

Lecture 9. Lovász Local Lemmas. Review of (discrete)
random variables. Chernoff bounds

Martin Klazar

December 1, 2020

LLL

Let A, A_1, \dots, A_n , $n \in \mathbb{N}$, be events in a probability space. We say that A is independent of the events A_1, \dots, A_n , if for every set $I \subset [n] = \{1, 2, \dots, n\}$ one has that

$$\Pr(A \cap \bigcap_{i \in I} A_i) = \Pr(A) \cdot \Pr(\bigcap_{i \in I} A_i) .$$

Recall that for $I = \emptyset$ we define the intersection over I as Ω . Thus if $\Pr(\bigcap_{i \in I} A_i) > 0$, the independence of A of the A_i s means that

$$\Pr(A \mid \bigcap_{i \in I} A_i) = \Pr(A) .$$

A lemma similar to one in the previous lecture holds:

Exercise. If an event A is independent of the events A_1, \dots, A_n , then A is also independent of the events A'_1, \dots, A'_n where each A'_i is either A_i or $\overline{A_i} = \Omega \setminus A_i$. \square

For events A_1, \dots, A_n , $n \in \mathbb{N}$, we say that a digraph (directed graph)

$$D = ([n], E), \quad E \subset [n] \times [n] ,$$

is their *dependency digraph* if for every $i \in [n]$ the event A_i is independent of the events $\{A_j \mid j \in [n], (i, j) \notin E\}$. So all events “problematic” with respect to A_i , on which A_i may depend, have their indices contained in the out-neighbors of i . Dependency digraph may not be unique. We define $\Delta(D)$ to be the maximum out-degree in D ,

$$\Delta(G) := \max_{i \in [n]} |\{j \in [n] \mid (i, j) \in E\}| .$$

We state two versions of LLL and prove a third one. Recall that $\overline{A} = \Omega \setminus A$.

Theorem (LLL, general form). *Let $n \in \mathbb{N}$, $x_1, \dots, x_n \in [0, 1]$, A_1, \dots, A_n be events in a probability space, and $D = ([n], E)$ be their dependency digraph. If for every $i \in [n]$ one has that*

$$\Pr(A_i) \leq x_i \prod_{\substack{j \in [n] \\ (i, j) \in E}} (1 - x_j), \text{ then } \Pr\left(\bigcap_{i=1}^n \overline{A}_i\right) \geq \prod_{i=1}^n (1 - x_i) > 0$$

— with a positive probability none of the events A_i occurs.

Theorem (LLL, symmetric form). *Let $d, n \in \mathbb{N}$, $p \in [0, 1]$, and A_1, \dots, A_n be events in a probability space with $\Pr(A_i) \leq p$ for every $i \in [n]$ and with a dependency digraph $D = ([n], E)$ such that $\Delta(D) \leq d$. Then ($e = 2.71828\dots$ is the Euler number)*

$$ep(d + 1) \leq 1 \Rightarrow \Pr\left(\bigcap_{i=1}^n \overline{A}_i\right) > 0.$$

LLL was motivated and illustrated by applications in Lecture 7, where the symmetric form was deduced from the general one. For a proof of the general form see pp. 64–65 in Alon and Spencer. However, that proof does not deal with the issue that conditional probability $\Pr(A|B)$ is defined only if $\Pr(B) > 0$. We give here the proof from the textbook

- M. Mitzenmacher and E. Upfal, *Probability and Computing. Randomization and Probabilistic Techniques in Algorithms and Data Analysis*, Cambridge U. Press, 2017,

which is recommended literature for this course, and prove a slightly weaker symmetric LLL with the inequality $4pd \leq 1$.

Proof. (Mitzenmacher and Upfal, pp. 148–150.) The assumptions are as in the above symmetric LLL, except that we assume more restrictively that

$$4pd \leq 1. \tag{0}$$

We prove by induction on $s = 0, 1, \dots, n - 1$ that for every $k \in [n]$

and $S \subset [n]$ with $|S| = s$ and $k \notin S$,

$$\Pr(A_k | \bigcap_{i \in S} \overline{A_i}) \leq 2p. \quad (*)$$

For $S = \emptyset$ we define the intersection as Ω . For $s = 0$ the bound $(*)$ holds as $\Pr(A_k | \Omega) = \Pr(A_k) \leq p \leq 2p$. We assume that $s > 0$ and that the bound $(*)$ holds for every S with less than s elements.

To perform the induction step we first prove by inner induction over $|S|$ that $\Pr(\bigcap_{i \in S} \overline{A_i}) > 0$. For $s = 1$ this holds as $\Pr(\overline{A_i}) = 1 - \Pr(A_i) \geq 1 - p > 0$ by the assumption (0). For $s > 1$ we assume, as we may, that $S = \{1, 2, \dots, s\}$ and compute that

$$\begin{aligned} \Pr(\bigcap_{i=1}^s \overline{A_i}) &\stackrel{a=\frac{a}{b} \frac{b}{c} \frac{c}{1}}{=} \prod_{i=1}^s \left(\Pr(\bigcap_{j=1}^i \overline{A_j}) / \underbrace{\Pr(\bigcap_{j=1}^{i-1} \overline{A_j})}_{> 0 \text{ by inner ind.}} \right) \\ &\stackrel{\text{prob. of } \overline{A}}{=} \prod_{i=1}^s \frac{\Pr(\bigcap_{j=1}^{i-1} \overline{A_j}) - \Pr(A_i \cap \bigcap_{j=1}^{i-1} \overline{A_j})}{\Pr(\bigcap_{j=1}^{i-1} \overline{A_j})} \\ &\stackrel{\text{cond. prob.}}{=} \prod_{i=1}^s (1 - \Pr(A_i | \bigcap_{j=1}^{i-1} \overline{A_j})) \\ &\stackrel{\text{induction } (*)}{\geq} \prod_{i=1}^s (1 - 2p) \stackrel{(0)}{>} 0. \end{aligned} \quad (1)$$

We return to the induction step for the bound $(*)$ and set $S_1 := \{j \in S \mid (k, j) \in E\}$ and $S_2 := S \setminus S_1$. If $S_2 = S$ then the event A_k is independent of the events $\{\overline{A_j} \mid j \in S\}$ (by the definition of the digraph D and by the previous exercise) and we get

$$\Pr(A_k | \bigcap_{j \in S} \overline{A_j}) = \Pr(A_k) \leq p \leq 2p,$$

the bound $(*)$.

Let $|S_2| < s$. We introduce the notation $F_S := \bigcap_{j \in S} \overline{A_j}$, and similarly for F_{S_1} and F_{S_2} . So $F_S = F_{S_1} \cap F_{S_2}$ and

$$\Pr(A_k | \bigcap_{i \in S} \overline{A_i}) = \frac{\Pr(A_k \cap F_S)}{\Pr(F_S)}.$$

Recall that $\Pr(F_S) > 0$ by the inner induction, and similarly $\Pr(F_{S_2}) > 0$. We have

$$\Pr(A_k \cap F_S) = \Pr(A_k \cap F_{S_1} \cap F_{S_2}) = \Pr(A_k \cap F_{S_1} | F_{S_2}) \Pr(F_{S_2})$$

and

$$\Pr(F_S) = \Pr(F_{S_1} \cap F_{S_2}) = \Pr(F_{S_1} | F_{S_2}) \Pr(F_{S_2}).$$

Thus

$$\Pr(A_k | F_S) = \frac{N}{J} := \frac{\Pr(A_k \cap F_{S_1} | F_{S_2})}{\Pr(F_{S_1} | F_{S_2})}.$$

We estimate N from above and J from below. In

$$N = \Pr(A_k \cap F_{S_1} | F_{S_2}) \leq \Pr(A_k | F_{S_2}) = \Pr(A_k) \leq p$$

the first inequality follows from the monotonicity of probability, the next equality from the definition of S_2 and D (and the previous exercise), and the last inequality is assumed. Since $|S_2| < |S| = s$, we get that

$$\begin{aligned} J &\stackrel{\text{def}}{=} \Pr(F_{S_1} | F_{S_2}) \stackrel{\text{def}}{=} \Pr\left(\bigcap_{i \in S_1} \overline{A_i} \mid \bigcap_{j \in S_2} \overline{A_j}\right) \\ &\stackrel{\text{de Morgan}}{=} \Pr\left(\overline{\bigcup_{i \in S_1} A_i} \mid \bigcap_{j \in S_2} \overline{A_j}\right) \\ &\stackrel{\text{union bound}}{\geq} 1 - \sum_{i \in S_1} \Pr(A_i \mid \bigcap_{j \in S_2} \overline{A_j}) \\ &\stackrel{\text{induction (*)}}{\geq} 1 - \sum_{i \in S_1} 2p \\ &\stackrel{|S_1| \leq d}{\geq} 1 - 2pd \stackrel{(0)}{\geq} 1/2. \end{aligned}$$

Thus the inequality (*) is proven,

$$\Pr(A_k \mid \bigcap_{i \in S} \overline{A_i}) = \Pr(A_k | F_S) = \frac{N}{J} \leq \frac{p}{1/2} = 2p.$$

Finally,

$$\Pr\left(\bigcap_{i=1}^n \overline{A_i}\right) = \prod_{i=1}^n (1 - \Pr(A_i \mid \bigcap_{j=1}^{i-1} \overline{A_j})) \geq \prod_{i=1}^n (1 - 2p) > 0$$

—the first equality is equality (1) for $s = n$, the next inequality follows from inequality (*), and the last inequality is due to $2p < 1$ which follows from (0). \square

The LLL appeared in a 1975 article by P. Erdős and L. Lovász.

- *László Lovász* is a Hungarian mathematician, specializing in combinatorics, and computer scientist. Of his books one has to mention *Combinatorial Problems and Exercises*, AMS, 2007 (previous editions in 1979 and 1993) and *Large Networks and Graph Limits*, AMS, 2012.

A nice application of LLL that you saw in Lecture 7 is as follows. For any $k, d \in \mathbb{N}$ with $e(d+1) \leq 2^{k-1}$, if X is a finite set of finite sets such that $|Y| \geq k$ for every $Y \in X$ and

$$\forall Y \in X : |\{Z \in X \mid Z \cap Y \neq \emptyset\}| \leq d + 1 ,$$

then

$$\exists \chi: \bigcup X \rightarrow [2] : Y \in X \Rightarrow |\chi(Y)| > 1 .$$

In words, every set system X in which each edge has k and more elements and intersects at most d other edges, has a proper vertex coloring with two colors (the vertices can be colored by two colors so that no edge is monochromatic). LLL ensures existence of such coloring, but can we efficiently — by a polynomial-time algorithm — find one? The first efficient algorithmic LLL was found by J. Beck in 1991. His results, and of others, were much improved and generalized by R. A. Moser and G. Tardos in 2010 (see their preprint arXiv:0903.0544 and the Wikipedia article *Algorithmic Lovász local lemma*). For their article A constructive proof of the general Lovász Local Lemma, *Journal of the ACM* **57(2)**, 11:1–11:15 (2010) they were awarded the 2020 Gödel Prize.

- *József Beck* is a Hungarian mathematician who is a professor of mathematics at the Rutgers University in the USA. He should not be confused with the Polish politician Józef Beck (1894–1944). His results in mathematical foundations of the kinetic theory of gases and in rigorous statistical physics are fundamental (see Lecture 13).
- “*Robin Moser* obtained his PhD in 2012 from the Swiss Federal Institute of Technology in Zurich (...) Since 2013, he has worked developing trading software and as a quantitative analyst for Circular Capital in the Basel area in Switzerland.” (from the citation to the 2020 Gödel Prize).
- *Gábor Tardos* is a Hungarian mathematician, currently a professor at Central European University.

Review of RVs (random variables)

We review random variables¹, especially discrete ones. Recall that a real-valued function

$$X: \Omega \rightarrow \mathbb{R}$$

on a probability space (Ω, Σ, \Pr) is a *random variable*, if for every $a \in \mathbb{R}$ one has that

$$\{\omega \in \Omega \mid X(\omega) \leq a\} \in \Sigma .$$

Following the custom in probability theory we write this set (event) briefly as $\{X \leq a\}$ or as $(X \leq a)$ or just as $X \leq a$. We use this compact notation also for other events defined in terms of values of random variables.

Exercise. Show that the above definition is equivalent to the variant with the strict inequality $X(\omega) < a$. Hint: see the proof of the next proposition. \square

In probability theory and in this course we work with sums and products of random variables, so we better check that these are again random variables.

Proposition (sums and products of RVs). *If X and Y are random variables on a probability space (Ω, Σ, \Pr) , then their (pointwise) sum and product*

$$X + Y \quad \text{and} \quad XY$$

are again random variables on the same probability space.

Proof. We show that for every $a \in \mathbb{R}$ the set $\{X + Y < a\} \in \Sigma$, which by the previous exercise proves that $X + Y$ is a RV. This follows at once from the fact that (Ω, Σ) is a σ -algebra, from the assumption that X and Y are random variables, and from the expression

$$\{X + Y < a\} = \bigcup_{\substack{b, c \in \mathbb{Q} \\ b + c < a}} \{X < b\} \cap \{Y < c\} .$$

The proof for the product function XY is similar (but a little bit more complicated because of signs of products of two numbers) and we leave it to the reader. \square

¹We do not review everything, for indicator RVs and linearity of expectation see Lecture 2, and for the inequalities of Markov and of Čebyšev (Chebyshev) see Lectures 4 and 5, respectively.

Exercise. Prove that if $f: \mathbb{R} \rightarrow \mathbb{R}$ is a continuous function and X is a random variable on a probability space then also $Y := f \circ X = f(X)$ is a random variable on the same probability space. \square

We say that random variables X_1, X_2, \dots, X_n are *independent*, if for every $a_1, \dots, a_n \in \mathbb{R}$ and every set $I \subset [n]$ we have that

$$\Pr\left(\bigcap_{i \in I} X_i \leq a_i\right) = \prod_{i \in I} \Pr(X_i \leq a_i) .$$

If (Ω, Σ, \Pr) is a probability space and U is any set, then a map $X: \Omega \rightarrow U$ is a *discrete random variable* if the image

$$X(\Omega) = \{X(\omega) \mid \omega \in \Omega\} \subset U$$

is an at most countable set (i.e. a finite or a countable set) and

$$\forall y \in U : X^{-1}(y) = \{\omega \in \Omega \mid X(\omega) = y\} \in \Sigma .$$

Thus we define general RVs as real-valued functions, but discrete RVs may have any values.

Exercise. Show that every real discrete RV $X: \Omega \rightarrow \mathbb{R}$ is a RV by the initial definition. \square

In our course we deal most of the time only with discrete RVs. It is easy to see that if

$$C = \{c_i \in [0, 1] \mid i \in I\}$$

is an at most countable collection of constants, then a discrete random variable X on some probability space (Ω, Σ, \Pr) exists such that $X(\Omega) = I$ and $\Pr(X = i) = c_i$ for every $i \in I$ if and only if $\sum_{i \in I} c_i = 1$.

For a real discrete random variable $X: \Omega \rightarrow \mathbb{R}$ on a probability space (Ω, Σ, \Pr) its *expected value (expectation, mean)* $\mathbb{E}X$ is defined as the sum of the series

$$\mathbb{E}X := \sum_{c \in X(\Omega)} c \cdot \Pr(X = c)$$

if the series absolutely converges, else $\mathbb{E}X$ does not exist.² Recall that *absolute convergence* of a series $\sum_{i \in I} a_i$, where each $a_i \in \mathbb{R}$ and I is

²The expectation of a general RV X is defined as the integral $\mathbb{E}X := \int_{\Omega} X(\omega) d\Pr$. Since it takes some effort to say in this generality what precisely the f is, we mention it here only in footnote.

an at most countable set, means that there is a constant $c > 0$ such that for every finite set $J \subset I$,

$$\sum_{i \in J} |a_i| < c .$$

If this condition holds and I is countable (the case of finite I is trivial), then for any ordering $I = (i_1, i_2, \dots)$ of I in a sequence (i.e. each $i_n \in I$ and every $i \in I$ appears exactly once in the sequence) the limit

$$\sum_{i \in I} a_i := \lim_{n \rightarrow \infty} (a_{i_1} + a_{i_2} + \dots + a_{i_n}) \in \mathbb{R}$$

exists, does not depend on the ordering, and is called the *sum of the series*.

It is not hard to see that independence of real discrete random variables X_1, \dots, X_n is equivalent to satisfaction, for every $c_i \in \mathbb{R}$ and every set $I \subset [n]$, of the equality

$$\Pr\left(\bigcap_{i \in I} X_i = c_i\right) = \prod_{i \in I} \Pr(X_i = c_i) .$$

One can prove the next theorem the proof but we skip its proof (it is a result on absolutely convergent series).

Theorem (on independent discrete RVs). *Let X_1, X_2, \dots, X_n be independent real discrete RVs such that each mean $\mathbb{E}X_i$ exists. Then the product*

$$X := X_1 X_2 \dots X_n$$

is a real discrete RV, its mean $\mathbb{E}X$ exists, and

$$\mathbb{E}X = \prod_{i=1}^n \mathbb{E}X_i .$$

See Lecture 2 or Matoušek and Vondrák (see below) for the case when $n = 2$ and the X_i have finite ranges.

We conclude our review of RVs with an example of a concrete and important distribution. A *distribution* of a discrete random variable X is the list of probabilities

$$(\Pr(X = y) \mid y \in X(\Omega)) .$$

Example (binomial distribution). We say that a discrete RV X has the *binomial distribution* with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$, and write $X = B(n, p)$, if $X(\Omega) = [n]_0 := \{0, 1, \dots, n\}$ and

$$\Pr(X = i) = \binom{n}{i} p^i (1 - p)^{n-i} .$$

It is a consistent definition as by the binomial theorem these nonnegative numbers sum up to $(p + (1 - p))^n = 1$. Such RV X is realized, for example, in the probability space of 2^n series of n independent flips of a coin with $\Pr(\text{head}) = p$, as the number of heads obtained in a particular series. We compute the mean of $B(n, p)$, its second moment, and its variance. By the definition of expectation and the binomial theorem,

$$\begin{aligned} \mathbb{E}X &= \sum_{i=0}^n i \cdot \binom{n}{i} p^i (1 - p)^{n-i} = \sum_{i=1}^n i \cdot \frac{n}{i} \binom{n-1}{i-1} p^i (1 - p)^{n-i} \\ &= pn \sum_{i=1}^n \binom{n-1}{i-1} p^{i-1} (1 - p)^{n-1-(i-1)} = pn \cdot (p + (1 - p))^{n-1} \\ &= pn . \end{aligned}$$

This is a computation of somebody who is unaware of the probabilistic method. We know it and notice that the X realized by n flips of the coin is a sum $X = X_1 + X_2 + \dots + X_n$ of random variables X_i , where X_i is the indicator RV of the event of head in the i -th flip, and compute more easily by linearity of expectation that indeed

$$\mathbb{E}X = \sum_{i=1}^n \mathbb{E}X_i = n\mathbb{E}X_1 = n(1 \cdot p + 0 \cdot (1 - p)) = np .$$

Similarly, now using also independence of the X_i s and the fact that $X_i^2 = X_i$,

$$\begin{aligned} \mathbb{E}X^2 &= \mathbb{E}(X_1 + \dots + X_n)^2 \\ &\stackrel{\text{lin. of exp.}}{=} \sum_{i=1}^n \mathbb{E}X_i^2 + 2 \sum_{i < j} \mathbb{E}(X_i X_j) \\ &\stackrel{\text{ind. of } X_i \text{ and } X_j}{=} n\mathbb{E}X_1 + n(n-1)\mathbb{E}X_1 \cdot \mathbb{E}X_2 \\ &= np + n(n-1)p^2 = np(1 + (n-1)p) . \end{aligned}$$

Thus $\text{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = np(1 - p)$. □

Chernoff bounds

After these preparations (which were necessary so that the next exposition be rigorous and complete) we can finally turn to Chernoff bounds. I follow the lecture notes J. Matoušek and J. Vondrák, *The Probabilistic Method*, 71 pp., which are available on-line.

- *Jiří Matoušek (1963–2015)* was a Czech mathematician and computer scientist specializing in the area of computational and discrete geometry. He contributed to many other areas like discrepancy, algebraic topology or linear programming. He was lecturer's colleague in the Department of Applied Mathematics of MFF UK.
- *Jan Vondrák* is a Czech theoretical computer scientist, alumnus of the Department of Applied Mathematics of MFF UK, and today an associate professor at the Stanford University in the USA. The Iranian-American mathematician Maryam Mirzakhani (1977–2017), the first female Fields medal laureate (in 2014), was his wife.
- *Herman Chernoff (born 1923)* is, by the Wikipedia, an American applied mathematician, statistician and physicist.

By Chernoff (or Hoeffding) type bounds one understands estimates of (very small) probability of (large) deviation from mean of a RV formed as a sum of independent, or almost independent, RVs. Matoušek and Vondrák note that the next result, the only one of this type we prove here (see Appendix A in Alon and Spencer for many more of them), is much older.

Theorem (S. Bernštejn, 1924). *Let $n \in \mathbb{N}$, $X_i \in \{-1, 1\}$ for $i = 1, 2, \dots, n$ be independent discrete RVs which attain the values -1 and 1 with equal probability $\frac{1}{2}$, and let $X := X_1 + X_2 + \dots + X_n$ be their sum. Then for every real $t \geq 0$ one has that*

$$\Pr(X \geq t) \leq e^{-t^2/2n} \quad \text{and} \quad \Pr(X \leq -t) \leq e^{-t^2/2n} .$$

Proof. We prove only the first bound, the second follows by symmetry; but see the exercise below. We set $Y := e^{uX}$ for some real $u \geq 0$ to be specified later. By the previous proposition and the exercise after it, this is a RV. Now $\Pr(X \geq t) = \Pr(Y \geq e^{ut})$. By Markov's inequality,

$$\Pr(Y \geq e^{ut}) \leq \frac{\mathbb{E}Y}{e^{ut}}.$$

But

$$\mathbb{E}Y = \mathbb{E}e^{u(X_1 + \dots + X_n)} = \prod_{i=1}^n \mathbb{E}e^{uX_i} = \left(\frac{e^u + e^{-u}}{2} \right)^n \leq e^{nu^2/2}.$$

In the second equality we used the basic property of the exponential function and the assumption that the X_i (and hence the e^{uX_i}) are independent. The third equality follows from the assumption on distribution of X_i and the last inequality follows from the Taylor series of e^x . Thus

$$\Pr(Y \geq e^{ut}) \leq e^{-ut + nu^2/2}.$$

Setting $u = \frac{t}{n}$ we get that

$$\Pr(X \geq t) \leq e^{-t^2/n + t^2/2n} = e^{-t^2/2n}.$$

□

Exercise. Prove that in the previous theorem one has in fact for every $t \geq 0$ equal probabilities

$$\Pr(X \geq t) = \Pr(X \leq -t).$$

□

- *Sergej N. Bernštejn (1880–1968)* was a Russian and Soviet mathematician who contributed to the areas of partial differential equations, differential geometry, probability theory, and approximation theory.

An obvious application of the theorem is to independent coin tosses, because this is what the RVs X_i are (more precisely, the independent coin tosses corresponding to them are $X_i^{-1}(1)$, $i = 1, 2, \dots, n$). For $n \in \mathbb{N}$ the RV X records the score after n independent flips of a fair coin, if each head counts for +1 point and each tail for −1 point.

What is the chance that the score deviates much from 0? By the theorem we get that, for example,

$$\Pr(|X| \geq n/4) \leq 2e^{-n/32} \quad \text{or} \quad \Pr(|X| \geq 10\sqrt{n}) \leq 2e^{-50} .$$

These are much, much stronger bounds than what one can get from Čebyšev's (Chebyshev's) inequality. We start the next lecture by a result that quantifies deviations like this for *infinitely* many coin tosses. We conclude the present lecture by presenting from Matoušek and Vondrák another Chernoff type bound, without proof.

Theorem. *Let X_1, X_2, \dots, X_n be independent RVs with $X_i \in [0, 1]$ (these need not be discrete RVs), $X = \sum_i X_i$, and $\sigma^2 = \text{Var}(X) = \sum_i \text{Var}(X_i)$. Then for any real $t \geq 0$,*

$$\Pr(X \geq \mathbb{E}X + t) < e^{-\frac{t^2}{2(\sigma^2+t/3)}} \quad \text{and} \quad \Pr(X \leq \mathbb{E}X - t) < e^{-\frac{t^2}{2(\sigma^2+t/3)}} .$$

Thank you!

(final version of January 13, 2021)