Math 3215: Lecture 14

Will Perkins

March 1, 2012

1 Buffon's Needle

2 Approximations to the Binomial Distribution

Recall the binomial (n, p) distribution:

$$\Pr[X = k] = \binom{n}{k} p^k (1-p)n - k$$

For small values of n and single values of k, calculating binomial probabilities is not hard. But what if we're interested in much larger n, and a large range of k's?

    examples:

- political polling: n might be $10,000$ or $20,000$

- medical studies

- quality control in a factory: how would you calculate the probability that more than 1% of the $1,000,000$ Apple laptops produced are defective?

In all of these examples, the binomial distribution is the right model, but the sums and numbers involved are too large for an exact calculation.

There are two different regimes of approximations, which we derive today. The first regime is the case in which n grows very large but p is gets smaller, so that the mean of the binomial, np, stays small, say less than 20 or so.

The second regime is the case in which n grows large, but p stays fixed, so that $np \to \infty$.

3 The Poisson Distribution

Say $X \sim Bin(n, p)$ and $n \cdot p \to \lambda$ as $n \to \infty$.

    Calculate:

- $\lim_{n \to \infty} \Pr[X = 0]$

- $\lim_{n \to \infty} \Pr[X = 1]$

- $\lim_{n \to \infty} \Pr[X = k]$

You've derived the Poisson distribution. It's a discrete distribution, taking values $0, 1 \ldots$ and it has one paramter, λ, its mean.

- Let $Y \sim Pois(\lambda)$. Calculate $var(Y)$.

4 The Birthday Problem

You may have heard of this brain-teaser before:

How many people must be in a room so that the probability that any two of them share the same birthday is at least $1/2$?

The usual way to solve this involves multiplying a lot of fractions. But what if you wanted to know the probability that at least three pairs of students shared the same birthday?

There's an easy way to get a very good approximation to the answer.

We'll do it in steps:

1. If there are n people in a room, what is the epxectation and variance of the number of pairs of people sharing a birthday? (you did this on the homework)

2. Do those values ring a bell?

3. Are the events that different pairs share a birthday indepedent?

4. How close are they to being independent?

To answer the last question we need a new definition:
The *covariance* of random variables X and Y is defined to be:

$$cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}X \cdot \mathbb{E}Y$$

- What is $cov(X, X)$?

- What is the covariance of independent random variables?

- give and upper an lower bound on $cov(X, Y)$ in terms of $var(X)$ and $var(Y)$

Back to the birthday problem:

- What is the covariance of two indicator rv's of pairs of people having the same birthday?

- Does this tell you if the random variables are close or far from being independent?

5 Poisson Approximation

6 Normal Approximation

In 1738, Abraham DeMoive wrote a book called *The Doctrine of Chances*. In it, he studied the probability distribution of the number of heads in 3600 flips of a fair coin. He proved this famous theorem: Let $X \sim Bin(n, p)$. Then

$$\Pr[X = k] \sim \frac{1}{\sqrt{2\pi p(1-p)}} \cdot e^{-(k-np)^2/(2np(1-p))}$$

as $n \to \infty$

- What does this remind you of?