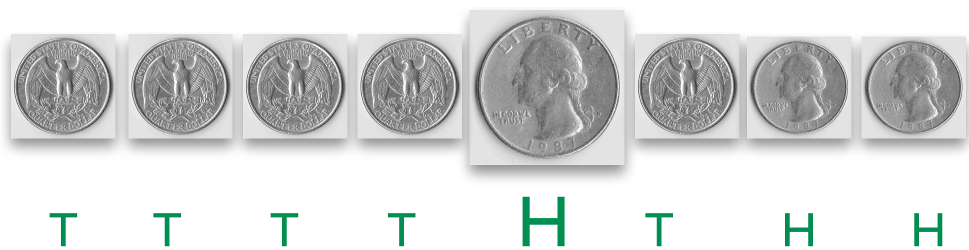
random variables



let X = index of

A *random variable* X assigns a real number to each outcome in a probability space.

Ex.

Let *H* be the number of Heads when 20 coins are tossed

Let *T* be the total of 2 dice rolls

Let *X* be the number of coin tosses needed to see 1st head

Note; even if the underlying experiment has "equally likely outcomes," the associated random variable may not

Outcome	*H*	*P(H)*
TT	0	P(H=0) = 1/4
TH	1	} P(H=1) = 1/2
HT	1	
HH	2	P(H=2) = 1/4

20 balls numbered 1, 2, ..., 20

Draw 3 without replacement

Let X = the maximum of the numbers on those 3 balls

What is P(X ≥ 17)

$$P(X = 20) = \binom{19}{2} / \binom{20}{3} = \frac{3}{20} = 0.150$$

$$P(X = 19) = \binom{18}{2} / \binom{20}{3} = \frac{18 \cdot 17/2!}{20 \cdot 19 \cdot 18/3!} \approx 0.134$$

$$\vdots$$

$$\sum_{i=17}^{20} P(X = i) \approx 0.508$$

Alternatively:

$$P(X \geq 17) = 1 - P(X < 17) = 1 - \binom{16}{3} \Big/ \binom{20}{3} \approx 0.508$$

Flip a (biased) coin repeatedly until 1st head observed

How many flips? Let X be that number.

P(X=1) = P(H) = p

P(X=2) = P(TH) = (1-p)p

P(X=3) = P(TTH) = $(1-p)^2p$

...

Check that it is a valid probability distribution:

$$P\left(\bigcup_{i\geq 1}\{X=i\}\right) = \sum_{i\geq 1}(1-p)^{i-1}p = p\sum_{i\geq 0}(1-p)^i = p\frac{1}{1-(1-p)} = 1$$

A *discrete* random variable is one taking on a countable number of possible values.

Ex:

X = sum of 3 dice, $3 \leq X \leq 18, X \in N$

Y = index of 1st head in seq of coin flips, $1 \leq Y, Y \in N$

Z = largest prime factor of (1+Y), $Z \in \{2, 3, 5, 7, 11, ...\}$

If X is a discrete random variable taking on values from a countable set $T \subseteq R$, then

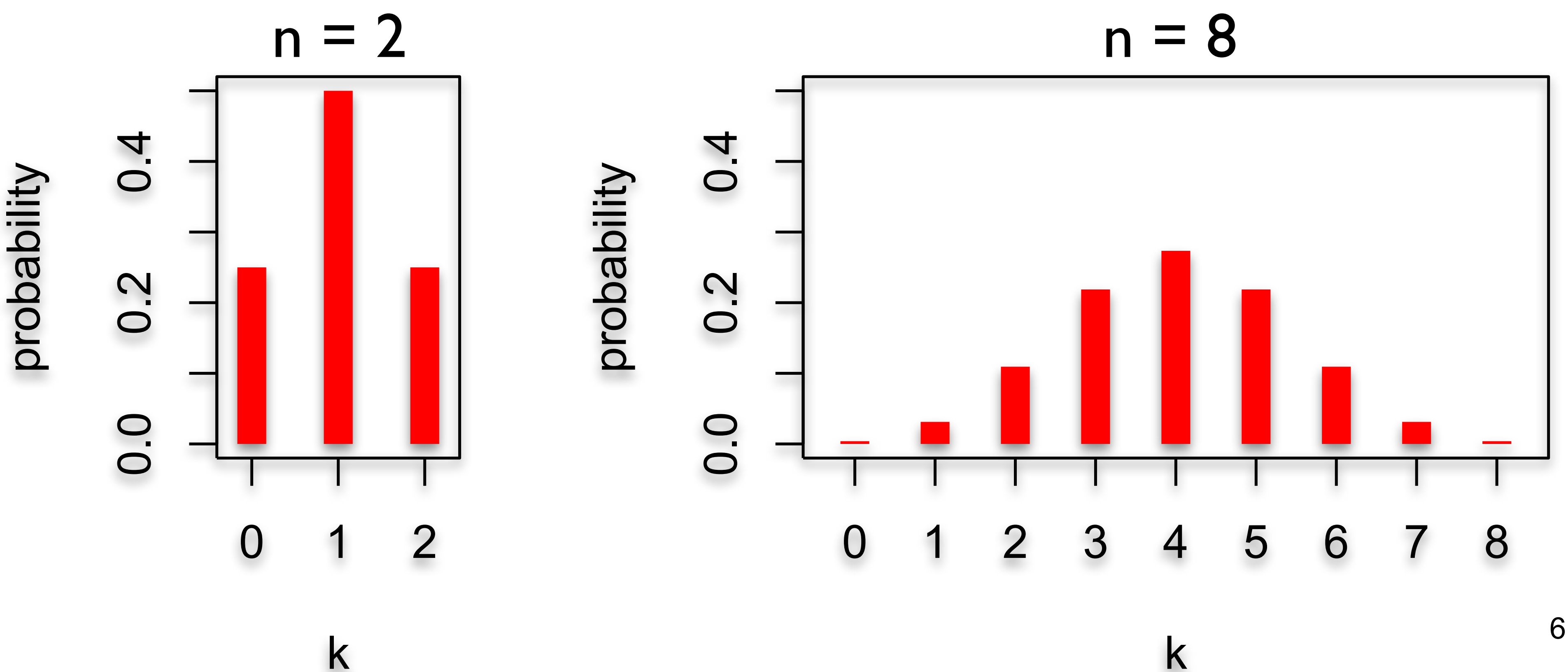
$$p(a) = \begin{cases} P(X = a) & \text{for } a \in T \\ 0 & \text{otherwise} \end{cases}$$

is called the *probability mass function*. Note: $\sum_{a \in T} p(a) = 1$

Let X be the number of heads observed in n coin flips

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \text{ where } p = P(H)$$

Probability mass function:



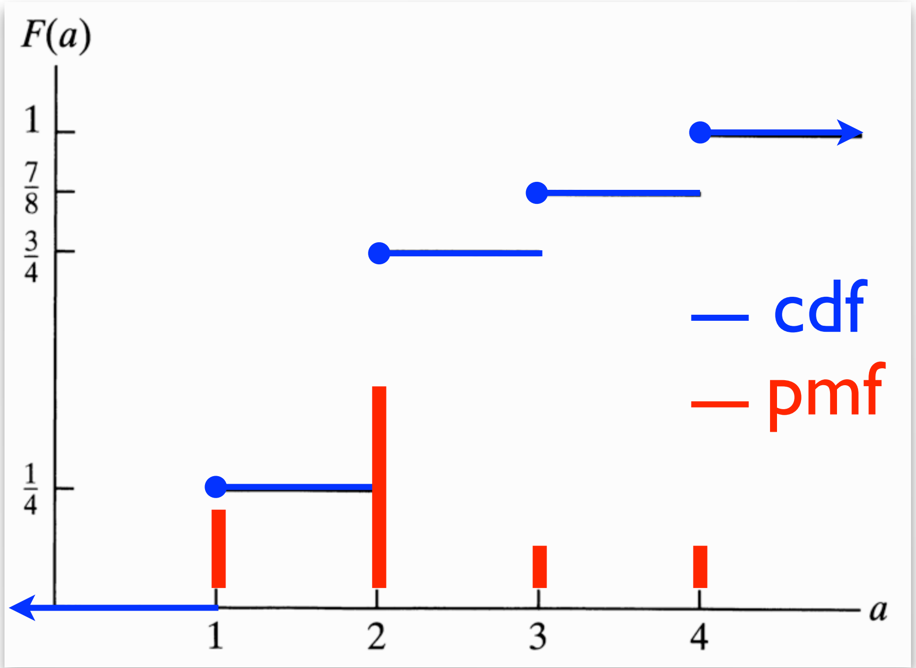
The *cumulative distribution function* for a random variable X is the function $F: \mathbb{R} \to [0,1]$ defined by

$$F(a) = P[X \le a]$$

Ex: if X has probability mass function given by:

$$p(1) = \frac{1}{4} \quad p(2) = \frac{1}{2} \quad p(3) = \frac{1}{8} \quad p(4) = \frac{1}{8}$$

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{4} & 1 \le a < 2 \\ \frac{3}{4} & 2 \le a < 3 \\ \frac{7}{8} & 3 \le a < 4 \\ 1 & 4 \le a \end{cases}$$



NB: for discrete random variables, be careful about "≤" vs "<"

For a discrete r.v. X with p.m.f. p(•), the *expectation of X*, aka *expected value* or *mean*, is

$E[X] = \Sigma_x xp(x)$

average of random values, weighted by their respective probabilities

For the equally-likely outcomes case, this is just the average of the possible random values of X

For *un*equally-likely outcomes, it is again the average of the possible random values of X, weighted by their respective probabilities

Ex 1: Let X = value seen rolling a fair die p(1), p(2), ..., p(6) = 1/6

$$E[X] = \sum_{i=1}^{6} ip(i) = \frac{1}{6}(1+2+\cdots+6) = \frac{21}{6} = 3.5$$

Ex 2: Coin flip; X = +1 if H (win \$1), -1 if T (lose \$1)

E[X] = (+1)•p(+1) + (-1)•p(-1) = 1•(1/2) +(-1)•(1/2) = 0

For a discrete r.v. X with p.m.f. p(•), the *expectation of X*, aka *expected value* or *mean*, is

$$E[X] = \Sigma_x xp(x)$$

average of random values, weighted by their respective probabilities

Another view: A gambling game. If X is how much you win playing the game once, how much would you expect to win, on average, per game when repeatedly playing?

Ex 1: Let X = value seen rolling a fair die p(1), p(2), ..., p(6) = 1/6

If you win X dollars for that roll, how much do you expect to win?

$$E[X] = \sum_{i=1}^{6} ip(i) = \frac{1}{6}(1 + 2 + \cdots + 6) = \frac{21}{6} = 3.5$$

Ex 2: Coin flip; X = +1 if H (win \$1), -1 if T (lose \$1)

E[X] = (+1)•p(+1) + (-1)•p(-1) = 1•(1/2) +(-1)•(1/2) = 0

"a fair game": in repeated play you expect to win as much as you lose. Long term net gain/loss = 0.

Let X be the number of flips up to & including 1st head observed in repeated flips of a biased coin. If I pay you $1 per flip, how much money would you expect to make?

$$\begin{aligned} P(H) &= p; \quad P(T) = 1 - p = q \\ p(i) &= pq^{i-1} \\ E(x) &= \textstyle\sum_{i \geq 1} ip(i) = \sum_{i \geq 1} ipq^{i-1} = p\sum_{i \geq 1} iq^{i-1} \quad (*) \end{aligned}$$

A calculus trick:

$$\sum_{i \geq 1} iy^{i-1} = \sum_{i \geq 1} \frac{d}{dy} y^i = \sum_{i \geq 0} \frac{d}{dy} y^i = \frac{d}{dy} \sum_{i \geq 0} y^i = \frac{d}{dy} \frac{1}{1-y} = \frac{1}{(1-y)^2}$$

$dy^0/dy = 0$

So $(*)$ becomes:

$$E[X] = p \sum_{i \geq i} iq^{i-1} = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

How much would you pay to play?

E.g.:

p=1/2; on average head every 2nd flip

p=1/10; on average, head every 10th flip.

Calculating E[g(X)]:

Y=g(X) is a new r.v. Calc P[Y=j], then apply defn:

X = sum of 2 dice rolls

i	p(i) = P[X=i]	i•p(i)
2	1/36	2/36
3	2/36	6/36
4	3/36	12/36
5	4/36	20/36
6	5/36	30/36
7	6/36	42/36
8	5/36	40/36
9	4/36	36/36
10	3/36	30/36
11	2/36	22/36
12	1/36	12/36

$E[X] = \Sigma_i \, ip(i) =$ 252/36 = 7

Y = g(X) = X mod 5

j	q(j) = P[Y = j]	j•q(j)
0	4/36+3/36 =7/36	0/36
1	5/36+2/36 =7/36	7/36
2	1/36+6/36+1/36 =8/36	16/36
3	2/36+5/36 =7/36	21/36
4	3/36+4/36 =7/36	28/36

$E[Y] = \Sigma_j \, jq(j) =$ 72/36 = 2

Calculating E[g(X)]: Another way – add in a different order, using P[X=...] instead of calculating P[Y=...]

X = sum of 2 dice rolls

i	p(i) = P[X=i]	g(i)•p(i)
2	1/36	2/36
3	2/36	6/36
4	3/36	12/36
5	4/36	0/36
6	5/36	5/36
7	6/36	12/36
8	5/36	15/36
9	4/36	16/36
10	3/36	0/36
11	2/36	2/36
12	1/36	2/36

$E[g(X)] = \Sigma_i\ g(i)p(i) =$ 72/36 = 2

Y = g(X) = X mod 5

j	q(j) = P[Y = j]	j•q(j)
0	4/36+3/36 =7/36	0/36
1	5/36+2/36 =7/36	7/36
2	1/36+6/36+1/36 =8/36	16/36
3	2/36+5/36 =7/36	21/36
4	3/36+4/36 =7/36	28/36

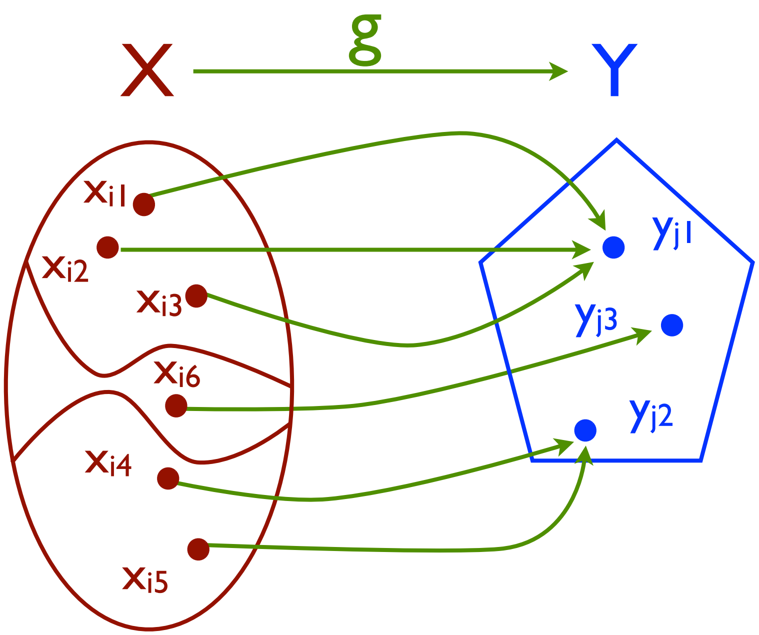
$E[Y] = \Sigma_j\ jq(j) =$ 72/36 = 2

expectation of a *function* of a random variable

Above example is not a fluke.

Theorem: if Y = g(X), then $E[Y] = \Sigma_i\, g(x_i)p(x_i)$, where x_i, i = 1, 2, ... are all possible values of X.

Proof: Let y_j, j = 1, 2, ... be all possible values of Y.



Note that $S_j = \{ x_i \mid g(x_i)=y_j \}$ is a partition of the domain of g.

$$
\begin{aligned}
\sum_i g(x_i)p(x_i) &= \sum_j \sum_{i:g(x_i)=y_j} g(x_i)p(x_i) \\
&= \sum_j \sum_{i:g(x_i)=y_j} y_j p(x_i) \\
&= \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i) \\
&= \sum_j y_j P\{g(X) = y_j\} \\
&= E[g(X)]
\end{aligned}
$$

A & B each bet \$1, then flip 2 coins:

HH	A wins \$2
HT	Each takes back \$1
TH	
TT	B wins \$2

Let X be A's net gain: +1, 0, -1, resp.:

$P(X = +1) = 1/4$
$P(X = 0) = 1/2$
$P(X = -1) = 1/4$

What is E[X]?

$$E[X] = 1 \cdot 1/4 + 0 \cdot 1/2 + (-1) \cdot 1/4 = 0$$

What is $E[X^2]$?

$$E[X^2] = 1^2 \cdot 1/4 + 0^2 \cdot 1/2 + (-1)^2 \cdot 1/4 = 1/2$$

Note:
$E[X^2] \neq E[X]^2$

Linearity of expectation, I

For any constants a, b: E[aX + b] = aE[X] + b

Proof:

$$
\begin{aligned}
E[aX+b] &= \sum_x (ax+b)\cdot p(x) \\
&= a\sum_x xp(x) + b\sum_x p(x) \\
&= aE[X]+b
\end{aligned}
$$

Example:

Q: In the 2-person coin game above, what is E[2X+1]?

A: E[2X+1] = 2E[X]+1 = 2•0 + 1 = 1

Linearity, II

Let X and Y be two random variables derived from outcomes of a single experiment. Then

$$E[X+Y] = E[X] + E[Y]$$

True even if X,Y *dependent*

Proof: Assume the sample space S is countable. (The result is true without this assumption, but I won't prove it.) Let X(s), Y(s) be the values of these r.v.'s for outcome s∈S.

Claim: $E[X] = \sum_{s \in S} X(s) \cdot p(s)$

Proof: similar to that for "expectation of a function of an r.v.," i.e., the events "X=x" partition S, so sum above can be rearranged to match the definition of $E[X] = \sum_x x \cdot P(X = x)$

Then:

$$E[X+Y] = \Sigma_{s \in S}(X[s] + Y[s])\, p(s)$$
$$= \Sigma_{s \in S} X[s]\, p(s) + \Sigma_{s \in S} Y[s]\, p(s) = E[X] + E[Y]$$

Example

X = # of heads in *one* coin flip, where P(X=1) = p.

What is $E(X)$?

$E[X] = 1 \cdot p + 0 \cdot (1-p) = p$

Let X_i, $1 \leq i \leq n$, be # of H in flip of coin with $P(X_i=1) = p_i$

What is the expected number of heads when all are flipped?

$E[\Sigma_i X_i] = \Sigma_i E[X_i] = \Sigma_i p_i$

Special case: $p_1 = p_2 = ... = p$:

$E[\#\text{ of heads in } n \text{ flips}] = pn$

Note:

Linearity is special!

It is *not* true in general that

$E[X \cdot Y] = E[X] \cdot E[Y]$

$E[X^2] = E[X]^2$ ← counterexample above

$E[X/Y] = E[X] / E[Y]$

$E[\text{asinh}(X)] = \text{asinh}(E[X])$

⋮

Alice & Bob are gambling (again). X = Alice's gain per flip:

$$X = \begin{cases} +1 & \text{if Heads} \\ -1 & \text{if Tails} \end{cases}$$

E[X] = 0

… Time passes …

Alice (yawning) says "let's raise the stakes"

$$Y = \begin{cases} +1000 & \text{if Heads} \\ -1000 & \text{if Tails} \end{cases}$$

E[Y] = 0, as before.

Are you (Bob) equally happy to play the new game?

E[X] measures the "average" or "central tendency" of X.

What about its *variability*?

Definition

The *variance* of a random variable X with mean $E[X] = \mu$ is $Var[X] = E[(X-\mu)^2]$, often denoted σ^2.

Alice & Bob are gambling (again). X = Alice's gain per flip:

$$X = \begin{cases} +1 & \text{if Heads} \\ -1 & \text{if Tails} \end{cases}$$

E[X] = 0

Var[X] = 1

... Time passes ...

Alice (yawning) says "let's raise the stakes"

$$Y = \begin{cases} +1000 & \text{if Heads} \\ -1000 & \text{if Tails} \end{cases}$$

E[Y] = 0, as before.

Var[Y] = 1,000,000

Are you (Bob) equally happy to play the new game?

E[X] measures the "average" or "central tendency" of X.

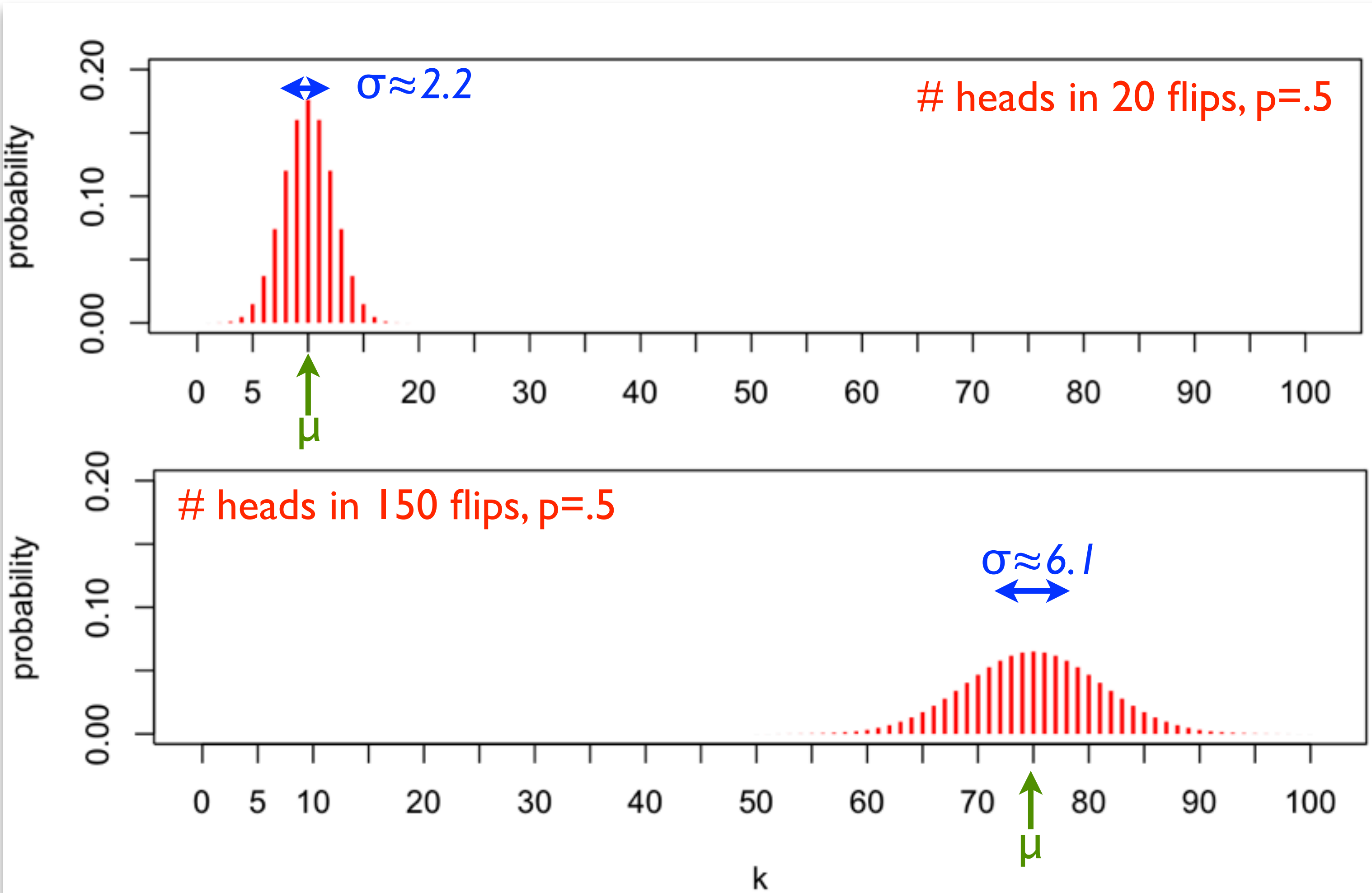
What about its *variability*?

Definition

The *variance* of a random variable X with mean $E[X] = \mu$ is $Var[X] = E[(X-\mu)^2]$, often denoted σ^2.

The *standard deviation* of X is $\sigma = \sqrt{Var[X]}$

μ = E[X] is about *location*; σ = $\sqrt{Var(X)}$ is about *spread*



Two games:

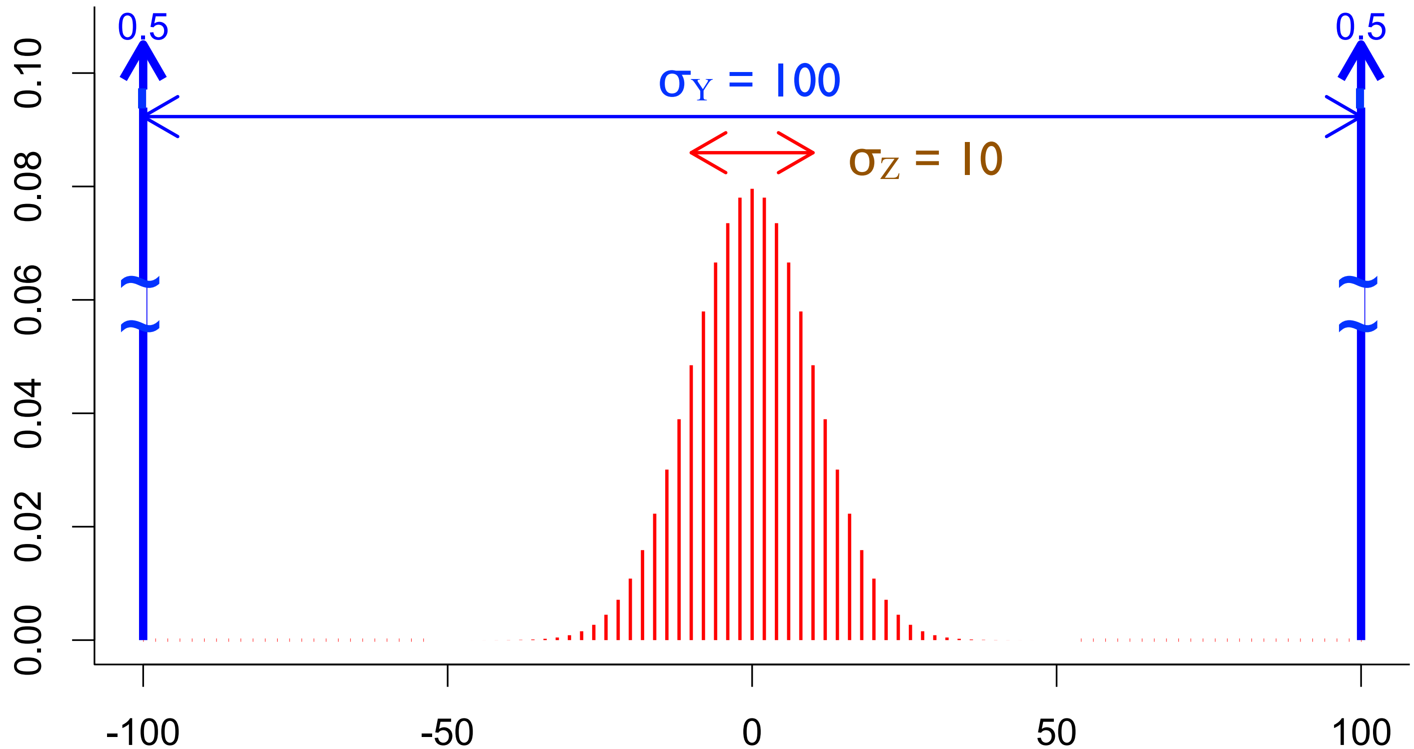
a) flip 1 coin, win Y = \$100 if heads, \$-100 if tails

b) flip 100 coins, win Z = (#(heads) - #(tails)) dollars

Same expectation in both: E[Y] = E[Z] = 0

Same extremes in both: max gain = \$100; max loss = \$100

But *variability* is very different:



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

$$\begin{aligned}
\mathrm{Var}(X) &= E[(X - \mu)^2] \\
&= \sum_x (x - \mu)^2 p(x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2) p(x) \\
&= \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x) \\
&= E[X^2] - 2\mu^2 + \mu^2 \\
&= E[X^2] - \mu^2
\end{aligned}$$

Example:

What is Var[X] when X is outcome of one fair die?

$$E[X^2] = 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{6}\right) + 4^2\left(\frac{1}{6}\right) + 5^2\left(\frac{1}{6}\right) + 6^2\left(\frac{1}{6}\right)$$
$$= \left(\frac{1}{6}\right)(91)$$

E[X] = 7/2, so

$$\mathrm{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

$$\mathrm{Var}[aX+b] = a^2\,\mathrm{Var}[X]$$

$$\begin{aligned}\mathrm{Var}(aX + b) &= E[(aX + b - a\mu - b)^2]\\ &= E[a^2(X - \mu)^2]\\ &= a^2E[(X - \mu)^2]\\ &= a^2\mathrm{Var}(X)\end{aligned}$$

Ex:

$$X = \begin{cases} +1 & \text{if Heads} \\ -1 & \text{if Tails} \end{cases}$$

E[X] = 0

Var[X] = 1

$$Y = \begin{cases} +1000 & \text{if Heads} \\ -1000 & \text{if Tails} \end{cases}$$

Y = 1000 X

E[Y] = E[1000 X] = 1000 E[x] = 0

Var[Y] = Var[1000 X]

$= 10^6 \mathrm{Var}[X] = 10^6$

In general: $\mathrm{Var}[X+Y] \neq \mathrm{Var}[X] + \mathrm{Var}[Y]$

Ex 1:

Let $X = \pm 1$ based on 1 coin flip

As shown above, $E[X] = 0$, $\mathrm{Var}[X] = 1$

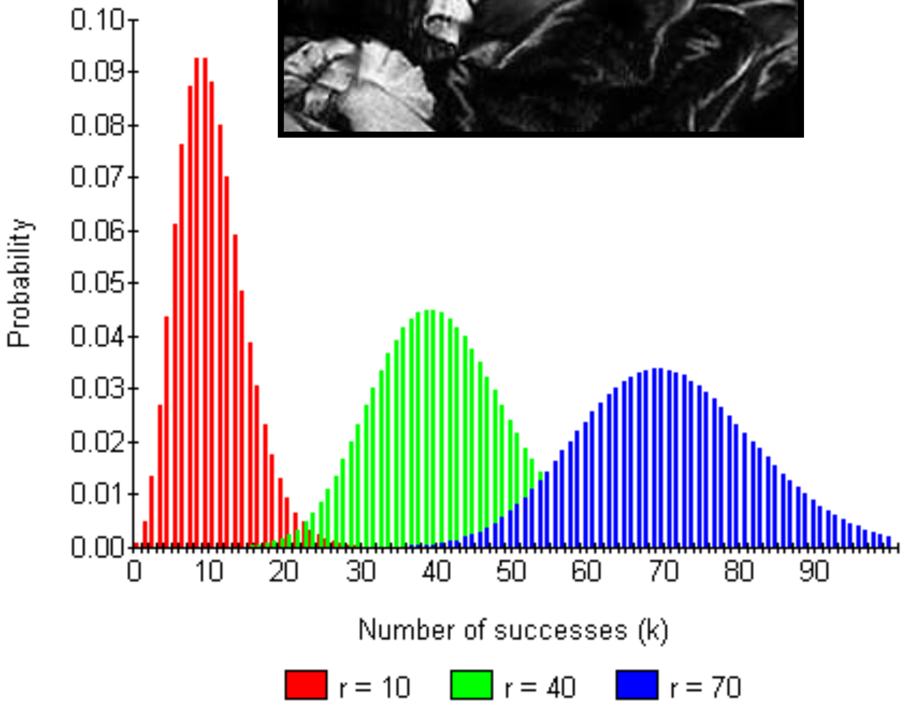
Let $Y = -X$; then $\mathrm{Var}[Y] = (-1)^2\mathrm{Var}[X] = 1$

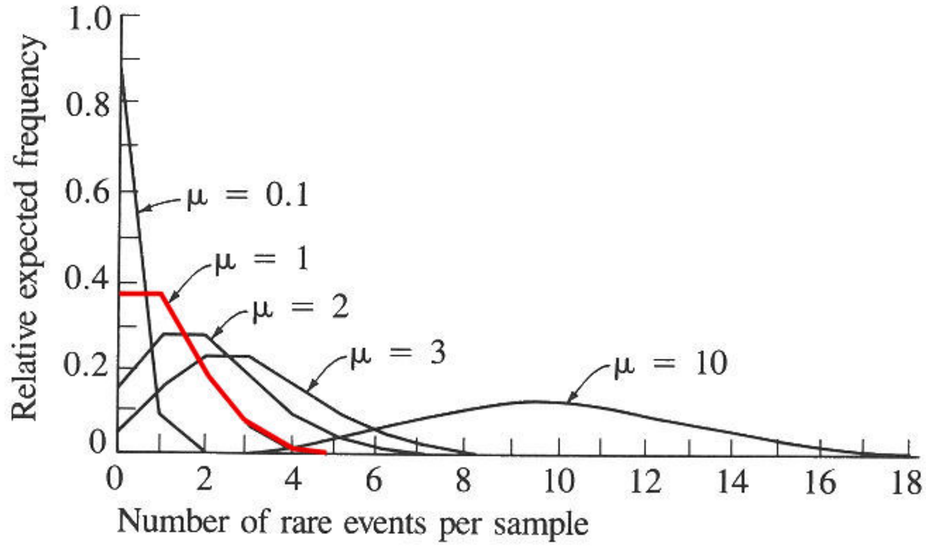
But $X+Y = 0$, always, so $\mathrm{Var}[X+Y] = 0$

Ex 2:

As another example, is $\mathrm{Var}[X+X] = 2\mathrm{Var}[X]$?

a zoo of (discrete) random variables





An experiment results in "Success" or "Failure"

X is a random *indicator variable* (1=success, 0=failure)

$P(X=1) = p$ and $P(X=0) = 1\text{-}p$

X is called a *Bernoulli* random variable: $X \sim Ber(p)$

$E[X] = E[X^2] = p$

$Var(X) = E[X^2] – (E[X])^2 = p – p^2 = p(1\text{-}p)$

Examples:

coin flip

random binary digit

whether a disk drive crashed

Jacob (aka James, Jacques) Bernoulli, 1654 – 1705

Consider n independent random variables $Y_i \sim \text{Ber}(p)$

$X = \Sigma_i Y_i$ is the number of successes in n trials

X is a *Binomial* random variable: $X \sim \text{Bin}(n,p)$

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \quad i = 0, 1, \ldots, n$$

By Binomial theorem, $\sum_{i=0}^{n} P(X = i) = 1$

Examples

\# of heads in n coin flips

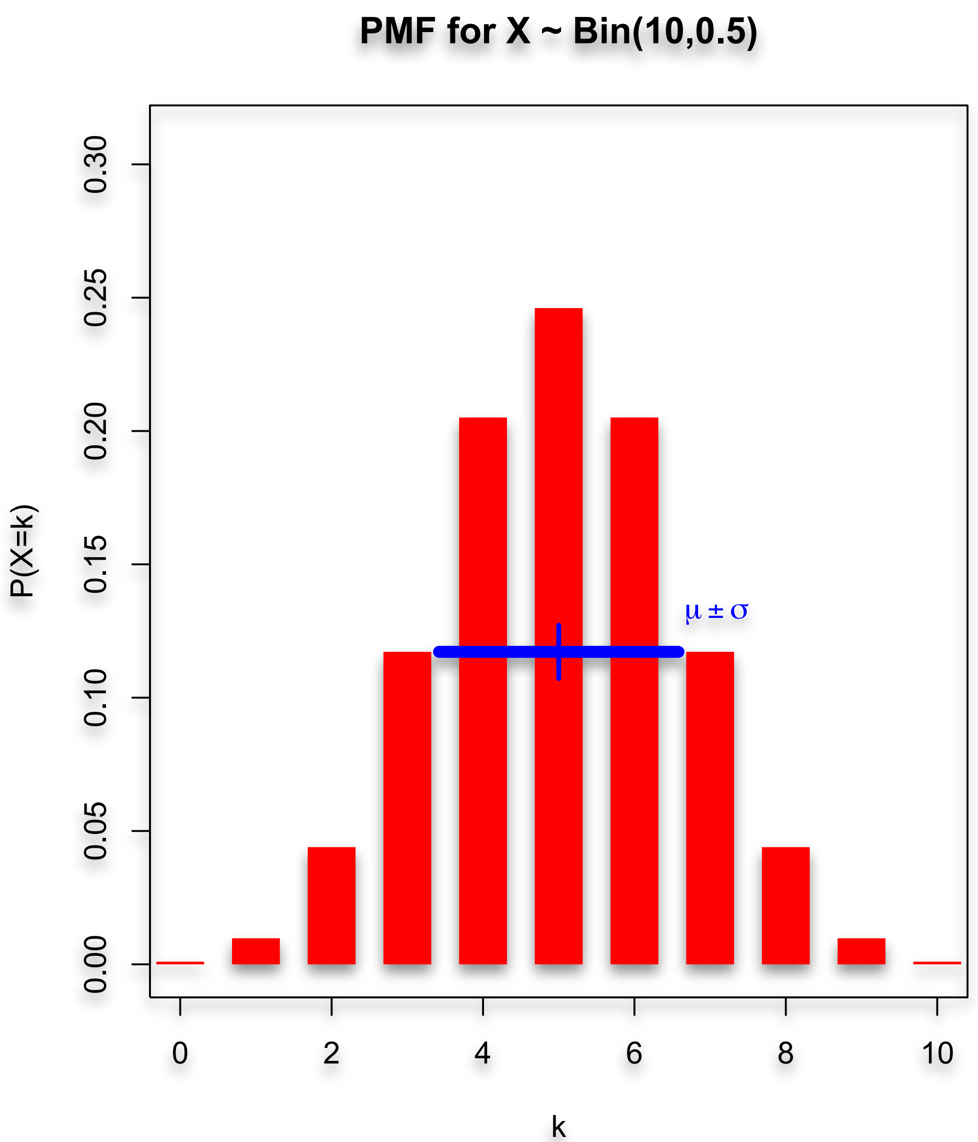
\# of 1's in a randomly generated length n bit string

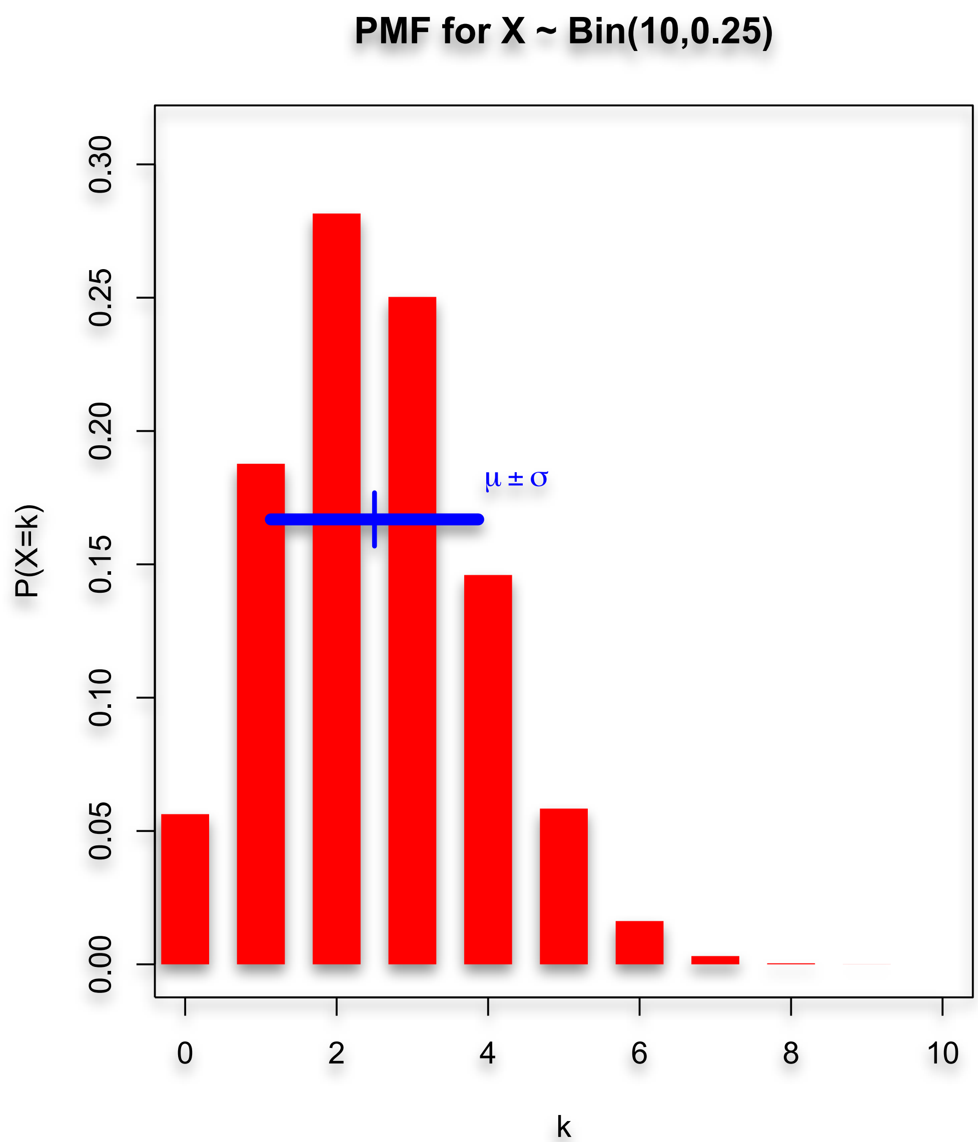
\# of disk drive crashes in a 1000 computer cluster

$E[X] = pn$

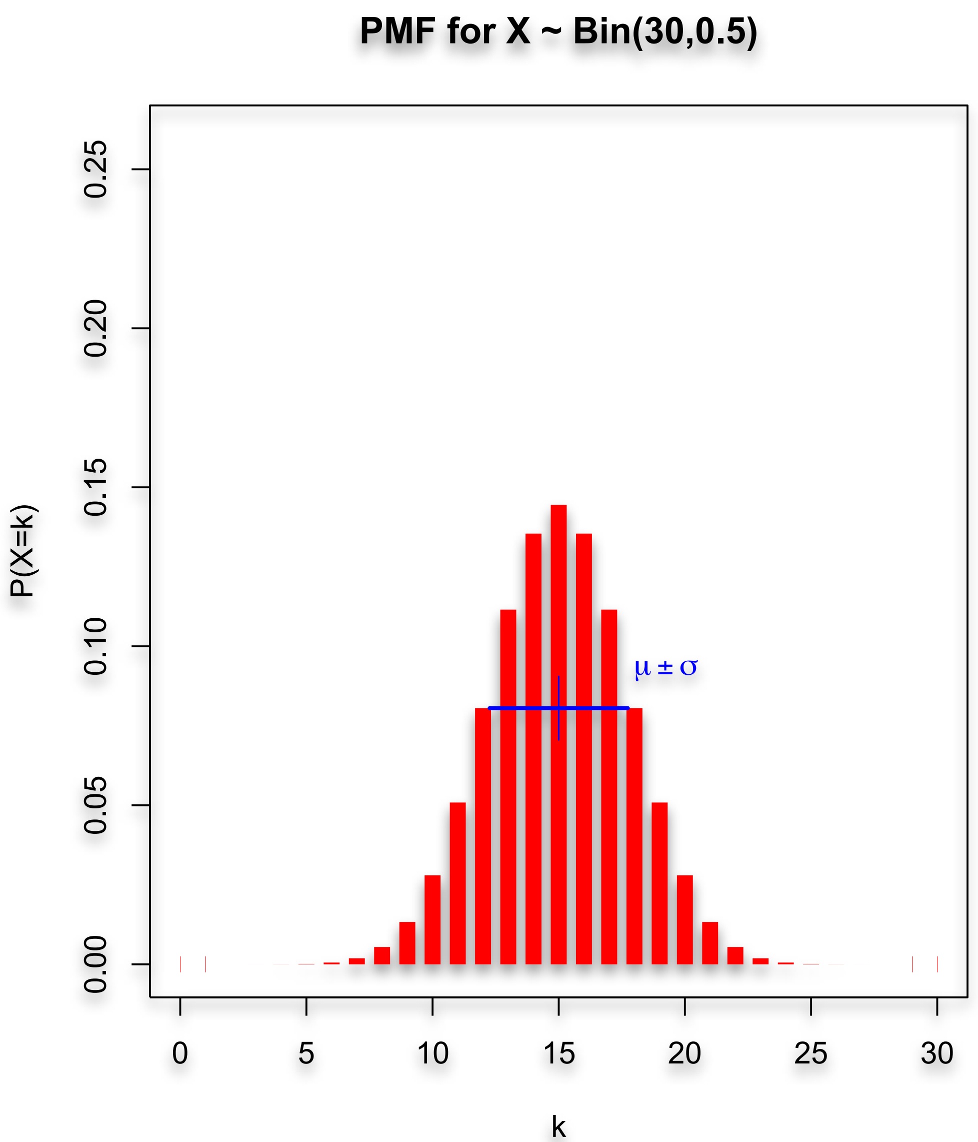
$\text{Var}(X) = p(1-p)n$ ←(proof below, twice)

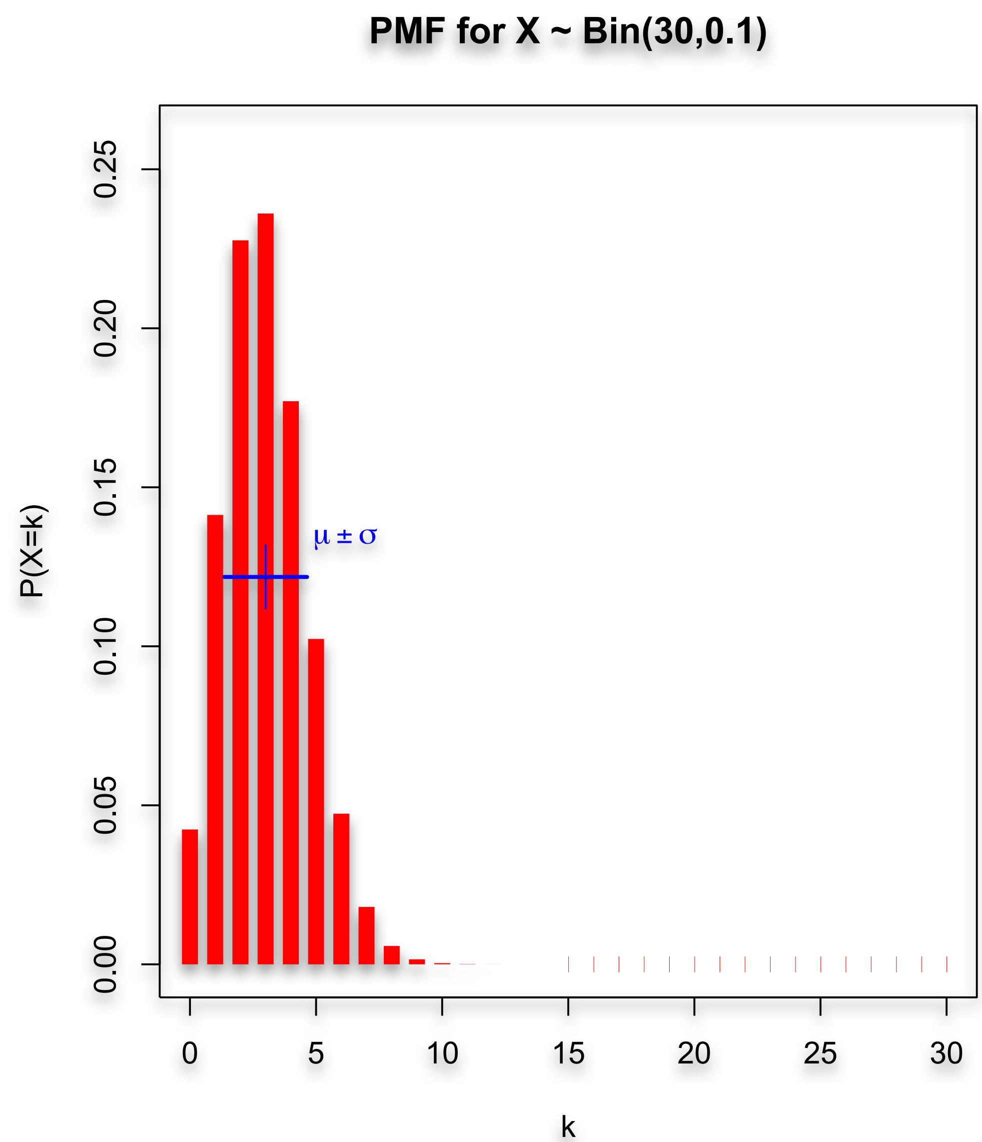
binomial pmfs





binomial pmfs





mean and variance of the binomial

$$E[X^k] = \sum_{i=0}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i}$$
$$= \sum_{i=1}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i}$$

using

$$i\binom{n}{i} = n\binom{n-1}{i-1}$$

$$E[X^k] = np \sum_{i=1}^{n} i^{k-1} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i}$$

letting
j = i-1

$$= np \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{n-1}{j} p^j (1-p)^{n-1-j}$$
$$= np E[(Y+1)^{k-1}]$$

where Y is a binomial random variable with parameters $n-1, p$.

k=1 gives: $E[X] = np$; k=2 gives $E[X^2]=np[(n-1)p+1]$

hence: $$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$
$$= np[(n-1)p+1] - (np)^2$$
$$= np(1-p)$$

Two random variables X and Y are independent if for every value i that X can take, and any value j that Y can take

$$Pr(X=i, Y=j) = Pr(X=i)Pr(Y=j)$$

Theorem: If X & Y are *independent*, then E[X•Y] = E[X]•E[Y]

Proof:

Let $x_i, y_i, i = 1, 2, \ldots$ be the possible values of X, Y.

$$
\begin{aligned}
E[X \cdot Y] &= \sum_i \sum_j x_i \cdot y_j \cdot P(X = x_i \wedge Y = y_j) \\
&= \sum_i \sum_j x_i \cdot y_j \cdot P(X = x_i) \cdot P(Y = y_j) \\
&= \sum_i x_i \cdot P(X = x_i) \cdot \left(\sum_j y_j \cdot P(Y = y_j) \right) \\
&= E[X] \cdot E[Y]
\end{aligned}
$$

independence

Note: *NOT* true in general; see earlier example $E[X^2] \neq E[X]^2$

variance of *independent* r.v.s is additive

(Bienaymé, 1853)

Theorem: If X & Y are *independent*, then

Var[X+Y] = Var[X]+Var[Y]

Proof: Let

$$\begin{aligned} \widehat{X} &= X - E[X] & \widehat{Y} &= Y - E[Y] \\ E[\widehat{X}] &= 0 & E[\widehat{Y}] &= 0 \\ Var[\widehat{X}] &= Var[X] & Var[\widehat{Y}] &= Var[Y] \end{aligned}$$

Var(aX+b) = a²Var(X)

$$\begin{aligned} Var[X+Y] &= Var[\widehat{X}+\widehat{Y}] \\ &= E[(\widehat{X}+\widehat{Y})^2] - (E[\widehat{X}+\widehat{Y}])^2 \\ &= E[\widehat{X}^2 + 2\widehat{X}\widehat{Y} + \widehat{Y}^2] - 0 \\ &= E[\widehat{X}^2] + 2E[\widehat{X}\widehat{Y}] + E[\widehat{Y}^2] \\ &= Var[\widehat{X}] + 0 + Var[\widehat{Y}] \\ &= Var[X] + Var[Y] \end{aligned}$$

If $Y_1, Y_2, \ldots, Y_n \sim \text{Ber}(p)$ and independent,

then $X = \sum_{i=1}^{n} Y_i \sim \text{Bin}(n, p)$.

$E[X] = E[\sum_{i=1}^{n} Y_i] = nE[Y_1] = np$

$\text{Var}[X] = \text{Var}[\sum_{i=1}^{n} Y_i] = n\text{Var}[Y_1] = np(1-p)$

A RAID-like disk array consists of *n* drives, each of which will fail independently with probability *p*. Suppose it can operate effectively if at least one-half of its components function, e.g., by "majority vote."

For what values of *p* is a 5-component system more likely to operate effectively than a 3-component system?

X_5 = # failed in 5-component system ~ Bin(5, p)

X_3 = # failed in 3-component system ~ Bin(3, p)

X_5 = # failed in 5-component system ~ Bin(5, p)
X_3 = # failed in 3-component system ~ Bin(3, p)

P(5 component system effective) = P(X_5 < 5/2)

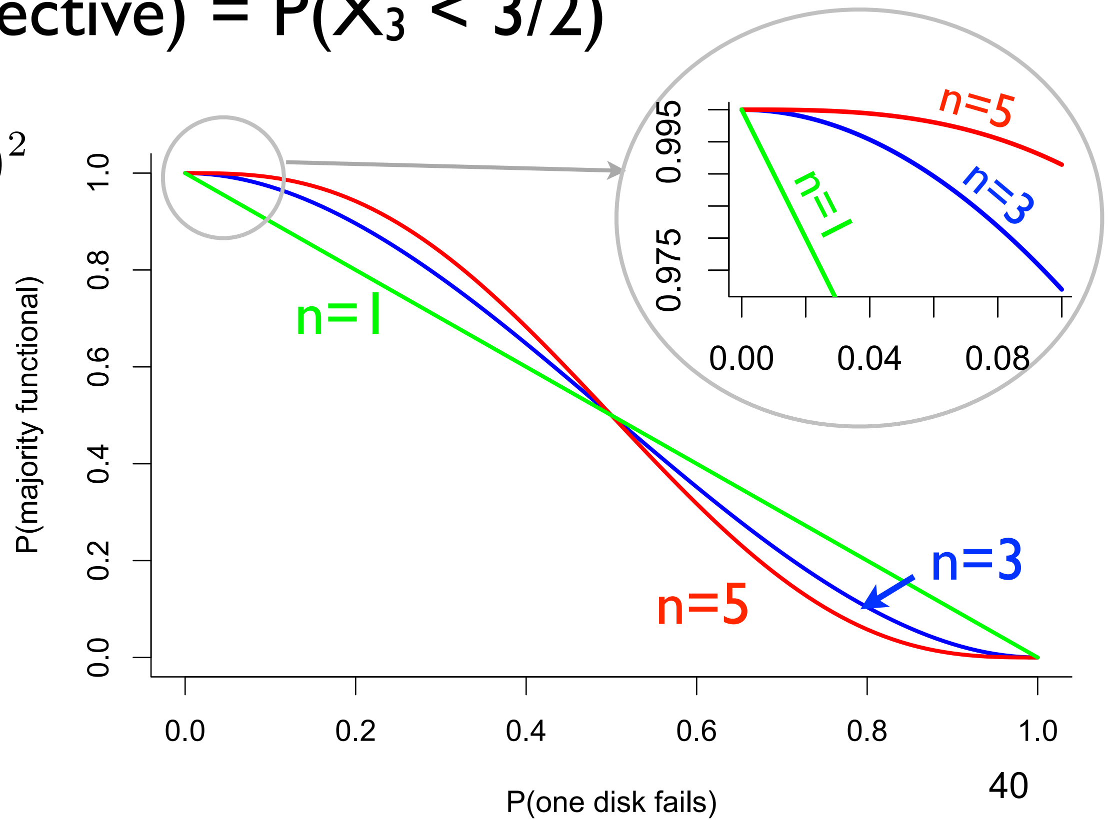
$$\binom{5}{0}p^0(1-p)^5 + \binom{5}{1}p^1(1-p)^4 + \binom{5}{2}p^2(1-p)^3$$

P(3 component system effective) = P(X_3 < 3/2)

$$\binom{3}{0}p^0(1-p)^3 + \binom{3}{1}p^1(1-p)^2$$

Calculation:

5-component system is better iff p < 1/2



Binomial distribution: models & reality

Sending a bit string over the network

- $n = 4$ bits sent, each corrupted with probability 0.1
- X = # of corrupted bits, $X \sim \text{Bin}(4, 0.1)$
- In real networks, large bit strings (length $n \approx 10^4$)
- Corruption probability is very small: $p \approx 10^{-6}$

Extreme n and p values arise in many cases

- # bit errors in file written to disk
- # of typos in a book
- # of elements in particular bucket of large hash table
- # of server crashes per day in giant data center
- # facebook login requests sent to a particular server

Binomial with parameters n and 1/m. Define $\lambda = n/m$

What is distribution of X, the number of successes?

$$Pr(X=0) = (1 - 1/m)^n \approx e^{-\frac{n}{m}} = e^{-\lambda}$$

$$1 - x \approx e^{-x} \quad \text{for x small}$$

$$Pr(X=1) = n\frac{1}{m}\left(1 - \frac{1}{m}\right)^{n-1} \approx \frac{n}{m}e^{-(\frac{n}{m})} = \lambda e^{-\lambda}$$

Siméon Poisson, 1781-1840

Suppose "events" happen, independently, at an *average* rate of λ per unit time. Let X be the *actual* number of events happening in a given time unit. Then X is a *Poisson* r.v. *with parameter* λ (denoted X ~ Poi(λ)) and has distribution (PMF):

$$P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}$$

Examples:

\# of alpha particles emitted by a lump of radium in 1 sec.

\# of traffic accidents in Seattle in one year

\# of roadkill per mile on a highway.

\# of white blood cells in a blood suspension

Poisson random variables

X is a Poisson r.v. with parameter λ if it has PMF:

$$P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}$$

Is it a valid distribution? Recall Taylor series:

$$e^{\lambda} = \frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \cdots = \sum_{0 \le i} \frac{\lambda^i}{i!}$$

So

$$\sum_{0 \le i} P(X = i) = \sum_{0 \le i} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \sum_{0 \le i} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1$$

Poisson approximates binomial when n is large, p is small, and $\lambda = np$ is "moderate"

Formally, Binomial is Poisson in the limit as $n \to \infty$ (equivalently, $p \to 0$) while holding $np = \lambda$

X ~ Binomial(n,p)

$$
\begin{aligned}
P(X=i) &= \binom{n}{i} p^i (1-p)^{n-i} \\
&= \frac{n!}{i!(n-i)!}\left(\frac{\lambda}{n}\right)^i \left(1-\frac{\lambda}{n}\right)^{n-i}, \text{ where } \lambda = pn \\
&= \frac{n(n-1)\cdots(n-i+1)}{n^i}\,\frac{\lambda^i}{i!}\,\frac{(1-\lambda/n)^n}{(1-\lambda/n)^i} \\
&= \underbrace{\frac{n(n-1)\cdots(n-i+1)}{(n-\lambda)^i}}\,\frac{\lambda^i}{i!}\,\underbrace{(1-\lambda/n)^n} \\
&\approx \quad 1 \cdot \frac{\lambda^i}{i!} \cdot e^{-\lambda}
\end{aligned}
$$

I.e., Binomial ≈ Poisson for large n, small p, moderate i, λ.

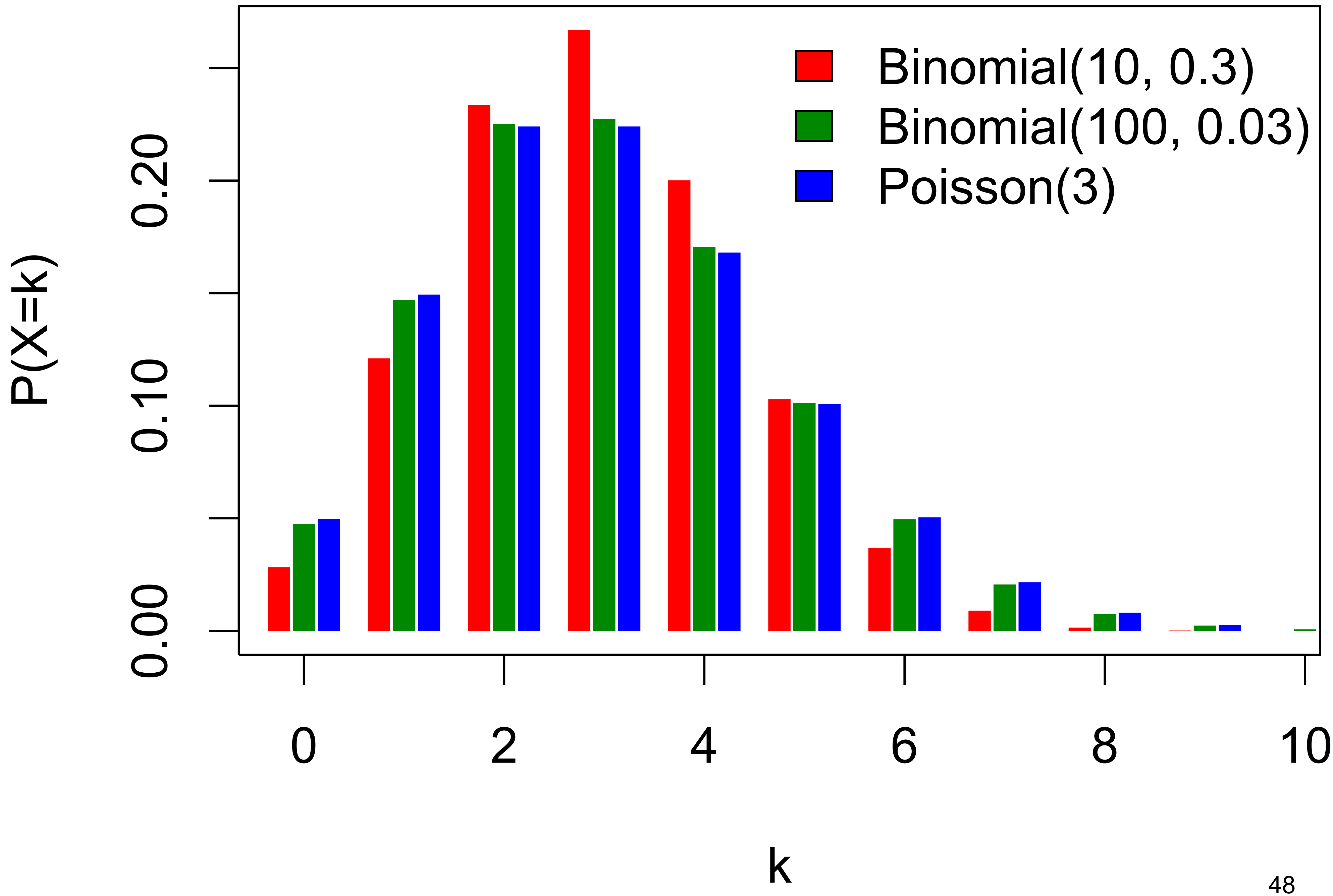
Recall example of sending bit string over a network

- Send bit string of length $n = 10^4$
- Probability of (independent) bit corruption is $p = 10^{-6}$
- $X \sim \text{Poi}(\lambda = 10^4 \cdot 10^{-6} = 0.01)$
- What is probability that message arrives uncorrupted?

$$P(X = 0) = e^{-\lambda}\frac{\lambda^0}{0!} = e^{-0.01}\frac{0.01^0}{0!} \approx 0.990049834$$

Using $Y \sim \text{Bin}(10^4, 10^{-6})$:

- $P(Y=0) \approx 0.990049829$


Binomial(10, 0.3)
Binomial(100, 0.03)
Poisson(3)
P(X=k)
0.00
0.10
0.20
0
2
4
6
8
10
k

expected value of Poisson r.v.s

$$
\begin{aligned}
E[X] &= \sum_{0 \le i} i \cdot e^{-\lambda} \frac{\lambda^i}{i!} \\
&= \sum_{1 \le i} i \cdot e^{-\lambda} \frac{\lambda^i}{i!} \\
&= \lambda e^{-\lambda} \sum_{1 \le i} \frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda} \sum_{0 \le j} \frac{\lambda^j}{j!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda
\end{aligned}
$$

i = 0 term is zero

j = i-1

As expected, given definition in terms of "average rate λ"

(Var[X] = λ, too; proof similar, see B&T example 6.20)

Recall: if $Y \sim \text{Bin}(n,p)$, then:

$E[Y] = pn$

$Var[Y] = np(1\text{-}p)$

And if $X \sim \text{Poi}(\lambda)$ where $\lambda = np$ ($n \to \infty$, $p \to 0$) then

$E[X] = \lambda = np = E[Y]$

$Var[X] = \lambda \approx \lambda(1\text{-}\lambda/n) = np(1\text{-}p) = Var[Y]$

Expectation and variance of Poisson are the same (λ)

Expectation is the same as corresponding binomial

Variance almost the same as corresponding binomial

Note: when two different distributions share the same mean & variance, it suggests (but doesn't prove) that one may be a good approximation for the other.

In a series $X_1, X_2, \ldots$ of Bernoulli trials with success probability p, let Y be the index of the first success, i.e.,

$$X_1 = X_2 = \ldots = X_{Y-1} = 0 \;\&\; X_Y = 1$$

Then Y is a *geometric* random variable with parameter p.

Examples:

Number of coin flips until first head

Number of blind guesses on SAT until I get one right

Number of darts thrown until you hit a bullseye

Number of random probes into hash table until empty slot

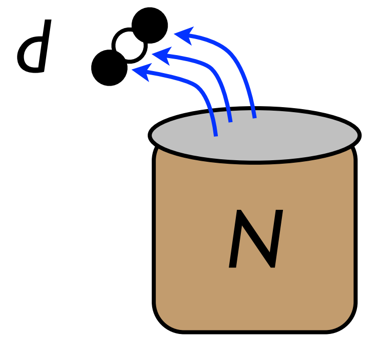
Number of wild guesses at a password until you hit it

$P(Y=k) = (1-p)^{k-1}p$; Mean $1/p$; Variance $(1-p)/p^2$

balls in urns – the hypergeometric distribution

B&T, exercise 1.61

Draw *d* balls (without replacement) from an urn containing *N*, of which *w* are white, the rest black.



Let X = number of white balls drawn

$$P(X = i) = \frac{\binom{w}{i}\binom{N-w}{d-i}}{\binom{N}{d}}, \quad i = 0, 1, \ldots, d$$

(note: n choose k = 0 if k < 0 or k > n)

E[X] = dp, where p = w/N (the fraction of white balls)

proof: Let X_j be 0/1 indicator for j-th ball is white, $X = \Sigma X_j$

The X_j are *dependent*, but $E[X] = E[\Sigma X_j] = \Sigma E[X_j] = dp$

Var[X] = dp(1-p)(1-(d-1)/(N-1))

Supreme Court case: Berghuis v. Smith

If a group is underrepresented in a jury pool, how do you tell?

Justice Breyer [Stanford Alum] opened the questioning by invoking the binomial theorem. He hypothesized a scenario involving "**an urn with a thousand balls, and sixty are red, and nine hundred forty are black, and then you select them at random… twelve at a time.**" According to Justice Breyer and the binomial theorem, if the red balls were black jurors then "**you would expect… something like a third to a half of juries would have at least one black person**" on them.

- Justice Scalia's rejoinder: "We don't have any urns here."

- Should model this combinatorially
  - Ball draws not independent trials (balls not replaced)
- Exact solution:

  P(draw 12 black balls) = $\binom{940}{12} \Big/ \binom{1000}{12} \approx 0.4739$

  P(draw ≥ 1 red ball) = 1 – P(draw 12 black balls) ≈ 0.5261
- Approximation using Binomial distribution
  - Assume P(red ball) constant for every draw = 60/1000
  - X = # red balls drawn. X ~ Bin(12, 60/1000 = 0.06)
  - P(X ≥ 1) = 1 – P(X = 0) ≈ 1 – 0.4759 = 0.5240

In Breyer's description, should actually expect just <u>over half</u> of juries to have at least one black person on them

RV: a numeric function of the outcome of an experiment

Probability Mass Function p(x): prob that RV = x; $\Sigma p(x)=1$

Cumulative Distribution Function F(x): *probability that RV ≤ x*

Expectation:

of a random variable: $E[X] = \Sigma_x xp(x)$

of a function: if $Y = g(X)$, then $E[Y] = \Sigma_x g(x)p(x)$

linearity:

$E[aX + b] = aE[X] + b$

$E[X+Y] = E[X] + E[Y]$; even if *dependent*

this interchange of "order of operations" is quite special to linear combinations. E.g. $E[XY] \neq E[X]*E[Y]$, in general (but see below)

Variance:

$Var[X] = E[(X-E[X])^2] = E[X^2] - (E[X])^2]$

Standard deviation: $\sigma = \sqrt{Var[X]}$

$Var[aX+b] = a^2 Var[X]$

If X & Y are *independent*, then

$E[X \cdot Y] = E[X] \cdot E[Y]$;

$Var[X+Y] = Var[X]+Var[Y]$

(These two equalities hold for *indp* rv's; but not in general.)

Important Examples:

Bernoulli: $P(X=1) = p$ and $P(X=0) = 1-p$ $\mu = p$, $\sigma^2 = p(1-p)$

Binomial: $P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$ $\mu = np$, $\sigma^2 = np(1-p)$

Poisson: $P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}$ $\mu = \lambda$, $\sigma^2 = \lambda$

Bin(n,p) ≈ Poi(λ) where $\lambda = np$ fixed, $n \to \infty$ (and so $p = \lambda/n \to 0$)

Geometric $P(X=k) = (1-p)^{k-1}p$ $\mu = 1/p$, $\sigma^2 = (1-p)/p^2$

Many others, e.g., hypergeometric