Exponential Concentration I: Part A

SDS 391P.6, Spring 2026

Pratik Patil

These notes are a work in progress and are provided as-is for instructional purposes only. They are not (yet) at the level of a scholarly document. In particular, the notes draw from various sources and do not (yet) have sufficient references to the original sources. Additionally, almost surely the notes have errors and they are only probably approximately correct. The notes will be updated regularly as the course progresses. Last updated: 2026-02-02.

1 Motivation

In the last two lectures, we developed tools for bounding the variance of random variables of the form $Z = f(X_1, \dots, X_n)$ with independent $X_1, \dots, X_n$. This led to variance bounds in terms of coordinatewise sensitivity (Efron–Stein–Steele, bounded differences, etc.), and in structured settings to Poincaré-type inequalities such as $\mathrm{Var}(f(X)) \leqslant C\,\mathbb{E}\|\nabla f(X)\|_2^2$. At a high level, these results showcase a basic manifestation of concentration: when f is not too sensitive to its inputs (as quantified by discrete changes or by gradients), the fluctuations of Z around $\mathbb{E}Z$ are small, as measured through the variance.

Variance control is often already informative as we saw last week. By Chebyshev's inequality,

$$\mathbb{P}\{|Z - \mathbb{E}Z| \geqslant t\} \leqslant \frac{\mathrm{Var}(Z)}{t^2},$$

so small variance forces Z to lie near $\mathbb{E}Z$ with high probability. The limitation is that Chebyshev produces only polynomial tail decay (t^{-2}). In many applications, especially when we take unions, maxima, or suprema over many events, we need much sharper tail control.

This motivates our next step: *exponential concentration inequalities*, where the tail probability decays exponentially fast, often like $\exp(-ct^2)$ (Gaussian-type) or $\exp(-ct)$ (exponential-type). Figure 1 compares these decay rates on a log scale.

A guiding example is an independent sum $S_n = \sum_{i=1}^n X_i$, or the average $n^{-1}\sum_i X_i$. The central limit theorem suggests that, after normalization, fluctuations can look approximately Gaussian, and therefore might exhibit Gaussian-like tail decay. However, the central limit theorem is not a good tool for proving tail bounds: it controls the distribution only weakly, and quantitative versions (e.g. Berry–Esseen) typically have errors of order $n^{-1/2}$, far larger than the exponentially small tail probabilities we want to certify.

In this lecture, we begin our study of exponential concentration. Today we focus on the simplest and most important setting: sums of independent random variables. Next time we will begin extending these ideas beyond sums to nonlinear functions, where the same philosophy reappears in more subtle forms.

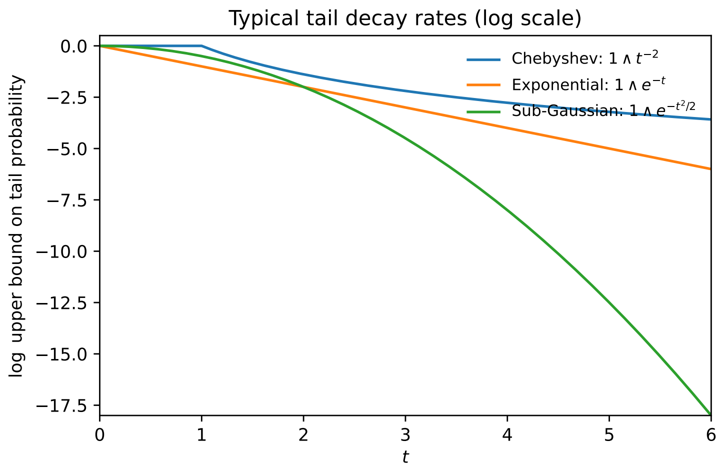

Figure 1: Tail decay (on a log scale). Chebyshev yields polynomial decay (t^{-2}), while exponential and sub-Gaussian bounds yield much faster decay (e^{-t} and $e^{-t^2/2}$, respectively).

2 From polynomial moments to exponential moments

Most concentration inequalities start from Markov's inequality: if $Y \geqslant 0$ and $a > 0$, then $\mathbb{P}\{Y \geqslant a\} \leqslant \mathbb{E}Y/a$. To turn this into tail bounds for a centered random variable $X - \mathbb{E}X$, we apply Markov not to $X - \mathbb{E}X$ itself, but to a nonnegative transform of it.

Let $\phi : \mathbb{R} \to [0, \infty)$ be nondecreasing. Then for any $t \in \mathbb{R}$,

$$\mathbb{P}\{X - \mathbb{E}X \geqslant t\} = \mathbb{P}\{\phi(X - \mathbb{E}X) \geqslant \phi(t)\} \leqslant \frac{\mathbb{E}\,\phi(X - \mathbb{E}X)}{\phi(t)}.$$

Different choices of ϕ give different *moment methods*:

- *First moment method.* Take $\phi(u) = u_+$. Then

  $$\mathbb{P}\{X - \mathbb{E}X \geqslant t\} \leqslant \frac{\mathbb{E}[(X - \mathbb{E}X)_+]}{t} \leqslant \frac{\mathbb{E}|X - \mathbb{E}X|}{t}.$$

  This requires only integrability, but yields weak t^{-1} decay.

- *Second moment method (Chebyshev).* Take $\phi(u) = u^2$. Then $\mathbb{E}\phi(X - \mathbb{E}X) = \mathrm{Var}(X)$ and

  $$\mathbb{P}\{|X - \mathbb{E}X| \geqslant t\} \leqslant \frac{\mathrm{Var}(X)}{t^2}.$$

  This requires finite variance, but provides a slightly better t^{-2} decay.

- *Higher polynomial moment method.* Take $\phi(u) = u_+^p$ for an integer $p \geqslant 1$. Then

  $$\mathbb{P}\{X - \mathbb{E}X \geqslant t\} \leqslant \inf_{p \in \mathbb{N}} \frac{\mathbb{E}[(X - \mathbb{E}X)_+^p]}{t^p}. \tag{1}$$

  As p increases, the decay in t improves now to t^{-p}, but estimating $\mathbb{E}|X - \mathbb{E}X|^p$ becomes harder and typically requires stronger assumptions (e.g. boundedness or additional structure).

- *Exponential moment method (Chernoff/Laplace).* Take $\phi(u) = e^{\lambda u}$ for $\lambda > 0$. Then

$$\mathbb{P}\{X - \mathbb{E}X \geqslant t\} \leqslant \inf_{\lambda > 0} \exp\Big(-\lambda t + \log \mathbb{E} e^{\lambda(X - \mathbb{E}X)} \Big).$$

  In many cases this yields Gaussian-type tails $\exp(-ct^2)$ when the log-MGF grows like λ^2.

There is a consistent "assume more, get more" pattern: better tail decay typically requires (i) more moment information and (ii) stronger independence assumptions when studying sums. For example, for $S_n = \sum_{i=1}^n X_i$: pairwise independence is enough to compute $\mathrm{Var}(S_n)$, k-wise independence can control $\mathbb{E}|S_n|^k$, and full independence makes exponential moments especially tractable (because MGFs factorize). Our goal is to identify natural assumptions under which exponential moments can be controlled cleanly, first for sums, and later for nonlinear functions.

3 Moment and cumulant generating functions

The exponential moment method is organized around the moment generating function (MGF) and its logarithm, the cumulant generating function (CGF).

Let X be a real random variable. Its moment generating function is

$$m_X(\lambda) := \mathbb{E} e^{\lambda X} \in (0, \infty] \qquad (\lambda \in \mathbb{R}),$$

and its cumulant generating function is

$$\kappa_X(\lambda) := \log m_X(\lambda) \in (-\infty, \infty] \qquad (\lambda \in \mathbb{R}),$$

with the convention $\log(+\infty) = +\infty$. We also define the centered log-MGF (centered CGF)

$$\psi_X(\lambda) := \log \mathbb{E} e^{\lambda(X - \mathbb{E}X)} = \kappa_X(\lambda) - \lambda\, \mathbb{E}X.$$

The MGF/CGF may be infinite for some λ. They are most useful when they are finite on a non-trivial interval around 0, which is a quantitative way of saying that X has at least some exponential tail behavior.

3.1 How MGFs "encode" polynomial moments

Whenever $m_X(\lambda) < \infty$ for λ in a neighborhood of 0, we may differentiate under the expectation (justified by dominated convergence on that neighborhood). Formally,

$$m_X(\lambda) = \mathbb{E} e^{\lambda X} = \mathbb{E} \sum_{k=0}^{\infty} \frac{\lambda^k X^k}{k!} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mathbb{E}[X^k].$$

Thus the MGF is the *exponential generating function* of the moments:

$$m_X^{(k)}(0) = \mathbb{E}[X^k].$$

In particular,

$$m_X'(0) = \mathbb{E}X, \qquad m_X''(0) = \mathbb{E}[X^2].$$

The CGF packages the same information in a different coordinate system. Its Taylor coefficients are the *cumulants*:

$$\kappa_X(\lambda) = \sum_{k\geqslant 1} \frac{\lambda^k}{k!} \operatorname{cumulant}_k(X), \qquad \operatorname{cumulant}_1(X) = \mathbb{E}X, \qquad \operatorname{cumulant}_2(X) = \operatorname{Var}(X).$$

So the variance that drove our previous lectures appears as the second derivative at zero of the CGF. This is the conceptual link between second-moment concentration and exponential-moment concentration: instead of controlling a single number (the variance), we control a whole function $\lambda \mapsto \psi_X(\lambda)$. Here λ is the "knob" that we get to control!

3.2 Convexity and tilted-variance identities

A key structural feature is convexity.

Proposition 3.1 (Convexity of the CGF)**.** *The functions $\lambda \mapsto \kappa_X(\lambda)$ and $\lambda \mapsto \psi_X(\lambda)$ are convex on $\mathbb{R}$ (where finite). If $m_X(\lambda) < \infty$ in a neighborhood of a point λ, then κ_X is twice differentiable there and*

$$\kappa_X''(\lambda) = \operatorname{Var}_{\mathbb{P}_\lambda}(X) \geqslant 0,$$

where $\mathbb{P}_\lambda$ is the exponentially tilted measure

$$d\mathbb{P}_\lambda := \frac{e^{\lambda X}}{\mathbb{E}e^{\lambda X}}\, d\mathbb{P}.$$

In particular, when m_X is finite near 0*,*

$$\psi_X(0) = 0, \qquad \psi_X'(0) = 0, \qquad \psi_X''(0) = \operatorname{Var}(X).$$

Proof. When m_X is finite near λ, differentiation under the expectation gives

$$\kappa_X'(\lambda) = \frac{\mathbb{E}[Xe^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} = \mathbb{E}_{\mathbb{P}_\lambda}[X],$$

and a second differentiation gives

$$\kappa_X''(\lambda) = \mathbb{E}_{\mathbb{P}_\lambda}[X^2] - \mathbb{E}_{\mathbb{P}_\lambda}[X]^2 = \operatorname{Var}_{\mathbb{P}_\lambda}(X) \geqslant 0.$$

The identities for ψ_X at 0 follow from centering. □

Figure 2 shows typical centered CGFs and their convexity.

3.3 A key structural advantage: additivity under independence

For independent sums, MGFs multiply and CGFs add. This simple fact is the main workforce behind many exponential concentration inequalities.

Proposition 3.2 (Additivity of CGFs for independent sums)**.** *If $X_1, \dots, X_n$ are independent and $S_n = \sum_{i=1}^n X_i$, then*

$$m_{S_n}(\lambda) = \prod_{i=1}^n m_{X_i}(\lambda), \qquad \kappa_{S_n}(\lambda) = \sum_{i=1}^n \kappa_{X_i}(\lambda), \qquad \psi_{S_n}(\lambda) = \sum_{i=1}^n \psi_{X_i}(\lambda).$$

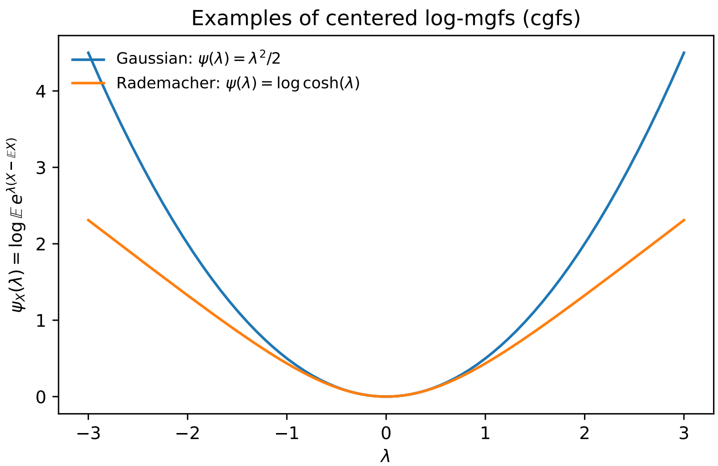

Figure 2: Examples of centered log-MGFs. For $G \sim \mathcal{N}(0,1)$, $\psi_G(\lambda) = \lambda^2/2$. For a Rademacher $R \in \{-1,+1\}$, $\psi_R(\lambda) = \log\cosh(\lambda)$. Both are convex, and $\log\cosh(\lambda) \leqslant \lambda^2/2$ reflects a sub-Gaussian MGF bound.

Proof. By independence,

$$m_{S_n}(\lambda) = \mathbb{E}e^{\lambda \sum_i X_i} = \mathbb{E}\prod_i e^{\lambda X_i} = \prod_i \mathbb{E}e^{\lambda X_i} = \prod_i m_{X_i}(\lambda).$$

Taking logarithms yields additivity of κ; subtracting $\lambda \mathbb{E}S_n$ yields additivity of ψ. □

4 The Laplace transform method (Chernoff bound)

We now state the exponential-moment analog of Chebyshev's inequality.

Lemma 4.1 (Chernoff bound)**.** *Let X be a real random variable such that $\psi_X(\lambda) = \log \mathbb{E}e^{\lambda(X-\mathbb{E}X)} < \infty$ for λ in a neighborhood of 0. Define the (one-sided) Legendre transform*

$$\psi_X^*(t) := \sup_{\lambda \geqslant 0}\{\lambda t - \psi_X(\lambda)\}, \qquad t \geqslant 0.$$

Then for all $t \geqslant 0$,

$$\mathbb{P}\{X - \mathbb{E}X \geqslant t\} \leqslant e^{-\psi_X^*(t)}.$$

Likewise,

$$\mathbb{P}\{X - \mathbb{E}X \leqslant -t\} = \mathbb{P}\{-X - \mathbb{E}(-X) \geqslant t\} \leqslant e^{-\psi_{-X}^*(t)}.$$

Proof. For any $\lambda \geqslant 0$, since $u \mapsto e^{\lambda u}$ is increasing,

$$\mathbb{P}\{X - \mathbb{E}X \geqslant t\} = \mathbb{P}\{e^{\lambda(X-\mathbb{E}X)} \geqslant e^{\lambda t}\} \leqslant e^{-\lambda t}\mathbb{E}e^{\lambda(X-\mathbb{E}X)} = \exp\big(-\lambda t + \psi_X(\lambda)\big).$$

Optimizing over $\lambda \geqslant 0$ finishes the proof. □

The same argument applied to $-X$ yields a lower tail bound, and the union bound then gives two-sided bounds:

$$\mathbb{P}\{|X - \mathbb{E}X| \geqslant t\} \leqslant \mathbb{P}\{X - \mathbb{E}X \geqslant t\} + \mathbb{P}\{X - \mathbb{E}X \leqslant -t\}.$$

The Chernoff bound is not restricted to Gaussian tails: different upper bounds on $\psi_X(\lambda)$ produce different tail behaviors. However, the method requires $\psi_X(\lambda) < \infty$ at least for λ near 0, so it is most naturally suited to random variables with at least exponential tail behavior.

For heavier tails, one can use the power-moment method (1). In fact, even when exponential moments exist, optimizing over p in (1) can be at least as good (up to constants) as optimizing over λ in Chernoff. (You will prove on your homework!)

Then, why do we emphasize Chernoff bounds? Because $\lambda \mapsto \psi_X(\lambda)$ is a smooth object that can be studied by calculus, and because for independent sums ψ is additive (Proposition 3.2). This combination makes Chernoff bounds remarkably powerful! Calculus for the win!

5 Sub-Gaussian MGF bounds and Gaussian tails

Let us derive Gaussian tails from the Laplace method. If $G \sim \mathcal{N}(0, \sigma^2)$, then

$$\mathbb{E}e^{\lambda G} = e^{\lambda^2\sigma^2/2}, \qquad \text{so} \qquad \psi_G(\lambda) = \frac{\lambda^2\sigma^2}{2}.$$

Optimizing the Chernoff bound yields

$$\mathbb{P}\{G \geqslant t\} \leqslant e^{-t^2/(2\sigma^2)} \qquad \text{for all } t \geqslant 0.$$

This suggests a general strategy: if we can show that a random variable has $\psi_X(\lambda)$ no larger than a Gaussian CGF, then it must have Gaussian-type tails.

A real random variable X is called σ^2-sub-Gaussian if

$$\psi_X(\lambda) = \log \mathbb{E}e^{\lambda(X - \mathbb{E}X)} \leqslant \frac{\sigma^2\lambda^2}{2} \qquad \text{for all } \lambda \in \mathbb{R}.$$

The parameter σ^2 is called a *variance proxy.*

Proposition 5.1 (Sub-Gaussian implies Gaussian tails)**.** *If X is σ^2-sub-Gaussian, then for all $t \geqslant 0$,*

$$\mathbb{P}\{X - \mathbb{E}X \geqslant t\} \leqslant \exp\Big(-\frac{t^2}{2\sigma^2}\Big), \qquad \mathbb{P}\{X - \mathbb{E}X \leqslant -t\} \leqslant \exp\Big(-\frac{t^2}{2\sigma^2}\Big),$$

and hence

$$\mathbb{P}\{|X - \mathbb{E}X| \geqslant t\} \leqslant 2\exp\Big(-\frac{t^2}{2\sigma^2}\Big).$$

Proof. Apply Lemma 4.1 and use $\psi_X(\lambda) \leqslant \sigma^2\lambda^2/2$:

$$\mathbb{P}\{X - \mathbb{E}X \geqslant t\} \leqslant \inf_{\lambda \geqslant 0} \exp\Big(-\lambda t + \frac{\sigma^2\lambda^2}{2}\Big).$$

The infimum occurs at $\lambda = t/\sigma^2$, giving $\exp(-t^2/(2\sigma^2))$. Apply the same argument to $-X$ for the lower tail. □

6 Hoeffding's lemma: bounded implies sub-Gaussian

The main ingredient behind Hoeffding's inequality for bounded independent sums is a CGF bound for a single bounded random variable.

Lemma 6.1 (Hoeffding's lemma)**.** *If $a \leqslant X \leqslant b$ almost surely, then for all $\lambda \in \mathbb{R}$,*

$$\log \mathbb{E} e^{\lambda(X - \mathbb{E}X)} \leqslant \frac{\lambda^2 (b-a)^2}{8}.$$

Equivalently, X is $(b-a)^2/4$-sub-Gaussian.

Proof. By shifting, assume $\mathbb{E}X = 0$ and write $\psi(\lambda) = \log \mathbb{E} e^{\lambda X}$. Differentiate:

$$\psi'(\lambda) = \frac{\mathbb{E}[X e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]}, \qquad \psi''(\lambda) = \frac{\mathbb{E}[X^2 e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} - \Big(\frac{\mathbb{E}[X e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]}\Big)^2.$$

Introduce the exponentially tilted probability measure

$$d\mathbb{P}_\lambda := \frac{e^{\lambda X}}{\mathbb{E} e^{\lambda X}} \, d\mathbb{P}.$$

Then $\psi'(\lambda) = \mathbb{E}_\lambda[X]$ and $\psi''(\lambda) = \mathrm{Var}_\lambda(X)$. Under $\mathbb{P}_\lambda$ the random variable X is still supported on $[a, b]$, so the range bound for variance yields

$$0 \leqslant \psi''(\lambda) = \mathrm{Var}_\lambda(X) \leqslant \frac{(b-a)^2}{4}.$$

Now integrate twice using $\psi(0) = 0$ and $\psi'(0) = \mathbb{E}X = 0$:

$$\psi(\lambda) = \int_0^\lambda \int_0^t \psi''(s) \, ds \, dt \leqslant \int_0^\lambda \int_0^t \frac{(b-a)^2}{4} \, ds \, dt = \frac{\lambda^2 (b-a)^2}{8}.$$

□

7 Hoeffding's inequality for independent bounded sums

Theorem 7.1 (Hoeffding's inequality)**.** *Let $X_1, \dots, X_n$ be independent random variables with $a_i \leqslant X_i \leqslant b_i$ almost surely. Let $S_n = \sum_{i=1}^n X_i$ and define the variance proxy*

$$v := \frac{1}{4} \sum_{i=1}^n (b_i - a_i)^2.$$

Then for all $t \geqslant 0$,

$$\mathbb{P}\{|S_n - \mathbb{E}S_n| \geqslant t\sqrt{v}\} \leqslant 2e^{-t^2/2}.$$

Equivalently, for all $u \geqslant 0$,

$$\mathbb{P}\{|S_n - \mathbb{E}S_n| \geqslant u\} \leqslant 2 \exp\Big(-\frac{u^2}{2v}\Big).$$

Proof. By additivity of centered log-MGFs (Proposition 3.2) and Hoeffding's lemma,

$$\psi_{S_n}(\lambda) = \sum_{i=1}^{n} \psi_{X_i}(\lambda) \leqslant \sum_{i=1}^{n} \frac{\lambda^2 (b_i - a_i)^2}{8} = \frac{\lambda^2}{2} v.$$

Thus S_n is v-sub-Gaussian, and Proposition 5.1 yields the stated tails. □

Remark 7.2 (Variance proxy vs variance)**.** From last week's variance bounds,

$$\operatorname{Var}(S_n) = \sum_{i=1}^{n} \operatorname{Var}(X_i) \leqslant \frac{1}{4} \sum_{i=1}^{n} (b_i - a_i)^2 = v.$$

So v is an upper bound on $\operatorname{Var}(S_n)$ that depends only on coordinate ranges. Hoeffding's inequality upgrades this variance control into a Gaussian tail bound on the scale $\sqrt{v}$.

8 Application: median-of-means mean estimation

Hoeffding's inequality is a basic tool in statistics and learning theory. Here is a classical application illustrating how exponential concentration can be obtained even when the underlying data may be heavy-tailed.

Let X be a real random variable with mean μ and variance $\sigma^2 < \infty$, and let $X_1, \ldots, X_N$ be independent copies of X. The sample mean $\bar{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i$ always satisfies

$$\operatorname{Var}(\bar{X}_N) = \frac{\sigma^2}{N},$$

so Chebyshev gives a polynomial tail bound. Remarkably, one can do better: there exists an estimator with *Gaussian-type tails* under only a finite-variance assumption.

Theorem 8.1 (Median-of-means estimator)**.** *Assume* $\mathbb{E}X = \mu$ *and* $\operatorname{Var}(X) = \sigma^2 < \infty$. *There exists an estimator* $\widehat{\mu} = \widehat{\mu}(X_1, \ldots, X_N)$ *and an absolute constant* $c > 0$ *such that for every* t *with* $0 \leqslant t \leqslant \sqrt{N}$,

$$\mathbb{P}\left\{ |\widehat{\mu} - \mu| \geqslant \frac{t\sigma}{\sqrt{N}} \right\} \leqslant 2e^{-ct^2}.$$

Proof. Assume for simplicity that $N = BL$ for integers B, L. Partition the data into B blocks of size L and form the block means

$$\mu_b = \frac{1}{L} \sum_{i=(b-1)L+1}^{bL} X_i, \qquad b = 1, \ldots, B.$$

Define $\widehat{\mu}$ to be a median of $\{\mu_1, \ldots, \mu_B\}$.

Each μ_b has mean μ and variance σ^2 / L. Fix $t \geqslant 0$ and choose

$$B = \left\lceil \frac{t^2}{4} \right\rceil, \qquad \text{so that} \qquad L = \frac{N}{B}.$$

Then, by Chebyshev,

$$\mathbb{P}\left\{ \mu_b - \mu \geqslant \frac{t\sigma}{\sqrt{N}} \right\} \leqslant \frac{\operatorname{Var}(\mu_b)}{(t\sigma/\sqrt{N})^2} = \frac{\sigma^2/L}{t^2\sigma^2/N} = \frac{B}{t^2} \leqslant \frac{1}{4},$$

and the same bound holds for the lower deviation $\mu_b - \mu \leqslant -t\sigma/\sqrt{N}$.

Let I_b be the indicator that block b is "bad" in the upper tail:

$$I_b = \mathbf{1}\left\{\mu_b - \mu \geqslant \frac{t\sigma}{\sqrt{N}}\right\}.$$

The random variables (I_b) are independent and satisfy $\mathbb{E}I_b \leqslant 1/4$. If $\widehat{\mu} \geqslant \mu + t\sigma/\sqrt{N}$, then at least half the block means must be bad, hence

$$\mathbb{P}\left\{\widehat{\mu} - \mu \geqslant \frac{t\sigma}{\sqrt{N}}\right\} \leqslant \mathbb{P}\left\{\sum_{b=1}^{B} I_b \geqslant \frac{B}{2}\right\}.$$

Apply Hoeffding's inequality (Theorem 7.1) to the bounded independent variables $I_b \in [0,1]$. With $\mathbb{E}I_b \leqslant 1/4$, the event $\sum I_b \geqslant B/2$ is a deviation of size at least $B/4$ above the mean, so

$$\mathbb{P}\left\{\sum_{b=1}^{B} I_b \geqslant \frac{B}{2}\right\} \leqslant \exp(-c_0 B)$$

for an absolute constant $c_0 > 0$ (one may take $c_0 = 1/8$ from the explicit Hoeffding bound). With $B \simeq t^2$, this yields

$$\mathbb{P}\left\{\widehat{\mu} - \mu \geqslant \frac{t\sigma}{\sqrt{N}}\right\} \leqslant e^{-ct^2}.$$

The lower tail is handled identically by considering $\mathbf{1}\{\mu_b - \mu \leqslant -t\sigma/\sqrt{N}\}$. Combine the two tails with the union bound. □

The key take-away message here is: Hoeffding's inequality provides exponential concentration for sums of bounded variables, and this can be used as a building block to create robust estimators with exponential tails even when the raw data are unbounded and far from Gaussian.

9 Look ahead

In the previous lectures, we controlled variance using discrete coordinate perturbations (resampling and bounded differences) and, in certain settings, by gradients (Poincaré inequalities). In this lecture we introduced the Laplace transform method, which upgrades variance-level control into exponential tail bounds once we can control the log-moment generating function.

Next time we will develop this viewpoint further: we will introduce systematic classes of light-tailed variables (sub-Gaussian and sub-exponential), prove additional inequalities (Chernoff and Bernstein-type bounds), and begin extending exponential concentration beyond sums to nonlinear functions.

Source material

Parts of this lecture are based on references: Vershynin (2018); Tropp (2023); van Handel (2016), in addition to the author's accumulated experience working on related topics.

References

Tropp, J. A. (2023). Probability in high dimensions. Caltech CMS Lecture Notes 2021-01.

van Handel, R. (2016). Probability in high dimension. Lecture Notes (Princeton University). APC 550.

Vershynin, R. (2018). *High-dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press.