LEcTUEr 7

COnvergence

7.1 Motivation

Suppose we wish to estimate the statistics of a random variable, e.g., its mean, variance, and distribution. To estimate such a statistic, we collect samples and use an estimator in the form of a sample average. But how good is the estimator? Does it “converge” to the true statistic? How many samples do we need to ensure with some confidence that we are within a certain range of the true value of the statistic? These are questions that often arise in statistics and learning, commonly referred to as parametric and nonparametric estimation.

In communications and signal processing, one faces another type of estimation, namely, estimation or detection of a signal from noisy observations. We then ask if a given estimator converges to the true signal or how many observations are needed to achieve a desired estimation accuracy.

The subject of convergence and limit theorems for random variables addresses such questions. As a motivating example, we consider the problem of estimating the mean of a random variable. Let \( X \) be a random variable with finite but unknown mean \( E[X] \). To estimate the mean we generate \( X_1, \ldots, X_n \) i.i.d. samples drawn according to the same distribution as \( X \) and compute the sample mean

\[
S_n = \frac{1}{n} \sum_{i=1}^{n} X_i.
\]

Does \( S_n \) converge to \( E[X] \) as we increase \( n \)? If so, how fast? More concretely, let \( X_1, \ldots, X_n \) be i.i.d. \( N(0,1) \). Figure plots the sample mean \( S_n \) from 6 sets of outcomes of \( X_1, \ldots, X_n \). Note that each \( s_n \) sequence appears to be converging to 0, the mean of the random variables, as \( n \) increases.

But what does it mean to say that a sequence of random variables \( S_1, S_2, \ldots \) converges to \( E[X] \)? For a sequence of real numbers \( a_1, a_2, \ldots \), the limit

\[
a = \lim_{n \to \infty} a_n
\]

exists if for every \( \epsilon > 0 \), there exists \( n(\epsilon) \) such that \( |a_m - a| < \epsilon \) for every \( m \geq n(\epsilon) \). For a sequence of random variables, there are several different notions of convergence.
7.2 ALMOST SURE CONVERGENCE

Consider a sequence of random variables $X_1, X_2, \ldots$, all defined on the same probability space $\Omega$. For every $\omega \in \Omega$, we obtain a sample sequence (or sample path) $X_1(\omega), X_2(\omega), \ldots$, which is a sequence of real numbers. We say that the sequence of random variables $X_1, X_2, X_3, \ldots$ converges to a random variable $X$ almost surely (or with probability $\nu$) if

$$\mathbb{P}\{\omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\} = 1.$$ 

In other words, the set of sample paths that converge to $X(\omega)$, in the sense of a sequence converging to a limit, has probability 1. Equivalently, the sequence $X_1, X_2, \ldots$ converges to $X$ almost surely if for every $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}\{|X_n - X| < \epsilon \text{ for every } n \geq m\} = \mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{|X_n - X| < \epsilon\}\right) = 1.$$ 

Figure 7.1. Convergence of the sample mean to the ensemble mean.

7.2 ALMOST SURE CONVERGENCE

Consider a sequence of random variables $X_1, X_2, \ldots$, all defined on the same probability space $\Omega$. For every $\omega \in \Omega$, we obtain a sample sequence (or sample path) $X_1(\omega), X_2(\omega), \ldots$, which is a sequence of real numbers. We say that the sequence of random variables $X_1, X_2, X_3, \ldots$ converges to a random variable $X$ almost surely (or with probability $\nu$) if

$$\mathbb{P}\{\omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\} = 1.$$ 

In other words, the set of sample paths that converge to $X(\omega)$, in the sense of a sequence converging to a limit, has probability 1. Equivalently, the sequence $X_1, X_2, \ldots$ converges to $X$ almost surely if for every $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}\{|X_n - X| < \epsilon \text{ for every } n \geq m\} = \mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{|X_n - X| < \epsilon\}\right) = 1.$$ 

Figure 7.1. Convergence of the sample mean to the ensemble mean.
Example 7.1. Let $X_1, \ldots, X_n$ be i.i.d. Bern(1/2), and define $Y_n = 2^n \prod_{i=1}^{n} X_i$. Then for any $\varepsilon > 0$ (such that $\varepsilon < 2^n$),

$$P(|Y_n - 0| < \varepsilon \text{ for all } n \geq m) = P(X_n = 0 \text{ for some } n \leq m) = 1 - P(X_n = 1 \text{ for all } n \leq m) = 1 - \left(\frac{1}{2}\right)^m,$$

which converges to 1 as $m \to \infty$. Hence, the sequence $Y_n$ converges to 0 almost surely.

An important example of almost sure convergence is the strong law of large numbers (SLLN), which states that if $X_1, \ldots, X_n$ are i.i.d. with finite mean $E[X]$, then the sequence of sample means $S_n$ converges to the mean $E[X]$ almost surely. The example shown in Figure 7.1 is a good demonstration of the SLLN — each of the sample paths appears to be converging to 0, which is $E[X]$, and the probability of such sample paths is 1. The proof of the SLLN is beyond the scope of this course and omitted.

7.3 CONVERGENCE IN MEAN SQUARE

We say that a sequence of random variables $X_1, X_2, \ldots$ converges to a random variable $X$ in mean square (m.s.) if

$$\lim_{n \to \infty} E[(X_n - X)^2] = 0.$$

Example 7.2. Let $X_1, X_2, \ldots$ be i.i.d. with finite mean $E[X]$ and variance $\text{Var}(X)$. Then, $S_n \to E[X]$ in m.s. as $n \to \infty$. In other words,

$$\lim_{n \to \infty} E[(S_n - E(X))^2] = 0.$$

To show this, first note that $S_n$ is an unbiased estimate of $E[X]$, i.e.,

$$E[S_n] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = E[X].$$

Now to prove the convergence in m.s., consider

$$E[(S_n - E[X])^2] = E[(S_n - E[S_n])^2]$$

$$= \frac{1}{n^2} E \left[ \left( \frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{n} \sum_{i=1}^{n} E[X] \right)^2 \right]$$

$$= \frac{1}{n^2} \text{Var} \left( \sum_{i=1}^{n} X_i \right)$$
\[
\frac{1}{n^2} \left( \sum_{i=1}^{n} \text{Var}(X_i) \right) = \frac{1}{n} \text{Var}(X),
\]

which tends to zero as \( n \to \infty \). Note that the proof works even if the random variables are only pairwise independent or even only uncorrelated.

**Example 7.3.** Consider the linear MMSE estimates Example 6.10 as a sequence of random variables \( \hat{X}_1, \hat{X}_2, \ldots \), where \( \hat{X}_n \) is the LMMSE estimate of \( X \) given the first \( n \) observations. This sequence converges in m.s. to \( X \) since the sequence of MSEs converges to 0 as \( n \to \infty \).

Mean square convergence does not necessarily imply almost sure convergence.

**Example 7.4.** Consider a sequence of random variables

\[
X_n = \begin{cases} 
0 & \text{w.p. } 1 - 1/n, \\
1 & \text{w.p. } 1/n.
\end{cases}
\]

Since \( \text{E}[X_n^2] = 1/n \), this sequence converges to 0 in m.s. It does not, however, converge almost surely, since for \( 0 < \epsilon < 1 \) and any \( m \)

\[
P(|X_n - 0| < \epsilon \text{ for all } n \geq m) = \lim_{n \to \infty} \prod_{i=m}^{n} \left( 1 - \frac{1}{i} \right) = \lim_{n \to \infty} \prod_{i=m}^{n} \left( \frac{i-1}{i} \right) = \lim_{n \to \infty} \left( \frac{m-1}{n} \right) = 0.
\]

Almost sure convergence does not imply mean square convergence either.

**Example 7.5.** Consider the sequence \( Y_n \) in Example 7.1. Since

\[
\text{E}[(Y_n - 0)^2] = \left( \frac{1}{2} \right)^n 2^{2n} = 2^n,
\]

the sequence does not converge in m.s. even though it converges almost surely.

### 7.4 Convergence in Probability

We say that a sequence of random variables \( X_1, X_2, \ldots \) converges to \( X \) in probability if for any \( \epsilon > 0 \),

\[
\lim_{n \to \infty} P(|X_n - X| < \epsilon) = 1.
\]
Clearly, almost sure convergence implies convergence in probability. The converse is not necessarily true.

**Example 7.6.** Let \( X_1, X_2, \ldots \) be independent and

\[
X_n = \begin{cases} 
0 & \text{w.p. } 1 - 1/n, \\
n & \text{w.p. } 1/n.
\end{cases}
\]

This sequence converges in probability to 0, since

\[
P(|X_n - 0| > \epsilon) = P(X_n > \epsilon) = \frac{1}{n} \to 0 \text{ as } n \to \infty.
\]

But it does not converge almost surely, since as in Example 7.4,

\[
\lim_{n \to \infty} P(|X_n - 0| < \epsilon \text{ for all } n \geq m) = 0.
\]

Convergence in mean square implies convergence in probability. Indeed, by the Markov inequality, for any \( \epsilon > 0 \),

\[
P(|X_n - X| > \epsilon) = P((X_n - X)^2 > \epsilon^2) \leq \frac{E(X_n - X)^2}{\epsilon^2}.
\]

Hence, if \( X_n \to X \) in m.s., i.e.,

\[
\lim_{n \to \infty} E[(X_n - X)^2] = 0,
\]

then \( X_n \to X \) in probability, i.e.,

\[
\lim_{n \to \infty} P(|X_n - X| > \epsilon) = 0.
\]

The converse is not necessarily true. In Example 7.6, \( X_n \) converges in probability, but

\[
E((X_n - 0)^2) = n
\]

and thus \( X_n \) does not converge in m.s. Hence, convergence in probability is weaker than both almost sure convergence and mean square convergence.

The most important example of convergence in probability is the weak law of large numbers (WLLN).

**Weak law of large numbers.** Let \( X_1, X_2, \ldots \) be a sequence of i.i.d. random variables with finite mean \( E[X] \) and variance \( Var(X) \). Then the sample mean

\[
S_n = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

converges to \( E[X] \) in probability.
We already proved that $S_n \to E[X]$ in m.s., and since convergence in m.s. implies convergence in probability, $S_n \to E[X]$ in probability. Moreover, the WLLN requires only that the random variables are uncorrelated (at the cost of finite variance), while the SLLN requires independence.

**Example 7.7 (Confidence interval).** Given $\epsilon, \delta > 0$, how large should $n$, the number of samples, be so that

$$P(|S_n - E[X]| \leq \epsilon) \geq 1 - \delta,$$

that is, $S_n$ is within $\pm \epsilon$ of $E[X]$ with probability $\geq 1 - \delta$? To answer this question, we use the Chebyshev inequality:

$$P(|S_n - E[X]| \leq \epsilon) = P(|S_n - E[S_n]| \leq \epsilon) \geq 1 - \frac{\text{Var}(S_n)}{\epsilon^2} = 1 - \frac{\text{Var}(X)}{n\epsilon^2}.$$

Hence, $n$ should satisfy

$$\frac{\text{Var}(X)}{n\epsilon^2} \leq \delta,$$

or equivalently,

$$n \geq \frac{\text{Var}(X)}{\delta\epsilon^2}.$$

For example, if $\epsilon = 0.1\sigma_X$ and $\delta = 0.001$, then the number of samples should satisfy

$$n \geq \frac{(0.1\sigma_X)^2}{0.001 \times 0.01\sigma_X^2} = 10^5.$$

In other words, $10^5$ samples ensure that one can be 99.9% confident that the unknown true mean $E[X]$ lies within $\pm 0.1\sigma_X$ of the sample mean $S_n$ from the data, regardless of the distribution of $X$.

### 7.5 CONVERGENCE IN DISTRIBUTION

We say that a sequence of random variables $X_1, X_2, \ldots$ converges *in distribution* (or *weakly*) to $X$ if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

for every $x$ at which $F_X(x)$ is continuous. Convergence in probability implies convergence in distribution — and convergence in distribution (weak convergence) is the weakest form of convergence we discuss.

The most important example of convergence in distribution is the central limit theorem (CLT).
Central limit theorem. Let $X_1, X_2, \ldots$ be i.i.d. random variables with finite mean $E[X]$ and variance $\sigma^2_X$. Then, the normalized sum

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( X_i - E(X) \right) / \sigma_X$$

converges to $Z \sim N(0, 1)$ in distribution, i.e., $\lim_{n \to \infty} F_{Z_n}(z) = \Phi(z)$.

Example 7.8. Let $X_1, X_2, \ldots$ be i.i.d. $\text{Unif}[-1, 1]$ random variables. Then the pdf of the normalized sum

$$Z_n = \frac{1}{\sqrt{n/3}} \sum_{i=1}^{n} X_i$$

converges to that of the standard normal random variable, as illustrated in Figure 7.2 for $n = 1, 2, 4, 16$. Note how quickly the pdf converges. It can be shown that convergence of the pdf implies convergence of the cdf (weak convergence).

Figure 7.2. The pdfs of the normalized sums $Z_n$ in Example 7.8.
Example 7.9. Let $X_1, X_2, \ldots$ be i.i.d. Bern(1/2). The normalized sum $Z_n = \sum_{i=1}^{n}(X_i - 0.5)/\sqrt{n/4}$ is discrete and thus has no pdf, but its cdf converges to the Gaussian cdf, as depicted in Figure 7.3 for $n = 10, 20, 160$.

![Figure 7.3](image)

**Figure 7.3.** The cdfs of the normalized sums $Z_n$ in Example 7.9

Example 7.10 (Confidence interval). Let $X_1, X_2, \ldots$ be i.i.d. with finite mean $E(X)$ and variance $\text{Var}(X)$ and let $S_n$ be the sample mean. Given $\epsilon, \delta > 0$, how large should $n$ be so that

$$P[|S_n - E(X)| \leq \epsilon] \geq 1 - \delta?$$

We can use the CLT to find an estimate of $n$:

$$P[|S_n - E[S_n]| \leq \epsilon] = P\left\{\left|\frac{1}{n} \sum_{i=1}^{n}(X_i - E[X])\right| \leq \epsilon\right\}$$

$$= P\left\{\left|\frac{1}{\sigma_X \sqrt{n}} \sum_{i=1}^{n}(X_i - E[X])\right| \leq \frac{\epsilon \sqrt{n}}{\sigma_X}\right\}$$

$$\approx 1 - 2Q\left(\frac{\epsilon \sqrt{n}}{\sigma_X}\right).$$
For $\varepsilon = 0.1\sigma_X$, $\delta = 0.001$, set $2Q(0.1\sqrt{n}) = 0.001$, so $0.1\sqrt{n} = 3.3$ or $n = 1089$ — much smaller than $n \geq 10^5$ obtained in Example 7.7 by the Chebyshev inequality.

The CLT applies to i.i.d. sequences of random vectors. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. $k$-dimensional random vectors with finite mean $\mu$ and nonsingular covariance matrix $\Sigma$. Define the sequence of random vectors $Z_1, Z_2, \ldots$ by

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu).$$

Then the CLT for random vectors states that as $n \to \infty$

$$Z_n \to Z \sim N(\mu, \Sigma)$$

in distribution.

Example 7.11. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. $2$-dimensional random vectors with

$$f_{X_1}(x_{11}, x_{12}) = \begin{cases} x_{11} + x_{12} & 0 < x_{11} < 1, \ 0 < x_{12} < 1, \\ 0 & \text{otherwise}. \end{cases}$$

Figure 7.2 shows the joint pdf of $Y_n = \sum_{i=1}^{n} X_i$ for $n = 1, 2, 3, 4$. Note how quickly it looks Gaussian.

![Figure 7.4. Plots of the joint pdf of $Y_n$ in Example 7.11](image-url)
PROBLEMS

7.1. Consider the sequence of i.i.d. random variables $X_1, X_2, \ldots$ with mean $E(X_i) = 2$ and finite variance. Define the sequence

$$Y_n = \begin{cases} 
X_n, & \text{for all } n \text{ w.p. } \frac{1}{3}, \\
\frac{1}{2}X_n, & \text{for all } n \text{ w.p. } \frac{1}{3}, \\
0, & \text{for all } n \text{ w.p. } \frac{1}{3}.
\end{cases}$$

Let

$$M_n = \frac{1}{n} \sum_{i=1}^{n} Y_i.$$ 

Define the random variable (or constant) that $M_n$ converges to (in probability) as $n$ approaches infinity and prove the convergence.

7.2. Roundoff errors. The sum of a list of 100 real numbers is to be computed. Suppose that these numbers are rounded off to the nearest integer so that each number has an error that is uniformly distributed in the interval $(-0.5, 0.5)$. Use the central limit theorem to estimate the probability that the total error in the sum of the 100 numbers exceeds 6.

7.3. The signal received over a wireless communication channel can be represented by two sums

$$X_{1n} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_j \cos \Theta_j,$$

$$X_{2n} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_j \sin \Theta_j,$$

where $Z_1, Z_2, \ldots$ are i.i.d. with mean $\mu$ and variance $\sigma^2$ and $\Theta_1, \Theta_2, \ldots$ are i.i.d. $U[0, 2\pi]$ independent of $Z_1, Z_2, \ldots$. Find the distribution of $\left[ \frac{X_{1n}}{X_{2n}} \right]$ as $n$ approaches $\infty$.

7.4. Polya's urn. An urn initially has one red ball and one white ball. Let $X_1$ denote the name of the first ball drawn from the urn. Replace that ball and one like it. Let $X_2$ denote the name of the next ball drawn. Replace it and one like it. Continue, drawing and replacing.

(a) Argue that the probability of drawing $k$ reds followed by $n - k$ whites is

$$\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{k}{k+1} \cdot \frac{1}{(k+2)} \cdot \frac{(n-k)}{(n+1)} = \frac{k!(n-k)!}{(n+1)!} = \frac{1}{(n+1)} \binom{n}{k}.$$
(b) Let $P_n$ be the proportion of red balls in the urn after the $n$-th drawing. Argue that $P[P_n = \frac{k}{n+2}] = \frac{1}{n+1}$, for $k = 1, 2, \ldots, n+1$. Thus all proportions are equally probable. This shows that $P_n$ tends to a uniformly distributed random variable in distribution, i.e.,

$$\lim_{n \to \infty} P[P_n \leq t] \to t, \ 0 \leq t \leq 1.$$ 

(c) What can you say about the behavior of the proportion $P_n$ if you started initially with one red ball in the urn and two white balls? Specifically, what is the limiting distribution of $P_n$? Can you show $P[P_n = \frac{k}{n+3}] = \frac{k}{n+2}$, for $k = 1, 2, \ldots, n+1$?

7.5. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with $X_i \sim \text{Exp}(\lambda)$. Show that the sequence of r.v.s $Y_n = \min\{X_1, X_2, \ldots, X_n\}$ converges w.p.1. What is the limit?

7.6. Let $X_1, X_2, \ldots$ be independent random variables with the same finite mean $\mu \neq 0$ and variance $\sigma^2$. Find the limit of $P[\frac{1}{n} \sum_{i=1}^{n} X_i < \frac{\mu}{2}]$ as $n$ approaches infinity.

7.7. Convergence Consider the following sequences of random variables defined on the probability space $(\Omega, \mathcal{F}, P)$, where $\Omega = \{0, 1, 2, \ldots, m-1\}$, $\mathcal{F}$ is the collection of all subsets of $\Omega$, and $P$ is the uniform distribution over $\Omega$.

$$X_n(\omega) = \begin{cases} \frac{1}{n}, & \omega = n \text{ mod } m \\ 0, & \text{otherwise} \end{cases}$$

$$Y_n(\omega) = \begin{cases} 2^n, & \omega = 1 \\ 0, & \text{otherwise} \end{cases}$$

$$Z_n(\omega) = \begin{cases} 1, & \omega = 1 \\ 0, & \text{otherwise} \end{cases}$$

Which of these sequences converges to zero

(a) with probability one?

(b) in mean square?

(c) in probability?

7.8. Consider a coin with random bias $P$. Flip it $n$ times independently to generate $X_1, X_2, \ldots, X_n$, where $X_i = 1$ if the $i$th outcome is heads, and $X_i = 0$ otherwise. Let $S_n$ be the sample average. Prove that $S_n$ converges to $P$ in probability. (Hint: First prove convergence in mean square.)