1) *The scale of the t-Distribution is inversely related to the* ***chi-squared distribution****. Recall that the chi-squared distribution has mean of n, where n is the degrees of freedom. If the degrees of freedom increase, will it be more or less likely that the variable with t-Distribution becomes extremely large?*

As degrees of freedom increase, outliers will be increasingly rare in a t-Distribution, because the chance that all those Zi in the chi-squared distribution (which determines the inverse variance of the t-distribution) are simultaneously zero becomes small.

Jonathan found a link to explain this: http://stats.stackexchange.com/questions/110359/why-does-the-t-distribution-become-more-normal-as-sample-size-increases.

Emiliana pointed out the relevance of t-distribution to copulas, which are also used to model tail dependence. Ruppert p. 186 demonstrates using copulas that Gaussian distributions do not have tail dependence, but t-distributions can have (depending on the value of the degree of freedom and the correlation coefficient.) In fact, a precise definition of tail dependence is given in p. 185.

A simple understanding of copulas is this: they are an extension of correlation matrices to distributions that are non-Gaussian and have higher moments.

Herberto shows

Suppose that $Z \sim N(0,1)$Z~N(0,1), and $X_n \sim \chi^2(n)$Xn~χ2(n), then a random variable $T_n \sim t(n)$Tn~t(n)can be though of as the following quotient

$$T_n = \frac{Z}{\sqrt{\frac{X_n}{n}}}.$$

Since $X_n \sim \chi^2(n)$Xn~χ2(n), this means that there exist $Z_1, \ldots, Z_n$Z1,<,Zn, n standard normal random variables such that $X_n = Z_1^2 + Z_2^2 + \ldots + Z_n^2$Xn=Z12+Z22+...+Zn2, and because of this $\frac{X_n}{n}$ is the average of the random variables $Z_1^2, \ldots, Z_n^2$Z12,<,Zn2

Last week we were asked to show that $E[X_n] = n$E[Xn]=n and $Var(X_n) = 2n$Var(Xn)=2n so now we can get that

$E\left[\frac{X_n}{n}\right] = 1$ E[Xnn]=1 and $Var\left(\frac{X_n}{n}\right) = \frac{2}{n}$Var(Xnn)=2n,

because of this fact the Weak Law of Large Numbers and the Continuous Mapping Theorem, we can show that as the number of degrees of freedom n increases, T_nTn will converge in probability to the standard normal distribution.

More interestingly it can be shown (using Chebyshev's inequality, and conditional probability) that as n increases

$$Pr(|T_n| \geq k) \approx Pr(Z \geq k)$$

2) *For a single stock, we believe that the t-Distribution is a more realistic model for its returns than a normal distribution. Do you remember why?*

Stock returns exhibit **left** fat-tail, implying non-zero skew and kurtosis. Normal distribution has neither.

2.1) Former student Steven found out that "…, it was recently proven that the t distribution provides one of the better fits for modelling returns of S&P 500 stocks. You can read more about that here: http://arxiv.org/abs/1305.4173"

2.2) Former student Amy wrote: "I read an excellent paper which discusses the multivariate t-distribution vs the multivariate normal distribution. The article states that the t-distribution is far superior because it can 1) best the normal distribution in special cases and 2) it can still capture the observed fat tails. Furthermore, their paper shows the rejection of the normality assumptions and acceptance of the t-distribution as it passes standard skewness and kurtosis tests. Ref: Kan, Raymond; Zhou, Ghofu. (2003). Modeling Non-normality Using Multivariate t: Implications for Asset Pricing."

2.3) T-distribution converges to the normal distribution as the degree of freedom goes to infinity. See the proof at http://courses.ncssm.edu/math/Stat_Inst/Stats2007/Stat%20and%20Calc/t%20Converges%20to%20N.pdf

2.4) Former student Joseph asked: " is there a t-distribution analog of Brownian motion? I'm continually suprised how much normal time series processes are still relied on."

My response: "The essential characterization of *log* stock prices as a random walk ("Brownian motion") is not much affected by the assumption of t-distribution. In the usual model of stock prices, the returns are normally distributed, and therefore the prices are log-normally distributed. Here, we model returns with t-distribution instead, and the prices have log-t (if there is such a term!) distribution. But in both cases, the log prices exhibit a random walk, which means it is not stationary, and has a variance that goes to infinity with time. The only question is whether the variance goes linearly with time or not. Based on the central limit theorem, we can say that if any distribution of returns has a finite variance, then indeed the random walk will have variance that scales linearly with time, and in fact indistinguishable from a price series with underlying normal distribution of returns. On the other hand, if we find that the distribution of returns actually does not have finite variance (e.g. can be fitted to a t-distribution with degrees of freedom <= 2), then the variance may not scale linearly with time. I have not worked out what the dependence on time is in the latter case, but maybe you can find out!"

2.5) Former student John pointed out that tail dependence of returns undermines the diversification motivation behind creating the portfolio.

3) *For a portfolio of multiple stocks, the multivariate t-Distribution is a more realistic model for predicting the returns of all these stocks than a multivariate normal distribution. The reason it is more realistic goes beyond that given in 2). Do you know what it is?*

Tail dependence of stock returns can be modeled by t-distribution, but not normal distribution. Tail dependence means that at times of extremely negative or positive returns, stock returns are highly correlated, and thus making the overall portfolio returns even more extreme. But during times of small returns, they can be uncorrelated. See Ruppert p. 158 (bottom paragraph) – p. 159, especially Figure 7.4.

3.1) Steven also added this: "Empirical studies show that multivariate normal distributions are not suitable for asset returns. Tests against the Fama/French model portfolios indicate that kurtosis levels are > 3, violating what is expected of a normal distribution. A multivariate skewness test also strongly rejects the assumption of normality under the same test conditions. You can read further about this study here: http://apps.olin.wustl.edu/faculty/zhou/KZ-t-06.pdf "

3.2) An application of t-distribution to a trading strategy can be found here: research.stlouisfed.org/wp/2001/2001-021.pdf

3.3) Former students Mark and Alex raised a good point: What is the distribution of the returns of a portfolio of stocks whose returns follow a multivariate t-distribution? This is the same as asking whether a linear combination (or weighted average) of variables with t-distribution is itself t-distributed. The answer is **No**: the resulting distribution is complicated. Please see http://arxiv.org/pdf/0906.3037v1.pdf and www.jstor.org/stable/2286298.

3.4) Former student Jason asked a good question: "The whole concept of tail dependent was very well explained in Ruppert and makes logical sense. Due to this fact, is this a good reason to have stop loss points? To cut off those large risk moves earlier?"

For a mean-reverting strategy, it is indeed necessary to impose an artificial stoploss to cut off these tail events to prevent them from wiping out the equity. But interestingly, these tail events are often especially profitable to momentum strategies. That is because once prices trend in a certain direction, a momentum strategy will automatically adopt a position in the same direction as the trend (possibly reversing a previously losing position with the opposite sign), and if the trend continues to become a large tail event, that position will be extraordinarily profitable.

4. Former student Jeff has a good explanation of the Maximum Likelihood Estimation procedure that we can use to find the optimal parameters of any parametric distribution given a set of sample data. In the homework, this method is illustrated by finding the optimal parameters of a t-distribution. I have edited his response as follows:

We have a bunch of returns that we do not believe are normally distributed. We could have a million of them, and no matter what, they have a much "fatter tailed" distribution than normal. So what we do is say "Lets assume the distribution is a t distribution" (because a t distribution has fatter tails than a normal distribution.) If that is the case, we can … start with a "best guess" for mu1, mu1, var.a, var.b, cov.ab, and DF, and then uses the optim function to find their best values …Simply, we are finding the values of the parameters so that the shape of the t distribution best fits the actual returns. The best fit is defined as one that gives the maximum log likelihood value of the distribution function given the sample data as input.

5) Former student John linked to a chart and an article to assist us in picking the right distribution: http://www.johndcook.com/distribution_chart.html, and
: http://pages.stern.nyu.edu/~adamodar/New_Home_Page/StatFile/statdistns.htm

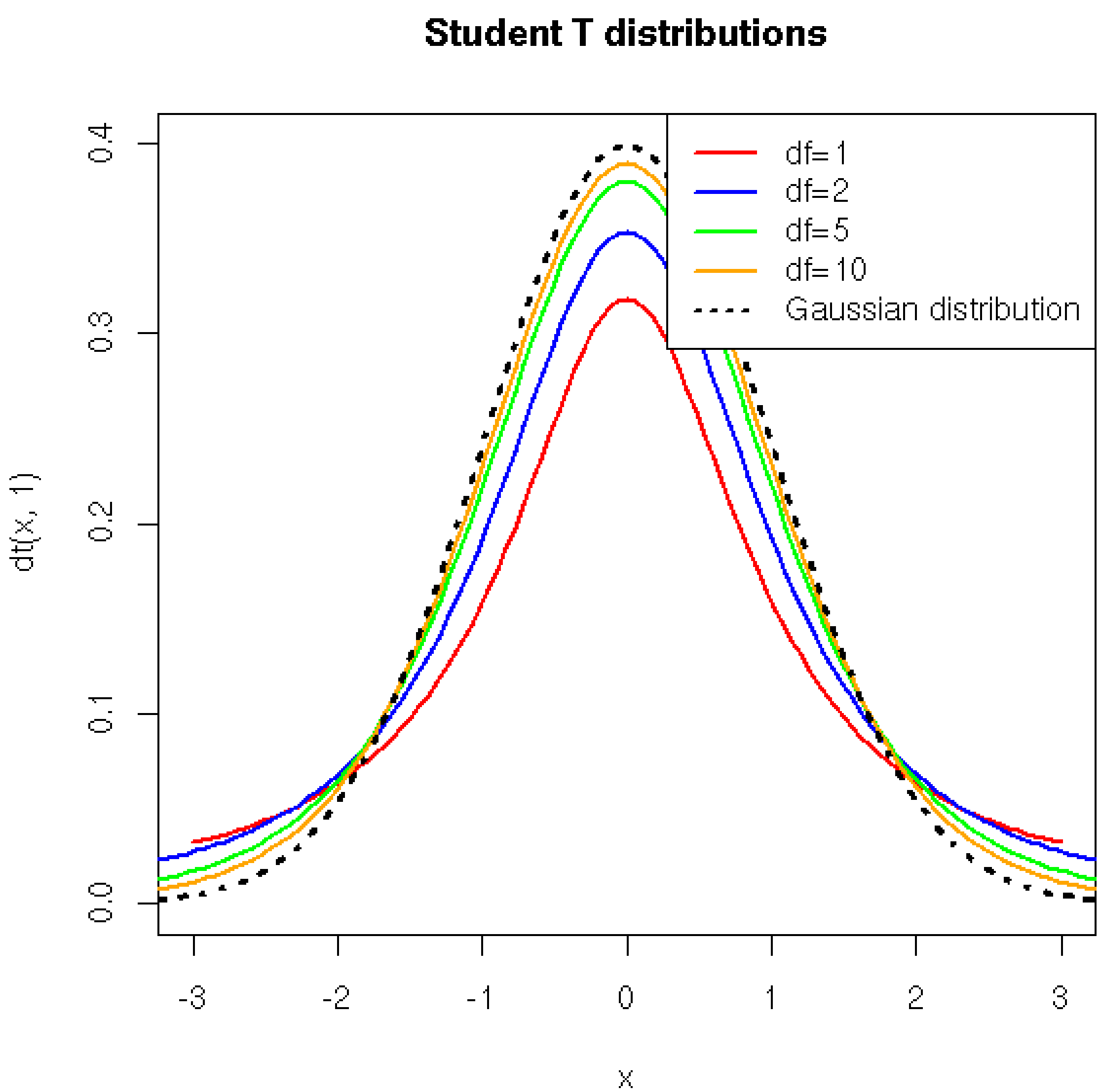
6) Justin shared his codes to demonstrate some of these observations:

```
# Just sharing the work I did in R to make sure I understood how the
# degrees of freedom impacted a T distribution
# if you get the error
#  "some 'x' not counted; maybe 'breaks' do not span range of 'x'
# just run it again.  I fixed the scale of the x axis and sometimes we get
# values bigger than I specified.

# I started by implementing eq 5.11 "Z/(W/v)^.5" (Ruppert, 2011, p.88)

# histogram of 10000 values from equation 5.11 with 5 degrees of freedom
hist(rnorm(10000,0,1)/(rchisq(10000,5)/5)^.5,breaks=seq(-25,25,1))

# histogram of 10000 values from equation 5.11 with 10 degrees of freedom
hist(rnorm(10000,0,1)/(rchisq(10000,10)/10)^.5,breaks=seq(-25,25,1))

# histogram of 10000 values from equation 5.11 with 100 degrees of freedom
hist(rnorm(10000,0,1)/(rchisq(10000,100)/100)^.5,breaks=seq(-25,25,1))

# histogram of 10000 values from equation 5.11 with 1000 degrees of freedom
hist(rnorm(10000,0,1)/(rchisq(10000,1000)/1000)^.5,breaks=seq(-25,25,1))

# as the degrees of freedom increase the tails thin out and the distribution
# approaches a normal distribution
hist(rnorm(10000,0,1),breaks=seq(-25,25,1))

# this is just a self check to ensure that my implementation of the t
# distribution matches the one in R
hist(rnorm(10000,0,1)/(rchisq(10000,5)/5)^.5,breaks=seq(-25,25,1))
hist(rt(10000,5),breaks=seq(-25,25,1))

# "The variance of a tv is fininte and equals v/(v-2) if v > 2" (Ruppert, 2011, p.89)
var(rnorm(10000,0,1)/(rchisq(10000,5)/5)^.5)
5/(5-2)

var(rnorm(10000,0,1)/(rchisq(10000,10)/10)^.5)
10/(10-2)

var(rnorm(10000,0,1)/(rchisq(10000,100)/100)^.5)
100/(100-2)

var(rnorm(10000,0,1)/(rchisq(10000,1000)/1000)^.5)
1000/(1000-2)

# References
```

Ruppert, David. (2011). Statistics and data analysis for financial
engineering. New York: Springer

7) A classical t-distributed variable ϵ can only have variance of 1 if the degree of freedom is infinite, in which case it becomes a normally distributed variable. If $\lambda\epsilon$ has variance of 1, and ϵ has a degree of freedom υ, then λ =sqrt((υ-2)/υ).

8) Student Jordan generated this chart to illustrate “As the degrees of freedom increase, the variable’s t-distribution approaches a normal distribution. Since t-distributions have fattier tails than a normal distribution, a variable with a t-distribution will be less likely to become extremely large.”



9) Student Robert asked: “how confident one can be about defining the df parameter for the t-distribution. If I understand correctly, in this context we are really using df to help us define the correct shape of the pdf for a given series of returns, it has less to do with the actual degree of freedoms that we would be defining in a traditional statistical sense, such as a regression coefficient t-test. So in that case, if we're really just using df as a lever to back-into the best approximation for the shape of historical returns, how confident can we be that our MLE is really going to stand the test of time; i.e. how do we know it is the best approximation for the entirety of possibilities (including those yet to be experienced), as opposed to just what we've seen thus far?”

I answered: “Indeed there is no reason to suppose that the t-distribution is the best fit for the returns data. Could it be that the Cauchy distribution is better? Or the Levy distribution? What about the Pareto distribution? Using MLE, one can try the best fit (the best parameters) with each distribution in turn, and see which has the highest likelihood. The model that has the highest likelihood is the best fit. Of course, whether MLE is a suitable method for finding best fit for a fat tail distribution is itself a subject of intense academic debate.

Nassim Taleb has been especially eloquent on fat tailed distributions: http://www.fooledbyrandomness.com/FatTails.html and so did Mandelbrot (the "father" of fractals) http://web.williams.edu/Mathematics/sjmiller/public_html/341Fa09/econ/Mandelbroit_VariationCertainSpeculativePrices.pdf.”

10) Robert also asked: “I have another question for you which is something I've never really understood about quant finance: how do practitioners get comfortable that historical statistical patterns are going to repeat themselves in the future (in the context of a given company and/or even a market or economy on the whole)?

As a fundamental market analyst my perspective is that the future is in fact quite unknown and ever-changing. Here are a few examples quickly off the top of my head as I write this: Amazon can make almost no money for 15 years and then seemingly flip the switch and report huge earnings, Greece is solvent for the entire existence of the Eurozone and then almost overnight loses market access, Phones4U (a UK example) maintains great wholesale relationships and runs quite a satisfactory business but one day out of nowhere goes bankrupt as 2 large suppliers pull their business, etc. I see examples like this all the time...to me they are more the rule than the exception.

So I've never quite understood the reliance that quants place on parametric market models. I presume the answer is that in the above I've selected a few extreme examples, but in most cases the historical experience well informs future developments and the majority of less extreme cases dull the minority of extreme results...this is essentially the law of large numbers. But I feel like there are extreme results all the time, and you never know which company/market/economy is going to experience a similar shock. It would be great to get your thoughts on this as the course unfolds. For me it is one of the major underlying concepts in this area of finance.”

Student Braden’s answer: “I'm not a quant in equities, but I'd imagine they aren't trying to model idiosyncratic risk for individual companies, since as you say above, those are extreme events and most of the time, no one has any prior knowledge, because if the market did, it would have been priced in long before the risk event occurred.

Without much knowledge of how quants work in equities, I would think I'd be categorizing a lot of factors and bucketing the universe of stocks into those categories. Then, I'd model off my cube of stratified samples on all these different factors. within each box in the cube there may be plays in value or momentum. I'd probably model off of that and build a portfolio with as much diversification as I can while still taking strategic tilts when I felt it optimal.

To the macro side of things, again, I doubt that anyone will know the next event that sets the market tumbling. I guess a few did in the credit crisis as is commonly known from "The Big Short". However, for macro, I don't think you're modeling for a specific risk, just knowing that whatever the event that sets the market falling, the market returns are in the left tail for awhile. So, maybe that distribution is useful."

My answer: "Certainly the stability of historical statistical distribution of returns is a big assumption. However, many of the examples you quoted concern individual stocks, where the assumption of stability of distribution is almost certainly wrong. Most of quant finance rely on the stability of distributions of the average instrument, not that of a specific stock. As quantitative investment models make bets on a large number of instruments, the law of large number took over as you pointed out. We will for sure lose money on a small number of instruments, but they should not wipe us out. They are counted as "specific risks", which are diversifiable. It is factor risk that we are most afraid of: factor risk is undiversifiable, and can indeed wipe out a portfolio. The distinction between specific and factor risks will be discussed in Module 9. But a quick example of factor risk is this: suppose one has a value model, betting that value stocks will appreciate faster than growth stocks. If the market is in a bubble where growth stocks appreciate faster, then we will incur large losses across whole portfolio.

Other more sophisticated financial models can assume dynamic evolution of the distribution of returns, taking into account current macroeconomic variables such as interest rate, oil prices, etc. But ultimately, even human traders depend on past experience to guide their experiences, and quantitative models are no different."

11) Scott asked "Is it ever appropriate to exclude extreme market events when fitting distributions for modeling?" to which Jonah gave an insightful reply:

"I would say that is highly situation dependent, but in general would argue that in most cases involving financial variables, the extreme events are often the most salient. As an actuary, a primary concern of mine is unexpected large losses due to various risks that my company assumes. We actually take an approach that is almost the opposite of what you mentioned - we isolate and model only the tail of the distribution. There is much literature on Extreme Value Theory and the application of such models as the Generalized Pareto or variations of the Weibull family that you may find interesting."