Probability & Computing

Concentration

Expectation Management

What does it mean?

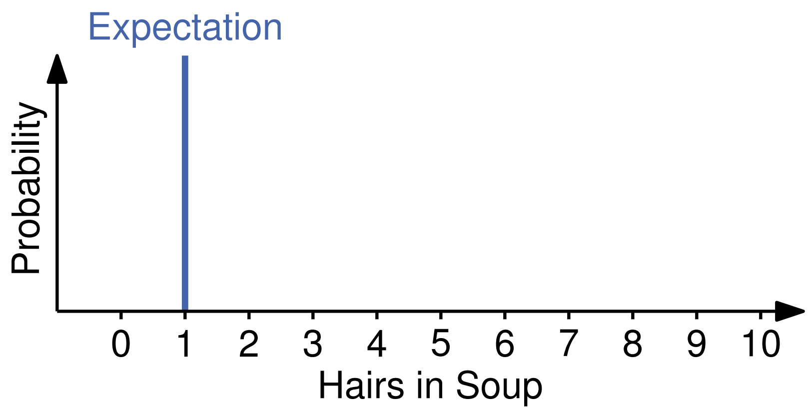
- "QuickSort has an *expected* running time of $O(n\log(n))$."
- "The vertex has an *expected* degree of c."
- "*In expectation* there is one hair in my soup."

Expectation

- The average of infinitly many trials
- How useful is that information in practice?
- Does not tell us much about the shape of the distribution
- Does not come with a level of certainty

Concentration

- In practice, expectation is often a good start
- But for meaningful statements, we need to know how likely we are close to the exepcation



Knowing that the expected value is 1 hair:

How likely is it that I get at least 10?

Not at all Somewhat

How likely is it that I get less than 2?

Extremely Somewhat

Definition: A **concentration inequality** bounds the probability of a random variable to deviate from a given value (typically its expectation) by a certain amount.

Markov's Inequality

Karlsruhe Institute of Technology

About Markov

- Andrei "The Furious" Andreyevich Markov (Russian mathematician)
- Unhappy with the state of living at the time (1921)
- Informed St. Petersburg Academy of Sciences that he could not attend meetings due to not having shoes
- After getting shoes from the Communist Party he replied:

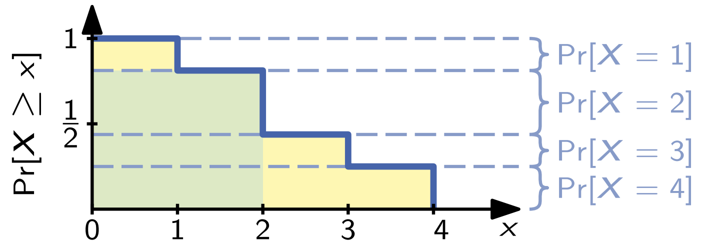
  Finally, I received footwear. However, it is stupidly stitched together and does not accord with my measurements. Thus I cannot attend the meetings. I propose placing the footwear in a museum, as an example of the material culture of the current time.

"Shape, The hidden geometry of absolutely everything", Jordan Ellenberg

Markov's Inequality

Theorem (Markov's inequality): Let X be a non-negative random variable and let $a > 0$. Then, $\Pr[X \geq a] \leq \mathbb{E}[X]/a$.

Visual Proof



$$\mathbb{E}[X] = \sum_x x \cdot \Pr[X = x] \geq a \cdot \Pr[X \geq a]$$

fits into

Proof

$$\mathbb{E}[X] = \underbrace{\mathbb{E}[X \mid X < a]}_{\geq 0} \cdot \underbrace{\Pr[X < a]}_{\geq 0} + \underbrace{\mathbb{E}[X \mid X \geq a]}_{\geq a} \cdot \Pr[X \geq a] \geq a \cdot \Pr[X \geq a] \checkmark$$

Corollary: Let X be a non-negative rand. var. and $a > 0$. Then, $\Pr[X \geq a \cdot \mathbb{E}[X]] \leq 1/a$.

- "*In expectation* there is one hair in my soup."
  - How likely is it that I get at least 10? $\Pr[X \geq 10] \leq 1/10$
  - How likely is that I get less than 2? $\Pr[X < 2] = 1 - \Pr[X \geq 2] \geq 1 - 1/2 = 1/2$

Oh no...

Application: Unfair Coins

- The sum of 20 unfair $\{0,1\}$-coin tosses: $X \sim \mathrm{Bin}(20, \frac{1}{5})$
- What is the probability of getting at least 16 ones?

1 0 1 0 0 1 0 0 1 1

0 0 0 0 1 0 0 0 1 1

$X = 8$

$$\Pr[X \geq 16] \leq \underbrace{\mathbb{E}[X]}_{20 \cdot \frac{1}{5} = 4}/16 = 0.25$$

Markov: X non-negative, $a > 0$: $\Pr[X \geq a] \leq \mathbb{E}[X]/a$.

- How tight is that bound? Not very?

$$\Pr[X \geq 16] = \sum_{k=16}^{20} \binom{20}{k} \left(\tfrac{1}{5}\right)^k \cdot \left(1 - \tfrac{1}{5}\right)^{20-k} \approx 0.0000000138$$

Maybe it is just a weak bound?

Fair Coin

- A single $\{0,1\}$-coin toss: $Y \sim \mathrm{Ber}(\frac{1}{2})$
- What is the probability of getting at least 1?
  - Clearly: $\Pr[Y \geq 1] = \Pr[Y = 1] = \frac{1}{2}$
  - Markov: $\Pr[Y \geq 1] \leq \mathbb{E}[Y]/1 = \mathbb{E}[Y] = \frac{1}{2}$

There exists a random variable and an $a > 0$ such that Markov's inequality is *exact*.

$\Rightarrow$ There is no better bound (that relies only on the expected value)

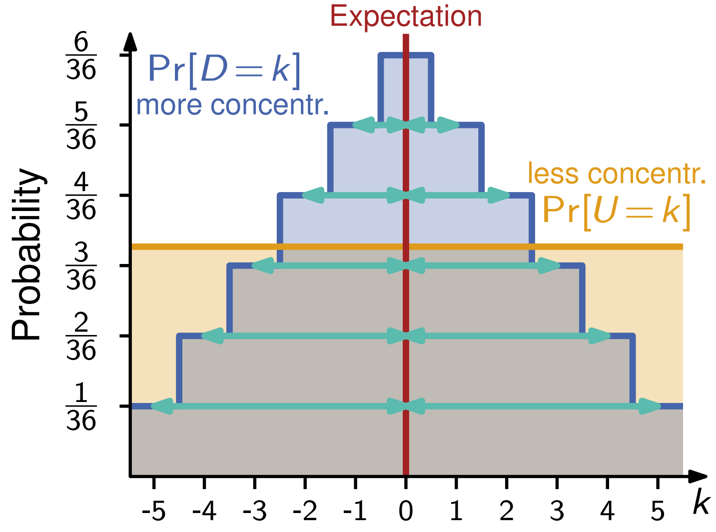
We need more information about the shape of the distribution!

Characterizing the Shape of a Distribution

- How much information do we need to characterize the shape of a distribution?

Example

- X, Y independent fair die-rolls, $D = X - Y$
- U uniform distribution over $\{-5, -4, \ldots, 5\}$
- Consider all probabilities individually Tedious... We need to aggregate!



Expectation? $f(k) = k$

$\mathbb{E}[D] = \sum_k \Pr[D = k] \cdot k = 0$

$\mathbb{E}[U] = \sum_k \Pr[U = k] \cdot k = 0$

Same value, different shapes
(also just seen with Markov: $\mathbb{E}$ not enough)

- Problem: $+$ & $-$ terms cancel

$\Rightarrow$ Fix: absolute value $f(k) = |k|$

$\mathbb{E}[|D|] = \sum_k \Pr[D = k] \cdot |k| \approx 1.945$

$\mathbb{E}[|U|] = \sum_k \Pr[U = k] \cdot |k| \approx 2.727$

Distance to $\mathbb{E}$

More concentrated!
Smaller expected distance to $\mathbb{E}$

- Problem: Nobody likes absolute value

$\Rightarrow$ Fix: square instead $f(k) = k^2$

$\mathbb{E}[D^2] = \sum_k \Pr[D = k] \cdot k^2 \approx 5.833$

$\mathbb{E}[U^2] = \sum_k \Pr[U = k] \cdot k^2 = 10.0$

Squared distance to $\mathbb{E}$

These are just expectations of functions of random variables!

Do you have a Moment?

Karlsruhe Institute of Technology

Expectation and Functions

- Random variable X taking values in a set S
- A function f, e.g. $f(X) = X^1$, $f(X) = |X|$, $f(X) = X^2$, $f(X) = \sqrt{X}$, $f(X) = X^3$, $f(X) = e^X$
- $\mathbb{E}[f(X)] = \sum_{x \in S} \Pr[X = x] \cdot f(x)$ *These turn out to be particularly useful!*

Moments

Definition: For random variable X and $n \in \mathbb{N}$ the ***n*-th raw moment** is $\mathbb{E}[X^n]$.

- Just seen: For $\mathbb{E}[X] = 0$, this captures distances to $\mathbb{E}[X]$ What if $\mathbb{E}[X] \neq 0$?

Definition: For random variable X and $n \in \mathbb{N}$ the ***n*-th central moment** is $\mathbb{E}[(X - \mathbb{E}[X])^n]$.

- Just seen: the $\underbrace{\text{2nd central moment}}_{\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathrm{Var}[X]}$ captures squared distances to the expected value

$$\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathrm{Var}[X]$$

- The smaller the variance, the more concentrated the random variable

... and with Markov's help, we can turn that insight into a concentration inequality!

Chebychev's Inequality

Markov's teacher! (Markov's inequality actually appeared earlier in Chebychev's works)

Theorem (Chebychev's inequality): Let X be a random variable with finite variance and let $b > 0$. Then, $\Pr[|X - \mathbb{E}[X]| \geq b] \leq \mathrm{Var}[X]/b^2$.

Proof

Markov: $Y \geq 0, a > 0$: $\Pr[Y \geq a] \leq \mathbb{E}[Y]/a$

$$\Pr[|X - \mathbb{E}[X]| \geq b] = \Pr\left[\overbrace{(X - \mathbb{E}[X])^2}^{\geq 0} \geq b^2\right] \leq \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]/b^2 = \mathrm{Var}[X]/b^2 \checkmark$$

Application: Unfair Coins

1 0 1 0 0 1 0 0 1 1 0 0 0 0 1 0 0 0 1 1

- $X \sim \mathrm{Bin}(20, \frac{1}{5})$, $\Pr[X \geq 16]$? | $\mathbb{E}[X] = 20 \cdot \frac{1}{5} = 4$ | $\mathrm{Var}[X] = 20 \cdot \frac{1}{5} \cdot (1 - \frac{1}{5}) = \frac{16}{5}$

  $\Pr[X \geq 16] = \sum_{k=16}^{20} \binom{20}{k} (\frac{1}{5})^k \cdot (1 - \frac{1}{5})^{20-k} \approx 0.0000000138$

  $X \sim \mathrm{Bin}(n, p) : \mathrm{Var}[X] = np(1-p)$

- Markov: $\Rightarrow \Pr[X \geq 16] \leq \mathbb{E}[X]/16 = 0.25$
- Chebychev:

$$\begin{aligned}\Pr[X \geq 16] &\leq \Pr[X \geq 16 \vee X \leq -8] \\ &= \Pr[|X - \mathbb{E}[X]| \geq 12] \\ &\leq \frac{\mathrm{Var}[X]}{12^2} = \frac{16}{5 \cdot 144} \approx 0.022\end{aligned}$$

Order of magnitude better than Markov!

$$\begin{aligned} & X \geq 16 \\ \Leftrightarrow\ & X - \mathbb{E}[X] \geq 16 - \mathbb{E}[X] \\ \Leftrightarrow\ & X - \mathbb{E}[X] \geq 12 \\ & |X - \mathbb{E}[X]| \geq 12 \Rightarrow X \geq 16 \text{ or } X \leq -8 \end{aligned}$$

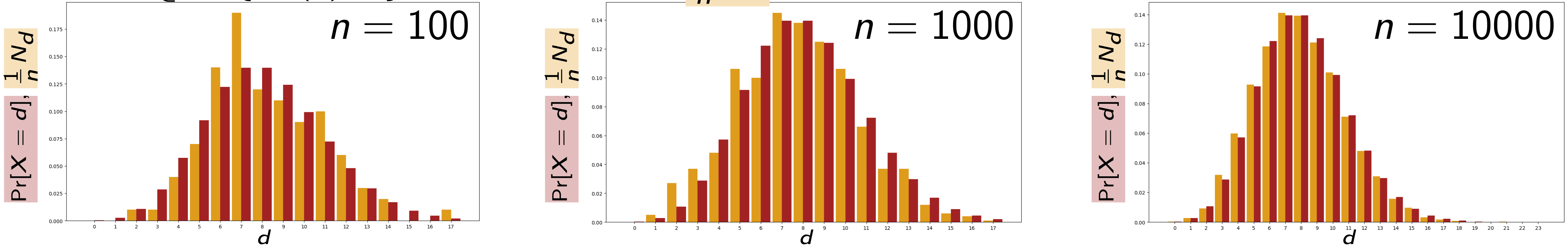
Application: ER – Degree Distribution

Karlsruhe Institute of Technology

Recap

- $G(n, p)$: Start with n nodes, connect any two with fixed probability p, independently
- Probability distribution of the degree of a *single* node v: $\deg(v) \sim \mathrm{Bin}(n-1, p)$
- For $p = c/n$ with $c \in \Theta(1)$ the degree of a vertex is approximately Poisson-distributed
  - Total variation distance of X, Y taking values in a set S:
    $d_{TV}(X, Y) = \frac{1}{2} \sum_{x \in S} |\Pr[X = x] - \Pr[Y = x]|$
  - For $\lambda = -n \log(1-p) = c + O(1/n)$ and $X \sim \mathrm{Pois}(\lambda)$ we have $d_{TV}(\deg(v), X) = o(1)$
- Empirical distribution of the degrees of *all* vertices in a graph $G = (V, E)$

$N_d = \sum_{v \in V} \mathbb{1}_{\{\deg(v) = d\}}$ (normalized: $\frac{1}{n} N_d$, for $n = |V|$)



Application: ER – Degree Distribution

Karlsruhe Institute of Technology

Theorem: Consider a $G(n,p)$ with $p = c/n$ for constant $c > 0$. For $\lambda = -n\log(1-p)$, let $X \sim \mathrm{Pois}(\lambda)$. Then for all $d > 0$ and every $\varepsilon > 0$ we have

$\lambda = c + O(1/n) \to c$ for $n \to \infty$

$$\lim_{n\to\infty} \Pr\left[\left|\Pr[X = d] - \tfrac{1}{n}N_d\right| \geq \varepsilon\right] = 0.$$

Proof

- Step 1: $\Pr[X{=}d]$ is close to the expectation of $\frac{1}{n}N_d$ $\quad \lim_{n\to\infty}\left|\Pr[X = d] - \mathbb{E}\left[\tfrac{1}{n}N_d\right]\right| = 0$ ✓

$$\left|\Pr[X = d] - \mathbb{E}\left[\tfrac{1}{n}N_d\right]\right| = \left|\Pr[X = d] - \Pr[\deg(v) = d]\right| \leq \textstyle\sum_{d\geq 0}\left|\Pr[X = d] - \Pr[\deg(v) = d]\right|$$

$$\mathbb{E}\left[\tfrac{1}{n}N_d\right] = \tfrac{1}{n}\mathbb{E}[N_d] = \tfrac{1}{n}\mathbb{E}\left[\textstyle\sum_{v\in V}\mathbb{1}_{\{\deg(v)=d\}}\right] = \tfrac{1}{n}\textstyle\sum_{v\in V}\mathbb{E}\left[\mathbb{1}_{\{\deg(v)=d\}}\right] = \tfrac{1}{n}\textstyle\sum_{v\in V}\Pr[\deg(v)=d] = \Pr[\deg(v) = d]$$

$$= 2\cdot d_{TV}(X, \deg(v))$$

$$= o(1) \xrightarrow{n\to\infty} 0 \text{ ✓}$$

Already shown last time!

$$d_{TV}(X,Y) = \tfrac{1}{2}\textstyle\sum_{x\in S}\left|\Pr[X = x] - \Pr[Y = x]\right|$$

- Step 2: $\frac{1}{n}N_d$ is concentrated $\quad \lim_{n\to\infty}\Pr\left[\left|\mathbb{E}\left[\tfrac{1}{n}N_d\right] - \tfrac{1}{n}N_d\right| \geq \varepsilon\right] = 0$

Step 2: Concentration of $\frac{1}{n} N_d$

$$\Pr\left[\left|\mathbb{E}\left[\tfrac{1}{n}N_d\right] - \tfrac{1}{n}N_d\right| \geq \varepsilon\right] \leq \underbrace{\mathrm{Var}\left[\tfrac{1}{n}N_d\right]} / \varepsilon^2$$

$$\lim_{n\to\infty} \Pr\left[\left|\mathbb{E}\left[\tfrac{1}{n}N_d\right] - \tfrac{1}{n}N_d\right| \geq \varepsilon\right] = 0$$

Chebychev: X finite variance, $b > 0$
$\Pr[|X - \mathbb{E}[X]| \geq b] \leq \mathrm{Var}[X]/b^2$

$$\left(\sum_i a_i\right)^2 = \sum_i a_i^2 + \sum_i \sum_{j\neq i} a_i a_j$$

$$\mathrm{Var}[\tfrac{1}{n}N_d] = \mathbb{E}[(\tfrac{1}{n}N_d)^2] - \mathbb{E}[\tfrac{1}{n}N_d]^2$$

$$= \frac{1}{n^2}\left(\underbrace{\mathbb{E}[(N_d)^2]} - \underbrace{\mathbb{E}[N_d]^2}_{= (n\Pr[\deg(v) = d])^2}\right)$$

(see Step 1)

$N_d = \sum_{v\in V} \mathbb{1}_{\{\deg(v)=d\}}$

$$= \mathbb{E}\left[\left(\sum_{v\in V} \mathbb{1}_{\{\deg(v)=d\}}\right)^2\right]$$

$$= \mathbb{E}\left[\sum_{v\in V}(\mathbb{1}_{\{\deg(v)=d\}})^2 + \sum_{v\in V}\sum_{u\neq v} \mathbb{1}_{\{\deg(v)=d\}} \cdot \mathbb{1}_{\{\deg(u)=d\}}\right]$$

Indicator RV X: $X^2 = X$, Lin. of Exp.

$$= \mathbb{E}\left[\sum_{v\in V} \mathbb{1}_{\{\deg(v)=d\}}\right] + \mathbb{E}\left[\sum_{v\in V}\sum_{u\neq v} \mathbb{1}_{\{\deg(v)=d\}} \cdot \mathbb{1}_{\{\deg(u)=d\}}\right]$$

Lin. of Exp.

$$= \sum_{v\in V} \underbrace{\mathbb{E}[\mathbb{1}_{\{\deg(v)=d\}}]}_{= \Pr[\deg(v) = d]} + \sum_{v\in V}\sum_{u\neq v} \underbrace{\mathbb{E}[\underbrace{\mathbb{1}_{\{\deg(v)=d\}} \cdot \mathbb{1}_{\{\deg(u)=d\}}}_{= 1 \text{ iff } \deg(v) = d \wedge \deg(u) = d}]}_{= \Pr[\deg(v) = d \wedge \deg(u) = d]}$$

$$= n \cdot \Pr[\deg(v) = d] + n(n-1) \cdot \Pr[\deg(v) = d \wedge \deg(u) = d]$$

Step 2: Concentration of $\frac{1}{n} N_d$

$$\Pr\left[\left|\mathbb{E}\left[\tfrac{1}{n}N_d\right] - \tfrac{1}{n}N_d\right| \geq \varepsilon\right] \leq \underbrace{\mathrm{Var}\left[\tfrac{1}{n}N_d\right]}/\varepsilon^2$$

$$\begin{aligned}
\mathrm{Var}[\tfrac{1}{n}N_d] &= \mathbb{E}[(\tfrac{1}{n}N_d)^2] - \mathbb{E}[\tfrac{1}{n}N_d]^2 \\
&= \tfrac{1}{n^2}\left(\mathbb{E}[(N_d)^2] - \mathbb{E}[N_d]^2\right) \\
&= \tfrac{1}{n^2}\big(n \Pr[\deg(v)=d] \\
&\qquad + n(n-1)\Pr[\deg(v)=d \wedge \deg(u)=d] \\
&\qquad - (n\Pr[\deg(v)=d])^2\big) \\
&= \tfrac{1}{n} \Pr[\deg(v)=d] \qquad \leq 1 \\
&\qquad + \tfrac{n-1}{n} \Pr[\deg(v)=d \wedge \deg(u)=d] \qquad \leq 1 \\
&\qquad - \Pr[\deg(v)=d]^2 \\
&\leq \tfrac{1}{n} + \Pr[\deg(v)=d \wedge \deg(u)=d] \\
&\qquad - \Pr[\deg(v)=d]^2
\end{aligned}$$

$$\lim_{n\to\infty} \Pr\left[\left|\mathbb{E}\left[\tfrac{1}{n}N_d\right] - \tfrac{1}{n}N_d\right| \geq \varepsilon\right] = 0$$

Chebychev: X finite variance, $b > 0$
$\Pr[|X - \mathbb{E}[X]| \geq b] \leq \mathrm{Var}[X]/b^2$

$$\left(\sum_i a_i\right)^2 = \sum_i a_i^2 + \sum_i \sum_{j\neq i} a_i a_j$$

Step 2: Concentration of $\frac{1}{n} N_d$

$$\Pr\left[\left|\mathbb{E}\left[\tfrac{1}{n}N_d\right] - \tfrac{1}{n}N_d\right| \geq \varepsilon\right] \leq \underbrace{\mathrm{Var}\left[\tfrac{1}{n}N_d\right]}/\varepsilon^2$$

$$\begin{aligned}
\mathrm{Var}[\tfrac{1}{n}N_d] &= \mathbb{E}[(\tfrac{1}{n}N_d)^2] - \mathbb{E}[\tfrac{1}{n}N_d]^2 \\
&= \tfrac{1}{n^2}\left(\mathbb{E}[(N_d)^2] - \mathbb{E}[N_d]^2\right) \\
&\leq \tfrac{1}{n} + \Pr[\deg(v) = d \wedge \deg(u) = d] \\
&\quad - \Pr[\deg(v) = d]\Pr[\deg(u) = d] \\
&= \tfrac{1}{n} + \Pr[X_1 + Y_1 = d \wedge X_1 + Y_2 = d] \\
&\quad - \Pr[X_1 + Y_1 = d]\Pr[X_2 + Y_2 = d] \\
&= \tfrac{1}{n} + \Pr[X_1 + Y_1 = d \wedge X_1 + Y_2 = d] \\
&\quad - \Pr[X_1 + Y_1 = d \wedge X_2 + Y_2 = d] \\
&\leq \tfrac{1}{n} + \Pr[X_1 + Y_1 = d \wedge X_1 + Y_2 = d \\
&\quad \wedge (X_1 + Y_1 \neq d \vee X_2 + Y_2 \neq d)]
\end{aligned}$$

For the whole event to occur, this needs to happen

Which excludes this from happening

$\deg(v) \overset{d}{=} \deg(u)$

$$\lim_{n\to\infty} \Pr\left[\left|\mathbb{E}\left[\tfrac{1}{n}N_d\right] - \tfrac{1}{n}N_d\right| \geq \varepsilon\right] = 0$$

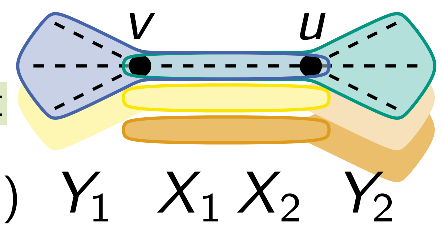
Chebychev: X finite variance, $b > 0$
$\Pr[|X - \mathbb{E}[X]| \geq b] \leq \mathrm{Var}[X]/b^2$

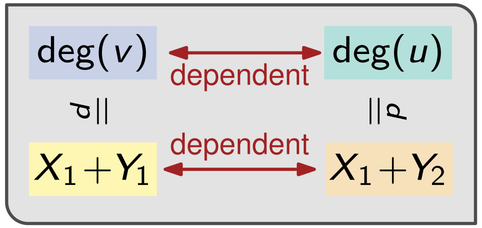
$$\left(\sum_i a_i\right)^2 = \sum_i a_i^2 + \sum_i \sum_{j \neq i} a_i a_j$$

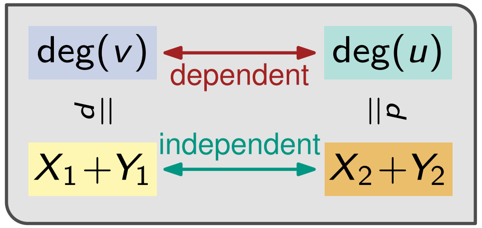
Fréchet: $\Pr[A] - \Pr[B] \leq \Pr[A \wedge \bar{B}]$

Couplings

- Consider $\deg(u)$ and $\deg(v)$
- $Y_1, Y_2 \sim \mathrm{Bin}(n-2, p)$ } independent
- $X_1, X_2 \sim \mathrm{Ber}(p)$
- $(\deg(v), \deg(u)) \overset{d}{=} (X_1 + Y_1, X_1 + Y_2)$







Step 2: Concentration of $\frac{1}{n} N_d$

$$\Pr\left[\left|\mathbb{E}\left[\tfrac{1}{n}N_d\right] - \tfrac{1}{n}N_d\right| \geq \varepsilon\right] \leq \underbrace{\mathrm{Var}\left[\tfrac{1}{n}N_d\right]}/\varepsilon^2$$

$$\begin{aligned}
\mathrm{Var}[\tfrac{1}{n}N_d] &= \mathbb{E}[(\tfrac{1}{n}N_d)^2] - \mathbb{E}[\tfrac{1}{n}N_d]^2 \\
&= \frac{1}{n^2}\left(\mathbb{E}[(N_d)^2] - \mathbb{E}[N_d]^2\right) \\
&\leq \frac{1}{n} + \Pr[\deg(v) = d \wedge \deg(u) = d] \\
&\quad - \Pr[\deg(v) = d]\Pr[\deg(u) = d] \qquad \deg(v) \stackrel{d}{=} \deg(u) \\
&= \frac{1}{n} + \Pr[X_1 + Y_1 = d \wedge X_1 + Y_2 = d] \\
&\quad - \Pr[X_1 + Y_1 = d]\Pr[X_2 + Y_2 = d] \\
&= \frac{1}{n} + \Pr[X_1 + Y_1 = d \wedge X_1 + Y_2 = d] \\
&\quad - \Pr[X_1 + Y_1 = d \wedge X_2 + Y_2 = d] \\
&\leq \frac{1}{n} + \Pr[X_1 + Y_1 = d \wedge X_1 + Y_2 = d \\
&\quad \wedge (X_1 + Y_1 \neq d \vee X_2 + Y_2 \neq d)] = \frac{1}{n} + \Pr[X_1 + Y_1 = d \wedge X_1 + Y_2 = d \wedge X_2 + Y_2 \neq d]
\end{aligned}$$

For the whole event to occur, this needs to happen

Which excludes this from happening

$$\lim_{n\to\infty} \Pr\left[\left|\mathbb{E}\left[\tfrac{1}{n}N_d\right] - \tfrac{1}{n}N_d\right| \geq \varepsilon\right] = 0$$

Chebychev: X finite variance, $b > 0$
$\Pr[|X - \mathbb{E}[X]| \geq b] \leq \mathrm{Var}[X]/b^2$

$$\left(\sum_i a_i\right)^2 = \sum_i a_i^2 + \sum_i \sum_{j \neq i} a_i a_j$$

Fréchet: $\Pr[A] - \Pr[B] \leq \Pr[A \wedge \bar{B}]$

Step 2: Concentration of $\frac{1}{n} N_d$

$$\Pr\left[\left|\mathbb{E}\left[\tfrac{1}{n}N_d\right] - \tfrac{1}{n}N_d\right| \geq \varepsilon\right] \leq \underbrace{\mathrm{Var}\left[\tfrac{1}{n}N_d\right]}/\varepsilon^2 \xrightarrow{n \to \infty} 0$$

$$\lim_{n\to\infty} \Pr\left[\left|\mathbb{E}\left[\tfrac{1}{n}N_d\right] - \tfrac{1}{n}N_d\right| \geq \varepsilon\right] = 0 \quad ✓$$

Chebychev: X finite variance, $b > 0$
$\Pr[|X - \mathbb{E}[X]| \geq b] \leq \mathrm{Var}[X]/b^2$

$$\left(\sum_i a_i\right)^2 = \sum_i a_i^2 + \sum_i \sum_{j \neq i} a_i a_j$$

Fréchet: $\Pr[A] - \Pr[B] \leq \Pr[A \wedge \bar{B}]$

$$\mathrm{Var}[\tfrac{1}{n}N_d] = \mathbb{E}[(\tfrac{1}{n}N_d)^2] - \mathbb{E}[\tfrac{1}{n}N_d]^2 \leq \tfrac{1}{n} + 2p = \tfrac{1}{n} + 2\tfrac{c}{n} \xrightarrow{n \to \infty} 0$$

$$= \tfrac{1}{n^2}\left(\mathbb{E}[(N_d)^2] - \mathbb{E}[N_d]^2\right)$$

$$\leq \tfrac{1}{n} + \Pr[\deg(v) = d \wedge \deg(u) = d]$$

$$- \Pr[\deg(v) = d] \Pr[\deg(u) = d] \quad \deg(v) \stackrel{d}{=} \deg(u)$$

$$\leq \tfrac{1}{n} + \Pr[X_1 + Y_1 = d \wedge X_1 + Y_2 = d \wedge X_2 + Y_2 \neq d]$$

$$= \tfrac{1}{n} + \Pr[X_1 + Y_1 = d \wedge X_1 + Y_2 = d \wedge X_2 + Y_2 \neq d \mid X_1 = 0] \overbrace{\Pr[X_1 = 0]}^{\leq 1}$$

$$+ \underbrace{\Pr[X_1 + Y_1 = d \wedge X_1 + Y_2 = d \wedge X_2 + Y_2 \neq d \mid X_1 = 1]}_{\leq 1} \Pr[X_1 = 1]$$

Law of total probability

$$\leq \tfrac{1}{n} + \Pr[Y_1 = d \wedge Y_2 = d \wedge X_2 + Y_2 \neq d \mid X_1 = 0]$$

$$+ \Pr[X_1 = 1]$$

$\Rightarrow X_2 = 1$

- $Y_1, Y_2 \sim \mathrm{Bin}(n-2, p)$
- $X_1, X_2 \sim \mathrm{Ber}(p)$

} independent

independent

$$= \tfrac{1}{n} + \Pr[Y_1 = d \wedge Y_2 = d \wedge X_2 = 1 \mid \cancel{X_1 = 0}] + \Pr[X_1 = 1] \leq \tfrac{1}{n} + \Pr[X_2 = 1] + \Pr[X_1 = 1]$$

Application: ER – Degree Distribution

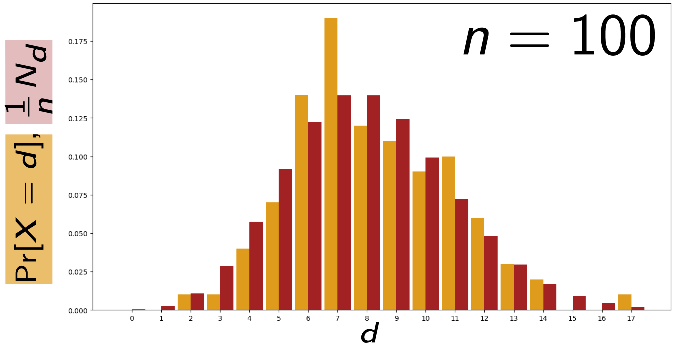
Theorem: Consider a $G(n,p)$ with $p = c/n$ for constant $c > 0$. For $\lambda = -n\log(1-p)$, let $X \sim \text{Pois}(\lambda)$. Then for all $d > 0$ and every $\varepsilon > 0$ we have

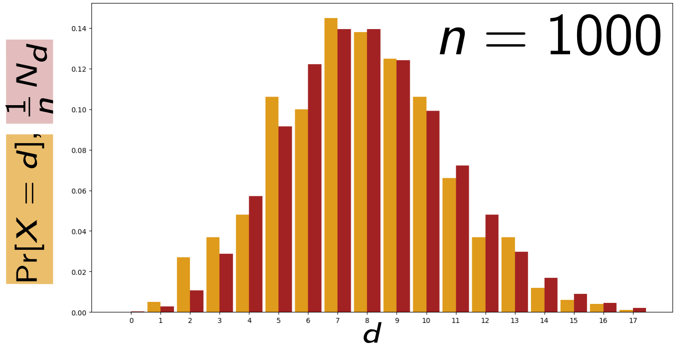
$\lambda = c + O(1/n) \to c$ for $n \to \infty$

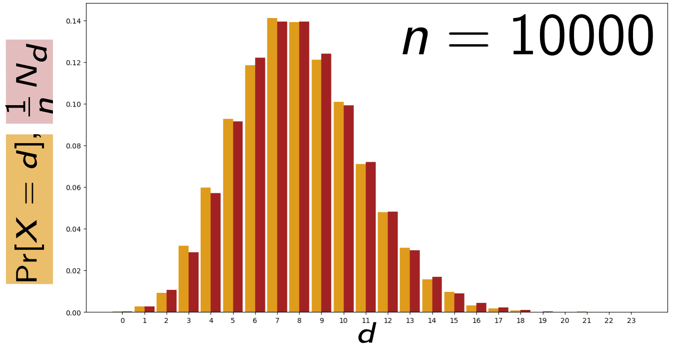
$$\lim_{n\to\infty} \Pr\left[\left|\Pr[X = d] - \tfrac{1}{n}N_d\right| \geq \varepsilon\right] = 0.$$

Proof

- Step 1: $\Pr[X = d]$ is close to the expectation of $\frac{1}{n}N_d$ $\quad \lim_{n\to\infty} \left|\Pr[X = d] - \mathbb{E}\left[\tfrac{1}{n}N_d\right]\right| = 0$ ✓
- Step 2: $\frac{1}{n}N_d$ is concentrated (via Chebychev) $\quad \lim_{n\to\infty} \Pr\left[\left|\mathbb{E}\left[\tfrac{1}{n}N_d\right] - \tfrac{1}{n}N_d\right| \geq \varepsilon\right] = 0$ ✓

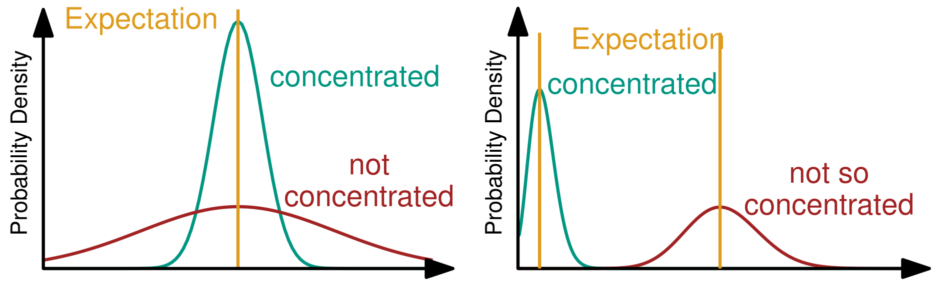






Concentration Bounds So Far

Definition: A **concentration inequality** bounds the probability of a random variable to deviate from a given value (typically its expectation) by a certain amount.



Markov

- based on expectation (first moment)
- X non-negative random variable and $a > 0$
  $\Pr[X \geq a] \leq \mathbb{E}[X]/a$
- tight

Chebychev

- based on variance (second moment)
- X random variable with finite variance and $b > 0$
  $\Pr[|X - \mathbb{E}[X]| \geq b] \leq \mathrm{Var}[X]/b^2$
- tight (stated without proof)

Can we utilize higher-order moments for even stronger bounds?

Another Moment Please

- The n-th raw moment of a random variable X is $\mathbb{E}[X^n]$
- We can capture *all* moments of X using a single function

Looks scary, but is again just $\mathbb{E}[f(X)]$ for $f(X) = e^{tX}$

Definition: For a random variable X the **moment generating function** is $M_X(t) = \mathbb{E}[e^{tX}]$

- Where the name comes from: For the n-th derivative $M_X^{(n)}(t)$ we have $M_X^{(n)}(0) = \mathbb{E}[X^n]$

(assuming the function exists in a neighborhood around 0)

Theorem: For independent random variables X, Y: $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$.

Proof $M_{X+Y}(t) = \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX} \cdot e^{tY}] = \mathbb{E}[e^{tX}] \cdot \mathbb{E}[e^{tY}] = M_X(t) \cdot M_Y(t)$ ✓

Concentration Inequality

Had his 100th birthday in 2023! Thought the bound (now named after him) to be so trivial that he didn't mention that it actually came from Herman Rubin.

"A conversation with Herman Chernoff", John Bather, Statist. Sci. 1996

Theorem (Chernoff Bounds): Let X be a random variable and $a > 0$.
Then, $\Pr[X \geq a] \leq \min_{t>0} \mathbb{E}[e^{tX}]/e^{ta}$ and $\Pr[X \leq a] \leq \min_{t<0} \mathbb{E}[e^{tX}]/e^{ta}$.

Proof for all $t > 0$: $\Pr[X \geq a] = \Pr[e^{tX} \geq e^{ta}] \leq \mathbb{E}[e^{tX}]/e^{ta}$

$\leq \min_{t>0} \mathbb{E}[e^{tX}]/e^{ta}$ ✓

for all $t < 0$: analogous. ✓

Markov: X non-negative, $b > 0$: $\Pr[X \geq b] \leq \mathbb{E}[X]/b$.

Get bounds for specific random variables by finding a good t!

Application: Binomial Distribution

Karlsruhe Institute of Technology

Theorem: Let $X \sim \text{Bin}(n,p)$. Then for any $\varepsilon > 0$

$$\Pr[X \geq (1+\varepsilon)\mathbb{E}[X]] \leq \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{(1+\varepsilon)}}\right)^{\mathbb{E}[X]}.$$

Proof Consider X as the sum of independent $X_i \sim \text{Ber}(p)$

$$M_{X_i}(t) = \mathbb{E}[e^{tX_i}] = \Pr[X_i = 0]\cdot e^{t\cdot 0} + \Pr[X_i = 1]\cdot e^{t\cdot 1}$$
$$= (1-p) + pe^t = 1 + (e^t - 1)p \leq e^{(e^t-1)p}$$

$1 + x \leq e^x$

$$M_X(t) = M_{\sum X_i}(t) = \prod_{i=1}^n M_{X_i}(t) \leq \prod_{i=1}^n e^{(e^t-1)p} = e^{(e^t-1)\cdot np}$$
$$= e^{(e^t-1)\mathbb{E}[X]}$$

$$\Pr[X \geq (1+\varepsilon)\mathbb{E}[X]] \leq \min_{t>0} \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\varepsilon)\mathbb{E}[X]}} \leq \min_{t>0} \frac{e^{(e^t-1)\mathbb{E}[X]}}{e^{t(1+\varepsilon)\mathbb{E}[X]}} = \min_{t>0}\left(\frac{e^{(e^t-1)}}{e^{t(1+\varepsilon)}}\right)^{\mathbb{E}[X]} \leq \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{(1+\varepsilon)}}\right)^{\mathbb{E}[X]} \checkmark$$

for $t = \log(1+\varepsilon)$

Chernoff: Random variable X and $a > 0$: $\Pr[X \geq a] \leq \min_{t>0} \mathbb{E}[e^{tX}]/e^{ta}$.

Mom. Gen. Function: $M_X(t) = \mathbb{E}[e^{tX}]$

Moment Addition: Independent X, Y: $M_{X+Y}(t) = M_X(t)\cdot M_Y(t)$.

Example 1 0 1 0 0 1 0 0 1 1 0 0 0 0 1 0 0 0 1 1

- Sum of 20 unfair $\{0,1\}$-coin tosses: $X \sim \text{Bin}(20, \frac{1}{5})$, $\mathbb{E}[X] = 4$
- $\Pr[X \geq 16] = \Pr[X \geq (1+3)\mathbb{E}[X]] \leq \left(\frac{e^3}{(1+3)^{1+3}}\right)^4 = \frac{e^{12}}{4^{4^4}} \approx 0.00003789$

Markov:	≤ 0.25
Chebychev:	$\lessapprox 0.022$
Actual:	≈ 0.0000000138

Chernoff – Simpler Versions

Theorem: Let $X \sim \text{Bin}(n, p)$. Then for any $\varepsilon > 0$

$$\Pr[X \geq (1+\varepsilon)\mathbb{E}[X]] \leq \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{(1+\varepsilon)}}\right)^{\mathbb{E}[X]}.$$

Chernoff: Random variable X and $a > 0$: $\Pr[X \geq a] \leq \min_{t>0} \mathbb{E}[e^{tX}]/e^{ta}$.

Corollary: Let $X \sim \text{Bin}(n, p)$. Then for any $t \geq 6\mathbb{E}[X]$, $\Pr[X \geq t] \leq 2^{-t}$.

Corollary: Let $X \sim \text{Bin}(n, p)$. Then for any $\varepsilon \in (0, 1]$, $\Pr[X \geq (1+\varepsilon)\mathbb{E}[X]] \leq e^{-\varepsilon^2/3 \cdot \mathbb{E}[X]}$.

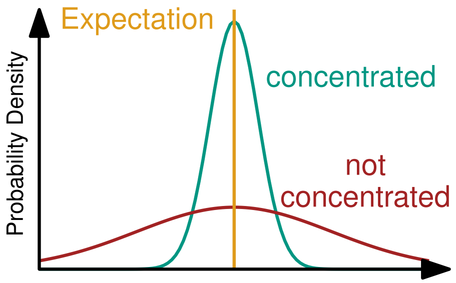
Corollary: Let $X \sim \text{Bin}(n, p)$. Then for any $\varepsilon \in (0, 1)$, $\Pr[X \leq (1-\varepsilon)\mathbb{E}[X]] \leq e^{-\varepsilon^2/2 \cdot \mathbb{E}[X]}$.

- In fact, these also work when the X_i are Bernoulli random variables with different success probabilities

Conclusion

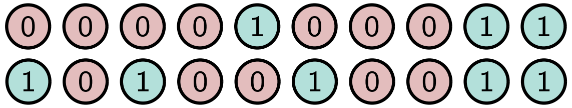
Concentration

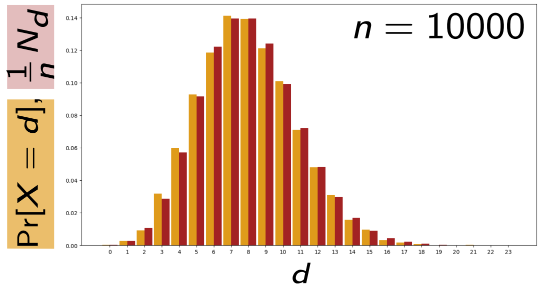
- Is a random variable likely to yield values close to the expectation?
- Concentration inequalities bound the probability for a random variable to deviate from its expectation



Moments

- Used to characterize the shape of a distribution
- First moment: expected value
- Second moment: variance
- Moment generating functions to determine higher-order moments





Concentration Inequalities

- Markov: Based on first moment
- Chebychev: Squaring within Markov (utilizing second moment)
- Chernoff: Exponentiating within Markov (utilizing moment generating functions)
- Examples: Sum of coin flips, empirical degree distribution of ER graphs