ESE 520 Probability and Stochastic Processes

Lecture 14

"The law of large numbers continued. Central limit theorem."

Lemma (Kolmogorov's inequality). Let $X_1, X_2, ...$ be independent random variables with $E(X_i) = 0$ and $E(X_i^2) < \infty$ for all $i = 1, 2, ...$ Then for any $\epsilon > 0$:

$$P(\max_{1\le k\le n} |\sum_{i=1}^{n} X_i| \geq \epsilon) \leq \frac{1}{\epsilon^2} E(\sum_{i=1}^{n} X_i)^2 = \frac{1}{\epsilon^2} \sum_{i=1}^{n} EX_i^2.$$

Note that $\frac{1}{\epsilon^2} E(\sum_{i=1}^{n} X_i)^2 = \frac{1}{\epsilon^2} \sum_{i=1}^{n} EX_i^2$ because of the independence of random variables (exercise!)

Proof. Let

$$S_k := \sum_{i=1}^{k} X_i$$

so that $ES_k = 0$.

Introduce (by fixed n)

$$A := \{w \in \Omega : \max_{1\le k\le n} |S_k| > \epsilon\}$$

and consider the events of "first occurrence"

$$A_k := \{w \in \Omega : |S_i| \leq \epsilon \text{ for } i = 1, 2, ..., k-1; |S_k| > \epsilon|\}.$$

Then

$$A = \cup_{k=1}^{n} A_k$$

where all events $A_1, A_2, ..., A_n$ are disjoint events.

We can then calculate (in case of integrals; similar - in discrete case using sums):

$$ES_n^2 \geq \int_A S_n^2 dP = \sum_{k=1}^{n} \int_{A_k} S_n^2 dP = \sum_{k=1}^{n} \int_{A_k} (S_k + \sum_{i=k+1}^{n} X_i)^2 dP =$$

$$\sum_{k=1}^{n} \{\int_{A_k} S_k^2 dP + 2\int_{A_k} S_k(\sum_{i=k+1}^{n} X_i) dP + \int_{A_k} (\sum_{i=k+1}^{n} X_i)^2 dP\}.$$

By independence,

$$\int_{A_k} S_k(\sum_{i=k+1}^{n} X_i)dP = E[S_k\mathbf{1}_{A_k}(\sum_{i=k+1}^{n} X_i)] =$$

$$E[S_k\mathbf{1}_{A_k}]E[\sum_{i=k+1}^{n} X_i] = 0$$

since the event A_k depends only on $X_1, X_2, ..., K_k$ and is independent of $X_i, i = k+1, .., n$.
Since

$$\int_{A_k} (\sum_{i=k+1}^{n} X_i)^2 dP \geq 0,$$

it follows that

$$ES_n^2 \geq \sum_{k=1}^{n} \int_{A_k} S_k^2 dP \geq \sum_{k=1}^{n} \int_{A_k} \epsilon_k^2 dP = \epsilon^2 \sum_{k=1}^{n} P(A_k) = \epsilon^2 P(A).$$

Now, we use the Kolmogorov's inequality to finish the proof of strong LLN. We had:

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

Noticing that the events

$$\{\max_{1\leq k\leq n} \{w : |S_{\tilde{N}+k}(w) - S_{\tilde{N}}(w)| > \frac{1}{r}\}$$

will increase monotonically (over n) to the event

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

and using the continuity of P, we obtain by Kolmogorov's inequality:

$$P(N_r) \leq r^2 \lim_{\tilde{N}\to\infty} \lim_{n\to\infty} \sum_{k=\tilde{N}+1}^{\tilde{N}+n} E(\frac{X_k - \mu}{k})^2 =$$

$$r^2 \lim_{\tilde{N}\to\infty} \sum_{k=\tilde{N}+1}^{\infty} \frac{\sigma^2}{k^2} = 0.$$

$\square$

Example 1. A monkey is trained to type, but he types at random, i.e. say there are k possible characters and the probability to hit any particular one is $1/k$. What is the probability that he will type a particular word consisting of N total characters if successive characters are independent? What if we let him try again?...and again?....and again?

$\circ$ If we denote by X_i the outcome of the i-th experiment (i-th try of typing) so that $X_i = 1$ if the word was typed correctly and $X_i = 0$ if not, then by the strong LLN:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to (1/k)^N \text{ as } n \to \infty.$$

with probability one.

Now we want to address the question about *the asymptotic distribution* of the sequence $\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$.

To be more precise, define

$$Z_n := \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma\sqrt{n}}\right)$$

(we just normalized $\bar{X}_n$)

Theorem (Central limit theorem). Assume that $\{X_i\}, i = 1, 2, \ldots$ is a sequence of i.i.d. random variables with $E(X_i) = \mu, Var(X_i) = \sigma^2, i = 1, 2, \ldots$ Then, for any $a \in \mathbb{R}$:

$$P(Z_n \leq a) \to \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

as $n \to \infty$. In other words, in limit the cdf of Z_n is *standard normal.*

In other words, for large n, we can say that $\bar{X}_n$ has *approximately* normal distribution $\mathcal{N}(\mu, \frac{\sigma^2}{n})$.

Proof. We use characteristic functions to prove the CLT. We write

$$Z_n = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma\sqrt{n}}\right) = \sum_{i=1}^{n} Y_i$$

where $Y_i = \frac{X_i - \mu}{\sigma\sqrt{n}}, i = 1, 2, ...$, are independent random variables.

Let $\phi(t)$ be the characteristic function of the random variable

$$(X_i - \mu), i = 1, 2, ...$$

It follows then that the random variable $\frac{X_i - \mu}{\sigma\sqrt{n}}$ has the characteristic function

$$\phi(\frac{t}{\sigma\sqrt{n}}).$$

Since $E(X_i^2) < \infty$, $\phi(t)$ is differentiable at $t = 0$ and can be written as (Taylor expansion):

$$\phi(t) = \phi(0) + \phi'(0)t + \frac{1}{2}\phi''(0)t^2 + o(t^2).$$

Since $\phi'(0) = E(X - \mu), i^2\phi''(0) = E(X - \mu)^2$, the last relation is written as

$$\phi(t) = 1 - \frac{1}{2}\sigma^2 t^2 + o(t^2)$$

or, equivalently,

$$\phi(\frac{t}{\sigma\sqrt{n}}) = 1 - \frac{1}{2}\frac{t^2}{n} + o(\frac{t^2}{n}).$$

Therefore,

$$\phi_{Z_n}(t) = \prod_{i=1}^{n} \phi_{Y_i}(t) = [\phi(\frac{t}{\sigma\sqrt{n}})]^n = [1 - \frac{1}{2}\frac{t^2}{n} + o(\frac{t^2}{n})]^n =$$

$$[1 - \frac{1}{2}\frac{t^2}{n}(1 + o(1))]^n = [1 + \frac{-\frac{1}{2}t^2(1 + o(1))}{n}]^n.$$

It follows then that

$$\psi(t) := \lim_{n\to\infty} \phi_{Z_n}(t) = \lim_{n\to\infty} [1 + \frac{-\frac{1}{2}t^2(1 + o(1))}{n}]^n = e^{-\frac{1}{2}t^2}.$$

But: $\psi(t) = e^{-\frac{1}{2}t^2}$ is the characteristic function of $\mathcal{N}(0, 1)$.

The Theorem is then proved if we use the following fact:

Theorem (Helly's theorem). Let F_n be a sequence of distribution functions and let ϕ_n be the corresponding sequence of their characteristic functions. Suppose $\lim_{n\to\infty} \phi_n(t) = \phi(t)$ exists for all $t \in \mathbb{R}$ and $\phi(t)$ is continuous at $t = 0$. Then $\phi(t)$ is the characteristic function of a cdf F and $\lim_{n\to\infty} F_n(x) = F(x)$ at every x where $F(x)$ is continuous.