MAT 2379, Introduction to Biostatistics

Chapter 7. Continuous Random Variables
7.1 Definition

Continuous random variables can assume an infinite number of values in some interval. The major *mathematical* difference between continuous and discrete random variables is that the probability that a continuous random variable X takes a particular value x is 0, that is:

$$P(X = x) = 0 \quad \text{for any numerical value } x.$$

Examples of continuous random variables:
1. The weight (or height) of an individual; 2. The blood pressure (or temperature) of a patient; 3. The weight gain of a woman during pregnancy.

For continuous random variables, the best thing that one can do is to estimate the probability that the variable will take values in a given interval $[a, b]$. The function f which gives these probabilities is called **the density function** of X:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

This can be represented graphically as the area under the graph of f between the points a and b. Moreover, since $f(x)$ are probabilities, we should always have $f(x) \geq 0$ and

$$\int f(x)dx = 1$$

where the integral is taken over the interval of all possible values for X, i.e. the density function f has to be chosen such that it takes only non-negative values and the total area under its graph is equal to 1.

The expectation $E(X)$ (or the "average") of a continuous random variable X can be calculated by means of the following formula:

$$\mu = E(X) = \int xf(x)dx$$

where the integral is taken over the interval of all possible values for X. The variance of X is

$$\sigma^2 = \mathrm{Var}(X) = E[(X - \mu)^2] = \int (x - \mu)^2 f(x)dx.$$

Cumulative Distribution Function

As in the discrete case, the cumulative distribution function $F(x_0)$ gives the probability that X takes values smaller or equal to x_0:

$$F(x_0) := P(X \leq x_0) = \int_{-\infty}^{x_0} f(x)dx$$

This integral is equal to the area under the graph of f at the left of x_0.

7.2 Normal Distribution

In this section we will learn about the well-known normal distribution, whose density function has a bell-shaped graph. This was "discovered" by De Moivre, Laplace and Gauss (in 1700's) as being a good approximation for the binomial distribution when the number n of trials becomes large. This is a very important distribution and most of the statistical methods that we will see in this course will be based on it.

To describe a particular normal distribution, we have to specify its parameters: μ denotes its mean (or expected value) and σ its standard deviation. The density function is

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

The graph of this density function is bell-shaped (symmetric), centered at μ and has inflection points at $\mu + \sigma$ and $\mu - \sigma$.

Standard Normal Distribution

The most important normal distribution is the one that has mean $\mu = 0$ and standard deviation $\sigma = 1$. This is called the *standard normal distribution*. A standard normal random variable is usually denoted with Z.

Tables 17.2 and 17.3 at the end of the book (pages 235-237) gives the areas under the standard normal distribution. It is extremely important that you learn very soon how to use this table.

Example 3. (to be solved in class) Let Z be a standard normal random variable. Using Tables 17.2 and 17.3, calculate the following probabilities:
(a) $P(Z \leq -1.52)$ (b) $P(Z \leq 1.37)$ (c) $P(Z \geq 1.98)$
(d) $P(-1.21 \leq Z \leq 1.73)$ (e) $P(Z = 1.50)$
Using Tables 17.2 and 17.3, find the point z such that:
(f) $P(Z \leq z) = 0.05$ (g) $P(Z \leq z) = 0.75$ (h) $P(-z \leq Z \leq z) = 0.95$.

Non-Standard Normal Distribution (Standardization Procedure)

If X has a normal distribution with mean μ and variance σ^2, then

$$\frac{X - \mu}{\sigma} \text{ has a standard normal distribution}$$

Example 4. The length of a fish is a normal random variable X which has a normal distribution with mean $\mu = 54$ mm and and standard deviation $\sigma = 4.5$ mm.
(a) What is the probability that a randomly chosen fish is less than 60 mm long?

$$P(X \leq 60) = P\left(\frac{X - 54}{4.5} \leq \frac{60 - 54}{4.5}\right) = P(Z \leq 1.33) = 0.9082$$

(b) What is the probability that a randomly chosen fish is more than 51 mm long?

$$P(X \geq 51) = P\left(\frac{X - 54}{4.5} \geq \frac{51 - 54}{4.5}\right) = P(Z \geq -0.67) = 1 - P(Z \leq -0.67) = 1 - 0.2514 = 0.7486$$

(c) What is the probability that a randomly chosen fish is less than 62 mm long but more than 55 mm long?

$$P(55 \leq X \leq 62) = P\left(\frac{55-54}{4.5} \leq \frac{X-54}{4.5} \leq \frac{62-54}{4.5}\right) = P(0.22 \leq Z \leq 1.78) =$$

$$P(Z \leq 1.78) - P(Z \leq 0.22) = 0.9625 - 0.5871 = 0.3754$$

(d) Find a length x_0 such that 20% of these fish have a length smaller than x_0. From

$$0.2 = P(X \leq x_0) = P\left(\frac{X-54}{4.5} \leq \frac{x_0-54}{4.5}\right) = P\left(Z \leq \frac{x_0-54}{4.5}\right)$$

we get $\frac{x_0-54}{4.5} = -0.84$ Solving for x_0, we get

$$x_0 = 54 + (4.5)(-0.84) = 54 - 3.78 = 50.22$$