**Combinatorics of Computation** Spring 2026

# Concentration Bounds

*Prof. Jara Uitto* *Antonio Cruciani*

In many randomized processes, the expected value alone is not enough: we also want to know that the random outcome is very close to its expectation with high probability. Concentration inequalities provide exactly this kind of guarantee. They are a central tool in the analysis of randomized algorithms.

In this lecture notes we are going to play with some of the concentration bounds discussed in class and some other applications that can be useful to understand their importance.

**Definition 1** (With high probability)**.** *An event $E_n$ holds* with high probability *(w.h.p.) if*

$$\Pr(E_n) \geq 1 - \frac{1}{n^c}$$

*for some constant $c > 0$ and all sufficiently large $n$.*

In other words, a w.h.p. statement says that the probability of failure goes to zero as the input size grows.

**Remark.** A statement that holds with high probability is stronger than a statement that holds with constant probability, but weaker than a statement that holds surely. In randomized algorithm design, w.h.p. guarantees are often strong enough to treat failures as negligible.

## 1 Markov's and Chebyshev's inequalities

We begin with two very general tail bounds. They apply to a wide range of random variables, but this generality comes at a cost: the bounds are often not very sharp.

**Lemma 2** (Markov's Inequality)**.** *For a non-negative random variable $X$, it holds that, for every $t > 0$,*

$$\Pr(X \geq t) \leq \frac{\mathbf{E}[X]}{t}.$$

Markov's inequality uses only one piece of information about the random variable, namely its expectation. Because of this, it is extremely general, but often quite weak. In particular, it is most useful as a first crude bound, or when the expectation is the only quantity we know how to control.

**Lemma 3** (Chebyshev's Inequality)**.** *Let $X$ be a random variable with $\mathbf{E}[X] = \mu$ and variance $\sigma^2$. Then for all $t > 0$, it holds that*

$$\Pr\left(|X - \mu| \geq t\right) \leq \frac{\sigma^2}{t^2}.$$

Chebyshev's inequality improves on Markov's inequality by also using the variance, and therefore captures how concentrated the random variable is around its mean. It is still completely general and does not require independence. However, the decay is only quadratic in $t$, so for sums of independent random variables one can often do much better using sharper concentration bounds such as Chernoff or Hoeffding.

## 2 Chernoff Bounds

In general Markov's Inequality is tight. However, it is particularly loose for sums of independent random variables. Here Chernoff bounds step in and usually give much stronger statements. There are plenty of variants of Chernoff bounds, we refer to [1, 2] for an extensive list of results. Here, we prove the following version.

**Lemma 4** (Chernoff Bound)**.** *Let* $X_1, \ldots, X_n$ *be independent indicator (or Bernoulli) random variables and let* $X = \sum_{i=1}^n X_i$ *be their sum. Moreover, let* $\mu \geq \mathbf{E}[X]$ *and* $\delta > 0$ *be constants. Then*

$$\Pr\left(X \geq (1+\delta)\mu\right) \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu .$$

Before we prove the lemma let us briefly consider what it states for the load balancing problem described above. Fix one machine and let $X_i$ be the indicator random variable being 1 if job $i$ is assigned to that machine. Then for $X = \sum_{i=1}^n X_i$, we get $E[X] = 1$ and thus for $\mu = 1$ and $\delta = 2\log n$, the lemma gives that the probability that there are more than $2\log n + 1 = O(\log n)$ jobs assigned to that machine is bounded by

$$\frac{e^\delta}{(1+\delta)^{1+\delta}} \leq \left(\frac{e}{1+\delta}\right)^\delta \leq \left(\frac{e}{2\log n}\right)^\delta \leq \left(\frac{1}{2}\right)^\delta = \frac{1}{n^2}.$$

So we can get a quadratic upper bound. This is good enough so that we can use the union bound over all machines and get a probability of at most $1/n$ that any machine is assigned more than $2\log n + 1$ jobs. We will see below that we can get an even better bound. Let us first prove the lemma.

*Proof of Lemma 4.* We define the random variable $\exp(t \cdot X)$ for some constant $t$ that we will fix later on. Note that $\exp(t \cdot X)$ is a non-negative random variable and thus Markov's inequality gives

$$\Pr\left(X \geq (1+\delta)\mu\right) = \Pr\left(\exp(t \cdot X) \geq \exp(t(1+\delta)\mu)\right) \leq \frac{\mathbf{E}[\exp(t \cdot X)]}{\exp(t(1+\delta)\mu)}.$$

We now bound $\mathbf{E}[\exp(t \cdot X)]$. To this end we observe that

$$\mathbf{E}[\exp(t \cdot X)] = \mathbf{E}\left[\exp\left(t\sum_{i=1}^n X_i\right)\right] = \mathbf{E}\left[\prod_{i=1}^n \exp(tX_i)\right] = \prod_{i=1}^n \mathbf{E}[\exp(t \cdot X_i)],$$

using independence of the $X_i$ variables. Now recall that $X_i$ was an indicator random variable and let us define $p_i = \Pr\left(X_i = 1\right)$. Then

$$\mathbf{E}[\exp(tX_i)] = p_i \cdot e^t + (1 - p_i) \cdot 1 = p_i \cdot (e^t - 1) + 1.$$

We now set $t = \ln(1+\delta)$ and obtain $\mathbf{E}[\exp(tX_i)] = p_i\delta + 1 \leq \exp(p_i\delta)$ using that $1 + x \leq e^x$. Hence, for the expectation of $\exp(tX)$, we get

$$\mathbf{E}[\exp(tX)] \leq \exp\left(\delta \sum_{i=1}^{n} p_i\right) = \exp\left(\delta \sum_{i=1}^{n} \mathbf{E}[X_i]\right) = \exp(\delta \mathbf{E}[X]) \leq \exp(\delta\mu).$$

Together with the definition of $t$, we obtain

$$\Pr\left(X \geq (1+\delta)\mu\right) \leq \frac{\exp(\delta\mu)}{\exp(\ln(1+\delta)(1+\delta)\mu)} = \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}.$$

□

**A simple example: coin flips.** Consider throwing a fair coin $n$ times and counting the number of heads we obtain. Let $X_i \in \{0,1\}$, $X = \sum_{i=1}^{n} X_i$ and $\mathbf{E}[X] = \frac{n}{2}$.

The Chernoff bound gives for any $\delta > 0$,

$$\Pr\left(X \geq (1+\delta)(n/2)\right) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{n/2}$$

This expression equals 1 only for $\delta = 0$, and then gives a value strictly less than 1 (check it as a warm-up.)

Notice that the inequality is *exponential in $n$* (for fixed $\delta$) which is much better than what we can get using Chebyshev's and Markov's inequalities. To visualize this, let's plug in some numbers.

Consider $n = 100$ independent coin flips. We wish to find an upper bound on the probability that the number of heads is greater or equal than 75.

Let's start with Markov's inequality: $\mathbf{E}[X] = 100/2 = 50$

$$\Pr\left(X \geq \frac{3}{2}\mathbf{E}[X]\right) \leq \frac{2}{3} = 0.666$$

Instead for Chebyshev's we have that $\mathbf{Var}[X] = \sum_{i=1}^{100} \mathbf{Var}[X_i] = 100 \cdot (1/2)^2 = 25$.

$$\Pr\left(|X - \mathbf{E}[X]| \geq t\right) \leq \frac{\mathbf{Var}[X]}{t^2}$$

and plugging in $t = 25$ fives an upper bound of $25/25^2 = 1/25 = 0.04$.

Finally, we use the above Chernoff bound. We set $\delta = 1/2$

$$\Pr\left(X \geq \frac{3}{2}\mathbf{E}[X]\right) \leq \left(\frac{e^{1/2}}{(3/2)^{3/2}}\right)^{50} \approx 0.004472$$

## 3 Hoeffding's Inequality

Chernoff bounds are very useful when we study multiplicative deviations of sums of independent Bernoulli random variables. However, in many applications we want an *additive* guarantee instead. For example, if we estimate the fraction of people in a population having some binary property, we may want our estimate to be within $\varepsilon$ of the true fraction. For this purpose, Hoeffding's inequality is often more convenient.

**Lemma 5** (Hoeffding's Inequality)**.** *Let $X_1, \dots, X_n$ be independent random variables such that*

$$a_i \le X_i \le b_i \qquad \text{for all } i \in [n].$$

*Then, for every $t > 0$,*

$$\Pr\left(\left|\sum_{i=1}^n X_i - \mathbf{E}\left[\sum_{i=1}^n X_i\right]\right| \ge t\right) \le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

**Corollary 6.** *Let $X_1, \dots, X_n$ be independent Bernoulli random variables with*

$$\Pr(X_i = 1) = p,$$

*and let*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i.$$

*Then, for every $\varepsilon > 0$,*

$$\Pr\left(|\overline{X} - p| \ge \varepsilon\right) \le 2\exp(-2n\varepsilon^2).$$

*Proof.* Since each $X_i$ is Bernoulli, we have

$$0 \le X_i \le 1 \qquad \text{for all } i.$$

Moreover,

$$\mathbf{E}[\overline{X}] = \frac{1}{n}\sum_{i=1}^n \mathbf{E}[X_i] = p.$$

Applying Hoeffding's inequality to the sum $X = \sum_{i=1}^n X_i$, we obtain

$$\Pr\left(|X - \mathbf{E}[X]| \ge \varepsilon n\right) \le 2\exp\left(-\frac{2(\varepsilon n)^2}{n}\right) = 2\exp(-2n\varepsilon^2).$$

Since $X = n\overline{X}$ and $\mathbf{E}[X] = pn$, this is equivalent to

$$\Pr\left(|\overline{X} - p| \ge \varepsilon\right) \le 2\exp(-2n\varepsilon^2).$$

$\square$

Suppose we want to estimate the fraction $p$ of people in a large population who hold some binary opinion. We sample $n$ people independently and uniformly at random with replacement. For each $i \in [n]$, let

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th sampled person has the opinion} \\ 0 & \text{otherwise} \end{cases}$$

Then $\overline{X}$ is the observed fraction of sampled people having the opinion, and the theorem below shows how many samples are sufficient to estimate $p$ within additive error $\varepsilon$ with confidence at least $1 - \delta$.

**Theorem 7.** *Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with*

$$\Pr(X_i = 1) = p,$$

*and let*

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

*If*

$$n \geq \frac{1}{2\varepsilon^2} \ln\left(\frac{2}{\delta}\right),$$

*then*

$$\Pr\left(|\overline{X} - p| \leq \varepsilon\right) \geq 1 - \delta$$

*for all $\varepsilon, \delta \in (0, 1)$.*

*Proof.* By the previous corollary,

$$\Pr\left(|\overline{X} - p| \geq \varepsilon\right) \leq 2\exp(-2n\varepsilon^2).$$

Thus it is enough to require

$$2\exp(-2n\varepsilon^2) \leq \delta.$$

Taking logarithms and rearranging gives

$$n \geq \frac{1}{2\varepsilon^2} \ln\left(\frac{2}{\delta}\right).$$

□

Thus, $\overline{X}$ is an $(\varepsilon, \delta)$-approximation of $p$ in the additive sense: with probability at least $1 - \delta$, the estimate differs from $p$ by at most $\varepsilon$.

## 3.1 Degree Splitting

In the degree splitting problem, we are given a graph $G = (V, E)$ and we want to partition the edges into two classes, for example by coloring each edge *red* or *blue*, in such a way that at every vertex the two colors are as evenly balanced as possible.

Formally, after coloring the edges, for each vertex $v \in V$ let

$$\deg_R(v) = \#\{\text{red edges incident to } v\}, \qquad \deg_B(v) = \#\{\text{blue edges incident to } v\}.$$

Since every edge incident to $v$ is colored either red or blue, we have

$$\deg_R(v) + \deg_B(v) = \deg(v).$$

Thus, the ideal situation would be

$$\deg_R(v) \approx \deg_B(v) \approx \frac{\deg(v)}{2} \qquad \text{for every } v \in V.$$

In general, an exact split may not always be possible. For example, if $\deg(v)$ is odd, then the two quantities $\deg_R(v)$ and $\deg_B(v)$ cannot both be exactly equal to $\deg(v)/2$. Therefore, the goal is usually to find a coloring such that, for every vertex $v$, the difference

$$|\deg_R(v) - \deg_B(v)|$$

is small.

A natural randomized strategy is to color each edge independently red or blue with probability $1/2$. For a fixed vertex $v$, the number of red incident edges is then a binomial random variable with expectation $\deg(v)/2$, so concentration inequalities such as Chernoff bounds suggest that the imbalance at $v$ should be small with high probability. We will use this idea to show that, if the minimum degree is sufficiently large, then a random coloring gives a good degree splitting with high probability.

For this problem, we will use the following Chernoff bound

**Lemma 8** (Chernoff bound)**.** *Let $X = \sum_{i=1}^{m} X_i$, where $X_1, \ldots, X_m$ are independent Bernoulli random variables, and let $\mu = \mathbf{E}[X]$. Then for every $0 < \varepsilon \leq 1$,*

$$\Pr(|X - \mu| > \varepsilon\mu) \leq 2\exp\left(-\frac{\varepsilon^2\mu}{3}\right).$$

**Theorem 9.** *Let $G = (V, E)$ be a graph on $n$ vertices, and let*

$$d_{\min} = \min_{v \in V} \deg(v).$$

*Color each edge independently red or blue with probability $1/2$ each. Then, for every $\varepsilon \in (0, 1]$, with probability at least*

$$1 - 2n\exp\left(-\frac{\varepsilon^2 d_{\min}}{6}\right),$$

*every vertex $v \in V$ satisfies*

$$\deg_R(v), \deg_B(v) \in \left[(1-\varepsilon)\frac{\deg(v)}{2},\ (1+\varepsilon)\frac{\deg(v)}{2}\right].$$

*Moreover, if*

$$d_{\min} \geq C\frac{\log n}{\varepsilon^2}$$

*for a sufficiently large constant $C > 0$, then with high probability every vertex $v \in V$ satisfies*

$$\deg_R(v), \deg_B(v) \in \left[(1-\varepsilon)\frac{\deg(v)}{2},\ (1+\varepsilon)\frac{\deg(v)}{2}\right].$$

*Proof.* Fix a vertex $v \in V$. Recall that $\deg_R(v)$ denotes the number of red edges incident to $v$. Since each edge incident to $v$ is colored red independently with probability $1/2$, the random variable

$$\deg_R(v)$$

has binomial distribution:

$$\deg_R(v) \sim \mathrm{Bin}\left(\deg(v), \frac{1}{2}\right).$$

Therefore,

$$\mathbf{E}[\deg_R(v)] = \frac{\deg(v)}{2}.$$

We now apply the multiplicative Chernoff bound. Since $\varepsilon \in (0, 1]$,

$$\begin{aligned}\Pr\left(\left|\deg_R(v) - \frac{\deg(v)}{2}\right| > \varepsilon\frac{\deg(v)}{2}\right) &\le 2\exp\left(-\frac{\varepsilon^2}{3} \cdot \frac{\deg(v)}{2}\right)\\ &= 2\exp\left(-\frac{\varepsilon^2 \deg(v)}{6}\right).\end{aligned}$$

Because $\deg(v) \ge d_{\min}$, this implies

$$\Pr\left(\left|\deg_R(v) - \frac{\deg(v)}{2}\right| > \varepsilon\frac{\deg(v)}{2}\right) \le 2\exp\left(-\frac{\varepsilon^2 d_{\min}}{6}\right).$$

Now take a union bound over all vertices $v \in V$. The probability that there exists at least one vertex for which the desired bound fails is at most

$$\begin{aligned}\Pr\left(\exists v \in V \,:\, \left|\deg_R(v) - \frac{\deg(v)}{2}\right| > \varepsilon\frac{\deg(v)}{2}\right) &\le \sum_{v \in V} \Pr\left(\left|\deg_R(v) - \frac{\deg(v)}{2}\right| > \varepsilon\frac{\deg(v)}{2}\right)\\ &\le 2n\exp\left(-\frac{\varepsilon^2 d_{\min}}{6}\right).\end{aligned}$$

Hence, with probability at least

$$1 - 2n\exp\left(-\frac{\varepsilon^2 d_{\min}}{6}\right),$$

every vertex $v \in V$ satisfies

$$\left|\deg_R(v) - \frac{\deg(v)}{2}\right| \le \varepsilon\frac{\deg(v)}{2}.$$

This is equivalent to

$$\deg_R(v) \in \left[(1-\varepsilon)\frac{\deg(v)}{2},\ (1+\varepsilon)\frac{\deg(v)}{2}\right].$$

Furthermore, since

$$\deg_B(v) = \deg(v) - \deg_R(v),$$

the same bound also holds for the blue degree:

$$\deg_B(v) \in \left[(1-\varepsilon)\frac{\deg(v)}{2},\ (1+\varepsilon)\frac{\deg(v)}{2}\right].$$

Moreover, for every constant $c > 0$, if

$$d_{\min} \geq 6(c+1)\frac{\log n}{\varepsilon^2},$$

then

$$\Pr\left(\exists v \in V : \left|\deg_R(v) - \frac{\deg(v)}{2}\right| > \varepsilon\frac{\deg(v)}{2}\right) \leq \frac{2}{n^c}.$$

In particular, if $d_{\min} = \Omega(\log n/\varepsilon^2)$, then the desired bound holds with high probability. □

## 4 Load Balancing

Assume that we have $n$ machines and $m$ jobs that we want to assign. If there is a central entity that coordinates everything we can achieve an assignment such that each machine gets at most $\lceil m/n \rceil$ jobs. We will now restrict our attention to the case where $m = n$. What can we do if we cannot coordinate? We can just randomly assign the jobs to machines. What is the maximum number of jobs that end up being assigned to one machine? We call this the maximum load, formally what is $\mathbf{E}[\max_{1\leq i\leq n} L_i]$, if $L_i$ is the load of machine $i$, i.e., the number of jobs assigned to machine $i$. We could try using the union bound to bound

$$\Pr\left(\max_{1\leq i\leq n} L_i \geq t\right) \leq \sum_{1\leq i\leq n} \Pr\left(L_i \geq t\right).$$

At this point we can apply Markov's inequality. We obtain a bound of

$$\Pr\left(\max_{1\leq i\leq n} L_i \geq t\right) \leq \frac{n}{t}.$$

This bound however is meaningless, as it only gives a probability less than 1 if $t > n$. Thus, we could, for example, conclude that the probability that more than $2n$ jobs are assigned to one machine is less than 1/2. As we have only $n$ jobs in total, this is not surprising. We will need to use a better tool than just Markov's inequality.

**Stronger bound for Load Balancing** Let us now show a stronger bound for the Load Balancing problem. First, we provide an Upper Bound.

**Theorem 10.** *The maximum load per machine is $\mathcal{O}(\log n/\log\log n)$ with high probability.*

*Proof.* Fix one machine and let $X_i$ be the indicator random variable defined as follows

$$X_i = \begin{cases} 1 & \text{If job } i \text{ is assigned to the fixed machine} \\ 0 & \text{Otherwise} \end{cases}$$

And let $X = \sum_{i=1}^{m} X_i$ be the number of jobs assigned to the fixed machine. Then $\mathbf{E}[X] = 1$ and thus for $\mu = 1$ and any $\delta > 0$, Lemma 4 gives that

$$\Pr\left(X \geq 1+\delta\right) \leq \frac{e^\delta}{(1+\delta)^{1+\delta}} = \frac{1}{e}\left(\frac{e}{1+\delta}\right)^{1+\delta}.$$

Now let $c > 0$ be arbitrary, choosing $\delta = \max\left\{e\sqrt{\ln n}, 2(c+1)\frac{\ln n}{\ln\ln n}\right\} - 1$ gives

$$\Pr\left(X \geq 1+\delta\right) \leq \frac{1}{e}\left(\frac{e}{e\sqrt{\ln n}}\right)^{2(c+1)\frac{\ln n}{\ln\ln n}} \leq \left(\frac{1}{\ln n}\right)^{\frac{(c+1)\ln n}{\ln\ln n}} = \frac{1}{n^{c+1}}.$$

The union bound over all $n$ machines gives the result with probability at most $1/n^c$. □

We now continue and provide a lower bound for the problem. Moreover, to show a lower bound we need the following basic concepts. Let the variance of a random variable be defined as $\mathbf{Var}(X) = \mathbf{E}[(X - \mathbf{E}[X])^2] = \sigma^2$, and the covariance of two random variables be defined as $\mathbf{Cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$.

**Observation 11.** *The variance of a Bernoulli random variable $X \sim Be(p)$ is*

$$\mathbf{Var}[X] = p(1-p).$$

**Observation 12.** *For random variables $Y_1, \ldots, Y_n$, let $Y = \sum_{j=1}^{n} Y_j$. Then it holds that*

$$\mathbf{Var}[Y] = \sum_{j=1}^{n} \mathbf{Var}[Y_j] + \sum_{j=1}^{n}\sum_{i\neq j} \mathbf{Cov}[Y_j, Y_i].$$

Recall Chebyshev's inequality.

**Lemma 13.** *Let $X$ be a random variable with $\mathbf{E}[X] = \mu$ and variance $\sigma^2$. Then for all $t > 0$, it holds that*

$$\Pr\left(|X - \mu| \geq t\right) \leq \frac{\sigma^2}{t^2}.$$

We now have all the tools to prove that the upper bound is also a lower bound four our problem with constant probability.

**Theorem 14.** *The maximum load per machine is $\Omega(\log n/\log\log n)$ with constant probability.*

*Proof.* The probability that machine $j$ gets exactly $k$ jobs is

$$\Pr\left(L_j = k\right) = \binom{n}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{n-k} \geq \left(\frac{n}{k}\right)^k\left(\frac{1}{n}\right)^k\frac{1}{e} = \frac{1}{ek^k}.$$

For $\left(k = \frac{\ln n}{3\ln\ln n}\right)$, we have $k^k \leq (\ln n)^{\frac{\ln n}{3\ln\ln n}} = n^{1/3}$. Now define random variables $Y_j$ as follows. We let

$$Y_j = \begin{cases} 1 & \text{If machine } j \text{ gets at least } k \text{ jobs} \\ 0 & \text{Otherwise} \end{cases}$$

Moreover, we let $Y = \sum_{j=1}^{n} Y_j$. Then

$$\Pr\left(Y_j = 1\right) \geq \Pr\left(L_j = k\right) \geq \frac{1}{en^{1/3}},$$

and thus the expected number of machines with at least $k$ jobs is

$$\mathbf{E}[Y] = \mathbf{E}[\sum_{j=1}^{n} Y_j] \geq \frac{n^{2/3}}{e}.$$

Although one might think so, this does not give us any information about

$$\Pr(Y \geq 1) = \Pr\left(\max_j L_j \geq k\right)$$

since the expectation can be large only because with small probability the random variable takes a very high value. In order to get an upper bound on $\Pr(Y = 0)$ we will now use Chebyshev's inequality.

If $Y = 0$, this implies also $|Y - \mu| \geq \mu$ for $\mu = \mathbf{E}[Y]$. Thus

$$\Pr(Y = 0) \leq \Pr(|Y - \mu| \geq \mu) \leq \frac{\sigma^2}{\mu^2},$$

where $\sigma^2$ is the variance of $Y$. It remains to upper bound $\sigma^2$. Using Observation 2.3, we get

$$\sigma^2 = \mathbf{Var}[Y] = \sum_{j=1}^{n} \mathbf{Var}[Y_j] + \sum_{j=1}^{n} \sum_{i \neq j} \mathbf{Cov}[Y_i, Y_j].$$

For the variances, from Observation 11 we have

$$\mathbf{Var}[Y_j] = \Pr(Y_j = 1)(1 - \Pr(Y_j = 1)) \leq \mathbf{E}[Y_j].$$

For the covariances, from Observation 12 we can conclude that

$$\begin{aligned}
\mathbf{Cov}(Y_i, Y_j) &= \mathbf{E}[Y_i Y_j] - \mathbf{E}[Y_i]\mathbf{E}[Y_j] \\
&= \Pr(Y_i = 1 \wedge Y_j = 1) - \Pr(Y_i = 1)\Pr(Y_j = 1) \\
&= \Pr(Y_i = 1)\Pr(Y_j = 1 \mid Y_i = 1) - \Pr(Y_i = 1)\Pr(Y_j = 1) \\
&= \Pr(Y_i = 1)\Big(\Pr(Y_j = 1 \mid Y_i = 1) - \Pr(Y_j = 1)\Big).
\end{aligned}$$

**Fact 15.** *For $i \neq j$, the indicator variables $Y_i$ and $Y_j$ are negatively correlated:*

$$\mathbf{Cov}(Y_i, Y_j) \leq 0.$$

In summary,

$$\Pr(Y = 0) \leq \frac{\sigma^2}{\mu^2} \leq \frac{1}{\mu} \leq \frac{e}{n^{2/3}},$$

and thus already for $n \geq (2e)^{3/2}$, we get that

$$\Pr\left(\max_j L_j \geq k\right) \geq 1/2$$

□

# References

[1] Mitzenmacher, Michael and Upfal, Eli, *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis, Cambridge university press, 2017*,

[2] Dubhashi, Devdatt P and Panconesi, Alessandro *Concentration of measure for the analysis of randomized algorithms*,