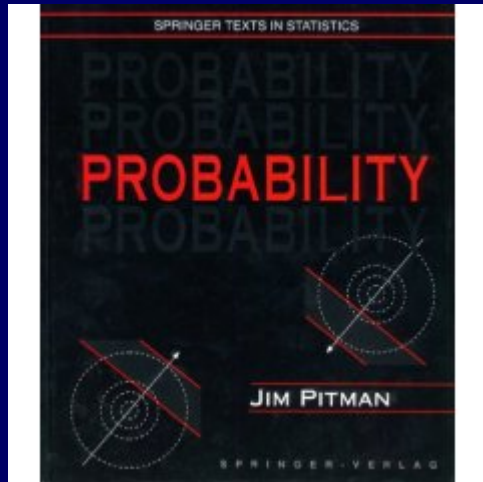


# Introduction to probability

Stat 134

FALL 2005

Berkeley



Lectures prepared by:  
Elchanan Mossel  
Yelena Shvets

Follows Jim Pitman's  
book:

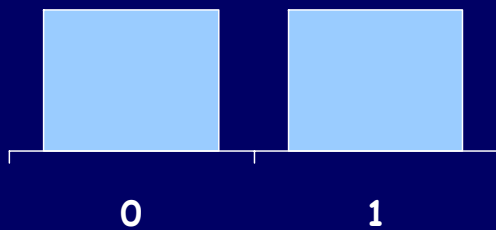
Probability  
Section 3.4

# Different types of Distributions

- **Finite distribution** take only finitely many values.
- Examples of **finite** distributions:

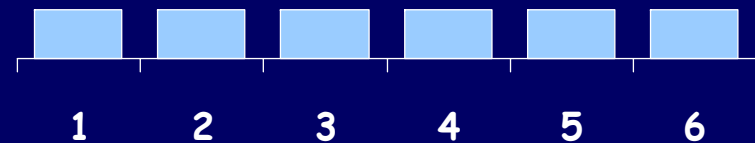
Bernoulli(1/2)

$P(0)=1/2, P(1)=1/2.$



Uniform on {1,2,3,4,5,6}

$P(i) = 1/6.$



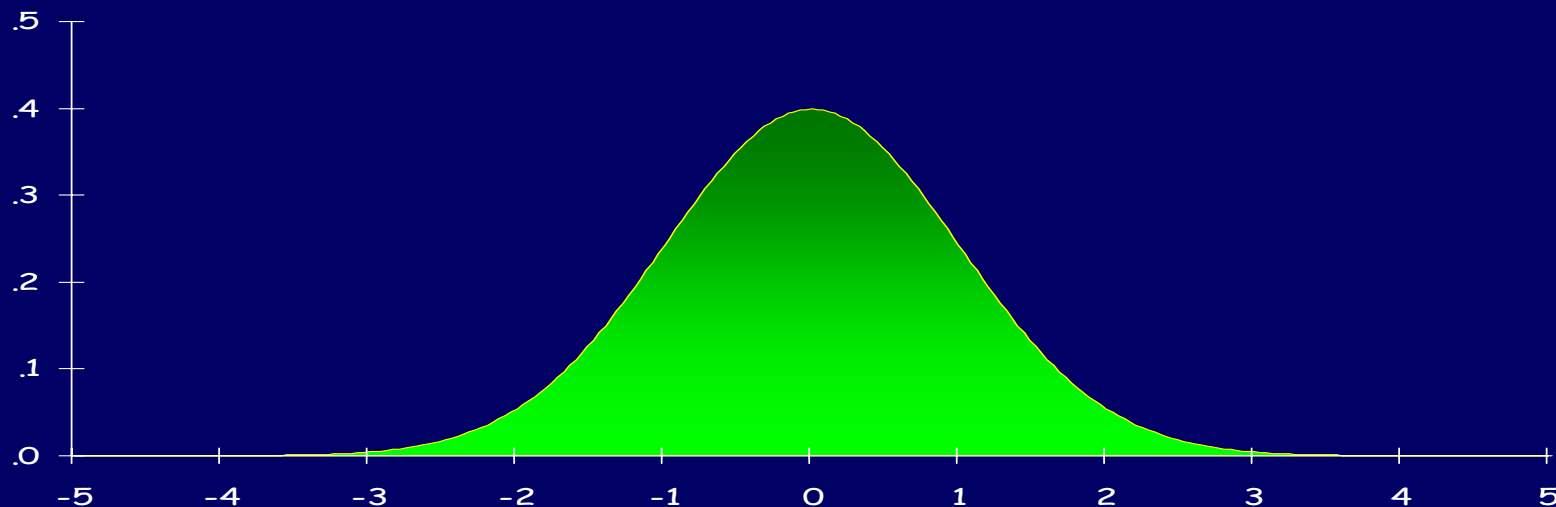
Number of heads in a coin flip.

Numbers on a fair die.

# Infinite Continuous Distributions

Example:  
Normal(0,1);

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$



We'll discuss continuous distributions in the 2<sup>nd</sup> half of the course.

# Infinite Discrete Distributions:

• An **infinite discrete** distribution on  $\{0,1,2,\dots\}$  is given by a sequence  $p_0, p_1, p_2$  of probabilities s.t  $\sum_{i=0}^{\infty} p_i = 1$  and  $p_i \geq 0$  for all  $i$ .

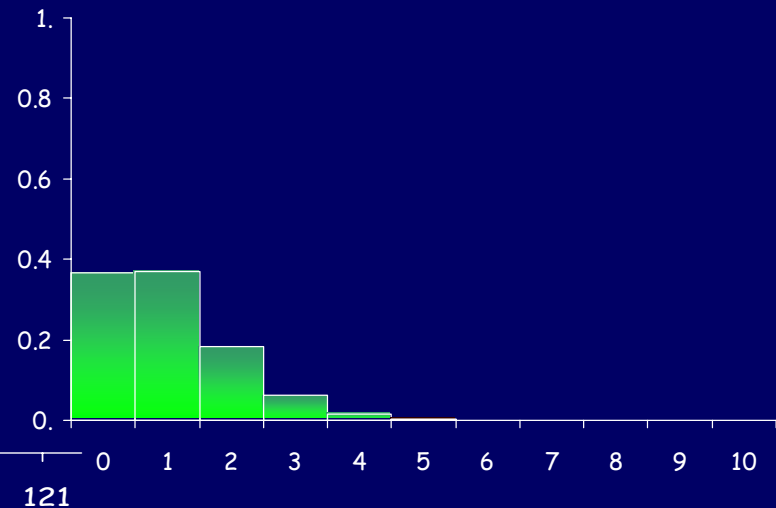
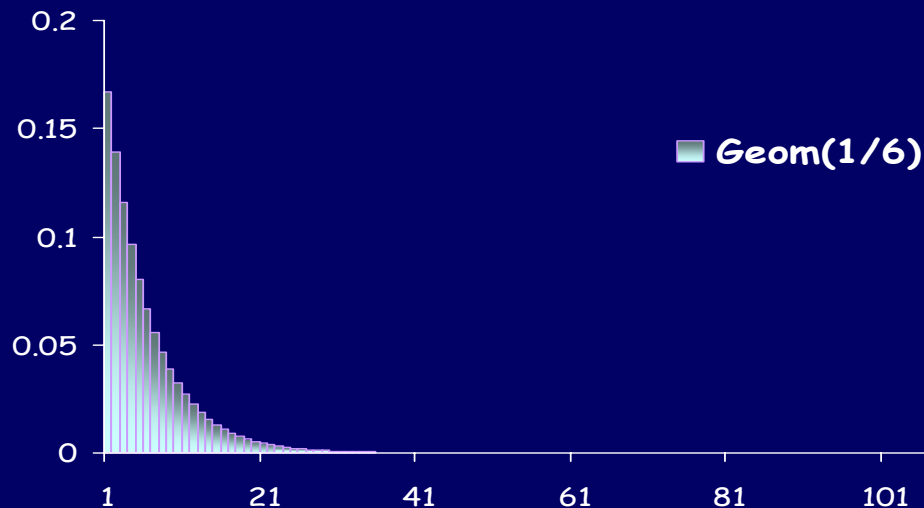
• Examples of **infinite discrete** distribution:

Geometric ( $p = 1/6$ )

$$P(n) = (5/6)^{n-1} 1/6$$

Poisson ( $\mu = 1$ );

$$P(n) = e^{-1}/n!$$



# Infinite Discrete Distributions:

• An **infinite discrete** distribution on  $\{0,1,2,\dots\}$  is given by a sequence  $p_0, p_1, p_2$  of probabilities s.t  $\sum_{i=1}^{\infty} p_i = 1$  and  $p_i \geq 0$  for all  $i$ .

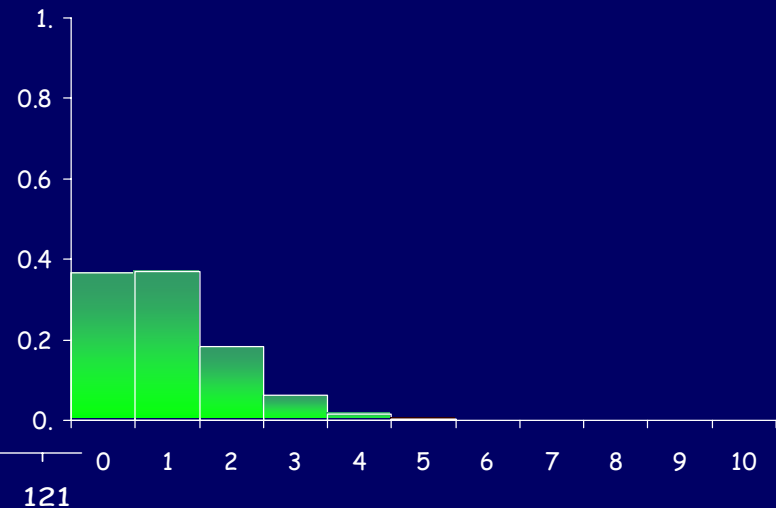
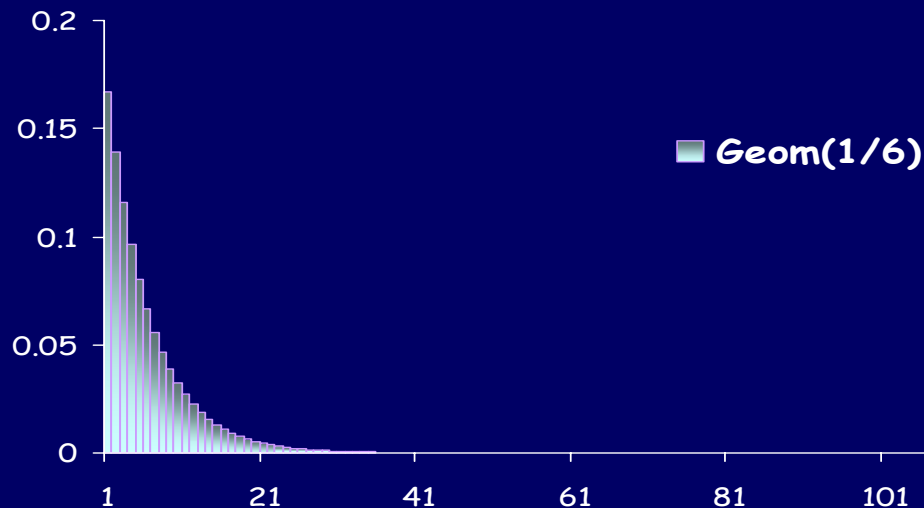
• Examples of **infinite discrete** distribution:

Geometric ( $p = 1/6$ )

$$P(n) = (5/6)^{n-1} 1/6$$

Poisson ( $\mu = 1$ );

$$P(n) = e^{-1}/n!$$



# Infinite Sum Rule

The infinite sum rule: If the event  $A$  is partitioned into  $A_1, A_2, A_3, \dots$

- $A = A_1 \cup A_2 \cup A_3 \cup \dots$  and

- $A_i \cap A_j = \emptyset$  for all  $i \neq j$  then

- $P(A) = P(A_1) + P(A_2) + P(A_3) + \dots$

- Ex: Let  $T = \#$  rolls needed to produce a 6

- What is the probability that  $T > 10$ ?

- What is the probability that  $T$  is even?

- What is the probability that  $T$  is finite?

# Infinite Sum Rule

• Let  $A$  be the event {six is rolled sometime}.

Then  $A$  can be partitioned into mutually exclusive sub-events:

$A_i = \{\text{the first six is on the } i^{\text{th}} \text{ roll}\}$

$$P(A_i) = \left(\frac{5}{6}\right)^{i-1} \frac{1}{6}$$

$$P(A) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} \left(\frac{5}{6}\right)^{i-1} \frac{1}{6} = \frac{1}{6} \sum_{i=0}^{\infty} \left(\frac{5}{6}\right)^i$$

$$P(A) = \frac{1}{6} \frac{1}{1 - \frac{5}{6}} = 1$$

$$P(\text{waiting forever}) = 1 - P(A) = 0$$

# Repeated games

- Faust wins each game with probability  $P(F)$ .
- Mephistopheles wins with probability  $P(M)$ .
- They draw with probability  $P(D)$ .
- They play until there is no draw.
- What is the probability Faust wins overall?

$P(F \text{ ultimately wins}) =$

$$\sum_{i=1}^{\infty} P(\text{draws for } i-1 \text{ games and } F \text{ wins } i^{\text{th}} \text{ game})$$

$$P(F \text{ wins}) = \sum_{i=1}^{\infty} (P(D))^{i-1} P(F)$$

$$= P(F) \sum_{i=0}^{\infty} (P(D))^i$$

$$= P(F) \frac{1}{1-P(D)} = \frac{P(F)}{P(F)+P(M)}$$



# Repeated games

• Claim: The number of games  $G$  they play has geometric distribution.

Let  $p = 1 - P(D)$ .

$$\begin{aligned} P(G=n) &= P(n-1 \text{ draws and the } n^{\text{th}} \text{ is not a draw}) \\ &= (1-p)^{n-1} p. \end{aligned}$$

So  $G$  has a  $\text{Geom}(p)$  distribution.

• Claim: Let  $X$  be the RV that is the identity of the overall winner. Then  $X$  and  $G$  are independent.

$$\begin{aligned} P(G=n \ \& \ X=\text{Faust}) &= P(n-1 \text{ draws} \ \& \ \text{Faust wins the } n^{\text{th}}) \\ &= (P(D))^{n-1} P(F) \\ &= (P(D))^{n-1} (1-P(D)) \frac{P(F)}{(1-P(D))} \\ &= P(G=n) P(X=\text{Faust}) \end{aligned}$$



# Expectation of a Discrete Random Variable

The expectation of the random variable  $X$  is defined by

$$E(X) = \sum x P(X=x)$$

provided that the  $\sum x$  series is absolutely convergent:

$$\sum_{x=0}^{\infty} |x| P(X=x) < \infty$$

## Expected number of Repeated games

Recall that the number  $G$  of games has  $\text{Geom}(1-P(D))$  distribution. So:

$$E(G) = \sum_{n=1}^{\infty} n (P(D))^{n-1} (1-P(D))$$

$$= (1-P(D)) \sum_{n=1}^{\infty} n (P(D))^{n-1}$$

$$= (1-P(D)) \frac{1}{(1-P(D))^2} = \frac{1}{1-P(D)}$$

## Interpretation

$$E(G) = \frac{1}{1-P(D)} = \frac{1}{P(F)+P(M)}$$

In the long run the average number of wins per game is  $(P(F) + P(M))$  and the average number of games per win is the reciprocal.

# Infinite Sequence of Coin Tosses

- Suppose we denote a head by 1 and a tail by 0.
- Then a sequence of coin tosses can be represented as a sequence of zeros and ones:

0 0 1 0 0 1 1 0 1 1...

- Let  $T_r$  denote the number of trials until the  $r$ th success.
- So:  $T_1=3$ ;  $T_2=6$ ;  $T_3=7$ ;  $T_4=9$ ;  $T_5=10$
- Question: What is the distribution of  $T_r$ ?

# Infinite coin tosses

- $T_r$  takes values in  $\{r, r+1, r+2, \dots\}$  and:

- $P(T_r=t) = P(r-1 \text{ successes in } t-1 \text{ trials}$

and  $t$ 'th trial - a success) =

$$= \binom{t-1}{r-1} p^{r-1} (1-p)^{t-r} p = \binom{t-1}{r-1} p^r (1-p)^{t-r}$$

- The distribution of  $T_r - r$ , the number of failures until the  $r$ th success is called **negative binomial** with parameters  $r$  and  $p$ .

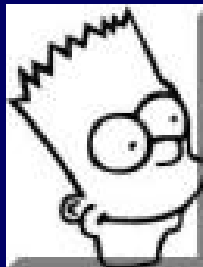
- Question: What is  $E[T_r]$ ? What is  $SD(T_r)$ ?

## Infinite coin tosses

- Let  $W_i$  denote the waiting time after the  $(i-1)$ st success until the  $i$ th success.
- Then  $T_r = W_1 + W_2 + \dots + W_r$  and
- The  $W_i$ 's are independent with  $\text{Geom}(p)$  distribution.
- So:  $E(T_r) = rE(W_i) = r/p$  and
- $SD(T_r) = \frac{\sqrt{r(1-p)}}{p}$

# Collector's Problem

Suppose each box of a particular brand of cereal contain 1 of  $n$  different Simpsons sticker and the sticker in each box is equally likely to be any one of the  $n$ , independently of what was found in the other boxes.



What is the expected number of boxes a collector must buy to get the complete set of stickers?

## Collector's Problem

- After getting  $k$  of the stickers, the additional number of boxes needed to get a different sticker is a  $\text{Geometric}(p_k)$  random variable with

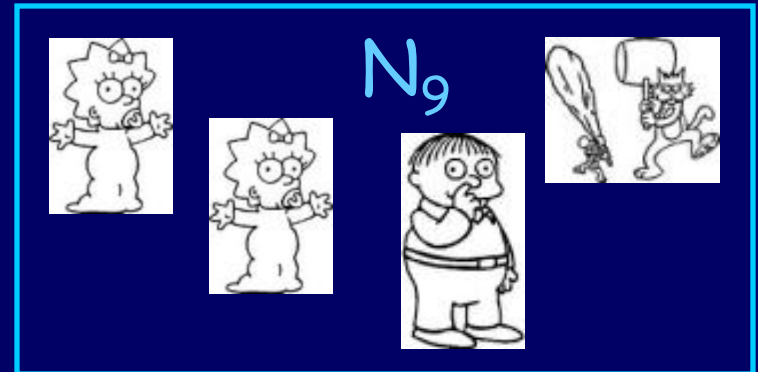
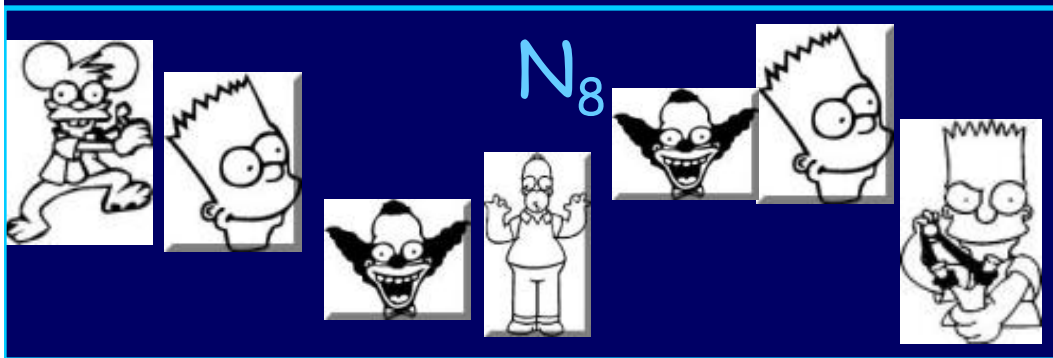
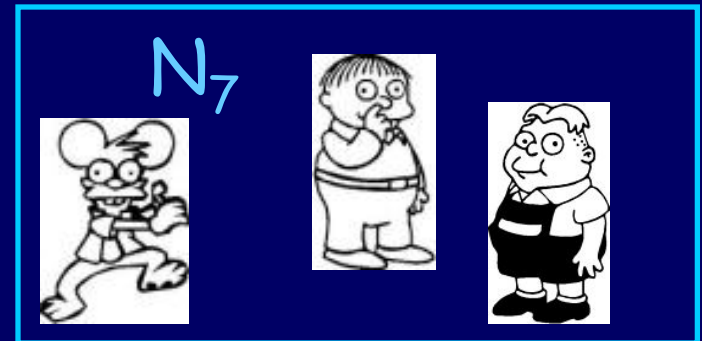
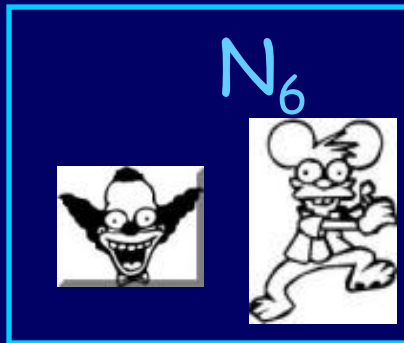
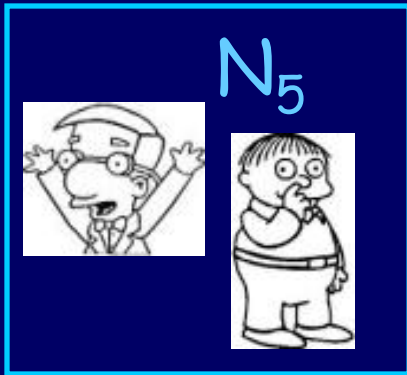
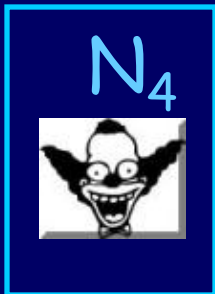
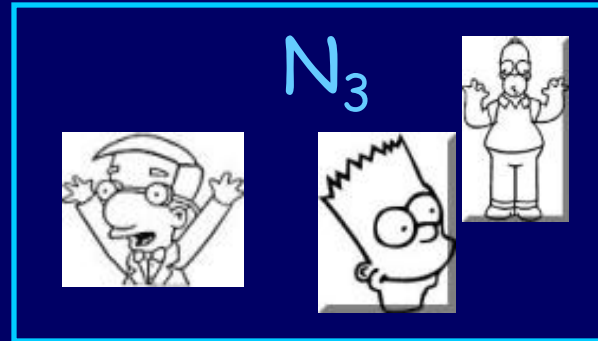
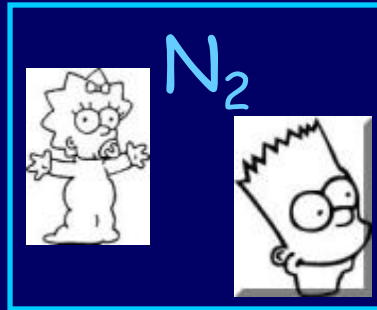
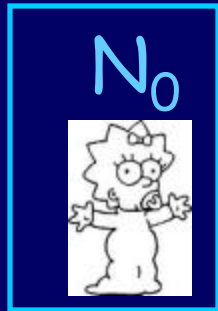
$$p_k = (n-k)/n.$$

- So the number of boxes can be written as a sum of  $n$  geometric r.v.'s:

- $T = N_0 + N_1 + \dots + N_n$ , with

- $N_k \sim \text{Geom}((n-k)/n)$ .

# Collector's Problem



# Collector's Problem

The expected value is:

$$\mu = E(T) = E(N_0) + E(N_1) + \dots + E(N_n)$$

$$\mu = 1 + n/(n-1) + \dots + n/1 = n(1/n + \dots + \frac{1}{2} + 1).$$