ESE 520 Probability and Stochastic Processes

Lecture 13

"Convergence of random variables. The law of large numbers."

Given a sequence of random variables

$$X_n : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n = 1, 2, ...,$$

it is natural to study the convergence of $\{X_n\}, n = 1, 2, ...$

There are several types of convergence we can introduce when talking about a sequence of random variables: *convergence in probability*, *convergence with probability one* *(almost everywhere - a.e.)*, *convergence in distribution*, *convergence in mean or mean square* etc.

Definition 1. a) A sequence of random variables $\{X_n\}, n = 1, 2, ...$ converges to a random variable X *in probability* if for any $\epsilon > 0$

$$\lim_{n\to\infty} P(w : |X_n(w) - X(w)| > \epsilon) = 0. \tag{1}$$

b) We say that a sequence of random variables $\{X_n\}, n = 1, 2, ...$ converges to X *almost everywhere (a.e.)* if there exist a set $N \in \mathcal{F}$ with $P(N) = 0$ so that for all $w \notin N$:

$$\lim_{n\to\infty} X_n(w) = X(w). \tag{2}$$

The condition (2) is then equivalent to

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

Convergence almost everywhere is also called convergence with probability one.

Our particular interest here is the following: consider a sequence of i.i.d. (independent and identically distributed) random variables $\{X_n\}, n = 1, 2, ...$ with $E(X_n) = \mu$ and $Var(X_n) = \sigma^2$ for all $n = 1, 2, ...$

Define

$$\bar{X}_n := \frac{X_1 + X_2 + ... + X_n}{n},$$

the moving average.

Question: What can be said about the *asymptotic behavior* of $\bar{X}_n$? In other words, what is

$$\lim_{n\to\infty} \bar{X}_n?$$

(if it exists).

First, we notice that

$$E[\bar{X}_n] = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu$$

and, using independence of random variables, we also get

$$Var[\bar{X}_n] = Var[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

We see that

$$Var(\bar{X}_n) \to 0 \text{ as } n \to \infty$$

so that it is naturally to expect that

$$\bar{X}_n \to \mu \text{ as } n \to \infty.$$

Indeed, we will show that it holds:

$$\bar{X}_n \to \mu \text{ as } n \to \infty \text{ in probability} \tag{3}$$

(*weak law of large numbers*: WLLN)

and

$$\bar{X}_n \to \mu \text{ as } n \to \infty \text{ a.e.} \tag{4}$$

(*strong law of large numbers*: SLLN).

Statements (3) and (4) play an important role in probability theory and its applications. In particular, law of large numbers provide the basis of so-called Monte Carlo simulations and Monte Carlo computations of probabilities and expected values.

Example 1. (*probability $P(A)$ of an event $A \subset \Omega$ is the limit of its relative frequency*). Let $A \in \mathcal{F}$ and define the random variable

$$X(w) = \begin{cases} 1, & w \in A \\ 0, & w \notin A. \end{cases}$$

Clearly,

$$\mu = E(X) = 1 \times P(A) + 0 \times P(A^c) = P(A).$$

Now, let $X_n, n = 1, 2, ...$ be the outcomes of independent repetitions of an experiment with A as "success" ("Bernoulli trial"). Then all X_n are independent random variables having the same distribution as X.

By the construction, we can interpret

$$\bar{X}_n = \frac{X_1 + X_2 + ... + X_n}{n} = \frac{number\ of\ successes\ in\ n\ experiments}{number\ of\ all\ experiments}$$

as the *relative frequency* of event A.

Strong LLN says that the relative frequency of an event A in independent repetitions of the experiment converges to the probability of A as $n \to \infty$ a.e.

Let us prove the weak LLN and the strong LLN.

Proposition. (*Chebyshev's inequality*). Assume that $EX^2 < \infty$. Then, for any $\epsilon > 0$:

$$P(|X| \geq \epsilon) \leq \frac{EX^2}{\epsilon^2}.$$

Proof. (for continuous case; in a discrete case - similar).

$$EX^2 = \int_{-\infty}^{\infty} x^2 f_X(x)dx \geq \int_{\{x:|x|\geq\epsilon\}} x^2 f_X(x)dx \geq$$

$$\int_{\{x:|x|\geq\epsilon\}} \epsilon^2 f_X(x)dx = \epsilon^2 \int_{\{x:|x|\geq\epsilon\}} f_X(x)dx = \epsilon^2 P(|X| \geq \epsilon)$$

implying the statement.

Using Chebyshev's inequality, one has for any $\epsilon > 0$:

$$P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{1}{\epsilon^2} E(\bar{X}_n - \mu)^2 =$$

$$\frac{1}{\epsilon^2} Var(\bar{X}_n) = \frac{1}{\epsilon^2}\frac{\sigma^2}{n} \to 0 \text{ as } n \to \infty$$

so that the weak LLN (3) is proven.

The proof of strong LLN (4) is more difficult and relies on the use of some inequalities of independent interest.

Lemma (Toeplitz): For $a_n \geq 0$, define $b_n := \sum_{i=1}^n a_i$ and assume that $\lim_{n\to\infty} b_n = +\infty$ (b_n diverges). Then, if $x_n \to x$ as $n \to \infty$, it also holds that

$$\frac{1}{b_n}\sum_{i=1}^n a_i x_i \to x \text{ as } n \to \infty.$$

Proof. Let $\epsilon > 0$. We can choose an integer $N_0 \in \mathbb{N}$ such that $|x_n - x| \leq \frac{\epsilon}{2}$ for all $n \geq N_0$. We also can choose $N_1 \geq N_0$ so that

$$\frac{1}{b_{N_1}}\sum_{i=1}^{N_0} a_i |x_i - x| \leq \frac{\epsilon}{2}.$$

As the result, for all $n \geq N_1$, it holds:

$$|\frac{1}{b_n}\sum_{i=1}^n a_i x_i - x| = |\frac{1}{b_n}\sum_{i=1}^n a_i (x_i - x)| \leq$$

$$\frac{1}{b_n}\sum_{i=1}^{N_0} a_i |x_i - x| + \frac{1}{b_n}\sum_{i=N_0+1}^{n} a_i |x_i - x| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

since $b_{N_1} \leq b_n$ and $\frac{1}{b_n}\sum_{i=N_0+1}^n a_i \leq 1$. $\square$

Corollary. If $x_n \to x$ as $n \to \infty$, then $\frac{1}{n}\sum_{i=1}^n x_i \to x$ as $n \to \infty$. (Choose $a_n = 1, b_n = n$ and apply the Lemma above).

Lemma (Kronecker). Assume $b_n > 0$ so that $b_n \uparrow \infty$ as $n \to \infty$ and $\sum_{i=1}^\infty x_i < \infty$. Then

$$\frac{1}{b_n}\sum_{i=1}^n b_i x_i \to 0 \text{ as } n \to \infty.$$

Proof. Let

$$S_n := \sum_{i=1}^n x_i \text{ with } S_0 = 0, b_0 = 0.$$

Summation by parts yields:

$$\sum_{i=1}^{n} b_i x_i = \sum_{i=1}^{n} b_i (S_i - S_{i-1}) = b_n S_n - S_0 b_0 - \sum_{i=1}^{n} (b_i - b_{i-1}) S_{i-1} =$$

$$b_n S_n - \sum_{i=1}^{n} (b_i - b_{i-1}) S_{i-1}.$$

Dividing both sides by b_n, we obtain

$$\frac{1}{b_n} \sum_{i=1}^{n} b_i x_i = S_n - \frac{1}{b_n} \sum_{i=1}^{n} a_i S_{i-1} \tag{5}$$

where $a_i := b_i - b_{i-1}$ so that $\sum_{i=1}^{n} a_i = b_n \uparrow \infty$ as $n \to \infty$.

By assumption, the sequence $\{S_n\}$ converges, say to S and, by Toeplitz's Lemma, the sequence $\frac{1}{b_n} \sum_{i=1}^{n} a_i S_{i-1}$ converges to the same limit S so that the right site in (5) goes to zero as $n \to \infty$. $\square$

Corollary. Let $\sum_{n=1}^{\infty} \frac{y_n}{n} < \infty$. Then

$$\frac{1}{n} \sum_{i=1}^{n} y_i \to 0 \text{ as } n \to \infty.$$

(Choose $b_n = n, x_n = \frac{y_n}{n}$)

Now, since

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

the strong LLN will be proven if we are able to show (by Kronecker's Lemma) that the series

$$\sum_{n=1}^{\infty} \left(\frac{X_n - \mu}{n}\right)$$

converges a.e.

Set

$$S_k := \sum_{n=1}^{k} \left(\frac{X_n - \mu}{n}\right).$$

Then we have to show that the sequence $\{S_k\}$ converges a.e.

Let $N := \{w : S_k(w)$ does not converge $\}$.

For all $w \notin N$, $S_k(w)$ is a Cauchy sequence. That is, for any $\epsilon > 0$ there is $\tilde{N} \in \mathbb{N}$ so that for all $m \geq \tilde{N}$ it holds $|S_m(w) - S_{\tilde{N}}(w)| \leq \epsilon$.

Using the negation argument, we can write that:

For all $w \in N$ there exists a number $\epsilon_0 > 0$ so that for all $\tilde{N} \in \mathbb{N}$ there exist $m \geq \tilde{N}$ so that $|S_m(w) - S_{\tilde{N}}(w)| > \epsilon_0$.

Without loss of generality, we can always choose $\epsilon_0 = \frac{1}{r}$ for some $r \in \mathbb{N}$.

Using the above, we can write N as

$$N = \cup_{r=1}^{\infty} \cap_{\tilde{N}=1}^{\infty} \cup_{m=\tilde{N}}^{\infty} \{w : |S_m(w) - S_{\tilde{N}}(w)| > \frac{1}{r}\}. \tag{6}$$

We note that the representation (6) shows that $N \in \mathcal{F}$ so that $P(N)$ exists.

We can write

$$N = \cup_{r=1}^{\infty} N_r.$$

Since

$$P(N) = P(\cup_{r=1}^{\infty} N_r) \leq \sum_{r=1}^{\infty} P(N_r),$$

it is enough to show that

$$P(N_r) = 0 \text{ for all } r \in \mathbb{N}.$$

Fix r. Then

$$N_r \subset \cup_{m=\tilde{N}}^{\infty} \{w : |S_m(w) - S_{\tilde{N}}(w)| > \frac{1}{r}\}$$

for all $\tilde{N} = 1, 2, \ldots.$

Therefore,

$$P(N_r) \leq P(\cup_{m=\tilde{N}}^{\infty} \{w : |S_m(w) - S_{\tilde{N}}(w)| > \frac{1}{r}\}) =$$

$$P(\sup_{m \geq \tilde{N}} \{w : |S_m(w) - S_{\tilde{N}}(w)| > \frac{1}{r}\})$$

for all $\tilde{N} = 1, 2, ...$ so that

$$P(N_r) \leq \lim_{\tilde{N}\to\infty} P(\sup_{m\geq\tilde{N}} \{w : |S_m(w) - S_{\tilde{N}}(w)| > \frac{1}{r}\}). \tag{7}$$

To finish the proof of the strong LLN, we have to show that the right site of (7) is equal to zero. For that, we need one more inequality.