Igor Kortchemski
Condensation phenomena in random trees
ETH, Spring 2024

Chapter 1: Large deviations in Cramer's regime

Outline
1) First observations
2) Cramer's theorem
3) Behavior under the conditional law
4) Local estimates and the local central limit theorem

In probability theory, many theorems concern "typical events", which have probability 1 or tending to one. Large deviations concern "atypical events" whose probability tends to 0. Typical questions then are:

- How fast is the convergence (rate of decay)?
- Given this atypical event, what are typical events of the system under the conditional law?

For example, if you release at ETH a capybara who wanders at random, if it goes to Paris and then back to ETH, what will be it's typical path?

Here we focus on the case of random walks. Let $(X_n)_{n\geq 1}$ be iid $\mathbb{R}$-valued r.v. with $\mathbb{E}[X_1^2]<\infty$.
Set $S_n = X_1 + \cdots + X_n$. We know that $\frac{S_n}{n} \xrightarrow[n\to\infty]{a.s.} \mathbb{E}[X_1]$ and $\frac{S_n - n\,\mathbb{E}[X_1]}{\sqrt{n\,\mathrm{Var}(X_1)}} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1)$

$\rightsquigarrow$ "Typically the deviations of S_n around $n\,\mathbb{E}[X_1]$ are of order $\sqrt{n}$"

But for fixed $a \in \mathbb{R}$, what can we say about $\mathbb{P}(S_n > an)$, $\mathbb{P}(S_n = an)$?
local probability

What can we say about $(S_1, \ldots, S_n)$ conditionally given $\{S_n > an\}$ or $\{S_n = an\}$?
(when these events have >0 probability).

1) First observations

We first focus on $\mathbb{P}(S_n > an)$. Clearly, by the central limit theorem, writing $\mathbb{P}(S_n > an) = \mathbb{P}\left(\frac{S_n - \mathbb{E}[X_1]n}{\sqrt{n}} > (a-\mathbb{E}[X_1])\sqrt{n}\right)$
we have $\mathbb{P}(S_n > an) \longrightarrow 1$ if $a < \mathbb{E}[X_1]$
$\longrightarrow \frac{1}{2}$ if $a = \mathbb{E}[X_1]$.

Now assume that $a > \mathbb{E}[X_1]$

- Assume that X_1 is $\mathcal{N}(0,1)$. Then $S_n \overset{(d)}{=} \mathcal{N}(0,n)$ and for $a>0$:

$$\mathbb{P}(S_n > na) = \frac{1}{\sqrt{2\pi}} \int_{a\sqrt{n}}^{\infty} e^{-y^2/2} dy \sim \frac{1}{a\sqrt{2\pi n}} e^{-na^2/2}, \text{ so } \mathbb{P}(S_n > na) = e^{-n\frac{a^2}{2} + o(n)}.$$

A similar computation shows that $\forall \varepsilon > 0$, $\mathbb{P}(S_n \in (an, an + n^{\varepsilon})) \underset{n\to\infty}{\sim} \mathbb{P}(S_n > an)$,

which shows that for $\mathbb{P}(\cdot \mid S_n > an)$, S_n is of order an

- If $\mathbb{P}(X_1 > a) \in (0,1)$, then $\mathbb{P}(S_n > an) \geq \mathbb{P}(X_1 > a, \dots, X_n > a) = \mathbb{P}(X_1 > a)^n$,

so $\mathbb{P}(S_n > an)$ cannot decay faster than exponentially.

- Assume that $-X_1 \overset{(d)}{=} X_1$, that X_1 has a density and $\mathbb{P}(X_1 > x) \geq c x^{-\alpha}$ for $x \geq 1$ for some $c, \alpha > 0$. Then

$\mathbb{P}(S_n > an) \geq \mathbb{P}(X_n > an)\, \mathbb{P}(S_{n-1} > 0) \geq c(an)^{-\alpha} \cdot \frac{1}{2}$

Thus $\mathbb{P}(S_n > an)$ does not decay exponentially fast.

If $\mathbb{P}(X_1 > x) \sim \exp(-\sqrt{x})$ the same holds (and X_1 has finite moments of all orders)

Conclusion: it is not always true that $\mathbb{P}(S_n > an)$ decays exponentially fast

2) Cramér's theorem

Assumptions X_1 is not constant and $\forall t \in \mathbb{R}$, $M(t) = \mathbb{E}[e^{tX_1}] < \infty$. $(*)$ [Cramer condition]

We shall see that $\mathbb{P}(S_n > an)$ decays exponentially fast for $\mathbb{E}[X_1] < a < \sup \text{Support } X_1$ (Support X_1 is the smallest closed set $F \subset \mathbb{R}$ such that $\mathbb{P}(X_1 \in F) = 1$, in particular $\mathbb{P}(S_n > an) = 0$ for $a > \sup \text{Support } X_1$).

The main idea/tool is "exponential tilting": by $(*)$ for every $\theta \in \mathbb{R}$ we can consider a r.v X_1^θ with law given by $\mathbb{E}[f(X_1^\theta)] = \mathbb{E}\left[f(X_1) \frac{e^{\theta X_1}}{M(\theta)}\right]$ for $f \geq 0$ measurable

Observe that for all $\theta > 0$, X_1^θ satisfies Cramer's condition

Set $(X_i^\theta)_{i\geq 1}$ be iid and set $S_n^\theta = X_1^\theta + \dots + X_n^\theta$. By definition one has:

Proposition For $f \geq 0$ measurable,

$$\mathbb{E}[f(X_1^\theta, \dots, X_n^\theta)] = \frac{\mathbb{E}[f(X_1, \dots, X_n)\, e^{\theta S_n}]}{M(\theta)^n} \text{ and } \mathbb{E}[f(X_1, \dots, X_n)] = M(\theta)^n\, \mathbb{E}[f(X_1^\theta, \dots, X_n^\theta) e^{-\theta S_n^\theta}]$$

In particular, $\mathbb{P}(S_n \geq an) = M(\theta)^n \, \mathbb{E}[\mathbb{1}_{S_n^\theta \geq an} e^{-\theta S_n^\theta}]$

Theorem (Cramer) Fix $a \in (\mathbb{E}[X_1], \sup \text{Support}(X_1))$.

There exists $\theta(a) > 0$ such that $\mathbb{E}[X_1^{\theta(a)}] = a$, and $\mathbb{P}(S_n \geq an) = e^{-I(a)n + o(n)}$ with $I(a) = \theta(a)a - \ln M(\theta(a))$

Intuition: the optimal strategy to achieve $S_n \geq an$ is that "each X_j contributes a little bit" and "tilts itself a little bit" in order to favor taking larger values. $e^{-I(a)}$ represents in some sense the "cost" of this tilt for every j, so $e^{-I(a)n}$ is the total cost

We first check the existence of $\theta(a)$. For $t \in \mathbb{R}$ set $L(t) = \ln M(t) = \ln \mathbb{E}[e^{tX_1}]$.

Observe that $L'(t) = \frac{M'(t)}{M(t)} = \frac{\mathbb{E}[X_1 e^{tX_1}]}{M(t)}$, so $\mathbb{E}[X_1^\theta] = \frac{\mathbb{E}[X_1 e^{\theta X_1}]}{M(\theta)} = L'(\theta)$

End of lecture 1

Lemma L is convex

(⚠ the log of a convex function is not always convex, for example x^2)

Proof take $t_1, t_2 \in \mathbb{R}$ and $p \in [0,1]$. Using Hölders inequality:

$L(pt_1 + (1-p)t_2) = \ln \mathbb{E}[(e^{t_1 X_1})^p (e^{t_2 X_1})^{1-p}] \leq \ln \mathbb{E}[e^{t_1 X_1}]^p \cdot \mathbb{E}[e^{t_2 X_2}]^{1-p} = pL(t_1) + (1-p)L(t_2)$

∼

Proof of existence of $\theta(a)$: Consider $g(\theta) = a\theta - L(\theta)$. We have $g(0) = 0$, g is continuous and $\lim_{\theta \to \infty} g < 0$.

Indeed, take $a < b < \sup \text{Support} X_1$. Then

$M(\theta) \geq \mathbb{E}[e^{\theta X_1} \mathbb{1}_{X_1 \geq b}] \geq e^{\theta b} \underbrace{\mathbb{P}(X_1 \geq b)}_{>0}$ so $L(\theta) \geq \theta b + \ln \mathbb{P}(\theta \geq b)$

Thus $g(\theta) \leq \theta(\underbrace{a - b + \frac{\ln \mathbb{P}(X_1 \geq b)}{\theta}}_{<0 \text{ for } \theta \text{ sufficiently large}})$

Thus, since g is concave, letting $\theta(a)$ be the point where g reaches it supremum, we get $g'(\theta(a)) = 0$, which gives $L'(\theta(a)) = a$.

∼

Remark The previous proof shows that $I(a) = g(\theta(a)) = \sup_{\theta \geq 0} (a\theta - L(\theta))$, which in particular implies that I is increasing.

End of the proof of the theorem

Upper bound: Just write $\mathbb{P}(S_n > an) = \mathbb{P}(e^{\theta(a) S_n} > e^{\theta(a) a n}) \leq e^{-\theta(a) a n} M(\theta(a))^n = e^{-n(\theta(a) a - \ln M(\theta(a)))}$.

Lower bound: Using the proposition, write (with $\theta = \theta(a)$ to simplify notation)

$$\mathbb{P}(S_n > an) = M(\theta)^n \, \mathbb{E}[\mathbb{1}_{S_n^\theta > an} \, e^{-\theta S_n^\theta}]$$
$$\geq M(\theta)^n \, \mathbb{E}[\mathbb{1}_{an < S_n^\theta < an + \sqrt{n}} \, e^{-\theta S_n^\theta}]$$
$$\geq M(\theta)^n \, e^{-\theta a n - \theta \sqrt{n}} \, \mathbb{P}(an < S_n^\theta < an + \sqrt{n})$$

But $\mathbb{E}[X_1^\theta] = a$ and $\mathbb{E}[(X_1^\theta)^2] < \infty$, so $\mathbb{P}(an < S_n^\theta < an + \sqrt{n})$ converges to a >0 number as $n \to \infty$.

Thus $M(\theta)^n e^{-\theta a n - \theta\sqrt{n}} \, \mathbb{P}(an < S_n^\theta < an + \sqrt{n}) = e^{-nI(a) + o(n)}$.

~

Examples:
- When $X_1 \sim N(0,1)$, $I(a) = \frac{a^2}{2}$ for $a \geq 0$
- When $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$, $I(a) = \frac{1+a}{2} \ln(1+a) + \frac{1-a}{2} \ln(1-a)$ for $0 < a < 1$.

Remarks:
- More generally, the study of $\mathbb{P}(S_n > x_n)$ has attracted a lot of interest.
- More generally, a family $(\mu_\varepsilon)_{\varepsilon > 0}$ of probability measures on a metric space E is said to satisfy a large deviation principle with rate function I if for every Borel set A,

$$-\inf_{x \in \mathring{A}} I(x) \leq \liminf_{\varepsilon \to 0} \varepsilon \, \mu_\varepsilon(A) \leq \limsup_{\varepsilon \to 0} \varepsilon \mu_\varepsilon(A) \leq -\inf_{x \in \bar{A}} I(x)$$

(See the book of Dembo & Zeitouni or of Klenke, chapter 23)

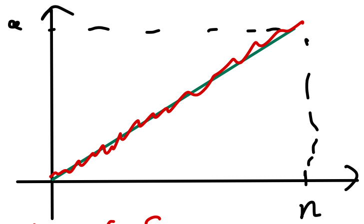
3) Behavior under the conditional law

(also called "Gibbs conditioning principle" in physics)

We keep the assumption X_1 not constant, $\forall t \in \mathbb{R}$ $\mathbb{E}[e^{tX}] < \infty$ and take $\mathbb{E}[X_1] < a < \sup \text{Support } X_1$.

What is the "typical" behavior of $(S_1, \dots, S_n)$ given $\mathbb{P}(\cdot \mid S_n > na)$?

Roughly speaking, we shall see that under $\mathbb{P}(\cdot \mid S_n > an)$ the typical picture is:



in red: $\left(\frac{S_k}{n} : 0 \le k \le n \right)$

Theorem (Folklore) Under the same assumptions as Cramer's theorem, for every $\varepsilon > 0$, $\mathbb{P}\left(\max_{0 \le k \le n} \left| \frac{S_k - ak}{n} \right| > \varepsilon \,\middle|\, S_n > an \right) \xrightarrow[n\to\infty]{} 0$.

Warm up: let us show that $\mathbb{P}\left(\frac{S_n}{n} > a+\varepsilon \,\middle|\, S_n > an \right) \xrightarrow[n\to\infty]{} 0$. We may assume that $a + \varepsilon < \sup \mathrm{Support}\, X_1$.
This probability is

$$\frac{\mathbb{P}(S_n > (a+\varepsilon)n)}{\mathbb{P}(S_n > an)} = \exp\left(-(I(a+\varepsilon) - I(a))n + o(n) \right)$$

But I is increasing, so $I(a+\varepsilon) - I(a) > 0$ and we get the result.

Lemma Let $(Y_i)_{i \ge 1}$ be iid $\mathbb{R}$-valued r.v. satisfying Cramer's condition. Set $W_0 = 0$, $W_n = Y_1 + \dots + Y_n$.
Fix $\varepsilon > 0$. There exists $c > 0$ such that $\mathbb{P}\left(\max_{0 \le k \le n} \frac{|W_k - \mathbb{E}[Y_1]k|}{n} \geq \varepsilon \right) \leq e^{-cn}$

Proof: Without loss of generality we may assume that $\mathbb{E}[Y_1] = 0$. Take $\lambda > 0$.
$(W_k)_{k \ge 0}$ is a martingale and $e^{\lambda x}$ is convex. Thus $(e^{\lambda W_k})_{k \ge 1}$ is a submartingale, and Doob's maximal inequality gives for every $\lambda > 0$:

$$\mathbb{P}\left(\max_{0 \le k \le n} W_k \geq \varepsilon n \right) = \mathbb{P}\left(\max_{0 \le k \le n} e^{\lambda W_k} \geq e^{\varepsilon \lambda n} \right) \le e^{-\varepsilon \lambda n} \mathbb{E}[e^{\lambda W_n}] = e^{-\lambda n \left(\varepsilon - \frac{\ln \mathbb{E}[e^{\lambda W_1}]}{\lambda} \right)}$$

Since $\frac{\ln \mathbb{E}[e^{\lambda W_1}]}{\lambda} \longrightarrow \frac{d}{d\lambda} \ln \mathbb{E}[e^{\lambda W_1}] \big|_{\lambda = 0} = \mathbb{E}[W_1] = 0$, we can choose $\lambda > 0$ such that $\varepsilon - \frac{\ln \mathbb{E}[e^{\lambda W_1}]}{\lambda} > 0$.
One similarly gets the result for $\mathbb{P}\left(\max_{0 \le k \le n} W_k \le -\varepsilon n \right)$ by considering $-W_n$.

Proof of the theorem Set $\theta = \theta(a)$
Write

$$\mathbb{P}\left(\max_{0 \le k \le n} \left| \frac{S_k - ak}{n} \right| > \varepsilon, S_n > an \right) = M(\theta)^n \, \mathbb{E}\left[\mathbb{1}_{\max_{0 \le k \le n} \left| \frac{S^\theta_k - ak}{n} \right| > \varepsilon, S^\theta_n > an} e^{-\theta S^\theta_n} \right]$$

$$\le e^{-nI(a)} \, \mathbb{P}\left(\max_{0 \le k \le n} \left| \frac{S^\theta_k - ak}{n} \right| > \varepsilon \right)$$

Thus by Cramer's theorem $\mathbb{P}\left(\max_{0 \le k \le n} \left| \frac{S_k - ak}{n} \right| > \varepsilon \,\middle|\, S_n > an \right) = \mathbb{P}\left(\max_{0 \le k \le n} \left| \frac{S^\theta_k - ak}{n} \right| > \varepsilon \right) e^{o(n)}$

and the result follows from the lemma (applied with $Y_1 = X_1^\theta$).

End of lecture 2

4) Local estimates and the local central limit theorem

Our goal is now to study analogous questions when "$S_n > an$" is replaced with "$S_n = an$". Since this event can be empty, we need some additional assumptions.

Definition A real-valued random variable X is called lattice if there exist $b \in \mathbb{R}$ and $h > 0$ with $\mathbb{P}(X \in b + h\mathbb{Z}) = 1$.
The largest such h is called the span of X (we will see below that it is well defined).
If the span is 1 then X is called aperiodic.

Example If $\mathbb{P}(X=1) = \mathbb{P}(X=-1) = \frac{1}{2}$, then X has span 2.

Lemma Let X be a $\mathbb{R}$-valued r.v. Set $\phi(t) = \mathbb{E}[e^{iXt}]$ for $t \in \mathbb{R}$.

(1) X is lattice iff $\exists t \neq 0$ with $|\phi(t)| = 1$

(2) X has span $h > 0$ iff $|\Phi(\frac{2\pi}{h})| = 1$ and $|\phi(t)| < 1$ for $t \in (0, \frac{2\pi}{h})$.

Proof: (1) If $\mathbb{P}(X \in b + h\mathbb{Z}) = 1$, then $\mathbb{P}(\frac{X-b}{h} \in \mathbb{Z})$, so $\mathbb{E}[e^{2\pi \frac{X-b}{h}}] = 1$ and $|\mathbb{E}[e^{\frac{2i\pi X}{h}}]| = |e^{2i\pi b}| = 1$.

Conversely, if $|\phi(t_0)| = 1$ for $t_0 \neq 0$, there is $a \in \mathbb{R}$ such that $\phi(t_0) = e^{ia}$. Then $\mathbb{E}[e^{i(t_0X - a)}] = 1$

Thus $\mathbb{E}[\cos(t_0X - a)] = 1$. Since $\cos \leq 1$, this implies a.s $\cos(t_0X - a) = 1$.

Thus a.s. $t_0X - a \in 2\pi\mathbb{Z}$, so $\mathbb{P}(X \in \frac{a}{t_0} + \frac{2\pi}{t_0}\mathbb{Z}) = 1$

(2) The proof of (1) shows that $\exists b \in \mathbb{R}$ and $h > 0$ s.t $\mathbb{P}(X \in b + h\mathbb{Z}) = 1$ iff $|\phi(\frac{2\pi}{h})| = 1$, which entails the result.

∞

Examples
- If $\mathbb{P}(X = \pm 1) = \frac{1}{2}$, $\phi(t) = \cos t$ and $|\phi(t)| = 1$ for $t \in \pi\mathbb{Z}$
- If $X \sim \mathcal{N}(0,1)$, $\phi(t) = e^{-t^2/2}$ and $|\phi(t)| < 1$ for $t \neq 0$.

Theorem (local central limit theorem) Let $(X_i)_{i\geq 1}$ be iid aperiodic random variables with values in $\mathbb{Z}$. Assume that $\mathbb{E}[X_1^2]<\infty$. Set $m=\mathbb{E}[X_1]$ and $\sigma^2=\mathbb{E}[X_1^2]-\mathbb{E}[X_1]^2$. Assume $\sigma^2>0$. Set $S_n=X_1+\dots+X_n$. Then

$$\sup_{k\in\mathbb{Z}}\left|\sqrt{n}\,\mathbb{P}(S_n=k)-\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2}\left(\frac{k-mn}{\sigma\sqrt{n}}\right)^2\right)\right|\xrightarrow[n\to\infty]{}0$$

We will first explore some consequences and prove this later.

In practice, we often write $\mathbb{P}(S_n=k)=\frac{1}{\sqrt{2\pi\sigma^2 n}}e^{-\frac{1}{2}\left(\frac{k-mn}{\sigma\sqrt{n}}\right)^2}+\frac{\varepsilon(k,n)}{\sqrt{n}}$

with $\sup_{k\in\mathbb{Z}}|\varepsilon(k,n)|\xrightarrow[n\to\infty]{}0$

As a consequence:

Corollary Under the same assumptions, when $\mathbb{E}[X_1]=0$:

(1) $\mathbb{P}(S_n=0)\sim\frac{1}{\sqrt{2\pi\sigma^2}}\cdot\frac{1}{\sqrt{n}}$

(2) $\exists\, C>0,\ \forall n\geq 1,\ \forall k\geq 1,\ \mathbb{P}(S_n=k)\leq\frac{C}{\sqrt{n}}$

(3) If (x_n) is a sequence of integers with $\frac{x_n}{\sigma\sqrt{n}}\xrightarrow[n\to\infty]{}x$, then $\sqrt{\sigma^2 n}\,\mathbb{P}(S_n=x_n)\xrightarrow[n\to\infty]{}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$.

Remark The local central limit theorem (LCLT) implies the central limit theorem (CLT): to simplify assume $\mathbb{E}[X_1]=0$. Then

$$\mathbb{P}\left(a<\frac{S_n}{\sigma\sqrt{n}}<b\right)=\sum_{a\sigma\sqrt{n}<k<b\sigma\sqrt{n}}\mathbb{P}(S_n=k)=\int_{a\sigma\sqrt{n}}^{b\sigma\sqrt{n}}\mathbb{P}(S_n=\lfloor u\rfloor)\,du+o(1)\quad(!)$$

$$=\int_a^b\underbrace{\sigma\sqrt{n}\,\mathbb{P}(S_n=\lfloor u\sigma\sqrt{n}\rfloor)}_{g_n(u)}\,du+o(1)$$

But $\begin{cases} g_n(u)\xrightarrow[n\to\infty]{}\frac{1}{\sqrt{2\pi}}e^{-u^2/2} & \text{for } u\in[a,b] \text{ fixed. (Corollary (3))}\\ \sup_{u\in[a,b]}|g_n(u)|\leq C & \text{(Corollary (2)).}\end{cases}$

We conclude that $\mathbb{P}\left(a<\frac{S_n}{\sigma\sqrt{n}}<b\right)\xrightarrow[n\to\infty]{}\int_a^b\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx$ by dominated convergence

Here we used the very useful trick of writing a sum as an integral: $\sum_{k=0}^{\infty} a_k = \int_0^{\infty} a_{\lfloor x \rfloor} dx$.

Let us now turn to the study of large deviations in a local setting.

Theorem Assume that X_1 satisfies Cramer's condition, and is $\mathbb{Z}$-valued aperiodic. Fix $a \in (\mathbb{E}[X_1], \sup \text{Support} X_1)$.
Let (x_n) be a sequence of integers s.t $x_n = an + o(\sqrt{n})$ (for example $x_n = \lfloor an \rfloor$).
Recall that $\theta(a)$ is chosen such that $L'(\theta(a)) = a$ with $L(t) = \ln \mathbb{E}[e^{tX_1}]$.
Then $\mathbb{P}(S_n = x_n) \underset{n\to\infty}{\sim} M(\theta(a))^n e^{-\theta(a) x_n} \frac{1}{\sqrt{2\pi \operatorname{Var}(X_1^{\theta(a)})n}}$

⚠ During the lecture this theorem was incorrectly stated with x_n st $\frac{x_n}{n} \to a$ (we actually need $x_n = an + o(\sqrt{n})$)

Remark that here we have an asymptotic equivalent (and that $M(\theta(a))^n e^{-\theta(a) an} = e^{-I(a)n}$, but $e^{-\theta(a) an} \underset{n\to\infty}{\not\sim} e^{-\theta(a) x_n}$ in general)

Proof As before, set $\theta = \theta(a)$ and write
$$\mathbb{P}(S_n = x_n) = M(\theta)^n \mathbb{E}[\mathbb{1}_{S_n^\theta = x_n} e^{-\theta S_n^\theta}] = M(\theta)^n e^{-\theta x_n} \mathbb{P}(S_n^\theta = x_n)$$
By the LCLT $\mathbb{P}(S_n^\theta = x_n) \sim \frac{1}{\sqrt{2\pi \operatorname{Var}(X_1^\theta) n}}$

∾

End of lecture 3

We now study the behavior under the conditional probability

Theorem Under the same assumptions and notation as the previous theorem,
for every $\varepsilon > 0$, $\mathbb{P}\left(\max_{0\le k\le n} \left|\frac{S_k - ak}{n}\right| > \varepsilon \,\middle|\, S_n = x_n\right) \underset{n\to\infty}{\longrightarrow} 0$.

The proof is essentially the same as in the "$S_n \geq an$" case:

Proof Set $\theta = \theta(a)$
Write $\mathbb{P}\left(\max_{0\le k\le n} \left|\frac{S_k - ak}{n}\right| > \varepsilon, S_n = x_n\right) = M(\theta)^n \mathbb{E}\left[\mathbb{1}_{\max_{0\le k\le n} \left|\frac{S_k^\theta - ak}{n}\right| > \varepsilon, S_n^\theta = x_n} e^{-\theta x_n}\right] \le M(\theta)^n e^{-\theta x_n} \mathbb{P}\left(\max_{0\le k\le n} \left|\frac{S_k^\theta - ak}{n}\right| > \varepsilon\right)$
Thus $\mathbb{P}\left(\max_{0\le k\le n} \left|\frac{S_k - ak}{n}\right| > \varepsilon \,\middle|\, S_n = x_n\right) \le C\sqrt{n} e^{-cn}$ for certain constants $c, C > 0$.

∾

Theorem Under the same assumptions and notation as the previous theorem, for every $k \geq 0$ and $i_1, \dots, i_k \in \mathbb{Z}$, $\mathbb{P}(X_1 = i_1, \dots, X_k = i_k \mid S_n = x_n) \xrightarrow[n\to\infty]{} \mathbb{P}(X_1 = i_1) \frac{e^{\theta i_1}}{M(\theta)} \dots \mathbb{P}(X_k = i_k) \frac{e^{\theta i_k}}{M(\theta)}$

This means that under $\mathbb{P}(\cdot \mid S_n = x_n)$, $X_1, \dots, X_k$ are asymptotically iid with law giving probability $\mathbb{P}(X_1 = i) \frac{e^{\theta i}}{M(\theta)}$ to i.

Proof Write $\mathbb{P}(X_1 = i_1, \dots, X_k = i_k \mid S_n = x_n)$

$$= \mathbb{P}(X_1 = i_1) \dots \mathbb{P}(X_k = i_k)\, \mathbb{P}(S_{n-k} = x_n - i_1 - \dots - i_k) / \mathbb{P}(S_n = x_n)$$

But $x_n - i_1 - \dots - i_k = a(n-k) + o(\sqrt{n-k})$ so by the theorem, setting $\theta = \theta(a)$ as usual,

$$\mathbb{P}(S_{n-k} = x_n - i_1 - \dots - i_k) \underset{n-k}{\sim} M(a) \, e^{-\theta(a)(x_n - i_1 - \dots - i_k)} \times \frac{1}{\sqrt{2\pi \operatorname{Var}(X_1^\theta)\, n}}$$

Thus $\mathbb{P}(X_1 = i_1, \dots, X_k = i_k \mid S_n = x_n) \xrightarrow[n\to\infty]{} \mathbb{P}(X_1 = i_1) \dots \mathbb{P}(X_k = i_k) \frac{e^{\theta(a)(i_1 + \dots + i_k)}}{M(\theta)^k}$

And the result follows

We finally turn to the proof of the LCLT. Recall the statement:

Theorem (local central limit theorem) Let $(X_i)_{i\geq 1}$ be iid aperiodic random variables with values in $\mathbb{Z}$. Assume that $\mathbb{E}[X_1^2] < \infty$. Set $m = \mathbb{E}[X_1]$ and $\sigma^2 = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2$. Assume $\sigma^2 > 0$. Set $S_n = X_1 + \dots + X_n$. Then

$$\sup_{k \in \mathbb{Z}} \left| \sqrt{n}\, \mathbb{P}(S_n = k) - \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{k - mn}{\sigma\sqrt{n}}\right)^2\right) \right| \xrightarrow[n\to\infty]{} 0$$

Proof: Set $X_i' = X_i - m$, $S_n' = S_n - mn$ and $\phi(t) = \mathbb{E}[e^{itX_1'}]$. Since X_1 is $\mathbb{Z}$-valued, observe that $|\phi(t)|$ is 2π periodic. The idea is to use "discrete Fourier inversion":

since $\mathbb{E}[e^{itS_n'}] = \sum_{j \in \mathbb{Z} - mn} e^{itj} \mathbb{P}(S_n' = j)$, we have $\mathbb{P}(S_n' = k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \mathbb{E}[e^{itS_n'}] dt$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \phi(t)^k \, dt \quad \text{for } k \in \mathbb{Z} - mn$$

Thus, for $u \in \mathbb{R}$ with $u\sigma\sqrt{n} + mn \in \mathbb{Z}$: $\sigma\sqrt{n}\,\mathbb{P}(S_n' = u\sigma\sqrt{n}) = \frac{1}{2\pi}\int_{-\pi\sigma\sqrt{n}}^{\pi\sigma\sqrt{n}} e^{-itu}\,\phi\left(\frac{t}{\sigma\sqrt{n}}\right)^n dt$

But $\forall u \in \mathbb{R}$, $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itu - t^2/2}\,dt$

Thus, for fixed $A > 0$, $0 < \varepsilon \leq 1$, for n sufficiently large:

$\left|\sigma\sqrt{n}\,\mathbb{P}(S_n' = u\sigma\sqrt{n}) - \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\right| \leq \frac{1}{2\pi}\left(|I_1^n(u,A)| + |I_2^n(u,A,\varepsilon)| + |I_3^n(u,\varepsilon)| + |I_4^n(u,A)|\right)$ with:

$I_1^n(u,A) = \int_{-A}^{A} e^{-itu}\left(\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)^n - e^{-t^2/2}\right) dt$, $I_2^n(u,A,\varepsilon) = \int_{A < |t| < \varepsilon\sigma\sqrt{n}} e^{-itu}\,\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)^n dt$, $I_3^n(u,\varepsilon) = \int_{\varepsilon\sigma\sqrt{n} < |t| < \pi\sigma\sqrt{n}} e^{-itu}\,\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)^n dt$

and $I_4(u,A) = \int_{|t| > A} e^{-itu - t^2/2}\,dt$. We shall just write I_1, I_2, I_3, I_4 to simplify notation.

We show that $\forall \varepsilon' > 0$ fixed, we can find $A > 0$ and $\varepsilon \in (0,1)$ such that for n large enough, for every $u \in \mathbb{R}$ with $u\sigma\sqrt{n} + mn \in \mathbb{Z}$ we have $|I_i| \leq \varepsilon'$ for $i = 1,2,3,4$

First write $\varphi(t) \overset{(*)}{=} 1 - \frac{t^2}{2}\sigma^2 + t^2\eta(t^2)$ with $\eta : \mathbb{R} \to \mathbb{C}$ continuous with $\eta(0) = 0$.

To see this, one can use the inequality $\left|e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!}\right| \leq \min\left(\frac{|x|^{n+1}}{(n+1)!}, 2\frac{|x|^n}{n!}\right)$

(cf Lemma 3.3.19 in Durrett's <u>Probability theory and examples</u> 5^{th} Ed), which entails

$$|\varphi(t) - 1 - t^2\sigma^2/2| \leq t^2\,\underbrace{\mathbb{E}[\min(|t||X_1|^3, 2|X_1|^2)]}_{\xrightarrow[t\to 0]{} 0 \text{ by dominated convergence}}$$

First choose $A > 0$ such that $2\int_A^{\infty} e^{-t^2/4}\,dt < \varepsilon'$. $(**)$ The choice of ε will be explained later.

$\boxed{\text{For } I_4}$ $|I_4| \leq 2\int_A^{\infty} e^{-t^2/2}\,dt < \varepsilon'$ by $(**)$.

$\boxed{\text{For } I_1}$ We have $|I_1| \leq \int_{-A}^{A} g_n(t)\,dt$ with $g_n(t) = \left|\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)^n - e^{-t^2/2}\right|$

We have $|g_n(t)| \leq 1 + e^{-t^2/2}$ and $g_n(t) \xrightarrow[n\to\infty]{} 0$ by $(*)$

Thus $|I_1| \leq \varepsilon'$ for n sufficiently large and all u by dominated convergence.

$\boxed{\text{For } I_2}$ We first check that $\exists \varepsilon > 0$ s.t $|\phi(t)| \overset{(***)}{\leq} \exp\left(-\frac{t^2\sigma^2}{4}\right)$ for $|t| \leq \varepsilon$.

By $(*)$, $|\phi(t)|^2 = \phi(t)\overline{\phi(t)} = \left(1 - \frac{t^2\sigma^2}{2} + t^2\eta(t^2)\right)\left(1 - \frac{t^2\sigma^2}{2} + t^2\overline{\eta(t^2)}\right)$

$= 1 - \sigma^2 t^2 + o(t^2)$

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

Since $\exp\left(-\frac{t^2\sigma^2}{4}\right) = 1 - \frac{\sigma^2}{4}t^2 + o(t^2)$ we get $(\ast\ast\ast)$

Then, for $|t| \leq \varepsilon$: $|I_2| \leq 2\int_A^{\varepsilon\sigma\sqrt{n}} \left(e^{-\left(\frac{t}{\sigma\sqrt{n}}\right)^2 \frac{\sigma^2}{4}}\right) dt \leq 2\int_A^{\infty} e^{-t^2/4} dt \leq \varepsilon'$

$\boxed{\text{For } I_3}$ By aperiodicity, $|\phi(t)| < 1$ for $t \in (0, 2\pi)$. Thus $\exists\, c > 0$ st $\left|\phi\left(\frac{t}{\sigma\sqrt{n}}\right)\right| \leq e^{-c}$ for $\varepsilon\sigma\sqrt{n} < |t| < \pi\sigma\sqrt{n}$, so that:

$$|I_3| \leq 2\int_{\varepsilon\sigma\sqrt{n}}^{\pi\sigma\sqrt{n}} e^{-cn} dt \leq 2\pi\sigma\sqrt{n}\, e^{-cn} \leq \varepsilon' \text{ for } n \text{ sufficiently large}$$

∞

Remarks • If X_1 has span h with $\mathbb{P}(X_1 = b + h\mathbb{Z}) = 1$, one gets an analog of the LCLT by considering $\frac{X_i - b}{h}$.

• It is possible to show that

$$\sup_{k \in \mathbb{Z}} \max\left(1, \left(\frac{k - mn}{\sqrt{n}}\right)^2\right) \left| \sigma\sqrt{n}\, \mathbb{P}(S_n = k) - \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-mn}{\sigma\sqrt{n}}\right)^2} \right| \xrightarrow[n\to\infty]{} 0$$

which gives a better error bound for $|k - mn| \gg \sqrt{n}$.

(see Principles of Random Walk, Spitzer, Chap II, Sec 7, P10)

End of Lecture 4