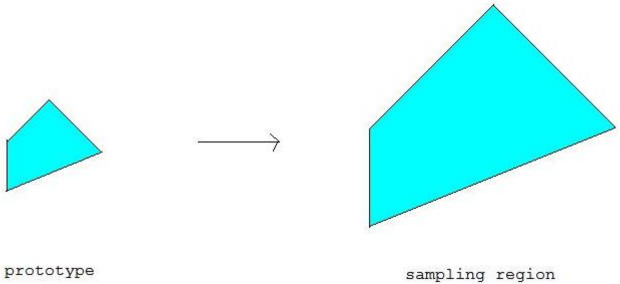
1. the basic orthogonality property of the sine- and cosine-transforms of gridded data at Fourier frequencies (e.g., $\omega_j = 2\pi j/n$) no longer holds.
2. one must deal with an unbounded frequency domain $\mathbb{R}^d$.
3. the periodogram of irregularly spaced data can be severely biased & must be pre-processed for inference on the underlying spectral density.
4. more than one asymptotic structure can arise, depending on the growth rate of the sample size n relative to the volume of a sampling region $\mathcal{D}_n \subset \mathbb{R}^d$.

Asymptotic Framework: Sampling Region

To allow sampling regions of differing shapes,

- let $\mathcal{D}_0 \subset (-1/2, 1/2]^d$ be an open connected set, containing the origin
- let $\{\lambda_n\}_{n \geq 1}$ be a positive real sequence where $\lambda_n \to \infty$ as $n \to \infty$.
- define a sampling region $\mathcal{D}_n$ by 'inflating' prototype $\mathcal{D}_0$ by a factor λ_n

$$\mathcal{D}_n = \lambda_n \mathcal{D}_0$$



Asymptotic Framework: Sampling Design

- Let $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$ be a zero mean second-order stationary random field

- Data $Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)$ are observed at sites $\{\mathbf{s}_i\}_{i=1}^n \subset \mathcal{D}_n = \lambda_n \mathcal{D}_0$, within the sampling region, obtained by

$$\mathbf{s}_i \equiv \mathbf{s}_{in} = \lambda_n \mathbf{X}_i, \quad 1 \leq i \leq n.$$

where $\{\mathbf{X_i}\}_{\mathbf{i}\geq\mathbf{1}}$ be iid $\sim f(\mathbf{x})$ ($\mathbf{X_i}$'s take values in $\mathcal{D}_0$)

- (unknown) $f(\mathbf{x})$ is continuous, positive probability density on $\mathcal{D}_0 \subset \mathbb{R}^d$

Possible Asymptotic Structures

1. Pure-increasing domain asymptotics (PID): (similar to gridded data)

   sampling region $\mathcal{D}_n$ expands as sample size $n \to \infty$, with $n \propto \mathrm{vol}(\mathcal{D}_n)$

   $$\text{or } n/\lambda_n^d \to c^* \in (0, \infty) \text{ as } n \to \infty$$

2. Mixed-increasing domain asymptotics (MID):

   expanding $\mathcal{D}_n$ but with a heavy infill of sampling sites: $\mathrm{vol}(\mathcal{D}_n) \ll n$

   $$\text{or } n/\lambda_n^d \to \infty \equiv c^* \text{ as } n \to \infty$$

OBSERVATION: These differing structures impact even simple statistics

$$\lambda_n^{d/2} \bar{Z}_n = \frac{\lambda_n^{d/2}}{n} \sum_{i=1}^{n} Z(\mathbf{s}_i) \xrightarrow{d} N\left(0, \ \sigma(\mathbf{0})/c^* + \int_{\mathbb{R}^d} \sigma(\mathbf{s}) d\mathbf{s}\right)$$

$\sigma(\mathbf{s}) = \mathrm{Cov}[Z(\mathbf{0}), Z(\mathbf{s})]$ (Lahiri, 2003)

Formulation of the FDEL method

Suppose

- $\phi(\boldsymbol{\omega})$, $\boldsymbol{\omega} \in \mathbb{R}^d$ is the spectral density of $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$
- parameter of interest $\theta \in \Theta \subset \mathbb{R}^p$ satisfies estimating equations:

$$\int_{\mathbb{R}^d} G_\theta(\boldsymbol{\omega})\phi(\boldsymbol{\omega})d\boldsymbol{\omega} = 0_p,$$

based on p spectral estimating functions $G_\theta(\boldsymbol{\omega}) : \mathbb{R}^d \times \Theta \to \mathbb{R}^p$

Some examples next

Example 1: Autocorrelations

Let $\sigma(\cdot)$ & $\rho(\cdot)$ be auto-covariance and -correlation functions of $Z(\cdot)$

For a given set of lags $\mathbf{h}_1, \cdots, \mathbf{h}_p \in \mathbb{R}^d$, define parameter of interest

$$\theta = (\rho(\mathbf{h}_1), \cdots, \rho(\mathbf{h}_p))'$$

& estimating functions

$$G_\theta(\boldsymbol{\omega}) = \left(\cos(\boldsymbol{\omega}'\mathbf{h}_1), \cdots, \cos(\boldsymbol{\omega}'\mathbf{h}_p)\right)' - \theta, \ \boldsymbol{\omega} \in \mathbb{R}^d$$

Then, spectral equations hold

$$\int G_\theta(\boldsymbol{\omega})\phi(\boldsymbol{\omega})d\boldsymbol{\omega} = (\sigma(\mathbf{h}_1), \cdots, \sigma(\mathbf{h}_p))' - \sigma(\mathbf{0})\theta = 0_p.$$

Formulation of the FDEL: Examples

Example 2: Variogram Model Fitting

Let $\{2\gamma(\cdot;\theta) : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^p$ denote a class of variogram models for the (scale-invariant) variogram $2\gamma(\mathbf{h}) \equiv \mathrm{Var}(Z(\mathbf{h}) - Z(\mathbf{0}))/\mathrm{Var}(Z(\mathbf{0}))$, $\mathbf{h} \in \mathbb{R}^d$

Least squares variogram fitting (Cressie, 1993) corresponds to minimizing a population criterion $\sum_{i=1}^m \left(2\gamma(\mathbf{h}_i) - 2\gamma(\mathbf{h}_i;\theta)\right)^2$

Taking partial derivatives ∇, the true θ_0 solves (cf. Lahiri, 2002)

$$\sum_{i=1}^m \Big(2\gamma(\mathbf{h}_i) - 2\gamma(\mathbf{h}_i;\theta)\Big)\nabla[2\gamma(\mathbf{h}_i;\theta)] = 0,$$

or an equivalent spectral estimating equation:

$$\int \Big[\sum_{i=1}^m \Big\{1 - \cos(\mathbf{h}_i'\boldsymbol{\omega}) - \gamma(\mathbf{h};\theta)\Big\}\nabla[2\gamma(\mathbf{h}_i;\theta)]\Big]\phi(\boldsymbol{\omega})d\boldsymbol{\omega} = 0.$$

Formulation of the Spatial FDEL

To re-iterate,

- $\phi(\boldsymbol{\omega})$, $\boldsymbol{\omega} \in \mathbb{R}^d$ is the spectral density of $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$

- parameter of interest $\theta \in \Theta \subset \mathbb{R}^p$ satisfies estimating equations:

$$\int_{\mathbb{R}^d} G_\theta(\boldsymbol{\omega})\phi(\boldsymbol{\omega})d\boldsymbol{\omega} = 0_p,$$

  based on p spectral estimating functions $G_\theta(\boldsymbol{\omega}) : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^p$

- to next formulate a FDEL method for θ, we use

  the discrete Fourier transform (DFT) of $\{Z(\mathbf{s}_i)\}_{i=1}^n$ &

  a concept of approximate independence of DFTs with such data

Formulation of the Spatial FDEL: DFT Asymptotics

Definition

The scaled DFT of $\{Z(\mathbf{s}_j)\}_{j=1}^{n}$ is given by

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

and periodogram is $I_n(\boldsymbol{\omega}) = |d_n(\boldsymbol{\omega})|^2$, $\quad \imath = \sqrt{-1}$

- Under standard moment/mixing conditions, Bandyopadhyay & Lahiri (2009) established asymptotic normality properties of DFTs
- In particular, for sequences of frequencies $\{\boldsymbol{\omega}_{1n}\}$ & $\{\boldsymbol{\omega}_{2n}\} \subset \mathbb{R}^d$, DFTs $d_n(\boldsymbol{\omega}_{1n})$ & $d_n(\boldsymbol{\omega}_{2n})$ are asymptotically independent if and only if $\boldsymbol{\omega}_{1n}$ & $\boldsymbol{\omega}_{2n}$ are asymptotically distant, i.e.,

$$\lambda_n \|\boldsymbol{\omega}_{1n} - \boldsymbol{\omega}_{2n}\| \to \infty \quad \text{as } n \to \infty$$

Formulation of the FDEL: DFT Asymptotics

FDEL involves $I_n(\cdot)$ over a grid of asymptotically distant frequencies in $\mathbb{R}^d$

For $0 < \kappa < \eta < 1$, $C \in (0, \infty)$, define a grid of N frequencies as

$$\{\boldsymbol{\omega}_{kn}\}_{k=1}^{N} \equiv \left\{ \lambda_n^{-\kappa}\mathbf{j} : \mathbf{j} \in \mathbb{Z}^d \cap [-C\lambda_n^{\eta}, C\lambda_n^{\eta}]^d \right\}$$

1. Frequencies $\{\boldsymbol{\omega}_{kn}\}_{k=1}^{N}$ form a regular lattice over the hyper-cube $\left[-C\lambda_n^{\eta-\kappa}, C\lambda_n^{\eta-\kappa}\right]^d \uparrow \mathbb{R}^d$ as $n \to \infty$ & covering the frequency domain
2. Any frequency-pairs in $\mathcal{N}_n$ are asymptotically distant

$$\lambda_n \|\boldsymbol{\omega}_{jn} - \boldsymbol{\omega}_{kn}\| \geq \lambda_n^{(1-\kappa)} \to \infty$$

& corresponding DFTs are asymptotically independent.

Formulation of the FDEL: Final Construction

- OBSERVATION: Periodogram is asymptotically biased under PID:

$$\mathrm{E} I_n(\omega) \to c_*^{-1}\mathrm{Var}(Z(\mathbf{0})) + K\phi(\boldsymbol{\omega}),\ \boldsymbol{\omega} \in \mathbb{R}^d, \quad K = (2\pi)^d \int f^2$$

  where $n/\lambda_n^d \to c^* \in (0,\infty)$ under PID (& $c* = \infty$ under MID)

- To define EL here, we use the bias corrected periodogram

$$\tilde{I}_n(\boldsymbol{\omega}) = I_n(\boldsymbol{\omega}) - n^{-1}\lambda_n^d S_Z^2 \quad (S_Z^2 \text{ is sample variance of } \{Z(\mathbf{s}_i)\}_{i=1}^N)$$

- With frequency grid/estimating functions, FDEL function for $\theta \in \mathbb{R}^p$:

$$L_n(\theta) = \sup\left\{\prod_{k=1}^N p_k : \sum_{k=1}^N p_k = 1, p_k \geq 0, \sum_{k=1}^N p_k G_\theta(\boldsymbol{\omega}_{kn})\tilde{I}_n(\boldsymbol{\omega}_{kn}) = 0_p\right\}.$$

  FDEL ratio $R_n(\theta) = L_n(\theta)/(N^{-1})^N$.

Main Results for the FDEL

Wilk's result (sort of) holds at true θ_0 with 3 versions, depending on

$$c_n = n/\lambda^d \quad \text{\& the volume } N\lambda_n^{-\kappa d} \text{ of the frequency grid}$$

PID: $c_n \longrightarrow c_* \in (0, \infty)$

$$-\log R_n(\theta_0) \xrightarrow{d} \chi^2_p \quad \text{as } n \to \infty \text{ a.s. } (P_{\mathbf{X}}).$$

MID with slow infill: $c_n \longrightarrow \infty$ with $c_n^2 \ll N\lambda_n^{-\kappa d}$

$$-\log R_n(\theta_0) \xrightarrow{d} \chi^2_p \quad \text{as } n \to \infty \text{ a.s. } (P_{\mathbf{X}}).$$

MID with fast infill: $c_n \longrightarrow \infty$ with $N\lambda_n^{-\kappa d} \ll c_n^2$

$$-2\log R_n(\theta_0) \xrightarrow{d} \chi^2_p \quad \text{as } n \to \infty \text{ a.s. } (P_{\mathbf{X}}).$$

Main Results for the FDEL: a modified FDEL

- Previous results show the standard calibration $-2\log(\cdot)$ of EL ratio statistic may be incorrect depending on the rate of infilling

- The choice of the correct scaling constant and, hence, the correct calibration may not be obvious in a finite sample application

- Define a modified (self-adjusting) FDEL ratio: $-2a_n(\theta_0)\log R_n(\theta_0)$

for $a_n(\theta) = \dfrac{\sum_{k=1}^{N} \|G_\theta(\omega_{kn})\|^2 \tilde{I}_n^2(\omega_{kn})}{\sum_{k=1}^{N} \|G_\theta(\omega_{kn})\|^2 I_n^2(\omega_{kn})}$ a ratio of biased corrected & uncorrected periodograms

In all previous cases, at true θ_0

$$-2a_n(\theta_0)\log R_n(\theta_0) \xrightarrow{d} \chi_p^2 \quad \text{as } n \to \infty \text{ a.s. } (P_{\mathbf{X}}).$$

Application Examples: Variogram model fitting

Variogram model: $2\gamma(\mathbf{h};\theta_1,\theta_2) = 1 - \exp[-\theta_1|h_1| - \theta_2|h_2|]$, $\mathbf{h} = (h_1, h_2)'$

- Gaussian r.f. $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{Z}^2\}$, $\theta_1 = \theta_2 = 1$
- iid uniform sites $n = 100, 400, 900, 1400$; region $\mathcal{D}_n = \lambda_n[-1/2, 1/2)^2$
- 90% modified FDEL regions for (θ_1, θ_2) using variogram estimating functions with $\mathbf{h}_1 = (1,1)', \mathbf{h}_2 = (1,-1)'$
- Frequency grid: $\{\lambda_n^{-\kappa}\mathbf{j} : \mathbf{j} \in \mathbb{Z}^2 \cap [-C\lambda_n, C\lambda_n]^2\}$, varying C, κ

		$\lambda_n = 12$				$\lambda_n = 24$			
C	κ	100	400	900	1400	100	400	900	1400
1	0.05	88.4	87.9	81.8	82.8	89.7	90.6	88.6	90.8
1	0.1	89.0	87.1	83.3	80.2	88.8	89.7	89.0	89.5
1	0.2	88.5	86.1	83.7	81.1	90.2	90.5	89.5	89.7
2	0.05	89.1	88.9	88.2	86.5	89.4	89.5	90.4	91.0
2	0.1	90.0	88.9	89.2	87.2	87.7	90.3	91.3	89.9
2	0.2	89.3	89.8	86.3	85.9	89.3	90.5	88.6	88.6
4	0.05	90.7	89.4	90.9	89.2	89.6	89.5	89.5	89.8
4	0.1	88.7	89.5	90.9	89.1	90.0	88.9	90.2	89.6
4	0.2	89.4	89.8	90.6	90.0	89.7	88.4	89.2	90.8

Applications: Whittle Estimation

- Let $\{\phi_\theta : \theta \in \Theta \subset \mathbb{R}^p\}$ be a parametric family of spectral density functions.

- Matsuda and Yajima (2009) introduced a version of the Whittle likelihood for Gaussian processes

$$\mathcal{L}_W(\theta) = \int_D \left[\log\left(\tilde{\phi}_\theta(\boldsymbol{\omega})\right) + \frac{I_n(\boldsymbol{\omega})}{\tilde{\phi}_\theta(\boldsymbol{\omega})} \right] d\boldsymbol{\omega}$$

where $\tilde{\phi}_\theta(\boldsymbol{\omega}) = \phi_\theta(\boldsymbol{\omega}) + n^{-1}\lambda_n^d \sigma(\mathbf{0};\theta)$, $D \in \mathbb{R}^d$ is a bounded domain.

- Under some standard smoothness conditions on ϕ_θ, this leads to estimating equations and a **Whittle estimator** $\hat{\theta}_{n,W}$ of θ.

Applications: Whittle Estimation (contd.)

- Matsuda and Yajima (2009) proved that, under MID conditions,

$$\lambda_n^{d/2}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} N(0, 2b\Gamma^{-1}(\theta_0))$$

with $b = (\int f^4)(\int f^2)^{-2}$ and

$$\Gamma(\theta_0) = (2\pi)^{-d} \int_D (\nabla \log(\phi_{\theta_0}(\boldsymbol{\omega})))(\nabla \log(\phi_{\theta_0}(\boldsymbol{\omega})))' \, d\boldsymbol{\omega}$$

- Thus one needs to further estimate the asymptotic variance to construct the confidence regions for θ_0.

- In contrast, the FDEL with the similar estimating functions yields a valid confidence region, under PID or MID,
without requiring estimation of the asymptotic variance.

Applications: Dependence structure assessments

- We discussed FDEL with p parameters & p estimating functions
- But, using $r > p$ estimating functions & p parameters, one can maximize the FDEL ratio $R_n(\theta)$ to obtain $\hat{\theta}_n$ and use

$$-2a_n(\hat{\theta}_n) \log R_n(\hat{\theta}_n) \xrightarrow{d} \chi^2_{r-p} \quad \text{as } n \to \infty \text{ a.s. } (P_{\mathbf{X}})$$

to test H_0 : moment $\int_{\mathbb{R}^d} G_{\theta_0}(\boldsymbol{\omega})\phi(\boldsymbol{\omega})d\boldsymbol{\omega} = 0_r$ holds, some θ_0

- **This allows tests of**

1. good-of-fit or model assessment

(e.g., functions based on variogram model fitting or Whittle estimation)

2. dependence structure such as spatial isotropy or separability

(e.g., estimating functions based on correlations or spectral distribution)