STAT 515 -- Chapter 5: Continuous Distributions

Probability distributions are used a bit differently for continuous r.v.'s than for discrete r.v.'s.

Continuous distributions typically are represented by a <u>probability density function</u> (pdf), or "density curve." (kind of a "theoretical" histogram)

A <u>density</u> curve is a representation of the <u>underlying population distribution</u> (not a description of actual sample data).

The <u>normal distribution</u> is a particular type of continuous distribution. Its density has a bell shape:

Properties of Density Functions:
(1) Density function always on or above the horizontal axis (curve can never have a negative value)
(2) Total area beneath the curve (between curve and horizontal axis) is exactly 1.
(3) An area under a density function represents a probability about the r.v. (or the proportion of observations we expect to have certain values).

With discrete r.v.'s we looked at probability function (table, graph) to find probability of the r.v. taking a particular value.

For continuous r.v.'s, the probability distribution will give us the probability that a value falls in an <u>interval</u> (for example, between two numbers).

That is, the probability distribution of a continuous r.v. X will tell us $P(a \leq X \leq b)$, where a and b are particular numbers.

Specifically, $P(a \leq X \leq b)$ is the area under the density function between $x = a$ and $x = b$.

Examples:

The Uniform Distribution

This is a simple example of a continuous distribution.

A uniform r.v. is equally likely to take any value between its lower limit (some number c) and its upper limit (some number d).

Density looks like a rectangle:

If total area is 1, then what is the height of the density function?

- Mean of a Uniform(c, d) r.v. $= (c + d) / 2$
- Std. deviation of a Uniform(c, d) r.v. $= (d - c) / \sqrt{12}$

Example: A machine designed to fill 16-ounce water bottles actually dispenses a random amount between 15.0 and 17.0 ounces. The amount X of water dispensed is a Uniform(15, 17) random variable:
Density:

What is the probability that the bottle has less than 15.5 ounces of water?

$$P(X < 15.5) = P(15 < X < 15.5) =$$

In general:

For $X \sim$ Uniform(c, d):

$$P(a < X < b) = \frac{b - a}{d - c}$$

The Normal Distribution

The density function for the normal distribution is complicated:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} \text{ for all } x$$

Note that the normal distribution changes depending on the values of the mean μ and the standard deviation σ.

Standard Normal Distribution [Notation: N(0, 1)]: The normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$.

Picture:

• Mound-shaped, symmetric, centered at 0.
• Density always positive, even in "tails."
• Area under curve is 0.5 to left of zero, 0.5 to right of zero.
• Almost all area under curve (99.7%) between -3 and 3.

Note: N(0, 1) distribution sometimes called the "z-distribution" and standard normal values are denoted by z.

Table IV in back of book gives areas between 0 and certain listed values of z.

Example: Area under N(0, 1) curve between 0 and 1.24:

Table IV: Go to row labeled 1.2, column labeled .04: Correct area =

What does this area mean?
• If Z is a r.v. with a standard normal distribution, then P(0 < Z < 1.24) =
[Note: Same as P(0 ≤ Z ≤ 1.24).]
• We expect that 39.25% of the values of data having a standard normal distribution will be between 0 and 1.24

Other Probabilities:

P(Z > 1.24) =

P(Z < 1.24) =

Values to the left of zero? Use symmetry!

$P(-0.54 \leq Z < 0) =$

$P(Z < -0.54) =$

$P(-1.75 < Z < -0.79) =$

$P(-0.79 < Z < 1.16) =$

Finding Probabilities for *any* Normal r.v.

Note: There are many different normal distributions (change μ and/or σ, get a different distribution).

• Changing μ shifts the distribution to the left or right.

• Increasing σ makes the normal distribution wider.

• Decreasing σ makes the normal distribution narrower.

So why so much emphasis on the <u>standard</u> normal?

<u>Standardizing</u>: If a r.v. *X* has a normal distribution with mean μ and standard deviation σ, then the standardized variable

$$Z = \frac{X - \mu}{\sigma}$$

has a standard normal distribution.

So: We can convert <u>any</u> normal r.v. to a standard normal and then use Table IV to find probabilities!

Example: Assume lengths of pregnancies are normally distributed with mean 266 days and standard deviation 16 days.

What proportion of pregnancies last less than 255 days?

What is the probability that a random pregnancy will last between 260 and 280 days?

We can also find the particular value of a normal r.v. that corresponds to a given proportion.

Example: Suppose the shortest tenth of pregnancies are classified as "unusually premature." What's the maximum pregnancy length that would be classified as such?

We need to "unstandardize" to get back to the X value (pregnancy length).

General Rule: To unstandardize a z-value, use:
$$X = Z\sigma + \mu$$

Normal Approximation to the Binomial

The normal distribution is very powerful --- can be used to approximate probabilities for r.v.'s that are not normal.

Calculating binomial probabilities using Table II: doesn't cover all values of *n* (only 5-10, 15, 20, 25)

Using the binomial probability formula can be tedious for large *n*.

Fortunately, when *n* is large, the binomial distribution closely resembles the normal distribution with mean *np* and standard deviation $\sqrt{npq}$.

<u>Rule of Thumb</u>: When can this normal approximation be applied?

<u>Continuity Correction</u>: Since the normal is a continuous distribution and the binomial distribution a discrete distribution, an adjustment of 0.5 is usually made to the value of interest.

Example: A hotel has found that 5 percent of its guests will steal towels. If there are 220 rooms with guests in a hotel on a certain night, what is the probability that at least 20 of the rooms will need the towels replaced?

What is the probability that between 10 and 20 rooms will need towels replaced?