Stat 6390.001: High Dimensional Probability and Statistics, Fall 2026 **Topic 1**

Distributions of Functions of Random Variables

Lecturer: Yun Wei *Scribe: Zechen Liu*

1 Introduction

A central challenge in high-dimensional statistics is the *curse of dimensionality*. As the dimension d grows, the sample space expands rapidly, so a fixed number of observations becomes increasingly sparse relative to the size of the space. As a result, methods and intuition that work well in low dimensions may become much less informative in high-dimensional settings.

High-dimensional problems also exhibit geometric and probabilistic behavior that can differ substantially from low-dimensional intuition. For example, if

$$X \sim N(0, I_d),$$

then the density is maximized at the origin, but a typical sample is not concentrated near the origin. Instead,

$$\|X\|_2 \approx \sqrt{d},$$

so most samples lie near a thin shell of radius approximately $\sqrt{d}$.

These phenomena motivate the study of probabilistic tools that remain useful when the dimension is large. In particular, this lecture develops tail and concentration bounds that quantify how likely a random variable is to deviate from its mean, with an emphasis on non-asymptotic guarantees.

1.1 One-dimensional and multivariate CLT

Let $X_1, \ldots, X_n$ be i.i.d. random variables with mean μ and finite variance σ^2. The one-dimensional CLT states that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} N(0, \sigma^2) \qquad \text{as } n \to \infty. \tag{1}$$

The multivariate version has the same form. Suppose now that $X_1, \ldots, X_n$ are i.i.d. random vectors in $\mathbb{R}^d$ with

$$\mathbb{E}[X_i] = \mu \in \mathbb{R}^d, \qquad \mathrm{Cov}(X_i) = \Sigma.$$

Then

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} N(0, \Sigma) \qquad \text{as } n \to \infty. \tag{2}$$

For a d-dimensional Gaussian vector $X \sim N(\mu, \Sigma)$ with positive definite covariance matrix Σ, the density is

$$f(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}. \tag{3}$$

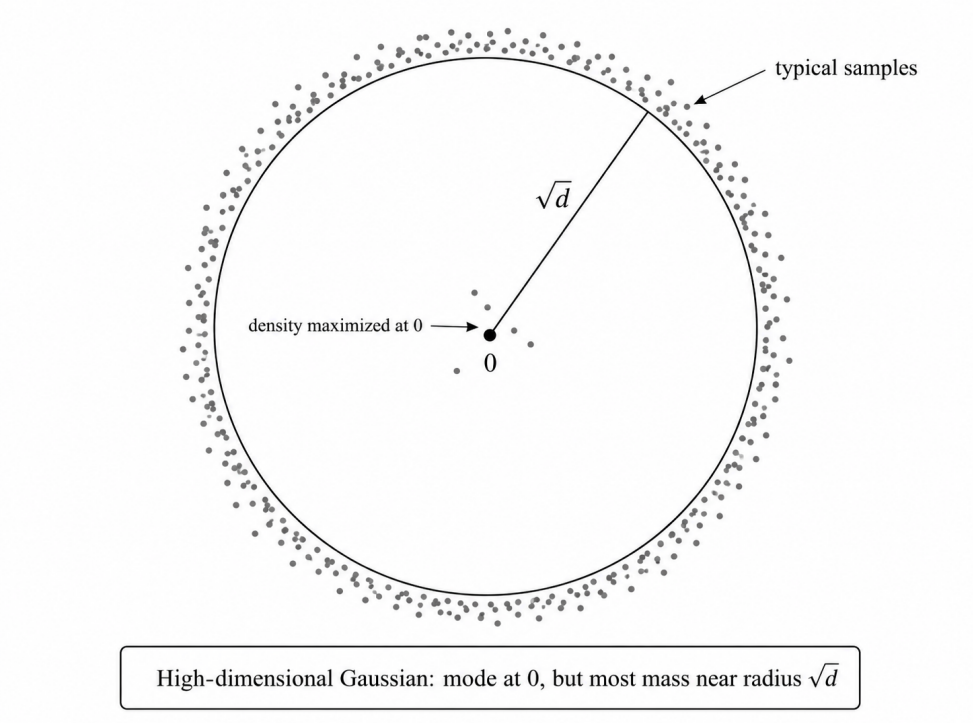

Figure 1: A schematic illustration of a high-dimensional standard Gaussian distribution.

Example 1.1 (A First High-Dimensional Phenomenon)**.** Consider the standard Gaussian case $X \sim N(0, I_d)$. The density is maximized at the origin, so the mode is $x = 0$. However, when d is large, a typical draw is not close to the origin. Indeed,

$$\mathbb{E}\|X\|_2^2 = \mathbb{E}\left[\sum_{j=1}^{d} X_j^2\right] = \sum_{j=1}^{d} \mathbb{E}[X_j^2] = d. \tag{4}$$

Thus the typical Euclidean radius is of order $\sqrt{d}$; more precisely, $\|X\|_2^2$ is a chi-square random variable with d degrees of freedom and is concentrated around d. Hence most of the Gaussian probability mass lies in a thin shell with radius approximately $\sqrt{d}$, even though the density itself is largest at the origin, shown in Figure 1. This illustrates a basic feature of high-dimensional geometry: the location of the mode need not describe where a typical sample lies.

2 Classical Tail Bounds

The purpose of a tail bound is to control probabilities such as $\mathbb{P}(X \geq t)$ or $\mathbb{P}(|X - \mathbb{E}X| \geq t)$. The amount of information available about the distribution determines the quality of the bound. Markov's inequality uses only a first moment; Chebyshev's inequality uses a second moment; and the Chernoff method exploits the moment generating function (MGF), thereby incorporating information from all moments when the MGF exists [1].

2.1 Markov and Chebyshev inequalities

Proposition 2.1 (Markov's inequality)**.** *If random variable X is nonnegative and $\mathbb{E}[X] < \infty$, then for every $t > 0$,*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}. \tag{5}$$

Proposition 2.2 (Chebyshev's inequality)**.** *If random variable X has finite variance and $\mu = \mathbb{E}[X]$, then for every $t > 0$,*

$$\mathbb{P}\big(|X - \mu| \geq t\big) \leq \frac{\mathrm{Var}(X)}{t^2}. \tag{6}$$

Chebyshev's inequality is an immediate consequence of Markov's inequality applied to the nonnegative random variable $(X - \mu)^2$. More generally, if the kth centered absolute moment exists, then

$$\mathbb{P}\big(|X - \mu| \geq t\big) = \mathbb{P}\big(|X - \mu|^k \geq t^k\big) \leq \frac{\mathbb{E}|X - \mu|^k}{t^k}. \tag{7}$$

2.2 From Markov to the Chernoff bound

The exponential function is particularly effective because it converts sums into products and leads naturally to the moment generating function.

Lemma 2.3 (Chernoff bound)**.** *Let $\mu = \mathbb{E}[X]$, and suppose*

$$\phi(\lambda) = \mathbb{E}\big[e^{\lambda(X-\mu)}\big]$$

exists for $\lambda \in [0, b]$. Then for every $t > 0$,

$$\log \mathbb{P}(X - \mu \geq t) \leq \inf_{\lambda \in [0,b]} \left\{ \log \phi(\lambda) - \lambda t \right\}. \tag{8}$$

Proof. Fox $\lambda > 0$. Since the exponential function is increasing,

$$\mathbb{P}(X - \mu \geq t) = \mathbb{P}\big(\lambda(X - \mu) \geq \lambda t\big) \tag{9}$$

$$= \mathbb{P}\big(e^{\lambda(X-\mu)} \geq e^{\lambda t}\big). \tag{10}$$

Applying Markov's inequality to the nonnegative random variable $e^{\lambda(X-\mu)}$ gives

$$\mathbb{P}(X - \mu \geq t) \leq \frac{\mathbb{E}[e^{\lambda(X-\mu)}]}{e^{\lambda t}} = \phi(\lambda) e^{-\lambda t}. \tag{11}$$

Taking logarithms yields

$$\log \mathbb{P}(X - \mu \geq t) \leq \log \phi(\lambda) - \lambda t.$$

Because the inequality holds for every admissible λ, minimizing the right-hand side over $\lambda \in [0, b]$ gives the result. □

Remark 2.4. The optimization over λ is essential. Different values of λ produce different valid bounds, and the infimum selects the sharpest one obtained from this argument. In applications, the main work is therefore to control the MGF and then solve a one-dimensional optimization problem.

3 Sub-Gaussian Random Variables

3.1 Gaussian Tail Bound

As a first example, consider a Gaussian random variable

$$X \sim N(\mu, \sigma^2).$$

Its moment generating function is

$$\mathbb{E}[e^{\lambda X}] = \exp\left(\mu\lambda + \frac{\sigma^2\lambda^2}{2}\right), \qquad \lambda \in \mathbb{R}. \tag{12}$$

Applying the preceding lemma, the upper-tail bound is determined by

$$\inf_{\lambda>0}\left\{\log \mathbb{E}[e^{\lambda(X-\mu)}] - \lambda t\right\} = \inf_{\lambda>0}\left\{\frac{\sigma^2\lambda^2}{2} - \lambda t\right\}. \tag{13}$$

Here we used the fact that

$$X - \mu \sim N(0, \sigma^2).$$

The quadratic expression is minimized at

$$\lambda^* = \frac{t}{\sigma^2},$$

which gives

$$\inf_{\lambda>0}\left\{\frac{\sigma^2\lambda^2}{2} - \lambda t\right\} = -\frac{t^2}{2\sigma^2}. \tag{14}$$

Therefore, for every $t > 0$,

$$\mathbb{P}(X - \mu \geq t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right). \tag{15}$$

(Normal distribution is light tail)

3.2 Definition and basic tail bounds

Definition 3.1 (Sub-Gaussian random variable)**.** A random variable X with mean $\mu = \mathbb{E}[X]$ is *sub-Gaussian with parameter* $\sigma > 0$ if

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq \exp\left(\frac{\sigma^2\lambda^2}{2}\right) \qquad \text{for all } \lambda \in \mathbb{R}. \tag{16}$$

The definition says that the centered MGF of X grows no faster than the centered MGF of a Gaussian $N(0, \sigma^2)$. Therefore, the same Chernoff calculation gives the Gaussian-type upper-tail bound

$$\mathbb{P}(X - \mu \geq t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right), \qquad t > 0. \tag{17}$$

Lemma 3.2 (Symmetry of sub-Gaussianity)**.** *A random variable X is sub-Gaussian with parameter σ if and only if $-X$ is sub-Gaussian with the same parameter.*

Proof. Let $\mu = \mathbb{E}[X]$. Recall that X is sub-Gaussian with parameter σ if

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq \exp\left(\frac{\sigma^2\lambda^2}{2}\right), \qquad \lambda \in \mathbb{R}. \tag{18}$$

Suppose first that X is sub-Gaussian. Since $\mathbb{E}[-X] = -\mu$,

$$\mathbb{E}\left[e^{\lambda((-X)-\mathbb{E}[-X])}\right] = \mathbb{E}\left[e^{-\lambda(X-\mu)}\right] \tag{19}$$

$$\leq \exp\left(\frac{\sigma^2(-\lambda)^2}{2}\right) \tag{20}$$

$$= \exp\left(\frac{\sigma^2\lambda^2}{2}\right). \tag{21}$$

Hence $-X$ is sub-Gaussian with parameter σ.

Conversely, suppose that $-X$ is sub-Gaussian. Then

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] = \mathbb{E}\left[e^{-\lambda((-X)-\mathbb{E}[-X])}\right] \tag{22}$$

$$\leq \exp\left(\frac{\sigma^2(-\lambda)^2}{2}\right) \tag{23}$$

$$= \exp\left(\frac{\sigma^2\lambda^2}{2}\right). \tag{24}$$

Therefore X is sub-Gaussian with parameter σ. □

Lemma 3.3 (Lower Tail)**.** *If X is sub-Gaussian with parameter σ, then for every $t > 0$,*

$$\mathbb{P}(X - \mathbb{E}[X] \leq -t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right). \tag{25}$$

Proof. By the symmetry lemma, $-X$ is also sub-Gaussian with parameter σ. Moreover,

$$\mathbb{P}(X - \mathbb{E}[X] \leq -t) = \mathbb{P}(-X - \mathbb{E}[-X] \geq t) \tag{26}$$

$$\leq \exp\left(-\frac{t^2}{2\sigma^2}\right), \tag{27}$$

where the last step follows from the upper-tail bound applied to $-X$. □

Now if we combine upper and lower tails with the union bound yields the two-sided concentration inequality we will get the following

Lemma 3.4. *If X is sub-Gaussian with parameter $\sigma > 0$,*

$$\mathbb{P}\big(|X - \mathbb{E}X| \geq t\big) \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right), \qquad t > 0. \tag{28}$$

Note: This much faster decay is one reason sub-Gaussian assumptions are central in non-asymptotic high-dimensional statistics [1].

Example 3.5 (Rademacher random variable)**.** A Rademacher random variable ε takes values in $\{-1, +1\}$ equiprobably:

In particular, $\mathbb{E}[\varepsilon] = 0$. Its MGF is

$$\mathbb{E}[e^{\lambda\varepsilon}] = \frac{1}{2}e^{\lambda} + \frac{1}{2}e^{-\lambda} \tag{29}$$

$$= \frac{1}{2}\left(\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} + \sum_{k=0}^{\infty}\frac{(-\lambda)^k}{k!}\right) \tag{30}$$

$$= \sum_{m=0}^{\infty}\frac{\lambda^{2m}}{(2m)!}, \tag{31}$$

where the odd-order terms cancel.

Since $(2m)! \geq 2^m m!$,

$$\mathbb{E}[e^{\lambda\varepsilon}] \leq \sum_{m=0}^{\infty} \frac{\lambda^{2m}}{2^m m!} \tag{32}$$

$$= e^{\lambda^2/2}. \tag{33}$$

Since $\mathbb{E}[\varepsilon] = 0$, it follows that

$$\mathbb{E}[e^{\lambda(\varepsilon - \mathbb{E}[\varepsilon])}] \leq e^{\lambda^2/2}.$$

Therefore, ε is sub-Gaussian with parameter $\sigma = 1$.

This example shows that sub-Gaussianity is not the same as Gaussianity. A Rademacher variable is discrete and bounded, yet its MGF is dominated by that of a standard Gaussian. The sub-Gaussian class is therefore broad enough to include many non-Gaussian random variables while retaining Gaussian-like tail behavior.

Note:"equiprobably" meaning

$$\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = \frac{1}{2}. \tag{34}$$

Therefore,

$$\mathbb{E}[\varepsilon] = \frac{1}{2}(1) + \frac{1}{2}(-1) = 0. \tag{35}$$

References

[1] Martin J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge University Press, 2019. See especially Chapters 1 and 2.