Math 3215: Lecture 15

Will Perkins

March 6, 2012

1 The Birthday Problem

Solving the Birthday Problem using Poisson approximation.

We'll do it in steps:

1. If there are n people in a room, what is the expectation and variance of the number of pairs of people sharing a birthday?
2. Compare that to the mean and variance of a Poisson random variable
3. Are the events that different pairs share a birthday independent?
4. How close are they to being independent - i.e. what is the covariance between the indicator random variables for two different pairs?

Review of covariance:

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

- What is $cov(X, X)$?
- What is the covariance of independent random variables?
- Let X be the number of heads in 10 flips of a fair coin, and Y the number of tails. What is $cov(X, Y)$? Before calculating, give a rough guess.

2 Poisson Approximation

Assume that the number of pairs of people in the room who share a birthday is distributed like a Poisson random variable.

- What's the probability no two people share a birthday?
- What's the probability that exactly three pairs share a birthday?

When is the Poisson approximation accurate?

- The random variable you're approximating is a count.
- There are many things being counted, but each is a 'rare' event
- The different occurrences are close to being independent

3 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 look like?

Written in a different way:

$$\lim_{n\to\infty} \Pr[s \leq \frac{X - np}{\sqrt{np(1-p)}} \leq t] = \Phi(t) - \Phi(s)$$

where $\Phi(t)$ is the CDF of the standard normal distribution.

- What is $\sqrt{np(1-p)}$?

Definition: The *standard deviation* of a random variable is the square root of its variance. Some useful facts about $\Phi(t)$:

- $\Phi(1) - \Phi(0) = .34$
- $\Phi(2) - \Phi(0) = .477$
- $\Phi(3) - \Phi(0) = .4987$

Now let $Z \sim N(0, 1)$. Calculate the following and sketch each corresponding area on a plot of the normal density function:

1. $\Pr[-1 \leq Z \leq 1]$
2. $\Pr[Z \geq 1]$
3. $\Pr[-2 \leq Z \leq 2]$
4. $\Pr[-1 \leq Z \leq 2]$
5. $\Pr[-3 \leq Z \leq 3]$
6. $\Pr[Z \leq -3]$

Pick two of the above and state the corresponding probability of the number of heads in n flips of a p-biased coin using the DeMoive-Laplace theorem.

4 Central Limit Theorem

The Central Limit Theorem says that the DeMoive-Laplace theorem is universal. The sum of any sequence of i.i.d. random variables (with finite variances) converges to a normal distribution when properly scaled.

Central Limit Theorem:

Let $X_1, \ldots X_n$ be i.i.d. random variables with mean μ and variance σ^2. Let $Y = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$. Then

$$\lim_{n\to\infty} \Pr[s \leq Y \leq t] = \Phi(t) - \Phi(s)$$