Stat205A: Probability Theory (Fall 2002) **Lecture: 7**

Almost sure limits for sums of independent random variables.

Lecturer: James W. Pitman *Scribe: Animesh Kumar* `animesh@eecs.berkeley.edu`

We first note a few general facts about the various types of convergence we know,

1. If $X_n \to X$ a.s. then $X_n \to X$ in P.
2. If $X_n \to X$ in P then there exists a fixed increasing subsequence n_k such that $X_{n_k} \to X$ a.s..
3. $X_n \to X$ in P iff for every subsequence n_k there exists a further subsequebce n'_k so that $X_{n'_k} \to X$ a.s..

Proof of 2 and 3 are in the textbook [1]. We first begin with a technique which uses the information about almost sure convergence of subsequence of a sequence of random variables, and then somehow getting control over a maximum. Let us start with the technique.

One can prove $X_n \to X$ a.s. by first showing $X_{n_k} \to X$ a.s. for some n_k (we choose n_k) and then getting control over

$$M_k = \max_{n_k \leq m < n_{k+1}} |X_m - X_{n_k}|$$

In particular we must be able to show that $M_k \to 0$ a.s. because if $\omega \in \Omega$ is such that both $X_{n_k}(\omega) \to 0$ and $M_k(\omega) \to 0$ then we get (using triangular inequality and max greater than the elements of set over which maximum is taken)

$$X_m(\omega) \to X(\omega)$$

for all ω in the set of significant probability. To illustrate how to use the technique and how easy it is to use, we start with the example of SLLN with a second moment condition,

Theorem 7.1. If $X, X_1, X_2, ...$ are IID[1] random variables with $E(X) = 0, \quad E(X^2) < \infty$, and $S_n := X_1 + X_2 + ... + X_n$, then,

$$\frac{S_n}{n} \to 0 \quad \text{a.s.} \tag{1}$$

Proof. First we find a subsequence converging almost surely to the mean. For that we use two tools,

- Convergence in Probability or P.
- Borel-Cantelli lemma.

WLOG[2] we can assume that $E(X) = 0$. From Chebychev's inequality we get,

$$P\left(\left|\frac{S_n}{n}\right| > \epsilon\right) < \frac{E(X^2)}{n\epsilon^2}$$

[1] Independent and Identically Distributed.
[2] Without Loss of Generality.

This means that $\frac{S_n}{n} \to 0 \quad$ in P. Notice that $\sum_k \frac{1}{k^2}$ converges to a finite value, therefore for the subsequence $n_k = k^2$ we get using Borel-Cantelli lemma

$$P\left(\left|\frac{S_{n^2}}{n^2}\right| > \epsilon \quad i.o.\right) = 0$$

which means that $\frac{S_{n^2}}{n^2} \to 0 \quad$ a.s..

Now let us try to control M_k as defined above. For convenience we define

$$D_n := \max_{n^2 \le k < (n+1)^2} |S_k - S_{n^2}|$$

for $n^2 \le k < (n+1)^2$, we have $|S_k| \le |S_{n^2}| + D_n$ and $\frac{1}{k} \le \frac{1}{n^2}$. So we have the following inequality,

$$\left|\frac{S_k}{k}\right| \le \left|\frac{S_{n^2}}{n^2}\right| + \frac{D_n}{n^2}$$

Finally, using definition of M_k, we get the following,

$$\begin{aligned} M_k &\le \max_{n^2 \le k < (n+1)^2} \left|\frac{S_k}{k}\right| + \left|\frac{S_{n^2}}{n^2}\right| \\ &\le 2\left|\frac{S_{n^2}}{n^2}\right| + \frac{D_n}{n^2} \end{aligned}$$

So all we need to prove is that $\frac{D_n}{n^2} \to 0 \quad$ a.s.. Let us define a new quantity $T_m = S_{n^2+m} - S_{n^2}$. Therefore,

$$\begin{aligned} D_n^2 &= \max_{1 \le m \le 2n} T_m^2 \\ &\le \sum_{m=1}^{2n} T_m^2 \end{aligned}$$

Taking expectations on both sides, we get that,

$$\begin{aligned} E(D_n^2) &\le \sum_{m=1}^{2n} m\sigma^2 = n(2n+1)\sigma^2 \\ &\le 4n^2\sigma^2 \end{aligned}$$

where $E(X^2) = \sigma^2$. Hence we get that

$$\begin{aligned} P\left(\left|\frac{D_n}{n^2}\right| > \epsilon\right) &\le \frac{E\left(\frac{D_n}{n^2}\right)}{\epsilon^2} \\ &\le \frac{4\sigma^2}{k^2\epsilon^2} \end{aligned}$$

Again in conjunction with Borel-Cantelli lemma and

$$\sum_n P\left(\left|\frac{D_n}{n^2}\right| > \epsilon\right) < \infty$$

we get that $\frac{D_n}{n^2} \to 0 \quad$ a.s.. Which completes the proof. □

Now we proceed to the **sums of independent random variables** which may not be identically distributed. We first start with Kronecker's lemma (see text for proof)

Lemma 7.2 (Kronecker). Let $\{x_n\}$ be a sequence of reals and $S_n = x_1 + x_2 + ... + x_n, \quad 0 \leq a_n \uparrow \infty$, then the lemma states that if $\sum_n \frac{x_n}{a_n}$ converges to a finite limit then $\frac{s_n}{a_n} \to 0$.

Now we start looking at sums like $\sum_{n=1}^{\infty} X_n$ where $\{X_n\}$ is a sequence of independent random variables. The first key fact that we will prove is Kolmogorov's Zero-One law, which will be written henceforth as K'rov 0-1 law for brevity. First key fact that we prove is that

$$P\left(\sum_{n=1}^{\infty} X_n \text{ converges}\right) = \left\{ \begin{array}{c} 1 \\ 0 \end{array} \right\} \tag{2}$$

which says that the set of ω for which the sum converges is either of probability 0 or 1.

Definition. Given a sequence of random variables $\{X_n\}$, the tail sigma field is defined as

$$\mathcal{T} := \bigcap_n \sigma(X_n, X_{n+1}, ...)$$

With this definition in mind, the K'rov 0-1 law says that if $X_i's$ are independent then for any $T \in \mathcal{T}$, we have $P(T) = 0$ or 1.

Proof. We start the proof of 0-1 law now. The trick is to show that any such T is independent of itself which sounds pretty bizarre but it turns out to be true. With that aid one can show that $P(T) = P(T \cap T) = P^2(T)$ and hence the result will follow.

Take $T \in \mathcal{T}$ and let $F_n \in \sigma(X_1, X_2, ..., X_n)$, and $T_n \in \sigma(X_{n+1}, X_{n+2}, ...)$. Then F_n and T_n are independent. So if $T \in \mathcal{T}$ then $T \in \sigma(X_{n+1}, X_{n+1}, ...)$, and hence

$$P(T \cap F_n) = P(T)P(F_n)$$

for all $F_n \in \sigma(X_1, X_2, ..., X_n)$. Now consider the set $\mathcal{F} = \{F \in \sigma(X_1, X_2, X_3, ...) : P(T \cap F) = P(T)P(F)\}$. This can be verified to be a λ system (use MCT for increasing sequences) and the λ system contains all sets like $\bigcup_n \sigma(X_1, X_2, X_3, ..., X_n)$ which is a field (and hence a π system). Therefore, using the $\pi - \lambda$ theorem, we get this property to be true for $\sigma(X_1, X_2, X_3, ...)$ which completes the proof. □

Finally we arrive at Kolmogorov's inequality. We formally state it as follows,

Theorem 7.3 (Kolmogorov's Inequality). Let $X_1, X_2, ...$ be independent with $E(X_i) = 0$ and $\sigma_i^2 = E(X_i^2) < \infty$, and define $S_k = X_1 + X_2 + ... + X_k$, then the inequality states that

$$P\left(\max_{1 \leq k \leq n} |S_k| \geq \epsilon\right) \leq \frac{E(S_n^2)}{\epsilon^2} \tag{3}$$

Proof. Decompose the event according to when we escape from the $\pm\epsilon$ strip. Let

$$A_k = \{|S_m| < \epsilon for 1 \leq m < k; |S_k| \geq \epsilon\}$$

Simply speaking or in words, A_k is the event of first escape out of ϵ strip and that too at the k-th step. Also notice that all these events are disjoint, and as a final remark we have $\bigcup_{k=1}^{n} A_k = (\max_{1 \leq k \leq n} |S_k| \geq \epsilon)$. Finally we start for a chain of inequality, since we have all the pieces ready,

$$E(S_n^2) \geq E\left(S_n^2 1\left(\bigcup_{k=1}^{n} A_k\right)\right) = \sum_{k=1}^{n} E\left(S_n^2 1_{A_k}\right)$$

We can split $S_n^2 = S_k^2 + (S_n - S_k)^2 + 2S_k(S_n - S_k)$, and write the following,

$$\begin{aligned} E\left(S_n^2 1_{A_k}\right) &= E\left(S_k^2 1_{A_k}\right) + E\left((S_n - S_k)^2 1_{A_k}\right) + E\left(2(S_n - S_k)S_k 1_{A_k}\right) \\ &\geq \epsilon^2 P(A_k) \end{aligned}$$

where the first term is larger than ϵ^2 second term is always positive, and the third term is expectation of product of two independent random variables (and hence product of expectation which is zero).

Finally put into the summation to get,

$$E(S_n^2) \geq \sum_{k=1}^{n} P(A_k)\epsilon^2 = P\left(\max_{1\leq k\leq n} |S_k| \geq \epsilon\right)\epsilon^2$$

which easily leads to the result. □

Hence the proof is complete.

References

[1] R. Durrett. *Probability: theory and examples.* Duxbury Press, Belmont, CA, second edition, 1996.