18.600: Lecture 16

Normal random variables

Scott Sheffield

MIT

Outline

Tossing coins

Normal random variables

Special case of central limit theorem

Outline

Tossing coins

- Suppose we toss a million fair coins. How many heads will we get?
- About half a million, yes, but how close to that? Will we be off by 10 or 1000 or 100,000?
- How can we describe the error?
- Let's try this out.

Tossing coins

- Toss n coins. What is probability to see k heads?
- Answer: $2^{-k}\binom{n}{k}$.
- Let's plot this for a few values of n.
- Seems to look like it's converging to a curve.
- If we replace fair coin with p coin, what's probability to see k heads.
- Answer: $p^k(1-p)^{n-k}\binom{n}{k}$.
- Let's plot this for $p = 2/3$ and some values of n.
- What does limit shape seem to be?

Outline

Outline

Tossing coins

Normal random variables

Special case of central limit theorem

Standard normal random variable

- Say X is a (standard) **normal random variable** if $f_X(x) = f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.
- Clearly f is always non-negative for real values of x, but how do we show that $\int_{-\infty}^{\infty} f(x)dx = 1$?
- Looks kind of tricky.
- Happens to be a nice trick. Write $I = \int_{-\infty}^{\infty} e^{-x^2/2}dx$. Then try to compute I^2 as a two dimensional integral.
- That is, write

$$I^2 = \int_{-\infty}^{\infty} e^{-x^2/2}dx \int_{-\infty}^{\infty} e^{-y^2/2}dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x^2/2}dxe^{-y^2/2}dy.$$

- Then switch to polar coordinates.

$$I^2 = \int_0^{\infty}\int_0^{2\pi} e^{-r^2/2}rd\theta dr = 2\pi\int_0^{\infty} re^{-r^2/2}dr = -2\pi e^{-r^2/2}\Big|_0^{\infty},$$

so $I = \sqrt{2\pi}$.

Standard normal random variable mean and variance

- Say X is a (standard) **normal random variable** if $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.
- Question: what are mean and variance of X?
- $E[X] = \int_{-\infty}^{\infty} xf(x)dx$. Can see by symmetry that this zero.
- Or can compute directly:

$$E[X] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} x dx = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \Big|_{-\infty}^{\infty} = 0.$$

- How would we compute $\mathrm{Var}[X] = \int f(x)x^2 dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} x^2 dx$?
- Try integration by parts with $u = x$ and $dv = xe^{-x^2/2}dx$. Find that $\mathrm{Var}[X] = \frac{1}{\sqrt{2\pi}} (-xe^{-x^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-x^2/2} dx) = 1$.

General normal random variables

- Again, X is a (standard) **normal random variable** if $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.
- What about $Y = \sigma X + \mu$? Can we "stretch out" and "translate" the normal distribution (as we did last lecture for the uniform distribution)?
- Say Y is normal with parameters μ and σ^2 if $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$.
- What are the mean and variance of Y?
- $E[Y] = E[X] + \mu = \mu$ and $\mathrm{Var}[Y] = \sigma^2 \mathrm{Var}[X] = \sigma^2$.

Cumulative distribution function

- Again, X is a standard normal random variable if $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.
- What is the cumulative distribution function?
- Write this as $F_X(a) = P\{X \le a\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx$.
- How can we compute this integral explicitly?
- Can't. Let's just give it a name. Write $\Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx$.
- Values: $\Phi(-3) \approx .0013$, $\Phi(-2) \approx .023$ and $\Phi(-1) \approx .159$.
- Rough rule of thumb: "two thirds of time within one SD of mean, 95 percent of time within 2 SDs of mean."

Outline

Tossing coins

Normal random variables

Special case of central limit theorem

Outline

Tossing coins

Normal random variables

Special case of central limit theorem

DeMoivre-Laplace Limit Theorem

- Let S_n be number of heads in n tosses of a p coin.
- What's the standard deviation of S_n?
- Answer: $\sqrt{npq}$ (where $q = 1 - p$).
- The special quantity $\frac{S_n - np}{\sqrt{npq}}$ describes the number of standard deviations that S_n is above or below its mean.
- What's the mean and variance of this special quantity? Is it roughly normal?
- **DeMoivre-Laplace limit theorem (special case of central limit theorem):**

$$\lim_{n \to \infty} P\{a \le \frac{S_n - np}{\sqrt{npq}} \le b\} \to \Phi(b) - \Phi(a).$$

- This is $\Phi(b) - \Phi(a) = P\{a \le X \le b\}$ when X is a standard normal random variable.

Problems

- Toss a million fair coins. Approximate the probability that I get more than $501,000$ heads.
- Answer: well, $\sqrt{npq} = \sqrt{10^6 \times .5 \times .5} = 500$. So we're asking for probability to be over two SDs above mean. This is approximately $1 - \Phi(2) = \Phi(-2) \approx .159$.
- Roll 60000 dice. Expect to see 10000 sixes. What's the probability to see more than 9800?
- Here $\sqrt{npq} = \sqrt{60000 \times \frac{1}{6} \times \frac{5}{6}} \approx 91.28$.
- And $200/91.28 \approx 2.19$. Answer is about $1 - \Phi(-2.19)$.