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.

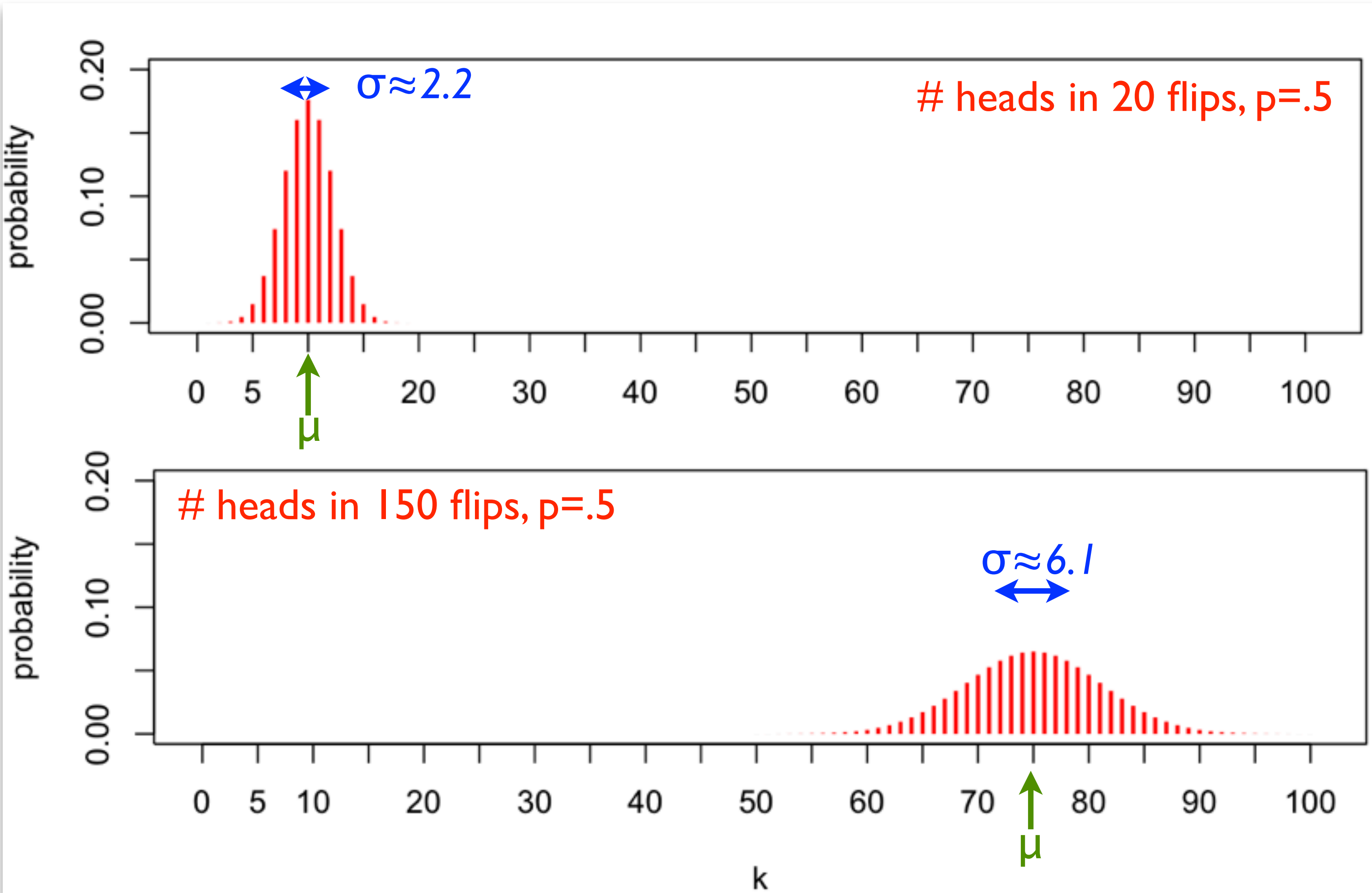
What about its *variability*?

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*; $\sigma = \sqrt{Var(X)}$ is about *spread*



Two games:

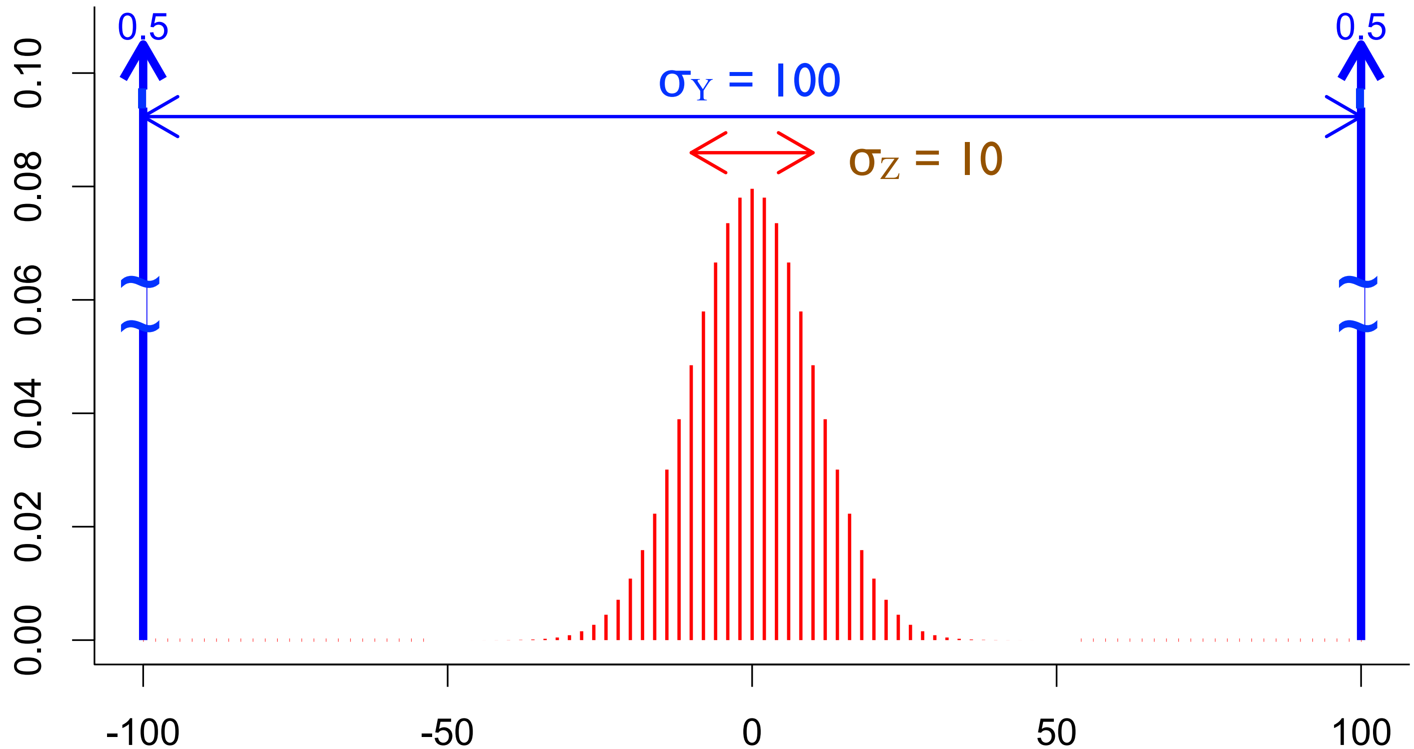
a) flip 1 coin, win Y = \$100 if heads, \$-100 if tails

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

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

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

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

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

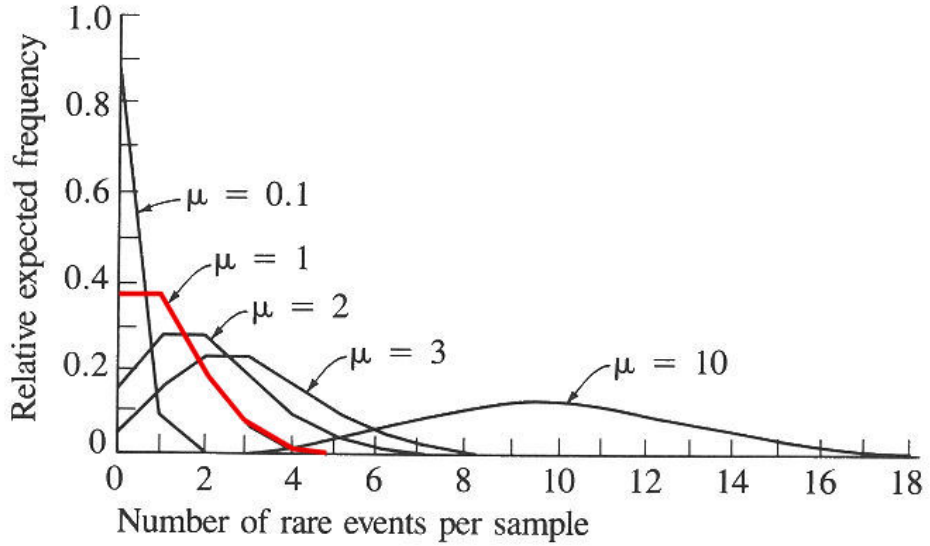
Let $Y = -X$; then $\mathrm{Var}[Y] = (-1)^2\mathrm{Var}[X] = 1$

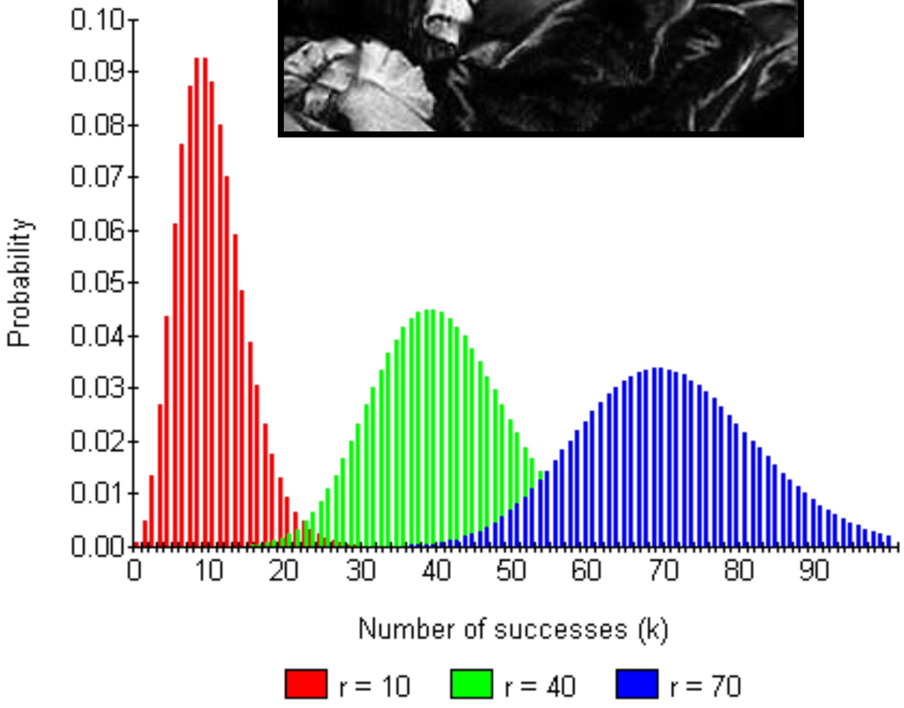
But $X+Y = 0$, always, so $\mathrm{Var}[X+Y] = 0$

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 $P(X=0) = 1-p$

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

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

$Var(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1-p)$

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 ~ 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

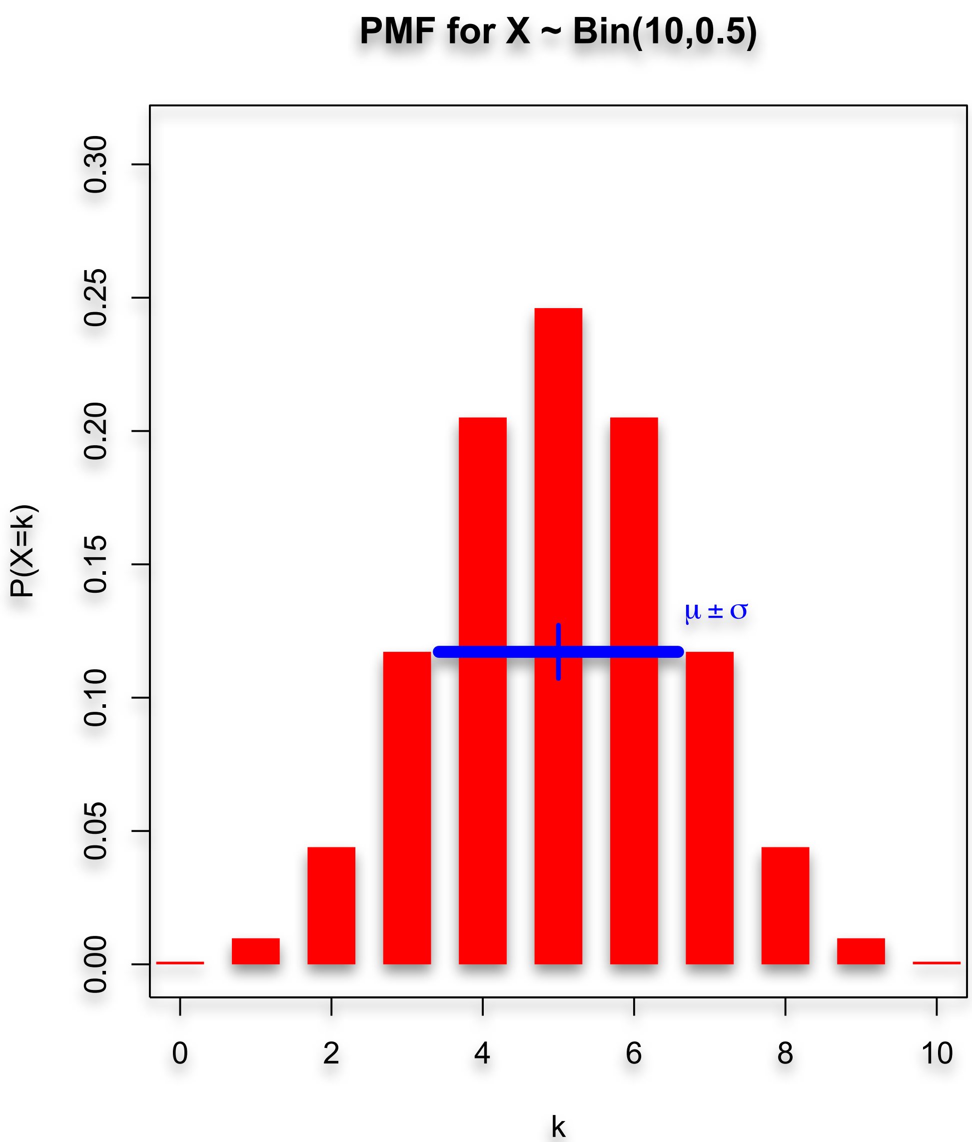
# of disk drive crashes in a 1000 computer cluster

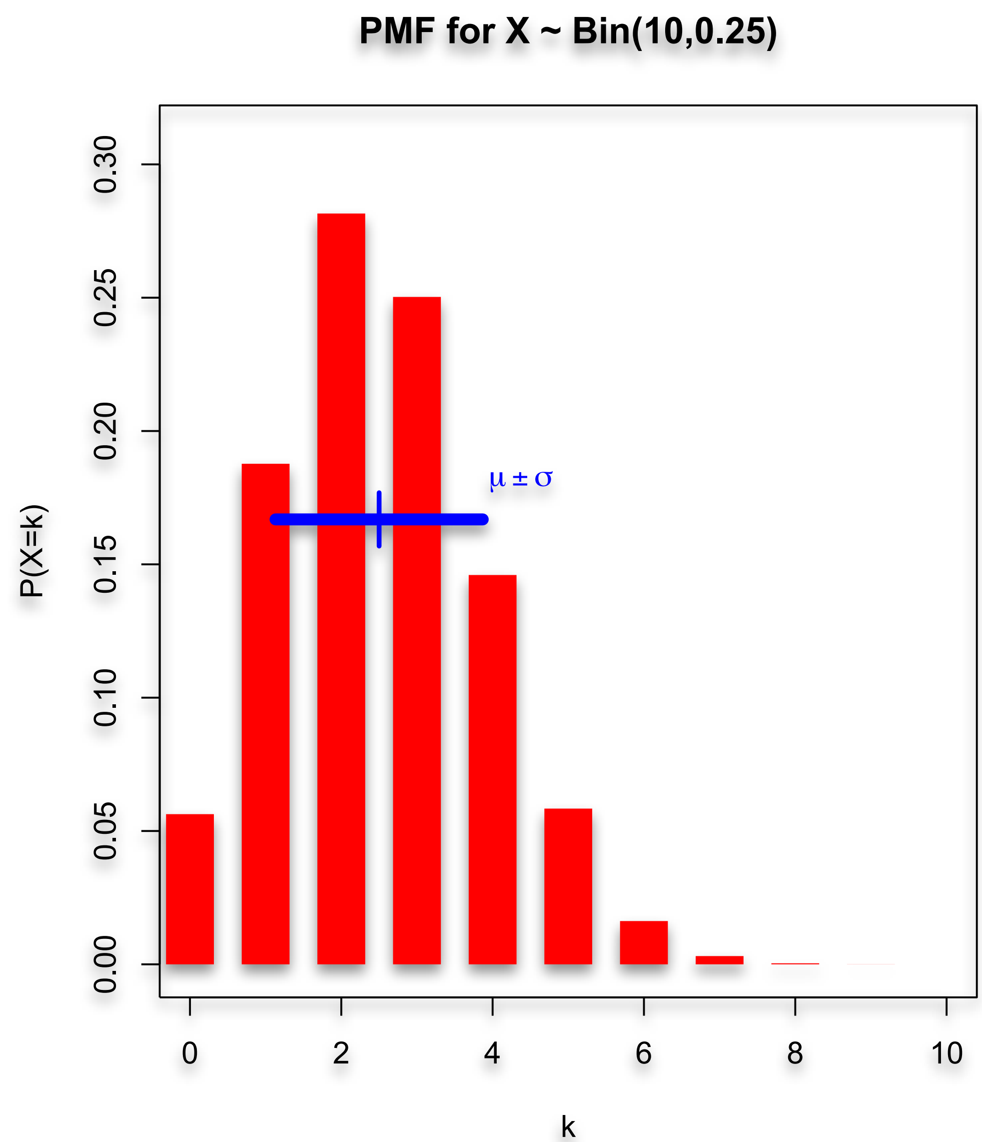
E[X] = pn

Var(X) = p(1-p)n

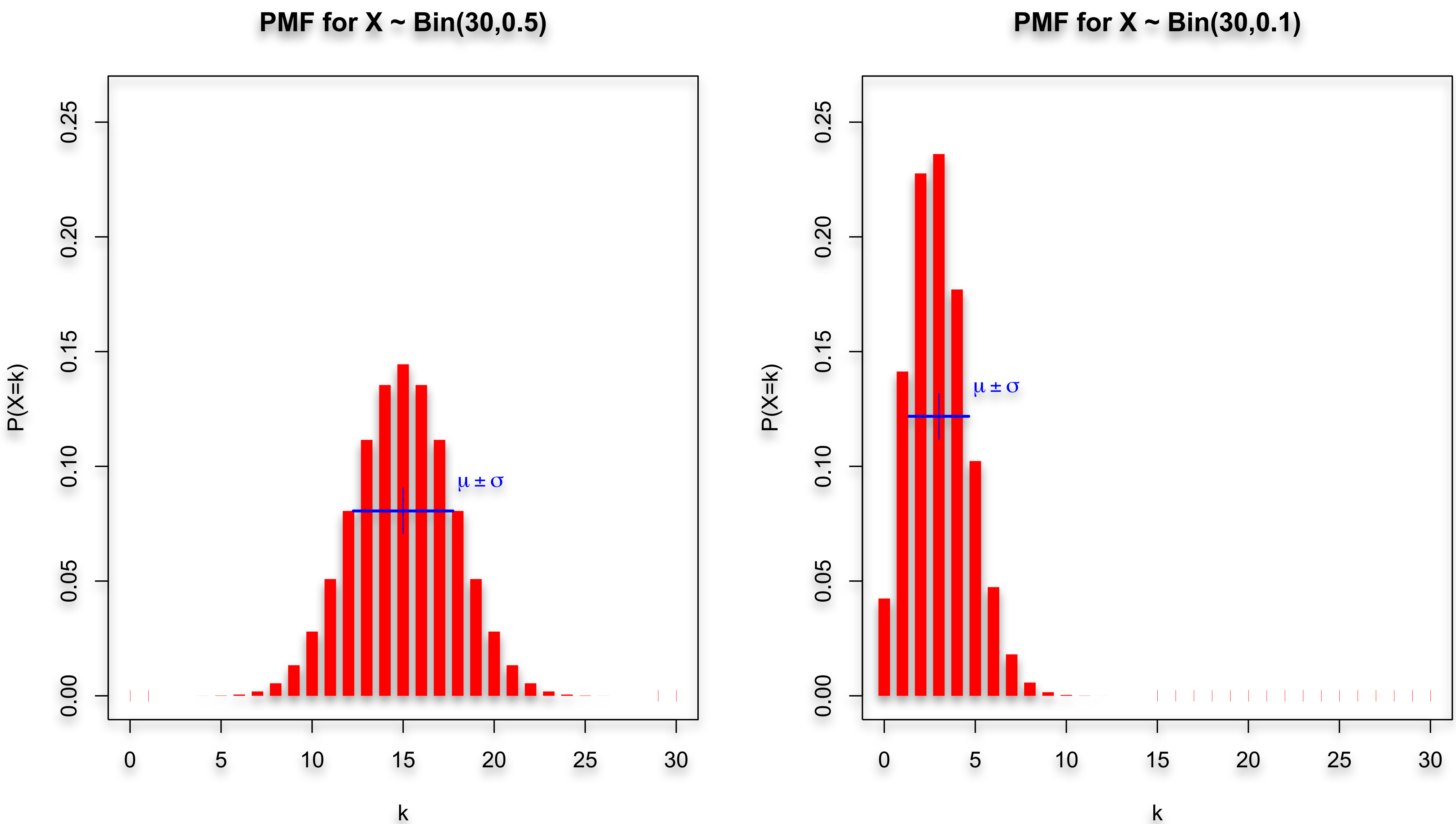
←(proof below, twice)

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
$j = i-1$

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

$$= npE[(Y+1)^{k-1}]$$

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

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