Statistics Review

This is a review of some basic concepts from probability theory. Most of the concepts will be encountered in Finance 406 so please at least read through this to (re)familiarize yourself with the material. Being comfortable with these concepts will help immensely with your upcoming homework and lectures.

There are two types of probability distributions

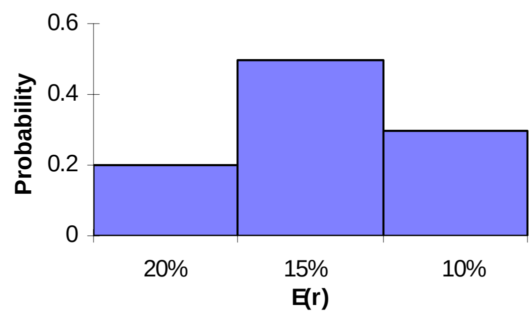
A. **Discrete random variables** give us a **frequency distribution**: when an experiment has a limited number of possible outcomes, i.e. the outcomes are countable and finite, then the probability distribution can be represented in a table or a histogram.

B. **Continuous random variables** give us a **density function:** when the number of possible outcomes from an experiment is infinite or countably infinite.
   1. The graph of the probability distribution function is smooth (unlike a histogram).
   2. For example, a stock price is an outcome that can take on values between zero and infinity, making it countably infinite. Hence, the random variable is continuous and must be described by a function (rather than a table of numbers).

Here is an example to help you discern between discrete and continuous random variables.

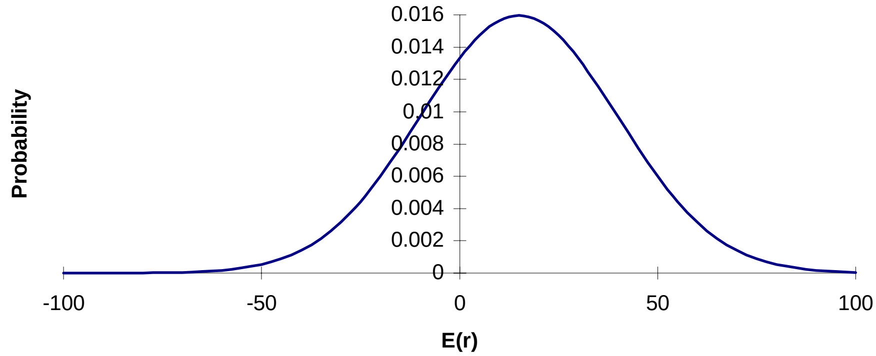
A. First think of a *scenario analysis*. If we limit the possible outcomes from investing in a project to three possible states of the economy we have a discrete random variable that we can tabulate and graph as follows. Both the table and the graph are representations of the frequency distribution. Note that the probabilities add up to one.

Economy	Probability	Return
good	.2	20%
normal	.5	15%
bad	.3	10%



Here we have a project that we expect to give us a 20% return if the economy is good and we think there is a 20% chance that the economy will be good. We think there is a 50% chance that the economy will be normal and in that case we expect a return of 15%, and so on.

B. If we let the possible outcomes take on *any* value and postulate a particular shape of the distribution of the outcomes we may represent the probability of the outcomes with some function and a graph as follows.



The function that I used to draw this probability density is

$$f(r) = \frac{1}{\sqrt{2\pi\sigma}} e^{\left(\frac{-(r-\mu)^2}{2\sigma^2}\right)}$$

This looks pretty nasty but it is just the probability density function for the Normal or Gaussian distribution. Here I have used μ = 15 and σ = 15. (More on this distribution in a bit)

Moments of Distributions:

A. **Measures of central tendency:**

1. The first *moment* of a distribution is the **mean or expected value**.
2. If we have a sample of data then the expected value over that sample is found as:

$$Mean = E(x) = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i ,$$

where n is the number of data points in the sample.

3. If we are looking forward and have assigned probabilities to possible outcomes in the future, then the expected value is found as:

$$Expected\ value = E(x) = p_1X_1 + p_2X_2 + ... + p_nX_n,$$

where x_i are the possible outcomes and p_i are the assigned probabilities.

4. The two formulas are equivalent. With a sample of historical data, we consider that each outcome has the same probability, namely 1/n.
5. The mean is the most common measure of central tendency and represents the most likely outcome. Other measures of central tendency are the
   a) Mode: the most often occurring outcome,
   b) Median: the outcome falling halfway between the endpoints of the sample.

NOTE: If a constant, k, is added to a random variable, the mean is also increased by the constant. That is, if $E(x) = \mu$, then $E(x+k) = \mu + k$. Also, if a random variable is multiplied by a constant, the mean is also multiplied by that constant. That is, if $E(x) = \mu$, then $E(x*k) = \mu * k$. In other words if I take all the grades for the class and add 10 points to each grade then the average grade in the class goes up by 10 points. If I multiply each grade by 10 then the average grade also gets multiplied by 10.

B. **Measures of Dispersion:**

1. The second *central moment* of a distribution is the **variance**.
2. As with the mean, we can calculate variance for a sample of historical data as

$$Variance = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

3. But, since we are finance people we are always looking to the future and assigning probabilities to future events. So, we may use the formula:

$$\sigma^2 = p_1(x_1-\bar{x})^2 + p_2(x_2-\bar{x})^2 + \ldots + p_n(x_n-\bar{x})^2 = \sum_{1}^{n} p_i(x_i-\bar{x})^2$$

4. If we define a *surprise* as a deviation from the expected value, then we can interpret the variance as a measure of the expected surprise. If we have a sample of data, the estimate of σ^2 is denoted as s^2.

NOTE: If a constant, k, is added to a random variable, the variance *does not change*. That is, if Var(x) = σ^2, then Var(x+k) = σ^2. Can you see why? Also, if a random variable is multiplied by a constant, the variance is multiplied by the square of the constant. That is, if Var(x) = σ^2, then Var(x*k) = $k^2\sigma^2$. Work through the last term in the variance formula to prove this to yourself.

5. A problem with the variance is that it gives the expected surprise in squared units, i.e. if we are dealing with returns; the variance is in percent squared. This is annoying to try and interpret. To get around this annoyance the standard deviation is used.

$$Standard\ deviation = \sqrt{Variance}$$

The standard deviation gives the expected surprise in the same units as the mean. By definition, the standard deviation is unaffected by addition of a constant to the random variable and is scaled directly by a constant that is multiplied to the random variable. That is *Stddev(x*k)* = $|k|\sigma$.

6. Another measure of dispersion that has become popular in finance recently is the **coefficient of variation**. This is just the ratio of the standard deviation to the expected value.

$$\mathbf{CV} = \frac{\sigma}{E(x)}$$

   a) This CV ratio gives the relationship between the standard deviation and the expected value.
   b) The bigger the CV ratio, the more variation relative to the expected value which is useful in comparing different assets.

7. Two other measures of dispersion
   a) The **range** is simply the difference between the maximum and minimum value of the sample. This measure of dispersion is highly susceptible to outliers.
   b) The **interquartile range** is the range of the central 50% of the data and is used primarily in box plots.

C. **Higher Order Central Moments:**

1. **Skewness** is a measure of symmetry and is derived from the third central moment.

$$\textit{Skewness Coefficient} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^3}{s^3} \text{ or } \frac{\sum_{i=1}^{n} p_i\left(x_i - \overline{x}\right)^3}{s^3}$$

2. A distribution is symmetric if it has the same shape on both sides of the mean.
3. The larger the skewness coefficient the less symmetric the distribution.
4. The sign of the coefficient represents the direction the skewness, e.g. if the coefficient is positive then more of the distribution lies to the right of the mean.

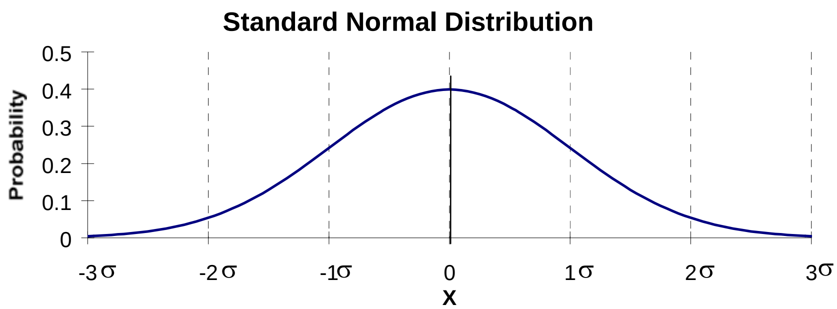
5. Why would a risk averse investor care about skewness?
    a) Most investors prefer small negative surprises and large positive surprises. The coefficient of skewness can help investors find assets with return probability distributions having these characteristics.

Risk

I. What Does Risk Look Like?

A. Random events in nature often occur with probabilities that follow a **normal distribution**. This is an extremely nice thing given the properties of the normal distribution.

1. A normal distribution is a symmetric distribution centered on its mean with approximately 96% of the distribution falling within two standard deviations of the mean.
2. What does all that mean?!? Look at the picture.



a) This distribution is symmetric with the same shape on both the left and right sides of the solid line (the mean) in the center (the skewness coefficient is zero).
   (1) A distribution is centered on the mean if the tallest part of the distribution is over the mean.
   (2) This also implies that for normal distributions, the mean equals the mode and the median. Here, the vertical solid line is the mean.

b) The dotted lines are the one and two standard deviations from the mean. It is easy to see that most of the distribution is within two standard deviations of the mean.

3. Another nice aspect of the normal distribution is that two parameters, the mean and the variance can describe the entire distribution. Look up the formula for the density function for a normal distribution if you don't see this (It is on page one of this review).
4. The truly convenient property of this distribution is that a linear combination of normal random variables is also a normal random variable. This property is called ***stability*** and makes this distribution ideal for dealing with portfolios and testing statistical hypotheses.

B. Conceptually, this figure above represents the probability of an event occurring.

1. Suppose that historically a stock return has a mean of 10% and a standard deviation of 20%.
   a) In other words, on average the stock price has been 10%. We would put 10% right below the solid vertical line.
   b) Since this is a graph of probabilities, what does the picture tell us?
      (1) It tells us that since the highest point of the curve is over 10%, that we will see a stock return of 10% most of the time. It is the most probable return that will occur for this stock.
      (2) The fact that most of the distribution lies within two standard deviations from 10%, tells us that 96% of the time we will observe a return within the 2σ (10% ± 40%) dotted lines.
      (3) We are less likely to observe a price the closer we get to the outermost dotted lines because the curve gets lower the further away we go from the mean.

C. The ***risk*** of an individual asset return comes down to the variability.

1. Since standard deviation is a measure of variability around the mean, large standard deviation in an asset return implies that the dotted lines in the above picture are far away from the mean.
   a) For example, say we have two stocks, both with a historical mean price of $100. And assume that the stock prices are distributed normally.
      (1) Stock A has a standard deviation of $5 and stock B has a standard deviation of $10.
      (2) So 96% percent of the time we will see a price that falls in between $90 and $110 for stock A, but for stock B the price will be between $80 and $120.
      (3) Stock B's price is more variable - has a higher standard deviation - and is therefore, more risky.

   b) Here is another example. Assume that the price of this stock is distributed as a normal random variable.

Price of stock A	Probability
$40	.10
$50	.20
$60	.40
$70	.20
$80	.10

E(P) = 40(.10) + 50(.20) + 60(.40) + 70 (.20) + 80(.10) = **$60**

σ^2 =.10(40-60)^2+.20(50-60)^2+.40(60-60)^2+.20(70-60)^2+.10(80-60)^2 = **120**

σ = **$10.95**

CV = 10.95/60 = **.183**

Skewness = [.10(40-60)^3+.20(50-60)^3+.40(60-60)^3+.20(70-60)^3+.10(80-60)^3]/(10.95)^3 = **0**

So, we *expect* the price of this stock to be $60 and 96% of the time we will see the price of this stock between $38.10 and $81.90. The skewness coefficient tells us that the distribution is perfectly symmetric around the expected value. The CV implies that the standard deviation of this stock is about one-fifth of the expected return. This may not be too bad depending on what other investments are available.

D. It is often more convenient deal with rates of return when talking about asset valuation. Here is the same example in rates or return given the price of stock A today is $50. Remember the formula for the rate of return? **r = (FV - Price) / Price**

Price of stock A	Rate of return %	Probability
$40	-20%	.10
$50	0%	.20
$60	20%	.40
$70	40%	.20
$80	60%	.10

And we have a new probability distribution in terms of rates of returns, calculated using the same formulas we used when dealing with prices. Check my answers.

	Stock A
Expected Return	20%
Standard Deviation	22%
CV	1.1
Skewness	0

Other Stuff:

Other concepts, which will be indispensable when we build a portfolio of assets, are the idea of the mean, variance, and covariance of two or more random variables. If the individual random variables are assumed to be normally distributed, then creating a new random variable as a weighted (linear) combination of the individual random variables also has a normal distribution. This comes from the stability property mentioned above.

For example, consider two stock returns (two normal random variables). If we form a portfolio such that we put half our money in each asset, then the expected return of the portfolio (the new normal random variable) can be calculated as: $E(R_p) = .5 * R_1 + .5 *R_2$.
More generally, the expected return is found as:

$$E\left(R_p\right) = \sum_{i=1}^{n} w_i E\left(R_i\right)$$

What's the risk of the portfolio (standard deviation of the new random variable)?

To calculate the **portfolio standard deviation** we cannot simply take the weighted average of the two standard deviations. That would fail to take into account the way the stocks move out of phase with each other.

The variance of a linear combination of two random variables is found using the formula:

$$\boldsymbol{\sigma^2_p = w^2_1\, \sigma^2_1 + w^2_2\, \sigma^2_2 + 2\, w_1\, w_2 \sigma_{12}}$$

What's this σ_{12} term? This is the covariance between the two returns. It is a direct measure of the relationship between the two returns; how the returns move in relation to each other.

Calculate the covariance with the following formula:

$$\sigma_{12} = \sum_{i=1}^{n} p_i \left[R_{1i} - E\left(R_1\right)\right]\left[R_{2i} - E\left(R_2\right)\right]$$

or

$$\sigma_{12} = \frac{1}{\left(n-1\right)} \sum_{i=1}^{n} \left[R_{1i} - E\left(R_1\right)\right]\left[R_{2i} - E\left(R_2\right)\right]$$

NOTE: If a constant is added to one or both of the random variables, the covariance *does not change*. If one or both of the random variables is multiplied by a constant, then the covariance is also multiplied by those constants. That is, *Cov(a+bx,c+dy) = b*d*Cov(x,k)*, where *a, b, c*, and *d* are constants.

Covariance is like variance in that it is scaled by the units of the variables involved. So often we will use correlation, which is a measure of the movements between two variables free of any scaling. Correlation will always fall between -1 and 1. If when one stocks return goes up, the other goes down we have negative correlation. If when one stocks return goes up, so does the other then we have positive correlation. We define correlation as follows:

$$\rho = \frac{\sigma_{AB}}{\sigma_A \sigma_B}$$

Regression Analysis:

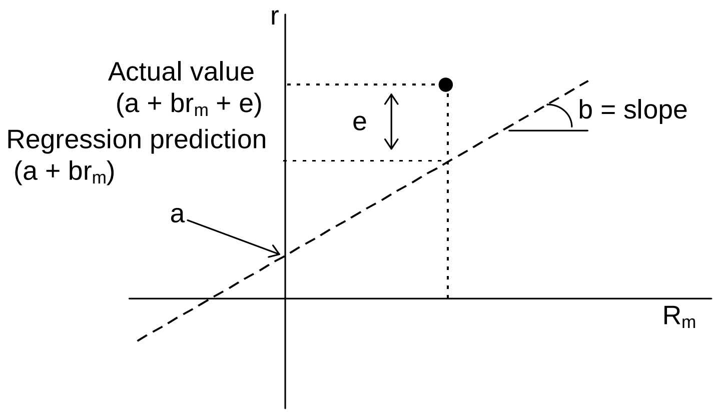
Regression analysis is an extremely popular and powerful tool for exploring causal relationships among random variables. Suppose "theory" tells us that any asset return in excess of a risk free interest rate can be described by the linear function

$$r_t = a + br_{m,t} + e_t$$

where r_t is the excess asset return at time t,
a is a constant intercept term,
b is a slope coefficient,
$r_{m,t}$ is the excess return on the market index such as the S&P 500, and
e_t is a random error term which is uncorrelated with $r_{m,t}$ and has an expected value of zero.

We will spend a great deal of time developing this model, but for now let's just accept this as a possible relationship between any asset return and the market index.

What the equation says is that the relationship between any asset return and the market index can be described by a line with intercept a and slope b plus some normally distributed error. If we know the terms a, b, and e we have a graph like the following:



More specifically, if we take the expected value of the equation to get $E(r_t) = a + bE(r_{m,t})$, the equation gives a direct relationship between the expected return of an asset and the expected return of the market index. We just need the coefficients for a and b. We can estimate the parameters a and b using a regression. The explained portion of the line is the regression prediction $(a + br_m)$. The difference between the actual value and the regression prediction is the error term e. If we have a time series of returns on the asset and the market we can estimate the relationship using ordinary least squares (OLS). OLS simply produces the line that gives the minimum distance between the line and the actual values. That is, *OLS minimizes the sum of squared error terms*. Mathematically, OLS is a simple optimization problem. Statistically, OLS is identical to estimating the parameters in a maximum likelihood problem when the error terms are assumed to be distributed normally with mean zero. From a user's standpoint it is the line of best fit. If you are interested in the mathematical and statistical derivations of OLS, I have many books in my library that you are welcome to review.

The regression equation is more than just a relation between the expected values of the asset and market returns it also relates the variances of the two. Taking the variance of both sides of the regression equation yields:

$$\sigma^2_{asset} = b^2\sigma^2_m + \sigma^2_e$$

The term $b^2\sigma^2_m$ is known as the explained variance and the term σ^2_e is the unexplained variance.

We can also derive the covariance and correlation between the asset return and the market index.

$$
\begin{aligned}
\textbf{Cov(r, r}_{\boldsymbol{m}}\textbf{)} &= Cov(a + br_m + e, r_m) \\
&= Cov(br_m, r_m) \\
&= bCov(r_m, r_m) \\
&= b\sigma^2_m
\end{aligned}
$$

$$\Rightarrow \boldsymbol{b = Cov(r, r_m) / \sigma^2_m}$$

Hence, the slope is a measure of the co-movement between the asset and the market as a fraction of the explanatory variable - the market.

A question that immediately arises when using OLS is how well does the explanatory variable "explain" the dependent variable? In this case, how well do movements in the market explain movements in the asset returns? To answer this question we can look at the amount of variance explained. The coefficient of determination, better known as R^2, is the ratio of explained variance to total variance.

$$\boldsymbol{R^2 = (explained\ variance)/(total\ variance) = b^2\sigma^2_m / \sigma^2_{asset}}$$

The coefficient of determination will take on values between zero and one. Values closer to one indicate that more of the total variance is explained and that the explanatory variable does a good job determining the value of the independent variable. As more explanatory variables are included in the regression, R^2 will increase simply by construction. To account for this increase many regression programs give an adjustment to R^2.

Suppose we are interested in testing a hypothesis concerning one of the estimated regression coefficients. For instance, theory states that the intercept term should be zero in our regression equation if we have used excess returns to estimate the equation parameters. Typically the estimate of *a* will not be exactly zero. Thus, we would like to devise a test to see whether *a* is statistically different from zero. The test we would like to perform is called the null hypothesis and is denoted as H_0. The values we think *a* may take if the null is not true is called the alternative. In this case

H_0: $a = 0$
H_a: *a* is not 0.

To test any hypothesis about a coefficient like the one above we can use a t-test. The statistic is formed by subtracting the value we are testing for from the coefficient and then dividing by the estimated standard error of the coefficient. In our example we have

$t = (a - 0) / s(a)$

where s(*a*) is calculated from the standard deviation of the error terms and the standard deviation of the coefficient *a*. Most computer packages for regression calculate the standard error for each coefficient.

Once we have the t statistic we can compare the value to a table of probabilities to check significance. That is we check to see if $t > t_{\alpha, T-k}$ or $t > t_{\alpha/2, T-k}$. Where T is the number of observations in the data set, k is the number of explanatory variables and α is the level of significance.

If a test with a 95% level of confidence is desired, then calculate the level of significance α, as 1 - .95 = .05. Whether we use α or $\alpha/2$ depends on whether the alternative is one sided or two sided. In the example the alternative is two sided since *a* can be greater than or less than 0. A one sided test would imply that the alternative was something like H_a: $a < 0$. While all this might seem a little confusing there is a good "rule of thumb" to use when dealing with t-tests of coefficients. If the test is like the one we have above, H_o: $a = 0$ and a two sided alternative, then if the t statistic we calculate is greater than 2 we can safely reject the null.