EE5137 : Stochastic Processes (Spring 2021)
Some Notes on Convergence of Random Variables

Vincent Y. F. Tan

January 23, 2021

In this document, we provide some supplementary material to Lecture 3 of EE5137. You do NOT need to know Sections 1–3. You only need to know Sections 4 and 5.

1 Cramér's Theorem

Let $X_i, i \geq 1$ be i.i.d. random variables. In class we were interested in understanding the probability that $\frac{1}{n}S_n = \frac{1}{n}(X_1 + \ldots + X_n)$ deviates from its mean $\mathbb{E}X$ by more than $\epsilon > 0$. We saw that

$$\Pr(S_n \geq na) \leq \exp(n\mu_X(a)) \tag{1}$$

where

$$\mu_X(a) = \inf_{r>0} \gamma_X(r) - ar \tag{2}$$

and where $\gamma_X(r) = \log g_X(r)$ is the cumulant generating function and $g_X(r) = \mathbb{E}[e^{rX}]$ is the moment generating function. Note that $\mu_X(a) < 0$ if $a > \mathbb{E}X$. Thus, if $a > \mathbb{E}X$, the probability that the sample mean $\frac{1}{n}S_n$ exceeds a decays exponentially fast. Please refer to Gallager's book for a full explanation of this, which is important for you to know.

The upper bound in (1) is, in fact, tight in the following sense.

Theorem 1 (Cramér's Theorem [DZ98]). *If the MGF of X is finite for all $r \in \mathbb{R}$, then*

$$\lim_{n\to\infty} \frac{1}{n} \log \Pr(S_n \geq na) = \mu_X(a) \tag{3}$$

Proof. We first prove the upper bound (1). The event $\{S_n \geq na\}$ is the same as $\{\exp(rS_n) \geq \exp(rna)\}$ for any $r > 0$. Thus, by Markov's inequality,

$$\Pr(S_n \geq na) = \Pr(\exp(rS_n) \geq \exp(rna)) \tag{4}$$

$$\leq \frac{\mathbb{E}[\exp(rS_n)]}{\exp(rna)}. \tag{5}$$

At the same time, by the independence of the X_i's,

$$\mathbb{E}[\exp(rS_n)] = \mathbb{E}[\exp(r(X_1 + \ldots + X_n))] = \prod_{i=1}^{n} \mathbb{E}[\exp(rX_i)]. \tag{6}$$

Combining (5) and (6), we have

$$\Pr(S_n \geq na) \leq \prod_{i=1}^{n} \exp(-ar)\mathbb{E}[\exp(rX_i)] = \exp(-nar + n \log \mathbb{E}[\exp(rX_i)]) = \exp(n(-ar + \gamma_X(r))). \tag{7}$$

Since $r > 0$ is a free parameter, we can minimize $-ar + \gamma_X(r)$ in the exponent on the right-hand-side over $r > 0$ to yield the desired upper bound.

We now prove the lower bound for (3), which is significantly more challenging. Let r^* achieve the infimum in (2). Consider

$$\begin{aligned}
\Pr(S_n \geq na) &= \int_{\sum_{i=1}^n x_i \geq na} \left(\prod_{i=1}^n p(x_i) \right) \mathrm{d}x_1 \ldots \mathrm{d}x_n && (8)\\
&= \int_{na \leq \sum_{i=1}^n x_i \leq n(a+\delta)} \left(\prod_{i=1}^n p(x_i) \right) \mathrm{d}x_1 \ldots \mathrm{d}x_n && (9)\\
&= \frac{g_X(r^*)^n}{e^{nr^*(a+\delta)}} \int_{na \leq \sum_{i=1}^n x_i \leq n(a+\delta)} \frac{e^{nr^*(a+\delta)}}{g_X(r^*)^n} \left(\prod_{i=1}^n p(x_i) \right) \mathrm{d}x_1 \ldots \mathrm{d}x_n && (10)\\
&\geq \frac{g_X(r^*)^n}{e^{nr^*(a+\delta)}} \int_{na \leq \sum_{i=1}^n x_i \leq n(a+\delta)} \left(\prod_{i=1}^n \frac{e^{r^* x_i} p(x_i)}{g_X(r^*)} \right) \mathrm{d}x_1 \ldots \mathrm{d}x_n && (11)\\
&= \frac{g_X(r^*)^n}{e^{nr^*(a+\delta)}} \int_{na \leq \sum_{i=1}^n x_i \leq n(a+\delta)} \left(\prod_{i=1}^n q(x_i) \right) \mathrm{d}x_1 \ldots \mathrm{d}x_n && (12)
\end{aligned}$$

where

$$q(y) = \frac{e^{r^* y} p(y)}{g_X(r^*)}. \tag{13}$$

Note that $q(y)$ is indeed a distribution since $\int q(y)\, \mathrm{d}y = 1$. The MGF of Y is

$$g_Y(r) = \mathbb{E}_q[e^{rY}] = \frac{g_X(r + r^*)}{g_X(r^*)}. \tag{14}$$

As a result, its espectation is

$$\mathbb{E}[Y] = \left. \frac{\mathrm{d}g_Y(r)}{\mathrm{d}r} \right|_{r=0} = \frac{g_X'(r^*)}{g_X(r^*)}. \tag{15}$$

From the assumption that r^* achieves the infimum in (2), we have

$$\frac{\mathrm{d}}{\mathrm{d}r} \log g_X(r) - ar = 0 \quad \text{at} \quad r = r^*. \tag{16}$$

From this, we obtain that $a = \frac{g_X'(r^*)}{g_X(r^*)}$. Therefore $\mathbb{E}[Y] = a$. Now, from (12) we have

$$\Pr(S_n \geq na) \geq e^{n \log g_X(r^*) - nr^*(a+\delta)} \Pr\left(\frac{1}{n} \sum_{i=1}^n Y_i \in [a, a+\delta] \right). \tag{17}$$

By the central limit theorem, it is easy to see that the final probability is $\geq \frac{1}{4}$ for n large enough (why?) so

$$\Pr(S_n \geq na) \geq \frac{1}{4} \exp(n\mu_X(a) - nr^*\delta), \tag{18}$$

where recall that $\mu_X(a) = \log g_X(r^*) - ar^*$. In other words,

$$\liminf_{n \to \infty} \frac{1}{n} \log \Pr(S_n \geq na) \geq \mu_X(a) - r^*\delta. \tag{19}$$

Letting $\delta \downarrow 0$ establishes the lower bound to (3) □

2 Rough Proof of Central Limit Theorem

We use *Curtiss' Theorem* [Cur42] which says that if for a sequence of random variables X_n

$$\mathbb{E}[e^{rX_n}] \to f(r), \quad \forall\, r \in \mathbb{R} \tag{20}$$

then X_n converges in distribution to a random variable X with moment generating function (MGF) $f(r)$. In particular, if $f(r) = e^{r^2/2}$, then X_n converges in distribution to a standard Gaussian. Recall that *convergence in distribution* of a sequence of random variables X_n to a limiting random variable X means that

$$\lim_{n\to\infty} \Pr(X_n \le a) = \Pr(X \le a) \tag{21}$$

at all points of continuity of $a \mapsto F_X(a) := \Pr(X \le a)$.

For the sake of simplicity, assume that X_i is a sequence of i.i.d. random variables with zero mean and variance 1. Let the MGF of X_i be $g_X(\cdot)$. Consider the sequence of random variables $Z_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$. Then we claim that Z_n converges in distribution to a standard Gaussian.

Consider,

$$\mathbb{E}\left[e^{rZ_n}\right] = \mathbb{E}\left[e^{r\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i}\right] = \prod_{i=1}^n \mathbb{E}\left[e^{\frac{r}{\sqrt{n}}X_i}\right] = g_X\left(\frac{r}{\sqrt{n}}\right)^n. \tag{22}$$

Now by Taylor expanding $g_X(r/\sqrt{n})$ around zero, we obtain

$$g_X\left(\frac{r}{\sqrt{n}}\right) = g_X(0) + g_X'(0)\frac{r}{\sqrt{n}} + \frac{g_X''(0)}{2}\frac{r^2}{n} + O\left(\frac{1}{n^{3/2}}\right). \tag{23}$$

But note that $g_X(0) = 1$, $g_X'(0) = 0$ (the mean is zero), $g_X''(0) = 1$ (the variance is 1) so

$$g_X\left(\frac{r}{\sqrt{n}}\right) = 1 + \frac{r^2}{2n} + O\left(\frac{1}{n^{3/2}}\right) \tag{24}$$

Thus,

$$\mathbb{E}[e^{rZ_n}] = \left(1 + \frac{r^2}{2n} + O\left(\frac{1}{n^{3/2}}\right)\right)^n \to e^{r^2/2}, \quad \text{as } n \to \infty, \tag{25}$$

and from Curtiss' theorem, we conclude that Z_n converges in distribution to a standard Gaussian.

3 More on Convergence with Probability One (or Almost Sure Convergence)

Let $(\Omega, \mathfrak{F})$ be a probability space, i.e., Ω is the sample space and $\mathfrak{F}$ is a σ-algebra (informally, set of all admissible events. A random variable X is a function from Ω to $\mathbb{R}$ satisfying the condition that $\{\omega \in \Omega : X(\omega) \le x\} \in \mathfrak{F}$ for all $x \in \mathbb{R}$.

Consider a sequence of random variables $\{X_n\}_{n=1}^\infty$ defined on the given probability space $(\Omega, \mathfrak{F})$. We say that X_n converges to X *almost surely* (or *with probability one*) if

$$\Pr\left(\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1. \tag{26}$$

It is often easier to express the condition in (26) in terms of its complement. Given a fixed element in the sample space $\omega \in \Omega$, $X_n(\omega)$is nothing more than a sequence of real numbers. Then $X_n(\omega)$ fails to converge to $X(\omega)$ if there exists an $\epsilon > 0$ such that $|X_n(\omega) - X(\omega)| \ge \epsilon$ holds infinitely often in n. (Why? Think of what it means for $\{a_n\}_{n=1}^{\infty} \subset \mathbb{R}$ to converge to a limit a.) Motivated by this, let us define the family of events

$$B_n^{\epsilon} := \{\omega \in \Omega : |X_n(\omega) - X(\omega)| \ge \epsilon\}. \tag{27}$$

That X_n does not converge to X almost surely means that there exists an $\epsilon > 0$ such that

$$\Pr(B_n^{\epsilon} \text{ i.o.}) = \Pr\left(\limsup_{n\to\infty} B_n^{\epsilon}\right) = \Pr\left(\bigcap_{n\ge 1} \bigcup_{j\ge n} B_j^{\epsilon}\right) > 0. \tag{28}$$

Thus, that X_n converges to X almost surely means that there exists $\epsilon > 0$ such that

$$\Pr\left(\bigcap_{n\ge 1} \bigcup_{j\ge n} B_j^{\epsilon}\right) = 0. \tag{29}$$

This is equivalent to: For all $\epsilon > 0$,

$$\Pr\left(\bigcup_{n\ge 1} \bigcap_{j\ge n} \left(B_j^{\epsilon}\right)^c\right) = 1. \tag{30}$$

This is equivalent to: For all $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr\big(\{\omega \in \Omega : |X_j(\omega) - X(\omega)| \le \epsilon \text{ for all } j \ge n\}\big) = 1, \tag{31}$$

which was what was stated in class.

4 Intuitive Difference Between Convergence with Probability One and Convergence in Probability

Convergence with probability one was discussed in Section 3 above. There are two equivalent definitions in (47) and (31). Recall that $\{X_n\}_{n=1}^{\infty}$ *converges in probability* to a limiting random variable X if for every $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr\big(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \le \epsilon\big) = 1. \tag{32}$$

In contrast to (47), the big difference is that the limit is *outside the probability* in (32) compared to *inside the probability* in (47). This is a big difference! It is well-known that if $\{X_n\}_{n=1}^{\infty}$ converges to X with probability one then $\{X_n\}_{n=1}^{\infty}$ converges to X in probability.

The following analogy captures some aspects of the difference: Suppose we have a committee with 100 members. Imagine there is one meeting of the committee everyday throughout the year, and that we want to know about attendance on the committee. Each member of the committee is a single $\omega \in \Omega$ and so there are 100 different ω's.

Convergence almost surely is asking whether *almost all members had perfect attendance*. Convergence in probability is asking whether *all meetings were almost full*.

If almost all members have perfect attendance, then each meeting must be almost full (convergence almost surely implies convergence in probability). But if all meetings were nearly full, it isn't necessary that any member has perfect attendance (e.g., each member might have been absent on one distinct day). Convergence in probability does not imply convergence almost surely.

Convergence almost surely means that $X_n(\omega)$ gets close to $X(\omega)$ as n increases for almost all elements ω of the sample space Ω. It's a statement about convergence of $X_n(\omega)$ for many (in fact, almost all) individual $\omega \in \Omega$.

Convergence in probability means that the set of elements ω of the sample space Ω for which $X_n(\omega)$ is close to $X(\omega)$ has probability approaching 1. It's a statement about the *size of the set* of ω which satisfy the closeness property, but not about any one ω.

5 Example of Convergence of Random Variables

Consider the sequence of independent random variables

$$X_n = \begin{cases} 0 & \text{w.p.} \quad 1 - \frac{1}{n} \\ 1 & \text{w.p.} \quad \frac{1}{n} \end{cases} \tag{33}$$

Does it converge in mean-square, probability and with probability 1 (almost surely)?

- Yes, it converges to the deterministic random variable 0 in mean square. Consider

$$\mathbb{E}[(X_n - 0)^2] = \frac{1}{n} \to 0 \tag{34}$$

So it converges to 0 in m.s.

- Yes, it converges to the deterministic random variable 0 in probability. Consider

$$\Pr(|X_n - 0| > \epsilon) = \frac{1}{n} \to 0 \tag{35}$$

So it converges to 0 in probability.

- No, it does not converge to anything with probability 1 (or almost surely). Recall that convergence with probability 1 to a random variable X is equivalent to

$$\Pr\left(\lim_{n\to\infty} X_n = X\right) = 1. \tag{36}$$

Equivalently, $X_n \to X$ with probability 1 if for all $\epsilon > 0$,

$$\lim_{m\to\infty} \Pr(|X_n - X| < \epsilon, \forall\, n \geq m) = 1. \tag{37}$$

See Section 3. Let's check whether this condition is satisfied for $\{X_n\}_{n=1}^{\infty}$. Consider

$$\begin{aligned}
\Pr(|X_n - 0| < \epsilon, \forall\, n \geq m) &= \lim_{n\to\infty} \prod_{i=m}^{n} \left(1 - \frac{1}{i}\right) && (38) \\
&= \lim_{n\to\infty} \prod_{i=m}^{n} \frac{i-1}{i} && (39) \\
&= \lim_{n\to\infty} \frac{m-1}{m} \cdot \frac{m}{m+1} \cdot \ldots \frac{n-1}{n} && (40) \\
&= \lim_{n\to\infty} \frac{m-1}{n} = 0. && (41)
\end{aligned}$$

So clearly $\{X_n\}_{n=1}^{\infty}$ does not converge to 0 almost surely.

5.1 Almost Sure Convergence Does Not Imply L_1 Convergence

In this note, we show that almost sure convergence does not mean convergence in L_1.

Let $X_1, X_2, \ldots, X_n$ be i.i.d. $\{0,1\}$-valued random variables with $\Pr(X_1 = 1) = 3/4$. Let

$$Z_n := 2^n \prod_{i=1}^{n} X_i. \tag{42}$$

- (Almost sure convergence to 0) Let us fix $0 < \epsilon < 2^m$. Then consider the probability

$$\begin{aligned}
\Pr(|Z_n - 0| < \epsilon,\ \forall\ n \geq m) &= \Pr(X_n = 0,\ \text{for some } n \leq m) && (43)\\
&= 1 - \Pr(X_n = 1, \forall\ n \leq m) && (44)\\
&= 1 - \prod_{n=1}^{m} \Pr(X_n = 1) && (45)\\
&= 1 - (3/4)^m \to 1, \qquad \text{as } m \to \infty && (46)
\end{aligned}$$

Thus, we have that

$$\lim_{m\to\infty} \Pr(|Z_n - 0| < \epsilon,\ \forall\ n \geq m) = 1 \tag{47}$$

Thus, Z_n converges to the deterministic random variable $Z = 0$ almost surely. Recall that almost sure convergence of Z_n to Z is defined as

$$\Pr\left(\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \{|Z_n - Z| < \epsilon\} \right) = 1, \tag{48}$$

which is fulfilled with $Z = 0$ in view of (47). We have used continuity of probability (or measure) here (this means that if $A_1 \subset A_2 \subset A_3 \subset \ldots$, then $\Pr(\cup_{i=1}^{\infty} A_i) = \lim_{i\to\infty} \Pr(A_i)$.)

- (Divergence in L_1) Now consider

$$\begin{aligned}
\mathbb{E}[|Z_n|] &= \mathbb{E}[Z_n] && (49)\\
&= \mathbb{E}\left[2^n \prod_{i=1}^{n} X_i \right] && (50)\\
&= 2^n \prod_{i=1}^{n} \mathbb{E}[X_i] && (51)\\
&= 2^n (3/4)^n && (52)\\
&= (6/4)^n \to \infty, \qquad \text{as } n \to \infty && (53)
\end{aligned}$$

where (51) follows from independence of the process $\{X_i\}$ and (52) follows from the fact that $\mathbb{E}[X_i] = 3/4$ for all i. So the sequence of rvs $\{Z_n\}$ is not L_1-integrable, i.e., the expectation of the absolute value does not converge.

References

[Cur42] J. H. Curtiss. A note on the theory of moment generating functions. *Ann. Math. Statist.*, 13(4):430–433, 1942.

[DZ98] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications.* Springer, 2nd edition, 1998.