Lecture 2 — Analysis of Randomized Algorithms; Sorting & Searching

> Course **2-INF-135/15 Pravdepodobnostné algoritmy**, LS 2025/26. Source slides: `RA_02.pdf` (16 pages).

What this lecture is about

Lecture 1 was a *gallery* — many algorithms, each analyzed with whatever trick fit. This lecture stops to build the **standard toolbox** that every later analysis reuses, and then spends it on two case studies.

It has two halves:

1. **The probability toolkit** — three *tail inequalities* (Markov, Chebyshev, Chernoff) that answer the recurring question "*how unlikely is it that a random quantity strays far from its average?*" This is the property called **concentration** (koncentrácia).
2. **Sorting & searching** — two payoffs:
   - **Randomized QuickSort runs in $O(n \log n)$ not just *on average* but *with high probability*** — a much stronger guarantee than Lecture 1's expectation bound.
   - **Randomized Selection (R-Select)** finds the median (or any rank) in $2n + o(n)$ comparisons — *linear* time, beating the $n \log n$ of sorting.

The spine of the whole lecture is one idea: **the three inequalities form a ladder. Each rung assumes more about the random variable and pays you back with a sharper bound.** Get that ladder straight and everything else follows.

---

1. The probability vocabulary (recap, with the right names)

A quick refresher of the objects we keep using. Let X be a **random variable** (náhodná premenná), and let E_i be the event "$X = i$".

- **Expected value** (stredná hodnota), also written μ: $E[X] = \sum_i i \cdot \Pr[X = i]$.
- **Variance** (rozptyl) and **standard deviation** (odchýlka) σ:
  $\mathrm{Var}(X) = \sum_i (i - E[X])^2 \Pr[X = i] = E\big[(X - E[X])^2\big], \qquad \sigma = \sqrt{\mathrm{Var}(X)}$. Variance measures *how spread out* X is around its mean.

- **Union bound** (the most-used inequality in the course): $\Pr\left[\bigcup_i E_i\right] \le \sum_i \Pr[E_i]$. *When is it an equality?* Exactly when the events are **pairwise disjoint** (no overlap is double-counted).
- **Linearity of expectation** — the engine of Lecture 1, holds **with or without independence**: $E\left[\sum_i a_i X_i\right] = \sum_i a_i\, E[X_i]$.

Independence and what it unlocks

X, Y are **independent** (nezávislé) if
$\Pr[X = x, Y = y] = \Pr[X = x]\Pr[Y = y] \quad \iff \quad \Pr[X = x \mid Y = y] = \Pr[X = x]$. ($k$-independence is the same statement for every k-tuple — this is the notion Lecture 5's derandomization will weaken on purpose.)

Independence buys two things linearity does **not**:
$E[XY] = E[X]\,E[Y], \qquad \mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y].$

> **Watch the asymmetry.** Expectation *adds* for free (linearity, always). Variance *adds* only under independence. That single fact is why the whole Chebyshev machinery below needs independence and the linearity arguments of Lecture 1 did not.

Two estimates we will reach for constantly

$$1 + x \le e^x \quad (\forall x), \qquad\qquad \left(1 - \tfrac{1}{n}\right)^n \le \tfrac{1}{e} \le \left(1 - \tfrac{1}{n}\right)^{n-1}.$$

The first ($1 + x \le e^x$, equivalently $1 + y < e^y$) is the workhorse — it is exactly the step that turns a product into a clean exponential in the Chernoff proof.

"With high probability" (s vysokou pravdepodobnosťou, s.v.p.)

> An event happens **with high probability** if its probability is $1 - O\left(\frac{1}{n^c}\right)$ for some constant $c > 0$.

This is the target we aim for. The exponent c matters: if we later want to **union-bound over n bad events** (e.g. n root-to-leaf paths in a sort tree), each must fail with probability $\le 1/n^2$ so the *total* failure is $\le n \cdot 1/n^2 = 1/n$ — still high probability. Keep that bookkeeping in mind; it dictates *how* sharp a tail bound we need.

---

2. The running example — balls into boxes (guličky do krabíc)

To compare the three inequalities we keep one concrete question in front of us:

> Throw m balls independently and uniformly into n boxes. Let X = the number of balls in **one fixed box**. Then $E[X] = m/n$. **How likely is that box to be much fuller than average?**

X is **binomial**: $X \sim \mathrm{Bin}(m, 1/n)$, so
$E[X] = \frac{m}{n} = \mu, \qquad \mathrm{Var}(X) = m \cdot \frac{1}{n}\left(1 - \frac{1}{n}\right) = \mu\left(1 - \frac{1}{n}\right) < \mu.$

We will ask the *same* question — *what is* $\Pr[X \geq 2\mu]$*?* — of all three inequalities and watch the answer get sharper each time.

---

3. Markov's inequality — the root of everything

> **Theorem (Markov).** If $X \geq 0$ (takes only non-negative values), then for every $k > 0$
> $\Pr[X \geq k] \leq \frac{E[X]}{k}, \qquad \text{equivalently} \qquad \Pr[X \geq k\,E[X]] \leq \frac{1}{k}.$

Proof — one line of bookkeeping. Throw away every term below k and underestimate the rest by k: $E[X] = \sum_i i\,\Pr[X = i] \;\geq\; \sum_{i \geq k} k\,\Pr[X = i] = k \sum_{i \geq k} \Pr[X = i] = k\,\Pr[X \geq k]$**.** Divide by k. ■

That's the whole thing. Markov knows **only the mean** and **only that X can't go negative**. With so little information it can't say much — but it is the seed from which the other two grow.

On the balls: $\Pr[X \geq 2\mu] \leq \dfrac{\mu}{2\mu} = \dfrac{1}{2}$. Almost useless, but honest.

> **The deep point to say aloud.** Markov is *the* tail inequality. Chebyshev and Chernoff are not new ideas — they are **Markov applied to a cleverly transformed version of X.** Keep that in mind through the next two sections.

---

4. Chebyshev's inequality — Markov on the squared deviation

Markov is weak because it ignores the *shape* of X. Chebyshev feeds it one more number — the **variance** — and in return gets a **two-sided** bound that needs **no non-negativity assumption**.

> **Theorem (Chebyshev).** For any random variable X and every $k > 0$,
> $\Pr\left[\,|X - E[X]| \geq k\,\right] \leq \frac{\mathrm{Var}(X)}{k^2}, \qquad \text{equivalently} \qquad \Pr\left[\,|X - E[X]| \geq k\sigma\,\right] \leq \frac{1}{k^2}.$

The second form is the memorable one: *the probability of being k standard deviations off the mean is at most $1/k^2$.*

Proof — it really is just Markov. The deviation can be negative, so square it to make it non-negative, then apply Markov to $Y = (X - E[X])^2$:
$\Pr\left[\,|X - E[X]| \geq k\,\right] = \Pr\left[\,(X - E[X])^2 \geq k^2\,\right] \leq \frac{E[Y]}{k^2} = \frac{\mathrm{Var}(X)}{k^2}.$ The last equality is just the definition $E[(X - E[X])^2] = \mathrm{Var}(X)$. ■

On the balls. Use $\mathrm{Var}(X) = \mu(1 - 1/n) < \mu$:
$\Pr[X \geq 2\mu] \leq \Pr\left[\,|X - \mu| \geq \mu\,\right] \leq \frac{\mathrm{Var}(X)}{\mu^2} < \frac{\mu}{\mu^2} = \frac{1}{\mu} = \frac{n}{m}.$

> **Sharper already.** Markov gave $\frac{1}{2}$ no matter what; Chebyshev gives n/m, which is *tiny* when there are many balls per box ($m \gg n$). The price was one extra assumption — that we know (and can bound) the variance.

---

5. Chernoff bound — Markov on the exponential

For a **sum of independent** indicators we can do dramatically better. Knowing the tail of such a sum exactly is hopeless — the exact formula is a sum over all large subsets,
$\Pr[X \geq k] = \sum_{\substack{A \subseteq \{1,\ldots,n\} \\ |A| \geq k}} \prod_{i \in A} p_i \prod_{i \notin A} (1 - p_i),$ which has exponentially many terms. Chernoff replaces that monster with a clean **exponentially small** bound.

> **Theorem (Chernoff).** Let $X_1, \ldots, X_n$ be **independent** $0/1$ variables with $p_i = \Pr[X_i = 1]$, let $X = \sum_i X_i$ and $\mu = E[X] = \sum_i p_i$. Then:
> $\forall \delta > 0: \quad \Pr[X \geq (1+\delta)\mu] < \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu \quad (\star)$
> $\forall \delta \in (0,1): \quad \Pr[X \geq (1+\delta)\mu] \leq e^{-\mu\delta^2/3}$ $\forall \delta \in (0,1): \quad \Pr[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}$
> $\forall R \geq 6\mu: \quad \Pr[X \geq R] \leq 2^{-R}$

The middle two are the **usable everyday forms** — "the probability of being a δ-fraction off the mean decays like $e^{-\mu\delta^2}$." The key feature: the bound shrinks **exponentially in** μ, not polynomially.

The proof — the "MGF + Markov" trick (worth memorizing)

The recipe is three moves: **make it non-negative → exponentiate → Markov.**

Step 1 — exponentiate, then Markov. e^{tX} is non-negative for any $t > 0$, and $x \mapsto e^{tx}$ is increasing, so the event $X \geq (1+\delta)\mu$ is the same event as $e^{tX} \geq e^{t(1+\delta)\mu}$. Apply Markov to e^{tX}: $\Pr[X \geq (1+\delta)\mu] = \Pr\left[e^{tX} \geq e^{t(1+\delta)\mu}\right] \leq \frac{E[e^{tX}]}{e^{t(1+\delta)\mu}}$.

Step 2 — the moment generating function factorizes (here is where **independence** is spent): $E[e^{tX}] = E\left[e^{t\sum_i X_i}\right] = E\left[\prod_i e^{tX_i}\right] \overset{\text{indep.}}{=} \prod_i E[e^{tX_i}] = \prod_i \left(p_i e^t + (1-p_i)\right)$. Now use $1 + y < e^y$ with $y = p_i(e^t - 1)$: $\prod_i \left(1 + p_i(e^t-1)\right) < \prod_i e^{p_i(e^t-1)} = e^{(e^t-1)\sum_i p_i} = e^{\mu(e^t-1)}$.

Step 3 — optimize t**.** Substituting back, $\Pr[X \geq (1+\delta)\mu] < \frac{e^{\mu(e^t-1)}}{e^{t(1+\delta)\mu}} = e^{\left(e^t - 1 - t(1+\delta)\right)\mu}$. Minimize the exponent over t: the derivative gives $e^t = 1+\delta$, i.e. $t = \ln(1+\delta)$. Plugging in yields exactly $(\star)$: $\Pr[X \geq (1+\delta)\mu] < \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu$. ■ The friendlier forms ($e^{-\mu\delta^2/3}$ etc.) come from bounding this expression for δ in the stated ranges.

On the balls. With $\mu = m/n$ and $\delta = 1$: $\Pr[X \geq 2\mu] \leq e^{-\mu\delta^2/3} = e^{-m/(3n)}$.

> **The escalation, side by side** (same event, $\Pr[X \geq 2\mu]$ for the fullest box):
> $\underbrace{\frac{1}{2}}_{\text{Markov}} \gg \underbrace{\frac{n}{m}}_{\text{Chebyshev}} \gg \underbrace{e^{-m/(3n)}}_{\text{Chernoff}}$. Constant → polynomially small → **exponentially** small.
> Each rung cost one more assumption.

---

6. The ladder (the unifying idea — say this in the oral)

The three inequalities are **one idea at three resolutions.** Read the table top to bottom: each row assumes strictly more and pays back a strictly sharper tail.

Inequality	Needs	Mechanism	Tail decay
Markov	$X \geq 0$, the mean	— (direct)	$\sim 1/k$
Chebyshev	the variance	Markov on $(X-\mu)^2$	$\sim 1/k^2$
Chernoff	independence of a sum	Markov on e^{tX}	$\sim e^{-k}$

> **The one sentence that ties it together.** *Chebyshev and Chernoff are both Markov in disguise — applied not to X but to a transformed variable that amplifies the tail before Markov sees it.* Squaring ($(X-\mu)^2$) turns a two-sided question into a one-sided non-negative one and earns a $1/k^2$. Exponentiating (e^{tX}) is even more aggressive: it blows the tail up so violently that, after optimizing the knob t, what survives decays **exponentially**. Stronger transform → stronger bound — but e^{tX} only factorizes when the terms are **independent**, which is the price Chernoff pays and the other two don't.

---

7. RQS is $O(n \log n)$ with high probability

Lecture 1 proved randomized QuickSort uses $\approx 2n \ln n$ comparisons **in expectation**. That is an average — a single run could (rarely) be much worse. Now we upgrade to a **with-high-probability** guarantee: *almost every* run is $O(n \log n)$, not just the average run. Chernoff is exactly the tool for this.

Good pivots vs. bad pivots

Model a run as the **recursion tree** $\mathrm{RQS}(S)$: root S, children $S_<$ and $S_>$ (the elements below / above the pivot). Total work is $O(n \cdot \mathrm{depth})$, so **it suffices to bound the depth**, i.e. the length of the longest root-to-leaf path.

> Call a pivot **good** if it splits its set in a roughly balanced way: $|S_<|,\ |S_>| \le \frac{2}{3}|S|$.

A pivot is good iff it lands in the middle third of the sorted order, so
$\Pr[\text{pivot is good}] = \frac{1}{3}, \qquad \Pr[\text{bad}] = \frac{2}{3}.$

Why good pivots cap the depth. Each good pivot shrinks the set by a factor $\le \frac{2}{3}$. If $S, S_1, S_2, \dots$ are the sets where good pivots occurred along a path, then $|S_i| \le \left(\frac{2}{3}\right)^i |S|$. The set hits size 1 after at most $c \log n$ **good pivots,** $\qquad c = \frac{1}{\ln(3/2)} \approx 2.43$ (good pivots on a path $\le c \ln n$). So a path can only be *long* if it is stuffed with **bad** pivots — and bad pivots are where the randomness can be pinned down by Chernoff.

The Chernoff step

Fix a single root-to-leaf path and look at its first $60 \ln n$ vertices. Let
$X_i = \begin{cases} 1 & i\text{-th vertex on the path has a bad pivot} \\ 0 & \text{otherwise} \end{cases}, \qquad \Pr[X_i = 1] = \frac{2}{3},$ and
$X = \sum_{i=1}^{60 \ln n} X_i$, so $\mu = E[X] = \frac{2}{3} \cdot 60 \ln n = 40 \ln n$.

A path with **more than** $60\ln n$ vertices must contain at least $60\ln n - 2.43\ln n = 57.57\ln n$ bad ones (only $\le 2.43\ln n$ can be good). So the path being too long forces X far above its mean: $\Pr[\text{path} > 60\ln n] \le \Pr[X \ge 57.57\ln n] < \Pr[X \ge 56\ln n] = \Pr\Big[X \ge \big(1+\tfrac{2}{5}\big)\underbrace{40\ln n}_{\mu}\Big]$. With $\delta = \frac{2}{5}$, the Chernoff form $\Pr[X \ge (1+\delta)\mu] \le e^{-\mu\delta^2/3}$ gives $\le \exp\Big(-40\ln n \cdot \frac{(2/5)^2}{3}\Big) = n^{-160/75} < n^{-2}$.

From one path to the whole tree (union bound)

One path is short except with probability $< n^{-2}$. There are at most n leaves, hence $\le n$ paths. Union bound: $\Pr[\text{some path} > 60\ln n] \le n \cdot n^{-2} = \frac{1}{n}$.

So **every** path has length $O(\log n)$ with probability $\ge 1 - 1/n$, i.e. the tree has depth $O(\log n)$ w.h.p., i.e. RQS does $O(n\log n)$ work w.h.p.

> **Punchline.** This is *strictly stronger* than the expectation bound. The argument is the canonical w.h.p. recipe: **(1)** isolate "good" events with a constant success probability; **(2)** Chernoff a single object to make its failure $\le n^{-2}$; **(3)** union-bound over the n objects to get total failure $\le 1/n$. The exponent-2 in step (2) is *engineered* precisely so that step (3) survives — this is why we needed Chernoff and not Chebyshev.

---

8. R-Select — finding the median in $2n + o(n)$ comparisons

Sorting finds the median in $O(n\log n)$. Can we do **linear**? Yes — and the idea is pure sampling: **look at a sublinear random sample, use it to trap the median inside a tiny window, then sort only that window.**

The algorithm (median = the $\frac{n}{2}$-th element)

Input: set S, $|S| = n$. Output: the median m.

1. **Sample.** Draw $R \leftarrow n^{3/4}$ elements from S, uniformly **with replacement**.
2. **Bracket.** Sort R. Let $\ell = \frac{n^{3/4}}{2} - \sqrt{n}, \qquad u = \frac{n^{3/4}}{2} + \sqrt{n},$ and let d, h be the ℓ-th and u-th smallest elements of R. These two sampled values are our guessed lower/upper fence around the median.
3. **Filter S against the fences:**
   $D = \{x \in S : x < d\}, \quad H = \{x \in S : x > h\}, \quad C = \{x \in S : d \le x \le h\}.$
4. **Decide.** If $|D| > \frac{n}{2}$ **or** $|H| > \frac{n}{2}$ **or** $|C| > 4n^{3/4}$, then **FAIL**. Otherwise the median lies in the small set C — find it there (by sorting C).

Why $2n + o(n)$ comparisons. Sorting R costs $O(n^{3/4}\log n) = o(n)$. Building D, H, C compares each of the n elements against d and h — that's the $2n$. Sorting the surviving window C (size $\le 4n^{3/4}$) costs $o(n)$ again. The two fence-tests dominate: $2n + o(n)$, genuinely linear.

The three FAIL conditions are exactly the three ways the plan can go wrong, and each is killed by **Chebyshev** (we don't even need Chernoff here — we only need the failure probability to vanish).

Bounding $\Pr[|D| > n/2]$ — Chebyshev on the sample count

$|D| > \frac{n}{2}$ means more than half of S lies below the fence d, i.e. the true median fell *below* d (we bracketed too high). Let $X_i = \begin{cases} 1 & i\text{-th sampled element} \le m \\ 0 & \text{otherwise} \end{cases}$, $X = \sum_{i=1}^{n^{3/4}} X_i$. Since m is the median, each sample lands $\le m$ with probability exactly $\frac{1}{2}$, so $X \sim \mathrm{Bin}\left(n^{3/4}, \frac{1}{2}\right)$ with $E[X] = \frac{n^{3/4}}{2}, \qquad \mathrm{Var}(X) = \frac{n^{3/4}}{4}$. The bad event "$d$ ended up above m" happens iff fewer than ℓ samples were $\le m$, i.e. $X < \frac{n^{3/4}}{2} - \sqrt{n}$. Apply Chebyshev with $k = \sqrt{n}$:
$\Pr\left[X < \frac{n^{3/4}}{2} - \sqrt{n}\right] \le \Pr\left[|X - E[X]| > \sqrt{n}\right] \le \frac{\mathrm{Var}(X)}{(\sqrt{n})^2} = \frac{n^{3/4}/4}{n} = \frac{n^{-1/4}}{4}$. By symmetry $\Pr[|H| > \frac{n}{2}] \le \frac{n^{-1/4}}{4}$ as well.

Bounding $\Pr[|C| > 4n^{3/4}]$ — the window stays small

If the window C is too big, then it bulges on one side of the median:
$|C| > 4n^{3/4} \implies |C_{\le m}| > 2n^{3/4}$ or $|C_{\ge m}| > 2n^{3/4}$. Take the upper bulge $|C_{\ge m}| > 2n^{3/4}$: it means the fence h sits a full $2n^{3/4}$ ranks above the median, i.e. h is among the top $\frac{n}{2} - 2n^{3/4}$ elements of S. Let $Y_i = \begin{cases} 1 & i\text{-th sample is among the top } \frac{n}{2} - 2n^{3/4} \text{ of } S \\ 0 & \text{otherwise} \end{cases}$, $Y = \sum_{i=1}^{n^{3/4}} Y_i$,
so $Y \sim \mathrm{Bin}(n^{3/4}, p)$ with $p = \frac{1}{2} - \frac{2}{n^{1/4}}$:
$E[Y] = n^{3/4}p = \frac{n^{3/4}}{2} - 2\sqrt{n}, \qquad \mathrm{Var}(Y) = n^{3/4}\,p(1-p) < \frac{n^{3/4}}{4}$. The bulge forces $Y > \frac{n^{3/4}}{2} - \sqrt{n}$, which is $\sqrt{n}$ above $E[Y]$. Chebyshev again:
$\Pr\left[Y > \frac{n^{3/4}}{2} - \sqrt{n}\right] = \Pr\left[Y - E[Y] > \sqrt{n}\right] \le \Pr\left[|Y - E[Y]| > \sqrt{n}\right] < \frac{n^{-1/4}}{4}$. Two sides, so $\Pr[|C| > 4n^{3/4}] < 2 \cdot \frac{n^{-1/4}}{4} = \frac{n^{-1/4}}{2}$.

Putting the three together

> **Theorem.** R-Select finds the median with probability $\ge 1 - n^{-1/4}$.

$$\Pr[\mathrm{FAIL}] \le \underbrace{\Pr[|D| > \tfrac{n}{2}]}_{\le n^{-1/4}/4} + \underbrace{\Pr[|H| > \tfrac{n}{2}]}_{\le n^{-1/4}/4} + \underbrace{\Pr[|C| > 4n^{3/4}]}_{\le n^{-1/4}/2} \;=\; n^{-1/4}.$$

On FAIL, just **restart**. Failures are independent, so $\Pr[\text{FAIL after } \ell \text{ runs}] \le n^{-\ell/4}$, and the expected number of runs is $E[\ell] < 2$. So the expected total cost stays $2n + o(n)$.

> **Punchline.** A *sublinear* sample of size $n^{3/4}$ is enough to pin the median's rank to within $\pm\sqrt{n}$ (the standard-deviation scale of a binomial), which traps the answer in a window of size $O(n^{3/4})$ that we can afford to sort outright. Chebyshev — the *middle* rung of the ladder — is already strong enough, because we only need failure $\to 0$, not n^{-c}. This $2n + o(n)$ is essentially the optimal comparison constant for selection.

Generalization — the k-th smallest element

The same template returns the k-th smallest S_k for any rank k (not just the median). Only the centering changes: set $x = k/n^{1/4}$ and put the sample fences at $\ell = \max\{\lfloor x - \sqrt{n} \rfloor, 1\}$ and $u = \min\{\lceil x + \sqrt{n} \rceil, n^{3/4}\}$. Near the ends ($k < n^{1/4}$ or $k > n - n^{1/4}$) you keep a one-sided window $\{x \le h\}$ resp. $\{x \ge d\}$; in the bulk you keep $\{d \le x \le h\}$. FAIL if the window grows past the $O(n^{3/4})$ budget. Same guarantee: $\Pr[\text{FAIL}] \le n^{-1/4}$, cost $2n + o(n)$.

Recurring themes to carry forward

Theme	Where it appeared	Reused later in
The tail-bound ladder (Markov ⊂ Chebyshev ⊂ Chernoff)	§3–6	every concentration argument in the course
"Markov on a transformed variable" (square → Chebyshev, exponentiate → Chernoff)	§4, §5	the meta-trick to remember
MGF + optimize t (the Chernoff proof)	§5	error amplification, expander walks
Chernoff + union bound → w.h.p. (engineer n^{-2} per object, sum over n)	§7 RQS depth	any "every one of n things behaves" claim
Variance adds only under independence	§1	why Chebyshev needs independent summands
Sample to estimate ranks (sublinear sample fixes the answer to $\pm\sqrt{n}$)	§8 R-Select	sampling-based algorithms generally
Restart-on-FAIL ($E[\ell] < 2$)	§8	turning a Monte-Carlo failure into Las-Vegas expected cost

The single sentence that ties the lecture together:

Concentration is the whole game: a random quantity almost never strays far from its mean, and the three inequalities are one idea — apply Markov after a transform — at escalating strength. Spend a sharper bound where you must (Chernoff to union-bound RQS over n paths) and a cheaper one where you can (Chebyshev to keep R-Select's sample window small).