Binomial distribution

Number of successes in n Bernoulli trials

Useful in many applications

μ, V, σ

Plot distribution, generate data

21 heads

19 tails

The Binomial Distribution

n independent Bernoulli experiments

$\overline{p}$, q

Each "success" with same probability p

"failure" with probability 1 - p

$B_{p,n}$ - distribution of # successes

n independent coin flips

P(heads) = p

$B_{p,n}$ - distribution of # heads

$B_{⅓,5}$

$B_{.2,10}$

$B_{n,p}$ more common: $B_{5,⅓}$ $B_{10,.2}$

No confusion: $n \in \mathbb{N}$, $0 \le p \le 1$

Use $B_{p,n}$ because:

- Generalizes B_p
- p main parameter
- Extends to Poisson Binomial

Applications

#

Positive responses to a treatment

Faulty components

Rainy days in a month

Delayed flights

Small n

n independent experiments

Each with success probability p

Failure probability $q = 1 - p$

$b_{p,n}(k)$ - probability of k successes

$0 \le k \le n$

$n = 0$

seq	k	$b_{p,0}(k)$
Λ	0	1

$n = 1$

seq's	k	$b_{p,1}(k)$
0	0	q
1	1	p

$p+q=1$

$n = 2$

seq's	k	$b_{p,2}(k)$
00	0	q^2
01,10	1	2pq
11	2	p^2

Each sequence has probability pq

$p^2+2pq+q^2=(p+q)^2=1^2=1$

General n

$\perp\!\!\!\perp$

n independent B_p experiments

1

k successes

n+1 values

$0 \le k \le n$

$b_{p,n}(k) = p(\text{k successes})$

0

k-successes

n-k failures

Each such sequence has prob. $p^k \cdot q^{n-k}$

$\binom{n}{k}$ such sequences

$$= \binom{n}{k} p^k q^{n-k}$$

ΣWILL IT ADD?

$0 \leq k \leq n$

$p(X = k) = b_{p,n}(k) = \binom{n}{k} p^k q^{n-k}$

$$\sum_{k=0}^{n} b_{p,n}(k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$$

Binomial Theorem

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

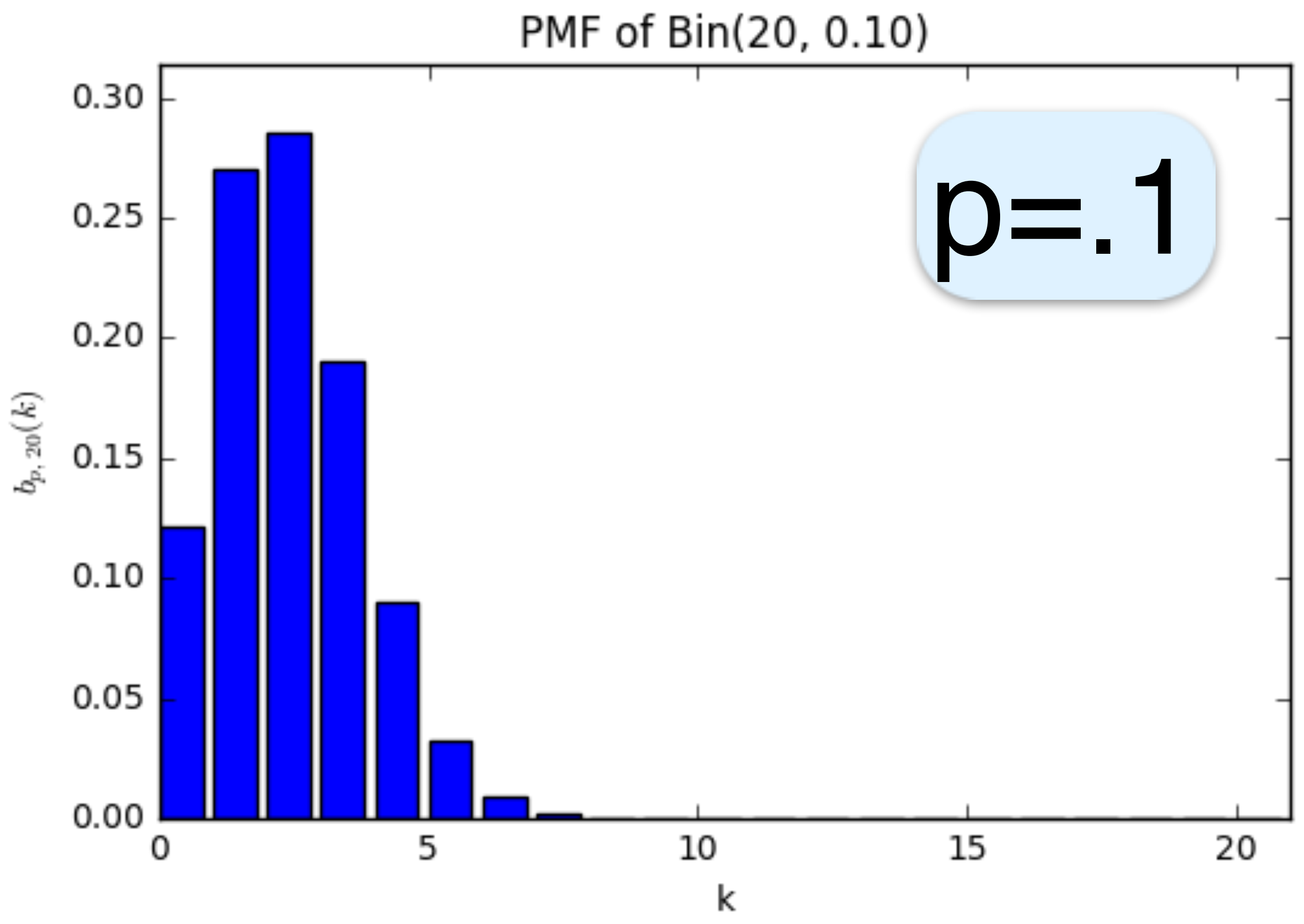
$$= (p+q)^n$$

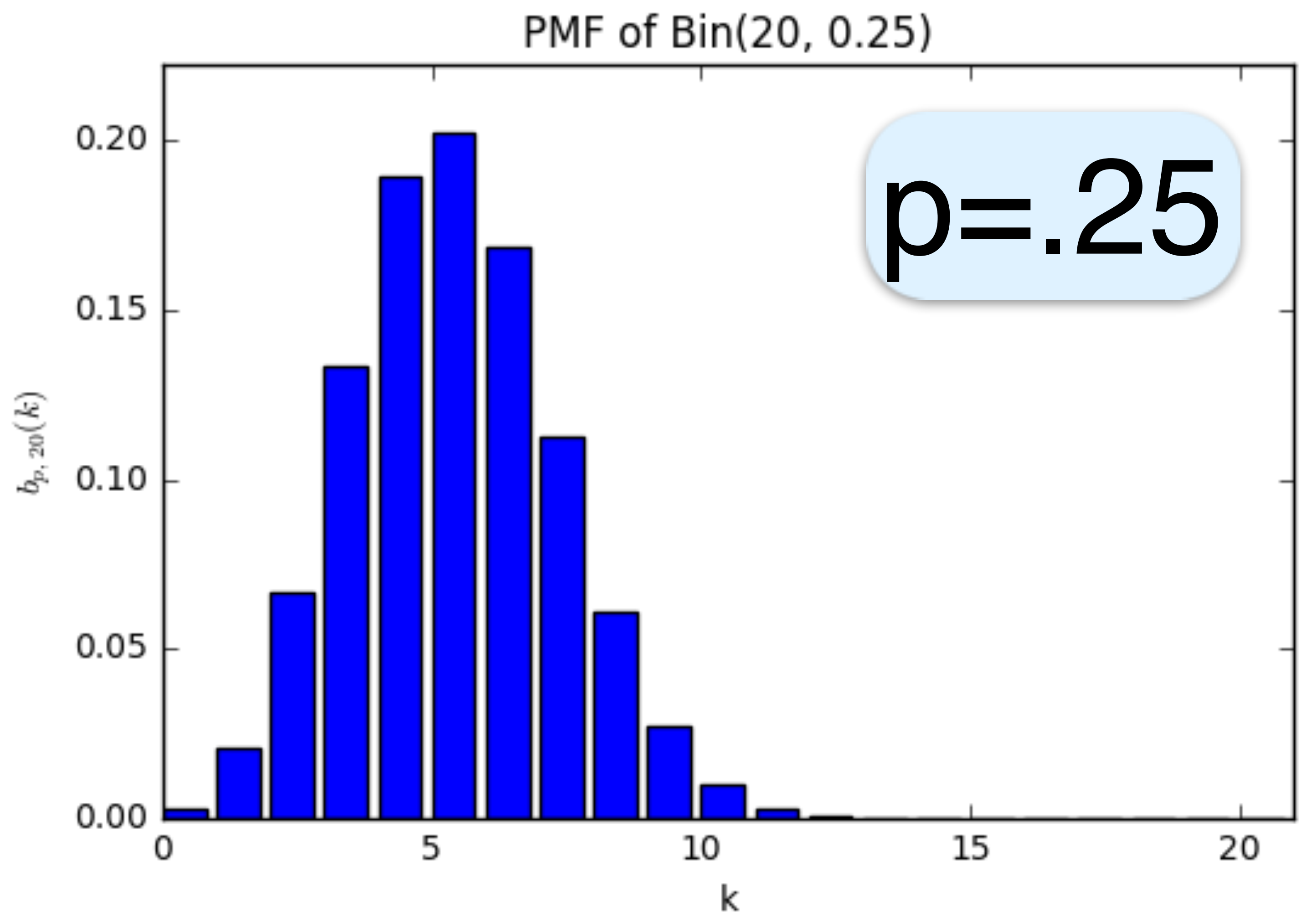
$$= 1^n = 1$$

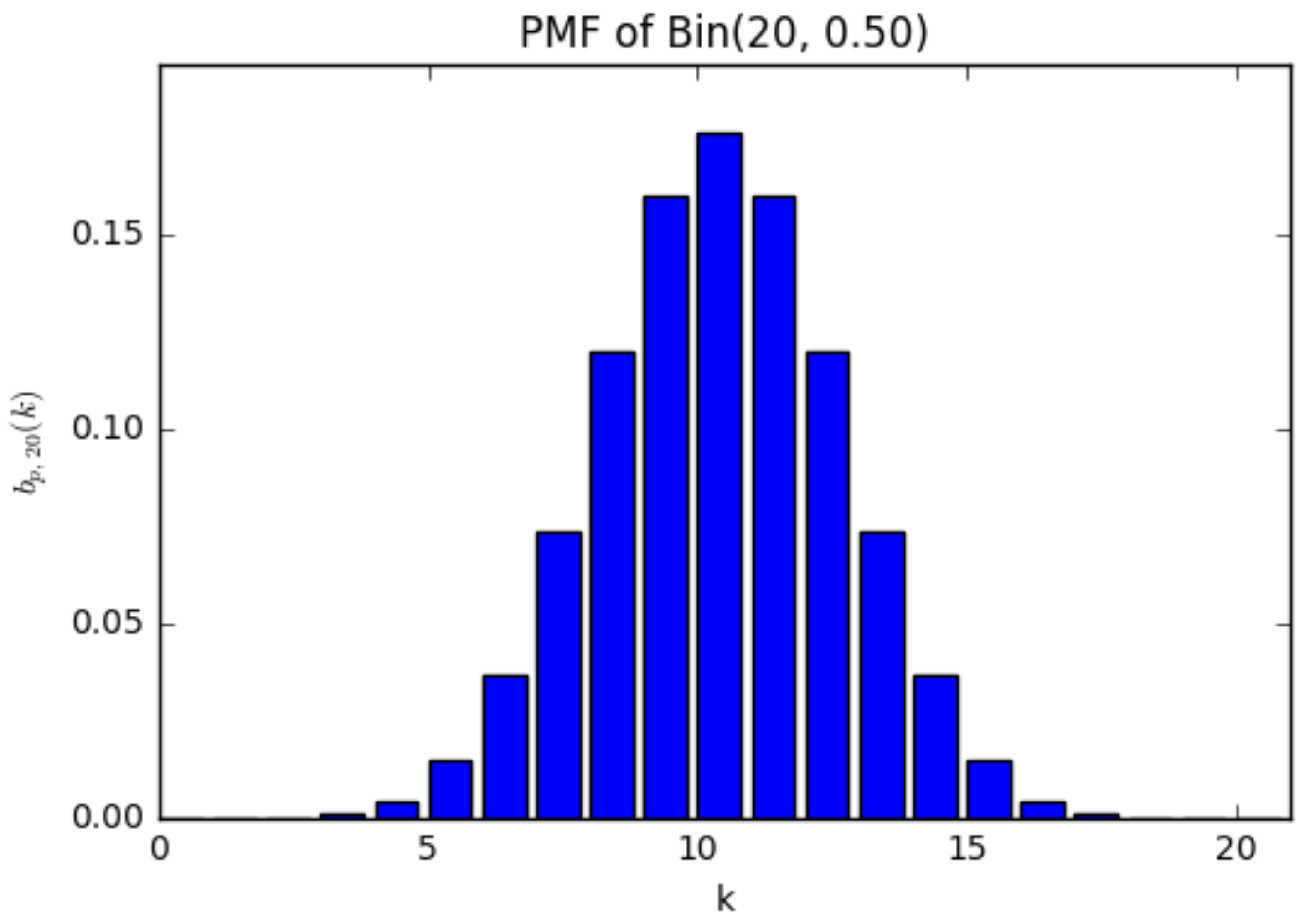
YES IT ADDS!

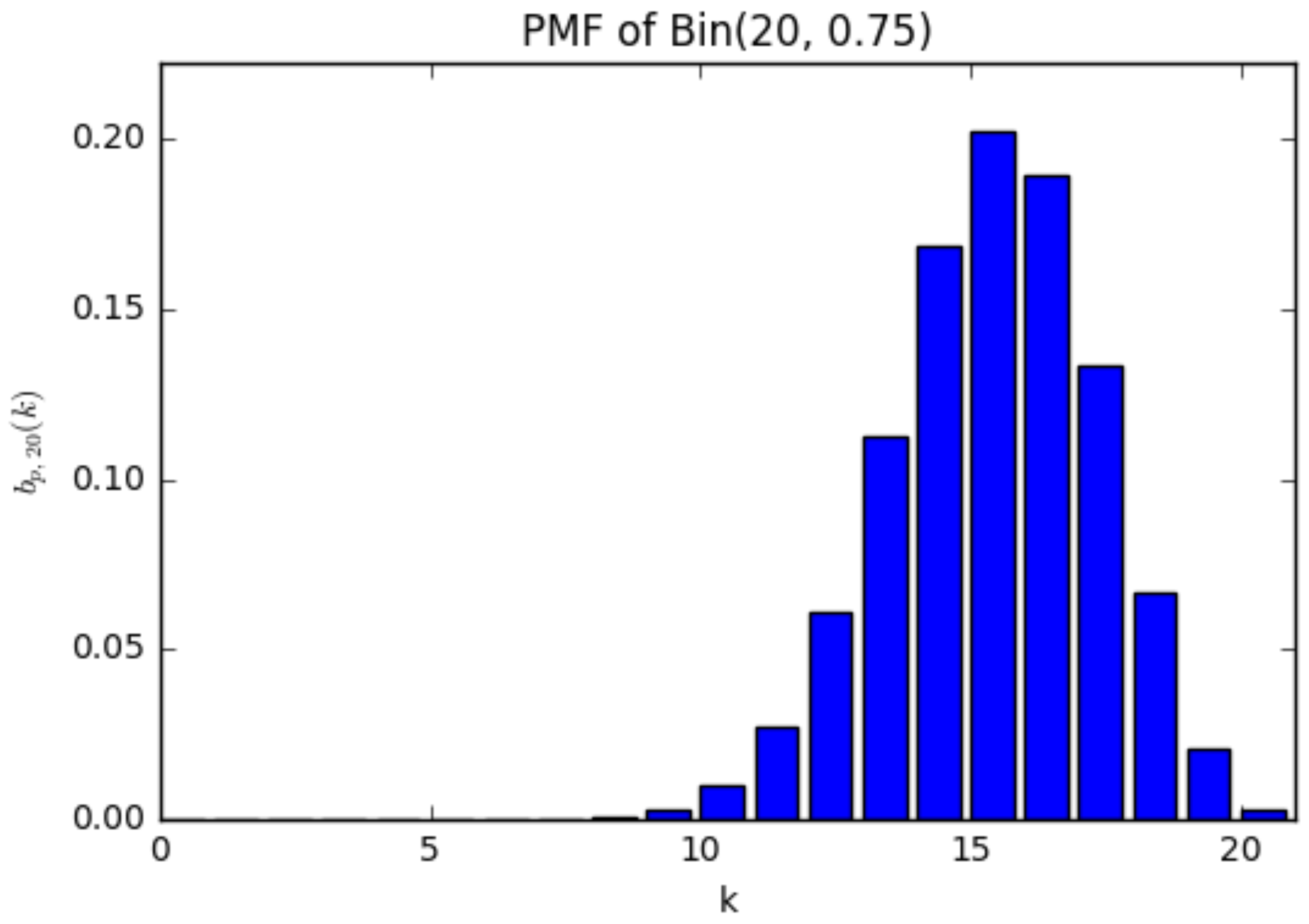
Typical Distributions

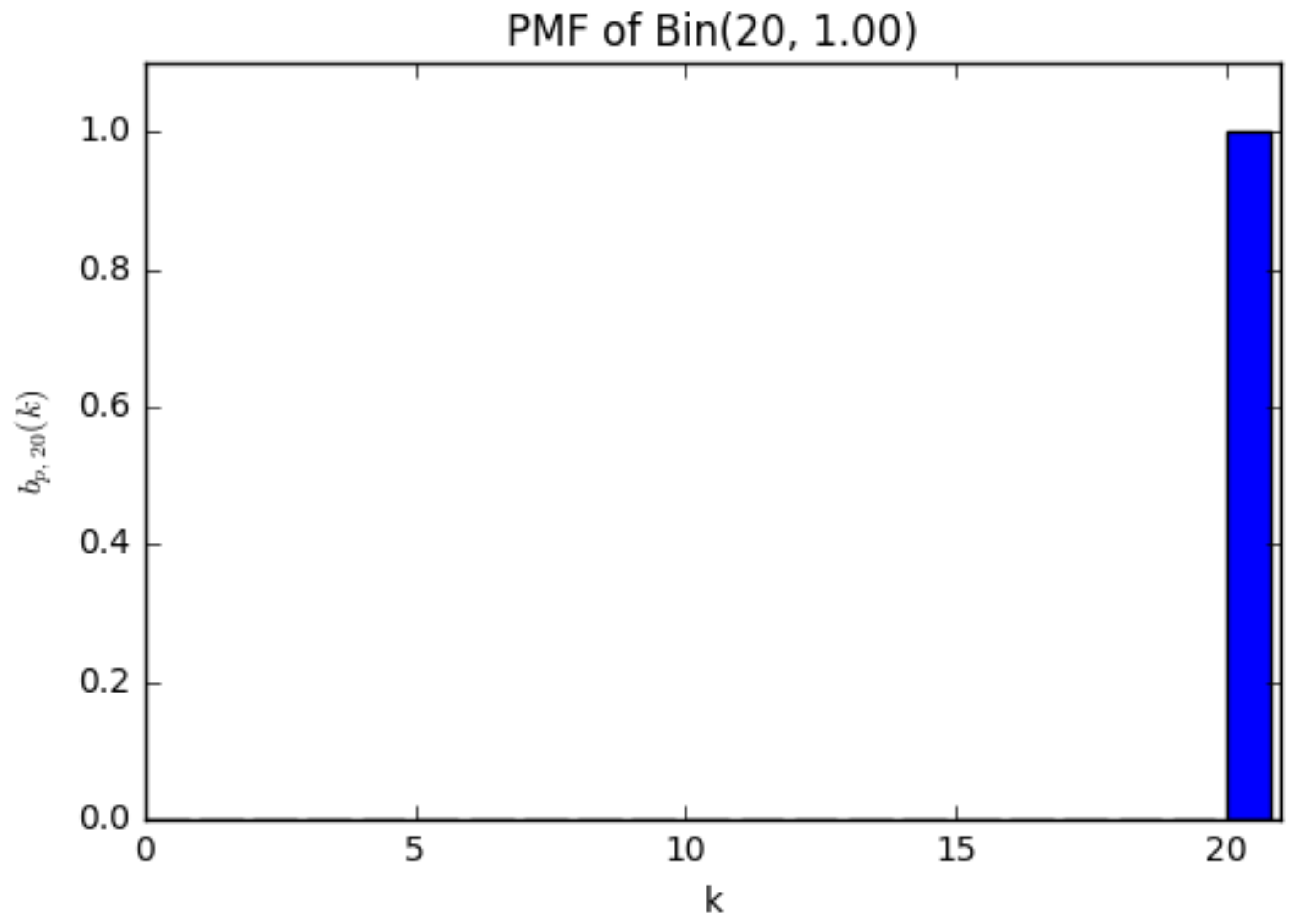
n=20 $b_{p,20}(k)$ Coin P(heads)=p 20 flips P(k heads)











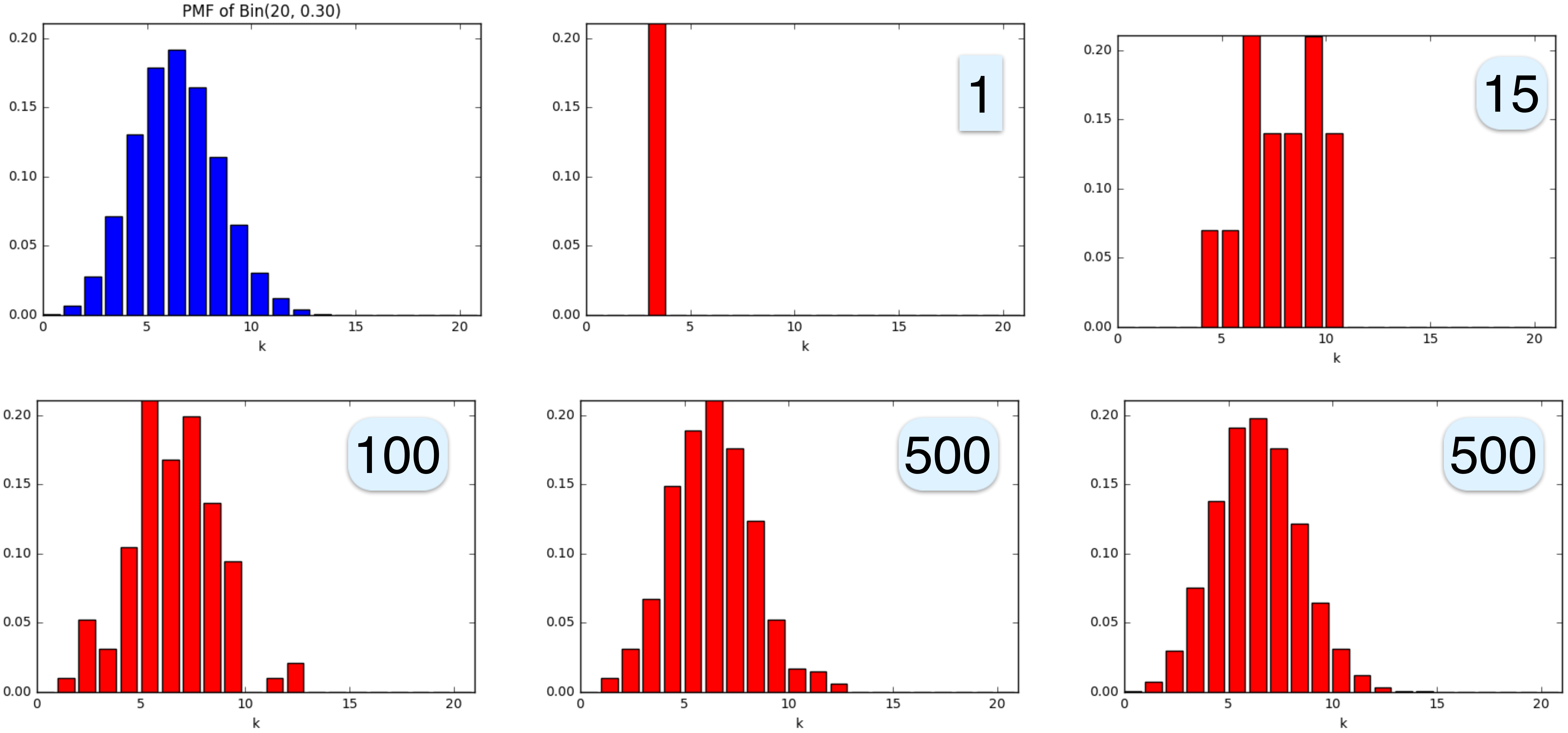
Code

Later

Typical Samples

PMF

Experiments



Multiple Choice

Exam has 6 multiple-choice questions, each with 4 possible answers

Each question, student selects one of the 4 answers randomly

X = # correct answers $\sim B_{¼,6}$

Passing: ≥ 4 correct answers P(passing) = ?

$$P(4) = \binom{6}{4} \cdot (\tfrac{1}{4})^4 \cdot (\tfrac{3}{4})^2 \approx 0.0329$$

$$P(5) = \binom{6}{5} \cdot (\tfrac{1}{4})^5 \cdot (\tfrac{3}{4})^1 \approx 0.00439$$

$$P(6) = \binom{6}{6} \cdot (\tfrac{1}{4})^6 \cdot (\tfrac{3}{4})^0 \approx 0.000244$$

$$P(\geq 4) = P(4) + P(5) + P(6) \approx 0.03759$$

Binomial as a Sum

$B_{p,n}$ a sum of n B_p

$X_1, \ldots, X_n \sim B_p \perp\!\!\!\perp$

$$X \stackrel{\text{def}}{=} \sum_{i=1}^{n} X_i$$

$P(X = k) = P(\text{exactly } k \text{ of } X_1, \ldots, X_n \text{ are } 1) = \binom{n}{k} p^k q^{n-k} = b_{p,n}(k)$

$X \sim B_{p,n}$

Apply to mean and variance

Mean and Variance

$X \sim B_{p,n}$

$$X = \sum_{i=1}^{n} X_i$$

$X_1, \ldots, X_n \sim B_p \perp\!\!\!\perp$

$$E(X) = E(\sum_{i=1}^{n} X_i) \overset{\text{LE}}{=} \sum EX_i \overset{B_p}{=} \sum p = np$$

$$V(X) = V(\sum_{i=1}^{n} X_i) \overset{\perp\!\!\!\perp}{=} \sum V(X_i) \overset{B_p}{=} \sum pq = npq$$

$$\sigma = \sqrt{npq}$$

Multiple Choice

Exam has 6 multiple-choice questions, each with 4 possible answers

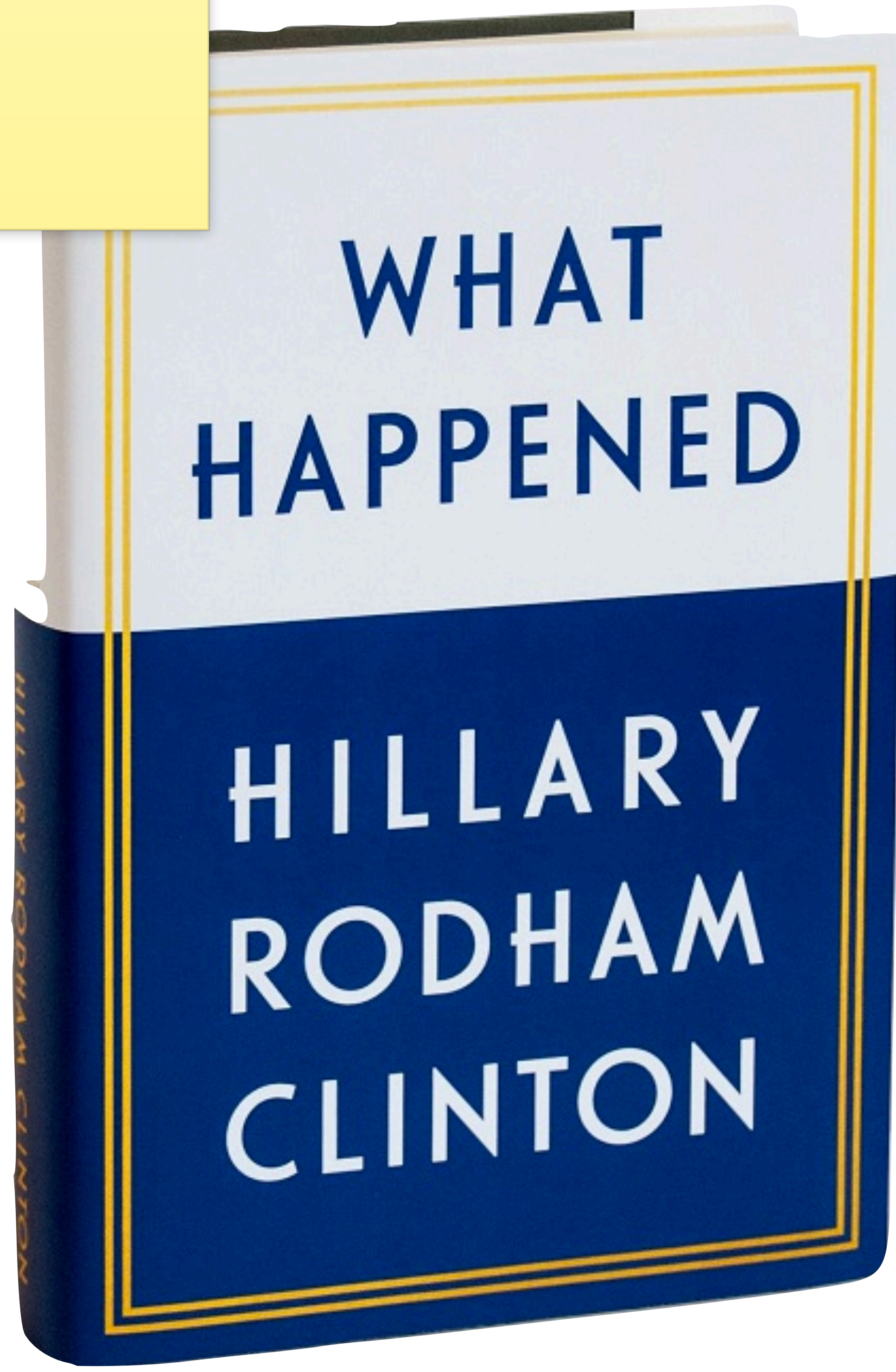
For each question, student selects one of the 4 answers randomly

X = # correct answers $\sim B_{1/4,6}$

Mean $EX = np = 6 \cdot \frac{1}{4} = 1.5$

Standard deviation $\sigma = \sqrt{npq} = \sqrt{6 \cdot \frac{1}{4} \cdot \frac{3}{4}} = \frac{\sqrt{18}}{4}$

Why Vote



For simplicity odd # voters: 2n + 1

Each equally likely D or R

P(voter makes a difference) = P(other 2n voters equally split)

$$b_{p,n}(k) = \binom{n}{k} p^k q^{n-k} = \binom{2n}{n} \frac{1}{2^n} \cdot \frac{1}{2^n}$$

$$= \frac{(2n)!}{n! \cdot n! \cdot 2^n \cdot 2^n}$$

Stirling's Approximation

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

$$\approx \frac{\sqrt{2\pi \cdot 2n}\left(\frac{2n}{e}\right)^{2n}}{\left(\sqrt{2\pi n}\left(\frac{n}{e}\right)^n\right)^2 2^{2n}}$$

$$= \frac{1}{\sqrt{\pi n}}$$

$\gg$ 1/n

Poisson Binomial

Generalizes the binomial distribution

$n \geq 1$	Binomial	$B_{p,n}$	For $1 \leq i \leq n$	$X_i \sim B_p$	$\perp\!\!\!\perp$	$X = \sum_{i=1}^{n} X_i$
	Poisson Binomial	$PB_{p_1, \ldots, p_n}$		$X_i \sim B_{p_i}$		

$PB_{1/4,\, 2/3}$ $X_1 \sim B_{1/4}$ $X_2 \sim B_{2/3}$ $\perp\!\!\!\perp$

X_1	X_2	P	X
0	0	$\frac{3}{4} \cdot \frac{1}{3} = \frac{1}{4}$	0
0	1	$\frac{3}{4} \cdot \frac{2}{3} = \frac{1}{2}$	1
1	0	$\frac{1}{4} \cdot \frac{1}{3} = \frac{1}{12}$	1
1	1	$\frac{1}{4} \cdot \frac{2}{3} = \frac{1}{6}$	2

X	P(x)
0	$\frac{1}{4}$
1	$\frac{7}{12}$
2	$\frac{1}{6}$

Expectation and Variance

$$X \sim PB_{p_1,p_2,\ldots,p_n} \qquad X = \sum_{i=1}^{n} X_i \qquad X_i \sim B_{p_i}$$

$$E(X) = E(\sum_{i=1}^{n} X_i) \overset{\text{LE}}{=} \sum_{i=1}^{n} EX_i \overset{\text{B}_{\text{pi}}}{=} \sum_{i=1}^{n} p_i$$

$$V(X) = V(\sum_{i=1}^{n} X_i) \overset{\perp\!\!\!\perp}{=} \sum_{i=1}^{n} V(X_i) \overset{\text{B}_{\text{pi}}}{=} \sum_{i=1}^{n} p_i(1 - p_i)$$

p(k) | No closed form | Computationally

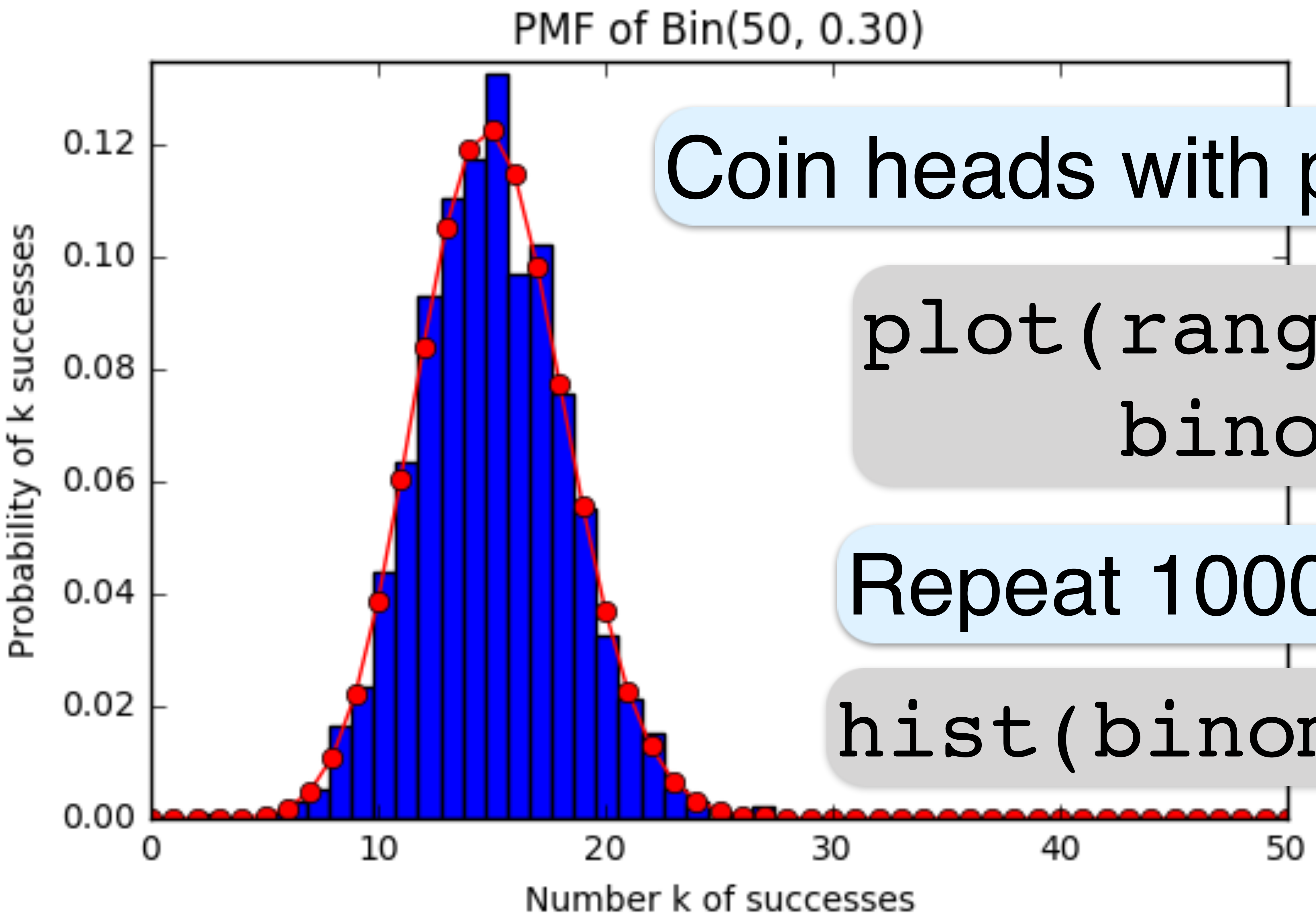
Plot & Sample

Notebook

Plot distribution

Generate data

```
from scipy.stats import binom
from matplotlib.pyplot import plot, hist
```



Coin heads with prob .3, 50 flips

Prob of k heads

$B_{.3,50}$

```
plot(range(51),
     binom.pmf(k=range(51),n=50,p=.3))
```

Repeat 1000 times

Histogram

```
hist(binom.rvs(n=50,p=.3,size=1000))
```

See notebook

Binomial distribution

$B_{p,n}$ Number of successes in n B_p trials

$$b_{p,n}(k) = \binom{n}{k} p^k \overline{p}^{n-k}$$

$\mu = np$

$V = np\bar{p}$

$\sigma = \sqrt{np\overline{p}}$

Plot distribution, generate data

Why vote

Coin Flips

Most basic convergence to average is B(p)

Flip n B(p) coins, average # 1's will approach np

Probability of a sequence with k 1's and n-k 0's is $p^k q^{n-k}$

Wolog assume p>0.5, then most likely is 1^n

Yet by WLLN with probability $\rightarrow$1 we see roughly pn 1's and qn 0's

Why do we observe these sequences and not the most likely ones?

Strength in #s. # sequences of a given composition increases near 1/2

pn balances # x probability.