Theory of Stochastic Processes

Lecture 9: Stationary processes

Tomonari SEI

June 15, 2017

1 Stationary processes

Consider a discrete-time process $X = \{X_n\}_{n=-\infty}^{\infty}$. The index set is sometimes restricted to non-negative integers: $\{X_n\}_{n=0}^{\infty}$.

Definition 1 (strong stationarity)**.** A process X is said to be *strongly stationary* if the joint distribution of $X_n, \ldots, X_{n+m}$ is the same as that of $X_0, \ldots, X_m$ for all n and m.

Example 1 (Markov chain). Let X be a Markov chain with the transition matrix (p_{ij}) and a stationary distribution (π_i). If $P(X_0 = i) = \pi_i$, then X is strongly stationary.

Today we focus on the following weaker version of stationarity. Here the state space is assumed to be $\mathbb{R}$.

Definition 2 (weak stationarity)**.** A process X is said to be *weakly stationary* if $E[X_n] = E[X_0]$ and $\mathrm{Cov}[X_n, X_{n+m}] = \mathrm{Cov}[X_0, X_m]$ for all n and m. For a weakly stationary process X, the *autocovariance function* is defined by

$$c(n) = \mathrm{Cov}[X_0, X_n], \quad n \in \mathbb{Z}.$$

The *autocorrelation function* (ACF) is defined by

$$\rho(n) = \frac{\mathrm{Cov}[X_0, X_n]}{\sqrt{V[X_0]V[X_n]}} = \frac{c(n)}{c(0)}.$$

whenever $c(0) > 0$.

It is easy to see that $\rho(0) = 1$ and $\rho(-n) = \rho(n)$ for all n. Note that the autocorrelation function is the autocovariance function of $X_n/\sqrt{c(0)}$.

Example 2 (white noise). Let $\{Z_n\}$ be a sequence of uncorrelated real-valued random variables with zero means and unit variances. Any process with this property is called *a white noise.*

The autocovariance function of a white noise is $c(n) = \delta_0(n)$, Kronecker's delta. See Figure 1.

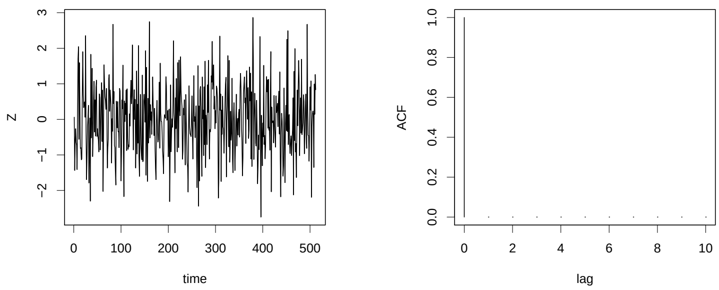


Figure1 A sample path (left) and the autocorrelation function (right) of the white noise.

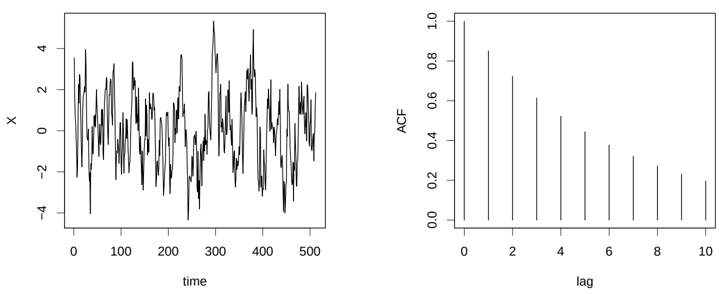
Example 3 (autoregressive process). Let $\{Z_n\}$ be a white noise. Define a process $\{X_n\}$ by

$$X_n = \alpha X_{n-1} + \sigma Z_n, \quad n \in \mathbb{Z},$$

where $\alpha \in \mathbb{R}$ and $\sigma > 0$. This process is called an *autoregressive process* of order 1 (AR(1)). Suppose $|\alpha| < 1$. If X_n is weakly stationary (and therefore $E[X_n^2]$ is bounded), we have

$$X_n = \sigma(Z_n + \alpha Z_{n-1} + \alpha^2 Z_{n-2} + \cdots)$$

by induction, where the limit on the right hand side is interpreted in L^2 sense[*1]. Then the autocovariance function is $c(n) = \sigma^2 \alpha^{|n|}/(1-\alpha^2)$. See Figure 2. Note that X_m and Z_n are uncorrelated if $m < n$. This property is called *causality*. See Section 3 and Problem 1.



sample path ACF

Figure2 A sample path (left) and the autocorrelation function (right)of AR(1) with $\alpha = 0.85$.

[*1] We say that a sequence of random variables $\{Y_n\}$ converges to a random variable Y in L^2 if $E[|Y_n - Y|^2] \to 0$ as $n \to \infty$.

2 Spectral distribution

The following theorem is powerful. See Section 4 for a sketch of proof.

Theorem 1. For any autocovariance function c, there exists a unique finite measure F on $(-\pi, \pi]$ such that

$$c(n) = \int_{-\pi}^{\pi} e^{i\lambda n} F(d\lambda), \quad n \in \mathbb{Z}, \quad i = \sqrt{-1}, \tag{1}$$

and F is symmetric in the sense that $F(A) = F(-A)$ for any subset of $(0, \pi)$. If c is an autocorrelation function ($c(0) = 1$), then F is a probability measure.

If you are not familiar with the notation $\int e^{i\lambda n} F(d\lambda)$, just replace it with $\int e^{i\lambda n} f(\lambda) d\lambda$ or $\sum_\lambda e^{i\lambda n} f(\lambda)$, in accordance with continuity or discreteness of F.

Definition 3 (spectral distribution)**.** The distribution F satisfying (1) is called the *spectral distribution*. If F has the density function f, then f is called the *spectral density function*.

Example 4. The spectral density of a white noise is $f(\lambda) = 1/2\pi$. Indeed,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\lambda n} d\lambda = \delta_0(n).$$

Example 5. Let

$$X_n = A \cos(\lambda_0 n) + B \sin(\lambda_0 n), \quad n \in \mathbb{Z}, \tag{2}$$

where A and B are uncorrelated random variables with zero mean and unit variance, and $\lambda_0 \in (0, \pi)$. The autocovariance function of X_n is $c(h) = \cos(\lambda_0 h)$, $h \in \mathbb{Z}$, and therefore the spectral distribution is a discrete measure $(\delta_{\lambda_0}(d\lambda) + \delta_{-\lambda_0}(d\lambda))/2$.

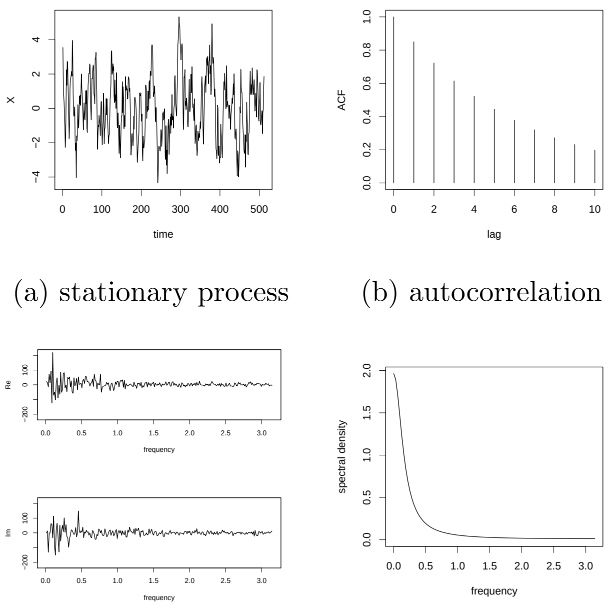
Intuitively speaking, any weakly stationary process is a superposition of (2) for infinitely many λ_0's. This is mathematically justified by *spectral processes*, but not discussed here[*2].

Example 6. As we shall see in the following section, the spectral density function of AR(1) process $X_n = \alpha X_{n-1} + \sigma Z_n$ is given by

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{1 - 2\alpha \cos \lambda + \alpha^2}.$$

Figure 3 shows the spectral density when $\alpha = 0.85$.

[*2] See e.g. Section 9.4 of PRP.



(a) stationary process (b) autocorrelation

(c) spectral process (not discussed here) (d) spectral density

Figure3 Spectral representation of AR(1) with $\alpha = 0.85$. The upper two figures are the same as Figure 2. The lower two figures are the "Fourier transform" of them.

3 Causal processes

We give a broad class of weakly stationary processes. Let $\{Z_n\}$ be a white noise. Define a lag operator L by $La_n = a_{n-1}$ for any sequence $\{a_n\}$. A *causal process*[*3] $\{X_n\}$ is defined by

$$X_n = g(L)Z_n. \tag{3}$$

where $g(L) = \sum_{m=0}^{\infty} g_m L^m$, $g_m \in \mathbb{R}$. More precisely, X_n is the output of a *causal system* $g(L)$ when the input is a white noise. We put a technical assumption[*4] that the convergence radius of a power series $g(z) = \sum_m g_m z^m$ is greater than 1.

Theorem 2. The spectral density function of (3) is given by

$$f(\lambda) = \frac{1}{2\pi}|g(e^{i\lambda})|^2. \tag{4}$$

[*3] For further details, See e.g. Brockwell and Davis (1991), *Time Series: Theory and Methods*, Springer.
[*4] The assumption implies that the right hand side of (3) is well-defined.

Proof. The autocovariance function of X_n is

$$\begin{aligned} c(h) = E[X_0 X_h] &= \sum_m \sum_n g_m g_n E[Z_{-m} Z_{h-n}] \\ &= \frac{1}{2\pi} \sum_m \sum_n g_m g_n \int_{-\pi}^{\pi} e^{i(h-n+m)\lambda} d\lambda \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_m g_m e^{im\lambda} \right|^2 e^{ih\lambda} d\lambda. \end{aligned}$$

Thus the spectral density is (4). □

Example 7. The stationary AR(1) process $X_n = \alpha X_{n-1} + \sigma Z_n$ is rewritten as

$$X_n = \sigma(1 - \alpha L)^{-1} Z_n$$

if $|\alpha| < 1$. Therefore its spectral density is

$$f(\lambda) = \frac{1}{2\pi} \frac{\sigma^2}{|1 - \alpha e^{i\lambda}|^2} = \frac{1}{2\pi} \frac{\sigma^2}{1 - 2\alpha \cos\lambda + \alpha^2}.$$

4 Bochner's theorem

This section may be skipped. We denote the complex conjugate of $z \in \mathbb{C}$ by $\overline{z}$.

Definition 4 (non-negative definiteness)**.** A complex-valued function $\rho(n)$, $n \in \mathbb{Z}$, is called non-negative definite if $\sum_{i=1}^k \sum_{j=1}^k \rho(n_i - n_j) w_i \overline{w_j} \geq 0$ for any $k \geq 1$, $n_1, \ldots, n_k \in \mathbb{Z}$ and $w_1, \ldots, w_k \in \mathbb{C}$.

Lemma 1. Any autocovariance function is non-negative definite.

Proof. $\sum_i \sum_j c(n_i - n_j) w_i \overline{w_j} = \sum_i \sum_j E[X_{n_i} X_{n_j}] w_i \overline{w_j} = E[|\sum_i X_{n_i} w_i|^2] \geq 0$. □

Lemma 2. If ρ is non-negative definite, then $\rho(0) \geq 0$, $|\rho(n)| \leq \rho(0)$, and $\rho(-n) = \overline{\rho(n)}$.

Proof. Let $k = 1$, $n_1 = 0$ and $w_1 = 1$ in the definition of non-negative definiteness. Then we have $\rho(0) \geq 0$. Let $k = 2$, $n_1 = 0$, $n_2 = n$, $w_1 = 1$ and $w_2 = \alpha \in \mathbb{C}$. Then we have $(1 + |\alpha|^2)\rho(0) + \alpha\rho(n) + \overline{\alpha}\rho(-n) \geq 0$. Take $\alpha = \sqrt{-1}$ to obtain $\mathrm{Re}(\rho(n)) = \mathrm{Re}(\rho(-n))$ and $\alpha = 1$ to obtain $\mathrm{Im}(\rho(n)) = -\mathrm{Im}(\rho(-n))$. Let $\alpha = -\overline{\rho(n)}/|\rho(n)|$ to obtain $|\rho(n)| \leq \rho(0)$. □

Lemma 3. Suppose that ρ is non-negative definite. Let N be a positive integer and $\rho_N(n) = (1 - |n|/N)_+ \rho(n)$, where $a_+ = \max(a, 0)$. Let $f_N(\lambda) = (2\pi)^{-1} \sum_n \rho_N(n) e^{-i\lambda n}$. Then $\rho_N(n) = \int_{-\pi}^{\pi} e^{i\lambda n} f_N(\lambda) d\lambda$ and $f_N(\lambda) \geq 0$.

Proof. It is easy to see that $\int_{-\pi}^{\pi} e^{i\lambda n} f_N(\lambda) d\lambda = (2\pi)^{-1} \sum_m \rho_N(m) \int_{-\pi}^{\pi} e^{i\lambda(n-m)} d\lambda = \rho_N(n)$. We also have $f_N(\lambda) = (2\pi N)^{-1} \sum_{j,k=1}^N \rho(j-k) e^{-i(j-k)\lambda} \geq 0$ by non-negative definiteness. □

Theorem 1 is a corollary of the following theorem.

Theorem 3 (Bochner[*5]). A sequence $\{\rho(n)\}$ is non-negative definite if and only if there exists a finite measure F on $(-\pi, \pi]$ such that

$$\rho(n) = \int_{-\pi}^{\pi} e^{i\lambda n} F(d\lambda). \tag{5}$$

In that case, the distribution F is unique.

Proof. It is easy to show that the equation (5) implies non-negative definiteness:

$$\sum_i \sum_j \rho(n_i - n_j) w_i \overline{w_j} = \sum_i \sum_j \int_{-\pi}^{\pi} e^{i\lambda(n_i - n_j)} F(d\lambda) w_i \overline{w_j} = \int_{-\pi}^{\pi} \left| \sum_i e^{i\lambda n_i} w_i \right|^2 F(d\lambda) \geq 0.$$

The converse is more technical. We only give a sketch here. Let ρ be a non-negative definite function and $\rho(0) = 1$ without loss of generality. For each positive integer N, define ρ_N and f_N as Lemma 3. Define $F_N(d\lambda) = f_N(\lambda)d\lambda$. It can be shown that the sequence $\{F_N\}_{N=1}^{\infty}$ is tight[*6] in the space of probability distributions and hence there exists a probability distribution F such that a subsequence F_{N_j} converges to F in distribution. Then we have the relation (5) as follows:

$$\rho(n) = \lim_{j\to\infty} \rho_{N_j}(n) = \lim_{j\to\infty} \int_{-\pi}^{\pi} e^{i\lambda n} F_{N_j}(d\lambda) = \int_{-\pi}^{\pi} e^{i\lambda n} F(d\lambda).$$

Finally, we prove the uniqueness of F. Suppose (5) holds. Define ρ_N and f_N as above. It is sufficient to prove an inversion formula

$$F((a,b]) = \lim_{N\to\infty} \int_a^b f_N(\lambda) d\lambda, \tag{6}$$

whenever $F(\{a\}) = F(\{b\}) = 0$. By definition of f_N and (5), we have

$$f_N(\lambda) = \int_{-\pi}^{\pi} \sum_{n=-(N-1)}^{N-1} \left(1 - \frac{|n|}{N}\right) e^{i(\mu-\lambda)n} F(d\mu) = \int_{-\pi}^{\pi} \underbrace{\frac{1}{N} \left| \sum_{n=0}^{N-1} e^{i(\mu-\lambda)n} \right|^2}_{K_N(\mu-\lambda)} F(d\mu).$$

Integrating both sides from $\lambda = a$ to b, we obtain

$$\int_a^b f_N(\lambda) d\lambda = \int_{-\pi}^{\pi} \left(\int_a^b K_N(\mu - \lambda) d\lambda \right) F(d\mu).$$

The function K_N converges to the "delta function". More precisely, it is shown that

$$\lim_{N\to\infty} \int_a^b K_N(\mu-\lambda) d\lambda = \begin{cases} 1 & \text{if } \mu \in (a,b), \\ 0 & \text{if } \mu \notin [a,b]. \end{cases}$$

It is also shown that $\int_a^b K_N(\mu - \lambda)d\lambda \leq \int_{-\pi}^{\pi} K_N(\mu-\lambda)d\lambda = 1$ for all N. Now the formula (6) follows from Lebesgue's dominated convergence theorem. □

[*5] e.g. W. Feller (1971), *An Introduction to Probability Theory and its Applications, Vol.2, 2nd ed., Wiley.*

[*6] For the definition of tightness and its implication, refer to any book on advanced probability theory, e.g., J. S. Rosenthal (2006), *A first look at rigorous probability theory*, 2nd ed., World Scientific.

5 Exercises

In the following, "stationary" refers to "weakly stationary".

Problem 1 (Non-causal process). Let $\{Z_n\}$ be a white noise. Show that even if $|\alpha| > 1$, there exists a stationary process $\{X_n\}$ such that $X_n = \alpha X_{n-1} + \sigma Z_n$, where X_{n-1} and Z_n are not necessarily uncorrelated[*7]. [Hint: represent X_{n-1} in terms of X_n and Z_n.]

Problem 2. Let $\{Z_n\}$ be a white noise. Define a process $X = \{X_n\}$ by

$$X_n = \sum_{j=1}^{p} \alpha_j X_{n-j} + \sigma Z_n,$$

where $\alpha_j \in \mathbb{R}$ and $\sigma > 0$. Suppose that all the roots of the equation $1 - \sum_{j=1}^{p} \alpha_j z^j = 0$ with respect to $z \in \mathbb{C}$ are outside the unit circle. This process is called an AR(p) process. Show that the spectral density function of X is

$$f(\lambda) = \frac{\sigma^2}{|1 - \sum_{j=1}^{p} \alpha_j e^{-i\lambda j}|^2}.$$

Problem 3. Let $c(n)$ and $d(n)$ be autocovariance functions.

(i) Show that $c(n) + d(n)$ is also an autocovariance function.
(ii) Show that $c(n)d(n)$ is also an autocovariance function.

[Hint: Consider processes $X_n + Y_n$ and $X_n Y_n$, respectively, where X_n and Y_n are independent.]

Problem 4. Let A, B, Ω be independent random variables. Assume that $P(A = \pm 1) = P(B = \pm 1) = 1/2$, and Ω is uniformly distributed on $(0, \pi)$. Define a process $\{X_n\}$ by

$$X_n = A\cos(\Omega n) + B\sin(\Omega n).$$

(i) Show that $\{X_n\}$ is a white noise.
(ii) Show that X_{-1}, X_0, X_1 determine the whole process $\{X_n\}_{n=-\infty}^{\infty}$.

Problem 5. Let X_n be a *circular* stationary process in the sense that there exists $N \geq 1$ such that the autocovariance function c satisfies $c(n) = c(n+N)$ for all n. Show that a matrix $C = \{c(j-k)\}_{j,k=1}^{N}$ is non-negative definite. Use the spectral decomposition of C to obtain the identity

$$c(n) = \sum_{m=0}^{N-1} e^{2\pi imn/N} f(m),$$

where $f(m) = N^{-1}\sum_{n=0}^{N-1} c(n)e^{-2\pi imn/N}$ is non-negative.

[*7] If we assume a priori that X_{n-1} and Z_n are uncorrelated, then there is no stationary solution X_n.

Problem 6 (Effective sample size). Let $\{X_n\}_{n=-\infty}^{\infty}$ be a stationary process with $E[X_n] = \mu$ and $E[X_m X_n] = \sigma^2 \rho(m-n)$, where ρ is an autocorrelation function. Denote the sample mean and sample variance of $\{X_n\}_{n=1}^{N}$ by

$$\bar{X} = \frac{1}{N} \sum_{n=1}^{N} X_n, \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^{N} (X_n - \bar{X})^2.$$

(i) Show that $E[\bar{X}] = \mu$ and $V[\bar{X}] = (\sigma^2/N) \sum_{n=-(N-1)}^{N-1} (1 - |n|/N)\rho(n)$.

(ii) Show that $E[\hat{\sigma}^2] = \sigma^2 - V[\bar{X}]$.

(iii) Assume $\sum_{n=-\infty}^{\infty} |\rho(n)| < \infty$. Show that

$$\lim_{N\to\infty} NV[\bar{X}] = \sigma^2 f(0), \quad \lim_{N\to\infty} E[\hat{\sigma}^2] = \sigma^2,$$

where $f(0) = \sum_{n=-\infty}^{\infty} \rho(n)$ is the spectral density at frequency zero.

Remark: the quantity $N_{\mathrm{eff}} = N/f(0)$ is called the *effective sample size*. If N_{eff} is given, the variance $V[\bar{X}]$ is estimated by $\hat{\sigma}^2/N_{\mathrm{eff}}$. This strategy is used in error estimate of MCMC[*8]

[*8] e.g. the ‘coda’ package in R language. `https://cran.r-project.org/web/packages/coda/coda.pdf`