Bayesian nonparametrics for time series

focustat
FOCUS DRIVEN STATISTICAL
INFERENCE WITH COMPLEX DATA

Gudmund Horn Hermansen with Nils Lid Hjort

Vårens Vakreste Variabler Conference

May 24, 2018

Develop Bayesian nonparametric framework for estimation and inference for stationary Gaussian time series with expectation zero, i.e.

$$(Y_1, \dots, Y_n) \sim \mathrm{N}(0, \Sigma(f)), \tag{1}$$

where f is the spectral density, and where P_n refers to the corresponding measure.

Introduction and summary

Develop Bayesian nonparametric framework for estimation and inference for stationary Gaussian time series with expectation zero, i.e.

$$(Y_1, \ldots, Y_n) \sim \mathrm{N}(0, \Sigma(f)), \tag{1}$$

where f is the spectral density, and where P_n refers to the corresponding measure.

In summary:

(i) develop a family of priors for the covariance functions, and

(ii) establish Bernshteĭn–von Mises type of results for this family of priors.

Introduction and summary

Develop Bayesian nonparametric framework for estimation and inference for stationary Gaussian time series with expectation zero, i.e.

$$(Y_1, \ldots, Y_n) \sim \mathrm{N}(0, \Sigma(f)), \tag{1}$$

where f is the spectral density, and where P_n refers to the corresponding measure.

In summary:

(i) develop a family of priors for the covariance functions, and

(ii) establish Bernshteĭn–von Mises type of results for this family of priors.

The main part of this talk will be used to motivate:

(a) an alternative contiguous measure P_n^* to P_n above, and

(b) a class of growing (so-called) piecewise constant priors.

Introduction and summary

Develop Bayesian nonparametric framework for estimation and inference for stationary Gaussian time series with expectation zero, i.e.

$$(Y_1, \ldots, Y_n) \sim \mathrm{N}(0, \Sigma(f)), \tag{1}$$

where f is the spectral density, and where P_n refers to the corresponding measure.

In summary:

(i) develop a family of priors for the covariance functions, and

(ii) establish Bernshteĭn–von Mises type of results for this family of priors.

The main part of this talk will be used to motivate:

(a) an alternative contiguous measure P_n^* to P_n above, and

(b) a class of growing (so-called) piecewise constant priors.

We will use (b) to break the main argument into smaller components, and (a) will be used to simplify the techniques and conditions needed.

The plan

1) Assumptions and basic notation for stationary time series
2) Summary of parametric Bernshteĭn–von Mises
3) Parametric models and the Whittle log-likelihood
4) Contiguity
5) Parametric Bernshteĭn–von Mises
6) Nonparametric modelling
7) Prior specification
8) Main conditions and for nonparametric Bernshteĭn–von Mises
9) Illustration

The main model, and some notation and conditions

Let $Y_1, \ldots, Y_n$ be a stationary Gaussian time series with covariance matrix Σ specified by the elements

$$\Sigma_{k,l} = \mathrm{Cov}(Y_k, Y_l) = C(|k-l|), \text{ for } k, l = 1, \ldots, n.$$

The main model, and some notation and conditions

Let $Y_1, \dots, Y_n$ be a stationary Gaussian time series with covariance matrix Σ specified by the elements

$$\Sigma_{k,l} = \mathrm{Cov}(Y_k, Y_l) = C(|k-l|), \text{ for } k, l = 1, \dots, n.$$

If the covariance function is absolute summable, i.e. if

$$\sum_{h=0}^{\infty} |C(h)| < \infty,$$

then there exist a spectral density f defined by

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-i\omega h} C(h) = \frac{C(0)}{2\pi} + \frac{1}{\pi} \sum_{h=1}^{\infty} \cos(\omega h) C(h),$$

The main model, and some notation and conditions

Let $Y_1, \ldots, Y_n$ be a stationary Gaussian time series with covariance matrix Σ specified by the elements

$$\Sigma_{k,l} = \mathrm{Cov}(Y_k, Y_l) = C(|k-l|), \text{ for } k, l = 1, \ldots, n.$$

If the covariance function is absolute summable, i.e. if

$$\sum_{h=0}^{\infty} |C(h)| < \infty,$$

then there exist a spectral density f defined by

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-i\omega h} C(h) = \frac{C(0)}{2\pi} + \frac{1}{\pi} \sum_{h=1}^{\infty} \cos(\omega h) C(h),$$

Furthermore, we have that

$$C(h) = C_f(h) = \int_{-\pi}^{\pi} e^{-i\omega h} f(\omega)\,\mathrm{d}\omega = 2\int_{0}^{\pi} \cos(\omega h) f(\omega)\,\mathrm{d}\omega,$$

and we will sometimes write Σ_f or $\Sigma(f)$ to make this connection clear.

Wold's Theorem and a simple example

Wold's Theorem: A function $C(h)$ is a covariance function for some real valued stationary process $\{Y_t\}$ if and only if there exists a positive non-decreasing and bounded function F on the interval $[0, \pi)$ such that

$$C(h) = 2\int_0^{\pi} \cos(\omega h)\, \mathrm{d}F(\omega), \qquad \text{for all } h \geq 0.$$

Wold's Theorem and a simple example

Wold's Theorem: A function $C(h)$ is a covariance function for some real valued stationary process $\{Y_t\}$ if and only if there exists a positive non-decreasing and bounded function F on the interval $[0, \pi)$ such that

$$C(h) = 2 \int_0^{\pi} \cos(\omega h) \, \mathrm{d}F(\omega), \qquad \text{for all } h \geq 0.$$

Wold's Theorem (above) has a central role in the prior construction.

Wold's Theorem and a simple example

Wold's Theorem: A function $C(h)$ is a covariance function for some real valued stationary process $\{Y_t\}$ if and only if there exists a positive non-decreasing and bounded function F on the interval $[0, \pi)$ such that

$$C(h) = 2\int_0^\pi \cos(\omega h)\,\mathrm{d}F(\omega), \qquad \text{for all } h \geq 0.$$

Wold's Theorem (above) has a central role in the prior construction.

Example: For an autoregressive model of order one, i.e.

$$Y_t = \rho Y_{t-1} + \sigma\epsilon_t, \text{ where } |\rho| < 1.$$

and ϵ_t are i.i.d. and $\epsilon_t \sim N(0, 1)$. Then, the spectral density is

$$f_{\sigma,\rho}(\omega) = \frac{\sigma^2}{2\pi(1 - 2\rho\cos\omega + \rho^2)} \tag{2}$$

for $\sigma > 0$ and $|\rho| < 1$, and the covariance function $C(h) = \sigma^2\rho^h/(1 - \rho^2)$.

What and how to place priors on covariances

The only part of the model (above) that is not specified is the covariance or spectral density.

What and how to place priors on covariances

The only part of the model (above) that is not specified is the covariance or spectral density.

Not easy to define priors on the space of covariance functions directly.

What and how to place priors on covariances

The only part of the model (above) that is not specified is the covariance or spectral density.

Not easy to define priors on the space of covariance functions directly.

The reason being that the resulting Σ has to be positive-semidefinite, i.e.

$$a^t \Sigma a \geq 0 \text{ for all vectors } a \in \mathbb{R}^n.$$

What and how to place priors on covariances

The only part of the model (above) that is not specified is the covariance or spectral density.

Not easy to define priors on the space of covariance functions directly.

The reason being that the resulting Σ has to be positive-semidefinite, i.e.

$$a^{t}\Sigma a \geq 0 \text{ for all vectors } a \in \mathbb{R}^n.$$

From Wold's Theorem (above), all positive non-decreasing and bounded function F on $[0, \pi]$ will generate a valid covariance functions and matrix.

What and how to place priors on covariances

The only part of the model (above) that is not specified is the covariance or spectral density.

Not easy to define priors on the space of covariance functions directly.

The reason being that the resulting Σ has to be positive-semidefinite, i.e.

$$a^{\mathrm{t}}\Sigma a \geq 0 \text{ for all vectors } a \in \mathbb{R}^n.$$

From Wold's Theorem (above), all positive non-decreasing and bounded function F on $[0, \pi]$ will generate a valid covariance functions and matrix.

And it is much easier to construct priors on the space of such functions.

What and how to place priors on covariances

The only part of the model (above) that is not specified is the covariance or spectral density.

Not easy to define priors on the space of covariance functions directly.

The reason being that the resulting Σ has to be positive-semidefinite, i.e.

$$a^{\mathrm{t}}\Sigma a \geq 0 \text{ for all vectors } a \in \mathbb{R}^n.$$

From Wold's Theorem (above), all positive non-decreasing and bounded function F on $[0, \pi]$ will generate a valid covariance functions and matrix.

And it is much easier to construct priors on the space of such functions.

The general workflow can be summarised by:

$$\begin{array}{ccc} \pi(C(\cdot)) & & \pi(C(\cdot)\,|\,\text{data}) \\ \downarrow & & \uparrow \\ \pi(F(\cdot)) & \longrightarrow & \pi(F(\cdot)\,|\,\text{data}) \end{array}$$

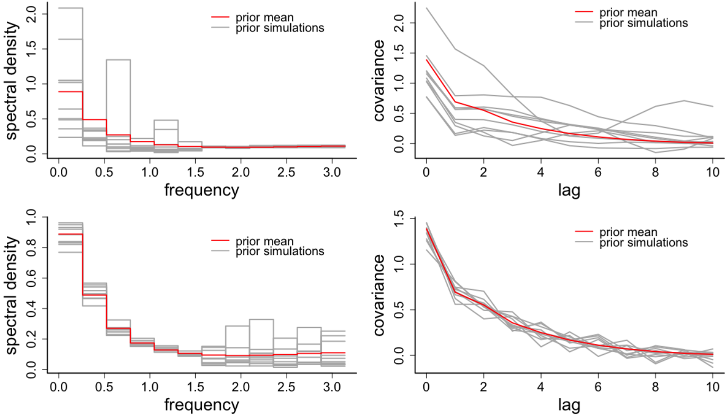
Why nonparametric Bayes?

Subjective and natural prior knowledge for the covariance is not (I suppose) so common.

Why nonparametric Bayes?

Subjective and natural prior knowledge for the covariance is not (I suppose) so common.

Therefore, we think of nonparametric priors as placing a nonparametric envelope around a parametric family.



Note, we do at this point see why it is called a piecewise constant prior.

What are Bernshteĭn–von Mises results?

Suppose $X_1, \ldots, X_n$ are i.i.d. with $X_i \sim G(x, \theta_0)$, then for a large class of parametric models

$$\widehat{\theta}_n = \arg\max_{\theta} \ell_n(\theta)$$

satisfy

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d \mathrm{N}(0, I(\theta_0)^{-1}), \tag{3}$$

where $I(\theta_0)$ is the Fisher information (matrix).

What are Bernshteĭn–von Mises results?

Suppose $X_1, \ldots, X_n$ are i.i.d. with $X_i \sim G(x, \theta_0)$, then for a large class of parametric models

$$\widehat{\theta}_n = \arg\max_\theta \ell_n(\theta)$$

satisfy

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d \mathrm{N}(0, I(\theta_0)^{-1}), \tag{3}$$

where $I(\theta_0)$ is the Fisher information (matrix).

For a Bayesian, most priors π on θ, will (under mild assumptions) satisfy

$$\sqrt{n}(\theta - \widehat{\theta}_n) \,|\, X_1, \ldots, X_n \to_d \mathrm{N}(0, I(\theta_0)^{-1}) \tag{4}$$

in probability.

What are Bernshteĭn–von Mises results?

Suppose $X_1, \dots, X_n$ are i.i.d. with $X_i \sim G(x, \theta_0)$, then for a large class of parametric models

$$\widehat{\theta}_n = \arg\max_{\theta} \ell_n(\theta)$$

satisfy

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d \mathrm{N}(0, I(\theta_0)^{-1}), \tag{3}$$

where $I(\theta_0)$ is the Fisher information (matrix).

For a Bayesian, most priors π on θ, will (under mild assumptions) satisfy

$$\sqrt{n}(\theta - \widehat{\theta}_n) \,|\, X_1, \dots, X_n \to_d \mathrm{N}(0, I(\theta_0)^{-1}) \tag{4}$$

in probability.

Bernshteĭn–von Mises generally means to establish (3) and (4).

What are Bernshteĭn–von Mises results?

Suppose $X_1, \ldots, X_n$ are i.i.d. with $X_i \sim G(x, \theta_0)$, then for a large class of parametric models

$$\widehat{\theta}_n = \arg\max_{\theta} \ell_n(\theta)$$

satisfy

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d \mathrm{N}(0, I(\theta_0)^{-1}), \tag{3}$$

where $I(\theta_0)$ is the Fisher information (matrix).

For a Bayesian, most priors π on θ, will (under mild assumptions) satisfy

$$\sqrt{n}(\theta - \widehat{\theta}_n) \,|\, X_1, \ldots, X_n \to_d \mathrm{N}(0, I(\theta_0)^{-1}) \tag{4}$$

in probability.

Bernshteĭn–von Mises generally means to establish (3) and (4).

Example: The prototype illustration is with i.i.d. $X_i \sim \mathrm{N}(\mu_0, 1)$ and a normal prior $\pi \sim \mathrm{N}(0, 1)$ for the location μ, where it can be show that

$$\sqrt{n}(\bar{X}_n - \mu_0) \text{ and } \sqrt{n}(\mu - \bar{X}_n) \,|\, X_1, \ldots, X_n$$

has the same limit distribution.

How to prove Bernshteĭn–von Mises results?

There are exist several strategies for proving Bernshteĭn–von Mises, from the direct (example above) to the very general.

How to prove Bernshteĭn–von Mises results?

There are exist several strategies for proving Bernshteĭn–von Mises, from the direct (example above) to the very general.

Our strategy is based on, or similar to, following technique/observation.

If

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d \mathrm{N}(0, I(\theta_0)^{-1})$$

it is clear that if $s = \sqrt{n}(\theta - \widehat{\theta}_n)$ and

$$B_n = \int \left| \pi(s \,|\, X_1, \ldots, X_n) - \frac{\sqrt{(\theta_0)}}{\sqrt{2\pi}} e^{-\frac{s^2 I(\theta_0)}{2}} \right| \mathrm{d}s \to_{\mathrm{pr}} 0$$

then we have Bernshteĭn–von Mises type of results.

The key observation is that Bernshteĭn–von Mises (at least for parametric models) can be related to statements about convergence in probability.

Parametric estimation and the Whittle log-likelihood

Let $Y_1, \ldots, Y_n$ be a stationary Gaussian time series with covariance matrix $\Sigma(f_{\theta_0})$ with elements

$$\Sigma(f_{\theta_0})_{k,l} = \mathrm{Cov}(Y_k, Y_l) = 2\int_0^{\pi} \cos(|k-l|) f_{\theta_0}(\omega)\,\mathrm{d}\omega$$

Parametric estimation and the Whittle log-likelihood

Let $Y_1, \ldots, Y_n$ be a stationary Gaussian time series with covariance matrix $\Sigma(f_{\theta_0})$ with elements

$$\Sigma(f_{\theta_0})_{k,l} = \mathrm{Cov}(Y_k, Y_l) = 2\int_0^{\pi} \cos(|k-l|) f_{\theta_0}(\omega)\,\mathrm{d}\omega$$

Then the full Gaussian log-likelihood is

$$\ell_n(\theta) = -\tfrac{1}{2}\left[n\log(2\pi) + \log|\Sigma(f_\theta)| + Y^{\mathrm{t}}\Sigma(f_\theta)^{-1}Y\right]$$

and the maximum likelihood estimator

$$\widehat{\theta}_n = \arg\max_{\theta} \ell_n(\theta),$$

and associated large-sample properties, are somewhat complicated because of the inverse covariance matrix.

Parametric estimation and the Whittle log-likelihood

Let $Y_1, \ldots, Y_n$ be a stationary Gaussian time series with covariance matrix $\Sigma(f_{\theta_0})$ with elements

$$\Sigma(f_{\theta_0})_{k,l} = \mathrm{Cov}(Y_k, Y_l) = 2\int_0^\pi \cos(|k-l|) f_{\theta_0}(\omega)\,\mathrm{d}\omega$$

Then the full Gaussian log-likelihood is

$$\ell_n(\theta) = -\tfrac{1}{2}\left[n\log(2\pi) + \log|\Sigma(f_\theta)| + Y^{\mathrm{t}}\Sigma(f_\theta)^{-1}Y\right]$$

and the maximum likelihood estimator

$$\widehat{\theta}_n = \arg\max_\theta \ell_n(\theta),$$

and associated large-sample properties, are somewhat complicated because of the inverse covariance matrix.

There is a popular alternative, the Whittle estimator $\widetilde{\theta}_n$, which is the maximiser of the so-called Whittle log-likelihood

$$\ell_n^*(f_\theta) = -\frac{n}{2}\left\{\log 2\pi + \frac{1}{\pi}\sum_{j=1}^{\lfloor n/2\rfloor}\log(2\pi f_\theta(\omega_j))\frac{2\pi}{n} + \frac{1}{\pi}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I_n(\omega_j)}{f_\theta(\omega_j)}\frac{2\pi}{n}\right\},$$

where $I_n(\omega) = |\sum_{t=1}^n Y_t \exp(-i\omega t)|^2/(2\pi n)$ is the periodogram.

Convergence in probability and contiguity of probability measures

A natural question is whether the Whittle pseudo maximum likelihood estimator is consistent, i.e. if

$$A_n = |\widetilde{\theta}_n - \theta_0| \to_{\mathrm{pr}} 0,$$

which is a question about convergence in probability with respect to P_n.

Convergence in probability and contiguity of probability measures

A natural question is whether the Whittle pseudo maximum likelihood estimator is consistent, i.e. if

$$A_n = |\widetilde{\theta}_n - \theta_0| \to_{\mathrm{pr}} 0,$$

which is a question about convergence in probability with respect to P_n.

This, however, becomes complicated (under P_n), since θ is hidden inside the inverse of the covariance matrix.

Convergence in probability and contiguity of probability measures

A natural question is whether the Whittle pseudo maximum likelihood estimator is consistent, i.e. if

$$A_n = |\widetilde{\theta}_n - \theta_0| \to_{\text{pr}} 0,$$

which is a question about convergence in probability with respect to P_n.

This, however, becomes complicated (under P_n), since θ is hidden inside the inverse of the covariance matrix.

One solution is to introduce an alternative (simpler) contiguous probability measure, say P_n^*, to the original P_n.

A natural question is whether the Whittle pseudo maximum likelihood estimator is consistent, i.e. if

$$A_n = |\widetilde{\theta}_n - \theta_0| \to_{\text{pr}} 0,$$

which is a question about convergence in probability with respect to P_n.

This, however, becomes complicated (under P_n), since θ is hidden inside the inverse of the covariance matrix.

One solution is to introduce an alternative (simpler) contiguous probability measure, say P_n^*, to the original P_n.

Definition: Let $(\Omega_n, \mathcal{A}_n)$ be a measurable space equipped with probability measures P_n and P_n^*. Then P_n and P_n^* are said to be mutually contiguous, if for every measurable sequence of sets $A_n \in \mathcal{A}_n$, we have that

$$P_n(A_n) \to 0 \text{ if and only if } P_n^*(A_n) \to 0.$$

Convergence in probability and contiguity of probability measures

A natural question is whether the Whittle pseudo maximum likelihood estimator is consistent, i.e. if

$$A_n = |\widetilde{\theta}_n - \theta_0| \rightarrow_{\text{pr}} 0,$$

which is a question about convergence in probability with respect to P_n.

This, however, becomes complicated (under P_n), since θ is hidden inside the inverse of the covariance matrix.

One solution is to introduce an alternative (simpler) contiguous probability measure, say P_n^*, to the original P_n.

Definition: Let $(\Omega_n, \mathcal{A}_n)$ be a measurable space equipped with probability measures P_n and P_n^*. Then P_n and P_n^* are said to be mutually contiguous, if for every measurable sequence of sets $A_n \in \mathcal{A}_n$, we have that

$$P_n(A_n) \rightarrow 0 \text{ if and only if } P_n^*(A_n) \rightarrow 0.$$

In summary, contiguity of probability measures ensures that convergence in probability transfers from P_n^* to P_n (and back again).

The Whittle (log-likelihood) model

Let $Y_1^*, \ldots, Y_n^*$ be a stationary Gaussian time series from the model with spectral measure (step-function)

$$F^*(t) = \frac{2\pi f(0)}{n} + \frac{4\pi}{n} \sum_{\omega_j \leq t} f(\omega_j), \text{ where } \omega_j = 2\pi j/n \text{ for } j = 0, 1, \ldots, \lfloor, n/2 \rfloor$$

which is valid (i.e. the series/model exist) by Wold's Theorem.

The Whittle (log-likelihood) model

Let $Y_1^*, \ldots, Y_n^*$ be a stationary Gaussian time series from the model with spectral measure (step-function)

$$F^*(t) = \frac{2\pi f(0)}{n} + \frac{4\pi}{n} \sum_{\omega_j \leq t} f(\omega_j), \text{ where } \omega_j = 2\pi j/n \text{ for } j = 0, 1, \ldots, \lfloor, n/2 \rfloor$$

which is valid (i.e. the series/model exist) by Wold's Theorem.

Then, the corresponding covariance matrix can be decomposed as

$$\Sigma^*(f) = 2\pi Q_n^{\mathrm{t}} D_n(f) Q_n$$

where Q_n are orthonormal and

$$D_n(f) = \mathrm{diag}(f(\omega_1), f((\omega_2), \ldots, f(\omega_{\lfloor n/2 \rfloor}))$$

The Whittle (log-likelihood) model

Let $Y_1^*, \dots, Y_n^*$ be a stationary Gaussian time series from the model with spectral measure (step-function)

$$F^*(t) = \frac{2\pi f(0)}{n} + \frac{4\pi}{n} \sum_{\omega_j \le t} f(\omega_j), \text{ where } \omega_j = 2\pi j/n \text{ for } j = 0, 1, \dots, \lfloor, n/2 \rfloor$$

which is valid (i.e. the series/model exist) by Wold's Theorem.

Then, the corresponding covariance matrix can be decomposed as

$$\Sigma^*(f) = 2\pi Q_n^{\mathrm{t}} D_n(f) Q_n$$

where Q_n are orthonormal and

$$D_n(f) = \mathrm{diag}(f(\omega_1), f((\omega_2), \dots, f(\omega_{\lfloor n/2 \rfloor}))$$

Furthermore, the log-likelihood function for the series $Y_1^*, \dots, Y_n^*$ is

$$\ell_n(f, \boldsymbol{Y}^*) = \widetilde{\ell}_n(f, \boldsymbol{Y}^*),$$

it is exactly equal to the Whittle approximation,

The Whittle (log-likelihood) model

Let $Y_1^*, \ldots, Y_n^*$ be a stationary Gaussian time series from the model with spectral measure (step-function)

$$F^*(t) = \frac{2\pi f(0)}{n} + \frac{4\pi}{n} \sum_{\omega_j \le t} f(\omega_j), \text{ where } \omega_j = 2\pi j/n \text{ for } j = 0, 1, \ldots, \lfloor, n/2 \rfloor$$

which is valid (i.e. the series/model exist) by Wold's Theorem.

Then, the corresponding covariance matrix can be decomposed as

$$\Sigma^*(f) = 2\pi Q_n^{\mathrm{t}} D_n(f) Q_n$$

where Q_n are orthonormal and

$$D_n(f) = \mathrm{diag}(f(\omega_1), f((\omega_2), \ldots, f(\omega_{\lfloor n/2 \rfloor}))$$

Furthermore, the log-likelihood function for the series $Y_1^*, \ldots, Y_n^*$ is

$$\ell_n(f, \boldsymbol{Y}^*) = \widetilde{\ell}_n(f, \boldsymbol{Y}^*),$$

it is exactly equal to the Whittle approximation, and the periodogram ordinates are (exact)

$$I_n^*(w_j) \sim f(\omega_j)\mathrm{Exp}(1).$$

Contiguity and consistency of the Whittle estimator

Let P_n^* be the probability measure for the new series $Y_1^*, \ldots, Y_n^*$.

Contiguity and consistency of the Whittle estimator

Let P_n^* be the probability measure for the new series $Y_1^*, \ldots, Y_n^*$.

Then, it can be shown that P_n and P_n^* are mutually contiguous; see Choudhuri et al. (2004) for details.

Contiguity and consistency of the Whittle estimator

Let P_n^* be the probability measure for the new series $Y_1^*, \ldots, Y_n^*$.

Then, it can be shown that P_n and P_n^* are mutually contiguous; see Choudhuri et al. (2004) for details.

For the parametric model with spectral density f_θ, proving consistency of the Whittle estimator is considerably simplified under P_n^*.

Contiguity and consistency of the Whittle estimator

Let P_n^* be the probability measure for the new series $Y_1^*, \ldots, Y_n^*$.

Then, it can be shown that P_n and P_n^* are mutually contiguous; see Choudhuri et al. (2004) for details.

For the parametric model with spectral density f_θ, proving consistency of the Whittle estimator is considerably simplified under P_n^*.

The main reason being that under P_n^*

$$I_n^*(w_0), \ldots, I_n^*(w_{\lfloor n/2 \rfloor})$$

are sufficient, and independent with $I_n^*(w_j) \sim f_{\theta_0}(\omega_j)\mathrm{Exp}(1)$.

Contiguity and consistency of the Whittle estimator

Let P_n^* be the probability measure for the new series $Y_1^*, \ldots, Y_n^*$.

Then, it can be shown that P_n and P_n^* are mutually contiguous; see Choudhuri et al. (2004) for details.

For the parametric model with spectral density f_θ, proving consistency of the Whittle estimator is considerably simplified under P_n^*.

The main reason being that under P_n^*

$$I_n^*(w_0), \ldots, I_n^*(w_{\lfloor n/2 \rfloor})$$

are sufficient, and independent with $I_n^*(w_j) \sim f_{\theta_0}(\omega_j)\mathrm{Exp}(1)$.

In a way, we have changed a problem with dependent observations to a problem with independent 'observations'.

Contiguity and consistency of the Whittle estimator

Let P_n^* be the probability measure for the new series $Y_1^*, \ldots, Y_n^*$.

Then, it can be shown that P_n and P_n^* are mutually contiguous; see Choudhuri et al. (2004) for details.

For the parametric model with spectral density f_θ, proving consistency of the Whittle estimator is considerably simplified under P_n^*.

The main reason being that under P_n^*

$$I_n^*(w_0), \ldots, I_n^*(w_{\lfloor n/2 \rfloor})$$

are sufficient, and independent with $I_n^*(w_j) \sim f_{\theta_0}(\omega_j)\text{Exp}(1)$.

In a way, we have changed a problem with dependent observations to a problem with independent 'observations'.

The nice thing about contiguity, is that consistency for $\widetilde{\theta}_n$ (derived under P_n^*) can be carried over (via the contiguity bridge) to P_n.

Contiguity and consistency of the Whittle estimator

Let P_n^* be the probability measure for the new series $Y_1^*, \ldots, Y_n^*$.

Then, it can be shown that P_n and P_n^* are mutually contiguous; see Choudhuri et al. (2004) for details.

For the parametric model with spectral density f_θ, proving consistency of the Whittle estimator is considerably simplified under P_n^*.

The main reason being that under P_n^*

$$I_n^*(w_0), \ldots, I_n^*(w_{\lfloor n/2 \rfloor})$$

are sufficient, and independent with $I_n^*(w_j) \sim f_{\theta_0}(\omega_j)\text{Exp}(1)$.

In a way, we have changed a problem with dependent observations to a problem with independent 'observations'.

The nice thing about contiguity, is that consistency for $\widetilde{\theta}_n$ (derived under P_n^*) can be carried over (via the contiguity bridge) to P_n.

The consistency of the Whittle estimator is already well known in the literature; the proof using contiguity might be new.

Furthermore, it is well known that actual maximum likelihood estimator in this setup (see e.g. Dzhaparidze (1986))

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d \mathrm{N}(0, J(f_{\theta_0})^{-1})$$

for a certain Fisher information matrix $J(f_{\theta_0})$.

Furthermore, it is well known that actual maximum likelihood estimator in this setup (see e.g. Dzhaparidze (1986))

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d \mathrm{N}(0, J(f_{\theta_0})^{-1})$$

for a certain Fisher information matrix $J(f_{\theta_0})$.

This result is also true if $\widehat{\theta}_n$ is replaced by $\widetilde{\theta}_n$.

Furthermore, it is well known that actual maximum likelihood estimator in this setup (see e.g. Dzhaparidze (1986))

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d \mathrm{N}(0, J(f_{\theta_0})^{-1})$$

for a certain Fisher information matrix $J(f_{\theta_0})$.

This result is also true if $\widehat{\theta}_n$ is replaced by $\widetilde{\theta}_n$.

Furthermore, we may use the same type of reasoning to simplify and use

$$B_n = \int \left| \pi(s \,|\, y_1, \ldots, y_n) - \frac{\sqrt{(\theta_0)}}{\sqrt{2\pi}} e^{-\frac{s^2 I(f_{\theta_0})}{2}} \right| \mathrm{d}s$$

to prove Bernshteĭn–von Mises under P_n^*, using, among others that

$$\pi(s \,|\, Y_1, \ldots, Y_n) = \pi(s \,|\, I_n^*(w_0), \ldots, I_n^*(w_{\lfloor n/2 \rfloor})).$$

Furthermore, it is well known that actual maximum likelihood estimator in this setup (see e.g. Dzhaparidze (1986))

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d \mathrm{N}(0, J(f_{\theta_0})^{-1})$$

for a certain Fisher information matrix $J(f_{\theta_0})$.

This result is also true if $\widehat{\theta}_n$ is replaced by $\widetilde{\theta}_n$.

Furthermore, we may use the same type of reasoning to simplify and use

$$B_n = \int \left| \pi(s \,|\, y_1, \ldots, y_n) - \frac{\sqrt{(\theta_0)}}{\sqrt{2\pi}} e^{-\frac{s^2 I(f_{\theta_0})}{2}} \right| \mathrm{d}s$$

to prove Bernshteĭn–von Mises under P_n^*, using, among others that

$$\pi(s \,|\, Y_1, \ldots, Y_n) = \pi(s \,|\, I_n^*(w_0), \ldots, I_n^*(w_{\lfloor n/2 \rfloor})).$$

Again, contiguity ensures that we also have (parametric) Bernshteĭn–von Mises under the original measure P_n.

From parametric to nonparametric

Let $Y_1, \ldots, Y_n$ be a series generated by a model with spectral measure F.

From parametric to nonparametric

Let $Y_1, \ldots, Y_n$ be a series generated by a model with spectral measure F.

One canonical nonparametric estimator for F is

$$\widehat{F}_n(t) = \frac{2\pi}{n} \sum_{j=0}^{\lfloor n/2 \rfloor} I_n(2\pi j/n),$$

with $I_n(w) = |\sum_{t=1}^{n} Y_t \exp\{i\omega t\}|^2/(2\pi n)$, has the following (normalised) limit

$$\sqrt{n}(\widehat{F}_n(t) - F(t)) \to_d W(2\pi \int_0^t f(\omega)^2 \, \mathrm{d}\omega), \tag{5}$$

see e.g. Ibragimov (1963) for details.

From parametric to nonparametric

Let $Y_1, \ldots, Y_n$ be a series generated by a model with spectral measure F.

One canonical nonparametric estimator for F is

$$\widehat{F}_n(t) = \frac{2\pi}{n} \sum_{j=0}^{\lfloor n/2 \rfloor} I_n(2\pi j/n),$$

with $I_n(w) = |\sum_{t=1}^{n} Y_t \exp\{i\omega t\}|^2/(2\pi n)$, has the following (normalised) limit

$$\sqrt{n}(\widehat{F}_n(t) - F(t)) \to_d W(2\pi \int_0^t f(\omega)^2 \, \mathrm{d}\omega), \tag{5}$$

see e.g. Ibragimov (1963) for details.

Let $B_n(t) = \sqrt{n}(F_n(t) - \widehat{F}_n(t))$, the goal is to obtain conditions such that

$$B_n(t) \,|\, Y_1, \ldots, Y_n \to_d W(2\pi \int_0^t f(\omega)^2 \, \mathrm{d}\omega) \tag{6}$$

in P_n-probability.

From parametric to nonparametric

Let $Y_1, \ldots, Y_n$ be a series generated by a model with spectral measure F.

One canonical nonparametric estimator for F is

$$\widehat{F}_n(t) = \frac{2\pi}{n} \sum_{j=0}^{\lfloor n/2 \rfloor} I_n(2\pi j/n),$$

with $I_n(w) = |\sum_{t=1}^{n} Y_t \exp\{i\omega t\}|^2/(2\pi n)$, has the following (normalised) limit

$$\sqrt{n}(\widehat{F}_n(t) - F(t)) \to_d W(2\pi \int_0^t f(\omega)^2 \, \mathrm{d}\omega), \tag{5}$$

see e.g. Ibragimov (1963) for details.

Let $B_n(t) = \sqrt{n}(F_n(t) - \widehat{F}_n(t))$, the goal is to obtain conditions such that

$$B_n(t) \,|\, Y_1, \ldots, Y_n \to_d W(2\pi \int_0^t f(\omega)^2 \, \mathrm{d}\omega) \tag{6}$$

in P_n-probability.

And nonparametric Bernshteĭn–von Mises means establishing (6).

Let $Y_1, \dots, Y_n$ be a series generated by a model with spectral measure F.

One canonical nonparametric estimator for F is

$$\widehat{F}_n(t) = \frac{2\pi}{n} \sum_{j=0}^{\lfloor n/2 \rfloor} I_n(2\pi j/n),$$

with $I_n(w) = |\sum_{t=1}^{n} Y_t \exp\{i\omega t\}|^2/(2\pi n)$, has the following (normalised) limit

$$\sqrt{n}(\widehat{F}_n(t) - F(t)) \to_d W(2\pi \int_0^t f(\omega)^2 \,\mathrm{d}\omega), \tag{5}$$

see e.g. Ibragimov (1963) for details.

Let $B_n(t) = \sqrt{n}(F_n(t) - \widehat{F}_n(t))$, the goal is to obtain conditions such that

$$B_n(t) \,|\, Y_1, \dots, Y_n \to_d W(2\pi \int_0^t f(\omega)^2 \,\mathrm{d}\omega) \tag{6}$$

in P_n-probability.

And nonparametric Bernshteĭn–von Mises means establishing (6).

First, we need to define $F_n(t)$ and an appropriate prior distribution.

The piecewise constant priors

Let $0 = w_0 < w_1 < \cdots < w_m = \pi$ be a growing partition, then a piecewise constant prior on the spectral density results in the prior

$$F_n(t) = \sum_{\{l:w_l \leq t\}} (w_{l+1} - w_l) f_l, \text{ with i.i.d. } f_l \sim \pi_j$$

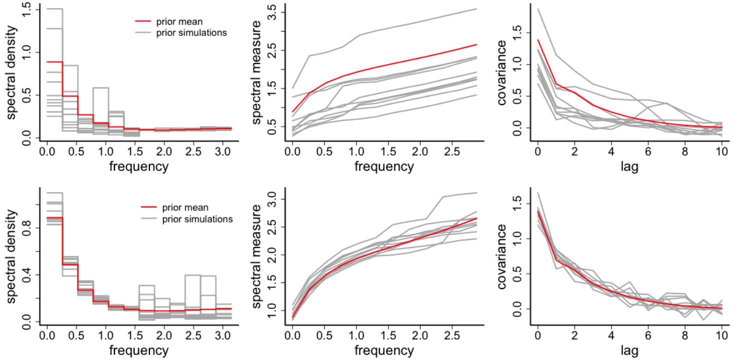
on the spectral measures, for $t = w_l$ and with linear interpolation between these points.

The piecewise constant priors

Let $0 = w_0 < w_1 < \cdots < w_m = \pi$ be a growing partition, then a piecewise constant prior on the spectral density results in the prior

$$F_n(t) = \sum_{\{l : w_l \leq t\}} (w_{l+1} - w_l) f_l, \text{ with i.i.d. } f_l \sim \pi_j$$

on the spectral measures, for $t = w_l$ and with linear interpolation between these points.



The number of cells/windows will grow with n; making it nonparametric.

Let $n = k \times m$, where m is the number of windows.

Let $n = k \times m$, where m is the number of windows.

Furthermore, let $w_l - w_{l-1} = 2\pi/m$, and define the estimator (for $t = w_l$ and with linear interpolation between these points)

$$\bar{F}_n(t) = \frac{2\pi}{m} \sum_{\{l \,:\, w_l \leq t\}} \bar{I}_{n,l},$$

where $\bar{I}_{n,l}$ is the average of the $I(\omega_j)$ inside window $[w_{l-1}, w_l)$.

Let $n = k \times m$, where m is the number of windows.

Furthermore, let $w_l - w_{l-1} = 2\pi/m$, and define the estimator (for $t = w_l$ and with linear interpolation between these points)

$$\bar{F}_n(t) = \frac{2\pi}{m} \sum_{\{l\,:\,w_l \le t\}} \bar{I}_{n,l},$$

where $\bar{I}_{n,l}$ is the average of the $I(\omega_j)$ inside window $[w_{l-1}, w_l)$.

Now, we may decompose

$$\begin{aligned} B_n(t) &= \sqrt{n}(F_n(t) - \widehat{F}_n(t)) \\ &= \sqrt{n}(F_n(t) - \bar{F}_n(t)) + \sqrt{n}(\bar{F}_n(t) - \widehat{F}_n(t)) \\ &= \sum_{\{l\,:\,w_l \le t\}} z_l + r_n(t) \end{aligned} \tag{7}$$

where $z_l = \sqrt{n}/m(f_l - \bar{I}_{n,l}) = \sqrt{k}(f_l - \bar{I}_{n,l})$.

Let $n = k \times m$, where m is the number of windows.

Furthermore, let $w_l - w_{l-1} = 2\pi/m$, and define the estimator (for $t = w_l$ and with linear interpolation between these points)

$$\bar{F}_n(t) = \frac{2\pi}{m} \sum_{\{l\,:\,w_l \le t\}} \bar{I}_{n,l},$$

where $\bar{I}_{n,l}$ is the average of the $I(\omega_j)$ inside window $[w_{l-1}, w_l)$.

Now, we may decompose

$$\begin{aligned} B_n(t) &= \sqrt{n}(F_n(t) - \widehat{F}_n(t)) \\ &= \sqrt{n}(F_n(t) - \bar{F}_n(t)) + \sqrt{n}(\bar{F}_n(t) - \widehat{F}_n(t)) \\ &= \sum_{\{l\,:\,w_l \le t\}} z_l + r_n(t) \end{aligned} \tag{7}$$

where $z_l = \sqrt{n}/m(f_l - \bar{I}_{n,l}) = \sqrt{k}(f_l - \bar{I}_{n,l})$.

In short, we now need show

(i) $r_n(t)$ to be small in probability (approximation quality), and

(ii) that $\sum_{\{l\,:\,w_l \le t\}} z_l \mid Y_1, \ldots, Y_n$ has the right limit.

The size of

$$r_n(t) = \sqrt{n}(\bar{F}_n(t) - \widehat{F}_n(t))$$

is essentially a question about approximation quality.

The size of

$$r_n(t) = \sqrt{n}(\bar{F}_n(t) - \widehat{F}_n(t))$$

is essentially a question about approximation quality.

And, if the underlying $F(t)$ is continuous, the size of $r_n(t)$ can be shown to be sufficiently small (in probability) provided

$$\sqrt{n}/m^2 \to 0,$$

again, a key part of the argument is the contiguity result; see also Hermansen & Hjort (2015) for details.

The second part of the argument is related to the limit distribution of the posterior, which should match the nonparametric estimator, i.e.

$$\sum_{\{l\,:\,w_l \leq t\}} z_l \,\Big|\, Y_1, \ldots Y_n \to_d W(2\pi \int_0^t f(\omega)^2 \, \mathrm{d}\omega),$$

in P_n-probability.

The second part of the argument is related to the limit distribution of the posterior, which should match the nonparametric estimator, i.e.

$$\sum_{\{l\,:\,w_l \leq t\}} z_l \,\Big|\, Y_1, \ldots Y_n \to_d W\left(2\pi \int_0^t f(\omega)^2 \,\mathrm{d}\omega\right),$$

in P_n-probability.

Essentially a question about finite dimensional convergence and tightness.

The second part of the argument is related to the limit distribution of the posterior, which should match the nonparametric estimator, i.e.

$$\sum_{\{l\,:\,w_l \leq t\}} z_l \,\big|\, Y_1, \ldots Y_n \to_d W(2\pi \int_0^t f(\omega)^2 \, \mathrm{d}\omega),$$

in P_n-probability.

Essentially a question about finite dimensional convergence and tightness.

Using Helland (1982) and the contiguity bridge, it is possible to break the problem down into statements of convergence in probability at the level of

$$z_l \,|\, Y_1, \ldots, Y_n \text{ for all } l = 1, \ldots, n.$$

The second part of the argument is related to the limit distribution of the posterior, which should match the nonparametric estimator, i.e.

$$\sum_{\{l\,:\,w_l \leq t\}} z_l \,\big|\, Y_1, \ldots Y_n \to_d W(2\pi \int_0^t f(\omega)^2 \,\mathrm{d}\omega),$$

in P_n-probability.

Essentially a question about finite dimensional convergence and tightness.

Using Helland (1982) and the contiguity bridge, it is possible to break the problem down into statements of convergence in probability at the level of

$$z_l \,|\, Y_1, \ldots, Y_n \text{ for all } l = 1, \ldots, n.$$

From results in Hermansen & Hjort (2015) the main condition (in addition to standard regularity assumptions) is

$$m/\sqrt{n} \to 0,$$

which (essentially) ensures sure that the prior is 'washed out' of the limit.

The illustration - Air pollutant data

Hourly measurements of fine particulate matter ($PM_{2.5}$), i.e. tiny particles, or droplets, that are two and one half microns or less in width.

The illustration - Air pollutant data

Hourly measurements of fine particulate matter ($PM_{2.5}$), i.e. tiny particles, or droplets, that are two and one half microns or less in width.

Exposure can cause eye, nose, throat and lung irritation, and in long-term may affect lung function and is also related to asthma and heart diseases.

The illustration - Air pollutant data

Hourly measurements of fine particulate matter ($PM_{2.5}$), i.e. tiny particles, or droplets, that are two and one half microns or less in width.

Exposure can cause eye, nose, throat and lung irritation, and in long-term may affect lung function and is also related to asthma and heart diseases.

For health reasons, government regulations typically restrict the daily average emission to 25-35 $\mu g/\mathrm{m}^3$.

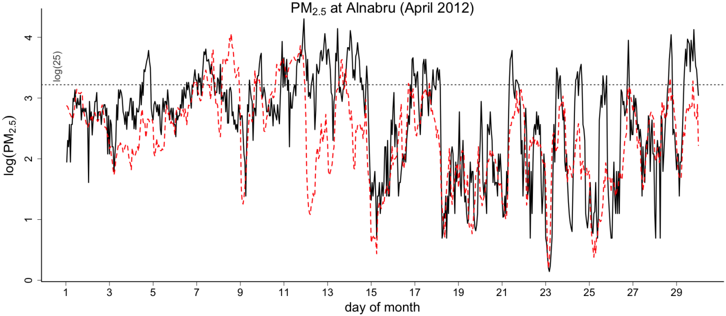


Figure: Observed and prognosis (24h in advanced) of $PM_{2.5}$ at Alnabru (Norway),

The illustration - Air pollutant data

The Norwegian Institute for Air Research (NILU) and Norwegian Meteorological Institute (MET) makes prognosis for $PM_{2.5}$ using the current history and computer simulations.

The illustration - Air pollutant data

The Norwegian Institute for Air Research (NILU) and Norwegian Meteorological Institute (MET) makes prognosis for $PM_{2.5}$ using the current history and computer simulations.

The prognosis can be used to predict periods with potential high emissions, enabling the legislature to take appropriate actions in advanced.

The Norwegian Institute for Air Research (NILU) and Norwegian Meteorological Institute (MET) makes prognosis for $PM_{2.5}$ using the current history and computer simulations.

The prognosis can be used to predict periods with potential high emissions, enabling the legislature to take appropriate actions in advanced.

There are certain systematic biases in the prognosis, however, which may be corrected using a stochastic model.

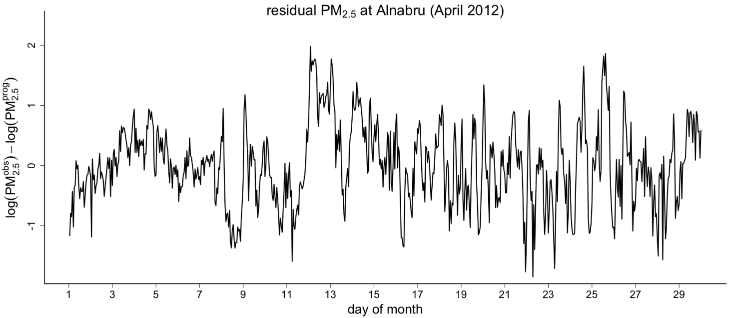


Figure: The difference between observed and prognosis (made 24h in advanced) of $PM_{2.5}$

Here, we will focus on the probability that at least one of three following $\mathrm{PM}_{2.5}$ measurements will exceed the critical value of 25 $\mu g/\mathrm{m}^3$, i.e.

$$1 - \Pr\{Y_{n+1} \leq a, Y_{n+2} \leq a, Y_{n+3} \leq a \,|\, \mathrm{data}_k\}$$

with $a = \log(25)$, $Y_i = \mathrm{PM}_{2.5,i}^{\mathrm{prog}} + \epsilon_i$ and $\mathrm{data}_k = \{Y_n, \ldots, Y_{n-k+1}\}$, by placing a nonparametric envelope around an AR(1) model.

The illustration - The focus and the posterior

Here, we will focus on the probability that at least one of three following $PM_{2.5}$ measurements will exceed the critical value of 25 $\mu g/m^3$, i.e.

$$1 - \Pr\{Y_{n+1} \le a, Y_{n+2} \le a, Y_{n+3} \le a \,|\, \text{data}_k\}$$

with $a = \log(25)$, $Y_i = \text{PM}^{\text{prog}}_{2.5,i} + \epsilon_i$ and $\text{data}_k = \{Y_n, \dots, Y_{n-k+1}\}$, by placing a nonparametric envelope around an AR(1) model.

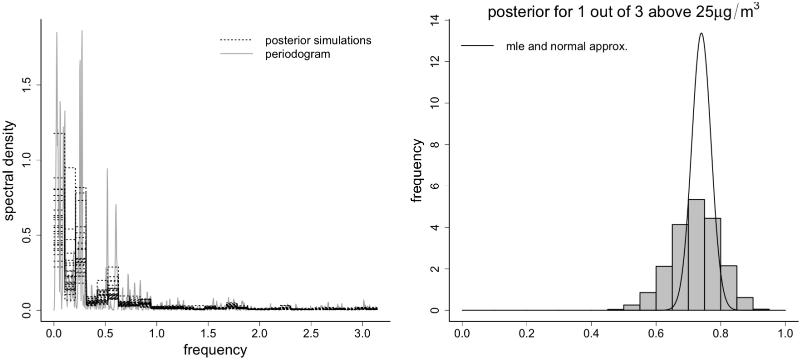


Figure: Empirical periodogram and samples from posterior (left), and posterior density for threshold probability and density based on mle for an AR(7) (AIC winner), normal approximation and delta method (right).

Concluding remarks

We have established Bernshteĭn–von Mises for a class of stationary time seires models with a class of growing piecewise constant priors, i.e.

$$F_n(t) = \sum_{\{l\,:\,w_l \le t\}} (w_{l+1} - w_l) f_l, \text{ with i.i.d. } f_l \sim \pi_j,$$

where $w_1, \ldots, w_m$ is a partition of $[0, \pi]$.

Concluding remarks

We have established Bernshteĭn–von Mises for a class of stationary time seires models with a class of growing piecewise constant priors, i.e.

$$F_n(t) = \sum_{\{l : w_l \leq t\}} (w_{l+1} - w_l) f_l, \text{ with i.i.d. } f_l \sim \pi_j,$$

where $w_1, \ldots, w_m$ is a partition of $[0, \pi]$.

The (non-standard) condition is related to the number m of components in the prior, requiring that $m = cn^\alpha$ with $1/4 < \alpha < 1/2$ and c a constant.

Concluding remarks

We have established Bernshteĭn–von Mises for a class of stationary time seires models with a class of growing piecewise constant priors, i.e.

$$F_n(t) = \sum_{\{l \,:\, w_l \leq t\}} (w_{l+1} - w_l) f_l, \text{ with i.i.d. } f_l \sim \pi_j,$$

where $w_1, \ldots, w_m$ is a partition of $[0, \pi]$.

The (non-standard) condition is related to the number m of components in the prior, requiring that $m = cn^\alpha$ with $1/4 < \alpha < 1/2$ and c a constant.

Most time series does not have expectation zero, a more natural model is

$$Y_t = \mu + \epsilon_t$$

where ϵ_t is a stationary process.

Concluding remarks

We have established Bernshteĭn–von Mises for a class of stationary time seires models with a class of growing piecewise constant priors, i.e.

$$F_n(t) = \sum_{\{l\,:\,w_l \leq t\}} (w_{l+1} - w_l) f_l, \text{ with i.i.d. } f_l \sim \pi_j,$$

where $w_1, \dots, w_m$ is a partition of $[0, \pi]$.

The (non-standard) condition is related to the number m of components in the prior, requiring that $m = cn^\alpha$ with $1/4 < \alpha < 1/2$ and c a constant.

Most time series does not have expectation zero, a more natural model is

$$Y_t = \mu + \epsilon_t$$

where ϵ_t is a stationary process.

It is possible to put a prior μ and use (e.g.) MCMC for posterior inference.

Concluding remarks

We have established Bernshteĭn–von Mises for a class of stationary time seires models with a class of growing piecewise constant priors, i.e.

$$F_n(t) = \sum_{\{l : w_l \leq t\}} (w_{l+1} - w_l) f_l, \text{ with i.i.d. } f_l \sim \pi_j,$$

where $w_1, \ldots, w_m$ is a partition of $[0, \pi]$.

The (non-standard) condition is related to the number m of components in the prior, requiring that $m = cn^\alpha$ with $1/4 < \alpha < 1/2$ and c a constant.

Most time series does not have expectation zero, a more natural model is

$$Y_t = \mu + \epsilon_t$$

where ϵ_t is a stationary process.

It is possible to put a prior μ and use (e.g.) MCMC for posterior inference.

A joint Bernshteĭn–von Mises for μ and $F_n(t)$ is not trivial.

Concluding remarks

We have established Bernshteĭn–von Mises for a class of stationary time seires models with a class of growing piecewise constant priors, i.e.

$$F_n(t) = \sum_{\{l : w_l \leq t\}} (w_{l+1} - w_l) f_l, \text{ with i.i.d. } f_l \sim \pi_j,$$

where $w_1, \ldots, w_m$ is a partition of $[0, \pi]$.

The (non-standard) condition is related to the number m of components in the prior, requiring that $m = cn^\alpha$ with $1/4 < \alpha < 1/2$ and c a constant.

Most time series does not have expectation zero, a more natural model is

$$Y_t = \mu + \epsilon_t$$

where ϵ_t is a stationary process.

It is possible to put a prior μ and use (e.g.) MCMC for posterior inference.

A joint Bernshteĭn–von Mises for μ and $F_n(t)$ is not trivial.

However, I see the outline of another contiguity bridge for the series

$$\widehat{Y}_t = Y_t - \widehat{\mu} = Y_t - \bar{Y}_n$$

Concluding remarks

We have established Bernshteĭn–von Mises for a class of stationary time seires models with a class of growing piecewise constant priors, i.e.

$$F_n(t) = \sum_{\{l\,:\,w_l \le t\}} (w_{l+1} - w_l) f_l, \text{ with i.i.d. } f_l \sim \pi_j,$$

where $w_1, \dots, w_m$ is a partition of $[0, \pi]$.

The (non-standard) condition is related to the number m of components in the prior, requiring that $m = cn^\alpha$ with $1/4 < \alpha < 1/2$ and c a constant.

Most time series does not have expectation zero, a more natural model is

$$Y_t = \mu + \epsilon_t$$

where ϵ_t is a stationary process.

It is possible to put a prior μ and use (e.g.) MCMC for posterior inference.

A joint Bernshteĭn–von Mises for μ and $F_n(t)$ is not trivial.

However, I see the outline of another contiguity bridge for the series

$$\widehat{Y}_t = Y_t - \widehat{\mu} = Y_t - \bar{Y}_n$$

And, similar for the canonical model with covariates, i.e.

$$Y_t = x_t^{\mathrm{t}} \beta + \epsilon_t.$$

References

[1] N. Choudhuri, S. Ghosal, and A. Roy. Contiguity of the Whittle measure for a Gaussian time series. *Biometrika*, 91(1):211–218, 2004.

[2] K. Dzhaparidze. *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. Springer, Berlin, 1986.

[3] I.S. Helland. Central limit theorems for martingales with discrete or continuous time. *Scandinavian Journal of Statistics*, 9:79–94, 1982.

[4] G. H. Hermansen and N. L. Hjort. Bernshteĭn–von Mises theorems for nonparametric function analysis via locally constant modelling: a unified approach. *Journal of Statistical Planning and Inference*, 166:138–157, 2015.

[5] I. A. Ibragimov. On estimation of the spectral function of a stationary Gaussian process. *Theory of Probability & Its Applications*, 8(4):366–401, 1963.

[6] P. Whittle. The analysis of multiple stationary time series. *Journal of the Royal Statistical Society Series B*, 15:125–139, 1953.