Random Variables

The Basics

The Definition

So we've talked about variables. And we've talked about things that are random. Now it's time to put the two together. A **Random Variable** measures the (quantitative) result of some random experiment. Pick a person at random, and measure his/her age: you've got a random variable (*age*). Pick a group of 10 cars and count how many are red: you've got a random variable (*number that are red*).

Properties

First of all, there are two types of random variables—discrete and continuous. We'll be focusing on the discrete case, for now. Thus, our random variables can only take on certain values (not ranges of values).

[1] The values of the random variable must represent disjoint events. If our random experiment involves measuring several things (10 cars, 5 people, 30 bags of sand, etc.), then the variable must hold some quantity that summarizes those measurements (number of red cars, mean age of people, etc.). It should not be possible to get two different summaries from one trial of an experiment!

[2] The sum of the probabilities must be 1. If you've got a list of everything that could happen, then the union of those events had better be the universe!

Examples

[1.] Let X represent the number of siblings that a student (of a particular statistics class) has. The probability distribution of X is (partially) given below.

Table 1 - Distribution of Siblings

x	**0**	**1**	**2**	**3**	**4**
$P(X=x)$	0.200	0.425	0.275	0.075	???

What is the value of $P(X = 4)$?

Since the probabilities must add to 1, $P(X = 4)$ must be 0.025 (1 – 0.2 – 0.425 – 0.275 – 0.075).

Describing the Distribution

The Mean

For discrete random variables, finding the mean is just like finding a weighted average; each value of the variable is weighted according to its probability.

Equation 1 - Mean of a Discrete Random Variable

$$\mu_X = \sum_{i=1}^{n} \left(x_i \cdot P\left(X = x_i \right) \right)$$

Note that the mean of a random variable is also called the **Expected Value**.

The Standard Deviation

Finding the standard deviation of a discrete random variable is very similar to that of a sample.

Equation 2 - Standard Deviation of a Discrete Random Variable

$$\sigma_X = \sqrt{\sum_{i=1}^{n} \left[\left(x_i - \mu_X \right)^2 \cdot P\left(X = x_i \right) \right]}$$

Of course, to find the variance (denoted Var(X) or σ_X^2), just square the standard deviation.

Examples

[2.] Back to Example 1—let's find the mean and variance.

$\mu_X = \sum_{i=1}^{n} \left(x_i \cdot P\left(X = x_i \right) \right)$ = (0)(0.2) + (1)(0.425) + (2)(0.275) + (3)(0.075) + (4)(0.025) = 1.3.

$\sigma_X^2 = \sum_{i=1}^{n} \left[\left(x_i - \mu_X \right)^2 \cdot P\left(X = x_i \right) \right]$ = $(0 - 1.3)^2(0.2) + (1 - 1.3)^2(0.425) + (2 - 1.3)^2(0.275) + (3 - 1.3)^2(0.075) + (4 - 1.3)^2(0.025) = 0.91$.

Transformations of Random Variables

Linear Transformations

There are times when we need to make adjustments to our random variables—perhaps they were measured with the wrong units; maybe we forgot to measure from the zero mark on the ruler, and accidentally added 1cm to each measurement. Fixing these errors (multiplying by 2.54 to change inches to centimeters; subtracting one) are examples of **Linear Transformations of Random Variables**.

Multiplying by a Constant

If X is a random variable, then $3X$ is also a random variable, where each datum of the set is multiplied by 3.

What will that do to the mean? $\mu_{3X} = \sum_{i=1}^{n} \left(3x_i \cdot P\left(3X = 3x_i \right) \right)$

$\Rightarrow \mu_{3X} = 3\sum_{i=1}^{n} \left(x_i \cdot P\left(X = x_i \right) \right) = 3\mu_X$. Looking at this equation, you can see that the 3 can be factored out (it isn't being factored out of the probability statement—that part simply matches the preceding variable), leaving you with the expression of μ_X. So, in general: $\mu_{aX} = a \cdot \mu_X$

What will that do to the standard deviation? $\sigma_{3X} = \sqrt{\sum_{i=1}^{n}\left[\left(3x_i - 3\mu_X\right)^2 \cdot P\left(3X = 3x_i\right)\right]}$

$\Rightarrow \sigma_{3X} = \sqrt{\sum_{i=1}^{n}\left[3^2\left(x_i - \mu_X\right)^2 \cdot P\left(X = x_i\right)\right]} \Rightarrow \sigma_{3X} = 3 \cdot \sqrt{\sum_{i=1}^{n}\left[\left(x_i - \mu_X\right)^2 \cdot P\left(X = x_i\right)\right]} = 3 \cdot \sigma_X$. Again, the 3 can be factored out. But be careful! We really took out $\sqrt{3^2}$, which is 3. If we had taken out $\sqrt{\left(-3\right)^2}$, it would come out as positive 3 (because of the squaring). So, in general: $\sigma_{aX} = |a| \cdot \sigma_X$.

Adding a Constant

If Y is a random variable, then Y+1 is also a random variable, where each datum of the set has 1 added to it.

What will this do to the mean? $\mu_{Y+1} = \sum_{i=1}^{n}\left(\left(y_i + 1\right) \cdot P\left(Y + 1 = y_i + 1\right)\right) =$ $\sum_{i=1}^{n}\left(y_i \cdot P\left(Y = y_i\right)\right) + \sum_{i=1}^{n}\left(1 \cdot P\left(Y = y_i\right)\right) = \mu_Y + 1$. The second term is one since the sum of the probabilities is 1 (and that's all the second term is). So adding one to every value simply added one to the mean. In general: $\mu_{Y+b} = \mu_Y + b$.

What will happen to the standard deviation? $\sigma_{Y+1} = \sqrt{\sum_{i=1}^{n}\left[\left(\left(y_i + 1\right) - \mu_{Y+1}\right)^2 \cdot P\left(Y + 1 = y_i + 1\right)\right]}$ $= \sqrt{\sum_{i=1}^{n}\left[\left(\left(y_i + 1\right) - \left(\mu_Y + 1\right)\right)^2 \cdot P\left(Y + 1 = y_i + 1\right)\right]} = \sqrt{\sum_{i=1}^{n}\left[\left(y_i - \mu_Y\right)^2 \cdot P\left(Y = y_i\right)\right]} = \sigma_Y$. So adding one to every value doesn't change the spread! Thus, in general: $\sigma_{Y+b} = \sigma_Y$.

Combining Random Variables

Sometimes, we want to combine random variables into a new random variable. For example, perhaps we want to add SAT-M scores to SAT-V scores. Each of these is a random variable, with a certain distribution—but what we want is a measure that combines the scores from each. So we need to somehow combine the distributions. Our earlier work with linear transformations can still apply, with one caveat: *the variables must be independent*. Also note that we must add variances, not standard deviations. Of course, once you have the variance, the standard deviation is easy.

Equation 3 - Mean of the Combined Random Variable

$$\mu_{X+Y} = \mu_X + \mu_Y$$

Equation 4 - Variance of the Combined Random Variable

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$$

Take note, however, that $\mu_{X+X} \neq \mu_{2X}$. To illustrate the difference, let X represent the number of pips showing when a fair six sided die is rolled. The variable $X + X$ represents adding the pips on two dice, which can give a number between 2 and 12. $2X$ represents doubling the results of a single die, which gives *only even numbers* from 2 to 12.

Examples

[3.] D is a random variable with mean 3.5 and standard deviation 1.708. C is another random variable (independent of D) with mean 1.5 and standard deviation 0.5. The random variable S is defined by $S = D + 2C$. What are the mean and standard deviation of S?

$\mu_S = \mu_{D+2C} = \mu_D + \mu_{2C} = \mu_D + 2\mu_C = 3.5 + 2(1.5) = 6.5$.

$\sigma_S^2 = \sigma_{D+2C}^2 = \sigma_D^2 + \sigma_{2C}^2 = \sigma_D^2 + 4\sigma_C^2 = (1.708)^2 + 4(0.5)^2 = 3.9167$. This is the variance; take the square root to obtain the standard deviation: 1.9791.