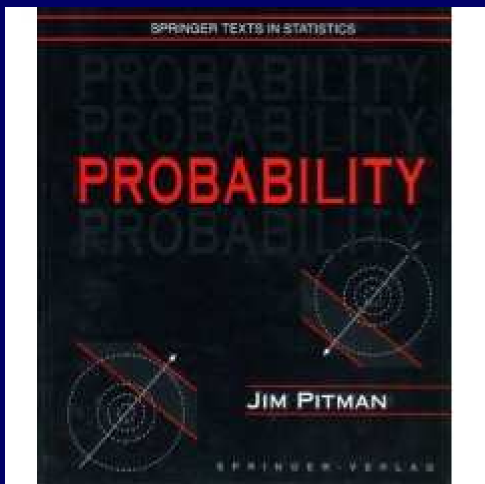
Introduction to probability

Stat 134 FAll 2005

Berkeley



Lectures prepared by:

Elchanan Mossel

Yelena Shvets

Follows Jim Pitman's book:

Probability

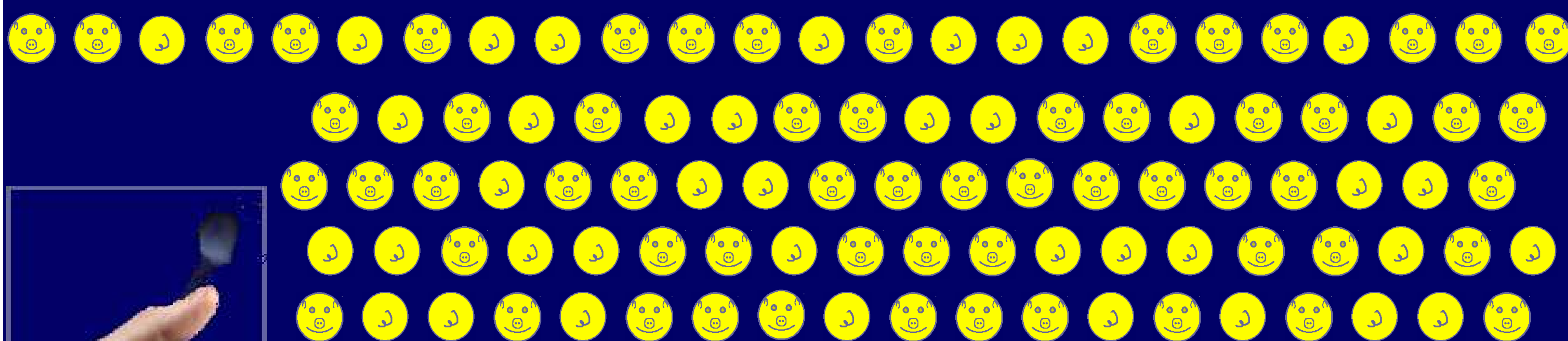
Section 2.1

Toss a coin 100, what's the chance of 60 ?

P(a sequence with 60) =

(# of sequences with 60)*P(particular sequence with 60)=

(# of sequences with 60)* $\frac{1}{2}^{100}$



hidden assumptions:

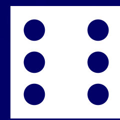
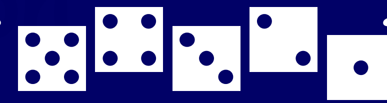
n-independence; probabilities are fixed.

Magic Hat:

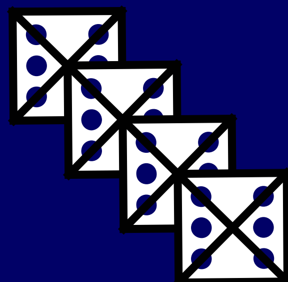
Each time we pull an item out of the hat it magically reappears. What's the chance of drawing 3 I-pods in 10 trials? ={ }

$$\frac{2}{3} \quad \frac{1}{3} \quad \frac{2}{3} \quad \frac{2}{3} \quad \frac{2}{3} \quad \frac{1}{3} \quad \frac{1}{3} \quad \frac{2}{3} \quad \frac{2}{3} \quad \frac{2}{3}$$

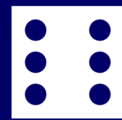
P(3 in 10 draws) = (# of sequences with 3) $(\frac{1}{3})^3(\frac{2}{3})^7$

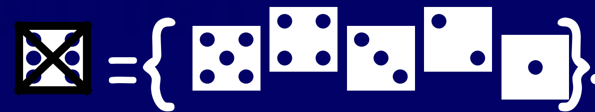
Suppose we roll a die 4 times. What's the chance of k  ? Let  ={  }.

k=0

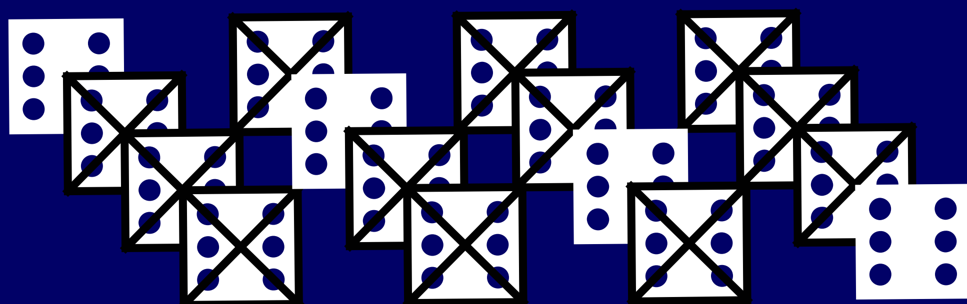


$$\frac{5}{6}\frac{5}{6}\frac{5}{6}\frac{5}{6}*1 = \frac{625}{1296}$$

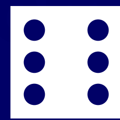
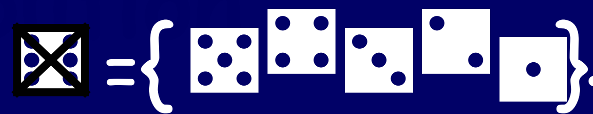
Suppose we roll a die 4 times. What's the chance of k  ?

Let .

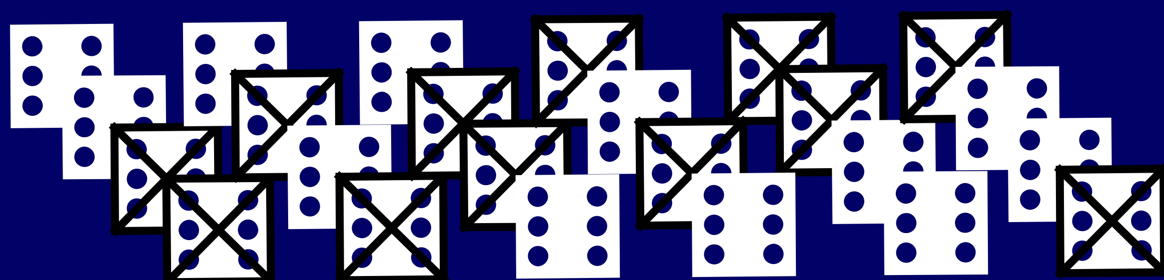
k=1



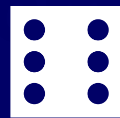
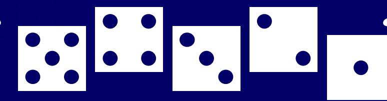
$$\frac{5}{6}\frac{5}{6}\frac{5}{6}\frac{1}{6}*4 = \frac{500}{1296}$$

Suppose we roll a die 4 times. What's the chance of k  ? Let .

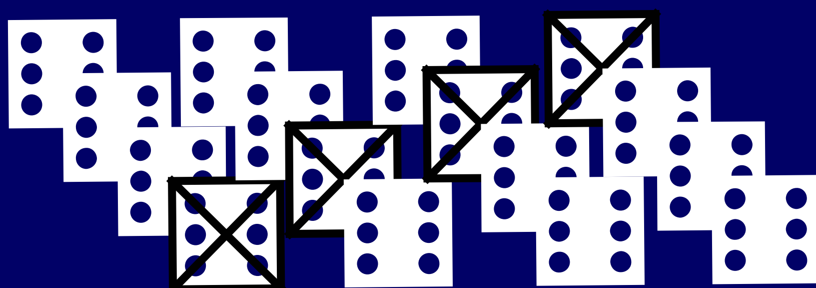
k=2



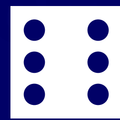
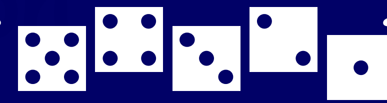
$$\frac{5}{6}\frac{1}{6}\frac{1}{6}\frac{5}{6}*6 = \frac{150}{1296}$$

Suppose we roll a die 4 times. What's the chance of k  ? Let  ={  }.

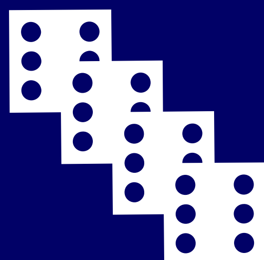
k=3



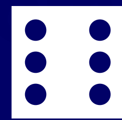
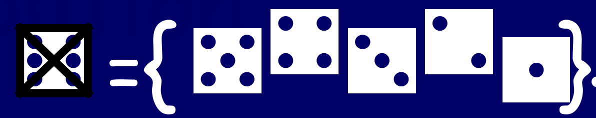
$$\frac{5}{6}\frac{1}{6}\frac{1}{6}\frac{1}{6}*4 = \frac{20}{1296}$$

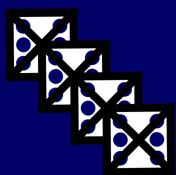
Suppose we roll a die 4 times. What's the chance of k  ? Let  ={  }.

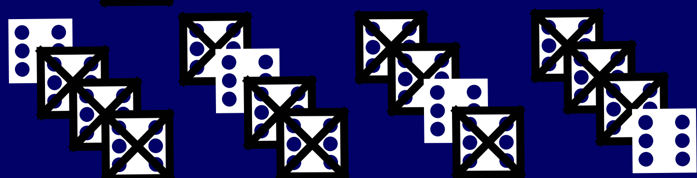
k=4

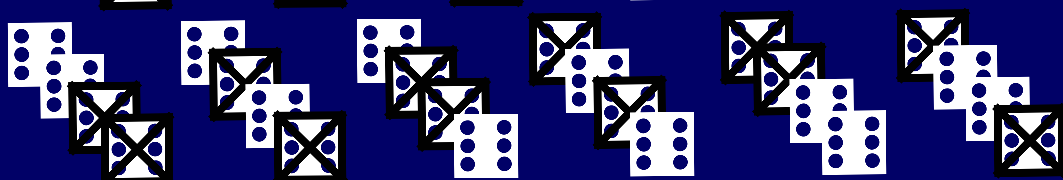


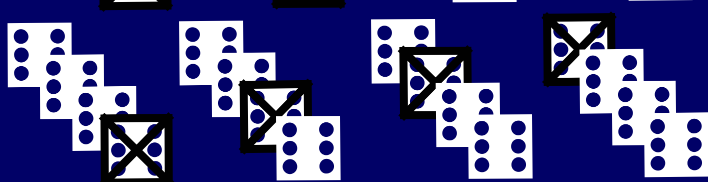
$$\frac{1}{6}\frac{1}{6}\frac{1}{6}\frac{1}{6}*1 = \frac{1}{1296}$$

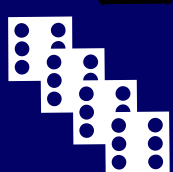
Suppose we roll a die 4 times. What's the chance of k  ? Let .

k=0  $\frac{5}{6}\frac{5}{6}\frac{5}{6}\frac{5}{6} * 1 = \frac{625}{1296}$

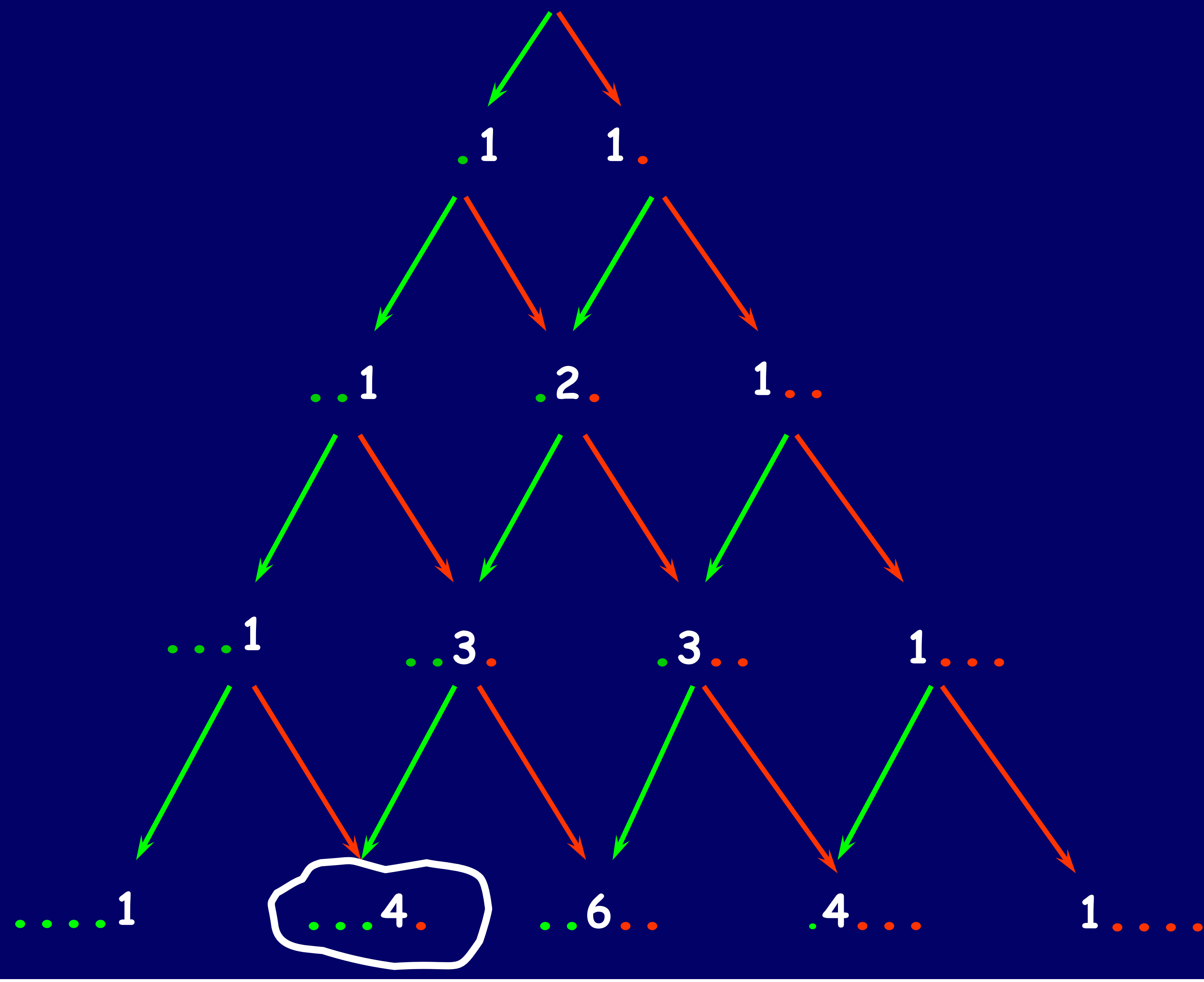
k=1  $\frac{5}{6}\frac{5}{6}\frac{5}{6}\frac{1}{6} * 4 = \frac{500}{1296}$

k=2  $\frac{5}{6}\frac{1}{6}\frac{1}{6}\frac{5}{6} * 6 = \frac{150}{1296}$

k=3  $\frac{5}{6}\frac{1}{6}\frac{1}{6}\frac{1}{6} * 4 = \frac{20}{1296}$

k=4  $\frac{1}{6}\frac{1}{6}\frac{1}{6}\frac{1}{6} * 1 = \frac{1}{1296}$

How do we count the number of sequences of length 4 with 3 . and 1 .?



Binomial Distribution

This is the Pascal's triangle, which gives, as you may recall the binomial coefficients.

Binomial Distribution

$$(a + b) = 1a \quad 1b$$

$$(a + b)^2 = 1a^2 \quad 2ab \quad 1b^2$$

$$(a + b)^3 = 1a^3 \quad 3a^2b \quad 3ab^2 \quad 1b^3$$

$$(a + b)^4 = 1a^4 \quad 4a^3b \quad 6a^2b^2 \quad 4ab^3 \quad 1b^4$$

Newton's Binomial Theorem

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

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Binomial Distribution

For *n* independent trials each with probability *p* of success and *(1-p)* of failure we have

$$P(\#successes=k) = \binom{n}{k} p^k (1-p)^{(n-k)}$$

This defines the *binomial(n,p)* distribution over the set of n+1 integers {0,1,…,n}.

Binomial Distribution

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

This represents the chance that in n draws there was some number of successes between zero and n.

A pair of coins will be tossed 5 times.
Find the probability of getting 
on k of the tosses, k = 0 to 5.

P() $= \frac{1}{4}$;

P() $= \frac{3}{4}$;

binomial(5,1/4)

A pair of coins will be tossed 5 times.
Find the probability of getting 
on k of the tosses, $k = 0$ to 5.

$P(\#$  $= k) = \binom{5}{k}\left(\frac{1}{4}\right)^{k}\left(\frac{3}{4}\right)^{5-k}$

To fill out the distribution table we could compute 6 quantities for k = 0,1,…5 separately, or use a trick.

Binomial Distributions

Consecutive odds ratio relates P(k) and P(k-1).

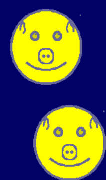
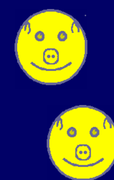
$$\frac{P(k)}{P(k-1)} = \frac{n!}{(k)!(n-k)!} \Big/ \frac{n!}{(k-1)!(n-k+1)!} \frac{p^{k+1}q^{n-k-1}}{p^{k}q^{n-k}}$$

$$= \frac{n-k+1}{k}\frac{p}{q}$$

Use consecutive odds ratio $\frac{P(k)}{P(k-1)} = \frac{n-k+1}{k}\frac{p}{q}$ to quickly fill out the distribution table for binomial(5,1/4):

k	0	1	2	3	4	5
$\frac{P(k)}{P(k-1)}$		$\frac{5}{1}\frac{1}{3}$	$\frac{4}{2}\frac{1}{3}$	$\frac{3}{3}\frac{1}{3}$	$\frac{2}{4}\frac{1}{3}$	$\frac{1}{5}\frac{1}{3}$
P(k)	$\left(\frac{3}{4}\right)^5$ .237	$P(0)\frac{5}{1}\frac{1}{3}$ .340	$P(1)\frac{4}{2}\frac{1}{3}$ .264	$P(2)\frac{3}{3}\frac{1}{3}$ .0879	$P(3)\frac{2}{4}\frac{1}{3}$ .0146	$P(4)\frac{1}{5}\frac{1}{3}$ .000977

We can use this table to find the following conditional probability:

P(at least 3  | at least 1  in first 2 tosses)

= P(3 or more  & 1 or 2  in first 2 tosses)

P(1 or 2  in first 2) bin(5,1/4):

k	0	1	2	3	4	5
$\frac{P(k)}{P(k-1)}$		$\frac{5}{1}\frac{1}{3}$	$\frac{4}{2}\frac{1}{3}$	$\frac{3}{3}\frac{1}{3}$	$\frac{2}{4}\frac{1}{3}$	$\frac{1}{5}\frac{1}{3}$
$P(k)$	.237	.340	.264	.0879	.0146	.000977

bin(2,1/4):

k	0	1	2
$\frac{P(k)}{P(k-1)}$		$\frac{2}{1}\frac{1}{3}$	$\frac{1}{2}\frac{1}{3}$
$P(k)$	.563 $\left(\frac{3}{4}\right)^2$	.375 $P(0)\frac{5}{1}\frac{1}{3}$	.0625 $P(1)\frac{4}{2}\frac{1}{3}$

How useful is the binomial formula?

Try using your calculators to compute P(500 H in 1000 coin tosses) directly:

$$P(500 \text{ in } 1000) = \frac{1000!}{500!500!}(\frac{1}{2})^{1000}$$

Your calculator may return error when computing 1000!. This number is just too big to be stored.

The following is called

the Stirling's approximation.

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

It is not very useful if applied directly : n^n is a very big number if n is 1000.

Stirling's formula: $n! \approx \sqrt{2\pi n}\,(\frac{n}{e})^n$

$$P(\text{500 in 1000}) = \frac{1000!}{500!500!}\frac{1}{2}^{1000}$$

$$P(\text{500 in 1000}) \approx \frac{1}{\sqrt{2\pi}}\sqrt{\frac{1000}{500*500}}\;\frac{(\frac{1000}{e})^{1000}}{(\frac{500}{e})^{500}(\frac{500}{e})^{500}}\frac{1}{2}^{1000}$$

$$P(\text{500 in 1000}) \approx \frac{1}{\sqrt{2\pi}}\sqrt{\frac{2}{500}}\;(\frac{1000}{2*500})^{1000}$$

$$P(\text{500 in 1000}) \approx \frac{1}{\sqrt{500\pi}}$$

$$P(\text{500 in 1000}) \approx .0252313252$$

Binomial Distribution

Toss coin 1000 times;

P(500 in 1000|250 in first 500)=

P(500 in 1000 & 250 in first 500)/P(250 in first 500)=

P(250 in first 500 & 250 in second 500)/P(250 in first 500)=

P(250 in first 500) P(250 in second 500)/ P(250 in first 500)=

$$P(250 \text{ in second } 500) = \frac{500!}{250!250!}\frac{1}{2}^{500} = \frac{1}{\sqrt{250\pi}} = 0.356824823$$

Binomial Distribution: Mean

Question: For a fair coin with p = ½, what do we expect in 100 tosses?

Binomial Distribution: Mean

Recall the frequency interpretation:

$p \approx$ #H/#Trials

So we expect about 50 H !

Expected value or Mean (μ) of a binomial(n,p) distribution

μ = #Trials $\times$ P(success)

= $n\,p$

Binomial Distribution: Mode

Question:

What is the most likely number of successes?

Mean seems a good guess.

Binomial Distribution: Mode

Recall that

$$P(50 \text{ in } 100) \approx \frac{1}{\sqrt{50\pi}} \approx .0797884561$$

To see whether this is the most likely number of successes we need to compare this to P(k in 100) for every other k.

Binomial Distributions
Mode

The most likely number of successes is called the **mode** of a binomial distribution.

Binomial Distribution: Mode

If we can show that for some m

$P(1) \leq \ldots P(m-1) \leq P(m) \geq P(m+1) \ldots \geq P(n)$,

then m would be the mode.

Binomial Distribution: Mode

$$P(k) \le P(l) \iff \frac{P(k)}{P(l)} \le 1,$$

so we can use successive odds ratio:

$$\frac{P(k)}{P(k-1)} = \frac{n-k+1}{k}\frac{p}{q},$$

to determine the mode.

successive odds ratio: $\frac{P(k)}{P(k-1)} = \frac{n-k+1}{k}\frac{p}{1-p},$

$$P(k-1) > P(k) \iff \frac{k}{n-k+1}\frac{1-p}{p} > 1,$$

$$P(k-1) > P(k) \iff k(1-p) > (n-k+1)p$$

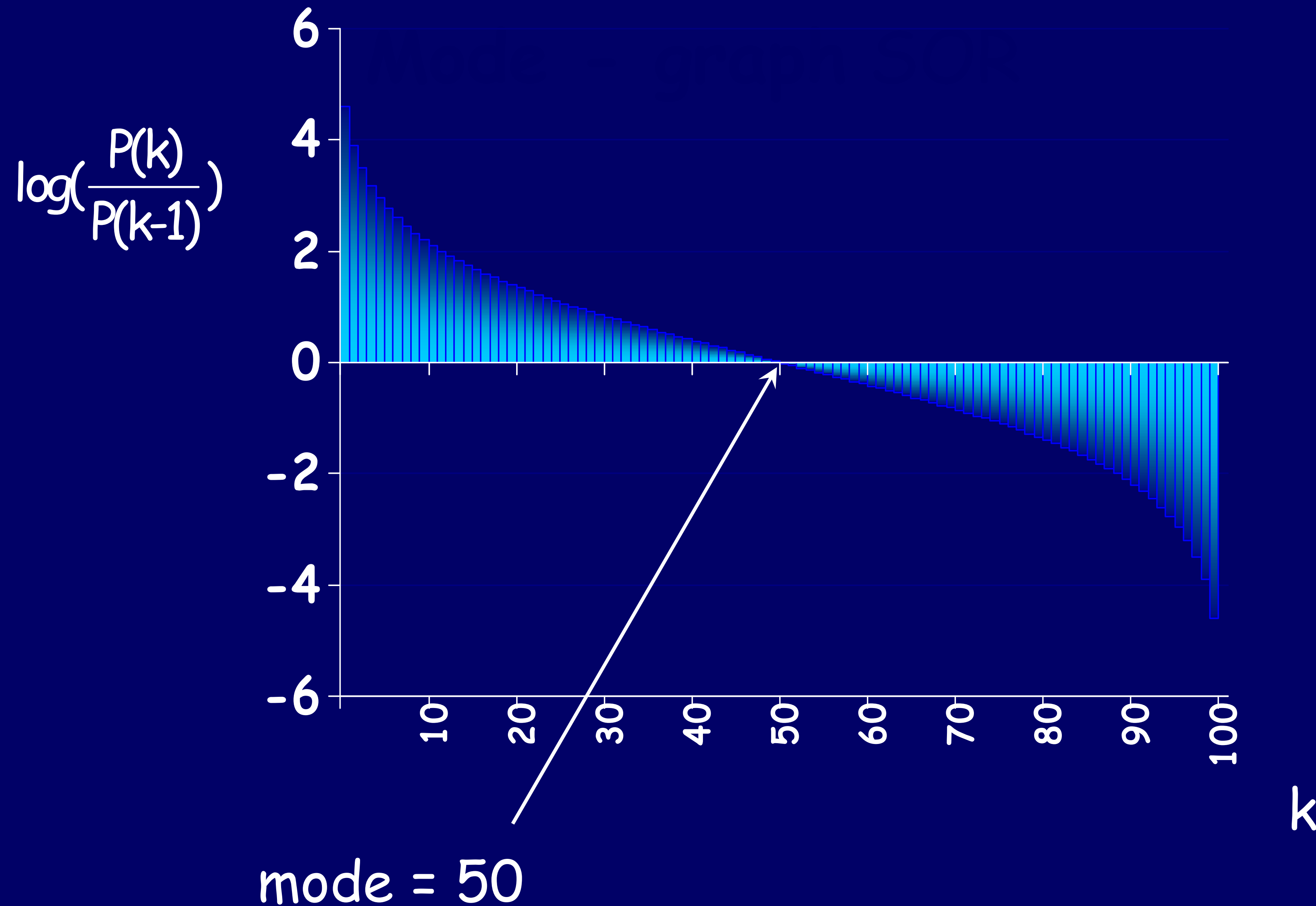
$$P(k-1) > P(k) \iff k > np+p;$$

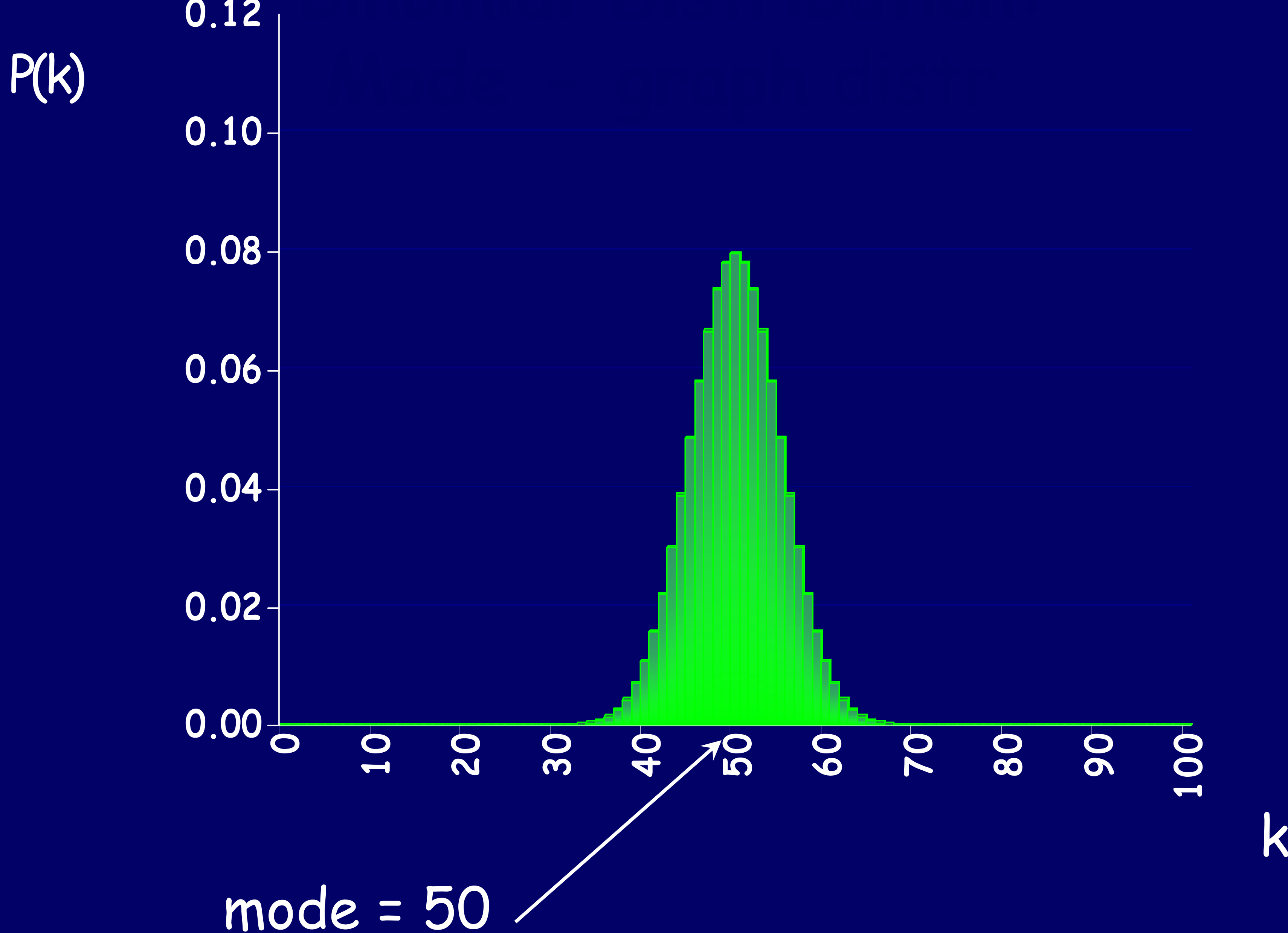
If we replace $\leq$ with $>$ the implications will still hold.

So for $m = \lfloor np+p \rfloor$ we get that

$$P(m-1) \leq P(m) > P(m+1).$$

Binomial Distribution Mode – graph SOR

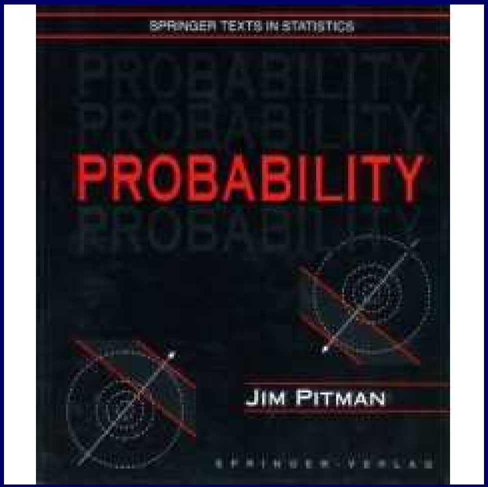



P(k)
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
10
20
30
40
50
60
70
80
90
100
k
mode = 50

Introduction to probability

Stat 134 FAll 2005

Berkeley



Follows Jim Pitman's book:

Probability

Section 2.1

Lectures prepared by:
Elchanan Mossel
Yelena Shvets