

Lecture 9: Concentration Inequalities

EECS 126 (UC Berkeley)

Spring 2026 — February 17, 2026

Reading: Bertsekas & Tsitsiklis, §5.1; supplementary notes on concentration inequalities.

These are my personal notes from Professor Kannan Ramchandran's EECS 126 (Probability and Random Processes) course at UC Berkeley.

Topics: Markov's inequality, Chebyshev's inequality, the Chernoff bound, Jensen's inequality, Hoeffding's inequality, and a first look at the law of large numbers and the central limit theorem. Each bound trades more information about X for a tighter estimate of a tail probability.

9.1 Markov's Inequality

Theorem 9.1 (Markov's inequality)

If $X \geq 0$ and $a > 0$, then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \quad (9.1)$$

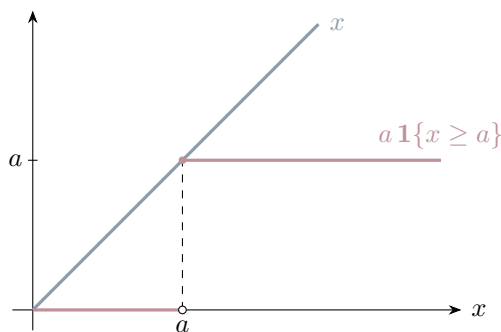
Proof. Compare X with the random variable $a \mathbf{1}\{X \geq a\}$ pointwise:

- if $X \geq a$: $a \mathbf{1}\{X \geq a\} = a \leq X$;
- if $X < a$: $a \mathbf{1}\{X \geq a\} = 0 \leq X$, since $X \geq 0$.

So $X \geq a \mathbf{1}\{X \geq a\}$ always. Taking expectations, and using $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A)$,

$$\mathbb{E}[X] \geq a \mathbb{E}[\mathbf{1}\{X \geq a\}] = a \mathbb{P}(X \geq a). \quad (9.2)$$

Dividing by $a > 0$ gives (9.1). □



For $x \geq 0$, the line x lies above the step $a \mathbf{1}\{x \geq a\}$; averaging both sides gives Markov.

Example 9.1 (Markov for an exponential)

Let $X \sim \text{Exponential}(1)$, so $\mathbb{E}[X] = 1$ (Lecture 8) and $\mathbb{P}(X \geq a) = \int_a^\infty e^{-x} dx = \left[-e^{-x}\right]_a^\infty = e^{-a}$. Markov gives

$$\mathbb{P}(X \geq a) \leq \frac{1}{a}. \quad (9.3)$$

At $a = 5$: Markov gives 0.2, while the exact value is $e^{-5} \approx 0.0067$.

Markov cannot be improved using only the mean. Fix $m = \mathbb{E}[X]$ and $a \geq m$. The random variable

$$X = \begin{cases} a & \text{w.p. } \frac{m}{a}, \\ 0 & \text{w.p. } 1 - \frac{m}{a} \end{cases} \quad (9.4)$$

has $\mathbb{E}[X] = a \cdot \frac{m}{a} = m$ and $\mathbb{P}(X \geq a) = \frac{m}{a} = \frac{\mathbb{E}[X]}{a}$, so (9.1) holds with equality. To do better, we need more information about X .

9.2 Chebyshev's Inequality

Theorem 9.2 (Chebyshev's inequality)

If X has mean μ and variance σ^2 , then for every $c > 0$,

$$\mathbb{P}(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}. \quad (9.5)$$

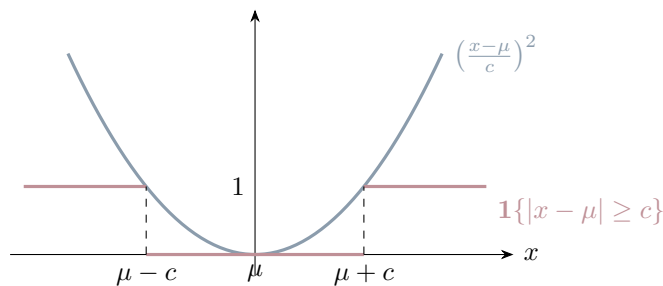
With $c = k\sigma$:

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}. \quad (9.6)$$

Proof. Both sides of $|X - \mu| \geq c$ are nonnegative, so squaring does not change the event: $\{|X - \mu| \geq c\} = \{(X - \mu)^2 \geq c^2\}$. Apply Markov to the nonnegative random variable $(X - \mu)^2$ with $a = c^2$:

$$\mathbb{P}(|X - \mu| \geq c) = \mathbb{P}((X - \mu)^2 \geq c^2) \leq \frac{\mathbb{E}[(X - \mu)^2]}{c^2} = \frac{\sigma^2}{c^2}. \quad (9.7)$$

□



The parabola $\left(\frac{x-\mu}{c}\right)^2$ lies above the indicator; averaging both sides gives Chebyshev. It uses the variance, not just the mean.

Example 9.2 (Chebyshev for a uniform)

Let $X \sim \text{Uniform}[0, 1]$. Then $\mu = \frac{1}{2}$, $\mathbb{E}[X^2] = \int_0^1 x^2 dx = \frac{1}{3}$, and

$$\sigma^2 = \mathbb{E}[X^2] - \mu^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}. \quad (9.8)$$

With $c = 0.4$, the exact value is $\mathbb{P}(|X - \frac{1}{2}| \geq 0.4) = \mathbb{P}(X \leq 0.1) + \mathbb{P}(X \geq 0.9) = 0.1 + 0.1 = 0.2$,

while

$$\mathbb{P}(|X - \tfrac{1}{2}| \geq 0.4) \leq \frac{1/12}{(0.4)^2} = \frac{1}{12 \cdot 0.16} = \frac{1}{1.92} \approx 0.52. \quad (9.9)$$

Example 9.3 (Polling with Chebyshev)

To estimate the fraction p of voters who support a candidate, poll n people independently. Let $X_i \sim \text{Bernoulli}(p)$ indicate support and $M_n = \frac{1}{n} \sum_{i=1}^n X_i$. How large must n be so that $\mathbb{P}(|M_n - p| \geq 0.01) \leq 0.05$?

$\mathbb{E}[M_n] = p$ and $\text{Var}(M_n) = \frac{p(1-p)}{n}$ (Lecture 4 and Lecture 10). Since p is unknown, bound the variance:

$$p(1-p) = \frac{1}{4} - \left(p - \frac{1}{2}\right)^2 \leq \frac{1}{4}. \quad (9.10)$$

By Chebyshev with $c = \varepsilon$,

$$\mathbb{P}(|M_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}. \quad (9.11)$$

Setting $\frac{1}{4n\varepsilon^2} \leq \delta$ with $\varepsilon = 0.01$, $\delta = 0.05$:

$$n \geq \frac{1}{4\varepsilon^2\delta} = \frac{1}{4(0.01)^2(0.05)} = \frac{1}{0.00002} = 50,000. \quad (9.12)$$

9.3 The Chernoff Bound

Theorem 9.3 (Chernoff bound)

For every b and every $s > 0$,

$$\mathbb{P}(X \geq b) \leq e^{-sb} M_X(s), \quad M_X(s) = \mathbb{E}[e^{sX}], \quad (9.13)$$

and therefore

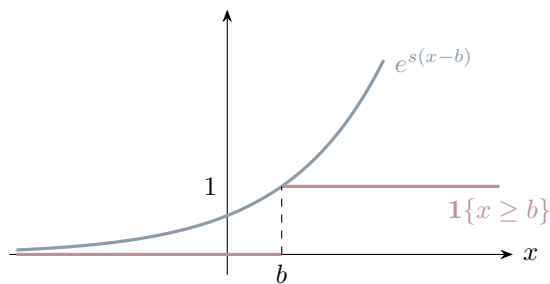
$$\mathbb{P}(X \geq b) \leq \min_{s>0} e^{-sb} M_X(s). \quad (9.14)$$

For the lower tail, $\mathbb{P}(X \leq b) \leq \min_{s<0} e^{-sb} M_X(s)$.

Proof. For $s > 0$, $x \mapsto e^{sx}$ is strictly increasing, so $\{X \geq b\} = \{e^{sX} \geq e^{sb}\}$. Apply Markov to the nonnegative random variable e^{sX} :

$$\mathbb{P}(X \geq b) = \mathbb{P}(e^{sX} \geq e^{sb}) \leq \frac{\mathbb{E}[e^{sX}]}{e^{sb}} = e^{-sb} M_X(s). \quad (9.15)$$

This holds for every $s > 0$, so it holds for the minimizing s . For $s < 0$, $x \mapsto e^{sx}$ is strictly decreasing, so $\{X \leq b\} = \{e^{sX} \geq e^{sb}\}$, and the same argument gives the lower-tail bound. $\square$



For any $s > 0$, the exponential $e^{s(x-b)}$ lies above the indicator $\mathbf{1}\{x \geq b\}$; averaging gives $\mathbb{P}(X \geq b) \leq e^{-sb} M_X(s)$, and we choose s to make the curve hug the indicator as tightly as possible.

Example 9.4 (Gaussian tail)

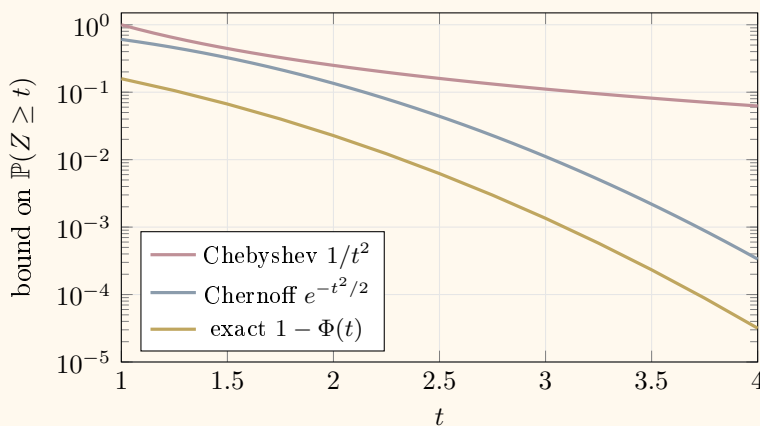
Let $Z \sim \mathcal{N}(0, 1)$, so $M_Z(s) = e^{s^2/2}$ (Lecture 8). For $t > 0$,

$$\mathbb{P}(Z \geq t) \leq \min_{s>0} e^{-st+s^2/2}. \quad (9.16)$$

The exponent $h(s) = \frac{s^2}{2} - st$ has $h'(s) = s - t = 0$ at $s^* = t$, and $h(t) = \frac{t^2}{2} - t^2 = -\frac{t^2}{2}$. Therefore

$$\mathbb{P}(Z \geq t) \leq e^{-t^2/2}. \quad (9.17)$$

Compare with Chebyshev, which gives $\mathbb{P}(Z \geq t) \leq \mathbb{P}(|Z| \geq t) \leq \frac{1}{t^2}$:



Chebyshev decays polynomially; Chernoff decays like the true tail, up to a smaller factor.

Example 9.5 (Chernoff for coin flips)

Let $S_n \sim \text{Binomial}(n, p)$, a sum of n independent Bernoulli(p) flips, so $M_{S_n}(s) = (1-p+pe^s)^n$: each flip has $\mathbb{E}[e^{sX_i}] = (1-p)e^0 + pe^s$, and the MGF of an independent sum is the product (Lecture 8). For $p < q < 1$,

$$\mathbb{P}(S_n \geq qn) \leq \min_{s>0} e^{-sqn} (1-p+pe^s)^n = \min_{s>0} \exp\left(n[-sq + \ln(1-p+pe^s)]\right). \quad (9.18)$$

Set the derivative of the bracket to zero:

$$-q + \frac{pe^s}{1-p+pe^s} = 0 \implies pe^s = q(1-p+pe^s) \implies pe^s(1-q) = q(1-p) \implies e^{s^*} = \frac{q(1-p)}{p(1-q)}, \quad (9.19)$$

and $s^* > 0$ since $q > p$. At s^* , $1-p+pe^{s^*} = 1-p + \frac{q(1-p)}{1-q} = \frac{(1-p)(1-q+q)}{1-q} = \frac{1-p}{1-q}$, so the bracket equals

$$\begin{aligned} -q \ln \frac{q(1-p)}{p(1-q)} + \ln \frac{1-p}{1-q} &= -q \ln \frac{q}{p} - q \ln \frac{1-p}{1-q} + \ln \frac{1-p}{1-q} \\ &= -q \ln \frac{q}{p} + (1-q) \ln \frac{1-p}{1-q} \\ &= -\left[q \ln \frac{q}{p} + (1-q) \ln \frac{1-q}{1-p} \right] \triangleq -D(q \| p). \end{aligned} \quad (9.20)$$

Therefore

$$\mathbb{P}(S_n \geq qn) \leq e^{-n D(q \| p)}. \quad (9.21)$$

Numbers. $n = 100$ fair flips ($p = \frac{1}{2}$), at least 75 heads ($q = \frac{3}{4}$):

$$D\left(\frac{3}{4} \| \frac{1}{2}\right) = \frac{3}{4} \ln \frac{3/4}{1/2} + \frac{1}{4} \ln \frac{1/4}{1/2} = \frac{3}{4} \ln 1.5 + \frac{1}{4} \ln 0.5 \approx 0.304 - 0.173 = 0.131. \quad (9.22)$$

Markov	$\frac{\mathbb{E}[S_n]}{75} = \frac{50}{75}$	≈ 0.67
Chebyshev	$\mathbb{P}(S_n - 50 \geq 25) \leq \frac{\text{Var}(S_n)}{25^2} = \frac{25}{625}$	$= 0.04$
Chernoff	$e^{-100 D(3/4 \ 1/2)} \approx e^{-13.1}$	$\approx 2.1 \times 10^{-6}$
Exact	$\sum_{k=75}^{100} \binom{100}{k} 2^{-100}$	$\approx 2.8 \times 10^{-7}$

9.4 Jensen's Inequality

Definition 9.1 (Convex function)

A function g is **convex** if for all x, y and $\lambda \in [0, 1]$,

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y) : \quad (9.23)$$

every chord lies above the graph. If g is differentiable, convexity is equivalent to every tangent line lying below the graph:

$$g(x) \geq g(c) + g'(c)(x-c) \quad \text{for all } x, c. \quad (9.24)$$

g is **concave** if $-g$ is convex. Examples: x^2 , e^{sx} , $|x|$ are convex; $\ln x$ and $\sqrt{x}$ are concave on $(0, \infty)$.

Theorem 9.4 (Jensen's inequality)

If g is convex, then

$$g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)]. \quad (9.25)$$

If g is concave, the inequality reverses: $g(\mathbb{E}[X]) \geq \mathbb{E}[g(X)]$.

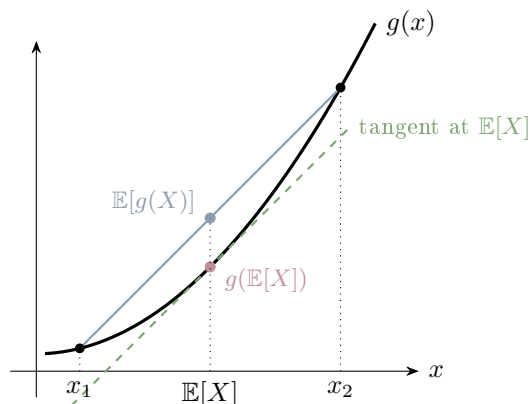
Proof. Take the tangent line (9.24) at $c = \mathbb{E}[X]$ and substitute $x = X$:

$$g(X) \geq g(\mathbb{E}[X]) + g'(\mathbb{E}[X])(X - \mathbb{E}[X]). \quad (9.26)$$

Take expectations. $g(\mathbb{E}[X])$ and $g'(\mathbb{E}[X])$ are constants, and $\mathbb{E}[X - \mathbb{E}[X]] = 0$:

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X]) + g'(\mathbb{E}[X]) \cdot 0 = g(\mathbb{E}[X]). \quad (9.27)$$

(For a convex g that is not differentiable, the same argument works with any supporting line through $(\mathbb{E}[X], g(\mathbb{E}[X]))$.) For concave g , apply the result to $-g$. $\square$



For X equal to x_1 or x_2 with probability $\frac{1}{2}$ each, $\mathbb{E}[g(X)]$ is the midpoint of the chord, which lies above $g(\mathbb{E}[X])$ on the curve.

Example 9.6 (Consequences of Jensen)

1. $g(x) = x^2$ (convex): $(\mathbb{E}[X])^2 \leq \mathbb{E}[X^2]$, i.e. $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \geq 0$.
2. $g(x) = e^{sx}$ (convex): $e^{s\mathbb{E}[X]} \leq M_X(s)$, so an MGF is never smaller than $e^{s\mathbb{E}[X]}$.
3. $g(x) = \ln x$ (concave): $\mathbb{E}[\ln X] \leq \ln \mathbb{E}[X]$. For X uniform on positive numbers $\{x_1, \dots, x_n\}$,

$$\frac{1}{n} \sum_{i=1}^n \ln x_i \leq \ln \left(\frac{1}{n} \sum_{i=1}^n x_i \right) \implies \left(\prod_{i=1}^n x_i \right)^{1/n} \leq \frac{1}{n} \sum_{i=1}^n x_i, \quad (9.28)$$

the arithmetic–geometric mean inequality.

4. In the St. Petersburg game (Lecture 5), $\mathbb{E}[\log_2 W] = 2$ is finite even though $\mathbb{E}[W] = \infty$, consistent with $\mathbb{E}[\log_2 W] \leq \log_2 \mathbb{E}[W]$.

9.5 Hoeffding's Inequality

Chernoff requires the MGF. For *bounded* random variables, we can bound the MGF without knowing the distribution.

Lemma 9.1 (Hoeffding's lemma)

If $a \leq X \leq b$ and $\mathbb{E}[X] = 0$, then for every $s \in \mathbb{R}$,

$$\mathbb{E}[e^{sX}] \leq e^{s^2(b-a)^2/8}. \quad (9.29)$$

Proof. Since $\mathbb{E}[X] = 0$, we have $a \leq 0 \leq b$. Write $x \in [a, b]$ as the convex combination $x = \lambda a + (1 - \lambda)b$ with $\lambda = \frac{b-x}{b-a}$. By convexity of e^{sx} (9.23),

$$e^{sx} \leq \frac{b-x}{b-a} e^{sa} + \frac{x-a}{b-a} e^{sb}. \quad (9.30)$$

Substitute $x = X$ and take expectations, using $\mathbb{E}[X] = 0$:

$$\mathbb{E}[e^{sX}] \leq \frac{b}{b-a} e^{sa} - \frac{a}{b-a} e^{sb}. \quad (9.31)$$

Let $p = \frac{-a}{b-a} \in [0, 1]$ and $h = s(b-a)$, so $sa = -ph$ and $sb = h - ph$. The right side becomes

$$(1-p)e^{-ph} + pe^{h-ph} = e^{-ph}(1-p+pe^h) = e^{L(h)}, \quad L(h) = -ph + \ln(1-p+pe^h). \quad (9.32)$$

Differentiate:

$$\begin{aligned} L'(h) &= -p + \frac{pe^h}{1-p+pe^h}, \quad L(0) = 0, \quad L'(0) = -p + p = 0, \\ L''(h) &= \frac{pe^h(1-p+pe^h) - pe^h \cdot pe^h}{(1-p+pe^h)^2} = \frac{pe^h}{1-p+pe^h} \cdot \frac{1-p}{1-p+pe^h} = u(1-u) \leq \frac{1}{4}, \end{aligned} \quad (9.33)$$

where $u = \frac{pe^h}{1-p+pe^h} \in [0, 1]$ and $u(1-u) \leq \frac{1}{4}$ (as for $p(1-p)$ above). By Taylor's theorem with remainder, for some ξ between 0 and h ,

$$L(h) = L(0) + L'(0)h + \frac{L''(\xi)}{2} h^2 \leq 0 + 0 + \frac{1}{8} h^2 = \frac{s^2(b-a)^2}{8}. \quad (9.34)$$

□

Theorem 9.5 (Hoeffding's inequality)

Let $X_1, \dots, X_n$ be independent with $a_i \leq X_i \leq b_i$, and $S_n = \sum_{i=1}^n X_i$. For every $t > 0$,

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right). \quad (9.35)$$

In particular, if the X_i are i.i.d. in $[a, b]$ with mean μ and $M_n = \frac{S_n}{n}$,

$$\mathbb{P}(M_n - \mu \geq \varepsilon) \leq e^{-2n\varepsilon^2/(b-a)^2}, \quad \mathbb{P}(|M_n - \mu| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2/(b-a)^2}. \quad (9.36)$$

Proof. Let $C = \sum_i (b_i - a_i)^2$. Each centered term $X_i - \mathbb{E}[X_i]$ has mean 0 and lies in an interval of length $b_i - a_i$. By Chernoff, independence (Lecture 8), and Hoeffding's lemma, for $s > 0$,

$$\begin{aligned} \mathbb{P}(S_n - \mathbb{E}[S_n] \geq t) &\leq e^{-st} \mathbb{E} \left[e^{s \sum_i (X_i - \mathbb{E}[X_i])} \right] = e^{-st} \prod_{i=1}^n \mathbb{E} \left[e^{s(X_i - \mathbb{E}[X_i])} \right] \\ &\leq e^{-st} \prod_{i=1}^n e^{s^2(b_i - a_i)^2/8} = \exp \left(-st + \frac{s^2 C}{8} \right). \end{aligned} \quad (9.37)$$

Minimize the exponent: $-t + \frac{sC}{4} = 0$ at $s^* = \frac{4t}{C}$, giving

$$-\frac{4t}{C}t + \frac{16t^2}{C^2} \cdot \frac{C}{8} = -\frac{4t^2}{C} + \frac{2t^2}{C} = -\frac{2t^2}{C}. \quad (9.38)$$

For the sample mean, $\{M_n - \mu \geq \varepsilon\} = \{S_n - n\mu \geq n\varepsilon\}$, so take $t = n\varepsilon$ and $C = n(b-a)^2$: $\frac{2t^2}{C} = \frac{2n^2\varepsilon^2}{n(b-a)^2} = \frac{2n\varepsilon^2}{(b-a)^2}$. The lower tail is the same bound applied to $-X_i$, and adding the two tails gives the factor 2. $\square$

Example 9.7 (Polling with Hoeffding)

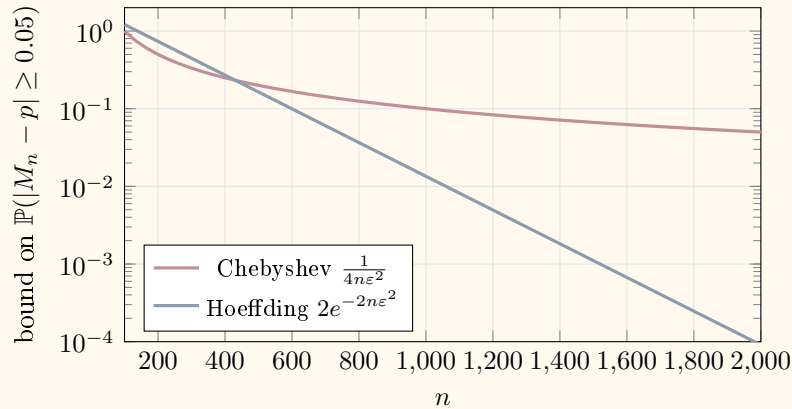
Return to polling: $X_i \in [0, 1]$, so $b - a = 1$. By (9.36),

$$\mathbb{P}(|M_n - p| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2} \leq \delta \iff n \geq \frac{\ln(2/\delta)}{2\varepsilon^2}. \quad (9.39)$$

With $\varepsilon = 0.01$, $\delta = 0.05$:

$$n \geq \frac{\ln 40}{2(0.01)^2} = \frac{3.689}{0.0002} \approx 18,445, \quad (9.40)$$

compared with 50,000 from Chebyshev. The two bounds as functions of n (with $\varepsilon = 0.05$):



Chebyshev decays like $\frac{1}{n}$; Hoeffding decays exponentially in n .

9.6 A First Look at the LLN and the CLT

Law of large numbers. For i.i.d. X_i with mean μ and variance σ^2 , Chebyshev gives

$$\mathbb{P}(|M_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0, \quad (9.41)$$

which is the weak law of large numbers (proved in full in Lecture 10).

Central limit theorem. $Z_n = \frac{\sqrt{n}(M_n - \mu)}{\sigma}$ is approximately $\mathcal{N}(0, 1)$ for large n (Lecture 10), so

$$\mathbb{P}(|M_n - \mu| \geq \varepsilon) = \mathbb{P}\left(|Z_n| \geq \frac{\varepsilon\sqrt{n}}{\sigma}\right) \approx 2\left(1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)\right). \quad (9.42)$$

For polling, $\sigma^2 \leq \frac{1}{4}$, and $2(1 - \Phi(1.96)) = 0.05$, so we need $\frac{\varepsilon\sqrt{n}}{\sigma} \geq 1.96$:

$$n \geq \frac{(1.96)^2 \sigma^2}{\varepsilon^2}, \quad \text{which holds for every } p \text{ once} \quad n \geq \frac{(1.96)^2}{4(0.01)^2} = 9,604. \quad (9.43)$$

This is an *approximation*, not a guaranteed bound like Chebyshev or Hoeffding.

Method	Requires	Decay	Polling n
Markov	$X \geq 0, \mathbb{E}[X]$	$1/a$	—
Chebyshev	$\text{Var}(X)$	$1/c^2, 1/n$	50,000
Chernoff	$M_X(s)$	exponential	—
Hoeffding	independent, bounded X_i	exponential in n	18,445
CLT	finite variance, large n	Gaussian (approximate)	9,604

End of Lecture 9.