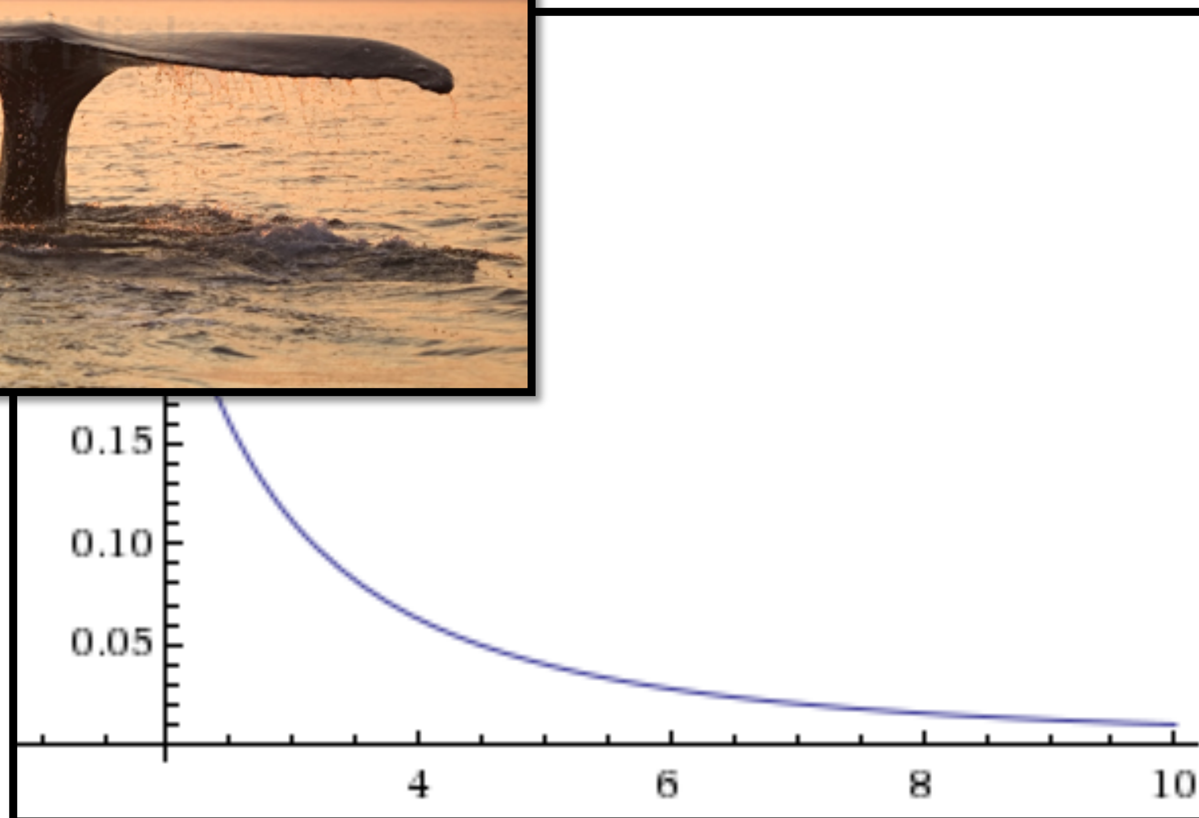
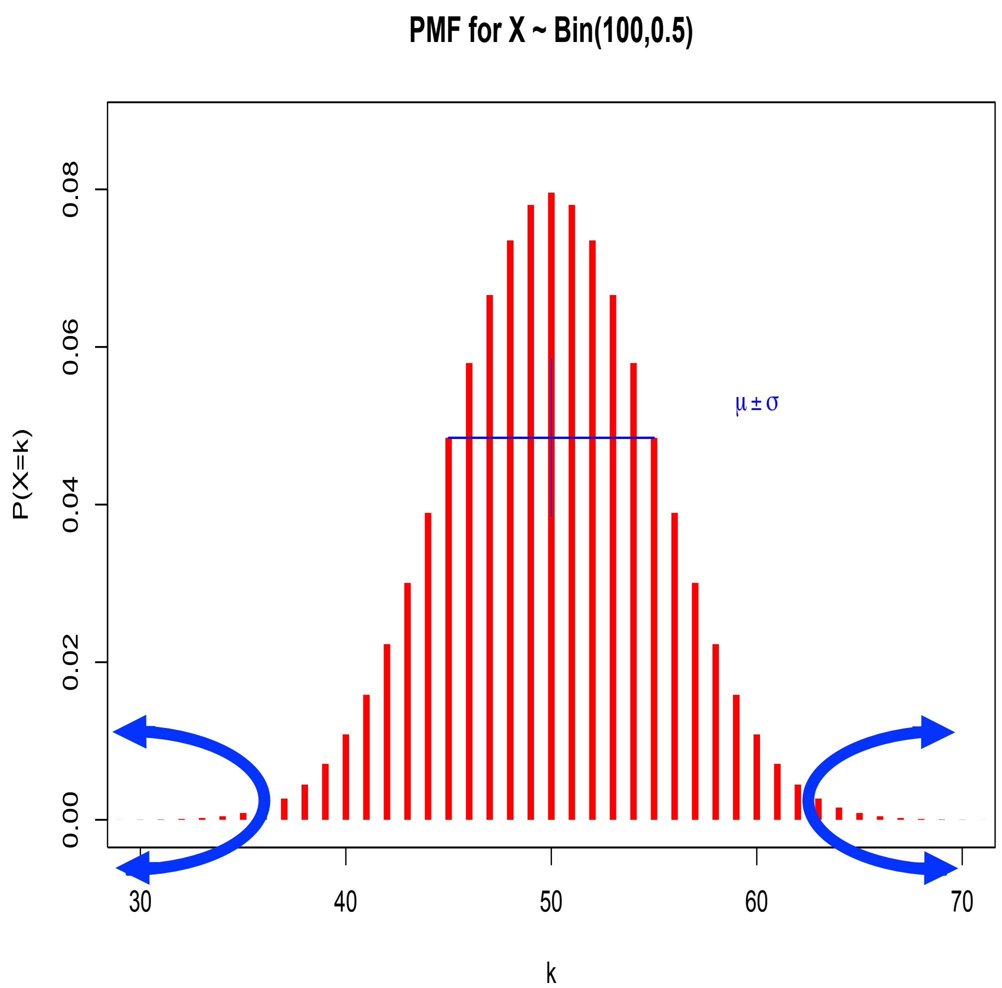
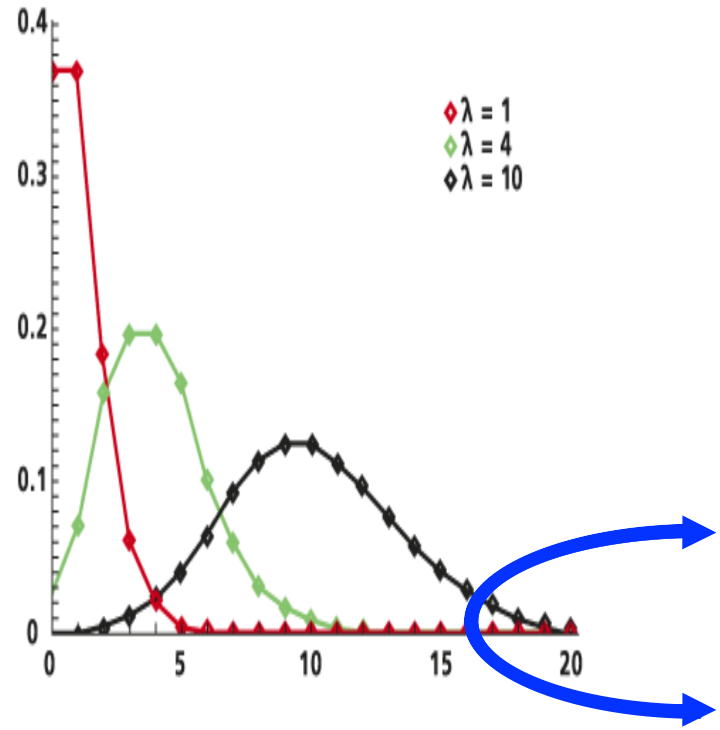
tail bounds



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





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

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

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

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

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?

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?

"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 = daily advertising expenses and

E[X] = 1500

Then, by Markov's inequality,

$$P(X \geq 6000) \leq \frac{1500}{6000} = 0.25$$

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:

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

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

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

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

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 $\alpha > 0$,

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

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

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

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

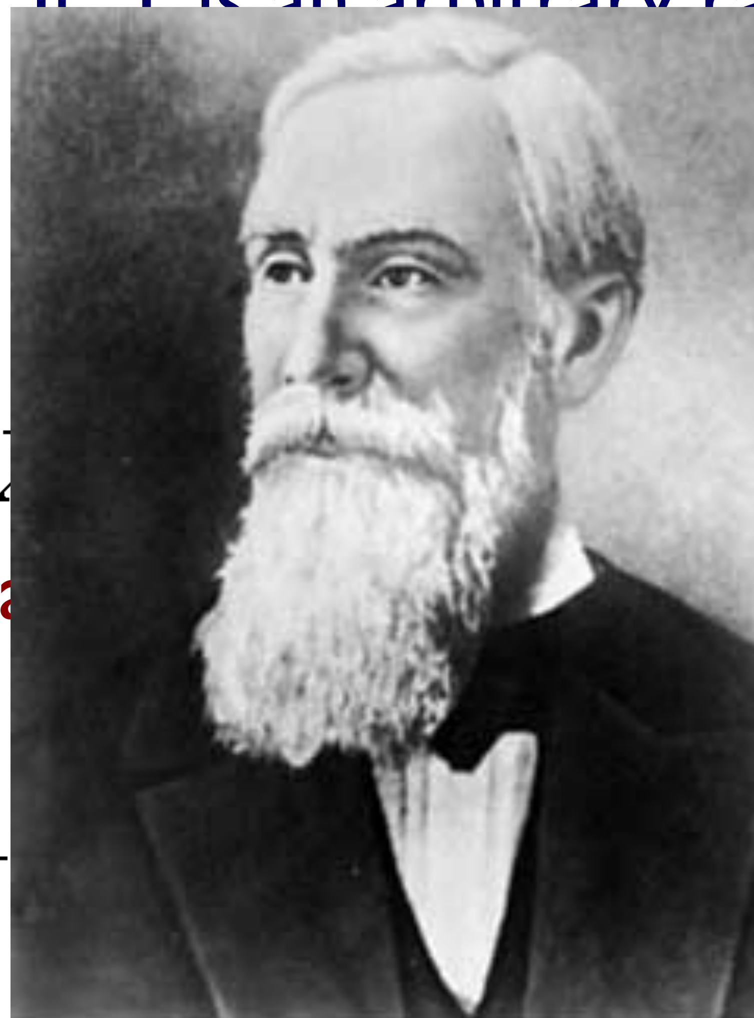
X is non-negative, so we can apply Markov's inequality:

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

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

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

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



Proof: Let X ...

X is non-nega... Markov's inequality:

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

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

Chebyshev's inequality

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

E.g., suppose:

Y = money spent on advertising in a day

E[Y] = 1500

Var[Y] = 500^2 (i.e. SD[Y] = 500)

$$P(Y \geq 6000) = P(|Y - \mu| \geq 4500)$$

$$\leq \frac{500^2}{4500^2} = \frac{1}{81} \approx 0.012$$

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:

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

$Y \sim \text{Bin}(15000, 0.1)$
$\mu = E[Y] = 1500, \sigma = \sqrt{\text{Var}(Y)} = 36.7$

1. $P(Y \geq 6000) = P(Y \geq 4\mu) \leq 1/4$ (Markov)
2. $P(Y \geq 6000) = P(Y\text{-}\mu \geq 122\sigma) \leq 7 \times 10^{-5}$ (Chebyshev)
3. $P(Y \geq 6000) \ll 10^{-1600}$ ($Y \sim \text{Poi}(1500)$)
4. The exact (binomial) value is $\approx 4 \times 10^{-2031}$

1,2,5 are easy calcs; 3 & 4 are *not* (underflow, etc.)

5. $P(Y \geq 6000) \lesssim 10^{-1945}$ (Chernoff, below; easy)

Method: B&T pp 284-7

Suppose X ~ Bin(n,p)
μ = E[X] = pn

Chernoff bound:

For any $0 < \delta < 1$,

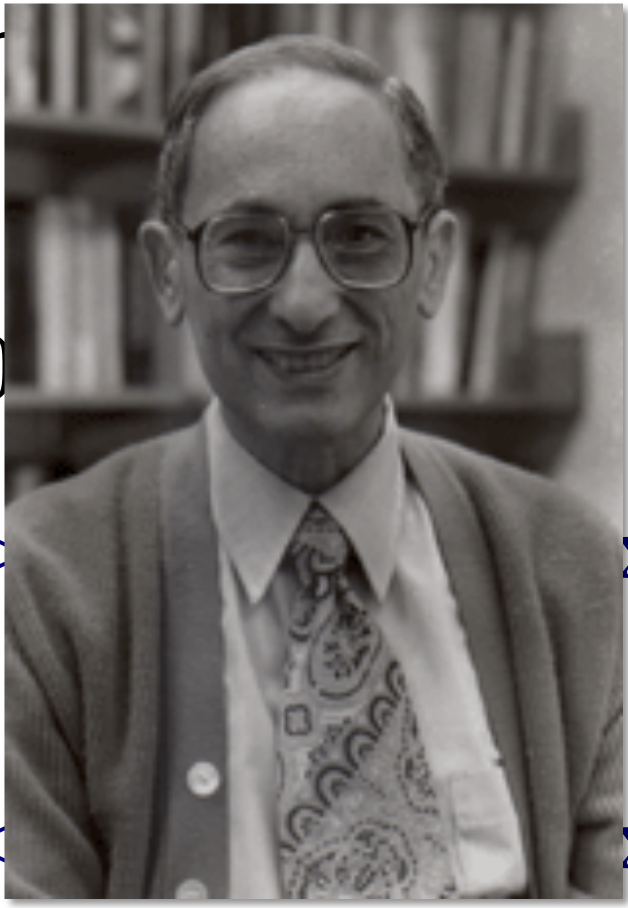
$$P(X > (1+\delta)\mu) \leq \exp\left(-\frac{\delta^2\mu}{3}\right)$$

$$P(X < (1-\delta)\mu) \leq \exp\left(-\frac{\delta^2\mu}{2}\right)$$

B&T pp 284-7

Suppose X ~ Bin(n,p)

μ = E[X] = pr



Chernoff bou

For any 0

$$P(X > \ldots \exp\left(-\frac{\delta^2\mu}{3}\right)$$

$$P(X < \ldots \exp\left(-\frac{\delta^2\mu}{2}\right)$$

router buffers

Model: n = 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 ≈1/2, every second?
  ≈1000 (P(X>E[X]) ≈ ½ when X ~ Binomial(100000, .01))
- With probability at most 10^{-6}, every hour?
  1257
- With probability at most 10^{-6}, every year?
  1305
- With probability at most 10^{-6}, since Big Bang?
  1404

Exercise: How would you formulate the exact answer to this problem in terms of binomial probabilities? Can you get a numerical answer?

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

Let p = probability of buffer overflow in 1 second

By the Chernoff bound

$$p = P(X > (1+\delta)\mu) \leq \exp\left(-\frac{\delta^2\mu}{3}\right)$$

Overflow probability in n seconds

$= 1-(1-p)^n \leq np \leq n \exp(-\delta^2\mu/3)$,

which is $\leq \varepsilon$ provided $\delta \geq \sqrt{(3/\mu)\ln(n/\varepsilon)}$.

For $\varepsilon = 10^{-6}$ per hour: $\delta \approx .257$, buffers = 1257

For $\varepsilon = 10^{-6}$ per year: $\delta \approx .305$, buffers = 1305

For $\varepsilon = 10^{-6}$ per 15BY: $\delta \approx .404$, buffers = 1404

Tail bounds – bound probabilities of extreme events

Important, e.g., for "risk management" applications

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 "≈"