Time Series
Ch2:Probability Models

Author: Ngai Hang Chan

Reporter: 蕭伯任

2017-3-31

Contents

2.2 STOCHASTIC PROCESSES

- **Definition 2.1** A collection of random variables $\{\mathbf{X}(t) : t \in \mathcal{R}\}$ is called a **stochastic process**.
- in general, $\{\mathbf{X}(t) : 0 \leqq t < \infty\}$ and $\{\mathbf{X_t} : t = 1, 2, ..., n\}$ are used to define a continuous-time and a discrete-time stochastic process, respectively.
- for a given $\omega \in \Omega$, $\mathrm{X}_t(\omega)$ can be considered as a function of t and as such, this function is called a sample function, a realization, or a sample path of the stochastic process.
- For a different ω, it will correspond to a different sample path. The collection of all sample paths is called an **ensemble**. All the time series plots we have seen are based on a single sample path. Accordingly, time series analysis is concerned with finding the probability model that generates the time series observed.
- To describe the underlying probability model, we can consider the joint distribution of the process; that is, for any given set of times $(t_1, ..., t_n)$, consider the joint distribution of $(X_{t_1}, ..., X_{t_n})$, called the **finite-dimensional distribution**.

2.2 STOCHASTIC PROCESSES

- **Definition 2.2** Let T be the set of all vectors $\{t = (t_1, ..., t_n)' \in T^n : t_1 < ... < t_n, n = 1, 2, ...\}$. Then the **(finite-dimensional) distribution functions** of the stochastic process $\{X_t, t \in T\}$ are the functions $\{F_t(\cdot), t \in T\}$ defined for $t = (t_1, ..., t_n)'$ by

$$F_t(x) = P(X_{t_1} \leqq x_1, ..., X_{t_n} \leqq x_n), \quad x = (x_1, ..., x_n)' \in R^n$$

2.2 STOCHASTIC PROCESSES

- **Theorem 2.1 (Kolmogorov's Consistency Theorem)** The probability distribution functions $\{F_t(\cdot), t \in T\}$ are the distribution functions of some stochastic process if and only if for any $n \in \{1, 2, ...\}$, $t = (ti, ..., tn)' \in \mathcal{T}$ and $1 \leqq i \leqq n$,

$$\lim_{x_i \to \infty} F_t(x) = F_{t(i)}(x(i)),$$

  where $t(i)$ and $x(i)$ are the $(n-1)$-component vectors obtained by deleting the ith components of t and x, respectively.
- This theorem ensures the existence of a stochastic process through specification of the collection of all finite-dimensional distributions . Condition ensures a consistency which requires that each finite-dimensional distribution should have marginal distributions that coincide with the lower finite-dimensional distribution functions specified.

2.2 STOCHASTIC PROCESSES

- **Definition 2.3** $\{X_t\}$ is said to be **strictly stationary** if for all n, for all $(t_1, ..., t_n)$, and for all τ,

$$(X_{t_1}, ..., X_{t_n}) \stackrel{\mathrm{d}}{=} (X_{t_1+\tau}, ..., X_{t_n+\tau}),$$

  where $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution.
- Intuitively, stationarity means that the process attains a certain type of statistical equilibrium and the distribution of the process does not change much. It is a very restrictive condition and is often difficult to verify.

2.2 STOCHASTIC PROCESSES

- **Definition 2.4** Let $\{X_t : t \in T\}$ be a stochastic process such that $var(X_t) < \infty, \quad$ for all $t \in T$. Then the **auto covariance function** $\gamma_X(\cdot, \cdot)$ of $\{X_t\}$ is defined by

$$\gamma_X(r, s) = cov(X_r, X_s) = E[(X_r - EX_r)(X_s - EX_s)], \quad r, s \in T$$

2.2 STOCHASTIC PROCESSES

- **Definition 2.5** $\{X_t\}$ is said to be **weakly stationary (second-orderstationary, wide-sense stationary)** if

$(i) E(X_t)=\mu$ for all t.

$(ii) cov(X_t, X_{t+\tau}) = \gamma(\tau)$ for all t and for all τ.

2.2 STOCHASTIC PROCESSES

- A couple of consequences can be deduced immediately from these definitions.
  1. Take $\tau = 0$, $cov(X_t, X_t) = \tau(0)$, for all t. The means and variances of a stationary process always remain constant.
  2. Strict stationarity implies weak stationarity . The converse is not true in general except in the case of a normal distribution .

2.2 STOCHASTIC PROCESSES

- **Definition 2.6** Let $\{X_t\}$ be a stationary process. Then

  $(i)\gamma(\tau) = cov(X_t, X_{t+\tau})$ is called the **autocovariance function**.

  $(ii)\rho(\tau) = \gamma(\tau)/\gamma(0)$ is called the **autocorrelation function**.

- For stationary processes, we expect that both $\tau(.)$ and $\rho(.)$ taper off to zero fairly rapidly. This is an indication of what is known as the short-memory behavior of the series.

2.3 EXAMPLES

- X_t are $i.i.d.$ random variables. Then

$$\rho(\tau) = \begin{cases} 1, & \tau = 0, \\ 0, & \text{otherwise.} \end{cases}$$

- Whenever a time series has this correlation structure , it is known as a **white noise** sequence and the whiteness will become apparent when we study the spectrum of this process.

2.3 EXAMPLES

- Let Y be a random variable such that var Y $= \sigma^2$. LetY$_1 =$ Y_2=...=Y_t=...=Y. Then

  $$\rho(\tau) = 1, \forall \tau.$$

  Hence the process is stationary. However, this process differs substantially from $\{X_t\}$ in example 1. For $\{X_t\}$, knowing its value at one time t has nothing to do with the other values. For $\{Y_t\}$, knowing Y_1 gives the values of all the other Y_t's. Furthermore,

  $$(X_1 + ... + X_n)/n \to EX_1 = \mu$$

  by the law of large numbers.
  But $(Y_i + ... + Y_n)/n = Y$. There is as much randomness in the nth sample average as there is in the first observation for the process $\{Y_t\}$. To prevent situations like this , we introduce the following definition.

2.3 EXAMPLES

- **Definition 2.7** If the sample average formed from a sample path of a process converges to the underlying parameter of the process, the process is called **ergodic**.
- For ergodic processes, we do not need to observe separate independent replications of the entire process in order to estimate its mean value or other moments. One sufficiently long sample path would enable us to estimate the underlying moments. In this book, all the time series studied are assumed to be ergodic.

2.3 EXAMPLES

- Let $X_t = A\cos\theta t + B\sin\theta t, \quad A, B \sim (0, \sigma^2)$ i.i.d. Since EX_t=0, it follow that

$$cov(X_{t+h}, X_t) = E(X_{t+h}X_t)$$

$$= E(A\cos\theta(t+h) + B\sin\theta(t+h))(A\cos\theta t + B\sin\theta t)$$

$$= \sigma^2\cos\theta h.$$

Hence the process is stationary.

2.4 SAMPLE CORRELATION FUNCTION

- In practice, $\gamma(\tau)$ and $\rho(\tau)$ are unknown and they have to be estimated from the data . This leads to the following definition.
- **Definition 2.8** Let $\{X_t\}$ be a given time series and be its sample mean. Then

$(i)\mathrm{C}_k = \sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})$ is called the **sample autocovariance function**.

$(ii) r_k = C_k/C_0$ is called the **sample autocorrelation function(ACF)**.

2.4 SAMPLE CORRELATION FUNCTION

- By definition, $r_0 = 1$. Intuitively, C_k approximates $\gamma(k)$ and r_k approximates $\rho(k)$.
- Although we can "identify" a time series through inspection of its ACF, it is not a procedure that is always free of error. When we calculate the ACF of any given series with a fixed sample size n , we cannot put too much confidence in the values of r_k for large k's , since fewer pairs of (X_t, X_{t-k}) will be available for computing r_k when k is large.
- One rule of thumb is not to evaluate r_k for $k > n/3$. Some authors even argue that only r_k's for $k = O(\log n)$ should be computed. In any case, precautions must be taken.
- Further more , if there is a trend in the data, $X_t = T_t + N_t$, then X_t becomes nonstationary and the idea of inspecting the ACF of X_t becomes questionable. It is therefore important to detrend the data before interpreting their ACF.

Thank You for your Attention