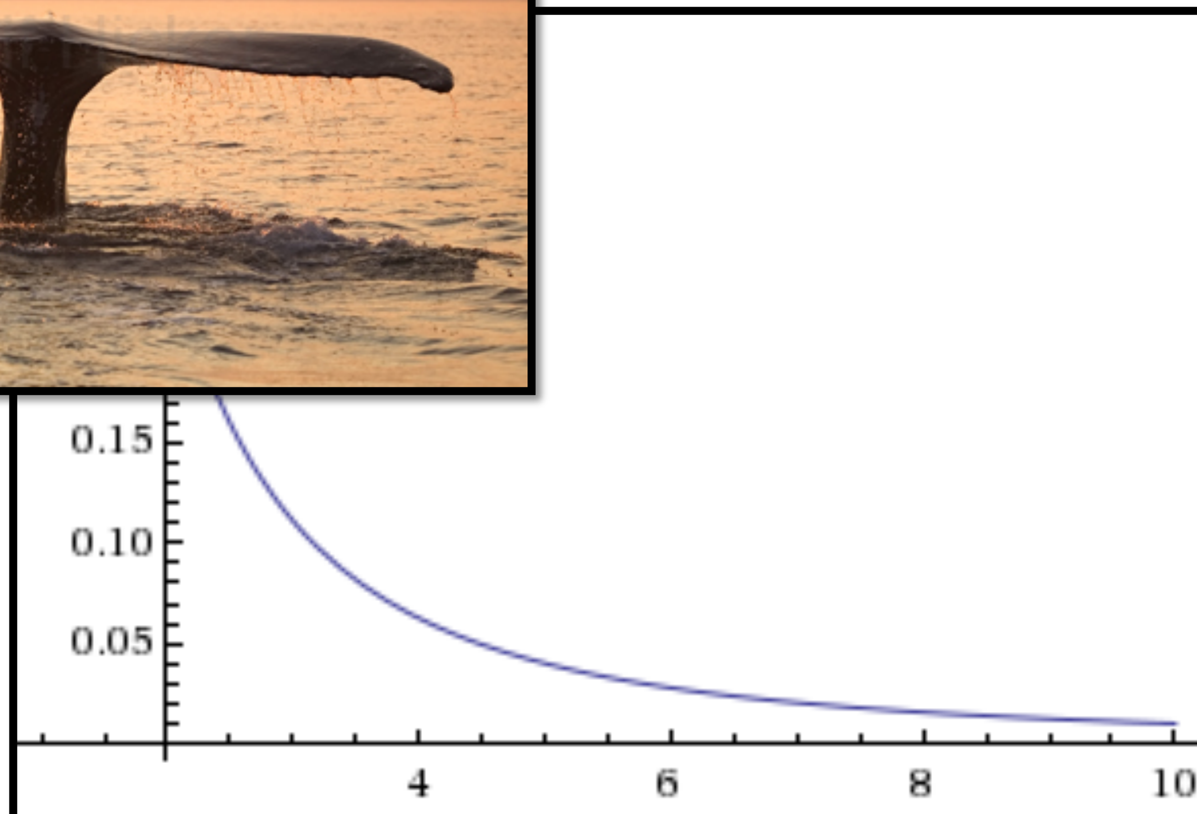
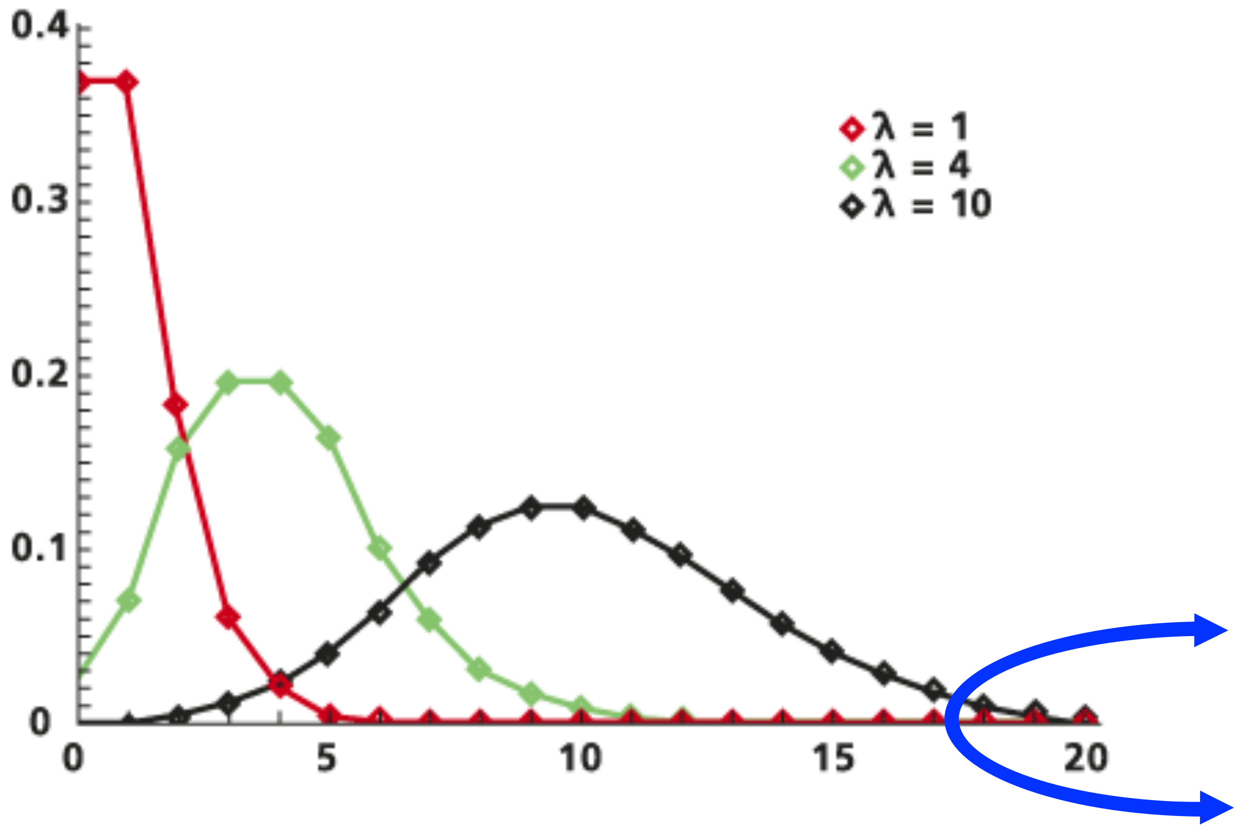
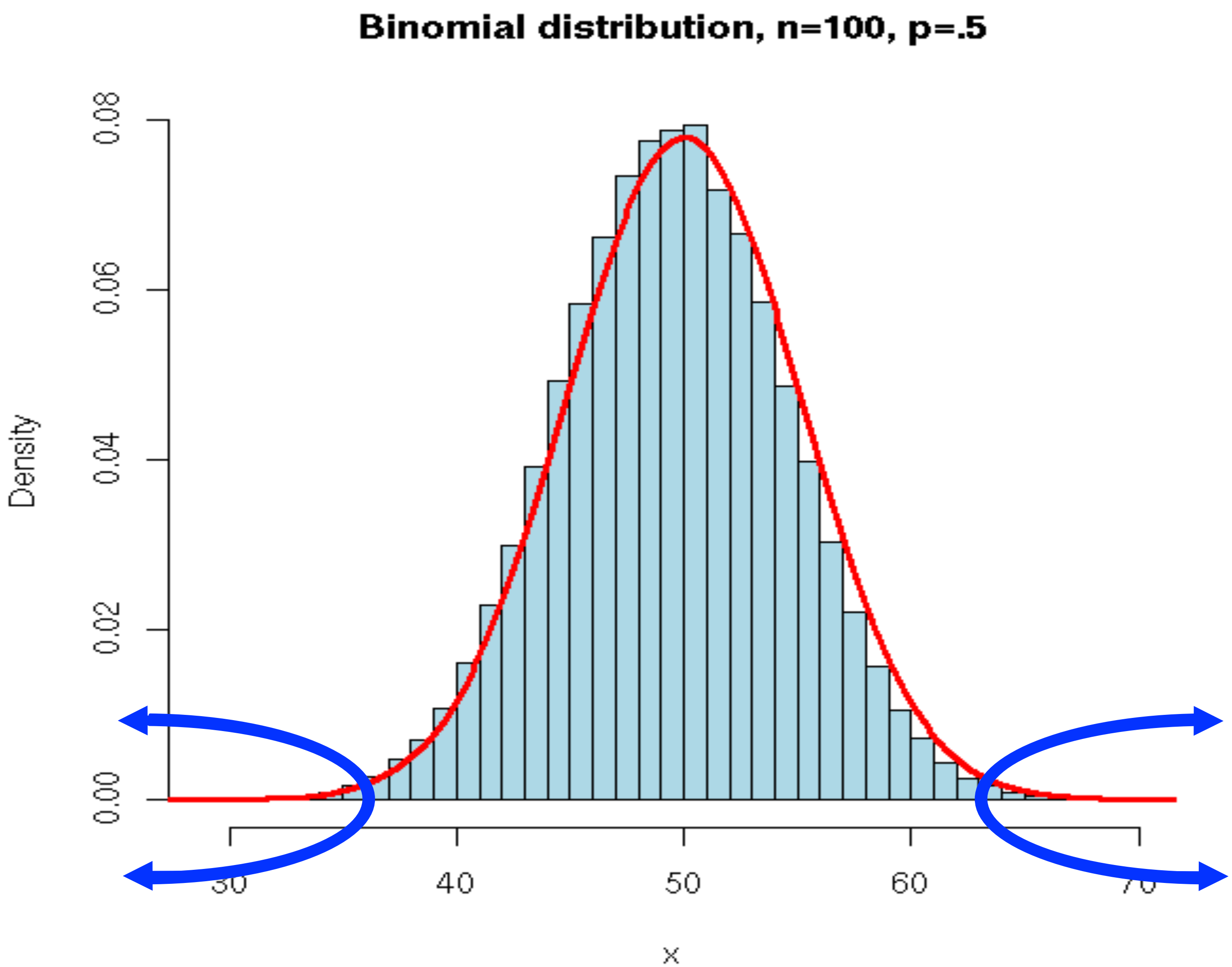
tail bounds



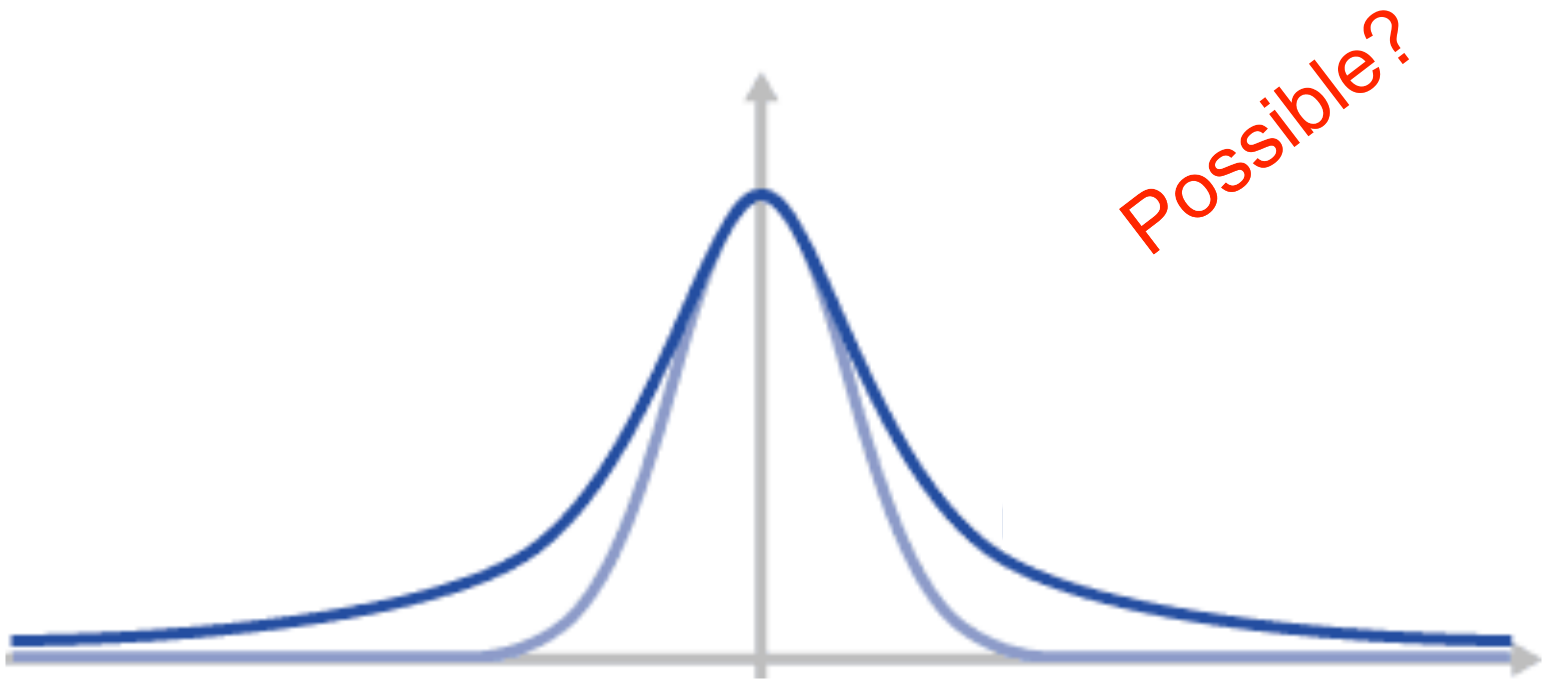
For a random variable X, the *tails* of X are the parts of the PMF that are "far" from its mean.



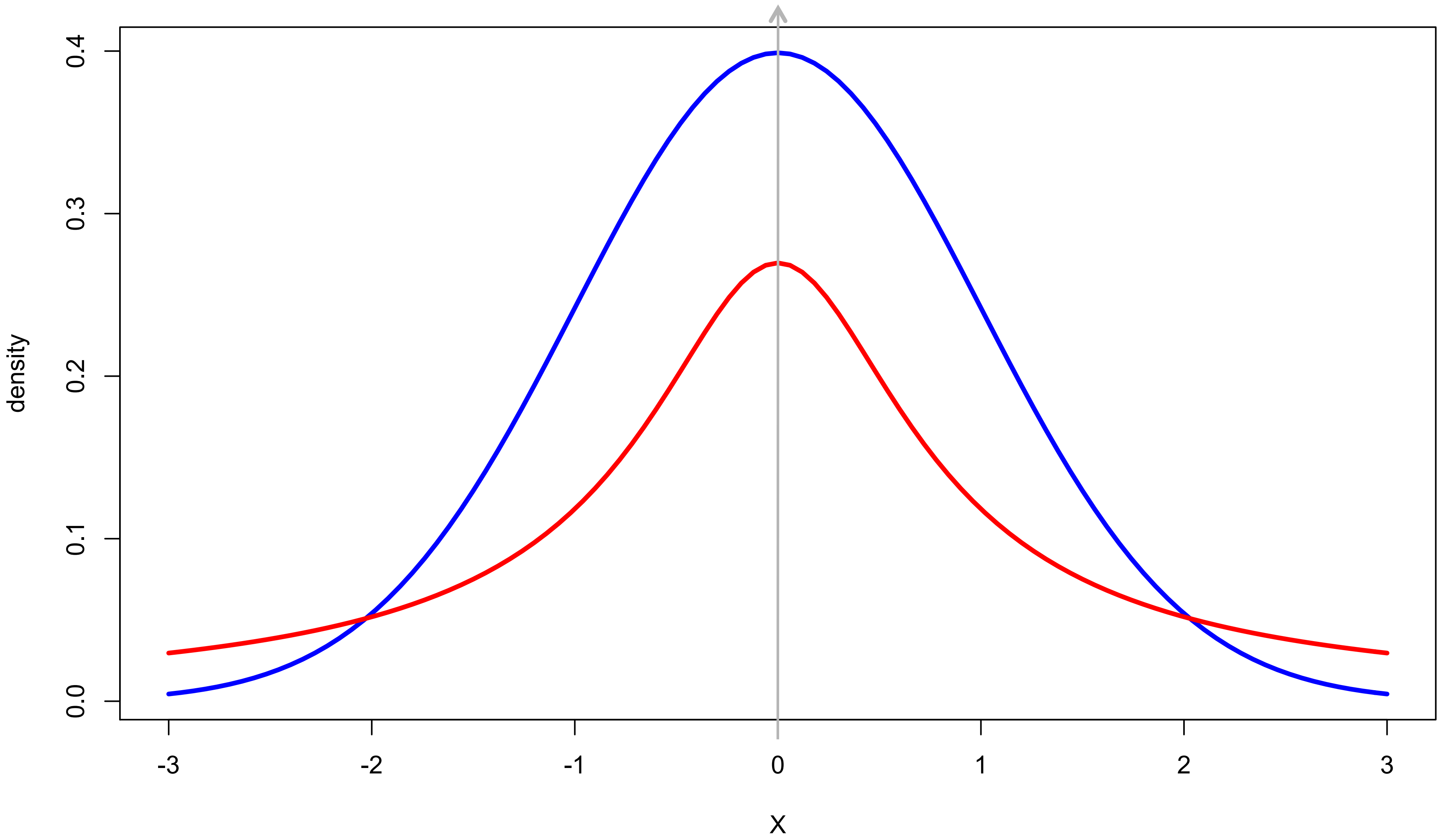
binomial tails



heavy-tailed distribution



heavy-tailed distribution



Often, we want to bound the probability that a random variable X is "large." Perhaps:

$$P(X > \alpha) < \frac{1}{\alpha^3}$$

$$P(X > E[X] + t) < e^{-t}$$

$$P(|X - E[X]| > t) < \frac{1}{\sqrt{t}}$$

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?

I 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…”

In general, 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:

$$E[X] = \Sigma_x \, xP(x)$$

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

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

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

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

Proof:

E[X] …

= …

≥ …) (x ≥ 0; α ≤ x)

= αP(X…

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 μ = E[Y], then, for any $\alpha > 0$,

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

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

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 $\alpha > 0$,

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

Proof: Let …

X is non-neg… Markov's inequality:

$P(|Y -$ … $\geq \alpha^2)$

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

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?: $\sigma/\mu \approx 5\%$, so $4\mu \approx \mu+60\sigma$

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

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

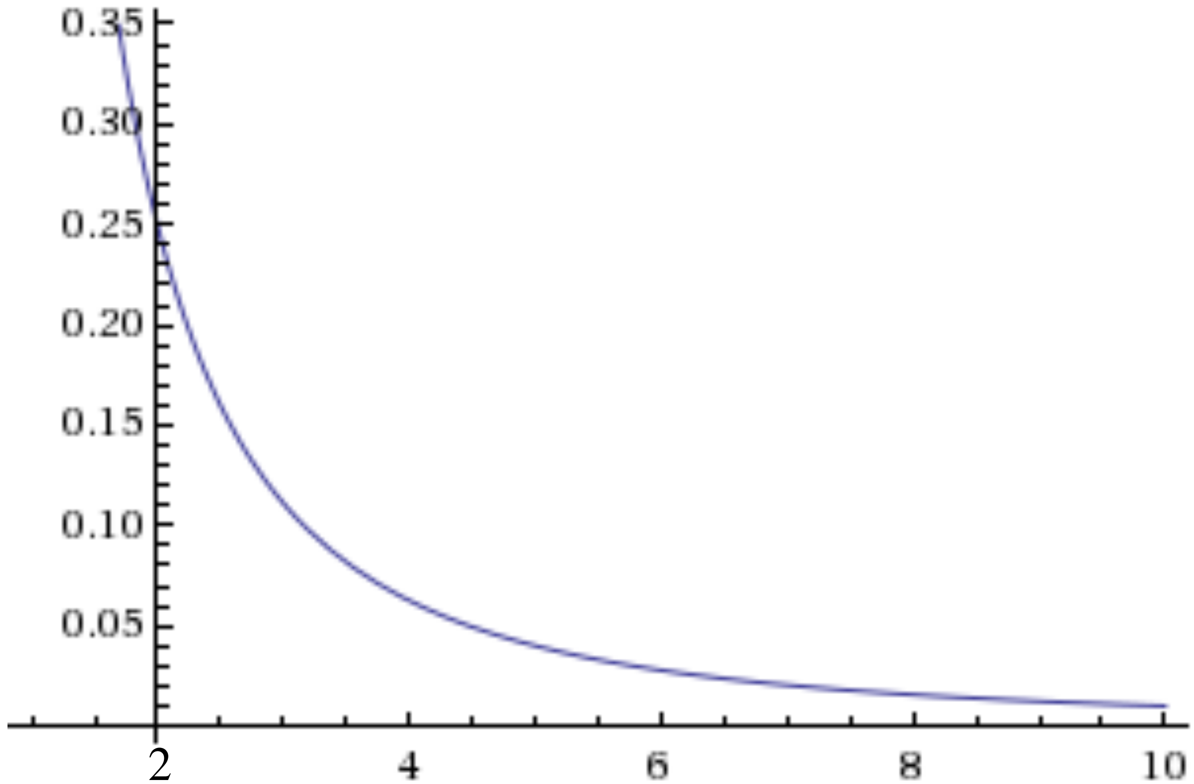
Corr: If

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

Then:

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

Chebyshev's inequality



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

$Y \sim \text{Bin}(15000, 0.1)$

$\mu = E[Y] = 1500, \sigma = \sqrt{\text{Var}(Y)} = 36.7$

$P(Y \geq 6000) = P(Y \geq 4\mu) \leq 1/4$ (Markov)

$P(Y \geq 6000) = P(Y\text{-}\mu \geq 122\sigma) \leq 7\text{x}10^{-5}$ (Chebyshev)

Poisson approximation: $Y \sim \text{Poi}(1500)$

Rough computer calculation:

$$P(Y > 6000) \ll 10^{-1600}$$

Suppose X ~ Bin(n,p)
μ = E[X] = pn

Chernoff bound:

For any δ with $0 < \delta < 1$,

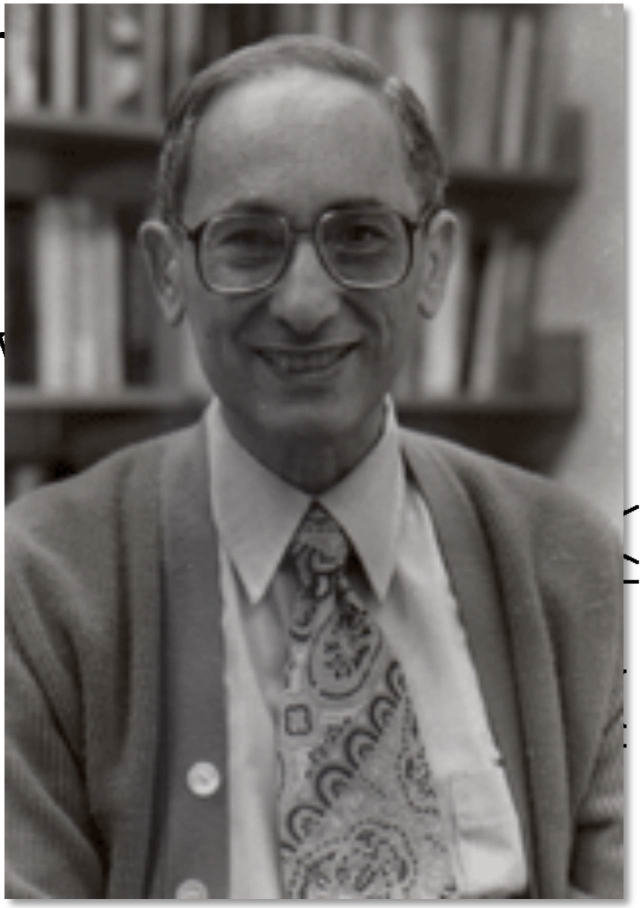
$$P(X > (1+\delta)\mu) \le e^{-\frac{\delta^2 \mu}{2}}$$
$$P(X < (1-\delta)\mu) \le e^{-\frac{\delta^2 \mu}{3}}$$

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

Model: 100,000 computers each independently send a packet with probability p = 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

$X \sim \mathrm{Bin}(100{,}000, 0.01)$, $\mu = E[X] = 1000$

Let p = probability of buffer overflow in 1 second

By the Chernoff bound

$$p = P(X > (1+\delta)\mu) \leq e^{-\frac{\delta^2 \mu}{2}}$$

Overflow probability in n seconds

$= 1-(1-p)^n \leq np \leq n \exp(-\delta^2\mu/2)$,

which is $\leq \varepsilon$ provided $\delta \geq \sqrt{(2/\mu)\ln(n/\varepsilon)}$.

For $\varepsilon=10^{-6}$ per hour: $\delta \approx .210$, buffers = 1210

For $\varepsilon=10^{-6}$ per year: $\delta \approx .250$, buffers = 1250

For $\varepsilon=10^{-6}$ per 15BY: $\delta \approx .331$, buffers = 1331

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 $\Rightarrow$ better bounds

"Better" than exact distribution?

- Maybe, e.g. if later is unknown or mathematically messy

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

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