CSE 312

Foundations of Computing II

Lecture 21: Chernoff Bound & Union Bound

Review Tail Bounds

Putting a limit on the probability that a random variable is in the "tails" of the distribution (e.g., not near the middle).

Usually statements in the form of

$$P(X \geq a) \leq b$$

or

$$P(|X - \mathbb{E}[X]| \geq a) \leq b$$

Review Markov's and Chebyshev's Inequalities

Theorem (Markov's Inequality). Let X be a random variable taking only non-negative values. Then, for any $t > 0$,

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

Agenda

- Markov's Inequality
- Chebyshev's Inequality
- Chernoff-Hoeffding Bound

Chebyshev's Inequality

Theorem. Let X be a random variable. Then, for any $t > 0$,

$$P(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$

Proof: Define $Z = X - \mathbb{E}[X]$. Then $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[Z^2]$.

$$P(|Z| \geq t) = P(Z^2 \geq t^2) \leq \frac{\mathbb{E}[Z^2]}{t^2} = \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2} = \frac{\mathrm{Var}(X)}{t^2}$$

$|Z| \geq t$ iff $Z^2 \geq t^2$

Markov's inequality ($Z^2 \geq 0$)

Example – Geometric Random Variable

Let X be geometric RV *with* parameter p

$$P(X = i) = (1-p)^{i-1}p \qquad \mathbb{E}[X] = \frac{1}{p} \qquad \mathrm{Var}(X) = \frac{1-p}{p^2}$$

What is the probability that $X \geq 2\mathbb{E}(X) = 2/p$*?*

Markov: $P(X \geq 2\mathbb{E}[X]) \leq \frac{1}{2}$

Chebyshev: $P(X \geq 2\mathbb{E}[X]) \leq P(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} = 1 - p$

Better if $p > 1/2$ ☺

Example

$P(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$.

Suppose that the average number of ads you will see on a website is 25 and the standard deviation of the number of ads is 4. Give an upper bound on the probability of seeing a website with 30 or more ads.

Poll: Where does that upper bound p lie?
pollev.com/rachel312

a. $0 \leq p < 0.25$
b. $0.25 \leq p < 0.5$
c. $0.5 \leq p < 0.75$
d. $0.75 \leq p$
e. Unable to compute

Chebyshev's Inequality – Repeated Experiments

"How many times does Alice need to flip a biased coin <u>until she sees heads n</u> times, if heads occurs with probability p?

X = # of flips until n times "heads"

X_i = # of flips between $(i-1)$-st and i-th "heads"

$$X = \sum_{i=1}^{n} X_i$$

Note: $X_1, \dots, X_n$ are independent and geometric with parameter p

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{p}$$

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{n(1-p)}{p^2}$$

Chebyshev's Inequality – Coin Flips

"How many times does Alice need to flip a biased coin until she sees heads n times, if heads occurs with probability p?

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{p} \qquad \mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{n(1-p)}{p^2}$$

What is the probability that $X \geq 2\mathbb{E}[X] = 2n/p$?

Markov: $P(X \geq 2\mathbb{E}[X]) \leq \frac{1}{2}$

Chebyshev: $P(X \geq 2\mathbb{E}[X]) \leq P(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} = \frac{1-p}{n}$

Goes to zero as $n \to \infty$ ☺

Tail Bounds

Useful for approximations of complex systems. How good the approximation is depends on the actual distribution and the context you are using it in.

- Very often loose upper-bounds are okay when designing for the worst case

Generally (but not always) making more assumptions about your random variable leads to a more accurate upper-bound.

Brain Break

Agenda

- Markov's Inequality
- Chebyshev's Inequality
- Chernoff-Hoeffding Bound

Chebyshev & Binomial Distribution

$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$.

Reformulated: $P(|X - \mu| \geq \delta\mu) \leq \frac{\sigma^2}{\delta^2\mu^2}$ where $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}(X)$

If $X \sim \mathrm{Bin}(n, p)$, then $\mu = np$ and $\sigma^2 = np(1-p)$

$$P(|X - \mu| \geq \delta\mu) \leq \frac{np(1-p)}{\delta^2 n^2 p^2} = \frac{1}{\delta^2} \cdot \frac{1}{n} \cdot \frac{1-p}{p}$$

E.g., $\delta = 0.1, p = 0.5$: $n = 200$: $P(|X - \mu| \geq \delta\mu) \leq 0.5$

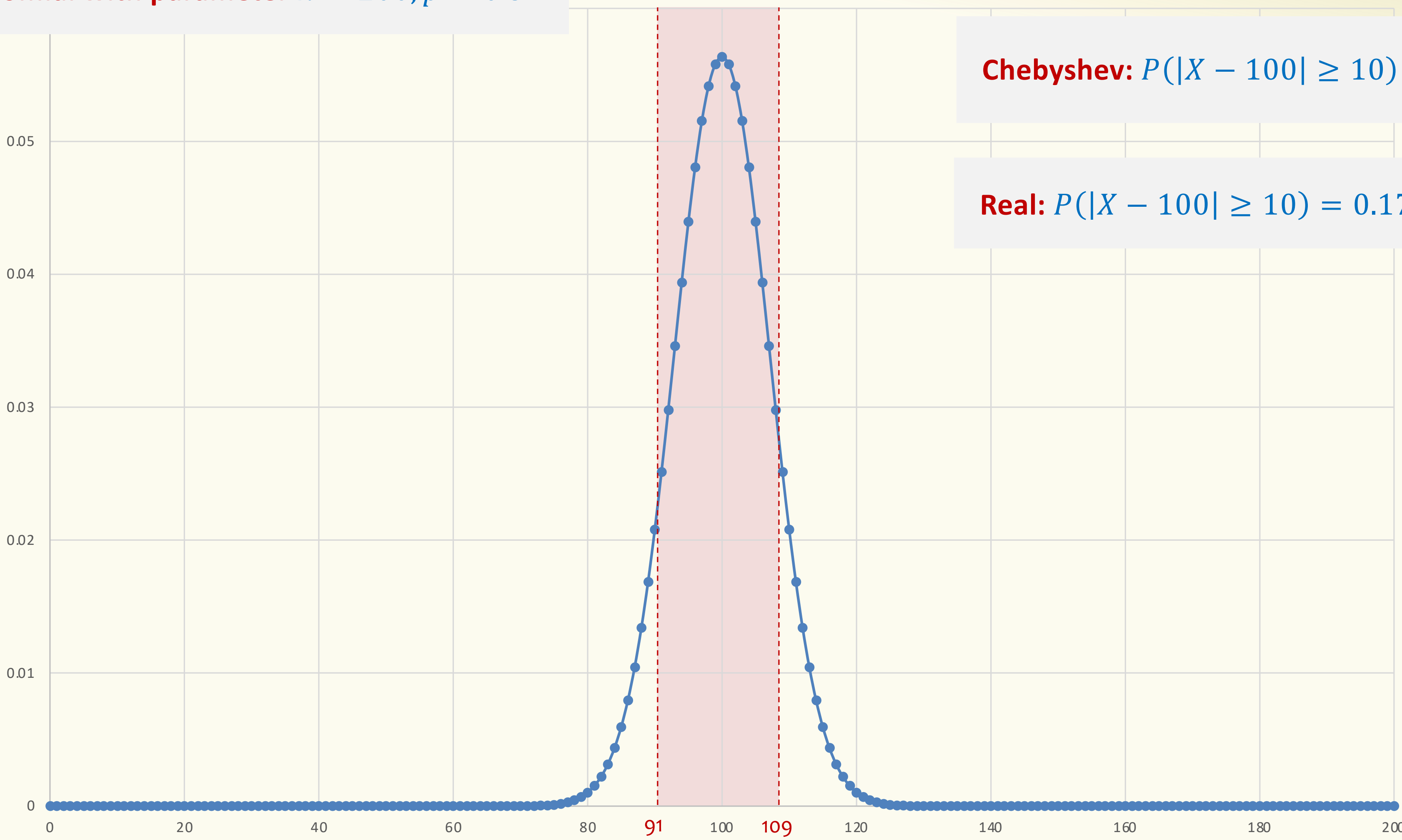
$n = 800$: $P(|X - \mu| \geq \delta\mu) \leq 0.125$

How good is it?

Binomial with parameter $n = 200, p = 0.5$

Chebyshev: $P(|X - 100| \geq 10) \leq \frac{1}{2}$

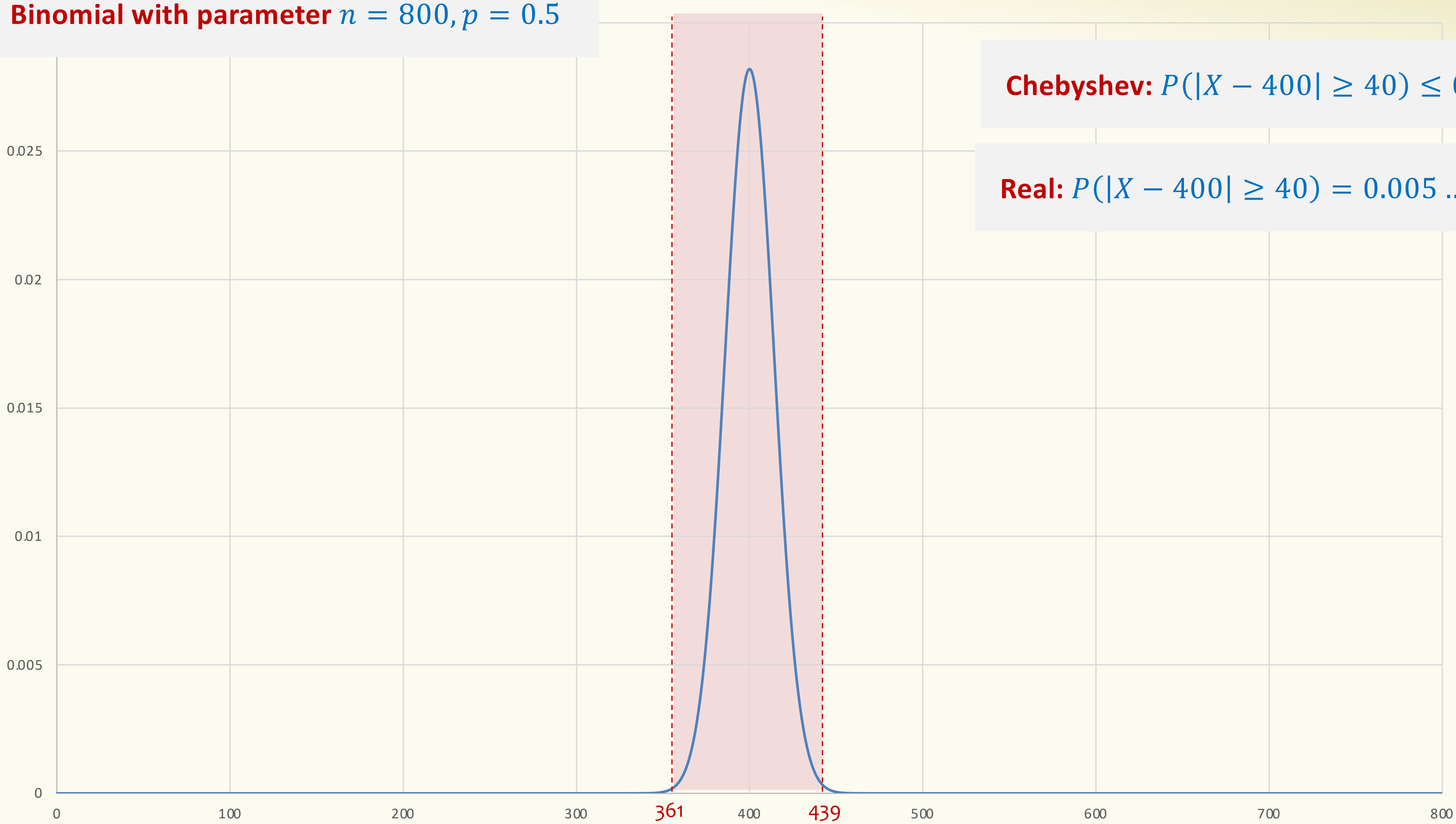
Real: $P(|X - 100| \geq 10) = 0.179 \ldots$



Binomial with parameter $n = 800, p = 0.5$

Chebyshev: $P(|X - 400| \geq 40) \leq 0.125$

Real: $P(|X - 400| \geq 40) = 0.005 \ldots$



Chernoff-Hoeffding Bound

Theorem. Let $X = X_1 + \cdots + X_n$ be a sum of independent RVs, each taking values in $[0,1]$, such that $\mathbb{E}[X] = \mu$. Then, for every $\delta \in [0,1]$,

$$P(|X - \mu| \geq \delta \cdot \mu) \leq e^{-\frac{\delta^2 \mu}{4}}.$$

Herman Chernoff, Herman Rubin, Wassily Hoeffding

Example: If $X \sim \mathrm{Bin}(n, p)$, then $X = X_1 + \cdots + X_n$ is a sum of independent $\{0,1\}$-Bernoulli variables, and $\mu = np$

Note: More accurate versions are possible, but with more cumbersome right-hand side (e.g., see textbook)

Chernoff-Hoeffding Bound – Binomial Distribution

Theorem. (CH bound, binomial case) Let $X \sim \mathrm{Bin}(n, p)$. Let $\mu = np = \mathbb{E}[X]$. Then, for any $\delta \in [0,1]$,

$$P(|X - \mu| \geq \delta \cdot \mu) \leq e^{-\frac{\delta^2 np}{4}}.$$

Example:

$p = 0.5$

$\delta = 0.1$

	Chebyshev	Chernoff
n	$\frac{1}{\delta^2} \cdot \frac{1}{n} \cdot \frac{1-p}{p}$	$e^{-\frac{\delta^2 np}{4}}$
800	0.125	0.3679
2600	0.03846	0.03877
8000	0.0125	0.00005
80000	0.00125	3.72×10^{-44}

Chernoff Bound – Example

$$\mathbb{P}(|X-\mu| \geq \delta \cdot \mu) \leq e^{-\frac{\delta^2 np}{4}}.$$

Alice tosses a fair coin n times, what is an upper bound for the probability that she sees heads at least $0.75 \times n$ times?

Poll: pollev.com/rachel312

a. $e^{-n/64}$
b. $e^{-n/32}$
c. $e^{-n/16}$
d. $e^{-n/8}$

Chernoff vs Chebyshev – Summary

$$\frac{1}{\delta^2} \cdot \frac{1}{n} \cdot \frac{1-p}{p}$$

Chebyshev, <u>linear</u> decrease in n

vs

Chernoff, <u>exponential</u> decrease in n

$$e^{-\frac{\delta^2 np}{4}}$$

Why is the Chernoff Bound True?

Proof strategy (upper tail): For any $t > 0$:

- $P(X \geq (1+\delta) \cdot \mu) = P\left(e^{tX} \geq e^{t(1+\delta)\cdot\mu}\right)$
- Then, apply Markov + independence:

$$P\left(e^{tX} \geq e^{t(1+\delta)\cdot\mu}\right) \leq \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mu}} = \frac{\mathbb{E}[e^{tX_1}] \cdots \mathbb{E}[e^{tX_n}]}{e^{t(1+\delta)\mu}}$$

- Find t minimizing the right-hand-side.

Agenda

- Chernoff Bound
  - Example: Server Load
  - The Union Bound
- Probability vs statistics
  - Estimation

Chernoff-Hoeffding Bound

Theorem. Let $X = X_1 + \cdots + X_n$ be a sum of independent RVs, each taking values in $[0,1]$, such that $\mathbb{E}[X] = \mu$. Then…

for every $\delta \in [0,1]$, $P(|X - \mu| \geq \delta \cdot \mu) \leq e^{-\frac{\delta^2 \mu}{4}}$ both tails

for every $\delta \geq 0$, $P(X - \mu \geq \delta \cdot \mu) \leq e^{-\frac{\delta^2 \mu}{4}}$ right/upper tail

Herman Chernoff, Herman Rubin, Wassily Hoeffding

Example: If $X \sim \mathrm{Bin}(n, p)$, then $X = X_1 + \cdots + X_n$ is a sum of independent $\{0,1\}$-Bernoulli variables, and $\mu = np$

Note: More accurate versions are possible, but with more cumbersome right-hand side (e.g., see textbook)

Chernoff vs Chebyshev – Summary

$$\frac{1}{\delta^2} \cdot \frac{1}{n} \cdot \frac{1-p}{p}$$

Chebyshev, <u>linear</u> decrease in n

vs

Chernoff, <u>exponential</u> decrease in n

$$e^{-\frac{\delta^2 np}{4}}$$

Agenda

- Chernoff Bound
  - Example: Server Load
  - The Union Bound
- Probability vs statistics
  - Estimation

Application – Distributed Load Balancing

We have k processors, and $n \gg k$ jobs.
We want to distribute jobs evenly across processors.

Strategy: Each job assigned to a randomly chosen processor!

X_i = load of processor i $\quad X_i \sim \text{Binomial}(n, 1/k) \quad \mathbb{E}[X_i] = n/k$

$X = \max\{X_1, \dots, X_k\}$ = max load of a processor

Question: How close is X to n/k?

Distributed Load Balancing

Claim. (Load of single server)

$$P\left(X_i > \frac{n}{k} + 4\sqrt{\frac{n \ln k}{k}}\right) \le 1/k^4.$$

Example:

- $n = 10^6 \gg k = 1000$
- Perfect load balancing would give load $\frac{n}{k} = 1000$ per server
- $\frac{n}{k} + 4\sqrt{n \ln k \,/k} \approx 1332$
- *"The probability that server i processes more than 1332 jobs is at most 1-over-one-trillion!"*

Distributed Load Balancing

Claim. (Load of single server)

$$P\left(X_i > \frac{n}{k} + 4\sqrt{\frac{n \ln k}{k}}\right) = P\left(X_i > \frac{n}{k}\left(1 + 4\sqrt{\frac{k \ln k}{n}}\right)\right) \le 1/k^4.$$

Proof. Set $\mu = \mathbb{E}[X_i] = \frac{n}{k}$ and $\delta = 4\sqrt{\frac{k}{n}\ln k}$

$$P\left(X_i > \mu\left(1 + 4\sqrt{\frac{k \ln k}{n}}\right)\right) = P(X_i > \mu(1+\delta))$$

$$= P(X_i - \mu > \delta\mu)$$ Upper tail

$$\le e^{-\frac{\delta^2 \mu}{4}} = e^{-4 \ln k} = \frac{1}{k^4}$$

$\delta^2 = 4^2 \cdot \frac{k \ln k}{n}$

so $\delta^2 \mu = 4^2 \ln k$

What about the maximum load?

Claim. (Load of single server)

$$P\left(X_i > \frac{n}{k} + 4\sqrt{\frac{n \ln k}{k}}\right) \leq 1/k^4.$$

What about $X = \max\{X_1, \dots, X_k\}$?

Note: $X_1, \dots, X_k$ are not (mutually) independent!

In particular: $X_1 + \cdots + X_k = n$

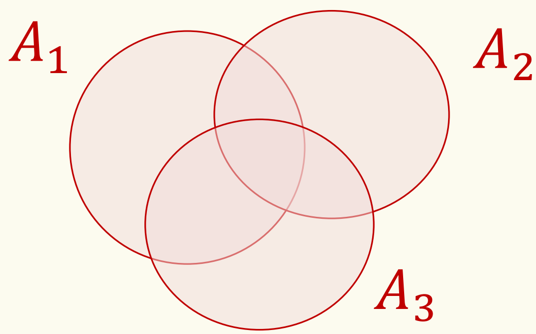
When non-trivial outcome of one RV can be derived from other RVs, they are non-independent.

Detour – Union Bound – A nice name for something you already know

Theorem (Union Bound). Let $A_1, \dots, A_n$ be arbitrary events. Then,

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

Intuition (3 evts.):



Detour – Union Bound - Example

Suppose we have $N = 200$ computers, where each one fails with probability 0.001.

What is the probability that at least one server fails?

Let A_i be the event that server i fails.

Then event that at least one server fails is $\bigcup_{i=1}^{n} A_i$

$$P\left(\bigcup_{i=1}^{N} A_i\right) \leq \sum_{i=1}^{N} P(A_i) = 0.001N = 0.2$$

What about the maximum load?

Claim. (Load of single server)

$$P\left(X_i > \frac{n}{k} + 4\sqrt{\frac{n \ln k}{k}}\right) \leq 1/k^4.$$

What about $X = \max\{X_1, \dots, X_k\}$?

$$P\left(X > \frac{n}{k} + 4\sqrt{n \ln k \,/k}\right) = P\left(\left\{X_1 > \frac{n}{k} + 4\sqrt{n \ln k \,/k}\right\} \cup \dots \cup \left\{X_k > \frac{n}{k} + 4\sqrt{n \ln k \,/k}\right\}\right)$$

$$\leq P\left(X_1 > \frac{n}{k} + 4\sqrt{\frac{n \ln k}{k}}\right) + \dots + P\left(X_k > \frac{n}{k} + 4\sqrt{n \ln k \,/k}\right)$$

Union bound

$$\leq \frac{1}{k^4} + \dots + \frac{1}{k^4} = k \times \frac{1}{k^4} = \frac{1}{k^3}$$

What about the maximum load?

Claim. (Load of single server)

$$P\left(X_i > \frac{n}{k} + 4\sqrt{\frac{n \ln k}{k}}\right) \le 1/k^4.$$

Claim. (Max load) Let $X = \max\{X_1, \dots, X_k\}$.

$$P\left(X > \frac{n}{k} + 4\sqrt{\frac{n \ln k}{k}}\right) \le 1/k^3.$$

Example:

- $n = 10^6 \gg k = 1000$
- $\frac{n}{k} + 4\sqrt{n \ln k / k} \approx 1332$
- *"The probability that some server processes more than* 1332 *jobs is at most 1-over-one-billion!"*

Using tail bounds

- Tail bounds are *guarantees*, unlike our use of CLT
- Often, we actually start with a target upper bound on failure probability
  – In the load-balancing example, the value of δ in terms of n and k was worked out in order to get failure probability $\leq 1/k^4$
    - We didn't start out with this weird value
  – See example in section and on homework
- We use these bounds to design (randomized) algorithms or analyze their guaranteed level of success.