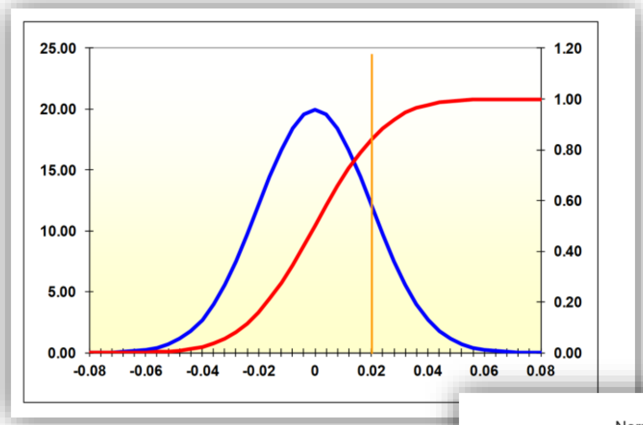


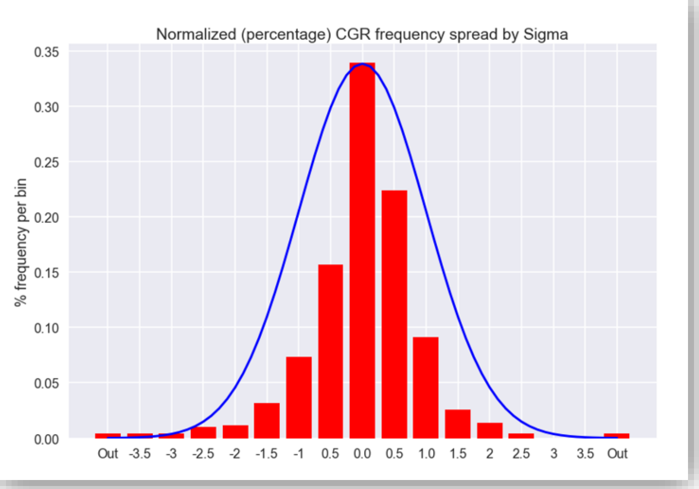
Economics 136
Harvey Mudd College

Fundamental Modeling Assumptions about

Stock Price Behavior over time ...

Part 1

Volatility and measures of risk



(again, how hard can it be??)

Topical: NFLX earnings ...

JAN 24 '20 (2 DAYS) | JAN 31 '20 (9 DAYS) | FEB 07 '20 (16 DAYS) | FEB 14 '20 (23 DAYS) | MORE

TABBED VIEW | All STRIKES | SMART | NFLX | 100 | IV: 2.6

CALLS											PUTS									
CHANGE %	VOLUME	OPTN OP...	HIGH	LOW	LAST	ASK SIZE	BID SIZE	ASK	BID	STRIKE	BID	ASK	BID SIZE	ASK SIZE	LAST	LOW	HIGH	OPTN OP...	VOLUME	CHANGE %
-65.75%	318	201	13.90	7.50	8.00	7	27	7.95	7.75	322.5	2.06	2.16	3	7	2.18	1.01	4.60	1.22K	1.63K	-71.16%
-71.34%	1.06K	632	12.70	5.80	6.20	20	100	6.35	6.10	325	2.90	3.10	8	168	3.10	1.43	5.50	2.42K	5.60K	-62.79%
-76.64%	2.05K	271	10.50	4.65	4.70	1	79	4.90	4.70	327.5	4.00	4.10	6	2	4.00	2.01	6.96	601	4.34K	-57.04%
-80.48%	7.86K	1.77K	9.15	3.55	3.63	1	1	3.70	3.55	330	5.30	5.45	2	6	5.35	2.72	8.10	2.41K	7.48K	-48.01%
-84.65%	5.54K	676	7.50	2.66	2.66	2	2	2.68	2.60	332.5	6.80	7.00	82	4	6.95	3.61	9.53	957	2.57K	-39.72%
-88.36%	7.01K	1.92K	6.15	1.95	1.87	1	3	1.95	1.86	335	8.55	8.80	9	49	8.50	4.75	11.25	3.02K	7.84K	-33.39%
-91.01%	4.37K	2.08K	4.90	1.33	1.33	8	7	1.40	1.31	337.5	10.50	10.70	4	1	10.55	6.02	13.05	1.40K	821	-24.59%

By Joe Flint and Micah Maidenberg

Updated Jan. 21, 2020 9:04 pm ET

SAVE PRINT TEXT 9

Netflix Inc. missed its forecast for U.S. subscriber growth for the third straight quarter, but blew through its expectations for overseas expansion, a mixed performance that comes as the streaming giant faces heightened competition from a gaggle of rivals.

The Los Gatos, Calif., company said Tuesday that it added 423,000 domestic subscribers in the fourth quarter, compared with its forecast of 600,000 additions. It also posted an increase of 8.3 million subscribers in overseas markets, more than the seven million the company was expecting. It now has 167 million subscribers world-wide, including 60.4 million in the U.S.

We decided Tuesday against taking a directional bet (the 335 Jan 24 Call at about $12) because we don't do directional bets. Good thing .. NFLX earnings were tepid and the call collapsed to less than $2, and a strangle would have lost hugely as well.

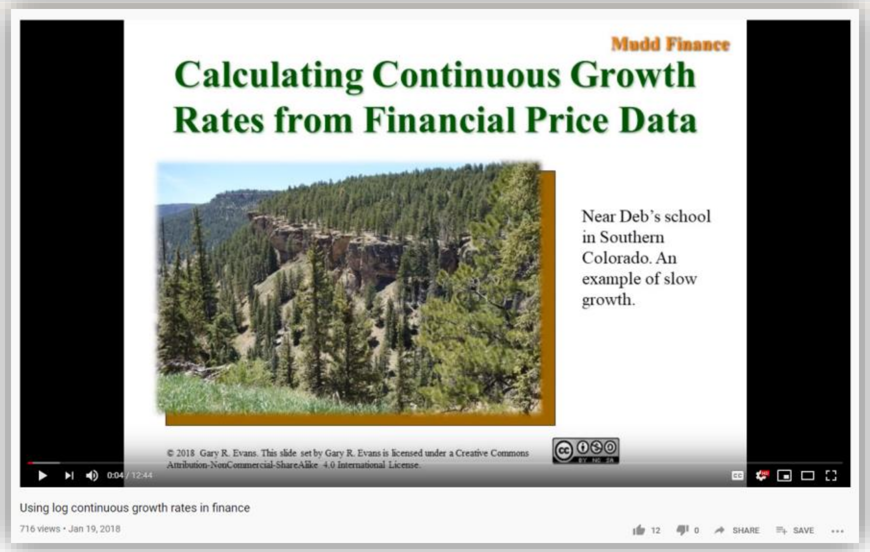
Remember these assignments for next week:

E136 Course Outline | E136 Course Calendar | E136 Support Material | Prof E's Courses

Dated Homework and other Assignments for Spring 2020.

Check this frequently! You should consult this page at least once per week to make sure that you are staying up with the class. Assignments, dated by when they should be completed, are listed below in order of assignment:

- By Tuesday, January 28, watch Downloading free stock and ETF data from IEXCloud and try downloading some data.
- By Tuesday, January 28, watch Using log continuous growth rates in finance (2018).



The relevance of this ...

Asset price volatility, such as volatility in the stock market, is a major source of risk. We must try to find reliable measures of risk if we want to minimize risk while targeting yields or other measures of performance from financial asset portfolios.

We often have access to time-series data (in other words, the history of the data series) and we can use common statistical techniques to find useful patterns of information in that data. When we have historical series like the graph shown to the right, whether of stock indexes, individual stocks, yields on bonds, or futures and options values, we can assess risk up to a point.

Common sense tells us that we are likely to make some use of measures of dispersion like variance or standard deviation ***and then gradually refine it.***

Common assumptions made about the price performance over time of primary assets and their derivatives

- (For purposes of mathematical ease and because historical data conform to this assumption *within limits*), since the time of Black, Scholes, and Merton, prices paths and their growth rates are assumed to be continuous.
- The price behavior of a financial asset (FA) is independent of it's past price behavior (Markov Chain)
  - also referred to as a "random number walk"
  - this is very debatable, but is the basis for a lot of modern modeling
  - this reduces the attraction of so-called "technical analysis" and "charting."
  - was this done to make the math models, like Black-Scholes, work, or because it is true?
  - is not meant to imply that the price of a share of stock is unrelated to its price the previous day – the previous close is the launch point for today.
- The past price behavior of a FA may be filtered in a way that gives some reliable indicator of the risk associated with the FA.

... more assumptions

- The previous assumption implies that we can use historical time series data for FAs to ***partly*** estimate measures of their risk (caveat: the past doesn't always repeat itself).
- We typically assume that the **rates of return** for FAs can be represented as random variables that conform to a normal Gaussian probability distribution
  - ... which further typically implies that the raw data from which the rates of return were calculated conform to a asymmetric distribution like lognormal (to be shown later).
  - ... and this assumption requires that when working with raw time series data that it be converted to continuous log growth rates before risk estimates are made.
  - ... and at some point, this assumption may be put to a test, like the Kolmogorov-Smirnov normalcy test.

The pure stock behavior model ..

- Assumed or implied in traditional options pricing models
- Easy to demonstrate with Monte Carlo simulations
- Has a lot of empirical weight
- Easy to model with Python and kind of fun too

A stock price over time follows geometric Brownian motion, where the stock price follows Brownian motion with drift:

$$dP_t = \mu P_t \, dt + \sigma P_t d\varepsilon_t$$

where:

μ	drift (daily log continuous growth rate, for us, having a value like 0.00812)
σ	volatility (for us, standard deviation of daily log continuous growth rate)
ε	Brownian motion (selecting from a Gaussian draw)

continued …

The differential equation solved (which has little meaning now but will matter later, and we will come back to it when it does):

$$P_t = P_0 e^{\left[\left(\mu - \sigma^2/_2\right)t + \sigma\varepsilon_t\right]}$$

Uncomfortable? See *Wikipedia's* treatment of "Geometric Brownian Motion" and related discussions.

which implies:

$$ln\frac{P_t}{P_0} = \left(\mu - \sigma^2/_2\right)\mathrm{t} + \sigma\varepsilon t$$

This derivation requires the use of *Ito's Formula*.

and where t equals 1:

$$ln\frac{P_{t+1}}{P_t} = \left(\mu - \sigma^2/_2\right) + \sigma\varepsilon$$

... an added note that you may not quite understand now, but will be useful later on (and we will return to it)

If μ is being drawn from a Gaussian (normal) distribution, then the solution P_t is a log-normally distributed random variable with

This expression here is very important ...

$$E(P_t) = P_0 e^{\mu t}$$

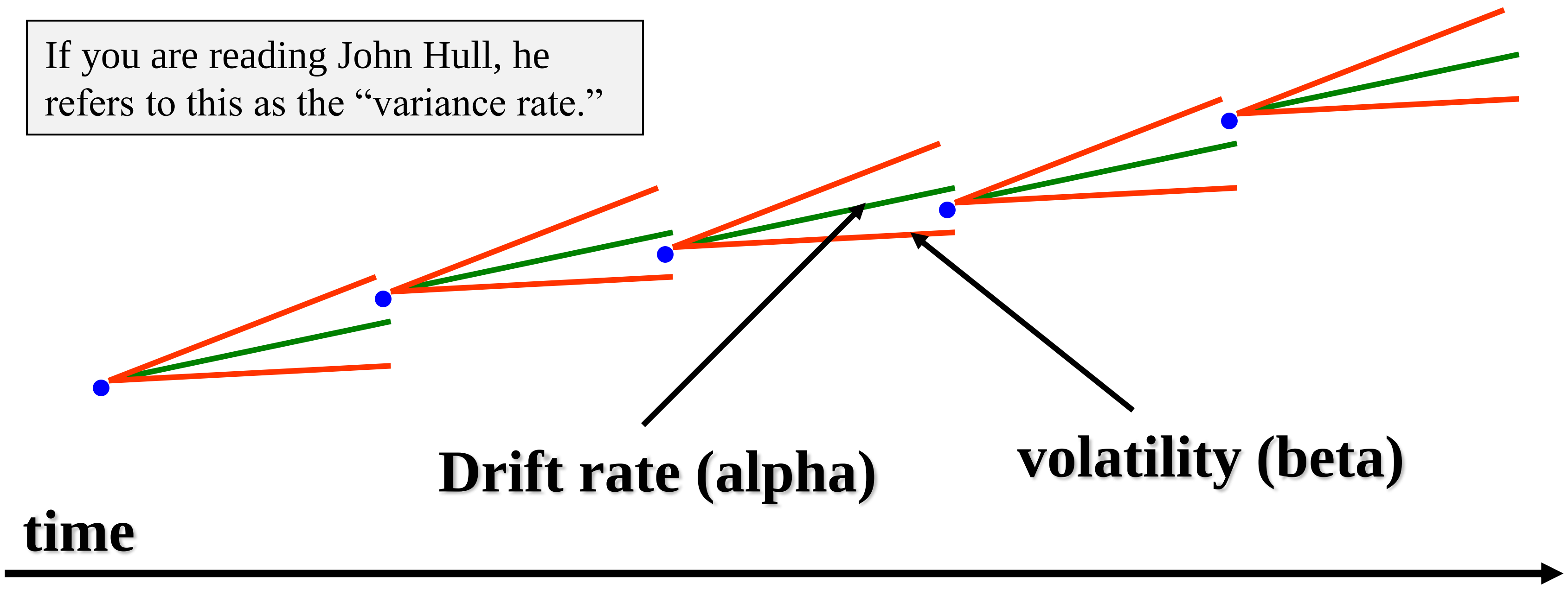
$$Var(P_t) = P_0^2 e^{2\mu t}\left(e^{\sigma^2 t} - 1\right)$$

$$SD(P^t) = \sqrt{Var(P_t)}$$

About drift and volatility

We are going to regard the path of stock prices as Geometric Brownian Motion (a Markov Process) of log growth rates reflecting ***drift*** and ***volatility***, where the latter is represented by a Gaussian (normal) distribution. The resulting pattern will reflect randomness with a trend. Here is a way of visualizing that.

If you are reading John Hull, he refers to this as the "variance rate."



Example: A Monte Carlo Simulation ...

$$P_{t+1} = P_t e^{\left(\mu - \sigma^2/_2\right) + \sigma\epsilon}$$

The drift term

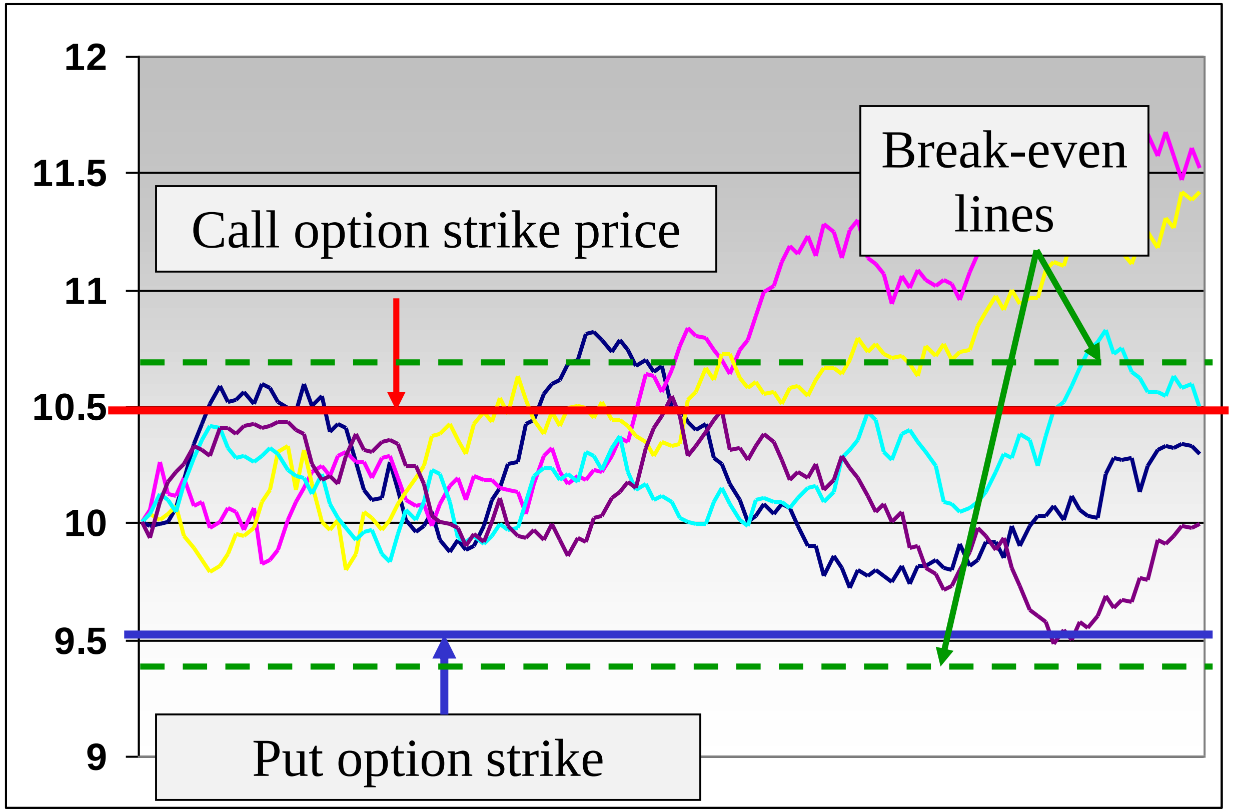
The volatility term

This is our gambling game: We have a special die. It has a Gaussian distribution with a mean μ and a standard deviation σ. At step "t" in our world, we role the die. Then we take the result of our roll, make that the power of an exponential (adjusted for half-variance)* and then multiply that times the value of P (price) at time t (now). Then we do it again, and again.

ε refers to a random selection from a standard normal probability distribution (mean of zero, variance of 1) and that is multiplied times our standard deviation.

*explained (justified) later .. I promise.

Monte Carlo Simulation of a Strangle



In this simulation suppose the stock is trading at 10 and we want to do a 6-month strangle at strike prices of 10.5 (call) and 9.5 (put). The stock has to go above or below these strike prices but we also have to cover our option costs (green lines).

Here we don't have to wait until expiration and normally wouldn't. If we did, two of these make money, one has value but we lose money, and two expire worthless. What clearly matters? Volatility.

A later, more realistic version of this model will include a Poisson distribution.

Therefore our typical first steps in empirical work:

Note: You have been assigned a video that explains this ... and you are expected to know this.

Given some *sample* from a *population* set of prices, such as the daily closing price of CSCO from Jan 2019 to Dec 2019, a typical first step is to convert the data to continuous growth rates:

$$CGR_{px} = R_p = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

for each paired observation.

In Excel:

Sample	Price	CGRP
1	22.34	
2	22.86	0.0230
3	23.01	0.0065
4	22.79	-0.0096
5	23.41	0.0268
6	23.56	0.0064

Original sample (in part)

Converted data

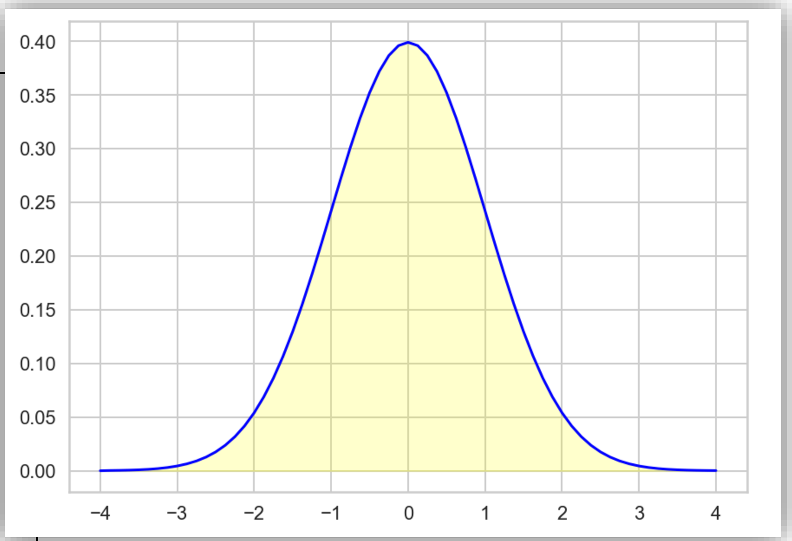
... see the dedicated video that discusses this in detail.

More about the assumption of the normal distribution

So far we have set up a model that assumes that the daily log continuous growth rate of a stock, calculated (typically) from OHLC data, fits a normal Gaussian distribution.

We can explore other distribution later if we want.

For the moment, though, let's explore the properties of our normal distribution assumption and what it might imply about the behavior of our stock and its option derivatives.



Elementary estimates of historical yield (*mu* – called *alpha* in some contexts) and risk (*sigma* – also called *beta*) for our normal distribution

Using the transformed continuous log price growth rates from a previous slide (CGR_i), which here we shorten to R_i we calculate the mean growth rate, which is our *mu* (drift) estimate …

$$\overline{R_p} = \frac{\sum_{i=1}^{n} R_i}{n} = \mu_p$$

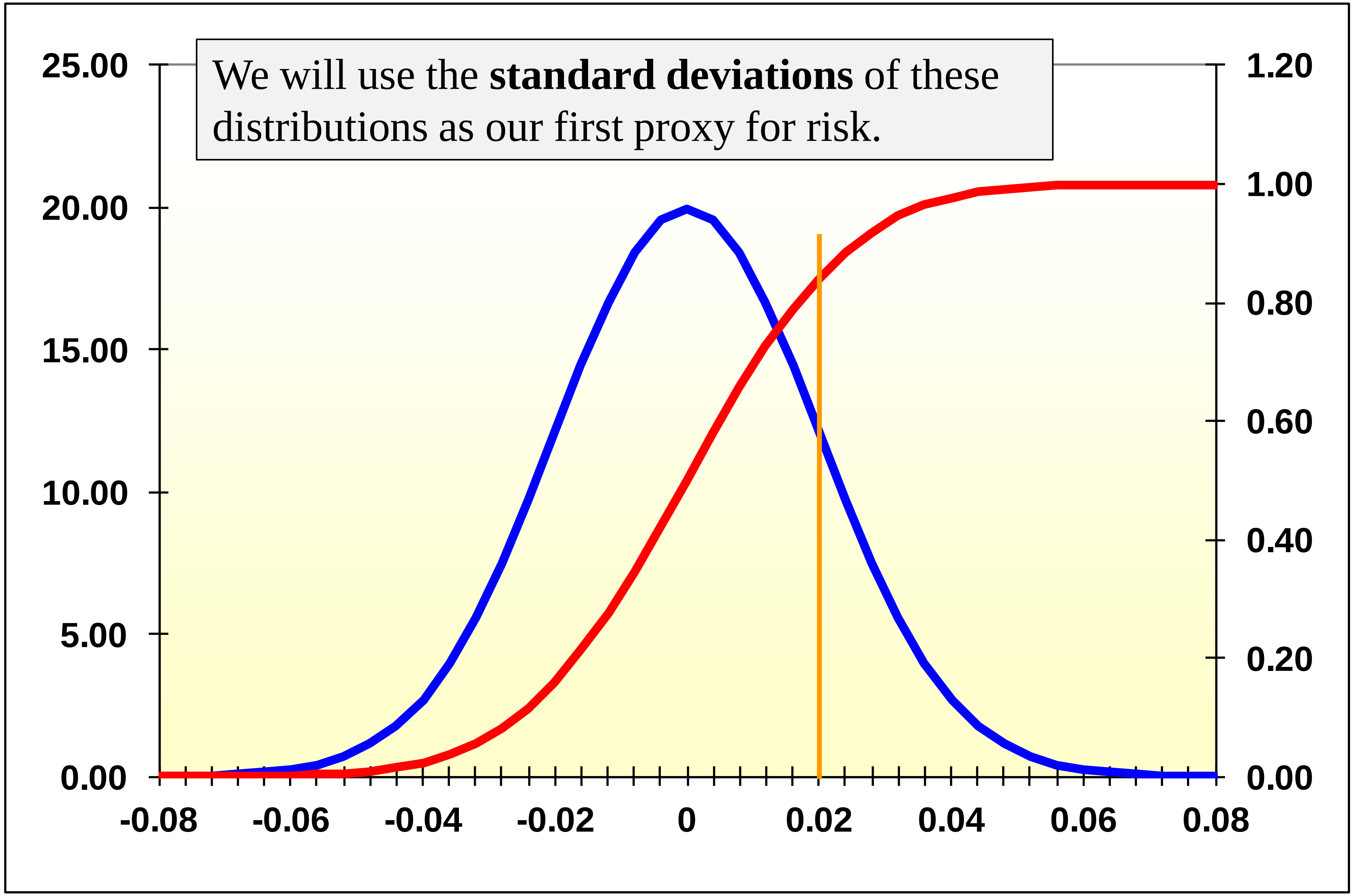
… then we calculate the variance of the same …

$$\overline{V_p} = \frac{\sum_{i=1}^{n} \left(R_i - \overline{R}\right)^2}{n} = \sigma_p^2$$

... and finally, the square root of variance, the standard deviation, is our *beta*, or risk proxy.

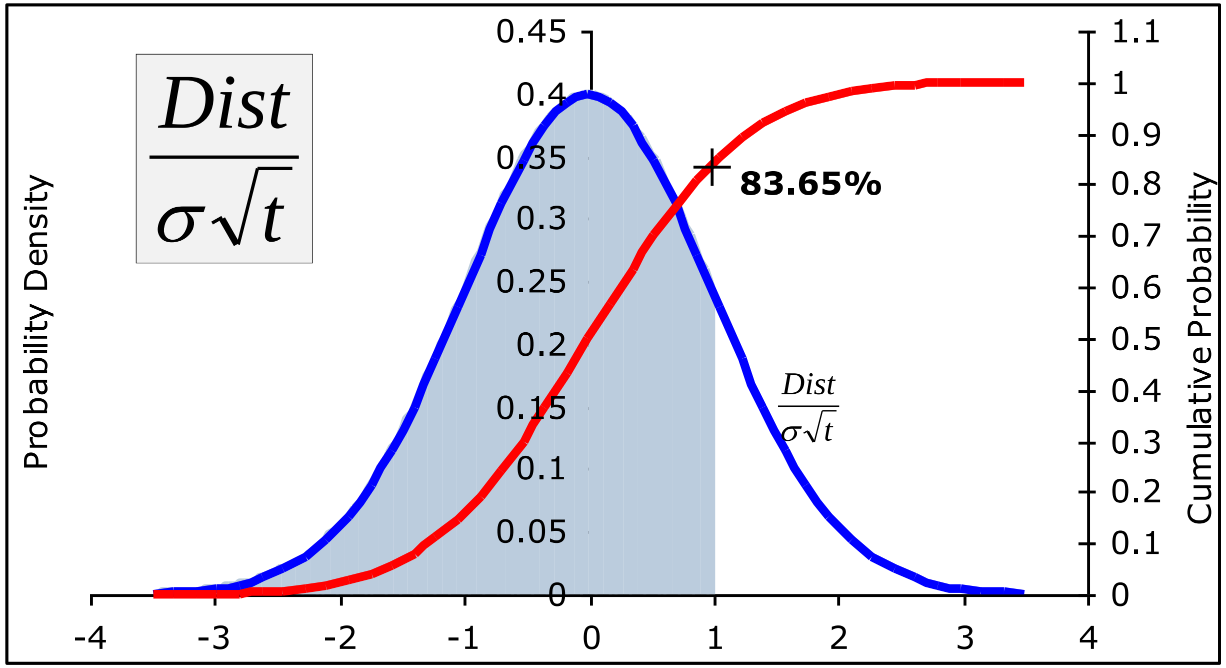
$$\overline{SD_p} = \sqrt{\overline{V_p}} = \sigma_p$$

The assumed probability distribution of FA continuous growth rates (or similar): Gaussian (normal)



The Standard Normal Distribution

which we will use a lot



σ	Π
-3.0	0.0013
-2.5	0.0062
-2.0	0.0228
-1.5	0.0668
-1.0	0.1587
-0.5	0.3085
0.0	0.5000
0.5	0.6915
1.0	0.8413
1.5	0.9332
2.0	0.9772
2.5	0.9938
3.0	0.9997
-1 to 1	0.6826

Dividing a normal distribution with mean 0 by its standard deviation produces the standard normal distribution, where we can describe the probability of a number being X standard deviations away from its mean. Shown is the probability of a value being less that +1 SD.

Memo slide: The Normal Distribution

Gaussian (Normal) Probability Density Function

NORMDIST[X,MEAN,SD,FALSE]

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

Probability Cumulative Distribution Function

NORMDIST[X,MEAN,SD,TRUE]

$$F\left(x;\mu,\sigma^2\right)=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{x}e^{-(x-\mu)^2/\left(2\sigma^2\right)}dx$$

or

$$F\left(x;\mu,\sigma^2\right)=\frac{1+errorfunction\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)}{2}$$

More memo:

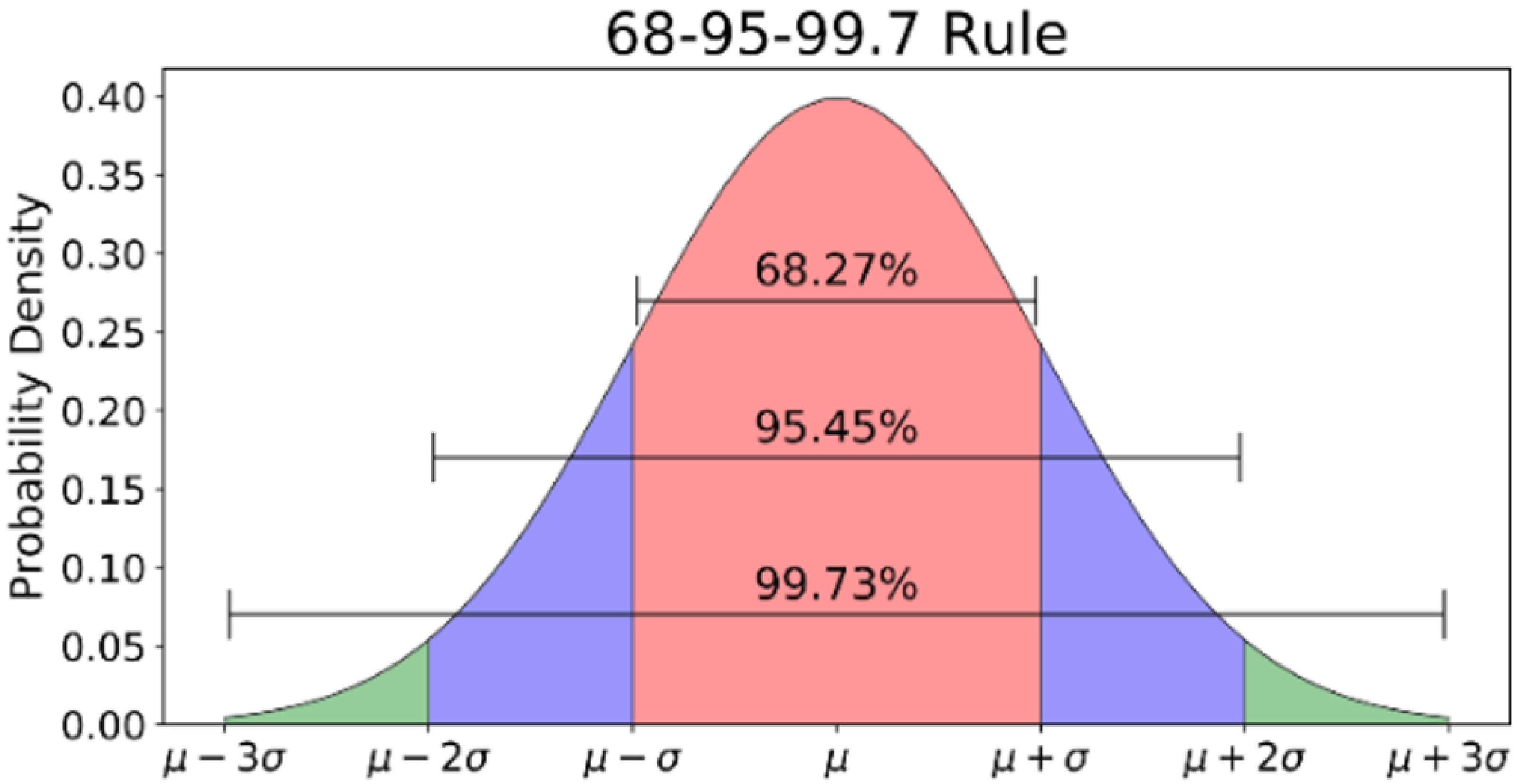
The Taylor Series expansion of the error function is:

$$ef(z) = \frac{2}{\sqrt{\pi}} \sum_{i=0}^{\infty} \frac{(-1)^n z^{2n+1}}{n!(2n+1)} = \frac{2}{\sqrt{\pi}} \left(z - \frac{z^3}{3} + \frac{z^5}{10} - \frac{z^7}{42} + \cdots - \cdots \right)$$

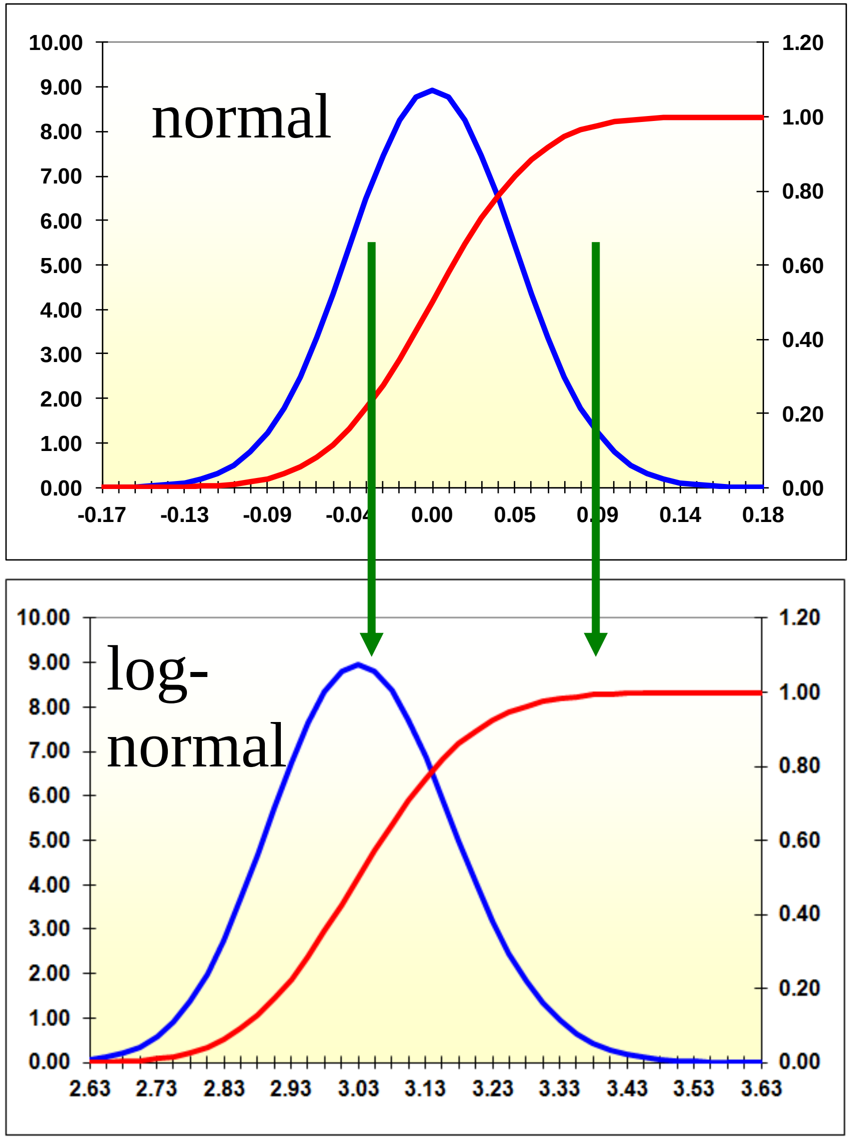
1.772454		PWR	Z PWR	Prod
		X	1	
		Mean	0	
		SD	1	
		Z	0.707107	
1	1	1	0.707107	0.707107
-1	3	3	0.353553	-0.11785
1	5	10	0.176777	0.017678
-1	7	42	0.088388	-0.0021
1	9	216	0.044194	0.000205
-1	11	1320	0.022097	-1.7E-05
1	13	9360	0.011049	1.18E-06
-1	15	75600	0.005524	-7.3E-08
1	17	685440	0.002762	4.03E-09
-1	19	6894720	0.001381	-2E-10
				0.605018
		erf(Z)		0.682689
		CumP		0.841345

You don’t have to go very far in the series before convergence and this is trivial to code.

The 68-95-99.7 rule ...



A very userful property: When we transform our normal distribution back to stock prices, the resulting distribution is log-normal!



Using an example from a later lecture:

How is this transformation being made (assuming the current price of the stock is \$3.03 per share)?

We center the new distribution at

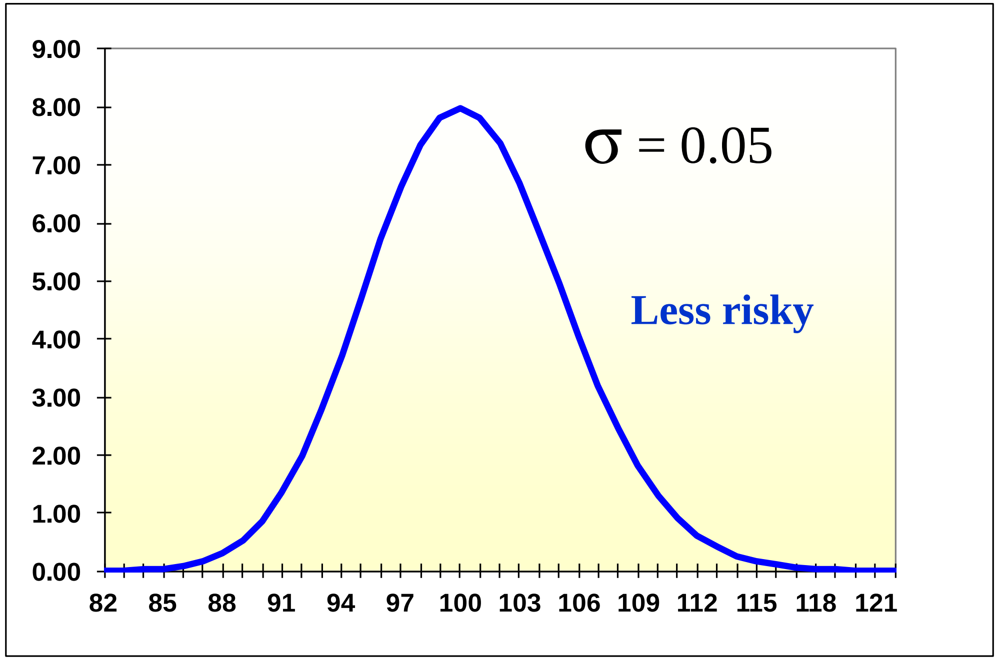
$$P_0 = 3.03e^0$$

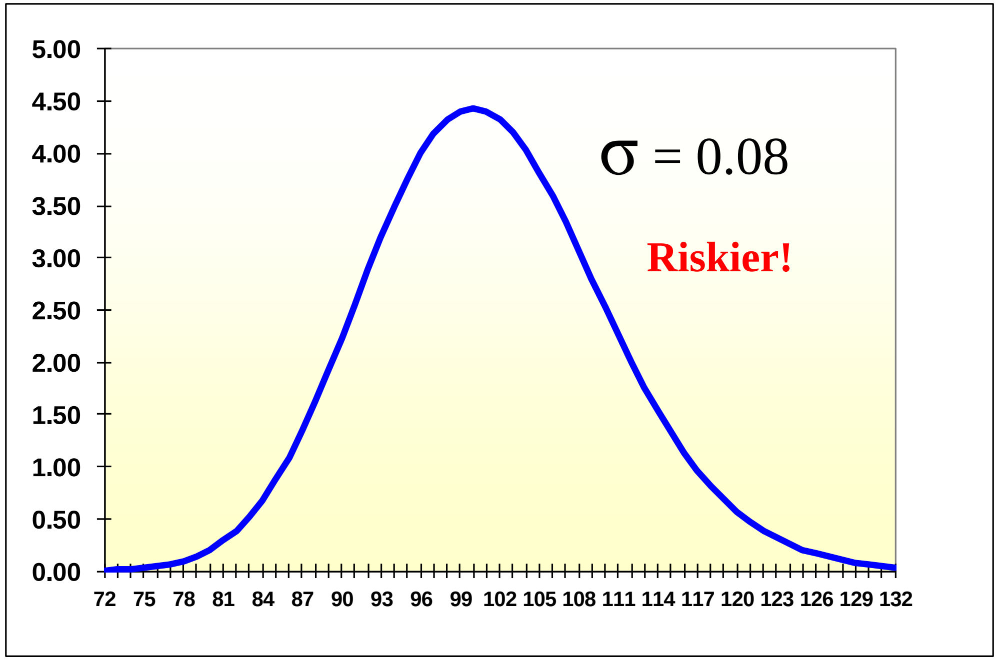
and each price point is plotted as

$$P_i = 3.03e^{r_i}$$

which is going to result in a log-normal distribution.

How do we represent relative risk?





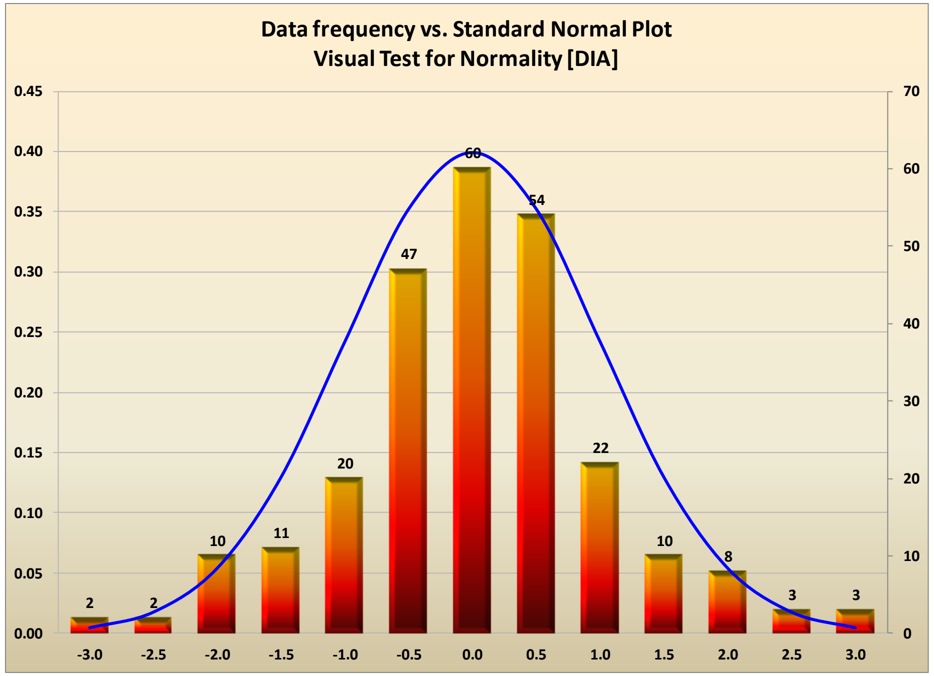
The log-normal distributions on the right are transformed from normal daily continuous growth distributions, with the same mean (zero), but the top has a standard deviation of 0.05 and the bottom has a standard deviation of 0.08.

The bottom has a wider dispersion, so although there is a higher probability of a great gain, there is also a higher probability of a great loss. That is seen as “riskier.”

Using a histogram …

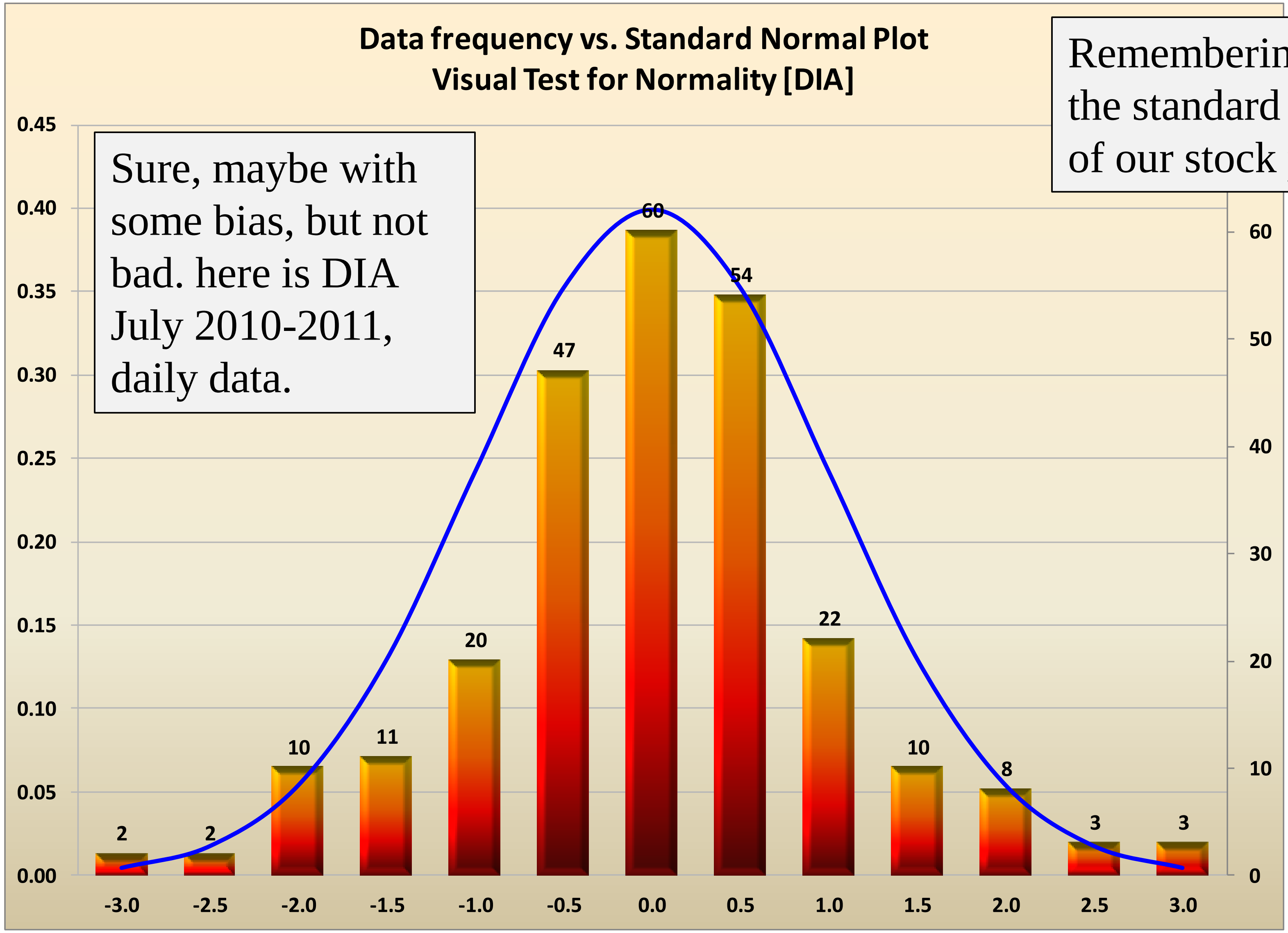
We can take our daily continuous growth rate data and, among other things, arrange it into histograms like the one shown on the right (the one shown here is in Excel, but we will be doing ours in Python).

A histogram divides the full range of your data into an arbitrary range of odd-numbered intervals of equal size, like 11 in our example.



Then the program counts the frequency of observations for each interval (such as 60 for the center interval in the diagram shown) and maps each interval as a bar. In our model we overlay the histogram with a mapping of the continuous density distribution taken from our estimate of the mean and standard deviation of the same data, assuming a normal distribution. The comparison shows how “close” we are to normal, and also identifies any anomalies like outliers. The Gaussian fit can also be checked with a normalcy test, like Kolmogorov-Smirnov or even Chi-squared.

Is anything actually going to fit? ...



Sure, maybe with some bias, but not bad. here is DIA July 2010-2011, daily data.

Remembering from Part 1 that we are using the standard deviation of the log growth rate of our stock price as our risk proxy ...

Using the old Excel model we used to use in this class ...

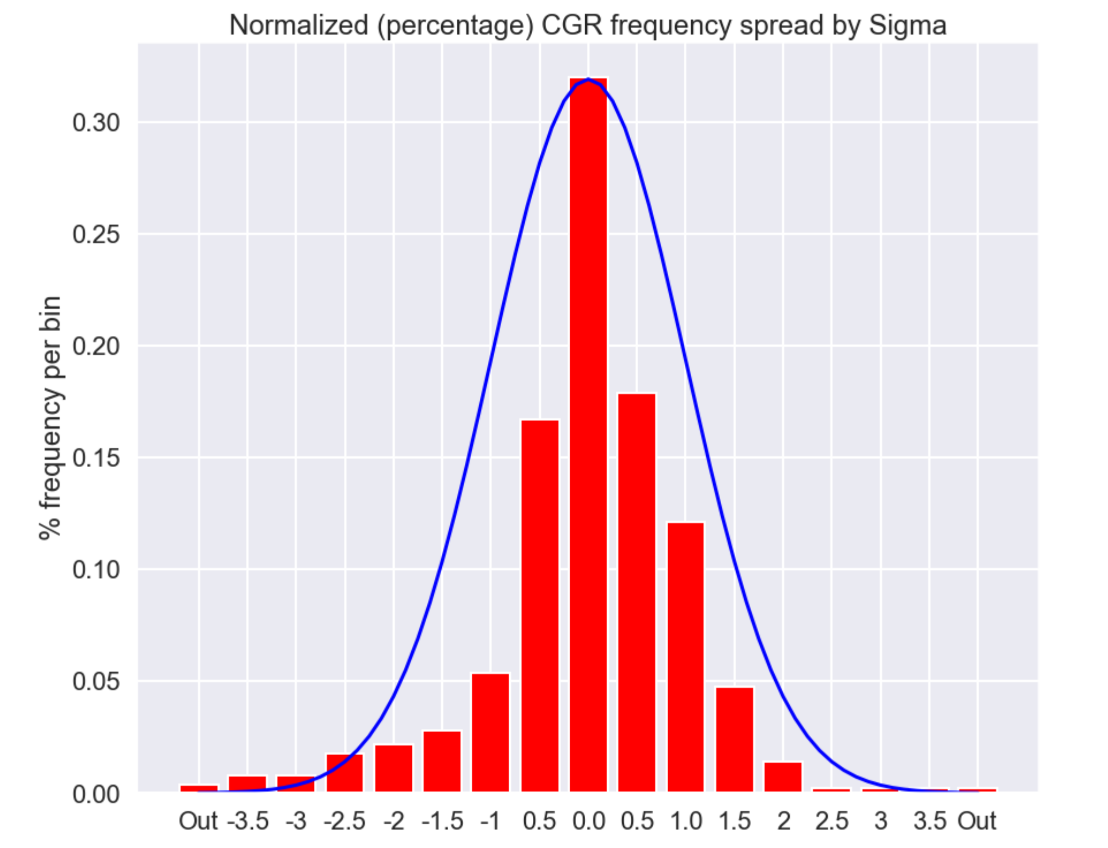
How “normal” can we expect our converted data to be? Evaluation by histogram: SPY

Example:

Here we are using Python ...

- SPY
- Daily data
- 1/22/2018 start
- 1/21/2020 end
- n = 503
- DCGR = 0.00032
- 1 year σ = 0.00743

	Drift	Volatility	qVolatil	Sharpe	Maxsigma	Minsigma
2 year	0.00032	0.00949	0.00757	0.03332	5.16303	-4.53728
1 year	0.00086	0.00743	0.00651	0.11627	2.77257	-4.22392
90 day	0.00109	0.00557	0.00557	0.19642	2.21811	-3.39796
30 day	0.00200	0.00437	0.00437	0.45745	1.67315	-2.19726



Dense around the center with “thin flanks and fat tails.”

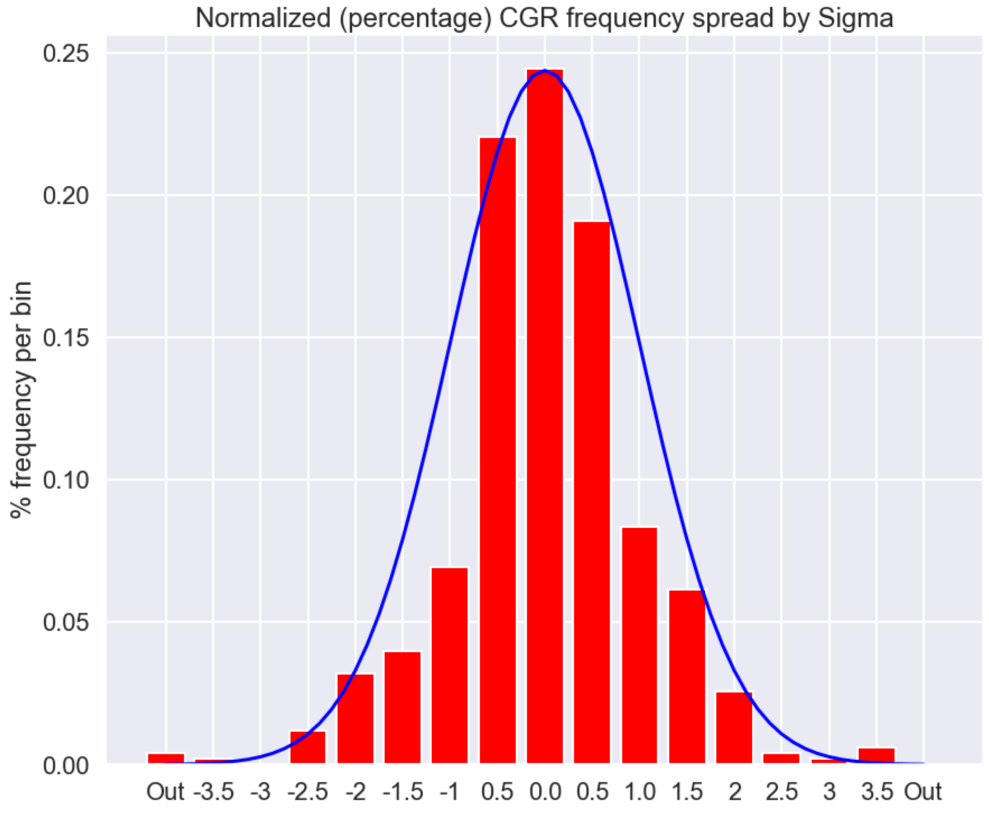
How "normal" can we expect our converted data to be?
Evaluation by histogram: NFLX

Example:

- NFLX
- Daily OHLC data
- 1/22/2018 start
- 1/21/2020 end
- n = 503
- DCGR = 0.00079
- 1 year σ = 0.02024

Much higher than SPY!

	Drift	Volatility	qVolatil	Sharpe	Maxsigma	Minsigma
2 year	0.00079	0.02570	0.02271	0.03069	3.67093	-4.24870
1 year	-0.00001	0.02024	0.01869	-0.00057	2.40221	-5.35378
90 day	0.00177	0.01970	0.01854	0.08996	2.30999	-3.31413
30 day	0.00367	0.01741	0.01741	0.21083	1.87892	-2.02037

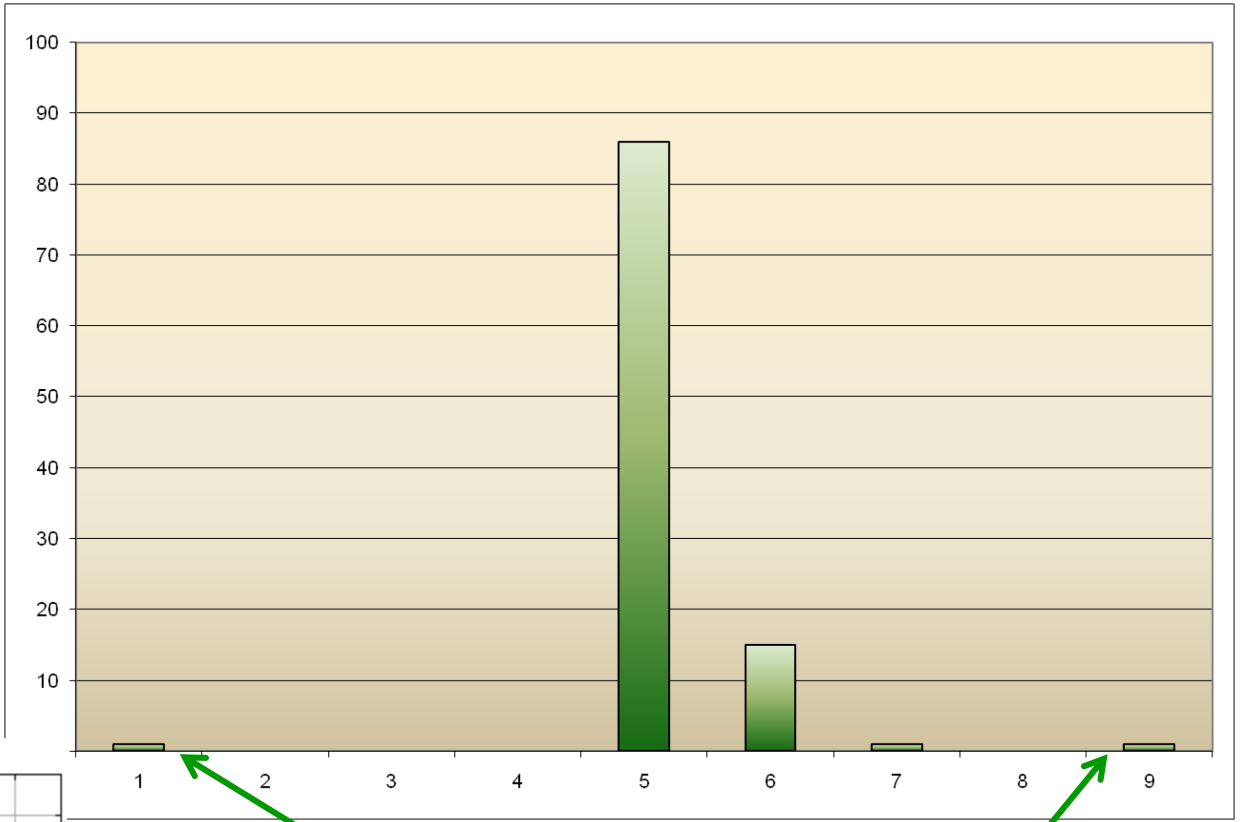


Also dense around the center with "thin flanks and fat tails."

Does everything fit? Always?? NO, we have to check ... and I love using this example ...

Example:

- DNDN
- Weekly data
- 8/15/05 start
- 8/6/07 end
- n = 104
- MWGR = 0.0024
- $\sigma = 0.17315$



... but this also helps explain why we can't get rid of outliers! This isn't GPS engineering!

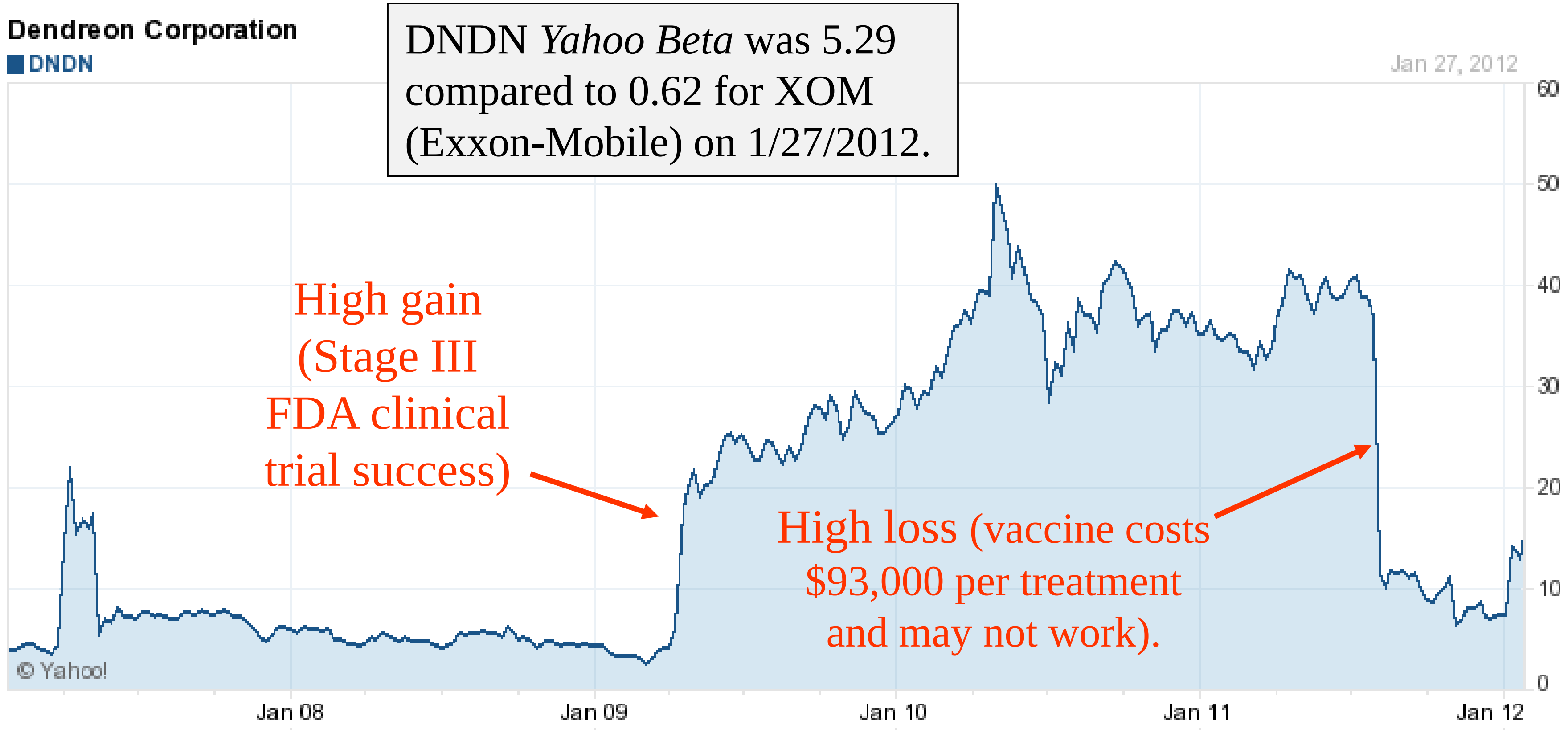
Getting rid of the **outliers** means getting rid of this.

Example of a high risk, high gain / high loss stock

… and a non-normal stock

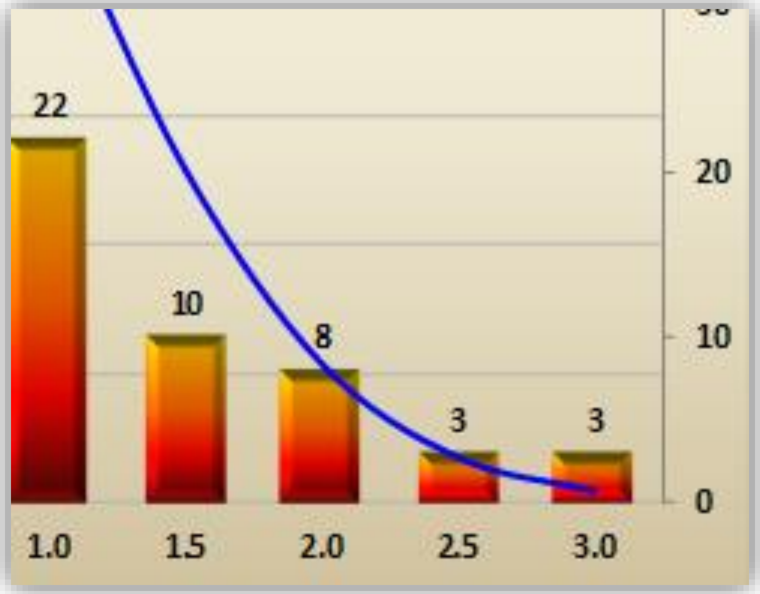
Dendreon (**DNDN**), a biotechnology research company who makes ***Provenge***, an experimental and controversial prostate cancer vaccine.

DNDN *Yahoo Beta* was 5.29 compared to 0.62 for XOM (Exxon-Mobile) on 1/27/2012.



Source: *finance.yahoo*

Introduction to “tail risk” [kurtosis]

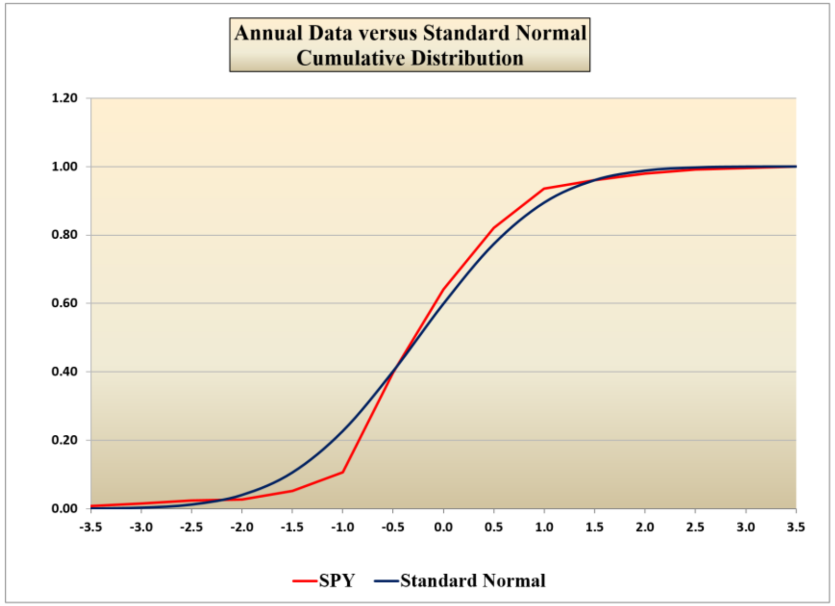


This segment from an earlier DIA histogram shows a typical fat tail bias, where the historical frequency of observations outside of 2 or even 3 standard deviations is a huge multiple above the expected statistical frequency for that range of observations.

In statistical jargon, this is called “Kurtosis.” What is the operational implications of this?? The implied “safety” of a calculated improbable event is not really there. The probability of a “six-sigma event” is much higher than six sigma.

But we are not tossing our approach out just because of Kurtosis. Our approach is more powerful than ever once we understand the implications of Kurtosis.

Testing for normality
… using Kolmogorov-Smirnov



We have this programmed into our Python model now, but we are not yet prepared to talk about it. This just indicates that we **do** test …

The old Excel model:

	Kolmogorov-Smirnov Test for Normality		
K-S Critical D (95%)*:	0.05906667		
Cum Spread:	**Test**		Note: The desired result is "Fail" in all rows. You are **failing** to **reject** the null hypothesis that the calculated distribution is **not** different from normal.
-0.0073911	Fail		
-0.0129565	Fail		
-0.0116799	Fail		
0.0121707	Fail		
0.0538569	Fail		
0.1190576	Pass		
0.0028873	Fail		
-0.0427279	Fail		
-0.0473445	Fail		
-0.0419048	Fail		
-0.0002185	Fail		
0.0076958	Fail		
0.0049884	Fail		
0.0034070	Fail		
-0.0002326	Fail		

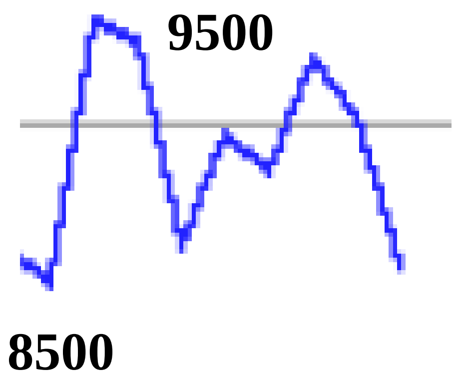
Hubert W. Lilliefors, ***On the Kolmogorov-Smirnov Test for Normaltiy with Mean and Variance Unknown,*** "Journal of the American Statistical Association, Vol. 62, No. 318. (june 1967), pp. 399-402.

What data frequency should we use?

Even when estimating weekly or monthly volatility over a span of years, we are going to use daily data and then use a formula to convert daily volatility to a larger aggregate.

Why? Because we are going to treat volatility as a fairly stable variable, but a variable none the less. It drifts over time and we have to be aware of that. Likewise many of our options trader are for relatively short duration. Daily volatility makes more sense.

Also, the more dense your data, the better the volatility information held within.



This is the DJIA in Oct 08. The endpoints are at more or less the same value (8500) But look at the volatility in between. In fact, this was the most volatile month in stock market history.

Time-adjusting volatility

(and introducing duration volatility)

(See note below): Duration volatility for a 16-day options contract (days remaining until expiry) equals four times daily volatility (and this will matter a lot).

1. $V_t = tV_d$

2. $\sigma_{dur} = \sqrt{t}\,\sigma_{day}$

3. $\sigma_{16} = \sqrt{16}\;\sigma_d \approx 4.0\,\sigma_d$

Because the variance of any interval that is *t* times longer than a shorter interval is simply equal to *t* times the variance of the shorter interval, the proportion of the standard deviations is equal to the square root of *t*. For an option, **duration volatility** is equal to daily volatility times the square root of the number of days until expiry. [**Important**]!

SD Volatility Time Conversions

When converting from standard deviation daily volatility to larger intervals, you take the daily standard deviation and multiply it times the square root of the number of days in the larger time interval. However, if there are, for example, only five trading days in a week (hence only five observations) and only 252 trading days in a year, then the number of days must be 5 and 252, not 7 and 365.

However, if you have a daily volatility measure for a stock and your are basing an option trade upon that measure, and the option expires in **one calendar month**, then the relevant multiplier for that trade is the square root of 30, not 22.

$$\sigma_{weekly} = \sigma_{daily}\sqrt{5} = 2.24\sigma_{daily}$$

$$\sigma_{monthly} = \sigma_{daily}\sqrt{22} = 4.69\sigma_{daily}$$

$$\sigma_{annual} = \sigma_{daily}\sqrt{252} = 15.9\sigma_{daily}$$

Modeling question: Volatility? Static or Variable?

Theoretically, how should we regard volatility? Many traditional models and textbook models treat volatility as a constant. Measuring volatility as “annual volatility,” common to classical options pricing models like Black-Scholes-Merton, implies static volatility over at least a year, if not constant volatility.

But efforts to measure daily volatility over different time intervals, such as that represented here to the right in one of our volatility models for SPY volatility, for different periods over a two-year sampling period, suggests that we cannot merely assume constant volatility.

There are two general approaches to this: (1) allowing for and trying to measure variable volatility, and (2) treating volatility as relatively static over long periods of time, but subject to a “shift in state” for the market.

SPY volatility and drift
21 Jan 2018 – 22 Jan 2020

	Drift	Volatility	qVolatil	Sharpe	Maxsigma	Minsigma
2 year	0.00032	0.00949	0.00757	0.03332	5.16303	-4.53728
1 year	0.00086	0.00743	0.00651	0.11627	2.77257	-4.22392
90 day	0.00109	0.00557	0.00557	0.19642	2.21811	-3.39796
30 day	0.00200	0.00437	0.00437	0.45745	1.67315	-2.19726

qVolatil is daily volatility with Kurtosis stripped out. This will be explained later.

Applications and extensions: The Sharpe Ratio

Historical Sharpe Ratio: The ratio of the stock's (or other FA) historical rate of return over it's volatility, over the same period:

$$\boxed{SR = \mu / \sigma}$$

Some versions make calculate this as an opportunity-cost return by subtracting the risk-free interest rate from yield:

$$\boxed{SR = (\mu - r) / \sigma}$$

Investment strategists may replace mu with their alpha and historical standard deviation with implied volatility.

Applications and extensions:
The elementary *Beta* risk estimator from 104

$$\mathrm{B}_x = \frac{COV_{xs}}{V_s}$$

and

$$COV_{xs} = CC_{xs} \times SD_x \times SD_s$$

which is the covariance of the financial asset and the index (or portfolio) divided by the variance of the index (or the portfolio). But variance and covariance of what? As we will now understand, never unconverted data, but either log growth rates or yields (the latter used when evaluating the volatility of yield-bearing financial assets).

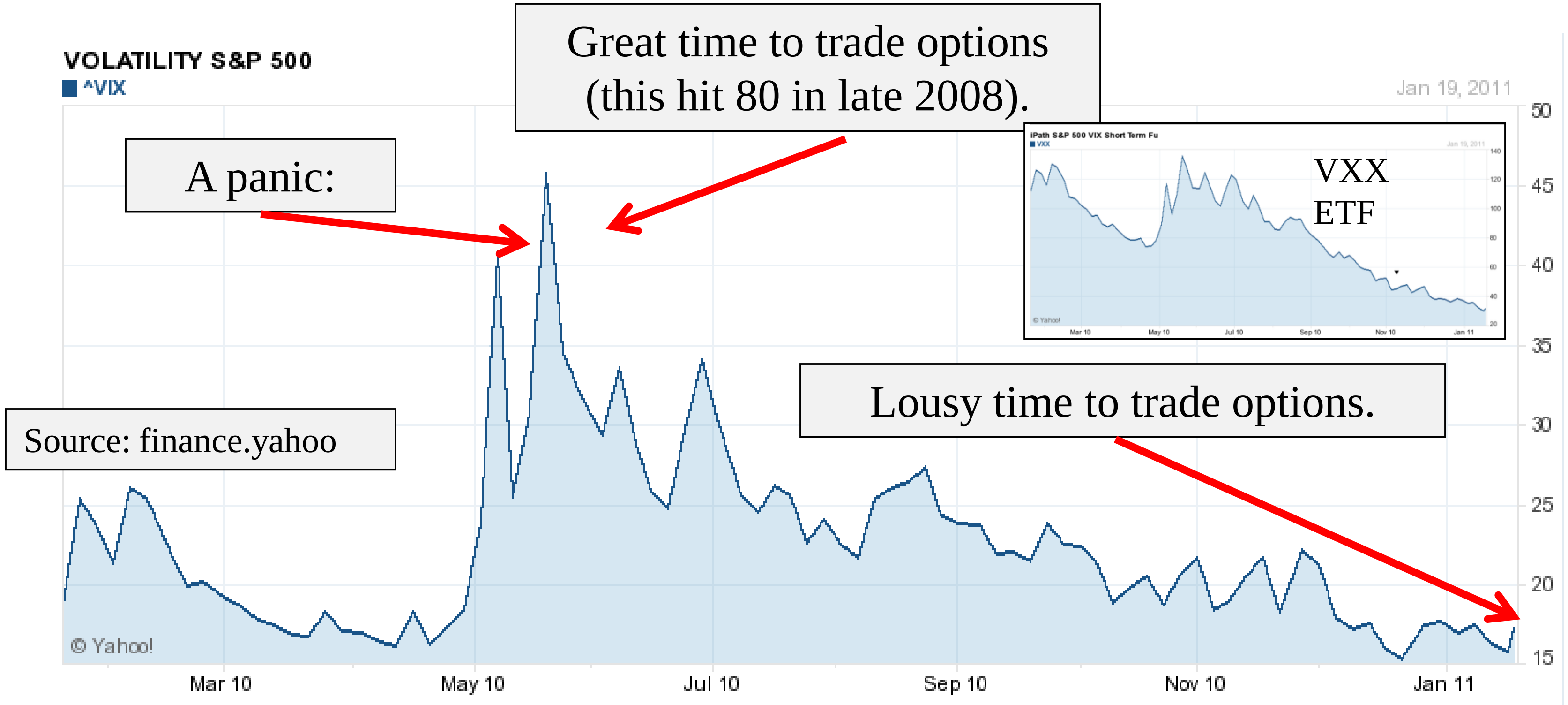
... and the Beta Rules of Thumb

When comparing the growth rate of a stock price to that of an index like the S&P500, then (subject to a qualification),

- If Beta > 1, then the asset is more volatile the the index .. if the asset were added to the index, the index would become more volatile.
- If Beta >0 and <1, the asset is less volatile than the index .. if the asset were added to the index, the index would become less volatile.
- If Beta is negative, the financial asset in question tends to move in the opposite direction of the index (the covariance is negative). Bond yields relative to the S&P 500 might be an example.

Can be more volatile not reflected in Beta if CC is very low but positive.

Market Volatility Proxy – The VIX



The VIX is a popular and very useful generic volatility indicator for the S&P 500. It is calculated using an options pricing model (which makes it comparable to calculating volatility for individual stocks). We review the technique later. If you trade options, you know where the VIX is.

Where do you find the tools to do this in Python (for HW1)?

Secure | https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.norm.html

SciPy.org Sponsored By ENTHOUGHT

Scipy.org | Docs | SciPy v1.0.0 Reference Guide | Statistical functions (`scipy.stats`)

scipy.stats.norm

Methods

rvs(loc=0, scale=1, size=1, random_state=None)	Random variates.
pdf(x, loc=0, scale=1)	Probability density function.
logpdf(x, loc=0, scale=1)	Log of the probability density function.
cdf(x, loc=0, scale=1)	Cumulative distribution function.
logcdf(x, loc=0, scale=1)	Log of the cumulative distribution function.
sf(x, loc=0, scale=1)	Survival function (also defined as `1 - cdf`, but *sf* is sometimes more accurate).
logsf(x, loc=0, scale=1)	Log of the survival function.
ppf(q, loc=0, scale=1)	Percent point function (inverse of `cdf` — percentiles).
isf(q, loc=0, scale=1)	Inverse survival function (inverse of `sf`).
moment(n, loc=0, scale=1)	Non-central moment of order n
stats(loc=0, scale=1, moments='mv')	Mean('m'), variance('v'), skew('s'), and/or kurtosis('k').
entropy(loc=0, scale=1)	(Differential) entropy of the RV.
fit(data, loc=0, scale=1)	Parameter estimates for generic data.
expect(func, args=(), loc=0, scale=1, lb=None, ub=None, conditional=False, **kwds)	Expected value of a function (of one argument) with respect to the distribution.
median(loc=0, scale=1)	Median of the distribution.
mean(loc=0, scale=1)	Mean of the distribution.
var(loc=0, scale=1)	Variance of the distribution.
std(loc=0, scale=1)	Standard deviation of the distribution.
interval(alpha, loc=0, scale=1)	Endpoints of the range that contains alpha percent of the distribution

... but, use Numpy instead of this wherever you can!

```python
# These show the scipy.stats norm methods
# for cdf and pdf.
# These are slow. Substitute the numpy error
# function method where you can.

from scipy.stats import norm

# Both cdf and pdf default to mean zero and
# SD zero so if you supply a single number it
# assumes standard normal.
# Be warned: Python operations of this kind
# are glacially slow. The erf approach , which
# uses C-based Numpy is much faster.
#
ex = norm.cdf(1)
print (ex)

ex2 = norm.cdf(2.6, 0.12, 2.8)
print (ex2)

ex3 = norm.pdf(1)
print (ex3)

ex4 = norm.pdf(2.6, 0.12, 2.8)
print (ex4)
```

I recommend the use of the error function (in math) because it is so much faster when used with a Numpy array ...!

```
#
import math
from datetime import date
import numpy as np
```

```
#
def snormdf(point):
    return (1/(math.sqrt(2*math.pi)))*math.pow(math.e,-0.5*point**2)
# csnd integrates a standard normal distribution up to some sigma.
#
def csnd(point):
    return (1.0 + math.erf(point/math.sqrt(2.0)))/2.0
#
# cnd integrates a Gaussian distribution up to some value.
#
def cnd(center,point,stdev):
    return (1.0 + math.erf((point - center)/(stdev*math.sqrt(2.0))))/2.0
#
```

Numpy and scipy.stats commands for inference

```python
# Experimental program to show a range of numpy and
# scipy.stats statistical calculators.
import numpy as np
from scipy.stats import norm, kurtosis, skew, sem
#
# Generate a numpy array of normal data
# Run this first with a sample size of 200, then 20,000
# to see the effects upon kurtosis, skew, etc.
#
data = np.random.normal(0.00326, 0.01824, 200)
#
# Mean, variance, standard deviation, median,max, min, skew, kurtosis
#
print("Mean: ", np.mean(data))
print("Variance: ", np.var(data))
print("Standard deviation: ", np.std(data))
print("Median: ", np.median(data))
print("Max: ", np.max(data))
print("Min: ", np.min(data))
print("Skew: ", skew(data))
print("Kurtosis: ", kurtosis(data))
print("Standard error of mean: ", sem(data))
# cdf are the cumulative density functions
# No parameters other than X assumes standard normal
print()
print("Probability of 1.2 sigma or less: ", norm.cdf(1.2))
print("Probability of 0.24 or less in 0, 0.18 : ",
    norm.cdf(0.0, 0.18, 0.24))
```

Other candidates for distributions for risk: The Lorentz (Cauchy) distribution

Cauchy distribution

From Wikipedia, the free encyclopedia

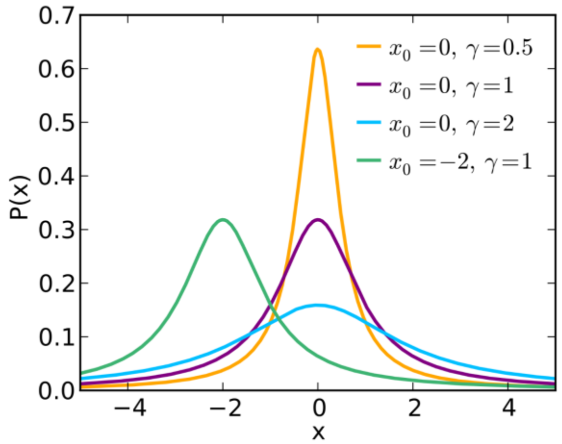
"Lorentz distribution" redirects here. It is not to be confused with Lorenz curve or Lorenz equation.

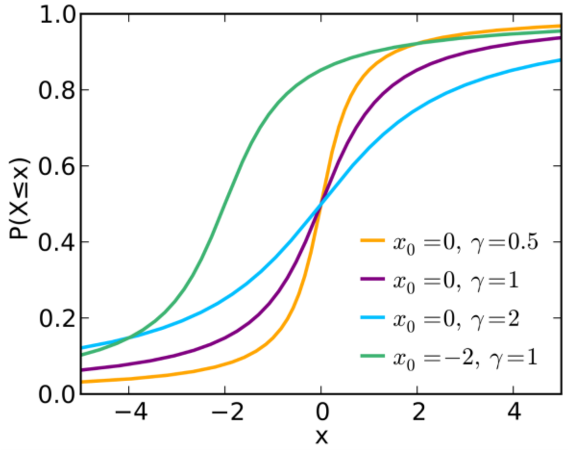
The **Cauchy distribution**, named after Augustin Cauchy, is a continuous probability distribution. It is also known, especially among physicists, as the **Lorentz distribution** (after Hendrik Lorentz), **Cauchy–Lorentz distribution**, **Lorentz(ian) function**, or **Breit–Wigner distribution**. The Cauchy distribution $f(x; x_0, \gamma)$ is the distribution of the x-intercept of a ray issuing from (x_0, γ) with a uniformly distributed angle. It is also the distribution of the ratio of two independent normally distributed random variables with mean zero.

The Cauchy distribution is often used in statistics as the canonical example of a "pathological" distribution since both its expected value and its variance are undefined. (But see the section *Explanation of undefined moments* below.) The Cauchy distribution does not have finite moments of order greater than or equal to one; only fractional absolute moments exist.[1] The Cauchy distribution has no moment generating function.

In mathematics, it is closely related to the Poisson kernel, which is the fundamental solution for the Laplace equation in the upper half-plane.

It is one of the few distributions that is stable and has a probability density function that can be expressed analytically, the others being the normal distribution and the Lévy distribution.





The clear advantage here is that we mitigate the problem of density at the center and "thin flanks."

Other candidates for distributions for risk: The slash distribution

Thanks to Patrick McKeen '17
https://en.m.wikipedia.org/wiki/Slash_distribution

Slash distribution

In probability theory, the **slash distribution** is the probability distribution of a standard normal variate divided by an independent standard uniform variate.[1] In other words, if the random variable *Z* has a normal distribution with zero mean and unit variance, the random variable *U* has a uniform distribution on [0,1] and *Z* and *U* are statistically independent, then the random variable *X* = *Z* / *U* has a slash distribution. The slash distribution is an example of a ratio distribution. The distribution was named by William H. Rogers and John Tukey in a paper published in 1972.[2]

The probability density function (pdf) is

$$f(x)=\frac{\varphi(0)-\varphi(x)}{x^2}.$$

where φ(*x*) is the probability density function of the standard normal distribution.[3] The result is undefined at *x* = 0, but the discontinuity is removable:

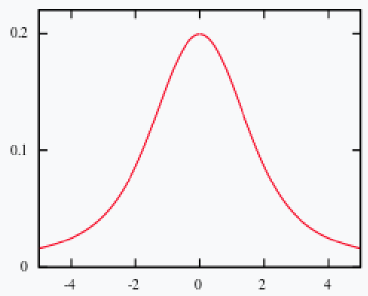
$$\lim_{x\to 0} f(x)=\frac{\varphi(0)}{2}=\frac{1}{2\sqrt{2\pi}}$$

The most common use of the slash distribution is in simulation studies. It is a useful distribution in this context because it has heavier tails than a normal distribution, but it is not as pathological as the Cauchy distribution.[3]

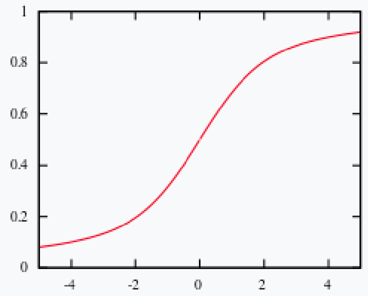
The clear advantage here is that we mitigate (maybe) the problem of kurtosis.

Slash

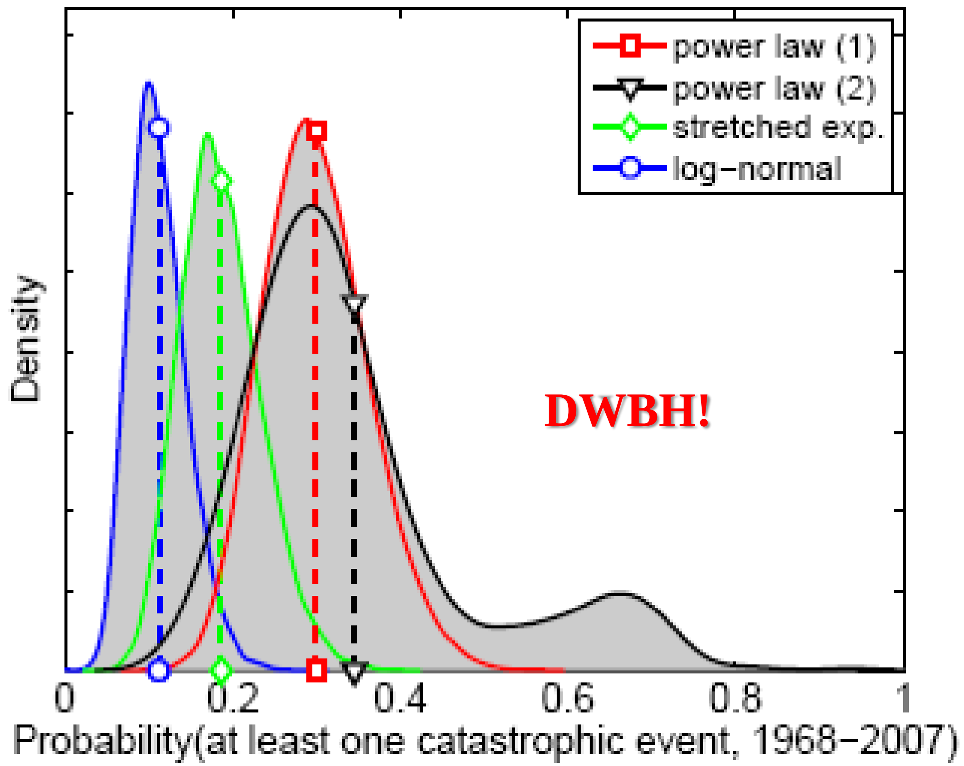
Probability density function



Cumulative distribution function



Memo: Power-law models for catastrophic risk, also called "tail risk," a hot subject these days



I found this is an article exploring models used to predict the probability of a rare event like 9/11:

Aaron Clauset and Ryan Woodard, "*Estimating the Historical and Future Probabilities of Large Terrorist Events*," physics.data.en, Sept. 1, 2012, [in class optional material].

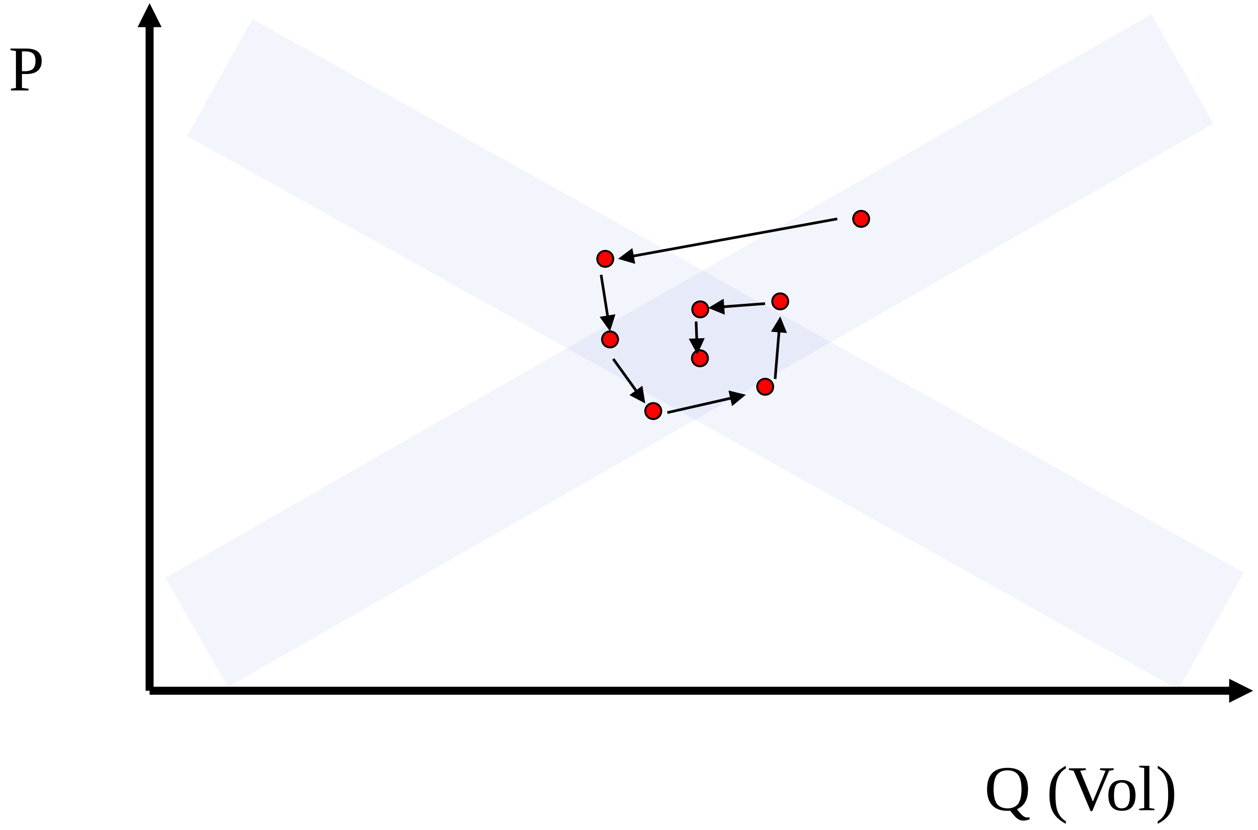
Fair and efficient markets??

We are going to make two more related assumptions that we will simply state here and return to evaluate later. If there was ever a subject that was misunderstood for its meaning and implications, it is the assumption of fair and efficient markets. It doesn't mean what many people think and it is not as important as people think. It is never worth going to the mat over the question of fair and efficient markets. We will return to this at the end of the semester.

A market is fair if no party to a transaction has an apriori avantage to the trade. For us who trade financial assets, assuming that there is a standard that we can accept as representing an unbiased expected value of a bet, the market is fair if either (a) for both parties, the expected value of the result of a bet is equal to the cost of the bet, or (b) the expected value of the result of a bet is equal to the cost of the bet adjusted for compensation for risk (the insurance value of the bet).

An efficient market (in finance) is a market that has a strong tendency toward a fair market, which implies a strong tendency toward settling prices that are fair. This does not imply that an efficient market instantaneously resolves settling prices that are fair.

The role of the cobweb theorem and equilbrium price discovery in fair and efficient markets ...



We have incomplete information in a vague world, and we don't actually know where supply and demand are so we don't where the market clears. But the market has incentives that punish us for being far away for the equilibium (inventory build-up and price punishment) which slowly gets resolved as we move closer to equilibrium.

What you should take away from this large lecture ...

1. Be able to explain the Markov assumption ..
   - ... and what it implies about stock price behavior over time
   - ... and what it does not imply
2. Why and how we are using log ratios and what do those represent?
3. How we can use the standard normal distribution.
4. What is the primary “violation” of normalcy in stock data, what does it mean, and how do we test for it?
5. What is the primary DE and how are its two components interpreted ... what are they?
6. When the DE is solved, what is the solution for price at “t” given initial price?
7. How is the DE converted into a MC simulation?
8. How can the MC simulation be applied to a strangle, what might it tell us about the strangle?
9. What is lacking in the MC simulation .. why is it unrealistic?
10. If the CGRs are assumed to be normal, what can we say about the transformed stock distribution?
11. What type of stock will not conform to normal?
12. How is daily volatility adjusted for longer time intervals?
13. What is duration volatility and how is it calculated?
14. Is volatility a constant?
15. What is the Sharpe Ratio and the Beta? How are they measured.
16. What are two alternative distributions that are sometimes recommended over Normal and Standard Normal?