Convergence in Mean Square
Tidsserieanalys SF2945
Timo Koski

- MEAN SQUARE CONVERGENCE.
- MEAN SQUARE CONVERGENCE OF SUMS.
- CAUSAL LINEAR PROCESSES
1  Definition of convergence in mean square

Definition 1.1 A random sequence \( \{X_n\}_{n=1}^{\infty} \) with \( E[X_n^2] < \infty \) is said to converge in mean square to a random variable \( X \) if

\[
E[|X_n - X|^2] \to 0 \quad (1.1)
\]
as \( n \to \infty \).

In Swedish this is called konvergens i kvadratiskt medel. We write also

\[ X = l.i.m._{n \to \infty} X_n. \]

This definition is silent about convergence of individual sample paths \( X_n(s) \).
By Chebyshev’s inequality we see that convergence in mean square implies convergence in probability.

2  Mean Ergodic Theorem

Although the definition of convergence in mean square encompasses convergence to a random variable, in many applications we shall encounter convergence to a degenerate random variable, i.e., a constant.

Theorem 2.1 The random sequence \( \{X_n\}_{n=1}^{\infty} \sim WN(\mu, \sigma^2) \) Then

\[ \mu = l.i.m._{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} X_n. \]

Proof: Let us set \( S_n = \frac{1}{n} \sum_{j=1}^{n} X_n \). We have \( E[S_n] = \mu \) and \( Var[S_n] = \frac{1}{n} \sigma^2 \), since the variables are non-correlated. For the claimed mean square convergence we need to consider

\[
E \left[ |S_n - \mu|^2 \right] = E \left[ (S_n - E[S_n])^2 \right] = Var[S_n] = \frac{1}{n} \sigma^2
\]
so that

\[
E \left[ |S_n - \mu|^2 \right] = \frac{1}{n} \sigma^2 \to 0
\]
as \( n \to \infty \), as was claimed.

Since convergence in mean square implies convergence in probability, we have the weak law of large numbers:
Theorem 2.2  The random sequence \( \{X_n\}_{n=1}^{\infty} \) is \( WN(\mu, \sigma^2) \). Then
\[
\frac{1}{n} \sum_{j=1}^{n} X_n \to \mu
\]
in probability.

3  Cauchy-Schwartz and Triangle Inequalities

Lemma 3.1
\[
E[|XY|] \leq \sqrt{E[|X|^2]} \cdot \sqrt{E[|Y|^2]}.
\tag{3.2}
\]
\[
\sqrt{E[|X \pm Y|^2]} \leq \sqrt{E[|X|^2]} + \sqrt{E[|Y|^2]}.
\tag{3.3}
\]

The inequality (3.2) is known as the Cauchy-Schwartz inequality, and the inequality (3.3) is known as the triangle inequality.

4  Properties of mean square convergence

Theorem 4.1  The random sequences \( \{X_n\}_{n=1}^{\infty} \) and \( \{Y_n\}_{n=1}^{\infty} \) are defined in the same probability space and \( E[X_n^2] < \infty \) and \( E[Y_n^2] < \infty \). Let
\[
X = l.i.m._{n \to \infty} X_n, Y = l.i.m._{n \to \infty} Y_n.
\]
Then it holds that
\( (a) \)
\[
E[XY] = \lim_{n \to \infty} E[X_n \cdot Y_n]
\]
\( (b) \)
\[
E[X] = \lim_{n \to \infty} E[X_n]
\]
\( (c) \)
\[
E[|X|^2] = \lim_{n \to \infty} E[|X_n|^2]
\]
(d) \[ E[X \cdot Z] = \lim_{n \to \infty} E[X_n Z] \]

if \( E[Z^2] < \infty \).

Proof: All the rest of the claims follow easily, if we can prove (a). First, we see that \( |E[X_n Y_n]| < \infty \) and \( |E[XY]| < \infty \) in view of Cauchy-Schwartz and the other assumptions. In order to prove (a) we consider

\[ |E[X_n Y_n] - E[XY]| \leq E[(X_n - X)Y_n + X(Y_n - Y)] \]

since \( |E[Z]| \leq E[|Z|] \). Now we can use the ordinary triangle inequality for real numbers and obtain:

\[ E[(X_n - X)Y_n + X(Y_n - Y)] \leq E[(X_n - X)Y_n] + E|[X(Y_n - Y)] \]

But Cauchy-Schwartz entails now

\[ E[(X_n - X)Y_n] \leq \sqrt{E[(X_n - X)^2]} \sqrt{E[|Y_n|^2]} \]

and

\[ E|[Y_n - Y]| \leq \sqrt{E[|Y_n - Y|^2]} \sqrt{E[|X|^2]} \]

But by assumption \( \sqrt{E[(X_n - X)^2]} \to 0 \) and \( \sqrt{E[|Y_n - Y|^2]} \to 0 \), and thus the assertion (a) is proved.

We shall often need Cauchy’s criterion for mean square convergence, which is the next theorem.

**Theorem 4.2** Consider the random sequence \( \{X_n\}_{n=1}^{\infty} \) with \( E[X_n^2] < \infty \) for every \( n \). Then

\[ E[|X_n - X_m|^2] \to 0 \quad (4.4) \]

as \( \min(m,n) \to \infty \) if and only if there exists a random variable \( X \) such that

\[ X = \lim_{n \to \infty} X_n. \]

This is left without a proof.

A useful form of Cauchy’s criterion is known as Loève’s criterion:

**Theorem 4.3**

\[ E[|X_n - X_m|^2] \to 0 \iff E[X_n X_m] \to C. \quad (4.5) \]

as \( \min(m,n) \to \infty \), where the constant \( C \) is finite and independent of the way \( m, n \to \infty \).
Proof: Proof of $\iff$: We assume that $E[X_nX_m] \to C$. Thus
\[
E[|X_n - X_m|^2] = E[X_n \cdot X_n + X_m \cdot X_m - 2X_n \cdot X_m] \\
\to C + C - 2C = 0.
\]
Proof of $\implies$: We assume that $E[|X_n - X_m|^2] \to 0$. Then
\[
E[X_nX_m] = E[(X_n - X)X_m] + E[XX_m].
\]
Here
\[
E[(X_n - X)X_m] \to E[l.i.m.(X_n - X)l.i.m.X_m] = 0,
\]
by theorem 4.1 (a), since $X = l.i.m._{n \to \infty} X_n$ according to Cauchy’s criterion. Also
\[
E[XX_m] \to E[Xl.i.m.X_n]
\]
by theorem 4.1 (d). Hence
\[
E[X_nX_m] \to 0 + C = C.
\]

\section{Applications}
Consider a random sequence $\{X_n\}_{n=0}^{\infty} \sim WN(\mu, \sigma^2)$. We wish to find conditions such that we may regard an infinite linear combination of random variables as a mean square convergent sum i.e.
\[
\sum_{i=0}^{\infty} a_i X_i = l.i.m._{n \to \infty} \sum_{i=0}^{n} a_i X_i
\]
The Cauchy criterion in theorem 4.2 gives for $Y_n = \sum_{i=0}^{n} a_i X_i$ and $n < m$ that
\[
E[|Y_n - Y_m|^2] = E \left[ \left| \sum_{i=n+1}^{m} a_i X_i \right|^2 \right] = \sigma^2 \sum_{i=n+1}^{m} a_i^2 + \mu^2 \left( \sum_{i=n+1}^{m} a_i \right)^2,
\]
since $EZ^2 = Var(Z) + (E[Z])^2$ for any random variable. Hence we see that $E[|Y_n - Y_m|^2]$ converges to zero if and only if
\[
\sum_{i=0}^{\infty} a_i^2 < \infty \text{ and } |\sum_{i=0}^{\infty} a_i| < \infty
\]
in case $\mu \neq 0$ and
\[ \sum_{i=0}^{\infty} a_i^2 < \infty \]
in case $\mu = 0$.

We note here that
\[ \sum_{i=0}^{\infty} |a_i| < \infty \Rightarrow \sum_{i=0}^{\infty} a_i^2 < \infty. \quad (5.6) \]

6 Causal Linear Processes

6.1 A Notation for White Noise

We say that $\{Z_t, t \in T \subseteq Z\}$ is $\text{WN}(0, \sigma^2)$ if $E[Z_t] = 0$ for all $t \in T$ and
\[ \gamma_Z(h) = \begin{cases} \sigma^2 & h = 0 \\ 0 & h \neq 0 \end{cases} \quad (6.1) \]

Next, Kronecker\(^1\) delta, denoted by $\delta_{i,j}$, is defined for integers $i$ and $j$ as
\[ \delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad (6.2) \]

Then we can write the ACVF in (6.1) above as
\[ \gamma_Z(h) = \sigma^2 \cdot \delta_{0,h}. \quad (6.3) \]

and even as
\[ \gamma_Z(r, s) = \sigma^2 \cdot \delta_{r,s}. \quad (6.4) \]

6.2 Causal Linear Processes

Let $\{X_t, t \in T \subseteq Z\}$ be given by
\[ X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad (6.5) \]

where $\{Z_t, t \in T \subseteq Z\}$ is $\text{WN}(0, \sigma^2)$.

\(^1\)Leopold Kronecker, a German mathematician, 1823-1891, has also deeper contributions to mathematics, see
http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Kronecker.html
Definition 6.1 If
\[ \sum_{j=0}^{\infty} |\psi_j| < \infty \] (6.6)
then we say that (6.5) is a causal linear process.

The condition (6.6) guarantees (c.f. (5.6)) that the infinite sum in (6.5) converges in mean square. By causality we mean that the current value \( X_t \) is influenced only by values of the white noise in the past, i.e., \( Z_{t-1}, Z_{t-2}, \ldots \), and its current value \( Z_t \), but not by values in the future. Alternatively,
\[ X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \] (6.7)
is causal if \( \psi_j = 0 \) for \( j < 0 \). Then we can also write
\[ X_t = \sum_{j=-\infty}^{t} \psi_j Z_{j}. \] (6.8)

Now we compute that ACVF of any causal linear process. We use the convergences in theorem 4.1 above. Assume \( h > 0 \).
\[ \gamma_X(t, t+h) = E[X_t X_{t+h}] = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \psi_k \psi_l E[Z_{t-k}Z_{t+h-l}] \]
\[ = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \psi_k \psi_l \sigma^2 \cdot \delta_{t-k, t+h-l} \]
\[ = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k-h} \] (6.9)
where we applied (6.4). Hence for \( h > 0 \)
\[ \gamma_X(h) = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k-h} \] (6.10)

One finds that \( \gamma_X(h) = \gamma_X(-h) \). In fact we have as above that if \( -h < 0 \)
\[ \gamma_X(t, t-h) = E[X_t X_{t-h}] = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \psi_k \psi_l E[Z_{t-k}Z_{t-h-l}] \]
\[ = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k-h} \] (6.11)
Now we have to recall that causality requires $\psi_{k-h}$ for $k-h < 0$, i.e., $k<h$. Thus

$$\sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k-h} = \sigma^2 \sum_{k=h}^{\infty} \psi_k \psi_{k-h}.$$  

By change of variable $s = k - h$ we get

$$\sigma^2 \sum_{k=h}^{\infty} \psi_k \psi_{k-h} = \sigma^2 \sum_{s=0}^{\infty} \psi_{s+h} \psi_s = \gamma_X(h).$$

Hence we have shown the following.

**Proposition 6.1** If $\sum_{j=0}^{\infty} |\psi_j| < \infty$, then a causal linear process

$$X_t = \sum_{j=-\infty}^{t} \psi_{t-j} Z_j,$$  

(6.12)

is (weakly) stationary and we have

$$\gamma_X(h) = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+h}.$$  

(6.13)

The process variance is thus

$$\gamma_Y(0) = \sigma^2 \sum_{k=0}^{\infty} \psi_k^2.$$  

(6.14)