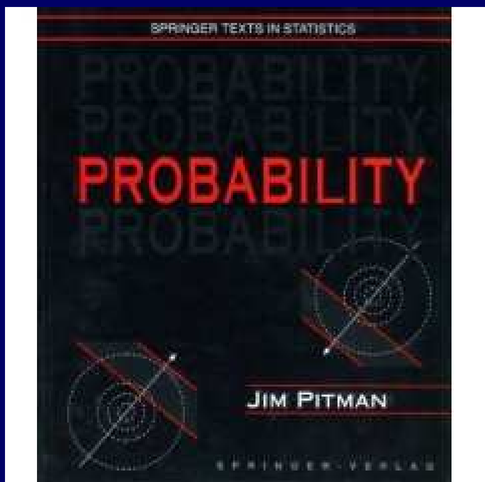
Introduction to probability

Stat 134 FAll 2005

Berkeley



Lectures prepared by:
Elchanan Mossel
Yelena Shvets

Follows Jim Pitman's book:

Probability

Section 3.4

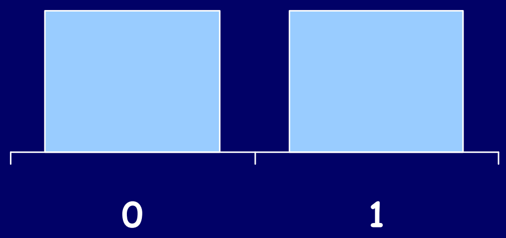
Different types of Distributions

• Finite distribution take only finitely many values.

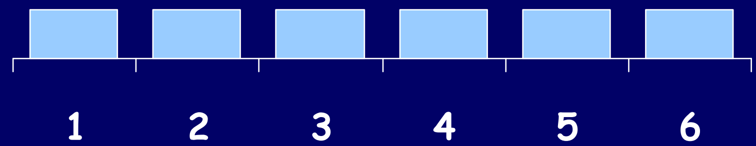
•Examples of finite distributions:

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.



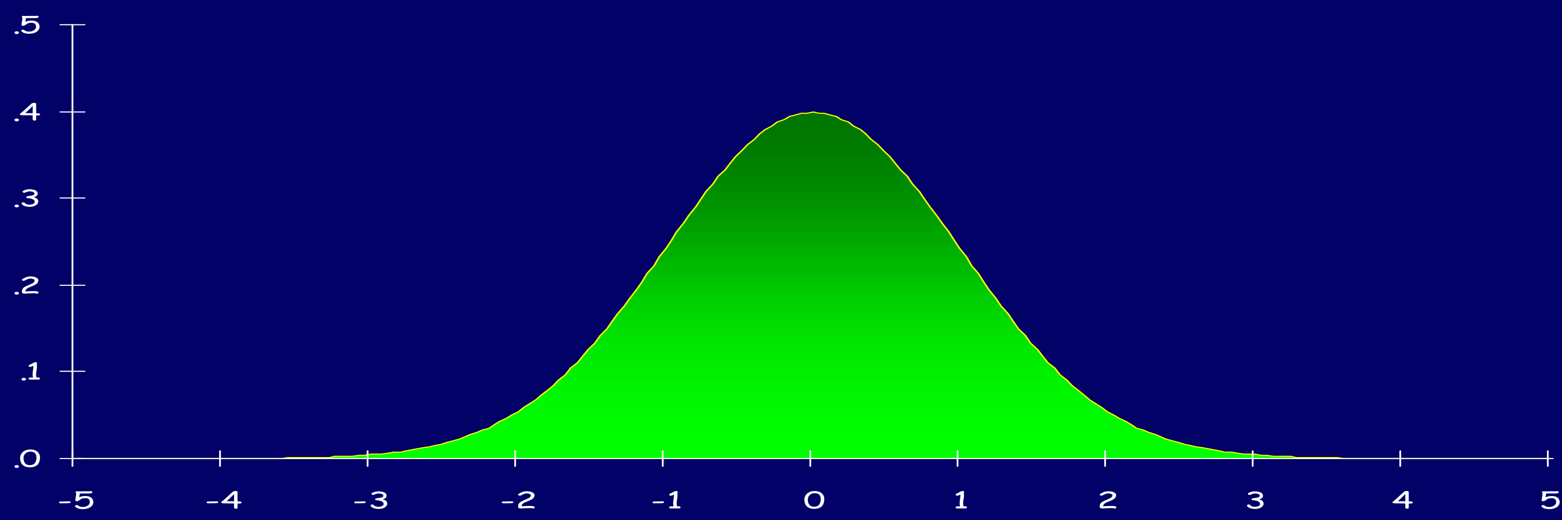
Numbers on a fair die.

Infinite Continuous Distributions

Example:

Normal(0,1);

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



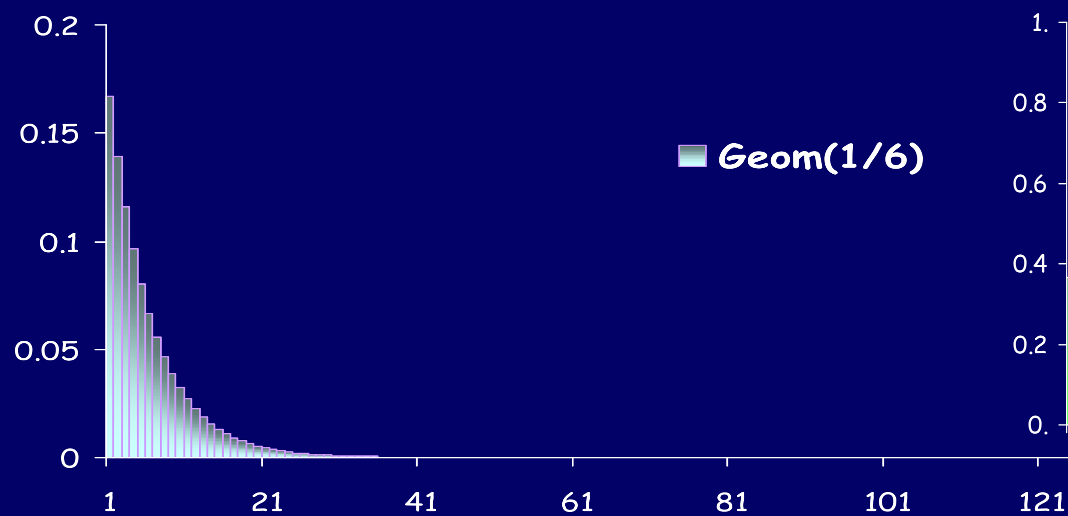
We'll discuss continuous distributions in the 2nd half of the course.

Infinite Discrete Distributions:

- An infinite discrete distribution on {0,1,2,…} 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:

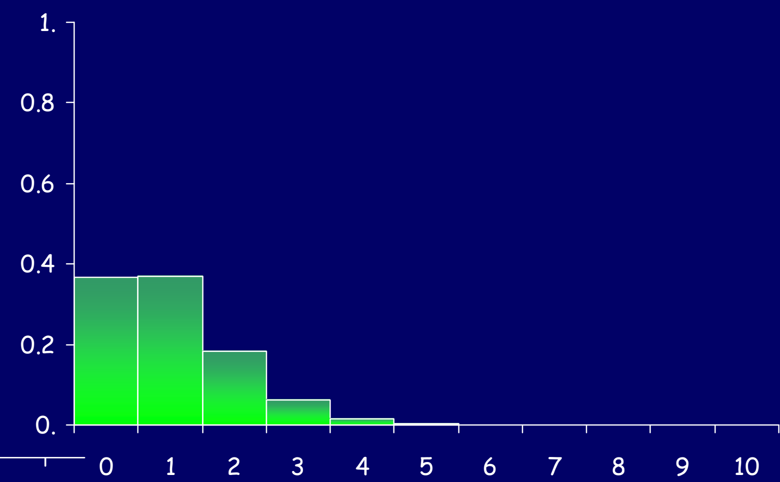
Geometric (p = 1/6)

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



Poisson (μ = 1);

$P(n) = e^{-1}/n!,\ n \geq 0$



Infinite Sum Rule

<u>The infinite sum rule:</u> If the event A is partitioned into A_1, A_2,$A_{3,}$ …

- $A = A_1 \cup A_2 \cup A_3 \cup \ldots$ and
- $A_i \cap A_j = \emptyset$ for all $i \neq j$ then
- $P(A) = P(A_1) + P(A_2) + P(A_3) + \ldots$

- 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={the first six is on the i^{th} roll}

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

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

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

P(waiting forever) = 1 - P(A) = 0

Infinite Sum Rule

$P(T>10) = \sum_{i=11}^{\infty}(P(T=i))$

$P(T \text{ is even}) = \sum_{i=1}^{\infty}(P(T=2i))$

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 ultimately wins) =

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

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

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

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

Repeated games

- <u>Claim:</u> The number of games G they play has geometric distribution.

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

$P(G=n) = P(\text{n-1 draws and the n}^{th} \text{ is not a draw})$

$= (1-p)^{n-1}p.$

So G has a Geom(p) distribution.

- <u>Claim:</u> Let X be the RV that is the identity of the overall winner. Then X and G are independent.

$$P(G=n \;\&\; X=\text{Faust}) = P(\text{n-1 draws \& Faust wins the n}^{th})$$

$$= (P(D))^{n-1} P(F)$$

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

$$= P(G=n)\, P(X=\text{Faust})$$

Expectation of a Discrete Random Variable

The expectation of the random variable X is defined by

$$E(X) = \sum_{x} xP(X=x)$$

provided that the series is convergent:

Expected number of Repeated games

Recall that the number G of games has Geom(1-P(D)) distribution. So:

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

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

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

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.

Geometric Distribution

Recall that for a X = Geom(p) distribution,

$P(X = x) = p(1-p)^{x-1}$ and $P(X \geq x) = (1-p)^{x-1}$

So $E(X) = \Sigma(1-p)^{x-1}$ by the tail sum formula.

This is a geometric series , which we should all know sums to $1/(1 - (1-p)) = 1/p$.

What about Var(X)?

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:

$$0\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 1\ 1\ldots$$

- Let T_r denote the number of trials until the rth success.
- So: T1=3; T2=6; T3=7; T4=9; T5=10
- <u>Question:</u> What is the distribution of T_r?

Infinite coin tosses

- T_r takes values in $\{r, r+1, r+2, \ldots\}$ and:
- $P(T_r=t)$ = P(r-1 successes in t-1 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 rth success is called negative binomial with parameters r and p.

- <u>Question:</u> 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 $T_r = W_1 + W_2 + \ldots + W_r$ and
- The W_i's are independent with Geom(p) distribution.
- So: $E(T_r) = rE(W_i) = r/p$ and
- $SD(T_r) = \dfrac{\sqrt{r(1-p)}}{p}$

Collector's Problem

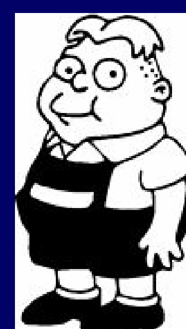
Suppose each box of a particular brand of cereal contain 1 of n different Simpsons sticker and the sticker in each box is equally likely to be any one of the n, independently of what was found in the other boxes.












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 Geometric(p_k) random variable with

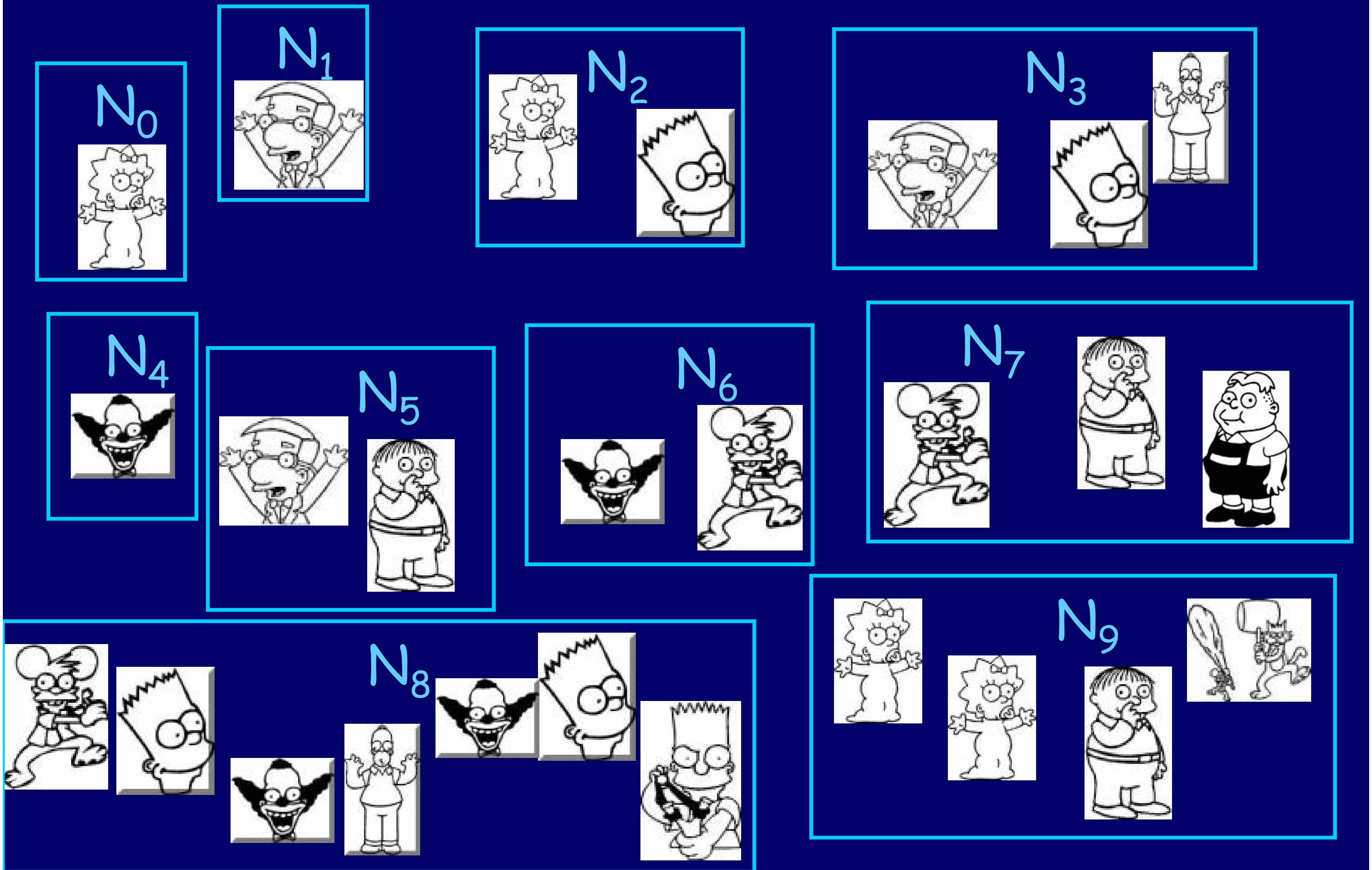
$$p_k = (n-k)/n\ .$$

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

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

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

Collector's Problem



Collector's Problem

The expected value is:

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

$$\mu = 1 + n/(n-1) + \ldots n/1 = n(1/n + \ldots + \tfrac{1}{2} + 1).$$

Collector's Problem – Variance

How to calculate Var?

Are N_i independent?

Yes!

σ^2 = Var(T) = Var(N_0) + Var(N_1) + … Var(N_n)

By Geometic, this is

$\sigma^2 = 0 + n/(n-1)^2 + 2n/(n-2)^2 + \ldots$