

Bernshtein–von Mises theorems for a class of Bayesian non-parametrics setups for stationary time series

Gudmund Horn Hermansen and Nils Lid Hjort, Department of Mathematics, University of Oslo, Norway

Introduction and summary

We construct a Bayesian nonparametric framework for estimation and inference for stationary Gaussian time series with expectation zero; detrending can be used to deal with series with regression type of trends. The idea is to use the **frequency representation**

$$C(h) = 2 \int_0^\pi e^{-i\omega h} dF(\omega),$$

to build a **nonparametric envelope** around a parametric covariance function by putting a prior on a family of spectral measures. The general workflow can be summarised as

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

In particular, we establish **Bernshtein–von Mises type of results**, in the sense that the posterior has the same limit distribution as the empirical analogue in (1). To achieve this we will: (i) use a **piecewise constant prior**, and (ii) work under an alternative **contiguous measure**; both are needed to simplify the construction and associated derivations and conditions.

Nonparametrics inference

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

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

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

$$\sqrt{n}(\hat{F}_n(t) - F^\circ(t)) \rightarrow_d W(2\pi \int_0^t f^\circ(\omega)^2 d\omega), \quad (1)$$

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

The piecewise constant priors

Let $0 = w_0 < w_1 < \dots < w_{m_n} = \pi$ be a growing partition, then a **piecewise constant prior** on the spectral density (Figure 1) results in the prior

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

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

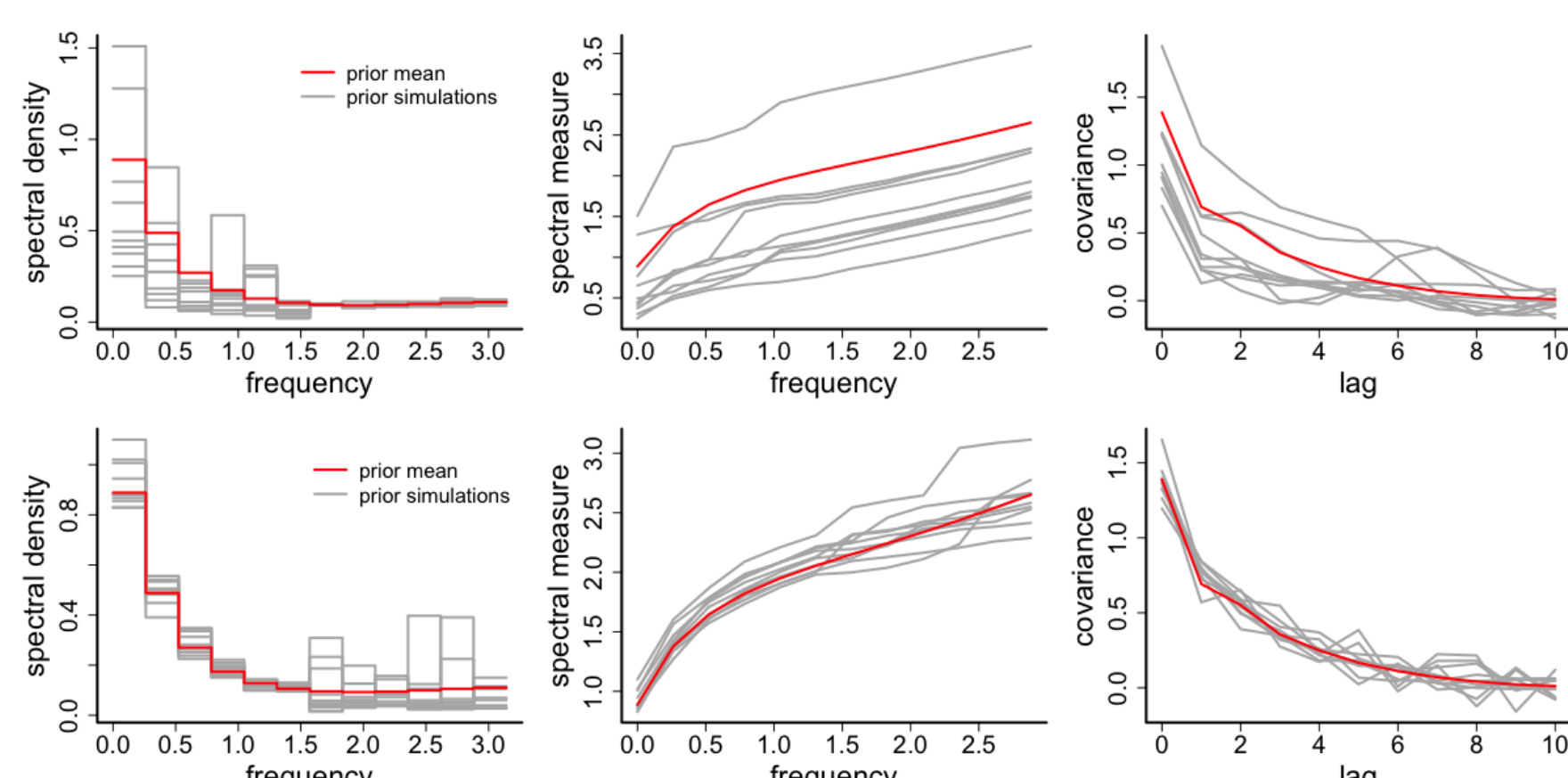


Figure 1: Samples from the prior with centred around an AR(2) with $\sigma = 1$ and $\rho = (0.3, 0.2)$ and two types of concentration.

Bernshtein–von Mises for time series

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

$$B_n(t) \mid \text{data} \rightarrow_d W(2\pi \int_0^t f^\circ(\omega)^2 d\omega) \quad (2)$$

in probability.

Air pollution

As an illustration we consider hourly measurements of fine particulate matter (PM_{2.5}), a type of air pollutant referring to tiny particles, or droplets, in the air that are two and one half microns or less in width. Exposure to fine particles can cause short-term effects, like eye, nose, throat and lung irritation, long-term exposure can affect lung function and has also been shown to be related to asthma and heart disease. For health reasons, government regulations typically restrict the daily average emission to 25-35 $\mu\text{g}/\text{m}^3$.

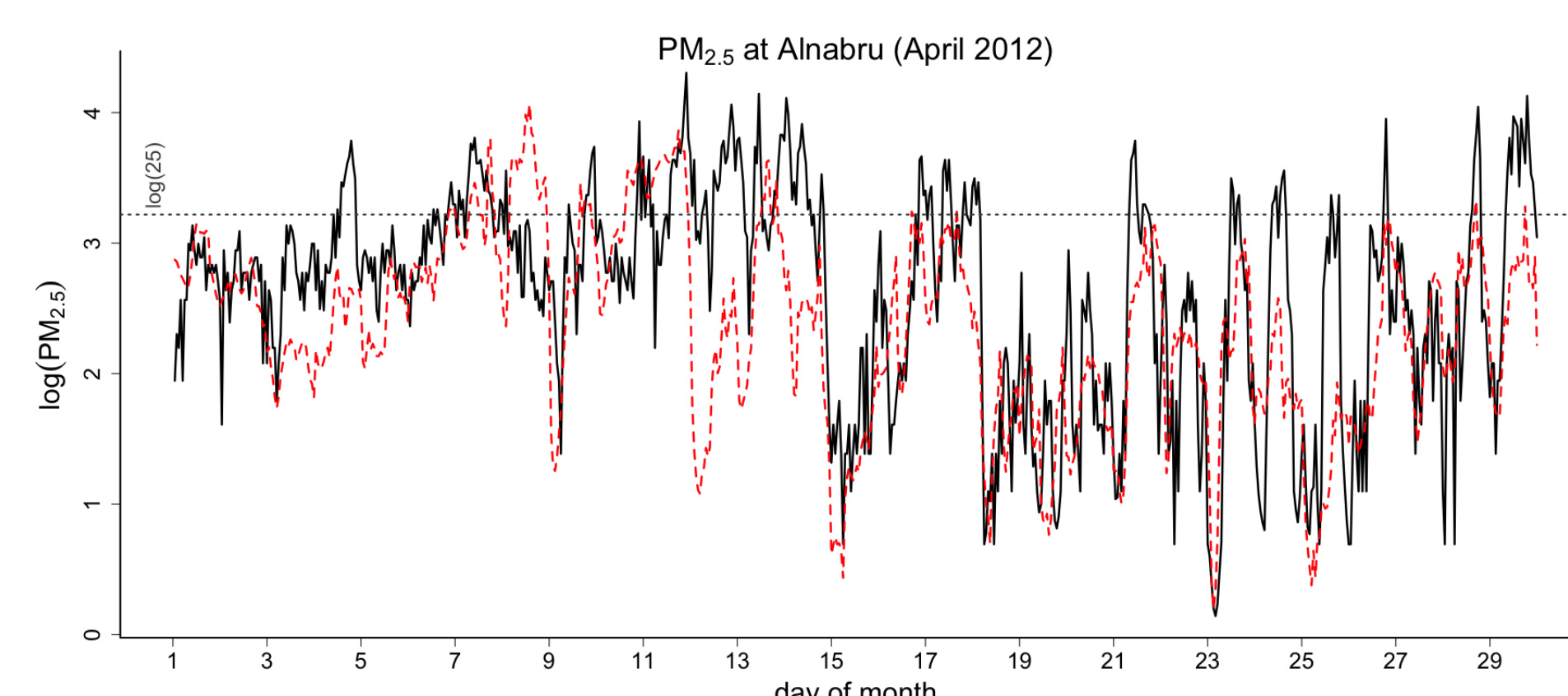


Figure 2: Observed and prognosis (made 24h in advanced) of PM_{2.5} at Alnabru (Norway),

The Norwegian Institute for Air Research (NILU) in collaboration with the Norwegian Meteorological Institute (MET) makes prognosis/forecasts for different kinds of air pollution. In particular they make prognosis for PM_{2.5} (Figure 2) using the current history and computer simulations. These forecast 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.

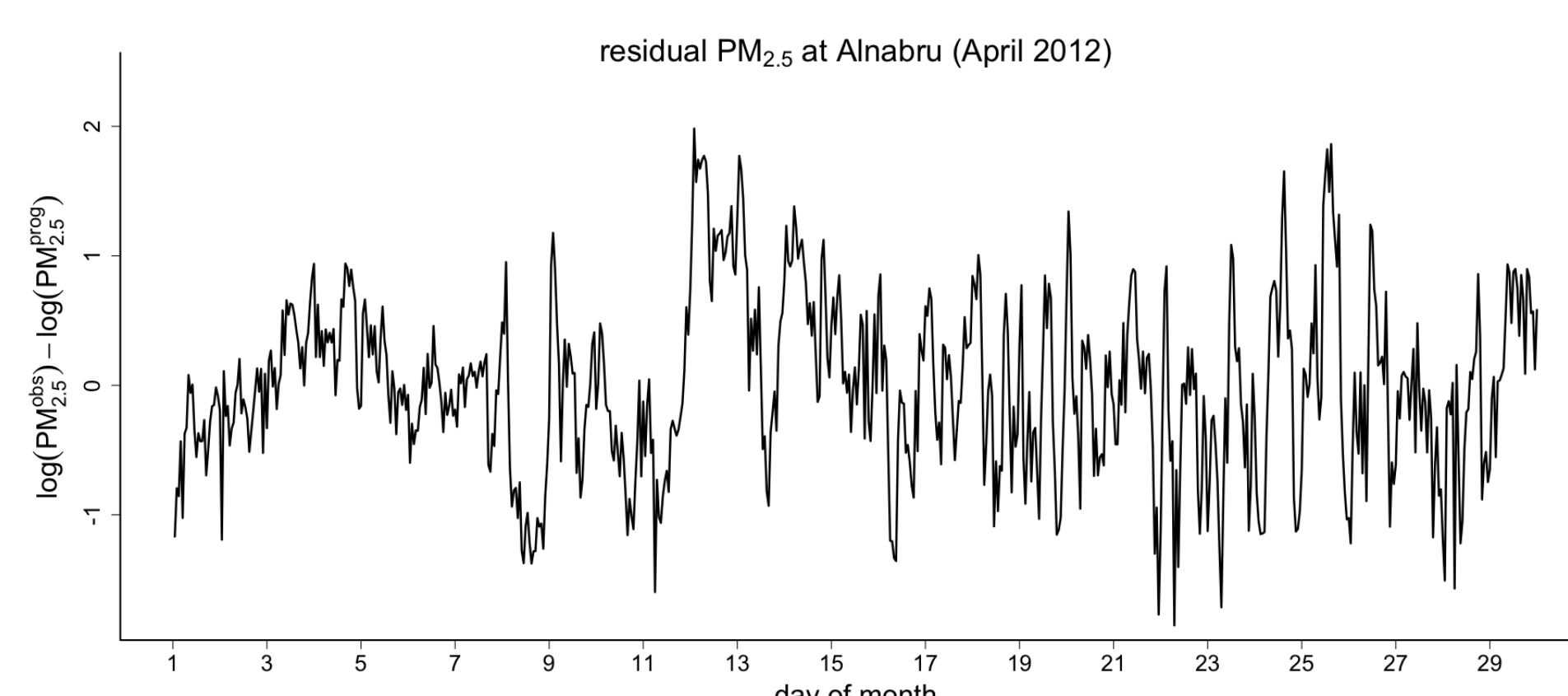


Figure 3: The difference between observed and prognosis (made 24h in advanced) of PM_{2.5} (from Figure 2) seems to be sufficiently stationary.

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 $\mu\text{g}/\text{m}^3$, by constructing a posterior for

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

with $a = \log(25)$, $Y_i = \text{PM}_{2.5,i}^{\text{prog}} + \epsilon_i$ and $\text{data}_k = \{Y_n, \dots, Y_{n-k+1}\}$, by placing a nonparametric envelope around the simple AR(1) model (Figure 4).

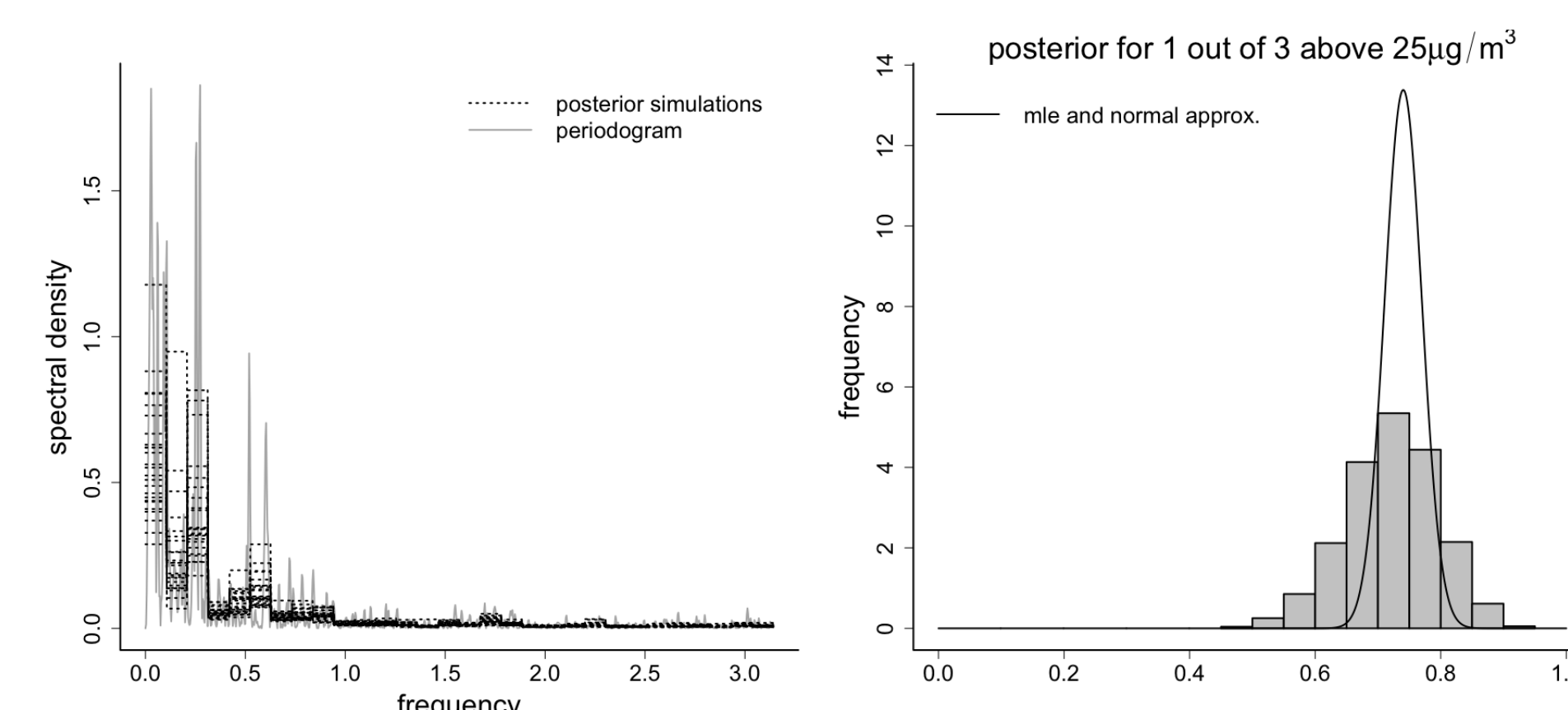


Figure 4: Empirical periodogram and samples from posterior (left), and posterior density for threshold probability and density based on mle for an AR(7) (AIC winner), normal approximation and delta method (right).

Contiguity

Let P_n and Q_n be two sequences of probability measures defined on the same measurable space $(\Omega_n, \mathcal{A}_n)$ for $n \geq 1$. Then P_n and Q_n are said to be mutually contiguous if, for every sequence of sets $A_n \in \mathcal{A}_n$, $P_n(A_n) \rightarrow 0 \Leftrightarrow Q_n(A_n) \rightarrow 0$.

Why Contiguity?

Working under an alternative measure may simplify derivations and conditions. As an example, let $\theta \in \mathbb{R}$ and $\pi(s \mid Y_1, \dots, Y_n)$ be the posterior of $s = \sqrt{n}(\theta - \hat{\theta}_{\text{ML}})$, then asymptotic normality of the posterior follows if

$$\int_{\Theta} |\pi(s \mid Y_1, \dots, Y_n) - \kappa \phi(s\kappa)| ds \rightarrow 0 \quad (3)$$

in P_n -probability, where $\phi(\cdot)$ is the standard normal. Then, if P_n and Q_n are contiguous, and Q_n is sufficiently simpler, we may benefit by working under Q_n to also prove (3) in P_n -probability.

From dependence to independence

If P_n is the probability measure associated with the series $Y_1, \dots, Y_n$, it follows from Choudhuri et al. (2004) that there exist an **contiguous measure** Q_n (Figure 5) which is such that

$$\pi(F_n \mid Y_1, \dots, Y_n) = \pi(F_n \mid I_n(\omega_1), \dots, I_n(\omega_m))$$

under Q_n , where $I_n(\omega_j) \sim f^\circ(\omega_j)E_j$ are independent and $E_j \sim \text{Exp}(1)$.

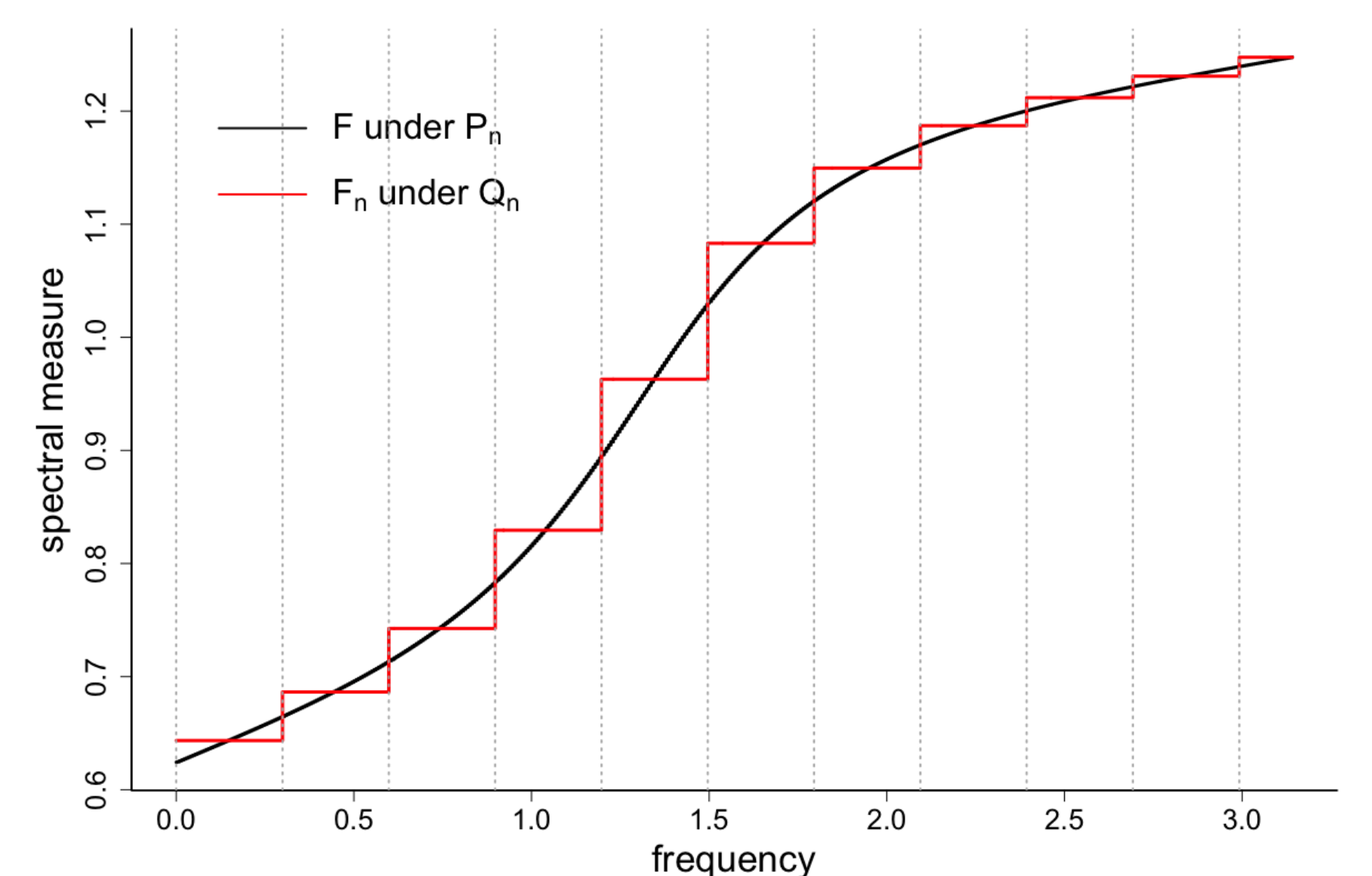


Figure 5: The spectral measures of the two contiguous measures P_n and Q_n . In particular, the log-likelihood under Q_n is exactly equal to the so-called Whittle approximation.

Conditions for Bernshtein–von Mises

Therefore, by working under Q_n (as illustrated in Figure 5) we may apply results from Hermansen and Hjort (2015) to derive simple conditions for Bernshtein–von Mises type of results for time series. In addition to standard regularity assumptions, we need that **the number of windows m_n** in the piecewise constant prior (Figure 1) satisfies

$$m_n/\sqrt{n} \rightarrow 0 \quad \text{and} \quad \sqrt{n}/m_n^2 \rightarrow 0$$

in order to secure the limit (2).

References

- Choudhuri, N., Ghosal, S., and Roy, A. (2004). Contiguity of the Whittle measure for a Gaussian time series. *Biometrika*, 91(1):211–218.
- Hermansen, G. H. and Hjort, N. L. (2015). Bernshtein–von Mises theorems for nonparametric function analysis via locally constant modelling: a unified approach. *Journal of Statistical Planning and Inference*, 166:138–157.
- Ibragimov, I. A. (1963). On estimation of the spectral function of a stationary Gaussian process. *Theory of Probability & Its Applications*, 8(4):366–401.