STAT 230
SECTION 2
LECTURE 27

Today's Agenda

Last time:

- Brief review exponential distribution
- Percentile
- Generating observations of random variables on a computer

Today (Lec 27, 07/06):

- The normal distribution

Definition

X is said to have a **normal distribution** (or Gaussian distribution) with mean μ and variance σ^2 if the density of X is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Shorthand, $X \sim N(\mu, \sigma^2)$

Figure: Carl Friedrich Gauss, 1777-1855.

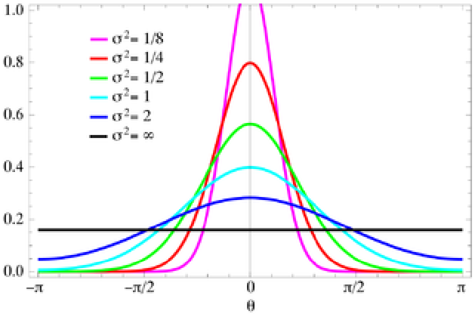

Figure: Normal densities with mean zero and increasing standard deviation.

Properties of the normal distribution:

a) Symmetric about its mean: If $X \sim N(\mu, \sigma^2)$

$$P(X \leq \mu + t) = P(X \geq \mu - t).$$

b) Density is unimodal: Peak is at μ.

c) Mean and Variance are the parameters:

$$E(X) = \mu$$

and

$$Var(X) = \sigma^2.$$

d) $N(\mu, \sigma^2)$ is also parametrised as Gaussian distribution, where

$$X \sim G(\mu, \sigma).$$

Many probabilists and statisticians would argue (myself included) that the normal distribution is the most important distribution.

a) The main reason for this is known as the Central Limit Theorem, which I hope to get to by the end of this course.
b) As described by the Central Limit Theorem, the normal distribution models many different phenomena: Test scores on exams, measurements of heights and weights, error measurements in experiments, etc..

Major problem with the normal random variable: If $X \sim N(\mu, \sigma^2)$, then

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = ???$$

functions of the form e^{-x^2} do not have elementary anti-derivatives... Curses!

This is a famous theorem of Liouville in the mid 1800's.

In practice these integrals must be evaluated numerically.

Definition

We say that X is a **standard normal** random variable if $X \sim N(0, 1)$.

Frequently in probability and statistics literature, the density of the standard normal random variable is denoted

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

and the cdf of a standard normal random variable is denoted

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy.$$

Why is there $1/\sqrt{2\pi}$ in the pdf?

That's because

$$\int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} dx = \sqrt{2\pi},$$

so we divide the integral by $\sqrt{2\pi}$ to obtain a valid pdf.

This can be shown through the Gamma function, see p175.

Though the values of Normal cdf are hard to compute, the cdf values of standard Normal distribution are tabulated thoroughly!

Values of the function $\Phi(x)$ are tabulated in what are called "Standard Normal Tables" or "Z-Tables" (see end of course notes).

Example

Suppose that $Z \sim N(0,1)$. Compute

a) $P(Z \leq 2.15)$

b) $P(Z \leq -1.29)$

Question

Suppose $X \sim N(0, 1)$. Compute $P(X > 2.0)$.

A .97725

B .98928

C 1-.97725

D 1-.98928

But wait! How do I use the Z-table if I don't have $N(0,1)$??

No worries! Mathematics saves the day again!

Theorem (Standardising normal random variable)

If $X \sim N(\mu, \sigma^2)$*, then*

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1),$$

and $P(X \leq x) = P\left(Z \leq \frac{x - \mu}{\sigma}\right)$.

Example

Let $X \sim N(4, 2)$. Compute the following:

- $P(X < 2.5)$
- $P(X > 1)$
- $P(-3 \leq X \leq 1)$
- $P(X \leq 4)$

Question

We know how to go from any normal distribution $N(\mu, \sigma^2)$ to standard normal $N(0, 1)$. On that note, can we go from standard normal $Z \sim N(0, 1)$ to an arbitrary normal random variable $X \sim N(\mu, \sigma^2)$?

A Yes

B No

C Maybe so

The Z-table can also be used to obtain the percentile. Let's see how.

Example

Compute the

a) 75th percentile of the standard normal distribution

b) 58th percentile of the $N(5, 9)$ distribution

c) Let $Z \sim N(0, 1)$. Find c such that

$$P(-c \leq Z \leq c) = 0.95$$

An interesting empirical rule about normal distribution is the **68-95-99.7** rule, which states:

If $X \sim N(\mu, \sigma^2)$, then

$$P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.68$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.997.$$