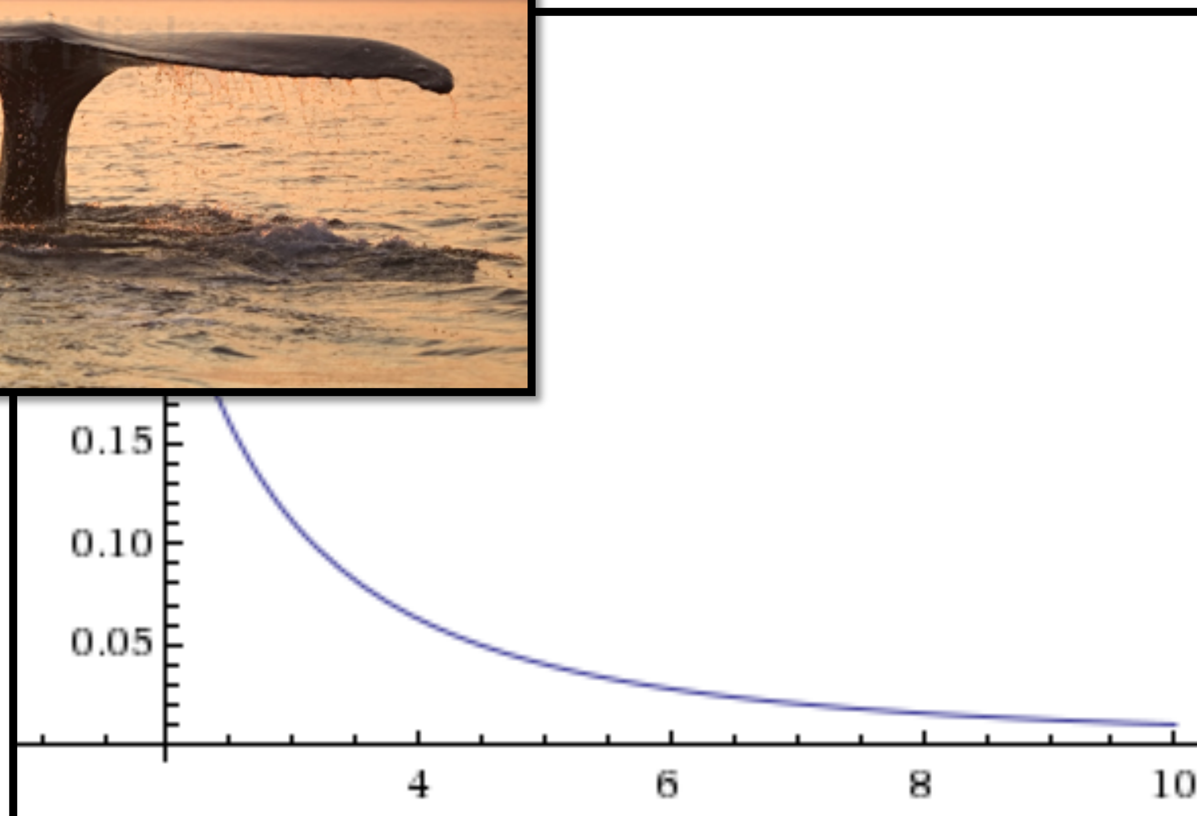
tail bounds



Often, we want to bound the probability that a random variable X is far from its expectation.

A random variable X has mean μ

Can we bound

$$Pr(X \geq 100\mu)$$

$$Pr(X \geq 1,000\mu)$$

$$Pr(X \geq 1,000,000\mu)$$

Not without additional information…

We know that randomized quicksort runs in O(n log n) *expected* time. But what's the probability that it takes more than 10 n log(n) steps? More than $n^{1.5}$ steps?

If we know the expected advertising cost is $1500/day, what's the probability we go over budget? By a factor of 4?

We only expect 10,000 homeowners to default on their mortgages. What's the probability that 1,000,000 homeowners default?

"Lake Wobegon, Minnesota, where
all the women are strong,
all the men are good looking,
and
all the children are above average…"

An *arbitrary* random variable could have very bad behavior. But knowledge is power; if we know *something*, can we bound the badness?

Suppose we know that X is always non-negative.

Theorem: If X is a non-negative random variable, then for every $\alpha > 0$, we have

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$

Corr:

$$P(X \geq \alpha E[X]) \leq 1/\alpha$$

Theorem: If X is a non-negative random variable, then for every $\alpha > 0$, we have

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$

Example: if X = time to quicksort *n* items, expectation E[X] ≈ 1.4 *n* log *n*. What's probability that it takes > 4 times as long as expected?

By Markov's inequality:

$$P(X \geq 4 \cdot E[X]) \leq E[X]/(4\, E[X]) = 1/4$$

Theorem: If X is a non-negative random variable, then for every $\alpha > 0$, we have

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$

Proof:

$$E[X] = \sum_x xP(x)$$

$$= \sum_{x<\alpha} xP(x) + \sum_{x\geq\alpha} xP(x)$$

$$\geq \quad 0 \quad + \sum_{x\geq\alpha} \alpha P(x) \qquad \underline{(x \geq 0;\ \alpha \leq x)}$$

$$= \alpha P(X \geq \alpha)$$

Theorem: If X is  tive random variable, then fo have

Proof:

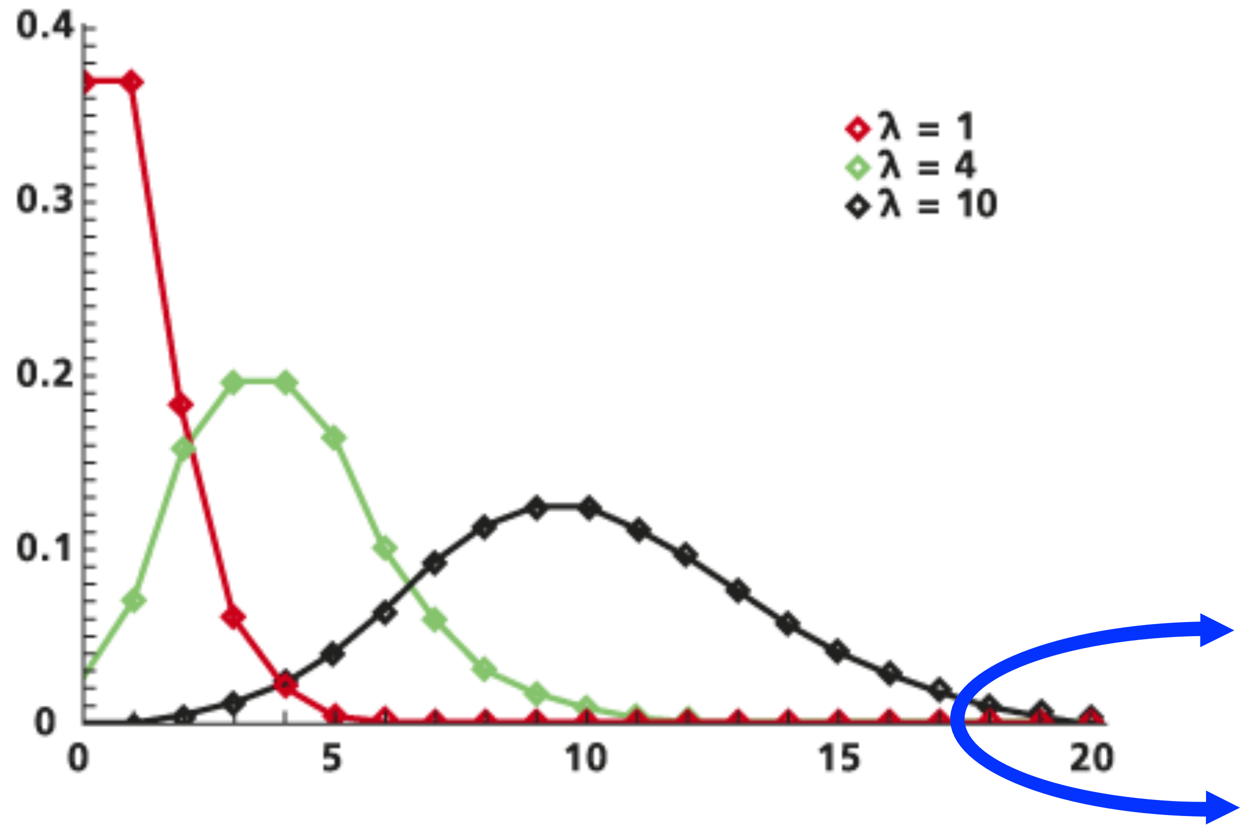
E[X]

=)

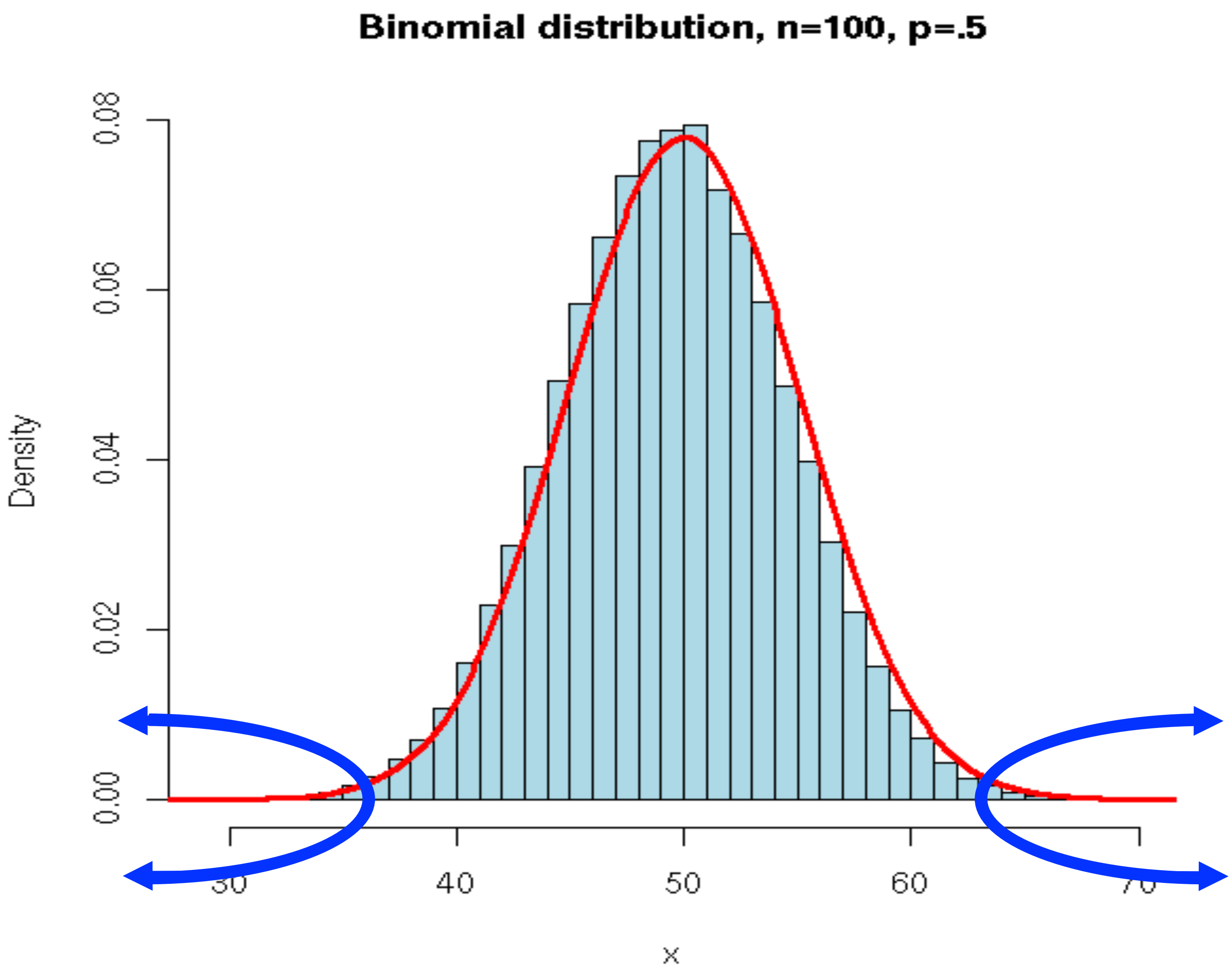
≥ P(x)   (x ≥ 0; α ≤ x)

= α P(

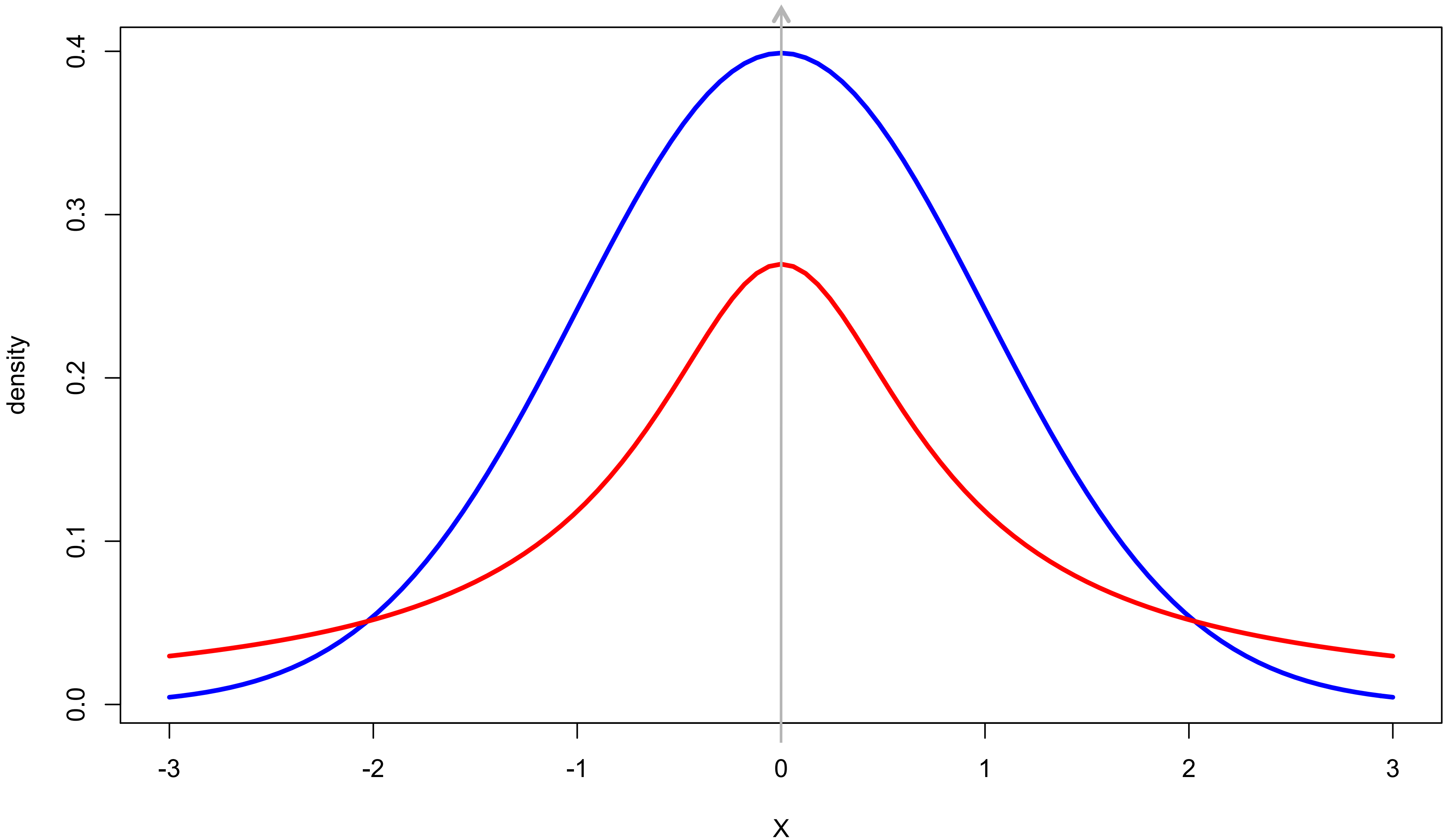
For a random variable X, the *tails* of X are the parts of the PMF that are "far" from its mean.



binomial tails



heavy-tailed distribution



If we know *more* about a random variable, we can often use that to get *better* tail bounds.

Suppose we *also* know the variance.

Theorem: If Y is an arbitrary random variable with E[Y] = μ, then, for any $\alpha > 0$,

$$P(|Y - \mu| \geq \alpha) \leq \frac{\mathrm{Var}[Y]}{\alpha^2}$$

Theorem: If Y is an arbitrary random variable with $\mu = E[Y]$, then, for any $\alpha > 0$,

$$P(|Y - \mu| \geq \alpha) \leq \frac{\mathrm{Var}[Y]}{\alpha^2}$$

Proof: $\text{Let } X = (Y - \mu)^2$

X is non-negative, so we can apply Markov's inequality:

$$P(|Y - \mu| \geq \alpha) = P(X \geq \alpha^2)$$

$$\leq \frac{E[X]}{\alpha^2} = \frac{\mathrm{Var}[Y]}{\alpha^2}$$

Theorem: If Y is an arbitrary random variable with … y α > 0,

$$P(|Y \quad \ldots \quad \frac{\mathrm{Var}[Y]}{\alpha^2}$$

Proof: Let …

X is non-nega… Markov's inequality:

$$P(|Y - \quad \ldots \quad \geq \alpha^2)$$

$$\leq \quad \ldots \quad = \frac{\mathrm{Var}[Y]}{\alpha^2}$$

Theorem: If Y is an arbitrary random variable with μ = E[Y], then, for any α > 0,

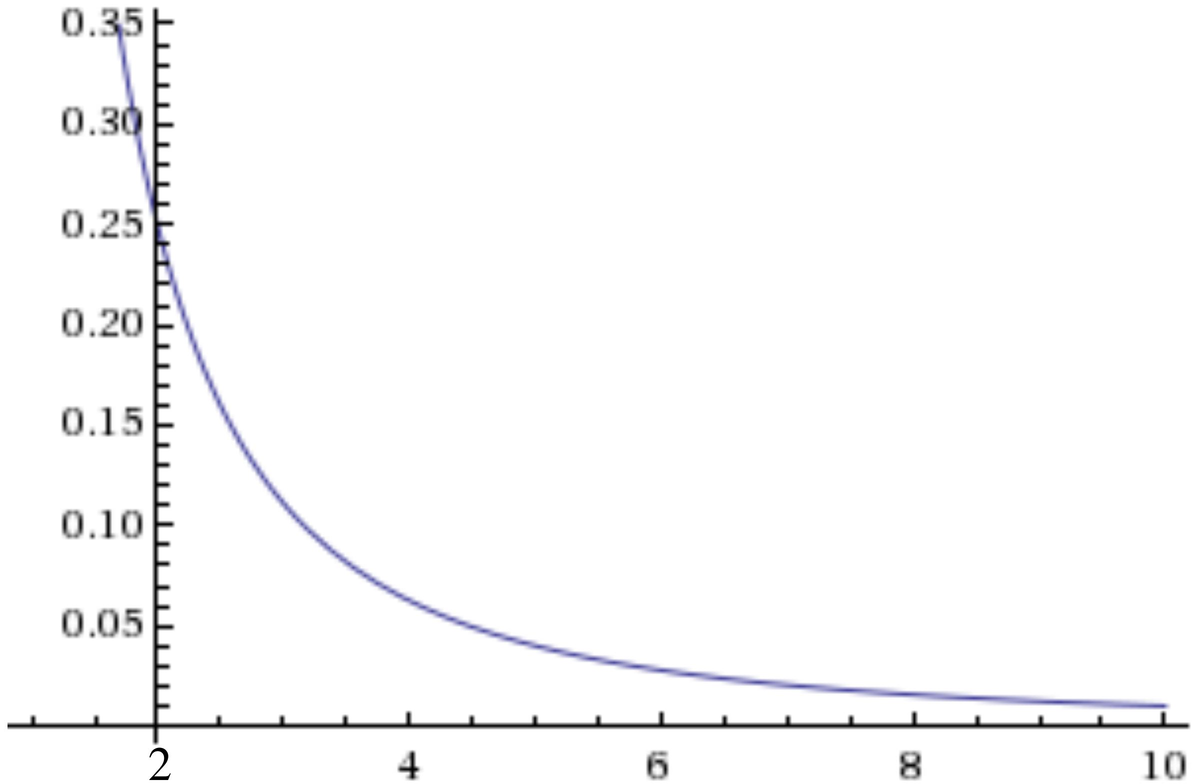
$$P(|Y - \mu| \geq \alpha) \leq \frac{\mathrm{Var}[Y]}{\alpha^2}$$

Corr: If

$$\sigma = SD[Y] = \sqrt{\mathrm{Var}[Y]}$$

Then:

$$P(|Y - \mu| \geq t\sigma) \leq \frac{\sigma^2}{t^2\sigma^2} = \frac{1}{t^2}$$



$$P(|Y - \mu| \geq t\sigma) \leq \frac{1}{t^2}$$

Chebyshev's inequality

$$P(|Y - \mu| \geq \alpha) \leq \frac{\mathrm{Var}[Y]}{\alpha^2}$$

Y = comparisons in quicksort for n=1024

$E[Y] = 1.4\, n \log_2 n \approx 14000$

$\mathrm{Var}[Y] = ((21-2\pi^2)/3)*n^2 \approx 441000$

(i.e. $SD[Y] \approx 664$)

$P(Y \geq 4\mu) = P(Y-\mu \geq 3\mu) \leq \mathrm{Var}(Y)/(9\mu^2) < .000242$

1000 times smaller than Markov
but still overestimated?: $\sigma/\mu \approx 0.05$, so $4\mu \approx \mu + 60\sigma$

X Binomial (n,1/2)

$$Pr(X \geq 3/4n)$$

Markov: $2/3$

Chebyshev: 2/n (because Chebyshev is 2-sided)

If n= 1000, Probability > 750 H's at most 0.002

Truth:

$$0.000000000000000000000000000000000000000000000000000000000067$$

What fraction of people approve of president?

Poll: call up n random people.

$$X = X_1 + X_2 + \ldots X_n$$

Define average X/n as our estimate.

What should n be? How good an estimate?
How confident are we?

Suppose X ~ Bin(n,p)
μ = E[X] = pn

Chernoff bound:

For any δ with $0 < \delta < 1$,

$$Pr(X > (1+\delta)\mu) \leq e^{-\frac{\mu\delta^2}{3}}$$

$$Pr(X < (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$$

Suppose X is the sum of independent random variables $X = X_1 + X_2 + \ldots + X_n$ where X_i in [0,1]
$\mu = E[X] = pn$

Another Chernoff bound:

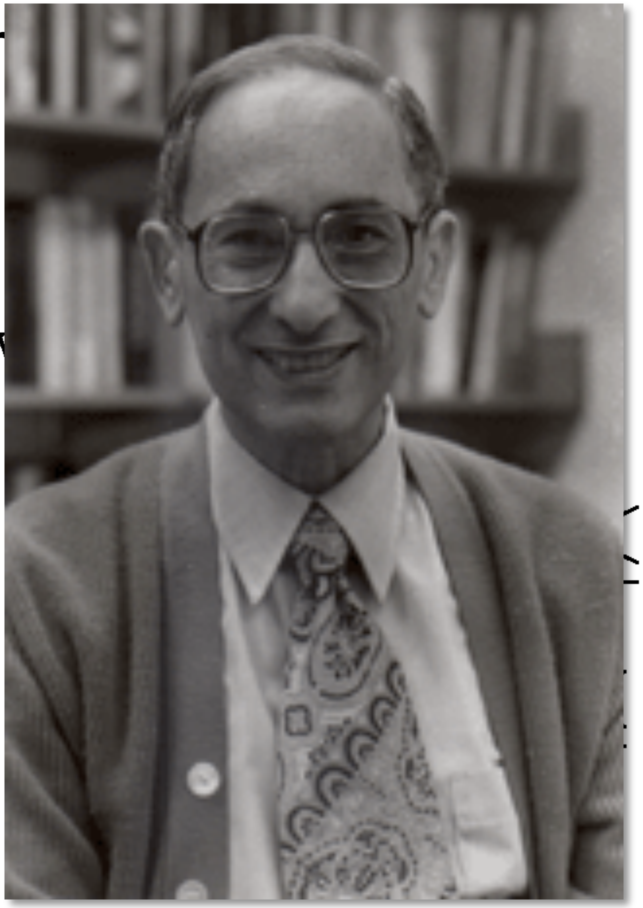
For any $\epsilon \geq 0$

$$Pr(|X - pn| \geq \epsilon pn) \leq 2e^{-\left(\frac{\epsilon^2}{2+\epsilon}\right)\cdot pn}$$

Many variants

Suppose X ~ Bin(n,p)

μ = E[X] = pr



Chernoff bou

For any δ w

$$P(X \quad \leq e^{-\frac{\delta^2 \mu}{2}}$$

$$P(X \quad \leq e^{-\frac{\delta^2 \mu}{3}}$$

router buffers

COURTESY: CISCO

Model: 100,000 computers independently of each other send a packet with probability q = 0.01 each second. The router processes its buffer every second. How many packet buffers so that router drops a packet:

- Never?
  100,000
- With probability at most 10^{-6}, every hour?
  1210
- With probability at most 10^{-6}, every year?
  1250
- With probability at most 10^{-6}, since Big Bang?
  1331

(these numbers may be slightly off.)

Tail bounds – bound probabilities of extreme events

Three (of many):

Markov: $P(X \geq k\mu) \leq 1/k$ (weak, but general; only need $X \geq 0$ and μ)

Chebyshev: $P(|X-\mu| \geq k\sigma) \leq 1/k^2$ (often stronger, but also need σ)

Chernoff: various forms, depending on underlying distribution; usually 1/exponential, *vs* 1/polynomial above

Generally, more assumptions/knowledge ⇒ better bounds

"Better" than exact distribution?

Maybe, e.g. if latter is unknown or mathematically messy

"Better" than, e.g., "Poisson approx to Binomial"?

Maybe, e.g. if you need rigorously "≤" rather than just "≈"