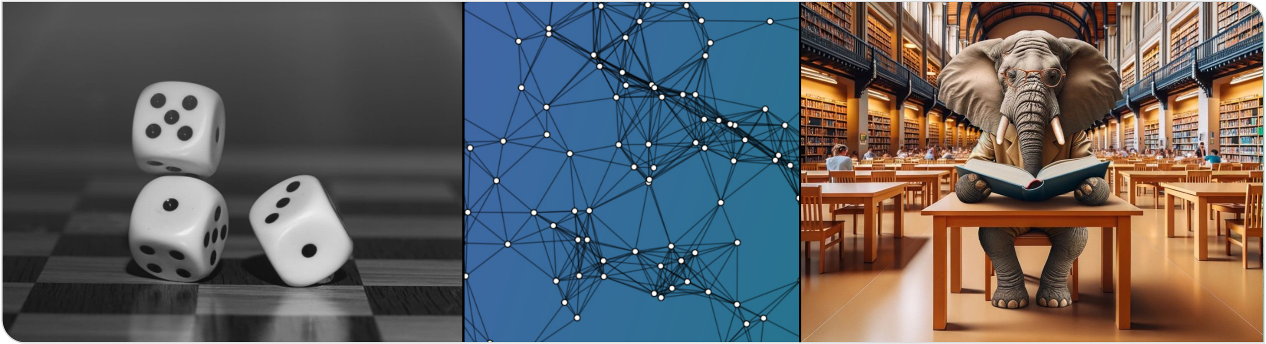
Karlsruhe Institute of Technology

Probability and Computing – Concentration Bounds

Stefan Walzer | WS 2025/2026

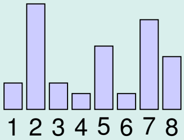


Content

How to describe the shape of a real-valued distribution?

Karlsruhe Institute of Technology

In general: Cannot be summarise in a few numbers



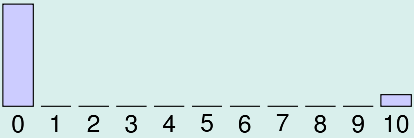
distribution on $[n]$ is given by $\vec{p} \in \mathbb{R}^n_{\geq 0}$ with $\sum_{i=1}^n p_i = 1$

- $n-1$ degrees of freedom
- a few numbers convey little information

Expectation is not always meaningful

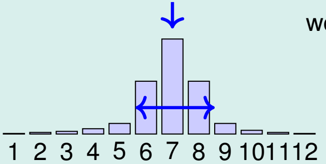
sniper "expected" to hit the target

I "expect" one hair in my soup



How to describe the shape of a real-valued distribution?

Many distributions: Unimodal, i.e. one “bump”



well summarised by

- where is the bump?
- how wide is the bump?
- how much is outside the bump?

Concentration Bound

Statement of the form:

$$\Pr[X \notin [a, b]] \le \varepsilon.$$

Bound is strong if $b - a$ and ε are small.

Why is this relevant for us?

Randomised algorithm might work only if $X \in [a, b]$. Want to bound the failure probability by ε.

Special Cases

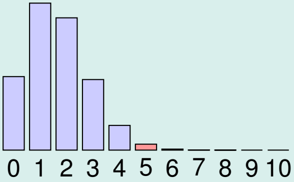
Concentration around Expectation: $\Pr[|X - \mathbb{E}[X]| > c] \le \varepsilon$ $([a, b] = [\mathbb{E}[X] - c, \mathbb{E}[X] + c])$

Tail bound: $\Pr[X > b] \le \varepsilon$ $(a = -\infty)$

Benchmark for What Follows

Karlsruhe Institute of Technology

Running Example



Let $X \sim \mathrm{Bin}(10, \frac{1}{6})$ the number of sixes when rolling a fair die 10 times.

$$\Pr[X \geq 5] \approx 0.016.$$

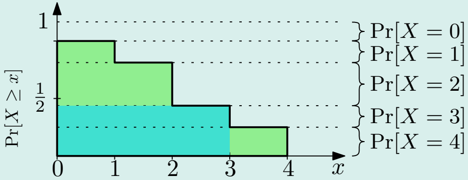
- 0.016 is calculated explicitly // only feasible for tiny examples
- later: we use general methods for obtaining bounds on $\Pr[X \geq 5]$

Markov: Tail bound for non-negative random variables

Markov Inequality

Let X be a non-negative random variable and $b \geq 0$. Then $\Pr[X \geq b] \leq \mathbb{E}[X]/b$.

Proof by Picture



From the picture ($b = 3$):

$$b \cdot \Pr[X \geq b] \leq \sum_{i \geq 1} \Pr[X \geq i] \overset{\text{TSF}}{=} \mathbb{E}[X]$$

Rearranging gives:
$\Pr[X \geq b] \leq \mathbb{E}[X]/b$.

Note: Tight if and only if $\Pr[X \in \{0, b\}] = 1$.

Actual Proof

$$\mathbb{E}[X] \overset{\text{LTE}}{=} \underbrace{\mathbb{E}[X \mid X \geq b]}_{\geq b} \cdot \Pr[X \geq b] + \underbrace{\mathbb{E}[X \mid X < b]}_{\geq 0} \cdot \Pr[X < b] \geq b \cdot \Pr[X \geq b]. \quad \square$$

Markov's Inequality: Benchmark

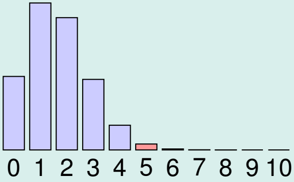
Markov Inequality

$\Pr[X \geq b] \leq \mathbb{E}[X]/b.$

Bound from Markov Inequality

$$\Pr[X \geq 5] \leq \frac{\mathbb{E}[X]}{5} = \frac{10/6}{5} \approx 0.333.$$

Running Example



Let $X \sim \mathrm{Bin}(10, \frac{1}{6})$ the number of sixes when rolling a fair die 10 times.

$$\Pr[X \geq 5] \approx 0.016.$$

Chebyshev's inequality ~~Corollaries to Markov's Inequality~~

Markov Inequality

Let X be a non-negative random variable and $b \geq 0$. Then $\Pr[X \geq b] \leq \mathbb{E}[X]/b$.

Corollary for non-negative X and strictly increasing $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$

$\Pr[X \geq b] = \Pr[f(X) \geq f(b)] \leq \mathbb{E}[f(X)]/f(b)$.

Corollary for any real-valued X

$\Pr[|X - \mathbb{E}X| \geq b] \leq \mathbb{E}[|X - \mathbb{E}[X]|]/b$. // but absolute values can be a pain to work with...

Chebyshev's Inequality

$\Pr[|X - \mathbb{E}X| \geq b] = \Pr[|X - \mathbb{E}X|^2 \geq b^2] \leq \mathbb{E}[|X - \mathbb{E}X|^2]/b^2 = \mathrm{Var}(X)/b^2$.

Moments[1]

Definitions

Let X be real-valued random variable and $n \in \mathbb{N}$. Then

$\mathbb{E}[X^n]$ is the nth **(raw) moment**, $\mathbb{E}[(X - \mathbb{E}[X])^n]$ is the nth **central moment**,
$\mathbb{E}[|X|^n]$ is the nth absolute[a] moment, $\mathbb{E}[|X - \mathbb{E}[X]|^n]$ is the nth absolute central moment.

[a] Note: $|\cdot|$ makes a difference only for odd n.

What's so special about the second central moment $\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$?

Intuitive Meaning

It's the mean squared distance.

Exercise: Nice mathematical properties

If X, Y are independent, then
$\mathrm{Var}(aX + bY) = a^2 \cdot \mathrm{Var}(X) + b^2 \cdot \mathrm{Var}(Y)$.

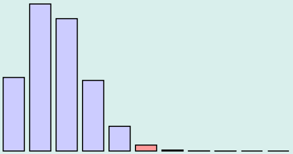
[1] deutsch: *das* Moment

Chebyshev's Inequality: Benchmark

Chebyshev's Inequality

$\Pr[|X - \mathbb{E}X| \geq b] \leq \mathrm{Var}(X)/b^2.$

Running Example



Let $X \sim \mathrm{Bin}(10, \frac{1}{6})$ the number of sixes when rolling a fair die 10 times.

$\Pr[X \geq 5] \approx 0.016.$

Variance Calculation

- let $X_i = [i\text{th roll is a six}]$, $X = \sum_{i=1}^{10} X_i$
- $\mathbb{E}[X_i] = \frac{1}{6}$, $\mathbb{E}[X] = \frac{10}{6} = \frac{5}{3}$
- $\mathrm{Var}(X_i) = \frac{1}{6} \cdot (\frac{5}{6})^2 + \frac{5}{6} \cdot (\frac{1}{6})^2 = \frac{5}{36}.$
- $\mathrm{Var}(X) = 10 \cdot \mathrm{Var}(X_1) = \frac{50}{36}.$

Bound from Chebyshev's Inequality

$\Pr[X \geq 5] = \Pr[X - \mathbb{E}[X] \geq 5 - \frac{5}{3} = \frac{10}{3}] \leq \Pr[|X - \mathbb{E}[X]| \geq \frac{10}{3}] \leq \mathrm{Var}(X)/(\frac{10}{3})^2 = \frac{50 \cdot 9}{36 \cdot 100} = \frac{1}{8} = 0.125.$

Moment Generating Function

Karlsruhe Institute of Technology

Definition

For real-valued random variable X call $M_X(t) = \mathbb{E}[e^{tX}]$ its *moment generating function*.

Remark: Why it's called "moment generating function".

$$M_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}\left[\sum_{i=0}^{\infty} \frac{(tX)^i}{i!}\right] = \sum_{i=0}^{\infty} \frac{t^i}{i!} \cdot \mathbb{E}[X^i].$$ // generating function[a] for $(\mathbb{E}[X^0], \mathbb{E}[X^1], \mathbb{E}[X^2], \dots)$

[a] https://en.wikipedia.org/wiki/Generating_function#Exponential_generating_function_(EGF)

Chernoff Bound

Let X be a real-valued RV and $b \geq 0$. Then

- $\Pr[X \geq b] \leq \inf_{t>0} \mathbb{E}[e^{tX}]/e^{tb}$ and
- $\Pr[X \leq b] \leq \inf_{t<0} \mathbb{E}[e^{tX}]/e^{tb}$.

Proof of first variant (second works simlarly)

Let $t > 0$ be arbitrary. Then $x \mapsto e^{tx}$ is increasing.

$$\Pr[X \geq b] = \Pr[e^{tX} \geq e^{tb}] \overset{\text{Markov}}{\leq} \mathbb{E}[e^{tX}]/e^{tb}.$$

Done, since $t > 0$ was arbitrary.

Generic Chernoff Bound: Benchmark

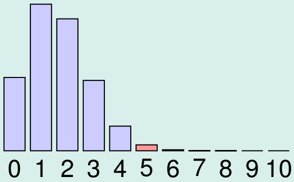
Chernoff Bound

$\Pr[X \geq b] \leq \inf_{t>0} \mathbb{E}[e^{tX}]/e^{tb}$

Calculations

- let $X_i = [i\text{th roll is a six}]$, $X = \sum_{i=1}^{10} X_i$
- $\mathbb{E}[e^{t \cdot X_i}] = \frac{1}{6} \cdot e^t + \frac{5}{6} = \frac{e^t+5}{6}$
- $\mathbb{E}[e^{t \cdot X}] = \mathbb{E}[e^{t \cdot X_1}]^{10} = (\frac{e^t+5}{6})^{10}$

Running Example



Let $X \sim \text{Bin}(10, \frac{1}{6})$ the number of sixes when rolling a fair die 10 times.

$$\Pr[X \geq 5] \approx 0.016.$$

Note: For independent $X_1, \ldots, X_n$

$\mathbb{E}[e^{t(X_1+\cdots+X_n)}] = \mathbb{E}[e^{tX_1} \cdot \ldots \cdot e^{tX_n}] = \mathbb{E}[e^{tX_1}] \cdot \ldots \cdot \mathbb{E}[e^{tX_n}]$.

Resulting Chernoff Bound

$$\Pr[X \geq 5] \leq \inf_{t>0} \mathbb{E}[e^{tX}]/e^{5t} = \inf_{t>0} \left(\frac{e^t+5}{6}\right)^{10} / e^{5t} \overset{\text{Wolfram Alpha}}{\approx} 0.053. \quad \text{// with } t = \ln(5)$$

What just happened?

Let $X \in \mathbb{N}$ be random variable and $p_i = \Pr[X = i]$ for $i \in \mathbb{N}$.
Consider bounds for $\Pr[X \geq b]$.

Markov

The bound is $\frac{\mathbb{E}[X]}{b}$. The contribution of "$X = i$" to is $\frac{p_i \cdot i}{b}$.

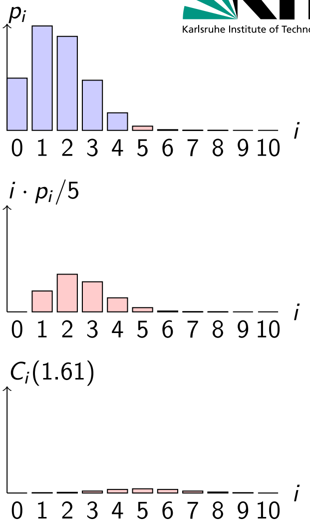
Chernoff

The bound is $\frac{\mathbb{E}[e^{tX}]}{e^{tb}}$. The contribution of "$X = i$" is $C_i(t) := \frac{p_i \cdot e^{ti}}{e^{tb}} = p_i \cdot e^{t(i-b)}$.

- if $i < b$ then $C_i(t)$ is decreasing in t
  $\hookrightarrow$ If $\Pr[X \geq b] = 0$ then Chernoff can prove this with $t \to \infty$.
- if $i > b$ then $C_i(t)$ is increasing in t
  $\hookrightarrow$ that's okay for a while as long as p_i are very small for $i > b$.

Remark: Finding the best t numerically is easy

$C_i''(t) > 0$ for all t, so $t \mapsto \mathbb{E}[e^{tX}]/e^{tb}$ is convex.



The Chernoff Bound for *your* Favourite Random Variable

Maybe someone already did the work...?

- The general Chernoff bound is hard to use
  - How to compute $\mathbb{E}[e^{tX}]$?
  - How to choose t?
- Many variants for special cases exist.

Multiplicative Form for Sums of Bernoulli RV

Theorem

Let $X = X_1 + \cdots + X_n$ with independent $X_i \sim \mathrm{Ber}(p_i)$ and $\mu := \mathbb{E}[X] = \sum_{i=1}^n p_i$. Then for any $\delta > 0$:

$$\Pr[X \geq (1+\delta)\mu] \leq \Bigg(\underbrace{\frac{e^\delta}{(1+\delta)^{1+\delta}}}_{f(\delta)}\Bigg)^\mu.$$

Note: $f(0) = 1$ and f is decreasing on $(0, \infty)$.

Hence for constant $\delta > 0$ the bound is exponentially small in μ.

Proof.

- $\mathbb{E}[e^{tX_i}] = p_i \cdot e^t + (1-p_i) = 1 + p_i(e^t - 1) \leq e^{p_i(e^t-1)}$.
- $\mathbb{E}[e^{tX}] = \prod_{i=1}^n \mathbb{E}[e^{tX_i}] \leq \prod_{i=1}^n e^{p_i(e^t-1)} = e^{\sum_{i=1}^n p_i(e^t-1)} = e^{\mu(e^t-1)}$.
- $\Pr[X \geq (1+\delta)\mu] \overset{\text{Chernoff}}{\leq} \frac{\mathbb{E}[e^{tX}]}{e^{(1+\delta)\mu t}} \leq \frac{e^{\mu(e^t-1)}}{e^{(1+\delta)\mu t}} = \left(\frac{e^{e^t-1}}{e^{(1+\delta)t}}\right)^\mu \overset{t=\ln(1+\delta)}{=} \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu$.

Consider $\ln(f(\delta)) = \delta - (1+\delta)\ln(1+\delta) =: g(\delta)$. Note that $g'(\delta) = 1 - \frac{1+\delta}{1+\delta} - \ln(1+\delta) = -\ln(1+\delta) < 0$ for $\delta > 0$.
Hence $g(\delta)$ is decreasing on $\mathbb{R}_{\geq 0}$ and so $f(\delta)$ is also decreasing on $\mathbb{R}_{\geq 0}$. $\square$

Multiplicative Form for Sums of Bernoulli RV (ii)

The function $f(\delta) = \frac{e^{\delta}}{(1+\delta)^{1+\delta}}$ is still inconvenient to work with...

Theorem (from last slide, slightly generalised)

Let $X = X_1 + \cdots + X_n$ with ind. $X_i \sim \mathrm{Ber}(p_i)$ and $\mu := \mathbb{E}[X]$.

Then $\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$ for any $\delta \geq 0$

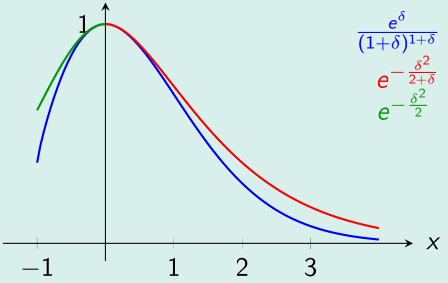
and $\Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}$ for any $\delta \in [0,1)$.

Corollary (in the same setting)

- $\Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\delta^2}{2+\delta}\mu}$ for $\delta \geq 0$,
- $\Pr[X \leq (1-\delta)\mu] \leq e^{-\frac{\delta^2}{2}\mu}$ for $\delta \in [0,1)$, and
- $\Pr[|X-\mu| \geq \delta\mu] \leq 2e^{-\frac{\delta^2}{3}\mu}$ for $0 \leq \delta \leq 1$

See also `https://en.wikipedia.org/wiki/Chernoff_bound`

"Proof by Plot"



Simplified Chernoff Bounds: Benchmark

Karlsruhe Institute of Technology

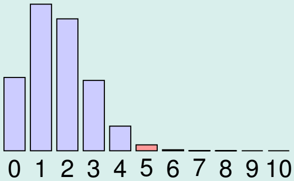
A Simplified Chernoff Bound

$\Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\delta^2}{2+\delta}\mu}$ for $0 \leq \delta$.

Calculations

- $\mu = \mathbb{E}[X] = \frac{10}{6}$
- $5 = 3 \cdot \frac{10}{6} = \mu + 2\mu \qquad // \; \delta = 2$

Running Example



Let $X \sim \mathrm{Bin}(10, \frac{1}{6})$ the number of sixes when rolling a fair die 10 times.

$\Pr[X \geq 5] \approx 0.016.$

Resulting Chernoff Bound

$$\Pr[X \geq 5] = \Pr[X \geq (1+2)\mu] \overset{\text{Chernoff}}{\leq} e^{-\frac{2^2}{2+2}\mu} = e^{-\frac{10}{6}} \approx 0.189.$$

How do the methods compare?

Karlsruhe Institute of Technology

Exercise

When throwing a fair die n times, bound the probability for seeing twice as many sixes as expected. . .

- . . . using Markov,
- . . . using Chebyshev,
- . . . using Chernoff (simplified).

Compare the results.

Conclusion

Summary: Methods for Obtaining Concentration Bounds for real-valued X

Method	Assumptions	Formula	Challenges	$\Pr[X \geq 5]$ in Benchmark
—	—	$\Pr[X \geq b] = \sum_{i=b}^{\infty} \Pr[X = i]$	tedious calculation	0.016
Markov	$X \geq 0$	$\Pr[X \geq b] \leq \frac{\mathbb{E}[X]}{b}$	compute $\mathbb{E}[X]$	0.333
Chebyshev	—	$\Pr[\lvert X - \mathbb{E}[X] \rvert \geq b] \leq \frac{\mathrm{Var}(X)}{b^2}$	compute $\mathrm{Var}(X)$	0.125
Chernoff	—	$\Pr[X \geq b] \leq \inf_{t>0} \frac{\mathbb{E}[e^{tX}]}{e^{tb}}$	compute $\mathbb{E}[e^{tX}]$, choose t	0.053
simpl. Ch.	X sum of ind. Ber-RVs	$\Pr[X \geq (1+\delta)\mathbb{E}[X]] \leq e^{-\frac{\delta^2}{2+\delta}\mathbb{E}[X]}$	compute $\mathbb{E}[X]$	0.189

Further Chernoff-flavoured concentration bounds (X aggregates independent $(X_i)_{i \in \mathbb{N}}$)

Hoeffding's inequality, McDiarmid's inequality, Bernstein inequalities.

Appendix: Possible Exam Questions I

- What is meant by a concentration bound?
- What do the Markov inequality / Chebyshev inequality / Chernoff inequality state?
  - What are the respective assumptions?
  - How difficult are the bounds to apply?
  - How good or bad are the resulting concentration bounds in comparison?
- Detailed questions:
  - Prove the Markov inequality.
  - Prove the Chebyshev inequality (from the Markov inequality).
    - Instead of the second moment, are higher moments suitable for deriving an inequality?
  - Prove the Chernoff inequality (from the Markov inequality).
  - What is the moment-generating function, and why is it called that?
  - How can we bound $\mathbb{E}[e^{tX}]$ when X is a sum of Bernoulli random variables?