Igor Kortchemski
Condensation phenomena in random trees
ETH, Spring 2024

Chapter 2: One big jump phenomenon

1) A maximal inequality
2) A local estimate
3) A one-big jump principle

Here we identify a regime where atypical events of the form $\{S_n = an\}$ with (S_n) random walk typically occur with one big jump (essentially when $\mathbb{P}(X_1 = n) \underset{n\to\infty}{\sim} \frac{c}{n^\gamma}$: heavy tail)

1) A maximal inequality

We start with proving an inequality for future use.

Theorem [Fuk-Nagaev '71, Denisov-Dieker-Schneer '08]
Assume that X is a $\mathbb{R}$-valued r.v with $\mathbb{E}[X^2]<\infty$, $\mathbb{E}[X]=0$. Let $(X_i)_{i\geq 1}$ be iid having same law as X. Set $S_n = X_1 + \cdots + X_n$. There $\exists K>0$ s.t $\forall n\geq 1$, $x>0$ and $c\geq 1$:

$$\mathbb{P}(S_n > x\sqrt{n}, X_1 \leq c\sqrt{n}, \ldots, X_n \leq c\sqrt{n}) \leq K \exp\left(-\frac{x}{c}\right)$$

Proof: The idea is to introduce the truncated random walk $\tilde{S}_n = \sum_{i=1}^{n} X_i \mathbb{1}_{X_i \leq c\sqrt{n}}$. Indeed,
$\mathbb{P}(S_n > x\sqrt{n}, X_1 \leq c\sqrt{n}, \ldots, X_n \leq c\sqrt{n}) \leq \mathbb{P}(\tilde{S}_n \geq x\sqrt{n})$.
To bound this probability we use the "exponential Markov" inequality:
$\mathbb{P}(\tilde{S}_n \geq x\sqrt{n}) = \mathbb{P}(e^{\frac{\tilde{S}_n}{c\sqrt{n}}} \geq e^{\frac{x}{c}}) \leq e^{-\frac{x}{c}} \mathbb{E}[e^{\frac{\tilde{X}}{c\sqrt{n}}}]^n$ with $\tilde{X} = X \mathbb{1}_{X \leq c\sqrt{n}}$ (⚠ $\tilde{X}$ depends on n)
We show that $\mathbb{E}[e^{\frac{\tilde{X}}{c\sqrt{n}}}] = 1 + O\left(\frac{1}{n}\right)$ and the result will follow.
The idea is to write $e^x = 1 + x + x^2 r(x)$ with $r(x) = \frac{e^x - 1 - x}{x^2}$, so that

$$\mathbb{E}[e^{\tilde{X}/c\sqrt{n}}] = 1 + \underbrace{\mathbb{E}\left[\frac{\tilde{X}}{c\sqrt{n}}\right]}_{m_n} + \underbrace{\mathbb{E}\left[\left(\frac{\tilde{X}}{c\sqrt{n}}\right)^2 r\left(\frac{\tilde{X}}{c\sqrt{n}}\right)\right]}_{s_n}$$

We show that $m_n = O\left(\frac{1}{n}\right)$ and $s_n = O\left(\frac{1}{n}\right)$

- For s_n: since r is bounded on $(-\infty, 1]$, we have $s_n = O\left(\mathbb{E}\left[\frac{\tilde{X}^2}{n}\right]\right)$ (here we use $c\geq 1$)
But $\mathbb{E}[\tilde{X}^2] \leq \mathbb{E}[X^2]$, thus $s_n = O\left(\frac{1}{n}\right)$.

• For m_n: write $m_n = \frac{1}{c\sqrt{n}} \mathbb{E}[X \mathbb{1}_{X \le c\sqrt{n}}] = -\frac{1}{c\sqrt{n}} \mathbb{E}[X \mathbb{1}_{X > c\sqrt{n}}]$ because $\mathbb{E}[X]=0$

But $\mathbb{E}[|X| \mathbb{1}_{|X|>c\sqrt{n}}] \le \mathbb{E}[|X| \cdot \frac{|X|}{c\sqrt{n}} \mathbb{1}_{|X|>c\sqrt{n}}] \le \frac{1}{c\sqrt{n}} \mathbb{E}[X^2]$

Thus $m_n = O(\frac{1}{n})$ (here we use $c \ge 1$)

$\sim$

Remark The result is false with "$c>0$" instead of "$c \ge 1$". Indeed, take $\mathbb{P}(X_1 = \pm 1) = \frac{1}{2}$, $x=1$, $c = n^{-1/4}$.

Then $\mathbb{P}(S_n \ge x\sqrt{n}, X_1 \le c\sqrt{n}, \ldots, X_n \le c\sqrt{n}) = \mathbb{P}(S_n \ge \sqrt{n}) \xrightarrow[n\to\infty]{} \mathbb{P}(\mathcal{N}(0,1) \ge 1)$

but $K \exp(-\frac{x}{c}) \xrightarrow[n\to\infty]{} 0$

2) A local estimate

We introduce a framework which will allow to treat "$S_n > an$" and "$S_n = an$" at the same time.

Fix $T \in (0, +\infty]$ and set $\Delta = [0, T)$. For $m \in \mathbb{R}$ we set $\Delta_m = m + \Delta = [m, m+T)$

Condition (H_Δ) There exist $c>0$ and $\beta > 2$ such that

- If $T = \infty$, $\mathbb{P}(X_1 \in \Delta_u) = \mathbb{P}(X_1 \ge u) \underset{u\to\infty}{\sim} \frac{c}{u^\beta}$
- If $T < \infty$, $\mathbb{P}(X_1 \in \Delta_u) = \mathbb{P}(X_1 \in [u, u+T)) \underset{u\to\infty}{\sim} \frac{c}{u^{1+\beta}}$

Example If X_1 is $\mathbb{Z}$-valued and if $\mathbb{P}(X_1 = n) \underset{n\to\infty}{\sim} \frac{c}{n^{1+\beta}}$ then X_1 satisfies (H_Δ) for $T=1$ and $T = \infty$, but not $T = 1.5$ (indeed, $\mathbb{P}(X_1 \in [n, n+1.5)) \underset{n\to\infty}{\sim} \frac{2c}{n^{1+\beta}}$ and $\mathbb{P}(X_1 \in [n+0.1, n+1.6)) \underset{n\to\infty}{\sim} \frac{c}{n^{1+\beta}}$)

Remark One can check that if X_1 satisfies (H_Δ) then X_1 has finite variance (because $\beta > 2$)

Lemma Assume that for $c>0, T>0, \beta>2$, $\mathbb{P}(X_1 \in [u, u+T)) \underset{u\to\infty}{\sim} \frac{c}{u^{1+\beta}}$. Then $\mathbb{P}(X_1 \ge u) \sim \frac{c}{\beta T} \cdot \frac{1}{u^\beta}$

In particular, if X_1 satisfies (H_Δ) for $T < +\infty$ then it satisfies (H_Δ) for $T = \infty$

Proof: Exercise for next week

End of lecture 5

We will use several times the following fact

Fact (*) Under (H_Δ), $\mathbb{P}(X_1 \in \Delta_u) \underset{u\to\infty}{\sim} \mathbb{P}(X_1 \in \Delta_{u+y})$ uniformly in $|y| \leq \frac{u}{\ln(u)}$,

i.e $\sup_{|y|\leq \frac{u}{\ln(u)}} \left| \frac{\mathbb{P}(X_1\in\Delta_u)}{\mathbb{P}(X_1\in \Delta_{u+y})} - 1 \right| \xrightarrow[u\to\infty]{} 0$

Theorem (Doney '89, Nagaev '57)

Assume that X_1 satisfies (H_Δ) and that $\mathbb{E}[X_1]=0$. Fix $\varepsilon>0$. Then, uniformly in $m \geq \varepsilon n$,

$$\mathbb{P}(S_n \in \Delta_m) \underset{n\to\infty}{\sim} n\,\mathbb{P}(X_1\in\Delta_m)$$

That is, $\sup_{m\geq \varepsilon n} \left| \frac{\mathbb{P}(S_n\in\Delta_m)}{n\,\mathbb{P}(X_1\in\Delta_m)} - 1\right| \xrightarrow[n\to\infty]{} 0$

Intuition: $S_n\in\Delta_m$ typically happens when one of the n jumps is in Δ_m (this will be made precise later)

Proof Set $\bar{m} = \frac{m}{\ln(m)^3}$. Write $n P_1^{m,n} \leq \mathbb{P}(S_n\in\Delta_m) \leq n P_1^{m,n} + P_0^{m,n} + P_2^{m,n}$ with

- $P_1^{m,n} = \mathbb{P}(S_n\in\Delta_m,\ X_n \geq \bar{m},\ \max_{1\leq k\leq n-1} X_k < \bar{m})$
- $P_0^{m,n} = \mathbb{P}(S_n\in\Delta_m,\ \max_{1\leq k\leq n} X_k < \bar{m})$
- $P_2^{m,n} = \mathbb{P}(S_n\in\Delta_m,\ \bigcup_{1\leq j<k\leq n} \{X_j\geq\bar{m},\ X_k\geq\bar{m}\})$

We show that uniformly in $m\geq\varepsilon n$, $P_1^{m,n} \sim \mathbb{P}(X_1\in\Delta_m)$, $P_0^{m,n} = o(n\mathbb{P}(X_1\in\Delta_m))$, $P_2^{m,n} = o(n\mathbb{P}(X_1\in\Delta_m))$

Recall that $\mathbb{P}(X_1\in\Delta_m) \sim \begin{cases} \frac{c}{m^\beta} & \text{if } T=\infty \\ \frac{c}{m^{1+\beta}} & \text{if } T<\infty \end{cases}$. We have $\mathbb{P}(X_1\geq u) \underset{u\to\infty}{\sim} \frac{K}{u^\beta}$ for some $K>0$ by the lemma.

For $P_2^{m,n}$: $P_2^{m,n} \leq \binom{n}{2}\mathbb{P}(X_1\geq\bar{m},\ X_2\geq\bar{m})$

$$\overset{(H_\Delta)}{\leq} C' n^2 \frac{1}{\bar{m}^\beta}\cdot\frac{1}{\bar{m}^\beta} = \frac{n^2}{m^{2\beta}}\cdot \ln(m)^{6\beta}, \quad \text{so} \quad \frac{P_2^{m,n}}{n\,\mathbb{P}(X_1\in\Delta_m)} \leq C' \frac{n}{m^{2\beta}}\cdot\ln(m)^{6\beta}\cdot m^{1+\beta} = C' \frac{n\cdot\ln(m)^{6\beta}}{m^{\beta-1}} \longrightarrow 0$$

because $\beta-1>1$

For $P_0^{m,n}$: We use the maximal inequality:

$$P_0^{m,n} \leq \mathbb{P}(S_n\geq m,\ \max_{1\leq k\leq n} X_k < \bar{m}) \leq K\exp\left(-\frac{m}{\bar{m}}\right) = K\exp(-\ln(m)^3) = o\left(\frac{n}{m^{1+\beta}}\right)$$

For $P_1^{m,n}$: This is more delicate, we need to consider cases according to the value of S_{n-1}.

Since $\frac{S_n}{\sqrt{n}}$ converges in distribution (CLT), $\frac{S_n}{n^{3/4}} \xrightarrow{\mathbb{P}} 0$. Then write

$P_1^{m,n} = Q_1^{m,n} + Q_2^{m,n} + Q_3^{m,n}$ with

$Q_1^{m,n} = \mathbb{P}(S_n \in \Delta_m, X_n \geq \bar{m}, \max_{1\leq k\leq n-1} X_k < \bar{m}, S_{n-1} > \frac{m}{\ln(m)})$

$Q_2^{m,n} = \mathbb{P}(S_n \in \Delta_m, X_n \geq \bar{m}, \max_{1\leq k\leq n-1} X_k < \bar{m}, -n^{3/4} < S_{n-1} < \frac{m}{\ln(m)})$

$Q_3^{m,n} = \mathbb{P}(S_n \in \Delta_m, X_n \geq \bar{m}, \max_{1\leq k\leq n-1} X_k < \bar{m}, S_{n-1} < -n^{3/4})$

We first show $Q_1^{m,n}, Q_3^{m,n} = o(\mathbb{P}(X_1 \in \Delta_m))$:

- For $Q_1^{m,n}$: By the maximal inequality, $Q_1^{m,n} \leq \mathbb{P}(S_{n-1} > \frac{m}{\ln(m)}, \max_{1\leq k\leq n-1} X_k < \bar{m}) \leq K \exp(-\ln(m)^2) = o(\mathbb{P}(X_1 \in \Delta_m))$.
- For $Q_3^{m,n}$:

$$\begin{aligned}Q_3^{m,n} &\leq \sum_{i<-n^{3/4}} \mathbb{P}(S_{n-1} = i, X_n \in \Delta_{m-i}) \\ &= \sum_{i<-n^{3/4}} \mathbb{P}(S_{n-1} = i)\mathbb{P}(X_n \in \Delta_{m-i}) \\ &\leq \underbrace{\sum_{i<-n^{3/4}} \mathbb{P}(S_{n-1}=i)}_{=\mathbb{P}(S_{n-1} < -n^{3/4}) = o(1)} \underbrace{\sup_{j \geq n^{3/4}} \mathbb{P}(X_1 \in \Delta_{m+j})}_{= O(\mathbb{P}(X_1 \in \Delta_m)) \text{ by } (H_\Delta)}\end{aligned}$$

- For $Q_2^{m,n}$:

$$\begin{aligned}Q_2^{m,n} &= \mathbb{P}(\max_{1\leq k\leq n-1} X_k < \bar{m}, -n^{3/4} < S_{n-1} < \frac{m}{\ln(m)}, X_n \geq \bar{m}, X_n \in \Delta_{m-S_{n-1}}) \\ &= \mathbb{P}(\max_{1\leq k\leq n-1} X_k < \bar{m}, -n^{3/4} < S_{n-1} < \frac{m}{\ln(m)}, X_n \in \Delta_{m-S_{n-1}})\end{aligned}$$

because $m - \frac{m}{\ln(m)} \geq \bar{m}$ for n large enough.

Observe that by assumption $\mathbb{P}(X_n \in \Delta_{m-j}) \sim \mathbb{P}(X_n \in \Delta_m)$ uniformly for $-n^{3/4} < j < \frac{m}{\ln(m)}$ (by fact (*))

Thus $Q_2^{m,n} \underset{n\to}{\sim} \mathbb{P}(\max_{1\leq k\leq n-1} X_k < \bar{m}, -n^{3/4} < S_{n-1} < \frac{m}{\ln(m)}) \mathbb{P}(X_1 \in \Delta_m)$

End of lecture 6

It remains to check that:

- $\mathbb{P}(-n^{3/4} < S_{n-1} < \frac{m}{\ln(m)}) \xrightarrow[n\to\infty]{} 1$, which is clear since $\frac{S_{n-1}}{\sqrt{n}}$ converges in distribution
- $\mathbb{P}(\max_{1\leq k\leq n-1} X_k < \bar{m}) \xrightarrow[n\to\infty]{} 1$;

(indeed, for two event A, B, $\mathbb{P}(A\cap B) = 1 - \mathbb{P}(A \cup B) \geq 1 - \mathbb{P}(A) - \mathbb{P}(B) = 1 - \mathbb{P}(A^c) - \mathbb{P}(B^c)$)

this comes from the fact that $\mathbb{P}(\max_{1\leq k\leq n-1} X_k < \bar{m}) = (1 - \mathbb{P}(X_1 > \bar{m}))^{n-1} = \exp(-n\,\mathbb{P}(X_1 > \bar{m})(1+o(1)))$

and $n\,\mathbb{P}(X_1 > \bar{m}) \sim c' \frac{n}{m^\beta} \cdot \ln(m)^{3\beta} \xrightarrow[n\to\infty]{} 0$ since $\beta > 2$

∾

Remark There has been quite some work to find the "best" sequence m_n such that the theorem holds uniformly for $m \geq m_n$ (we have shown that $m_n = \varepsilon n$ works).

3) A one-big-jump principle

We keep the previous framework : $T \in (0,+\infty]$, we set $\Delta = [0,T)$, $\Delta_u = [u, u+T)$ and assume that X_1 satisfies (H_Δ) and that $\mathbb{E}[X_1]=0$.

<u>Notation</u> • Set $V_n = \min\{1 \leq j \leq n : X_j = \max(X_1, \dots, X_n)\}$

• Set $(\hat{X}_1, \dots, \hat{X}_{n-1}) = (X_1, \dots, X_{V_n-1}, X_{V_n+1}, \dots, X_n)$

Fix a sequence (x_n) such that $\liminf\limits_{n\to\infty} \frac{x_n}{n} > 0$.

<u>Theorem</u> (one big jump principle, Armendariz & Loulakis '11)

We have $d_{TV}\big((\hat{X}_1, \dots, \hat{X}_{n-1}) \text{ under } \mathbb{P}(\cdot \mid S_n \in \Delta_{x_n}), (X_1, \dots, X_{n-1})\big) \xrightarrow[n\to\infty]{} 0$, that is

$$\sup_{A \in \mathcal{B}(\mathbb{R}^{n-1})} \left| \mathbb{P}((\hat{X}_1, \dots, \hat{X}_{n-1}) \in A \mid S_n \in \Delta_{x_n}) - \mathbb{P}((X_1, \dots, X_{n-1}) \in A) \right| \xrightarrow[n\to\infty]{} 0$$

This means that under $\mathbb{P}(\cdot \mid S_n \in \Delta_{x_n})$, once the biggest jump is removed, the remaining r.v are asymptotically iid with same law as X_1 !

In practice, to show that a property holds with probability tending to 0 or 1 for $(\hat{X}_1, \dots, \hat{X}_{n-1})$ under $\mathbb{P}(\cdot \mid S_n \in \Delta_{x_n})$ one can show that it holds for $(X_1, \dots, X_{n-1})$ (which are iid !)

Before proving this, let us see a striking consequence.

<u>Corollary</u> Set $D_n = \max(X_1, \dots, X_n)$ and let $D_n^{(2)}$ be the second largest element of $(X_1, \dots, X_n)$. Then:

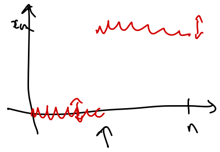
① $\forall u \geq 0$ $\mathbb{P}\left(\frac{D_n^{(2)}}{n^{1/\beta}} \leq u \,\middle|\, S_n \in \Delta_{x_n} \right) \xrightarrow[n\to\infty]{} \exp\left(-\frac{c_0}{u^\beta}\right)$ with $c_0 = \begin{cases} c & \text{if } T = \infty \\ \frac{c}{\beta T} & \text{if } T = \infty \end{cases}$

Now assume $T < \infty$

② $\frac{D_n - x_n}{\sigma \sqrt{n}}$ under $\mathbb{P}(\cdot \mid S_n \in \Delta_{x_n}) \xrightarrow{(d)} N(0,1)$ with $\sigma^2 = \mathrm{Var}(X_1)$

③ $\frac{D_n}{x_n}$ under $\mathbb{P}(\cdot \mid S_n \in \Delta_{x_n}) \xrightarrow{\mathbb{P}} 1$

Thus for $T<\infty$ we have



uniform time between 1 and n

fluctuations of order $\sqrt{n}$; second, third, etc. largest jumps of order $n^{1/\beta}$.

Proof ① By the theorem, it is enough to show that $\mathbb{P}\left(\frac{\max(X_1,\dots,X_{n-1})}{n^{1/\beta}} \leq u\right) \xrightarrow[n\to\infty]{} \exp\left(-\frac{c_0}{u^\beta}\right)$

For this, the probability is $\left(1-\mathbb{P}(X_1 \geq u\, n^{\frac{1}{\beta}})\right)^{n-1}$

$$= \exp\left((n-1)\ln\left(1-\mathbb{P}(X_1 \geq u\, n^{\frac{1}{\beta}})\right)\right)$$

But we have seen that $\mathbb{P}(X_1 \geq u\, n^{1/\beta}) \underset{n\to\infty}{\sim} \frac{c_0}{u^\beta n}$, and the result follows.

② Observe that $\frac{D_n - x_n}{\sigma\sqrt{n}} = \frac{S_n - x_n}{\sigma\sqrt{n}} - \frac{(\hat{X}_1 + \dots + \hat{X}_{n-1})}{\sigma\sqrt{n}}$

Under $\mathbb{P}(\cdot \mid S_n \in \Delta_{x_n})$, $|S_n - x_n| \leq T$, so $\frac{S_n - x_n}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{\mathbb{P}} 0$ and $\frac{\hat{X}_1 + \dots + \hat{X}_{n-1}}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{(d)} N(0,1)$ by the central limit theorem. The result follows

③ Since $\frac{D_n}{x_n} - 1 = \frac{\sigma\sqrt{n}}{x_n} \cdot \frac{D_n - x_n}{\sigma\sqrt{n}}$ and $\frac{\sigma\sqrt{n}}{x_n} \xrightarrow[n\to\infty]{} 0$, this follows from ②.

Proof of the theorem

Let $\hat{\mu}_n$ be the law of $(\hat{X}_1,\dots,\hat{X}_{n-1})$ under $\mathbb{P}(\cdot \mid S_n \in \Delta_{x_n})$

Let μ_n be the law of $(X_1,\dots,X_{n-1})$

To show that $\sup_{A\in\mathcal{B}(\mathbb{R}^{n-1})} |\mu_n(A) - \hat{\mu}_n(A)| \xrightarrow[n\to\infty]{} 0$ the idea is to find a "good" event E_n with

① $\mu_n(E_n) \xrightarrow[n\to\infty]{} 1$

② $\sup_{\substack{A\subset E_n \\ A\in\mathcal{B}(\mathbb{R}^{n-1})}} |\mu_n(A) - \hat{\mu}_n(A)| \xrightarrow[n\to\infty]{} 0$

Indeed, by ② we then have $\hat{\mu}_n(E_n) \xrightarrow[n\to\infty]{} 1$, so

$$\sup_{A\in\mathcal{B}(\mathbb{R}^{n-1})} |\mu_n(A) - \hat{\mu}_n(A)| \leq \sup_{A\in\mathcal{B}(\mathbb{R}^{n-1})} |\mu_n(A\cap E_n) - \hat{\mu}_n(A\cap E_n)| + \mu_n(E_n^c) + \hat{\mu}_n(E_n^c) \xrightarrow[n\to\infty]{} 0$$

To do this, set $E_n = \{ a = (a_1, \dots, a_{n-1}) \in \mathbb{R}^{n-1} : |a_1 + \dots + a_{n-1}| \leq n^{3/4}$ and $\max_{1 \leq i \leq n-1} a_i \leq n^{3/4} \}$

We check ① $\mu_n(E_n^c) \leq \mathbb{P}(|S_{n-1}| > n^{3/4}) + \mathbb{P}(\max_{1 \leq j \leq n-1} X_j > n^{3/4})$

- the first term is $o(1)$ since $\frac{S_{n-1}}{\sqrt{n}}$ converges in distribution
- The second term is $1 - (1 - \underbrace{\mathbb{P}(X_1 > n^{3/4})}_{\sim \frac{c_0}{n^{\frac{3}{4}\beta}}})^{n-1} \xrightarrow[n\to\infty]{} 0$ since $\frac{n}{n^{\frac{3}{4}\beta}} \to 0$ $(\beta > 2)$

We check ②: for $a = (a_1, \dots, a_{n-1}) \in \mathbb{R}^{n-1}$, observe that

$$\{ (\hat{X}_1, \dots, \hat{X}_{n-1}) = a, S_n \in \Delta_{x_n} \} = \bigcup_{i=1}^{n} \{ (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n) = a, X_i \in \Delta_{x_n - a_1 - \dots - a_{n-1}} \}$$

and that the union is disjoint. Thus

$$\hat{\mu}_n(a) = \mu_n(a) \times \underbrace{\frac{n \mathbb{P}(X_1 \in \Delta_{x_n - a_1 - \dots - a_{n-1}})}{\mathbb{P}(S_n \in \Delta_{x_n})}}$$

$$= 1 + \varepsilon_n(a) \quad \text{with} \quad \sup_{a \in E_n} |\varepsilon_n(a)| \xrightarrow[n\to\infty]{} 0 \quad \text{by the theorem of 2).}$$

This entails $\sup_{A \subset E_n} |\hat{\mu}_n(A) - \mu_n(A)| = 1 + o(1)$.

Indeed, for $A \subset E_n$, $|\hat{\mu}_n(A) - \mu_n(A)| \leq \sum_{a \in E_n} |\hat{\mu}_n(a) - \mu_n(a)| \leq \sum_{a \in E_n} \mu_n(a) \varepsilon_n(a) \leq \sup_{a \in E_n} |\varepsilon_n(a)|$.

End of lecture 7