5.4 PROBABILITY

Probability 101

Probability is the likelihood of an event occurring. Probability is given as a fraction, a percentage, or a decimal between 0 and 1. A probability of 0 means something will never happen, and a probability of 1 means the event is guaranteed to happen. When you flip a coin you have a 50% chance of the coin landing on heads. If you play the Colorado Lotto, you have a 1 in 5,245,786 chance of winning the jackpot. (Which, by the way is 0.0000001906)

Example: Suppose a jar is contains 7 black marbles, 6 yellow marbles, 4 green marbles, and 3 red marbles that are exactly the same except for their color. If you randomly select a marble from the jar find the following probabilities:

a) The marble is black.

b) The marble is yellow.

c) The marble is green.

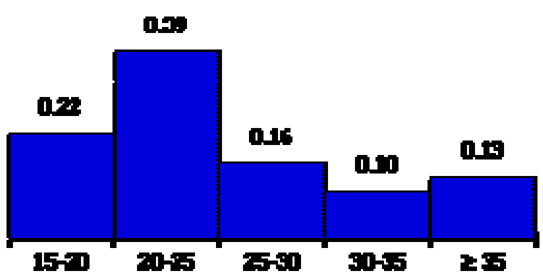
d) The marble is red.

Example: Make a histogram of the probabilities above with each bar having width = 1. Add all the probabilities. What do you get?

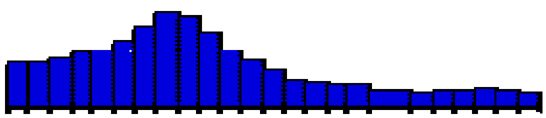
Example: The following table shows the distribution of U.S. residents (16 years old and over) attending college in 1980 according to age.

Age	15-20	20-25	25-30	30-35	≥35
Probability	0.22	0.39	0.16	0.10	0.13

The histogram below represents this data. Add all the probabilities. What do you get?



If the data had been collected at intervals that were only 1 apart instead of 5, then the groupings would have been smoother and the bars shorter, as shown below. Why?



A **random variable** is a function X that assigns to each possible outcome in an experiment a real number. If X may assume any value in some given interval I (the interval may be bounded or unbounded), X is called a **continuous** random variable. If it can assume only a number of separated values, X is called a **discrete** random variable.

From [http://en.wikipedia.org/wiki/Probability_density_function]

> "Informally, a probability density function can be seen as a "smoothed out" version of a histogram: if one empirically samples enough values of a continuous random variable, producing a histogram depicting relative frequencies of output ranges, then this histogram will resemble the random variable's probability density, assuming that the output ranges are sufficiently narrow."

What Does All This Mean for Us?

If a probability distribution function is just a "smoothed out" version of a histogram, and the probability of each outcome can be represented by a rectangle, then if we could assign a function that would hit the top of each bar of the histogram ("smoothing it out"), then the area under the curve would represent the probability!

Probability Density Function Defined

Let x be a continuous random variable. A function f is said to be a probability density function for x if:

1. For all x in the domain of f, we have $f(x) \geq 0$. … (So the curve must always be above the x-axis.)
2. The total area under the curve is 1. … (100% = 1)
3. For any subinterval $[c, d]$ in the domain of f, the probability that x will be in a given subinterval is

$$\int_c^d f(x)\,dx$$ … (the area under the curve from c to d)

Example: In a psychology experiment, the time t, in seconds, that it takes a rat to learn its way through a maze is an exponentially distributed random variable with the probability density function $f(t) = 0.02e^{-0.02t}$, $0 \leq t < \infty$.

a) Show that this function meets criteria #2 above.

b) What is the probability that a rat will learn its way through a maze in 150 seconds or less?

Example: Let $f(x) = kx^2$. Find k such that the function is a probability density function over the interval [–2, 2].

An Extension into 5.5 …

Have you ever seen a “bell curve”? Have you ever asked your teacher to “curve” a test. A “normal probability density function” is the mathematical name of a “bell curve”. The probability of an event occurring between two values is the area under the “bell curve” between two points on the x – axis.

The general form of a normal probability density function (whose mean is 0) is given by

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

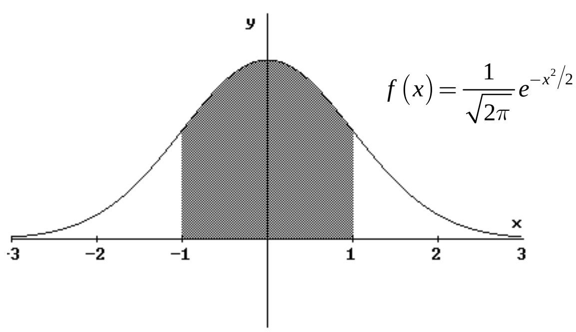
where σ is the standard deviation (σ is the lowercase Greek letter sigma).

Luckily for you this is not a statistics class, or you would have to memorize this function! Let’s use what we know to obtain a more “familiar” result.

A *standard normal distribution* has a mean = 0 and a standard deviation = 1. Any data that is normally distributed can be “standardized”. This simplifies the above equation to

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

Example: Use your calculator to evaluate the probability of an event occurring within 1 standard deviation of a normally distributed function (standard deviation = 1 and mean = 0). The picture is below.



Just fyi … if you ask to “curve” a test … mathematically speaking … anyone within 1 standard deviation of the mean (average) would get a C. That means the area under the curve above represents the percentage of students who would earn a C on the exam.

Example: Anyone between 1 and 2 standard deviations from the mean would earn a B. How would you find this percentage?

Example: Anyone more than 3 standard deviations less than the mean would receive an F. How could you determine this percentage?