课外资料 2

李 东

清华大学

统计学研究中心

March 19, 2019

Fundamental concepts

In this section, we illustrate some concepts in the framework of stochastic processes. Let $\{X_t : t \in T\}$ be a stochastic process.

Definition 1.1 (ACVF and ACF)

For any $t_1, t_2 \in T$, $\gamma_{t_1,t_2} := \mathrm{cov}(X_{t_1}, X_{t_2})$ *is called the autocovariance function* (ACVF) *of* $\{X_t : t \in T\}$. $\rho_{t_1,t_2} :=$ $\mathrm{corr}(X_{t_1}, X_{t_2})$ *is called the autocorrelation function* (ACF) *of* $\{X_t : t \in T\}$.

Note: $E(X_t^2) < \infty$ is prerequisite for each fixed $t \in T$.

ACVF $\gamma : T \times T \mapsto \mathbb{R}$,

ACF $\rho : T \times T \mapsto [-1, 1]$.

Fundamental concepts

Example 1.1 (continuous ACVF and ACF)

Let $\{\mathbb{B}(t) : t \in [0,\infty)\}$ be the standard Brownian motion, then its ACVF *and* ACF *are*

$$\gamma_{t,s} = \mathrm{cov}(\mathbb{B}(t), \mathbb{B}(s)) = E(\mathbb{B}(t), \mathbb{B}(s)) = t \wedge s;$$
$$\rho_{t,s} = \frac{t \wedge s}{\sqrt{ts}}, \qquad s, t \in [0,\infty),$$

which both are continuous function of $(t,s) \in T \times T$.

Example 1.2 (discrete ACVF and ACF)

Let $X_t = \sum_{i=1}^{t} \eta_i$, where $\{\eta_i\}$ is i.i.d. with mean 0 and finite variance σ^2, and $t \in \mathbb{N}$. Then its ACVF *and* ACF *are*

$$\gamma_{t,s} = \mathrm{cov}(X_t, X_s) = (t \wedge s)\sigma^2; \quad \rho_{t,s} = \frac{t \wedge s}{\sqrt{ts}}, \qquad s, t \in \mathbb{N},$$

which both are discrete function of $(t,s) \in \mathbb{N} \times \mathbb{N}$.

Fundamental concepts

Example 1.3 (continuous ACVF and ACF)

Let $X_t = \sin(\lambda t + U)$, $t \in [0, \infty)$, $\lambda \in (-\pi, \pi)$ *is a real number*, $U \sim \mathcal{U}(-\pi, \pi)$. *Then its* ACVF *and* ACF *are*

$$\gamma_{t,s} = \operatorname{cov}(X_t, X_s) = \frac{1}{2}\cos(\lambda(t-s)).$$

Example 1.4 (discrete ACVF)

Let $X_t = \eta_t + \theta\eta_{t-1}$, *where* $\{\eta_i\}$ *is i.i.d. with mean* 0 *and finite variance* σ^2, *and* $t \in \mathbb{N}$. *Then its* ACVF *is*

$$\gamma_{t,s} = \operatorname{cov}(X_t, X_s) = \begin{cases} (1+\theta^2)\sigma^2, & \text{if } |t-s| = 0, \\ \theta\sigma^2, & \text{if } |t-s| = 1, \\ 0, & \text{if } |t-s| \geq 2. \end{cases}$$

- Decomposition of TS: $X_t = \underbrace{T_t}_{\text{trend}} + \underbrace{S_t}_{\text{seasonal}} + \underbrace{R_t}_{\textit{stochastic part}}.$

Fundamental concepts

Definition 1.2 (weak stationarity/covariance stationarity)

A TS $\{X_t : t \in T\}$ is weakly stationary if $EX_t^2 < \infty$ for each $t \in T$, and

1. $EX_t \equiv \mu$ *is a constant, independent of* t,
2. $\text{cov}(X_t, X_{t+k})$ *is independent of* t *for each* k.

♣ Write: $\gamma_k = \text{cov}(X_t, X_{t+k})$ and $\rho_k = \gamma_k/\gamma_0$.

Definition 1.3 (strict stationarity)

A TS $\{X_t : t \in T\}$ is strictly stationary if $(X_{t_1}, X_{t_2}, ..., X_{t_n})$ and $(X_{t_1+k}, X_{t_2+k}, ..., X_{t_n+k})$ have the same joint distribution for any integer $n \geq 1$ and any integer k.

Fundamental concepts

Example 1.5 (WS but not SS)

Let $\{X_t : t \in T\} = \{\varepsilon, \eta, \varepsilon, \eta, \ldots\}$, i.e., $X_{2t-1} = \varepsilon$, $X_{2t} = \eta$, where $\varepsilon \sim \mathcal{N}(0,1)$, $\eta \sim \mathcal{U}(-\sqrt{3}, \sqrt{3})$, and ε and η are independent. Then $\{X_t\}$ is WS, not SS.

Solution. Clearly, $EX_t = 0$, $EX_t^2 = 1 < \infty$, and

$$\gamma_{t,t+s} = \mathrm{cov}(X_t, X_{t+s}) = \begin{cases} \mathrm{cov}(X_t, X_t) = 1, & \text{if } s \text{ is even,} \\ \mathrm{cov}(\varepsilon, \eta) = 0, & \text{if } s \text{ is odd.} \end{cases}$$

Hence, $\gamma_{t,t+s}$ is independent of t and only depends on s.

Thus, $\{X_t\}$ is WS. However, $\{X_t\}$ is not SS since ε and η do not have the same distribution functions.

Fundamental concepts

Example 1.6 (SS but not WS)

A TS $\{X_t : t \in T\} = \{\varepsilon_t : t \in T\}$ *is i.i.d. standard Cauchy distribution with density* $f(x) = \frac{1}{\pi(1+x^2)}$. *Then,* $\{X_t\}$ *is SS, not WS in that* $EX_t^2 = \infty$.

♣: Can you give a TS with finite second moments such that it is SS not WS?

Fundamental concepts

Example 1.6 (SS but not WS)

A TS $\{X_t : t \in T\} = \{\varepsilon_t : t \in T\}$ is i.i.d. standard Cauchy distribution with density $f(x) = \frac{1}{\pi(1+x^2)}$. Then, $\{X_t\}$ is SS, not WS in that $EX_t^2 = \infty$.

♣: Can you give a TS with finite second moments such that it is SS not WS?

Theorem 1.1

If a TS $\{X_t : t \in T\}$ is SS with $EX_t^2 < \infty$, then $\{X_t\}$ is WS.

Fundamental concepts

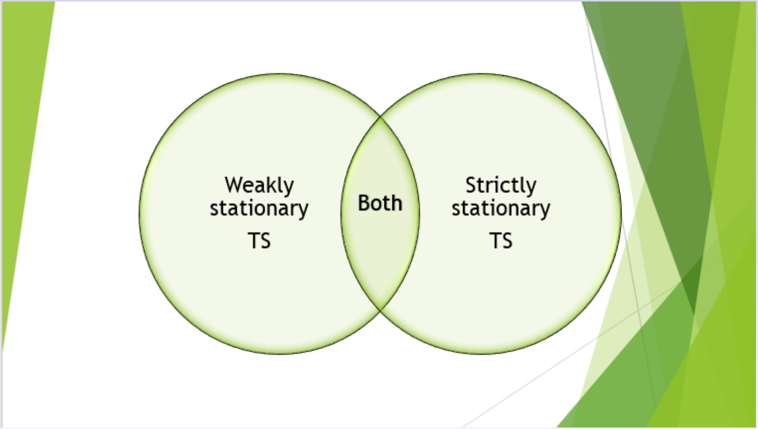
Example 1.7 (WS & SS)

A TS $\{X_t : t \in T\}$ *is i.i.d. with* $X_t \sim \mathcal{N}(0,1)$. *Then* $\{X_t : t \in T\}$ *is both WS and SS.*

Example 1.8 (neither SS nor WS)

A TS $\{\varepsilon, \eta, \xi_3, \xi_4, \dots\}$, *where* $\varepsilon \sim \mathcal{N}(0,1)$, $\eta \sim t_1$, *and* $\{\xi_j : j = 3, 4, \dots\}$ *is i.i.d.* $\sim \mathcal{E}(1)$. *All of them are independent. Then it is neither SS nor WS.*

Fundamental concepts



Fundamental concepts

■ For a WS process $\{X_t\}$, $\gamma_k = \text{cov}(X_t, X_{t+k})$ and ρ_k have the following properties.

Property 1.1 (ACVF and ACF)

1. $\gamma_0 = \text{var}(X_t); \quad \rho_0 = 1.$
2. $|\gamma_k| \leq \gamma_0; \quad |\rho_k| \leq 1.$
3. $\gamma_k = \gamma_{-k}; \quad \rho_k = \rho_{-k}.$
4. *They are positive semidefinite in the sense that*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j \gamma_{|t_i - t_j|} \geq 0, \quad \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j \rho_{|t_i - t_j|} \geq 0,$$

for any $n \in \mathbb{N}$, any set of time point $\{t_1, t_2, ..., t_n\}$ and any real number $\alpha_1, ..., \alpha_n$.

Fundamental concepts

Theorem 1.2 (Characterization of ACVF)

A real-valued function $\gamma_{(\cdot)}$ defined on the integers is the ACVF of a stationary TS if and only if it is even and positive semidefinite in the sense that

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\gamma_{|t_i-t_j|}\geq 0,$$

for any integer $n \geq 1$, any set of time point $\{t_1, t_2, ..., t_n\}$ and any real number $\alpha_1, ..., \alpha_n$.

Fundamental concepts

For WS TS, generally, the following matrix (a special Toeplitz matrix) is positive semidefinite.

$$\Gamma_n = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-2} & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-3} & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-4} & \gamma_{n-3} \\ \cdots & \cdots & \cdots & \ddots & \cdots & \cdots \\ \gamma_{n-2} & \gamma_{n-3} & \gamma_{n-4} & \cdots & \gamma_0 & \gamma_1 \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_1 & \gamma_0 \end{pmatrix}_{n\times n}$$

for any $n \geq 1$.

Example 1.9

The real function on $\mathbb{Z}$

$$\kappa(h) = \begin{cases} 1, & \text{if } h = 0, \\ \rho, & \text{if } h = \pm 1, \\ 0, & \text{otherwise}, \end{cases}$$

is an ACF if and only if $|\rho| \le 1/2$.

Solution. If $|\rho| \le 1/2$, construct a process, that is okay. If $\rho > 1/2$, $\mathbf{K} = (\kappa(i-j))_{i,j=1}^n$ and $\mathbf{a}$ is the n-component vector $\mathbf{a} = (1, -1, 1, -1, \ldots)'$, then

$$\mathbf{a}'\mathbf{K}\mathbf{a} = n - 2(n-1)\rho < 0 \quad \text{for} \quad n > 2\rho/(2\rho - 1),$$

which shows that $\kappa(\cdot)$ is not positive semidefinite.
If $\rho < -1/2$, using $\mathbf{a} = (1, 1, \ldots)'$ again shows that $\kappa(\cdot)$ is not positive semidefinite.

Fundamental concepts

♣ Alternative proof by the determinant of $\mathbf{K}_n$:

$$\mathbf{K}_n = \begin{pmatrix} 1 & \rho & 0 & \cdots & 0 & 0 \\ \rho & 1 & \rho & \cdots & 0 & 0 \\ 0 & \rho & 1 & \cdots & 0 & 0 \\ \ldots & \ldots & \ldots & \ddots & \ldots & \ldots \\ 0 & 0 & 0 & \cdots & 1 & \rho \\ 0 & 0 & 0 & \cdots & \rho & 1 \end{pmatrix}_{n\times n}$$

for any $n \geq 1$. Clearly, $\det(\mathbf{K}_n) = \det(\mathbf{K}_{n-1}) - \rho^2 \det(\mathbf{K}_{n-2})$ with $\det(\mathbf{K}_1) = 1, \det(\mathbf{K}_2) = 1 - \rho^2$.

Fundamental concepts

■ Background. The correlation between X_t and X_{t+k} after their **mutual linear dependency** on the intervening variables $X_{t+1}, \ldots, X_{t+k-1}$ has been removed.

Definition 1.4 (PACF)

Let $\{X_t\}$ be WS with $EX_t = 0$. The partial autocorrelation function (PACF) is defined as $\phi_{11} = \rho_1$ and

$$\phi_{kk} = \operatorname{corr}(R_{t|t+1,\ldots,t+k-1}, R_{t+k|t+1,\ldots,t+k-1}) \quad \textit{for } k \geq 2,$$

where $R_{j|t+1,\ldots,t+k-1}$ is the residual from the linear regression of X_j on $(X_{t+1}, \ldots, X_{t+k-1})$, namely,

$$R_{j|t+1,\ldots,t+k-1} = X_j - (\alpha_{j1}X_{t+1} + \ldots + \alpha_{j,k-1}X_{t+k-1}),$$

and $(\alpha_{j1}, \ldots, \alpha_{j,k-1})$

$$= \arg\min_{\beta_1,\ldots,\beta_{k-1}} E\{X_j - (\beta_{j1}X_{t+1} + \ldots + \beta_{j,k-1}X_{t+k-1})\}^2.$$

Fundamental concepts

♣ In the definition above, we assume that $EX_t = 0$ to simplify the notation.

♣ For a Gaussian process, the PACF is in fact equal to

$$\phi_{kk} = \operatorname{corr}(X_t, X_{t+k} | X_{t+1}, ..., X_{t+k-1}).$$

■ Formula for ϕ_{kk}:

$$\phi_{kk} = \frac{\gamma_k - \operatorname{cov}(X_{t+k}, \mathbf{X}^{\tau}_{k-1:1})\Sigma^{-1}_{k-1:1}\operatorname{cov}(\mathbf{X}_{k-1:1}, X_{t+k})}{\gamma_0 - \operatorname{cov}(X_t, \mathbf{X}^{\tau}_{k-1:1})\Sigma^{-1}_{k-1:1}\operatorname{cov}(\mathbf{X}_{k-1:1}, X_t)},$$

where $\mathbf{X}_{k-1:1} = (X_{t+k-1}, X_{t+k-2}, ..., X_{t+1})^{\tau}$ and $\Sigma_{k-1:1} = \operatorname{var}(\mathbf{X}_{k-1:1}) := \operatorname{cov}(\mathbf{X}_{k-1:1}, \mathbf{X}_{k-1:1})$.

Fundamental concepts

- Simple formula for ϕ_{kk}:

$$\phi_{11} = \rho_1,$$

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}_{k\times k}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}_{k\times k}}, \quad k \geq 2.$$

Note: Because by definition $\rho_0 = \phi_{00} = 1$ for any process, when we talk about the ACF and PACF, we refer only to ρ_k and ϕ_{kk} for $k \neq 0$.

Fundamental concepts

Durbin's recursive formula for ϕ_{kk}:

$$\phi_{11} = \rho_1,$$

$$\phi_{k+1,k+1} = \frac{\rho_{k+1} - \sum_{j=1}^{k} \phi_{k,j}\rho_{k+1-j}}{1 - \sum_{j=1}^{k} \phi_{k,j}\rho_j},$$

$$\phi_{k+1,j} = \phi_{k,j} - \phi_{k+1,k+1}\phi_{k,k+1-j}, \quad j = 1, ..., k.$$

Some important stochastic processes

Definition 2.1 (white noise)

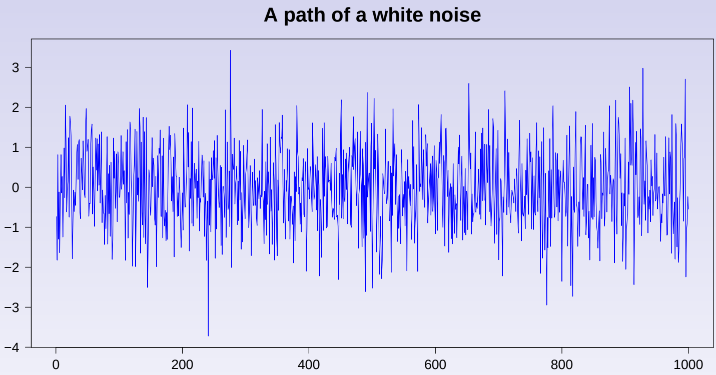
Let $\{\varepsilon_t\}$ *be a stationary TS. If for any* $s, t \in \mathbb{N}$,

$$E\varepsilon_t = \mu, \quad \mathrm{cov}(\varepsilon_t, \varepsilon_s) = \begin{cases} \sigma^2, & t = s, \\ 0, & t \neq s, \end{cases} = \sigma^2 \delta_{t-s},$$

then $\{\varepsilon_t\}$ *is called a* **white noise**, *write* $\{\varepsilon_t\} \sim \mathrm{WN}(\mu, \sigma^2)$.

1. *if* $\{\varepsilon_t\}$ *is independent, then call* $\{\varepsilon_t\}$ **independent white noise**, *write* $\mathrm{IWN}(\mu, \sigma^2)$.
2. *if* $\mu = 0$, *then call* $\{\varepsilon_t\}$ **zero-mean white noise**, *write* $\mathrm{WN}(0, \sigma^2)$. *(most commonly used)*
3. *if* $\mu = 0, \sigma^2 = 1$, *then call* $\{\varepsilon_t\}$ **standard white noise**, *write* $\mathrm{WN}(0, 1)$.
4. *For* $\mathrm{IWN}(\mu, \sigma^2)$, *if* ε_t *is normal, then call* $\{\varepsilon_t\}$ **normal white noise**. *(i.e.,* $\{\varepsilon_t\}$ *i.i.d.* $\mathcal{N}(\mu, \sigma^2)$*)*

Fundamental concepts



■ In signal processing, WS is a random signal having equal intensity at different frequencies, giving it a constant power spectral density. WS refers to a statistical model for signals and signal sources, rather than to any specific signal.

Some important stochastic processes

Definition 2.2 (Poisson process)

If a continuous-time stochastic process $\{N(t) : t \in [0, \infty)\}$ *satisfies for any* $t > s \geq 0$ *and any nonnegative integer* k

1. $N(0) = 0$ *a.s.*
2. $\{N(t)\}$ *has independent increments, i.e., for any* $n > 1$ *and* $0 \leq t_0 < t_1 < \cdots < t_n$*, r.v.s* $\{N(t_j) - N(t_{j-1}) : j = 1, 2, \ldots, n\}$ *are independent,*
3. $\mathbb{P}(N(t) - N(s) = k) = \frac{(\lambda(t-s))^k}{k!} \exp(-\lambda(t-s))$*, where* $\lambda > 0$,

then $\{N(t) : t \in [0, \infty)\}$ *is called a* **Poisson process** *with jump rate* λ.

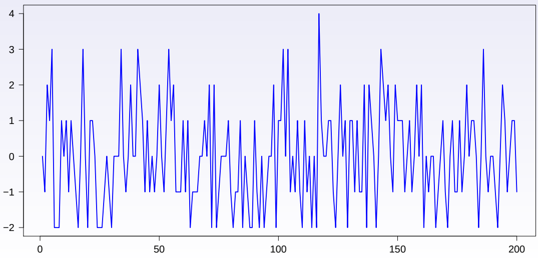
■ Facts: $E\{N(t)\} = \lambda t$ and $\mathrm{var}(N(t)) = \lambda t$.

Definition 2.3 (Poisson white noise)

Let $\{N(t) : t \in [0, \infty)\}$ *be a Poisson process with jump rate* λ*, define*

$$\varepsilon_n = N(n+1) - N(n) - \lambda, \quad n = 1, 2, \ldots.$$

Clearly, $\{\varepsilon_n\}$ *is an* $\mathrm{IWN}(0, \lambda)$ *and is called a* **Poisson white noise***.*



Some important stochastic processes

Definition 2.4 (Brownian motion/Wiener process)

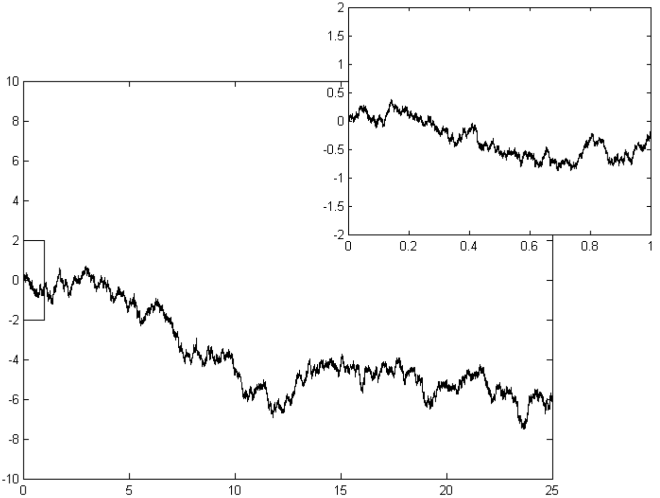
If a continuous-time stochastic process $\{\mathbb{B}(t) : t \in [0, \infty)\}$ *satisfies*

1. $\mathbb{B}(0) = 0$ *a.s.*
2. $\mathbb{B}(t)$ *has independent increments, i.e.,* $\mathbb{B}(t+u) - \mathbb{B}(t)$ *and* $\{\mathbb{B}(s) : s \le t\}$ *are independent for any* $u \ge 0$,
3. $\mathbb{B}(t+u) - \mathbb{B}(t) \sim \mathcal{N}(0, u\sigma^2)$,
4. $\mathbb{B}(t)$ *has continuous paths: with probability* 1, $\mathbb{B}(t)$ *is continuous in* t.

then $\{\mathbb{B}(t) : t \in [0, \infty)\}$ *is called a* **Brownian motion** *or* **Wiener process**.

■ Facts: $E\mathbb{B}(t) = 0$ and $\mathrm{var}(\mathbb{B}(t)) = \sigma^2 t$.

Some important stochastic processes



Some important stochastic processes

Definition 2.5 (orthogonal TS)

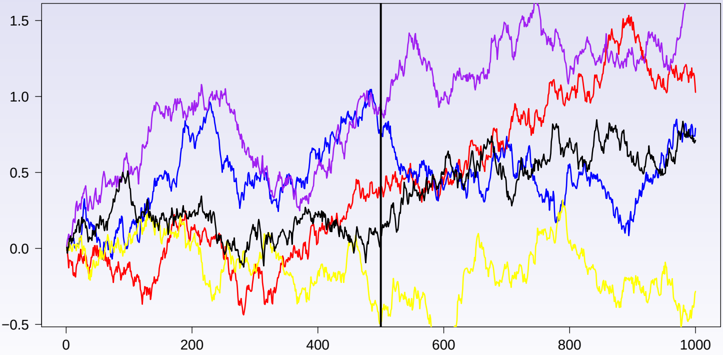
For WS TS $\{X_t\}$ *and* $\{Y_t\}$, *for any* $t, s \in T$, *if* $E(X_s Y_t) = 0$, *then call* $\{X_t\}$ *and* $\{Y_t\}$ *are* **orthogonal**. *If* $\operatorname{cov}(X_s, Y_t) = 0$, *then call* $\{X_t\}$ *and* $\{Y_t\}$ *are* **uncorrelated**.

Definition 2.6 (Gaussian TS)

For a TS $\{X_t\}$, *if for any* $n \geq 1$ *and any* $t_1, ..., t_n \in T$, $(X_{t_1}, ..., X_{t_n})$ *is multivariate normal, then* $\{X_t\}$ *is called* **Gaussian time series.** *Further, if it is stationary, then* $\{X_t\}$ *is called* **Gaussian stationary time series.**

Estimation of mean, ACVF, ACF, PACF

■ A WS TS is characterized by its mean μ, variance σ^2, ACF ρ_k and PACF ϕ_{kk}. The exact values of these parameters can be calculated if the ensemble of all possible realizations is known.



Estimation of mean

Question 1

Assume only one realization $\{X_1, X_2, ..., X_n\}$ is available for a WS TS $\{X_t\}$, how to estimate μ, σ^2, ρ_k, γ_k and ϕ_{kk}?

■ Sample mean: $\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$.

Property 3.1 ($\bar{X}_n$)

1. *Unbiasedness:* $E(\bar{X}_n) = \frac{1}{n}\sum_{t=1}^{n} E(X_t) = \mu$.
2. *Consistency: if* $\mathrm{var}(\bar{X}_n) \to 0$, *then* $\bar{X}_n \to \mu$ *in* L_2.

Estimation of mean

Note that

$$\mathrm{var}(\bar{X}_n) = \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\mathrm{cov}(X_t, X_s) = \frac{\gamma_0}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\rho_{t-s}$$

$$= \frac{\gamma_0}{n}\sum_{k=-(n-1)}^{n-1}\Big(1 - \frac{|k|}{n}\Big)\rho_k.$$

If $\rho_k \to 0$, then $\mathrm{var}(\bar{X}_n) \to 0$ by L'Hôpital's rule, which implies that

$$\bar{X}_n \to \mu, \quad \text{in } L_2.$$

Theorem 3.1 (The mean square ergodic theorem)

Suppose $\{X_t\}$ is WS, then,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\rho_k = 0 \Longleftrightarrow \lim_{n\to\infty} E[|\bar{X}_n - \mu|^2] = 0.$$

Estimation of mean

Theorem 3.2 (General mean square ergodic theorem)

Suppose $\{X_t\}$ is WS. Then there exists a random variable X^* such that

$$\lim_{n\to\infty} E[|\bar{X}_n - X^*|^2] = 0.$$

Theorem 3.3 (Strong ergodic theorem)

Suppose $\{X_t\}$ is SS with $E|X_1| < \infty$. Then there exists a random variable X^* such that

$$\bar{X}_n \to X^* \quad a.s.$$

♣ Reference: Chapter 9 in the book 《A First Course in Stochastic Processes》 (2nd) by Samuel Karlin and Howard M. Taylor.

Estimation of mean

Stationarity and Ergodicity

Theorem 3.4

If $\{X_t\}$ is strictly stationary and ergodic with $E|X_1| < \infty$, then the strong law of large numbers still holds, i.e.,

$$\frac{1}{n}\sum_{t=1}^{n} X_t \to \mu, \quad a.s.$$

Illustration:

$$\begin{tikzcd}
\frac{1}{n}\sum_{t=1}^{n} X_t \arrow[r, "SS"] \arrow[dr, "SS\&Er"'] & Y^* \arrow[d, "Er"] \\
 & \mu
\end{tikzcd}$$

Estimation of ACVF

- Sample ACVF:

$$\widehat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n)$$

or

$$\widehat{\widehat{\gamma}}_k = \frac{1}{n-k}\sum_{t=1}^{n-k}(X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n).$$

By a simple calculation, we have

$$E(\widehat{\gamma}_k) \approx \gamma_k - \frac{k}{n}\gamma_k - \left(1 - \frac{k}{n}\right)\mathrm{var}(\bar{X}_n),$$

$$E(\widehat{\widehat{\gamma}}_k) \approx \gamma_k - \mathrm{var}(\bar{X}_n).$$

Estimation of ACVF

Question 2

Consider the asymptotic limiting distribution of the joint $\{\widehat{\gamma}_k\}$ *or* $\{\widehat{\rho}_k\}$*:*

$$\sqrt{n}\left\{\begin{pmatrix}\widehat{\gamma}_0\\ \widehat{\gamma}_1\\ \widehat{\gamma}_2\\ \vdots\\ \widehat{\gamma}_m\end{pmatrix}-\begin{pmatrix}\gamma_0\\ \gamma_1\\ \gamma_2\\ \vdots\\ \gamma_m\end{pmatrix}\right\}\Longrightarrow ?$$

$$\sqrt{n}\left\{\begin{pmatrix}\widehat{\rho}_1\\ \widehat{\rho}_2\\ \vdots\\ \widehat{\rho}_m\end{pmatrix}-\begin{pmatrix}\rho_1\\ \rho_2\\ \vdots\\ \rho_m\end{pmatrix}\right\}\Longrightarrow ?$$

Estimation of ACVF

Further, if $\{X_t\}$ is Gaussian, Bartlett (1946) has shown

$$\operatorname{cov}(\widehat{\gamma}_k, \widehat{\gamma}_{k+j}) \approx \frac{1}{n} \sum_{i=-\infty}^{\infty} (\gamma_i \gamma_{i+j} + \gamma_{i+k+j} \gamma_{i-k}),$$

$$\operatorname{var}(\widehat{\gamma}_k) \approx \frac{1}{n} \sum_{i=-\infty}^{\infty} (\gamma_i^2 + \gamma_{i+j} \gamma_{i-k}).$$

Similarly,

$$\operatorname{cov}(\widehat{\widehat{\gamma}}_k, \widehat{\widehat{\gamma}}_{k+j}) \approx \frac{1}{n-k} \sum_{i=-\infty}^{\infty} (\gamma_i \gamma_{i+j} + \gamma_{i+k+j} \gamma_{i-k}),$$

$$\operatorname{var}(\widehat{\widehat{\gamma}}_k) \approx \frac{1}{n-k} \sum_{i=-\infty}^{\infty} (\gamma_i^2 + \gamma_{i+j} \gamma_{i-k}).$$

In practice, $\widehat{\gamma}_k$ is better than $\widehat{\widehat{\gamma}}_k$ since $\operatorname{var}(\widehat{\gamma}_k) < \operatorname{var}(\widehat{\widehat{\gamma}}_k)$.

■ More importantly, $\{\widehat{\gamma}_k : k = 0, \pm 1, ...\}$ is positive semidefinite, however, $\{\widehat{\widehat{\gamma}}_k : k = 0, \pm 1, ...\}$ is not.

Estimation of ACF

- Sample ACF:

$$\widehat{\rho}_k = \frac{\widehat{\gamma}_k}{\widehat{\gamma}_0} = \frac{\sum_{t=1}^{n-k}(X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n)}{\sum_{t=1}^{n}(X_t - \bar{X}_n)^2}.$$

Further, if $\{X_t\}$ is Gaussian, Bartlett (1946) has shown

$$\operatorname{cov}(\widehat{\rho}_k, \widehat{\rho}_{k+j}) \approx \frac{1}{n}\sum_{i=-\infty}^{\infty}(\rho_i\rho_{i+j} + \rho_{i+k+j}\rho_{i-k} - 2\rho_k\rho_i\rho_{i-k-j} - 2\rho_{k+j}\rho_i\rho_{i-k} + 2\rho_k\rho_{k+j}\rho_i^2)$$

for $k > 0$ and $k + j > 0$.

For large n, $\sqrt{n}(\widehat{\rho}_k - \rho_k) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$, where

$$\sigma^2 = \frac{1}{n}\sum_{i=-\infty}^{\infty}(\rho_i^2 + \rho_{i+k}\rho_{i-k} - 4\rho_k\rho_i\rho_{i-k} + 2\rho_k^2\rho_i^2). \quad (\approx \operatorname{var}(\widehat{\rho}_k)).$$

Estimation of ACF

For processes in which $\rho_k = 0$ for $k > q$, (e.g. MA(q) model), Bartlett's approximation becomes

$$\mathrm{var}(\widehat{\rho}_k) \approx \frac{1}{n}(1 + 2\rho_1^2 + \ldots + 2\rho_q^2).$$

In practice, $\rho_i (i = 1, \ldots, q)$ are unknown and are replaced by their sample estimates $\widehat{\rho}_i$ and we have the following large-lag standard error of $\widehat{\rho}_k$:

$$s_{\widehat{\rho}_k} \approx \sqrt{\frac{1}{n}(1 + 2\widehat{\rho}_1^2 + \ldots + 2\widehat{\rho}_q^2)}.$$

Particularly, to test a white noise process, we use

$$s_{\widehat{\rho}_k} \approx \sqrt{1/n}.$$

Estimation of PACF

- Sample PACF: Formula for $\widehat{\phi}_{kk}$:

$$\widehat{\phi}_{11} = \widehat{\rho}_1,$$

$$\widehat{\phi}_{kk} = \frac{\begin{vmatrix} 1 & \widehat{\rho}_1 & \widehat{\rho}_2 & \cdots & \widehat{\rho}_{k-2} & \widehat{\rho}_1 \\ \widehat{\rho}_1 & 1 & \widehat{\rho}_1 & \cdots & \widehat{\rho}_{k-3} & \widehat{\rho}_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \widehat{\rho}_{k-1} & \widehat{\rho}_{k-2} & \widehat{\rho}_{k-3} & \cdots & \widehat{\rho}_1 & \widehat{\rho}_k \end{vmatrix}_{k\times k}}{\begin{vmatrix} 1 & \widehat{\rho}_1 & \widehat{\rho}_2 & \cdots & \widehat{\rho}_{k-2} & \widehat{\rho}_{k-1} \\ \widehat{\rho}_1 & 1 & \widehat{\rho}_1 & \cdots & \widehat{\rho}_{k-3} & \widehat{\rho}_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \widehat{\rho}_{k-1} & \widehat{\rho}_{k-2} & \widehat{\rho}_{k-3} & \cdots & \widehat{\rho}_1 & 1 \end{vmatrix}_{k\times k}}, \qquad k \geq 2.$$

Estimation of PACF

Durbin's recursive formula for $\widehat{\phi}_{kk}$:

$$\widehat{\phi}_{11} = \widehat{\rho}_1,$$

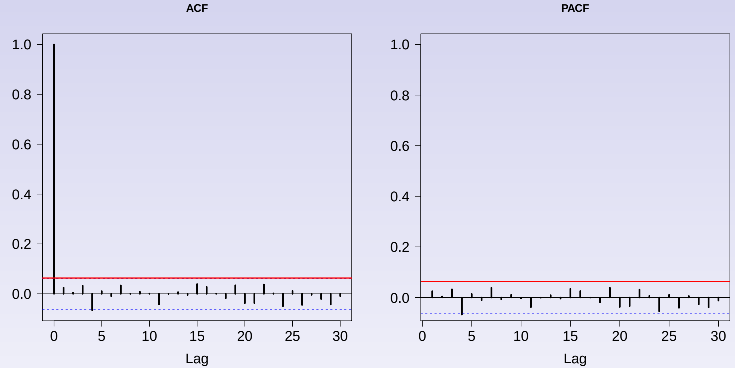
$$\widehat{\phi}_{k+1,k+1} = \frac{\widehat{\rho}_{k+1} - \sum_{j=1}^{k} \widehat{\phi}_{k,j}\widehat{\rho}_{k+1-j}}{1 - \sum_{j=1}^{k} \widehat{\phi}_{k,j}\widehat{\rho}_j},$$

$$\widehat{\phi}_{k+1,j} = \widehat{\phi}_{k,j} - \widehat{\phi}_{k+1,k+1}\widehat{\phi}_{k,k+1-j}, \quad j = 1, ..., k.$$

It was shown by Quenouille(1949) that on the hypothesis that underlying process is a white noise sequence,

$$\mathrm{var}(\widehat{\phi}_{kk}) \approx \frac{1}{n}.$$

Estimation of ACF and PACF



R code:

```
par(mfrow=c(1,2))
x=rnorm(1000)
acf(x,cex.lab=1.5,cex.axis=1.5,las=1,main="ACF",ylab="",lwd=3)
abline(h=2/sqrt(n),col="red",lwd=2)
pacf(x,cex.lab=1.5,cex.axis=1.5,las=1,main="PACF",ylim=c(-0.05,1),ylab=“”,lwd=3)

abline(h=2/sqrt(n),col="red",lwd=2)
```

MA(∞) representation

■ MA(∞) representation

X_t as a linear combination of a sequence of uncorrelated r.v.s:

$$X_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + ... := \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

where $\psi_0 = 1$, $\{\varepsilon_t\}$ is a zero mean white noise process, and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$.

♣ Backshift operator B: $B^k X_t = X_{t-k}$.

In compact form:

$$X_t - \mu = \psi(B)\varepsilon_t,$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$.

MA(∞) representation

Property 4.1 (some mathematical characteristics)

1. $E(X_t) = \mu;$
2. $\text{var}(X_t) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j^2;$
3. $\gamma_k = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}; \quad k \geq 0.$
4. $\rho_k = \dfrac{\sum_{j=0}^{\infty} \psi_j \psi_{j+k}}{\sum_{j=0}^{\infty} \psi_j^2}.$

MA(∞) representation

Definition 4.1 (autocovariance generating function)

For a given sequence of ACVF $\{\gamma_k\}$, the autocovariance generating function is defined as

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k.$$

Then

$$\begin{aligned}
\gamma(B) &= \sigma_\varepsilon^2 \sum_{k=-\infty}^{\infty} \sum_{i=0}^{\infty} \psi_i \psi_{i+k} B^k \\
&= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j B^{j-i} \\
&= \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j B^j \sum_{i=0}^{\infty} \psi_i B^{-i} \\
&= \sigma_\varepsilon^2 \psi(B) \psi(B^{-1}),
\end{aligned}$$

MA(∞) representation

Definition 4.2 (autocorrelation generating function)

For a given sequence of ACF $\{\rho_k\}$*, the autocorrelation generating function is defined as*

$$\rho(B) = \sum_{k=-\infty}^{\infty} \rho_k B^k = \frac{\gamma(B)}{\gamma_0}.$$

AR(∞) representation

■ AR(∞) representation

$$X_t - \mu = \pi_1(X_{t-1} - \mu) + \pi_2(X_{t-2} - \mu) + ... + \varepsilon_t$$

or $\pi(B)(X_t - \mu) = \varepsilon_t$, where $\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$, and $\sum_{j=1}^{\infty} |\pi_j| < \infty$.

♣ Explanation:

1. Regard the value of X at time t on its own past values plus a random shock.
2. The random shock can be recovered by the values of X's.

♣ Usefulness: understanding the mechanism of forecasting.

■ **Invertibility** if it can be written in this AR(∞) form.

Linear difference equation

- optional material: Linear difference equation. See the textbook(page 26-30).

ARMA model

Definition 6.1 (ARMA(p,q) model)

The TS/process $\{y_t : t = 0, \pm 1, \pm 2, \dots\}$ is said to be an autoregressive and moving average model/process if it satisfies

$$y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \quad (1)$$

where $\{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2)$, $p, q \geq 0$ are integers, and (p,q) are called the order of the model. Write $\{y_t\} \sim \mathrm{ARMA}(p,q)$.

Say that $\{y_t\}$ is an ARMA(p,q) with mean μ if $\{y_t - \mu\} \sim$ ARMA(p,q).

- B denotes the backshift operator: $B^k y_t = y_{t-k}$, and $\phi(\cdot)$ and $\theta(\cdot)$ are polynomials defined as

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p, \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q.$$

Then $\phi(B) y_t = \theta(B) \varepsilon_t$ if $\{y_t\} \sim \mathrm{ARMA}(p,q)$.

AR model

- **The first-order autoregressive model: AR(1)**
  $\{y_t\} \sim \text{AR}(1)$, i.e., $y_t = \phi y_{t-1} + \varepsilon_t$.

- Invertibility: AR(1) process is **always invertible**.

- Stationarity ???

  Consider four cases:

  1. if $|\phi| < 1$, then $y_t = \varepsilon_t + \sum_{j=1}^{\infty} \phi^j \varepsilon_{t-j}$.
  2. if $\phi = 1$, then $y_t = \varepsilon_t + \varepsilon_{t-1} + \ldots + \varepsilon_1 + y_0$.
  3. if $\phi = -1$, then $y_t = \varepsilon_t + \sum_{j=1}^{t-1} (-1)^j \varepsilon_{t-j} + (-1)^t y_0$.
  4. if $|\phi| > 1$, then $y_t = -\frac{\varepsilon_{t+1}}{\phi} + \frac{y_{t+1}}{\phi} = -\sum_{j=1}^{\infty} \frac{\varepsilon_{t+j}}{\phi^j}$.

Thus, AR(1) model is weak stationary if and only if $|\phi| \neq 1$.

Noncausal autoregressive model

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2017)
79, *Part* 3, *pp.* 737–756

Local explosion modelling by non-causal process

Christian Gouriéroux

Centre de Recherche en Economie et Statistique, Malakoff, France, and University of Toronto, Canada

and Jean-Michel Zakoïan

Centre de Recherche en Economie et Statistique, Malakoff, and Université de Lille, France

[Received June 2015. Final revision May 2016]

Summary. The non-causal auto-regressive process with heavy-tailed errors has non-linear causal dynamics, which allow for local explosion or asymmetric cycles that are often observed in economic and financial time series. It provides a new model for multiple local explosions in a strictly stationary framework. The causal predictive distribution displays surprising features, such as higher moments than for the marginal distribution, or the presence of a unit root in the Cauchy case. Aggregating such models can yield complex dynamics with local and global explosion as well as variation in the rate of explosion. The asymptotic behaviour of a vector of sample auto-correlations is studied in a semiparametric non-causal AR(1) framework with Pareto-like tails, and diagnostic tests are proposed. Empirical results based on the Nasdaq composite price index are provided.

AR model

Econometric Theory, 2019, Page 1 of 37.
doi:10.1017/S0266466618000452

MIXED CAUSAL-NONCAUSAL AR PROCESSES AND THE MODELLING OF EXPLOSIVE BUBBLES

SÉBASTIEN FRIES
CREST and Paris-Saclay University

JEAN-MICHEL ZAKOIAN
CREST and University of Lille

ARMA model

Definition 6.2 (Causal TS)

A TS/process $\{y_t : t = 0, \pm 1, \pm 2, ...\}$ is causal if for all t

$$y_t = f(\varepsilon_{t-j} : j \geq 0)$$

where f is a measurable function.

■ Causality means that y_t is caused by the white noise process (from the past) up to time t. For ARMA (p, q) process defined in (1), causality is equivalent to the condition that $\phi(z) \neq 0$ for all $|z| \leq 1$, and therefore it implies stationarity, but the converse is not true.

■ In fact, the model (1) admits a unique stationary solution if and only if $\phi(z) \neq 0$ for all complex numbers z on the unit cycle $|z| = 1$.

ARMA model

■ under the condition $\phi(z) \neq 0$ for all $|z| > 1$, the stationary solution of (1) with $q = 0$, for example, is of the form

$$y_t = \sum_{k=0}^{\infty} d_k \varepsilon_{t+k},$$

which is not causal. One may argue whether such a process should be called a time series since y_t depends on 'future' noise ε_{t+k} for $k \geq 1$.

■ However, any stationary noncausal ARMA process can be represented as a causal ARMA process (with the same orders) in terms of a newly defined white noise, and both processes have identical first two moments (Proposition 3.5.1 of Brockwell and Davis 1991).

■ Reference: p.82-83 & 105 in 《Time Series: Theory and Methods》(2nd) by P.J. Broackwell and R. A. Davis.

ARMA model

Theorem 6.1

Let $\{y_t\}$ be the ARMA(p,q) *process satisfying the equations*

$$y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

where $\phi(z) \neq 0$ and $\theta(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| = 1$. Then there exist polynomials, $\tilde{\phi}(z)$ and $\tilde{\theta}(z)$, nonzero for $|z| \leq$ 1, of degree p and q respectively, and a white noise sequence $\{\varepsilon_t^*\}$ such that $\{y_t\}$ satisfies the causal invertible equations

$$\tilde{\phi}(B) y_t = \tilde{\theta}(B) \varepsilon_t^*.$$

■ Therefore, we lose no generality by restricting our attention to the subset of causal processes in the class of stationary ARMA processes. But we should be aware of the fact that even if the original process is defined in terms of an i.i.d. process $\{\varepsilon_t\}$, the white noise in the new representation is no longer i.i.d.

ARMA model

- Claim: In what following, we only discuss causal TS.

Thank you!