Stationary processes

Last updated by Serik Sagitov: May 13, 2013

Abstract

A course based on the book Probability and Random Processes by Geoffrey Grimmett and David Stirzaker. Chapter 9. Stationary processes.

1 Weakly and strongly stationary processes

Definition 1 *The real-valued process $\{X(t), t \geq 0\}$ is called strongly stationary if the vectors $(X(t_1), \ldots, X(t_n))$ and $(X(t_1 + h), \ldots, X(t_n + h))$ have the same joint distribution for all $t_1, \ldots, t_n$ and $h > 0$.*

Definition 2 *The real-valued process $\{X(t), t \geq 0\}$ with $\mathbb{E}(X^2(t)) < \infty$ for all t is called weakly stationary if for all t_1, t_2 and $h > 0$*

$$\mathbb{E}(X(t_1)) = \mathbb{E}(X(t_2)), \qquad \mathbb{C}ov(X(t_1), X(t_2)) = \mathbb{C}ov(X(t_1 + h), X(t_2 + h)).$$

Its autocovariance and autocorrelation functions are

$$c(t) = \mathbb{C}ov(X(s), X(s + t)), \qquad \rho(t) = \frac{c(t)}{c(0)}.$$

Example 3 Consider an irreducible Markov chain $\{X(t), t \geq 0\}$ with countably many states and a stationary distribution $\boldsymbol{\pi}$ as the initial distribution. This is a strongly stationary process since

$$\mathbb{P}(X(h + t_1) = i_1, X(h + t_1 + t_2) = i_2, \ldots, X(h + t_1 + \ldots + t_n) = i_n) = \pi_{i_1} p_{i_1,i_2}(t_2) \ldots p_{i_{n-1},i_n}(t_n).$$

Example 4 The process $\{X_n, n = 1, 2, \ldots\}$ formed by iid Cauchy r.v is strongly stationary but not a weakly stationary process.

2 Linear prediction

Task: knowing the past $(X_r, X_{r-1}, \ldots, X_{r-s})$ predict a future value X_{r+k} by choosing $(a_0, \ldots, a_s)$ that minimize $\mathbb{E}(X_{r+k} - \hat{X}_{r+k})^2$, where

$$\hat{X}_{r+k} = \sum_{j=0}^{s} a_j X_{r-j}. \tag{1}$$

Theorem 5 *For a stationary sequence with zero mean and autocovariance $c(m)$ the best linear predictor (1) satisfies the equations*

$$\sum_{j=0}^{s} a_j c(|j - m|) = c(k + m), \quad 0 \leq m \leq s.$$

Proof. Geometrically, the best linear predictor $\hat{X}_{r+k}$ makes an error, $X_{r+k} - \hat{X}_{r+k}$, which is orthogonal to the past $(X_r, X_{r-1}, \ldots, X_{r-s})$:

$$\mathbb{E}((X_{r+k} - \hat{X}_{r+k})X_{r-m}) = 0, \quad m = 0, \ldots, s.$$

Plugging (1) into the last relation, we arrive at the claimed equations.

Example 6 AR(1) process Y_n satisfies

$$Y_n = \alpha Y_{n-1} + Z_n, \quad -\infty < n < \infty,$$

where Z_n are independent r.v. with zero means and unit variance. If $|\alpha| < 1$, then $Y_n = \sum_{m\geq 0} \alpha^m Z_{n-m}$ is weakly stationary with zero mean and autocovariance $c(m) = \frac{\alpha^{|m|}}{1-\alpha^2}$. The best linear predictor is $\hat{Y}_{r+k} = \alpha^k Y_r$. The mean squared error of prediction is $\mathbb{E}(\hat{Y}_{r+k} - Y_{r+k})^2 = \frac{1-\alpha^{2k}}{1-\alpha^2}$.

Example 7 Let $X_n = (-1)^n X_0$, where X_0 is -1 or 1 equally likely. The best linear predictor is $\hat{X}_{r+k} = (-1)^k X_r$. The mean squared error of prediction is zero.

3 Linear combination of sinusoids

Example 8 For a sequence of fixed frequencies $0 \leq \lambda_1 < \ldots < \lambda_k < \infty$ define a continuous time stochastic process by

$$X(t) = \sum_{j=1}^{k} (A_j \cos(\lambda_j t) + B_j \sin(\lambda_j t)),$$

where $A_1, B_1, \ldots, A_k, B_k$ are uncorrelated r.v. with zero means and $\mathbb{V}ar(A_j) = \mathbb{V}ar(B_j) = \sigma_j^2$. Its mean is zero and its autocovariancies are

$$\begin{aligned}\mathbb{C}ov(X(t), X(s)) = \mathbb{E}(X(t)X(s)) &= \sum_{j=1}^{k} \mathbb{E}(A_j^2 \cos(\lambda_j t)\cos(\lambda_j s) + B_j^2 \sin(\lambda_j t)\sin(\lambda_j s)) \\ &= \sum_{j=1}^{k} \sigma_j^2 \cos(\lambda_j (s-t)), \\ \mathbb{V}ar(X(t)) &= \sum_{j=1}^{k} \sigma_j^2.\end{aligned}$$

Thus $X(t)$ is weakly stationary with autocovariance and autocorrelation functions

$$c(t) = \sum_{j=1}^{k} \sigma_j^2 \cos(\lambda_j t), \quad c(0) = \sum_{j=1}^{k} \sigma_j^2,$$

$$\rho(t) = \frac{c(t)}{c(0)} = \sum_{j=1}^{k} g_j \cos(\lambda_j t) = \int_0^\infty \cos(\lambda t) dG(\lambda),$$

where G is a distribution function defined as

$$g_j = \frac{\sigma_j^2}{\sigma_1^2 + \ldots + \sigma_k^2}, \qquad G(\lambda) = \sum_{j:\lambda_j \leq \lambda} g_j.$$

We can write

$$X(t) = \int_0^\infty \cos(t\lambda) dU(\lambda) + \int_0^\infty \sin(t\lambda) dV(\lambda),$$

where

$$U(\lambda) = \sum_{j:\lambda_j \leq \lambda} A_j, \qquad V(\lambda) = \sum_{j:\lambda_j \leq \lambda} B_j.$$

Example 9 Let specialize further and put $k = 1$, $\lambda_1 = \frac{\pi}{4}$, assuming that A_1 and B_1 are iid with

$$\mathbb{P}(A_1 = \frac{1}{\sqrt{2}}) = \mathbb{P}(A_1 = -\frac{1}{\sqrt{2}}) = \frac{1}{2}.$$

Then $X(t) = \cos(\frac{\pi}{4}(t+\tau))$ with

$$\mathbb{P}(\tau = 1) = \mathbb{P}(\tau = -1) = \mathbb{P}(\tau = 3) = \mathbb{P}(\tau = -3) = \frac{1}{4}.$$

This stochastic process has only four possible trajectories. This is not a strongly stationary process since

$$\mathbb{E}(X^4(t)) = \frac{1}{2}\Big(\cos^4\Big(\frac{\pi}{4}t + \frac{\pi}{4}\Big) + \sin^4\Big(\frac{\pi}{4}t + \frac{\pi}{4}\Big)\Big) = \frac{1}{4}\Big(2 - \sin^2\Big(\frac{\pi}{2}t + \frac{\pi}{2}\Big)\Big) = \frac{1+\sin^2(\frac{\pi}{2}t)}{2}.$$

Example 10 Put

$$X(t) = \cos(t+Y) = \cos(t)\cos(Y) - \sin(t)\sin(Y),$$

where Y is uniformly distributed over $[0, 2\pi]$. In this case $k = 1, \lambda = 1, \sigma_1^2 = (4\pi)^{-1}$. What is the distribution of $X(t)$? For an arbitrary bounded measurable function $\phi(x)$ we have

$$\begin{aligned}\mathbb{E}(\phi(X(t))) = \mathbb{E}(\phi(\cos(t+Y))) = \frac{1}{2\pi}\int_0^{2\pi}\phi(\cos(t+y))dy = \frac{1}{2\pi}\int_t^{t+2\pi}\phi(\cos(z))dz = \frac{1}{2\pi}\int_0^{2\pi}\phi(\cos(z))dz \\ = \frac{1}{2\pi}\Big(\int_0^{\pi}\phi(\cos(z))dz + \int_{\pi}^{2\pi}\phi(\cos(z))dz\Big) = \frac{1}{2\pi}\Big(\int_0^{\pi}\phi(\cos(\pi - y))dy + \int_0^{\pi}\phi(\cos(\pi+y))dy\Big) \\ = \frac{1}{\pi}\int_0^{\pi}\phi(-\cos(y))dy.\end{aligned}$$

The change of variables $x = -\cos(y)$ yields $dx = \sin(y)dy = \sqrt{1-x^2}dx$, hence

$$\mathbb{E}(\phi(X)) = \frac{1}{\pi}\int_{-1}^{1}\frac{\phi(x)dx}{\sqrt{1-x^2}}.$$

Thus $X(t)$ has the so-called arcsine density $f(x) = \frac{1}{\pi\sqrt{1-y^2}}$ over the interval $[-1, 1]$. Notice that $Z = \frac{X+1}{2}$ has a Beta$(\frac{1}{2}, \frac{1}{2})$ distribution, since

$$\mathbb{E}(\phi(Z)) = \frac{1}{\pi}\int_{-1}^{1}\frac{\phi(\frac{x+1}{2})dx}{\sqrt{1-x^2}} = \frac{1}{\pi}\int_0^1\frac{\phi(z)dz}{\sqrt{z(1-z)}}.$$

This is a strongly stationary process, since $X(t+h) = \cos(t+Y')$, where Y' is uniformly distributed over $[h, 2\pi + h]$, and

$$(X(t_1+h), \ldots, X(t_n+h)) = (\cos(t_1+Y'), \ldots, \cos(t_n+Y')) \stackrel{d}{=} (\cos(t_1+Y), \ldots, \cos(t_n+Y)).$$

Example 11 In the *discrete time setting* for $n \in \mathbb{Z}$ put

$$X_n = \sum_{j=1}^{k}(A_j\cos(\lambda_j n) + B_j\sin(\lambda_j n)),$$

where $0 \leq \lambda_1 < \ldots < \lambda_k \leq \pi$ is a set of fixed frequencies, and again, $A_1, B_1, \ldots, A_k, B_k$ are uncorrelated r.v. with zero means and $\mathbb{V}ar(A_j) = \mathbb{V}ar(B_j) = \sigma_j^2$. Similarly to the continuous time case we get

$$\mathbb{E}(X_n) = 0, \quad c(n) = \sum_{j=1}^{k}\sigma_j^2\cos(\lambda_j n), \quad \rho(n) = \int_0^{\pi}\cos(\lambda n)dG(\lambda),$$

$$X_n = \int_0^{\pi}\cos(n\lambda)dU(\lambda) + \int_0^{\pi}\sin(n\lambda)dV(\lambda).$$

4 The spectral representation

Any weakly stationary process $\{X(t) : -\infty < t < \infty\}$ with zero mean can be approximated by a linear combination of sinusoids. Indeed, its autocovariance function $c(t)$ is non-negative definite since for any $t_1, \ldots, t_n$ and $z_1, \ldots, z_n$

$$\sum_{j=1}^{n}\sum_{k=1}^{n} c(t_k - t_j) z_j z_k = \mathbb{V}ar\Big(\sum_{k=1}^{n} z_k X(t_k)\Big) \geq 0.$$

Thus due to the Bochner theorem, given that $c(t)$ is continuous at zero, there is a probability distribution function G on $[0, \infty)$ such that

$$\rho(t) = \int_0^\infty \cos(t\lambda) dG(\lambda).$$

In the discrete time case there is a probability distribution function G on $[0, \pi]$ such that

$$\rho(n) = \int_0^\pi \cos(n\lambda) dG(\lambda).$$

Definition 12 *The function G is called the spectral distribution function of the corresponding stationary random process, and the set of real numbers λ such that*

$$G(\lambda + \epsilon) - G(\lambda - \epsilon) > 0 \text{ for all } \epsilon > 0$$

is called the spectrum of the random process. If G has density it is called the spectral density function.

Example 13 Consider an irreducible continuous time Markov chain $\{X(t), t \geq 0\}$ with two states $\{1, 2\}$ and generator

$$\mathbf{G} = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}.$$

Its stationary distribution is $\boldsymbol{\pi} = (\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta})$ will be taken as the initial distribution. From

$$\begin{pmatrix} p_{11}(t) & p_{12}(t) \\ p_{21}(t) & p_{22}(t) \end{pmatrix} = e^{t\mathbf{G}} = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbf{G}^n = \mathbf{I} + \mathbf{G} \sum_{n=1}^{\infty} \frac{t^n}{n!} (-\alpha - \beta)^{n-1} = \mathbf{I} + (\alpha + \beta)^{-1}(1 - e^{-t(\alpha+\beta)})\mathbf{G}$$

we see that

$$p_{11}(t) = 1 - p_{12}(t) = \frac{\beta}{\alpha + \beta} + \frac{\alpha}{\alpha + \beta} e^{-t(\alpha+\beta)},$$
$$p_{22}(t) = 1 - p_{21}(t) = \frac{\alpha}{\alpha + \beta} + \frac{\beta}{\alpha + \beta} e^{-t(\alpha+\beta)},$$

and we find for $t \geq 0$

$$c(t) = \frac{\alpha\beta}{(\alpha + \beta)^2} e^{-t(\alpha+\beta)}, \qquad \rho(t) = e^{-t(\alpha+\beta)}.$$

Thus this process has a spectral density corresponding to a one-sided Cauchy distribution:

$$g(\lambda) = \frac{2(\alpha + \beta)}{\pi((\alpha + \beta)^2 + \lambda^2)}, \quad \lambda \geq 0.$$

Example 14 Discrete white noise: a sequence $X_0, X_1, \ldots$ of independent r.v. with zero means and unit variances. This stationary sequence has the uniform spectral density:

$$\rho(n) = 1_{\{n=0\}} = \pi^{-1} \int_0^\pi \cos(n\lambda) d\lambda.$$

Theorem 15 *If $\{X(t) : -\infty < t < \infty\}$ is a weakly stationary process with zero mean, unit variance, continuous autocorrelation function and spectral distribution function G, then there exists a pair of orthogonal zero mean random process $(U(\lambda), V(\lambda))$ with uncorrelated increments such that*

$$X(t) = \int_0^\infty \cos(t\lambda) dU(\lambda) + \int_0^\infty \sin(t\lambda) dV(\lambda)$$

and $\mathbb{V}ar(U(\lambda)) = \mathbb{V}ar(V(\lambda)) = G(\lambda)$.

Theorem 16 *If $\{X_n : -\infty < n < \infty\}$ is a discrete-time weakly stationary process with zero mean, unit variance, and spectral distribution function G, then there exists a pair of orthogonal zero mean random process $(U(\lambda), V(\lambda))$ with uncorrelated increments such that*

$$X_n = \int_0^\pi \cos(n\lambda) dU(\lambda) + \int_0^\pi \sin(n\lambda) dV(\lambda)$$

and $\mathbb{V}ar(U(\lambda)) = \mathbb{V}ar(V(\lambda)) = G(\lambda)$.

5 Stochastic integral

Let $\{S(t) : t \in \mathbb{R}\}$ be a complex-valued process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that

- $\mathbb{E}(|S(t)|^2) < \infty$ for all t,
- $\mathbb{E}(|S(t+h) - S(t)|^2) \to 0$ as $h \searrow 0$ for all t,
- $\mathbb{E}([S(v) - S(u)][\bar{S}(t) - \bar{S}(s)]) = 0$ whenever $u < v \leq s < t$.

Put

$$F(t) := \begin{cases} \mathbb{E}(|S(t) - S(0)|^2), & \text{if } t \leq 0, \\ -\mathbb{E}(|S(t) - S(0)|^2), & \text{if } t < 0. \end{cases}$$

Since the process has orthogonal increments we obtain

$$\mathbb{E}(|S(t) - S(s)|^2) = F(t) - F(s), \quad s < t \tag{2}$$

implying that F is monotonic and right-continuous.

Let $\psi : \mathbb{R} \to \mathbb{C}$ be a measurable complex-valued function for which

$$\int_{-\infty}^\infty |\psi(t)|^2 dF(t) < \infty.$$

Next comes a two-step definition of a stochastic integral of ψ with respect to S,

$$I(\psi) = \int_{-\infty}^\infty \psi(t) dS(t),$$

possessing the following important property

$$\mathbb{E}(I(\psi_1)I(\psi_2)) = \int_{-\infty}^\infty \psi_1(t)\psi_2(t) dF(t). \tag{3}$$

1. For an arbitrary step function

$$\phi(t) = \sum_{j=1}^{n-1} c_j 1_{\{a_j \leq t < a_{j+1}\}}, \quad -\infty < a_1 < \ldots < a_n < \infty$$

put

$$I(\phi) := \sum_{j=1}^{n-1} c_j (S(a_{j+1}) - S(a_j)).$$

Due to orthogonality of increments we obtain (3) and find that "integration is distance preserving"

$$\mathbb{E}(|I(\phi_1) - I(\phi_2)|^2) = \mathbb{E}((I(\phi_1 - \phi_2))^2) = \int_{-\infty}^\infty |\phi_1 - \phi_2|^2 dF(t).$$

2. There exists a sequence of step functions such that

$$\|\phi_n - \psi\| := \Big(\int_{-\infty}^{\infty} |\phi_n - \psi|^2 dF(t)\Big)^{1/2} \to 0.$$

Thus $I(\phi_n)$ is a mean-square Cauchy sequence and there exists a mean-square limit $I(\phi_n) \to I(\psi)$.

A sketch of the proof of Theorem 16 for the complex-valued processes.

Step 1. Let H_X be the set of all r.v of the form $\sum_{j=1}^n a_j X_{m_j}$ for $a_1, a_2, \ldots \in \mathbb{C}$, $n \in \mathbb{N}$, $m_1, m_2, \ldots \in \mathbb{Z}$. Similarly, let H_F be the set of linear combinations of sinusoids $f_n(x) := e^{inx}$. Define the linear mapping $\mu : H_F \to H_X$ by $\mu(f_n) := X_n$.

Step 2. The closure $\overline{H}_X$ of H_X is defined to be the space H_X together with all limits of mean-square Cauchy-convergent sequences in H_X. Define the closure $\overline{H}_F$ of H_F as the space H_F together with all limits of Cauchy-convergent sequences $u_n \in H_F$, with the latter meaning by definition that

$$\int_{(-\pi,\pi]} (u_n(\lambda) - u_m(\lambda))\overline{(u_n(\lambda) - u_m(\lambda))} dF(\lambda) \to 0, \qquad n, m \to \infty.$$

For $u = \lim u_n$, where $u_n \in H_F$, define $\mu(u) = \lim \mu(u_n)$ thereby defining a mapping $\mu : \overline{H}_F \to \overline{H}_X$.

Step 3. Define the process $S(\lambda)$ by

$$S(\lambda) = \mu(h_\lambda), \quad -\pi < \lambda \le \pi, \quad h_\lambda(x) := 1_{\{x \in (-\pi,\lambda]\}}$$

and show that it has orthogonal increments and satisfies (2). Prove that

$$\mu(\psi) = \int_{(-\pi,\pi]} \psi(t) dS(t)$$

first for step-functions and then for $\psi(x) = e^{inx}$.

6 The ergodic theorem for the weakly stationary processes

Theorem 17 *Let $\{X_n, n = 1, 2, \ldots\}$ be a weakly stationary process with mean μ and autocovariance function $c(m)$. There exists a r.v. Y with mean μ and variance*

$$\mathrm{Var}(Y) = \lim_{n\to\infty} n^{-1} \sum_{j=1}^n c(j) = c(0)(G(0) - G(0-)),$$

such that

$$\frac{X_1 + \ldots + X_n}{n} \to Y \text{ in square mean.}$$

Proof. Suppose that $\mu = 0$, then using a spectral representation

$$X_n = \int_0^\pi \cos(n\lambda) dU(\lambda) + \int_0^\pi \sin(n\lambda) dV(\lambda).$$

we get

$$\bar{X}_n := \frac{X_1 + \ldots + X_n}{n} = \int_0^\pi g_n(\lambda) dU(\lambda) + \int_0^\pi h_n(\lambda) dV(\lambda), \quad \begin{cases} g_n(\lambda) = n^{-1}(\cos(\lambda) + \ldots + \cos(n\lambda)), \\ h_n(\lambda) = n^{-1}(\sin(\lambda) + \ldots + \sin(n\lambda)). \end{cases}$$

We have that $|g_n(\lambda)| \le 1$, $|h_n(\lambda)| \le 1$, and $g_n(\lambda) \to 1_{\{\lambda=0\}}$, $h_n(\lambda) \to 0$ as $n \to \infty$. It can be shown that

$$\int_0^\pi g_n(\lambda) dU(\lambda) \to \int_0^\pi 1_{\{\lambda=0\}} dU(\lambda) = U(0) - U(0-), \qquad \int_0^\pi h_n(\lambda) dV(\lambda) \to 0$$

in square mean. Thus $\bar{X}_n \to Y := U(0) - U(0-)$ in square mean and it remains to find the mean and variance of Y.

7 The ergodic theorem for the strongly stationary processes

Let $\{X_n, n = 1, 2, \ldots\}$ be a strongly stationary process defined on $(\Omega, \mathcal{F}, \mathbb{P})$. The vector $(X_1, X_2, \ldots)$ takes values in $\mathcal{R}^T$, where $T = \{1, 2, \ldots\}$. We write $\mathcal{B}^T$ for the appropriate number of copies of the Borel σ-algebra $\mathcal{B}$ of subsets of $\mathcal{R}$.

Definition 18 *A set $A \in \mathcal{F}$ is called invariant, if for some $B \in \mathcal{B}^T$ and all n*

$$A = \{\omega : (X_n, X_{n+1}, \ldots) \in B\}.$$

The collection of all invariant sets forms a σ-algebra and denoted $\mathcal{I}$. The strictly stationary process is called ergodic, if for any $A \in \mathcal{I}$ either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

Example 19 Any invariant event is a tail event, so that $\mathcal{I} \subset \mathcal{T}$. If $\{X_n, n = 1, 2, \ldots\}$ are iid with a finite mean, then according to Kolmogorov's zero-one law such a stationary process is ergodic. The classical LLN follows from the next ergodic theorem.

Theorem 20 *If $\{X_n, n = 1, 2, \ldots\}$ is a strongly stationary sequence with a finite mean, then*

$$\bar{X}_n := \frac{X_1 + \ldots + X_n}{n} \to \mathbb{E}(X_1|\mathcal{I}) \text{ a.s. and in mean.}$$

In the ergodic case

$$\bar{X}_n \to \mathbb{E}(X_1) \text{ a.s. and in mean.}$$

Proof in the ergodic case. To prove the a.s. convergence it suffices to show that

$$\text{if } \mathbb{E}(X_1) < 0, \text{ then } \limsup_{n\to\infty} \bar{X}_n \le 0 \text{ a.s.} \tag{4}$$

Indeed, applying (4) to $X'_n = X_n - \mathbb{E}(X_1) - \epsilon$ and $X''_n = \mathbb{E}(X_1) - X_n - \epsilon$, we obtain

$$\mathbb{E}(X_1) - \epsilon \le \liminf_{n\to\infty} \bar{X}_n \le \limsup_{n\to\infty} \bar{X}_n \le \mathbb{E}(X_1) + \epsilon.$$

To prove (4) assume $\mathbb{E}(X_1) < 0$ and put

$$M_n := \max\{0, X_1, X_1 + X_2, \ldots, X_1 + \ldots + X_n\}.$$

Clearly $\bar{X}_n \le M_n/n$, and it is enough to show that $\mathbb{P}(M_\infty < \infty) = 1$. Suppose the latter is not true. Since $\{M_\infty < \infty\}$ is an invariant event, we get $\mathbb{P}(M_\infty < \infty) = 0$ or in other words $M_n \nearrow \infty$ a.s. To arrive to a contradiction observe that

$$M_{n+1} = \max\{0, X_1 + M'_n\} = M'_n + \max\{-M'_n, X_1\}, \quad M'_n := \max\{0, X_2, X_2 + X_3, \ldots, X_2 + \ldots + X_n\}.$$

Since $\mathbb{E}(M_{n+1}) \ge \mathbb{E}(M_n) = \mathbb{E}(M'_n)$ it follows that $\mathbb{E}(\max\{-M'_n, X_1\}) \ge 0$. This contradicts the assumption $\mathbb{E}(X_1) < 0$ as due to the monotone convergence theorem $\mathbb{E}(\max\{-M'_n, X_1\}) \to \mathbb{E}(X_1)$. Thus almost-sure convergence is proved.

To prove the convergence in mean we verify that the family $\{\bar{X}_n\}_{n\ge1}$ is uniformly integrable, that is for all $\epsilon > 0$, there is $\delta > 0$ such that, for all n, $\mathbb{E}(|\bar{X}_n|1_A) < \epsilon$ for any event A such that $\mathbb{P}(A) < \delta$. This follows from

$$\mathbb{E}(|\bar{X}_n|1_A) \le n^{-1} \sum_{i=1}^n \mathbb{E}(|X_i|1_A),$$

and the fact that for all $\epsilon > 0$, there is $\delta > 0$ such that $\mathbb{E}(|X_i|1_A) < \epsilon$ for all i and for any event A such that $\mathbb{P}(A) < \delta$.

Example 21 Let $Z_1, \ldots, Z_k$ be iid with a finite mean μ. Then the following cyclic process

$$X_1 = Z_1, \ldots, X_k = Z_k,$$
$$X_{k+1} = Z_1, \ldots, X_{2k} = Z_k,$$
$$X_{2k+1} = Z_1, \ldots, X_{3k} = Z_k, \ldots,$$

is a strongly stationary process. The corresponding limit in the ergodic theorem is not the constant μ like in the strong LLN but rather a random variable

$$\frac{X_1 + \ldots + X_n}{n} \to \frac{Z_1 + \ldots + Z_k}{k}.$$

Example 22 Let $\{X_n, n = 1, 2, \ldots\}$ be an irreducible positive-recurrent Markov chain with the state space $S = \{0, \pm 1, \pm 2, \ldots\}$. Let $\boldsymbol{\pi} = (\pi_j)_{j \in S}$ be the unique stationary distribution. If X_0 has distribution $\boldsymbol{\pi}$, then X_n is strongly stationary.

For a fixed state $k \in S$ let $I_n = 1_{\{X_n = k\}}$. The stronlgy stationary process I_n has autocovariance function

$$c(m) = \mathbb{C}ov(I_n, I_{n+m}) = \mathbb{E}(I_n I_{n+m}) - \pi_k^2 = \pi_k (p_{kk}^{(m)} - \pi_k).$$

Since $p_{kk}^{(m)} \to \pi_k$ as $m \to \infty$ we have $c(m) \to 0$ and the limit in Theorem 20 has zero variance. It follows that $n^{-1}(I_1 + \ldots + I_n)$, the proportion of $(X_1, \ldots, X_n)$ visiting state k, converges to π_k as $n \to \infty$.

Example 23 Binary expansion. Let X be uniformly distributed on $[0, 1]$ and has a binary expansion $X = \sum_{j=1}^{\infty} X_j 2^{-j}$. Put $Y_n = \sum_{j=n}^{\infty} X_j 2^{n-j-1}$ so that $Y_1 = X$ and $Y_{n+1} = (2^n X) \bmod 1$. From

$$\mathbb{E}(Y_1 Y_{n+1}) = \sum_{j=0}^{2^n-1} \int_{j2^{-n}}^{(j+1)2^{-n}} x(2^n x - j)dx = 2^{-2n} \sum_{j=0}^{2^n-1} \int_0^1 (y + j)ydy = 2^{-2n} \sum_{j=0}^{2^n-1} \Big(\frac{1}{3} + \frac{j}{2}\Big) = \frac{1}{4} + \frac{2^{-n}}{12}$$

we get $c(n) = \frac{2^{-n}}{12}$ implying that $n^{-1} \sum_{j=1}^{n} Y_j \to 1/2$ almost surely.

8 Gaussian processes

Bivariate normal distribution with parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\frac{(\frac{x-\mu_1}{\sigma_1})^2 - 2\rho(\frac{x-\mu_1}{\sigma_1})(\frac{y-\mu_2}{\sigma_2}) + (\frac{y-\mu_2}{\sigma_2})^2}{2(1-\rho^2)} \right\}.$$

Marginal distributions

$$f_1(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, \quad f_2(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}},$$

and conditional distributions

$$f_1(x|y) = \frac{f(x, y)}{f_2(y)} = \frac{1}{\sigma_1\sqrt{2\pi(1-\rho^2)}} \exp\left\{ -\frac{(x - \mu_1 - \frac{\rho\sigma_1}{\sigma_2}(y - \mu_2))^2}{2\sigma_1^2(1-\rho^2)} \right\},$$

$$f_2(y|x) = \frac{f(x, y)}{f_1(x)} = \frac{1}{\sigma_2\sqrt{2\pi(1-\rho^2)}} \exp\left\{ -\frac{(y - \mu_2 - \frac{\rho\sigma_2}{\sigma_1}(x - \mu_1))^2}{2\sigma_2^2(1-\rho^2)} \right\}.$$

The covariance matrix of a random vector $(X_1, \ldots, X_n)$ with means $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$

$$\mathbf{V} = \mathbb{E}\big(\mathbf{X} - \boldsymbol{\mu}\big)^{\mathrm{t}}\big(\mathbf{X} - \boldsymbol{\mu}\big) = \|\mathrm{cov}(X_i, X_j)\|$$

is symmetric and nonnegative-definite. For any vector $\mathbf{a} = (a_1, \ldots, a_n)$ the r.v. $a_1 X_1 + \ldots + a_n X_n$ has mean $\mathbf{a}\boldsymbol{\mu}^{\mathrm{t}}$ and variance

$$\mathbb{V}ar(a_1 X_1 + \ldots + a_n X_n) = \mathbb{E}\big(\mathbf{a}\mathbf{X}^{\mathrm{t}} - \mathbf{a}\boldsymbol{\mu}^{\mathrm{t}}\big)\big(\mathbf{X}\mathbf{a}^{\mathrm{t}} - \boldsymbol{\mu}\mathbf{a}^{\mathrm{t}}\big) = \mathbf{a}\mathbf{V}\mathbf{a}^{\mathrm{t}}.$$

A multivariate normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ and covariance matrix $\mathbf{V}$ has density

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det\mathbf{V}}} e^{-(\mathbf{x}-\boldsymbol{\mu})\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{t}}}$$

and moment generating function $\phi(\boldsymbol{\theta}) = e^{\boldsymbol{\theta}\boldsymbol{\mu}^{\mathrm{t}} + \frac{1}{2}\boldsymbol{\theta}\mathbf{V}\boldsymbol{\theta}^{\mathrm{t}}}$. It follows that given a vector $\mathbf{X} = (X_1, \ldots, X_n)$ with a multivariate normal distribution any linear combination $\mathbf{a}\mathbf{X}^{\mathrm{t}} = a_1 X_1 + \ldots + a_n X_n$ is normally distributed since

$$\mathbb{E}(e^{\theta\mathbf{a}\mathbf{X}^{\mathrm{t}}}) = \phi(\theta\mathbf{a}) = e^{\theta\mu + \frac{1}{2}\theta^2\sigma^2}, \quad \mu = \mathbf{a}\boldsymbol{\mu}^{\mathrm{t}}, \quad \sigma^2 = \mathbf{a}\mathbf{V}\mathbf{a}^{\mathrm{t}}.$$

Definition 24 *A random process* $\{X(t), t \geq 0\}$ *is called Gaussian if for any* $(t_1, \ldots, t_n)$ *the vector* $(X(t_1), \ldots, X(t_n))$ *has a multivariate normal distribution.*

A Gaussian random process is strongly stationary iff it is weakly stationary.

Theorem 25 *A Gaussian process* $\{X(t), t \geq 0\}$ *is Markov iff for any* $0 \leq t_1 < \ldots < t_n$

$$\mathbb{E}(X(t_n)|X(t_1), \ldots, X(t_{n-1})) = \mathbb{E}(X(t_n)|X(t_{n-1})). \tag{5}$$

Proof. Clearly, the Markov property implies (5). To prove the converse we have to show that in the Gaussian case (5) gives

$$\mathbb{V}ar(X(t_n)|X(t_1), \ldots, X(t_{n-1})) = \mathbb{V}ar(X(t_n)|X(t_{n-1})).$$

Indeed, since $X(t_n) - \mathbb{E}\{X(t_n)|X(t_1), \ldots, X(t_{n-1})\}$ is orthogonal to $(X(t_1), \ldots, X(t_{n-1})$, which in the Gaussian case means independence, we have

$$\begin{aligned}
&\mathbb{E}\Big\{\Big(X(t_n) - \mathbb{E}\{X(t_n)|X(t_1), \ldots, X(t_{n-1})\}\Big)^2 |X(t_1), \ldots, X(t_{n-1})\Big\} \\
&\quad = \mathbb{E}\Big\{\Big(X(t_n) - \mathbb{E}\{X(t_n)|X(t_1), \ldots, X(t_{n-1})\}\Big)^2\Big\} = \mathbb{E}\Big\{\Big(X(t_n) - \mathbb{E}\{X(t_n)|X(t_{n-1})\}\Big)^2\Big\} \\
&\quad = \mathbb{E}\Big\{\Big(X(t_n) - \mathbb{E}\{X(t_n)|X(t_{n-1})\}\Big)^2 |X(t_{n-1})\Big\}.
\end{aligned}$$

Example 26 A stationary Gaussian Markov process is called the Ornstein-Uhlenbeck process. It is characterized by the auto-correlation function $\rho(t) = e^{-\alpha t}$, $t \geq 0$ with a positive α. This follows from the equation $\rho(t + s) = \rho(t)\rho(s)$ which is obtained as follows. From the property of the bivariate normal distribution

$$\mathbb{E}(X(t + s)|X(s)) = \theta + \rho(t)(X(s) - \theta)$$

we derive

$$\begin{aligned}
\rho(t + s) = c(0)^{-1}\mathbb{E}((X(t + s) - \theta)(X(0) - \theta)) &= c(0)^{-1}\mathbb{E}\{\mathbb{E}((X(t + s) - \theta)(X(0) - \theta)|X(0), X(s))\} \\
&= \rho(t)c(0)^{-1}\mathbb{E}((X(s) - \theta)(X(0) - \theta)) \\
&= \rho(t)\rho(s).
\end{aligned}$$

The OU–process $X(t)$ with a fixed initial value X_0 is described by the stochastic differential equation

$$dX(t) = -\alpha(X(t) - \theta)dt + \sigma dB(t), \quad X(0) = X_0, \tag{6}$$

which is a continuous version of an AR(1) process $X_n = aX_{n-1} + Z_n$. This process can be interpreted as the evolution of a phenotypic trait value (like logarithm of the body size) along a lineage of species in terms of the adaptation rate $\alpha > 0$, the optimal trait value θ, and the noise size $\sigma > 0$. The distribution of $X(t)$ is normal with

$$\mathbb{E}(X(t)) = \theta + e^{-\alpha t}(X_0 - \theta), \quad \mathbb{V}ar(X(t)) = \sigma^2(1 - e^{-2\alpha t})/2\alpha \tag{7}$$

implying that $X(t)$ looses the effect of the ancestral state X_0 at an exponential rate. In the long run X_0 is forgotten, and the OU–process acquires a stationary normal distribution with mean θ and variance $\sigma^2/2\alpha$.

Proof of (7). Ito's lemma: if $dX(t) = \mu_t dt + \sigma_t dB(t)$, then for any nice function $f(t, x)$

$$df(t, X(t)) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}(\mu_t dt + \sigma_t dB(t)) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\sigma_t^2 dt,$$

where $(dB(t))^2$ is replaced by dt. We apply Ito's lemma to the process $X(t)$ described by (6) using $f(t, x) = xe^{\alpha t}$

$$df(t, X(t)) = \alpha X(t)e^{\alpha t}dt + e^{\alpha t}\Big(\alpha(\theta - X(t))dt + \sigma dB(t)\Big) = \theta e^{\alpha t}\alpha dt + \sigma e^{\alpha t}dB(t).$$

Integration gives

$$X(t)e^{\alpha t} - X_0 = \theta(e^{\alpha t} - 1) + \sigma \int_0^t e^{\alpha u} dB(u),$$

implying

$$X(t) = \theta + e^{-\alpha t}(X_0 - \theta) + \sigma \int_0^t e^{\alpha(u-t)} dB(u).$$

It remains to notice that in view of (3) with $G(t) = t$ we have

$$\mathbb{E}\Big(\int_0^t e^{\alpha u} dB(u)\Big)^2 = \int_0^t e^{2\alpha u} du = \frac{e^{2\alpha t} - 1}{2\alpha}.$$

Observe that the correlation coefficient between $X(s)$ and $X(s+t)$ equals

$$\rho(s, s+t) = e^{-\alpha t}\sqrt{\frac{1 - e^{-2\alpha s}}{1 - e^{-2\alpha(s+t)}}} \to e^{-\alpha t}, \qquad s \to \infty.$$