Resampling for nonstationary stochastic models

Jacek Leśkow

Anna Dudek Łukasz Lenart Rafał Synowiecki

Plan of the presentation

* Nonstationary stochastic models
  - PC and APC models, time domain
  - PC and APC frequency domain
  - periodic counting process models
* Limit results
* Why resampling?
* Selected results
* Future directions of research

Time domain approach

$\{X_t : t \in \mathbb{Z}\}$ - APC, when

$$\mu_X(t) = E(X_t)$$

and the autocovariance function

$$B_X(t,\tau) = \text{cov}(X_t, X_{t+\tau})$$

are almost periodic function at t for every $\tau \in \mathbb{Z}$.
Put $B_X(t,\tau) = \sum_{\lambda\in\Lambda} a(\lambda,\tau)e^{i\lambda\tau}$
Time domain approach ($\mu_X(t) \equiv 0$):

$$\hat{a}_n(\lambda,\tau) = \frac{1}{n-\tau}\sum_{t=1}^{n-\tau} X(t+\tau)X(t)e^{-i\lambda t}$$

Frequency domain approach

Harmonizable time series $\{X(t) : t \in \mathbb{Z}\}$

$$X(t) = \int_0^{2\pi} e^{i\xi t} Z(d\xi).$$

Spectral bimeasure is defined as

$$R((a,b] \times (c,d]) = E[(Z(b) - Z(a))(Z(d) - Z(c))],$$

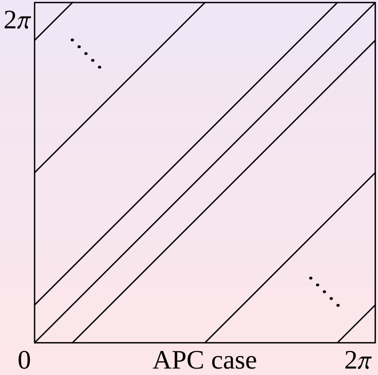
with a support

$$S = \bigcup_{\lambda \in \Lambda} \{(\xi_1, \xi_2) \in (0, 2\pi]^2 : \xi_2 = \xi_1 \pm \lambda\}.$$

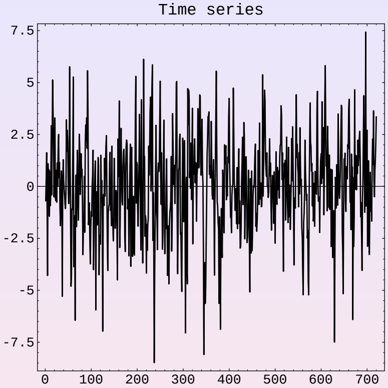
Spectral density estimator

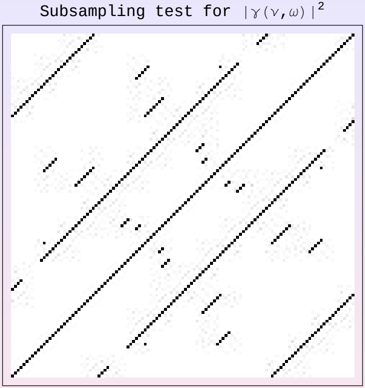
$$\hat{G}_n(\nu, \omega) = \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{s=1}^{n} K_n(s-t) X_t X_s e^{-i\nu t} e^{i\omega s}. \tag{1}$$

Support lines



Simulation example





$$X_t = (2 + \sin(2\pi t/4))Y_{t-1} + Y_t,$$
where Y_t are i.i.d. from $N(0,1)$.

Nonstationary counting process

X - counting process on $[0, T]$.
Intensity of X is of the form

$$\lambda(t) = \lambda_0(t) Y(t) \qquad , \qquad t \in [0, T]$$

- $\lambda_0(t)$ – nonnegative deterministic periodic function
- $Y(t)$ – nonnegative stochastic process

Nonstationary counting process

Sieve estimator of $\lambda_0(t)$

The histogram maximum likelihood estimator of the periodic $\lambda_0(t)$ function is of the form:

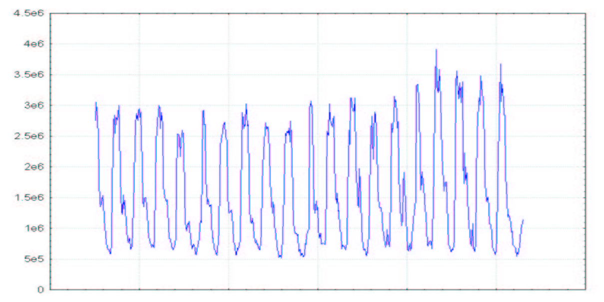
$$\widehat{\lambda}_n(s) = \frac{\sum_{k=1}^{n} X_k(B_n^s)}{\sum_{k=1}^{n} \int_{B_n^s} Y_k(u)du} \mathbf{1}_{D_n}(s), \quad s \in [0, P],$$

where $s \in B_n^s$ is the interval of the length P/b that contains s and

$$D_n = \{\sum_{k=1}^{n} \int_{B_n^s} Y_k(u)du > 0\}.$$

Real data example

Incoming packets number in one hour non-overlapping bins - border between the network of University of Waikato and the internet provider.



Limit results

Asymptotic normality, PC time domain domain

We have

$$\sqrt{n}\,(\hat{a}_n(\lambda,\tau) - a(\lambda,\tau)) \xrightarrow{d} \mathcal{N}_2(0,\Sigma), \qquad (2)$$

where

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix},$$

$$\sigma_{11} = \frac{1}{T}\sum_{s=1}^{T}\sum_{k=-\infty}^{\infty} B_{Z_\tau}(s,k)\cos(\lambda s)\cos(\lambda k),$$

$$\sigma_{22} = \frac{1}{T}\sum_{s=1}^{T}\sum_{k=-\infty}^{\infty} B_{Z_\tau}(s,k)\sin(\lambda s)\sin(\lambda k),$$

$$\sigma_{12} = \sigma_{21} = \frac{1}{T}\sum_{s=1}^{T}\sum_{k=-\infty}^{\infty} B_{Z_\tau}(s,k)\cos(\lambda s)\sin(\lambda k),$$

and $Z(t,\tau) = X(t)X(t+\tau) - B_X(t,\tau)$,
$B_{Z_\tau}(s,k) = Cov\,(Z(s,\tau), Z(s+k,\tau))$.

Asymptotic covariance, APC case

Lemma (Lenart 2008)

If

(i) there exists $\delta > 0$ such that $\sup_{t\in\mathbb{Z}} \|X_t\|_{6+3\delta} \leq \Delta < \infty$,

(ii) $\sum_{k=1}^{\infty} k^2 \alpha(k)^{\frac{\delta}{2+\delta}} \leq K < \infty$,

(iii) $K_n(s-t) = I\{|s-t| \leq w_n\}$ + additional regularity assumptions

then we have a convergence

$$\lim_{n\to\infty} \frac{n}{w_n} cov\left(\hat{G}_n(\nu_1,\omega_1), \hat{G}_n(\nu_2,\omega_2)\right) = P(\nu_1,\nu_2)\overline{P(\omega_1,\omega_2)} + P(\nu_1, 2\pi-\omega_2)\overline{P(\nu_2, 2\pi-\omega_1)},$$

for any $(\nu_1,\omega_2), (\nu_2,\omega_2) \in (0, 2\pi]^2$.

Asymptotic normality, APC case

Theorem (Lenart 2008)

If

(i) there exists $\delta > 0$ such that $\sup_{t\in\mathbb{Z}} \|X_t\|_{6+3\delta} \leq \Delta < \infty$,

(ii) $w_n = O(n^{\kappa})$ for some $\kappa \in (0, \delta/(4+4\delta)$),

(iii) $\sum_{h=1}^{\infty} h^{2r} \alpha(h)^{\frac{\delta}{2(r+1)+\delta}} < \infty$, where r is the integer such that $r > \max\left\{1+\delta, \frac{1-\kappa}{4\kappa}, \frac{\kappa(1+\delta)}{\delta-2\kappa(1+\delta)}\right\}$,

then

$$\sqrt{\frac{n}{w_n}}\left(\hat{G}_n(\nu,\omega) - P(\nu,\omega)\right) \longrightarrow N(0,\Sigma(\nu,\omega)),$$

where matrix $\Sigma(\nu,\omega)$ can be obtained by previous Theorem.

Asymptotic normality, counting process case

Theorem (Dudek, 2008)
If

(i) The Y process is periodically correlated with period P and its mean function $E(Y(s))$ is bounded away from zero.

(ii) Process Y has the third moment bounded.

(iii) Process Y is α–mixing with $\alpha(k) = o(k^{-3})$.

(iv) Each period P is divided into b parts, where $b = O\left(\sqrt{n}\right)$.

(v) The periodic λ_0 function (with the period length equal to P) and $EY(t)$ fulfill the Lipschitz condition on $[0, P]$

then

$$\sqrt{\frac{n}{b}}\left(\widehat{\lambda}_n(s) - \lambda_0(s)\right) \Rightarrow N\left(0, \frac{\lambda_0(s)}{E(Y(s))}\right).$$

Why resampling?

* APC time series case: too complicated asymptotic covariance matrix
* periodic counting process case: slow convergence, need for simultaneous confidence bands

Subsampling for Fourier coefficient of autocovariance function

Consistency holds for the estimator $\hat{\theta}_n = |\hat{a}_n(\lambda, \tau)|$. Let

$$J_n(x, P) = Prob_P\left(\sqrt{n}(|\hat{a}_n(\lambda, \tau)| - |a(\lambda, \tau)|) \leq x\right).$$

By CLT for $\hat{a}_n(\lambda, \tau)$ and the delta method we have

$$J_n(P) \xrightarrow{d} J(P).$$

We define correspondingly subsampling distribution in the form

$$L_{n,b}(P) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} \mathbf{1}\{\sqrt{b}(|\hat{a}_{n,b,t}(\lambda, \tau)| - |\hat{a}_n(\lambda, \tau)|) \leq x\}$$

Subsampling for Fourier coefficient of autocovariance function

Theorem (Lenart, Leśkow, Synowiecki, 2008)

Let $\{X(t) : t \in \mathbb{Z}\}$ be APC time series. Assume that

(i) $b \to \infty$ but $b/n \to 0$,

(ii) $\sup_t E|X(t)|^{4+4\delta} < \infty$,

(iii) $\sum_{k=0}^{\infty} (k+1)^2 \alpha(k)^{\frac{\delta}{4+\delta}} < \infty$,

(iv) the function $V(t, \tau_1, \tau_2, \tau_3) = E\left(X(t)X(t+\tau_1)X(t+\tau_2)X(t+\tau_3)\right)$ is almost periodic.

Then subsampling is consistent, which means that

$$\sup_x |J_n(x, P) - L_{n,b}(x)| \xrightarrow{P} 0.$$

Application of subsampling procedure for PC time series

Testing problem:

$$H_0 : B(\cdot, \tau) \text{ is periodic with period } T_0,$$
$$H_1 : B(\cdot, \tau) \text{ is periodic with period } T_1.$$

Test statistics (Lenart, Leskow, Synowiecki, 2008):

$$U_n(\tau) = \sqrt{n} \left(\sum_{\lambda \in \Lambda_{T_1} \setminus \Lambda_{T_0}} |\hat{a}_n(\lambda, \tau)| \right).$$

Application of subsampling procedure for PC time series

Under H_0:

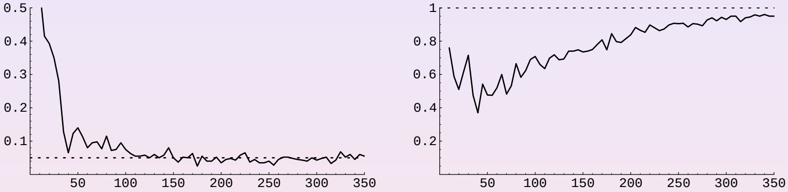
$$U_n(\tau) \xrightarrow{d} J.$$

Under H_1:

$$U_n(\tau) \longrightarrow \infty.$$

Large values of $U_n(\tau)$ suggest that hypothesis H_1 is true. The rejection area is of the form $[c_{1-\alpha}, \infty)$. In order to find $c_{1-\alpha}$ subsampling may be applied.

Application of subsampling procedure for PC time series



(a) Probability of rejection H_0 provided that H_0 is true.

(b) Probability of rejection H_0 provided that H_1 is true.

Figure: Monte Carlo approximations of test errors.

Theorem (Synowiecki, 2007)

Let $\{X_t : t \in \mathbb{Z}\}$ be APC and α-mixing, let $(X_1^*, \ldots, X_n^*)$ be MBB sample, $b \to \infty$ ale $b/n \to 0$. Assume that

(i) $\Lambda = \{\lambda : [0, 2\pi) : M_t(EX_t e^{-i\lambda t}) \neq 0\}$ is finite,

(ii) autocovariance is uniformly summable

(iii) $\sup_{s=1,\ldots,n-b+1} E\left(\frac{1}{\sqrt{b}}\sum_{t=s}^{s+b-1}(X_t - EX_t)\right)^4 < K$

(iv) CLT holds, i.e. $\sqrt{n}\left(\overline{X}_n - M_t(EX_t)\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$

Then MBB procedure is consistent, which means that

$$Var^*(\sqrt{n}\,\overline{X}_n^*) \xrightarrow{P} \sigma^2$$

and

$$\sup_{x\in\mathbb{R}} \left| P\left(\sqrt{n}\left(\overline{X}_n - \mu\right) \le x\right) - P^*\left(\sqrt{n}\left(\overline{X}_n^* - E^*\overline{X}_n^*\right) \le x\right)\right| \xrightarrow{p} 0.$$

Consistency of subsampling - APC case, spectral coherence

Theorem (Lenart, 2008)

Under regularity conditions the subsampling confidence intervals for coherence are consistent

$$P\left(\sqrt{n/w_n}\,(|\hat{\gamma}_n(\nu,\omega)| - |\gamma(\nu,\omega)|) \leq c^{\gamma}_{n,b}(1-\alpha)\right) \longrightarrow 1-\alpha,$$

where $b = b(n) \to \infty$, and $b/n \to 0$,

$$c^{\gamma}_{n,b}(1-\alpha) = \inf\{x :\ L^{\gamma}_{n,b}(x) \geq 1-\alpha\}.$$

$$L^{\gamma}_{n,b}(x) = \frac{1}{n-b+1}\sum_{t=1}^{n-b+1} \mathbf{1}\{\sqrt{b/w_b}(|\hat{\gamma}_{n,b,t}(\nu,\omega)| - |\hat{\gamma}_n(\nu,\omega)|) \leq x\}.$$

Consistency of bootstrap - counting process case

Theorem (Dudek, 2008)

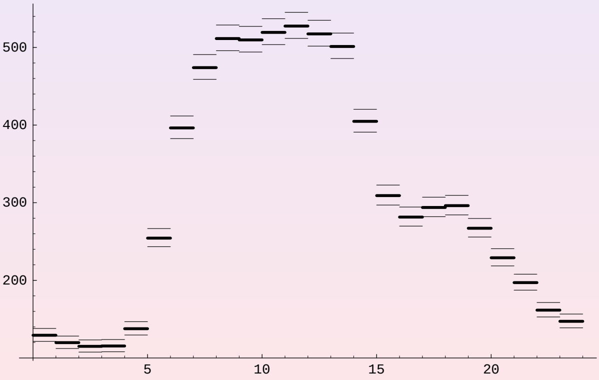
$$\sup_{u \in R} \left| P^* \left(\sqrt{\frac{n}{b}} (\widehat{\lambda}_n^*(s) - \widehat{\lambda}_n(s)) \leq u \right) - P \left(\sqrt{\frac{n}{b}} (\widehat{\lambda}_n(s) - \lambda_0(s)) \leq u \right) \right| = o_P(1),$$

where

$$\widehat{\lambda}_n^*(s) = \frac{\sum_{k=1}^n X_k^*(B_n^s)}{\sum_{k=1}^n \int_{B_n^s} Y_k(u) du} 1_{D_n}(s).$$

Real data example - counting process case

Estimator of the intensity of the number of packets being received by one host together with 90% confidence region:



Future directions of research

* Subsampling - optimal selection of block size
* Resampling in GACS signals
* nonstationary random fields

References

- Chan V., Lahiri S., Meeker W. (2004)
  Block bootstrap estimation of the distribution of cumulative outdoor degradation
  Technometrics
- Dudek A., Goćwin M., Leśkow J., (2008)
  Simultaneous confidence bands for periodic hazard function,
  submitted
- Lenart Ł., Leśkow J., Synowiecki R. (2007)
  Subsampling in testing autocovariance for periodically correlated time series
  Journal of Time Series Analysis, to appear
- Lenart Ł., (2008)
  Asymptotic properties of periodogram for almost periodically correlated time series
  Probability and Mathematical Statistics, to appear

References

Politis D. (2001)
Resampling time series with seasonal components
Proceedings of the 33rd Symposium on the Interface of Computing Science and Statistics

Synowiecki R. (2007)
Consistency and application of MBB for nonstationary time series with periodic and almost periodic structure
Bernoulli

Thank you for your attention