Price and Return Distributions

- Return types: GDR, NDR and LGDR.
- Arithmetic and geometric averages.
- Log normal and normal distributions.
- Mean > median > mode for log normal distributions.
- Ensemble and time weighted averages.
- Calculation example: mean and median.
- Calculation example: probability of price exceeding the mean and median.

Sum and Product functions

The capital Greek letter sigma denotes the sum of T discrete values (from t=1 to T) of some variable x:

$$\sum_{t=1}^{T} x = x_1 + x_2 + \cdots + x_T$$

The capital Greek letter pi denotes the product of T discrete values (from t=1 to T) of some variable x:

$$\prod_{t=1}^{T} x = x_1 \times x_2 \times \ldots \times x_T$$

Single Period Returns

Gross discrete return, often denoted by capital R:

$$GDR_{0\to1} = \frac{p_1}{p_0}$$

Net discrete return, also called the effective return (r_{eff}) or relative return, often denoted by lower case r:

$$NDR_{0\to1} = \frac{p_1}{p_0} - 1 = \frac{p_1 - p_0}{p_0} = GDR_{0\to1} - 1$$

Log gross discrete return, also called the continuously compounded return (r_{cc}) or force of interest (y):

$$LGDR_{0\to1} = \ln\left(\frac{p_1}{p_0}\right) = \ln(GDR_{0\to1})$$

Geometric and Arithmetic Averages

Arithmetic average over T periods:

$$\bar{x}_{arithmetic\ average,0\to T} = \frac{\sum_{t=1}^{T}(x_{t-1\to t})}{T}$$

$$= \frac{x_{0\to 1} + x_{1\to 2} + x_{2\to 3} + \cdots + x_{T-1\to T}}{T}$$

Geometric average over T periods:

$$\bar{x}_{geometric\ average,0\to T} = \left(\prod_{t=1}^{T}(x_{t-1\to t})\right)^{\frac{1}{T}}$$

$$= (x_{0\to 1}.\, x_{1\to 2}.\, x_{2\to 3} \ldots x_{T-1\to T}\ )^{\frac{1}{T}}$$

Geometric and Arithmetic Average Returns

Two types of average (also called mean):

Arithmetic average net discrete return (NDR) from time 0 to n:

$$\bar{r}_{arithm\ NDR\ 0-n} = \frac{\sum_{i=1}^{n}(r_i)}{n} = \frac{r_1 + r_2 + \cdots + r_n}{n}$$

Geometric average net discrete return (NDR) from time 0 to n:

$$\bar{r}_{geom\ NDR\ 0-n} = [(1 + r_1)(1 + r_2) \ldots (1 + r_n)]^{1/n} - 1$$

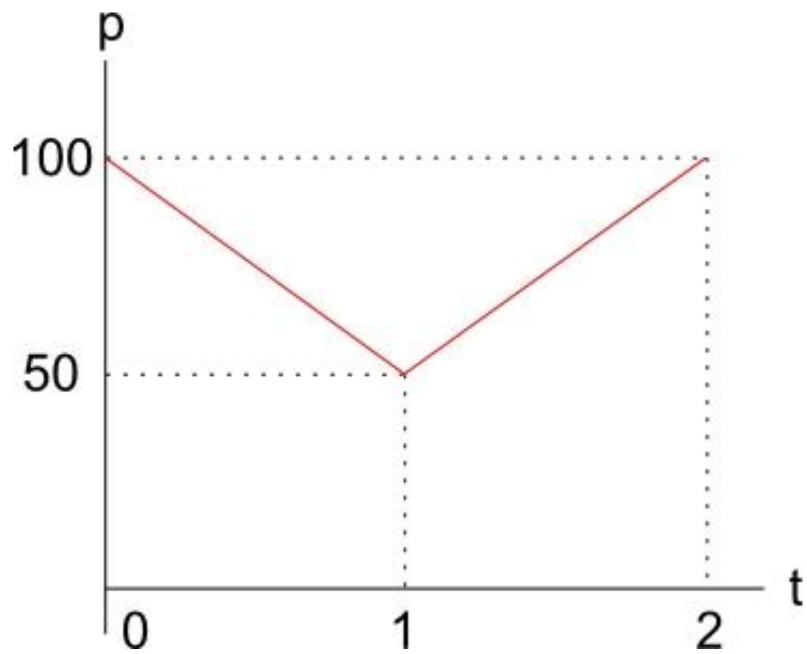
$$= \left(\frac{p_n}{p_0}\right)^{1/n} - 1$$

Average Returns – A Curious Example

Time (year)	**0**	**1**	**2**
Share price ($)	100	50	100
Net discrete return pa		-0.5	1.0

Note that we will express all returns as pure decimals, not %, unless marked as so.

Similarly for standard deviation and variance.



$$\bar{r}_{geom\ NDR\ 0-2,p.a.} = [(1+r_1)(1+r_2)\ldots(1+r_n)]^{1/n} - 1$$
$$= [(1+r_{0-1})(1+r_{1-2})]^{1/2} - 1$$
$$= [(1+-0.5)(1+1)]^{1/2} - 1$$
$$= 0$$

$$\bar{r}_{arithm\ NDR\ 0-2,p.a.} = \frac{r_1 + r_2 + \cdots + r_n}{n}$$
$$= \frac{r_{0-1} + r_{1-2}}{2}$$
$$= \frac{-0.5 + 1}{2}$$
$$= 0.25$$

Which average return tells the true story?

If an investor wants to buy the share at time zero and sell it two years later, then the geometric average return makes more sense since it takes compounding into account over time.

If the investor wants to buy for one year and sell a year later, so there's a 1/2 probability of losing 50% and a 1/2 probability of gaining 100% then the arithmetic average is arguably better.

In finance we frequently use both types of averages.

- Arithmetic average returns are often used in Markowitz mean-variance portfolio analysis to find stocks' average returns from past data.
- Geometric averages are often used in the debt markets for computing the term structure of interest rates. Notice that the quantity $(1 + r_{0 \to 1\ eff})$ is the gross discrete return (GDR).

Term Structure of Interest Rates: The Expectations Hypothesis

Expectations hypothesis is that long term spot rates (plus one) are the geometric average of the shorter term spot and forward rates (plus one) over the same time period. Mathematically:

$$1 + r_{0\to T} = \left((1 + r_{0\to 1})(1 + r_{1\to 2})(1 + r_{2\to 3})\dots\left(1 + r_{(T-1)\to T}\right)\right)^{\frac{1}{T}}$$

or

$$(1 + r_{0\to T})^T = (1 + r_{0\to 1})(1 + r_{1\to 2})(1 + r_{2\to 3}) \dots (1 + r_{(T-1)\to T})$$

Where T is the number of periods and all rates are **effective** rates over each period.

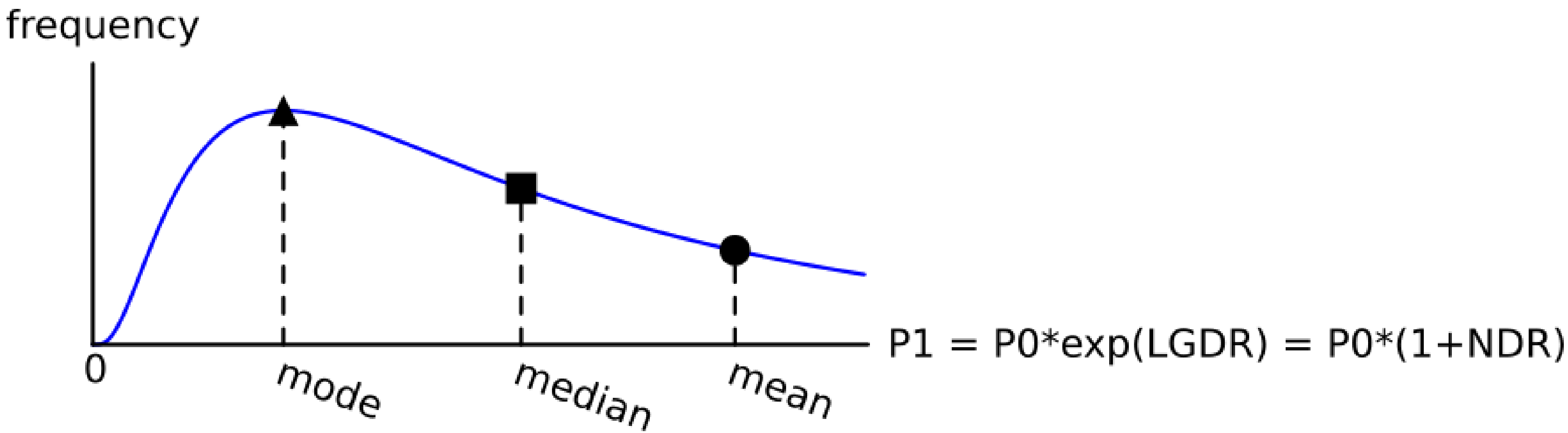
Prices are Log-Normally Distributed

Since security prices (p) cannot fall below zero due to limited liability, they're normally assumed to be log-normally distributed with a left-most value of zero.

$$p_t \sim lnN(mean, variance)$$

$$0 \leq p_t < \infty$$

Of course, in reality this assumption of log-normal prices (and normal continuously compounded returns) is broken, but keeping it makes the mathematics tractable.



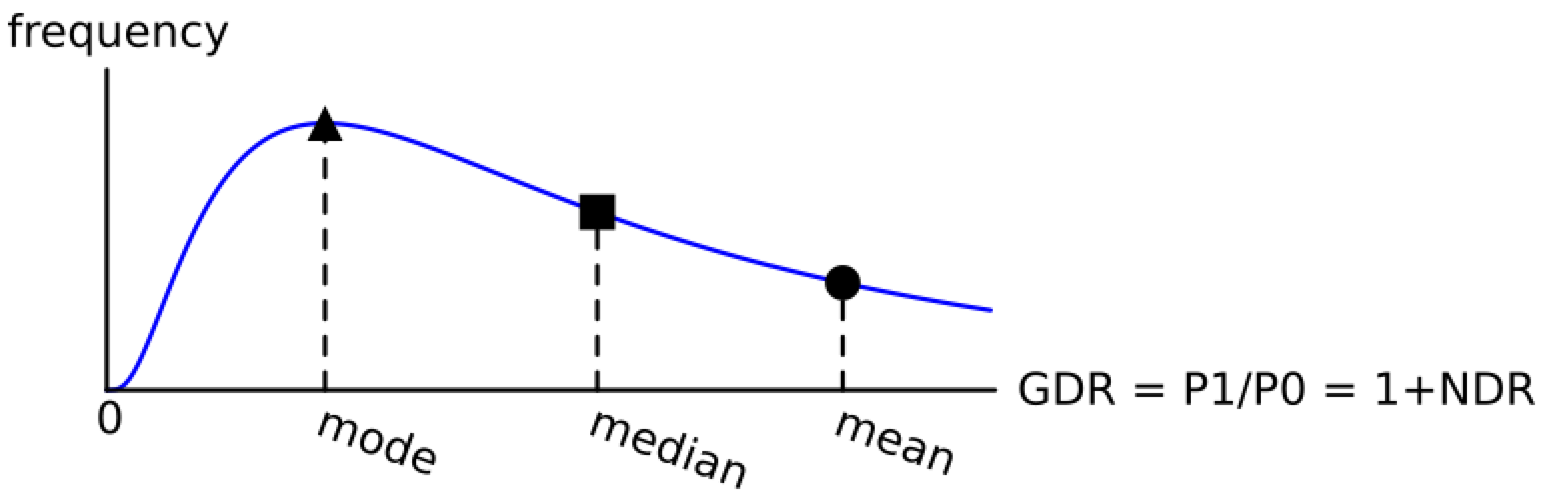
GDR's are Log-Normally distributed

Since gross discrete returns (GDR's) are linear functions of the price, they are also log-normally distributed.

GDR's have a left-most point of zero, same as prices.

$GDR \sim lnN(mean, variance)$

$0 \leq GDR_{t \to t+1} < \infty$



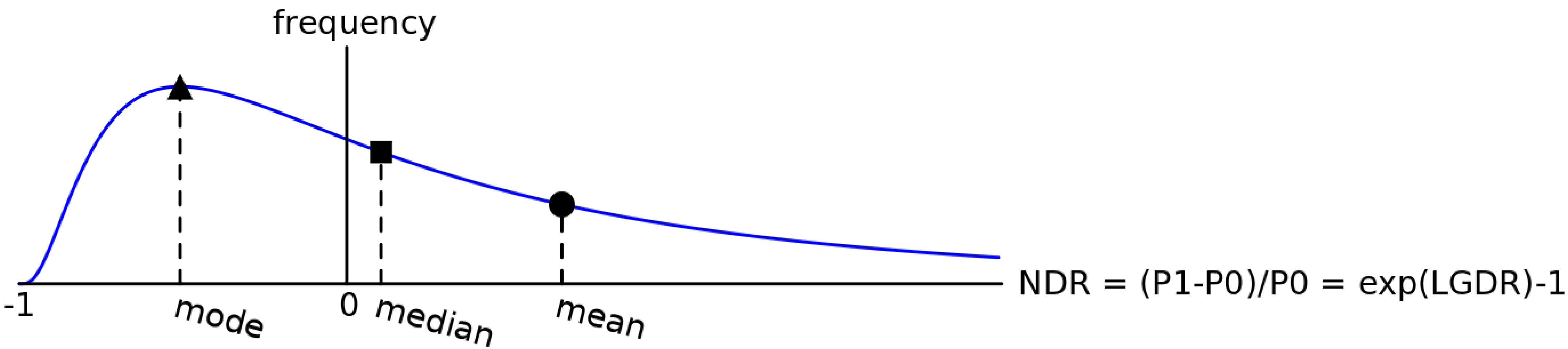
NDR's are Log-Normally distributed

Net discrete returns (NDR's, also known as effective returns) are equal to GDR’s minus 1, so they’re shifted to the left by one.

NDR’s have a left-most point of negative one.

$NDR \sim lnN(mean, variance)$

$-1 \leq NDR_{t \to t+1} < \infty$

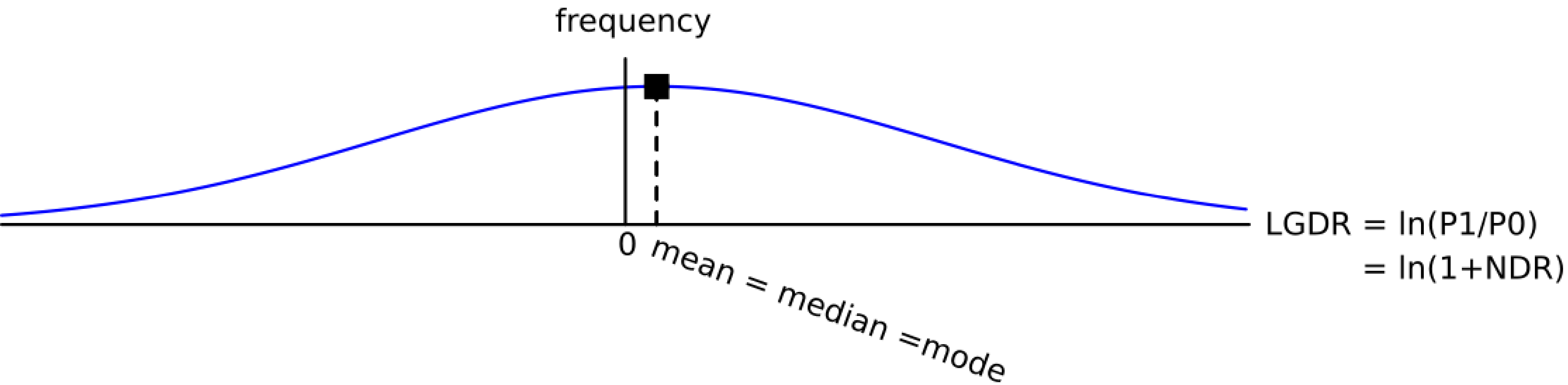


LGDR's are Normally distributed

Log gross discrete returns $\left(LGDR = \ln\left(\frac{p_t}{p_0}\right)\right)$ are normally distributed and they are unbounded in the positive and negative directions.

$LGDR \sim N(mean, variance)$

$-\infty \leq LGDR_{t \to t+1} < \infty$



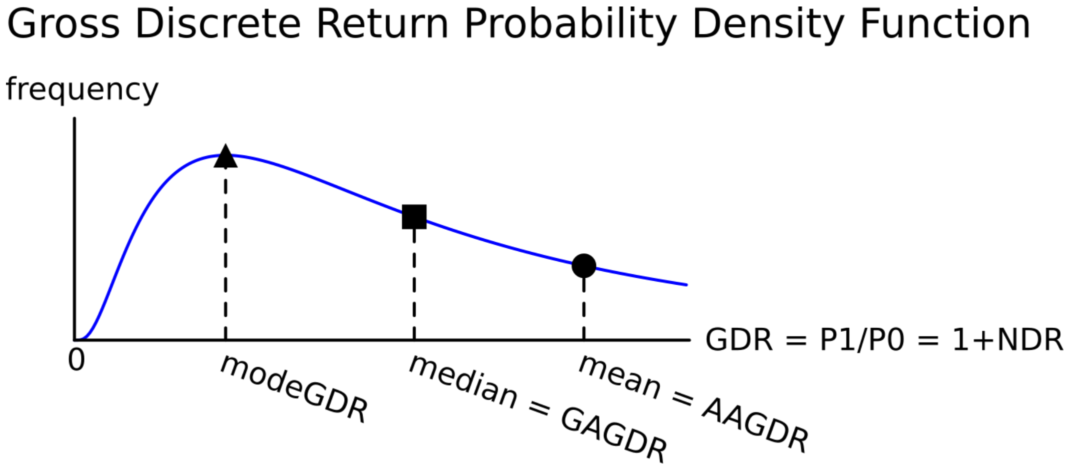
AAGDR = Mean GDR

Arithmetic average of the gross discrete returns (AAGDR) is the mean:

$$AAGDR_{0\to T} = \frac{1}{T} \cdot \sum_{t=1}^{T} \left(\frac{p_t}{p_{t-1}} \right)$$

$$= \frac{\frac{p_1}{p_0} + \frac{p_2}{p_1} + \cdots + \frac{p_T}{p_{T-1}}}{T}$$

$$= \frac{GDR_{0\to 1} + GDR_{1\to 2} + \cdots + GDR_{T-1\to T}}{T}$$



GAGDR = Median GDR

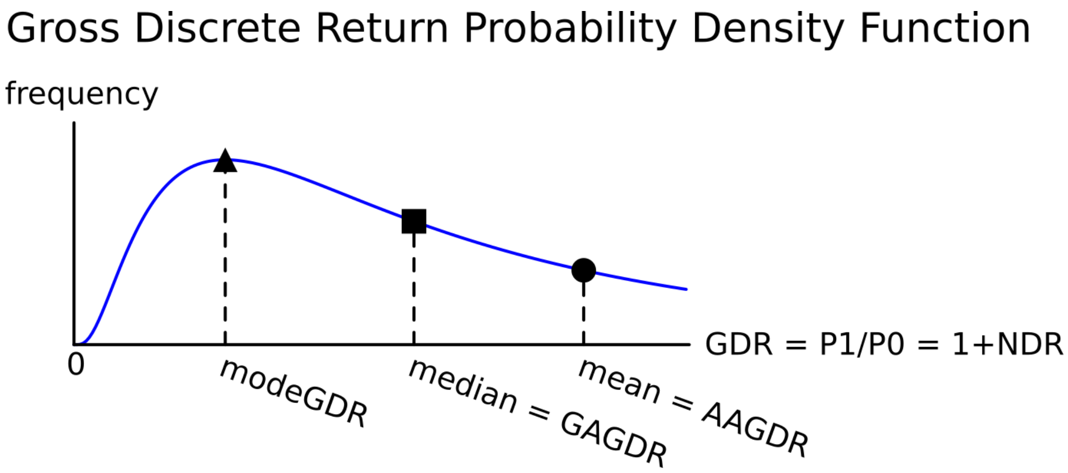
Geometric average of the gross discrete returns (GAGDR):

$$GAGDR_{0\to T} = \left(\prod_{t=1}^{T}\left(\frac{p_t}{p_{t-1}}\right)\right)^{\frac{1}{T}}$$

$$= \left(GDR_{0\to 1} \times GDR_{1\to 2} \times \ldots \times GDR_{T-1\to T}\right)^{\frac{1}{T}}$$

$$= \left(\frac{p_1}{p_0} \times \frac{p_2}{p_1} \times \frac{p_3}{p_2} \times \ldots \times \frac{p_T}{p_{T-1}}\right)^{\frac{1}{T}}$$

$$= \left(\frac{p_T}{p_0}\right)^{\frac{1}{T}} = \left(GDR_{0\to T}\right)^{\frac{1}{T}}$$



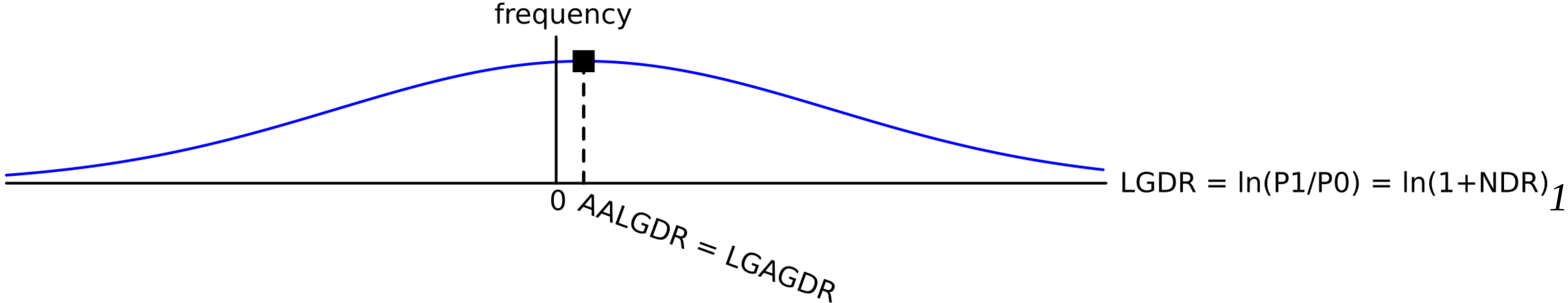
LGAGDR = AALGDR = Mean, Median and Mode LGDR

The log of the geometric average of the gross discrete returns (LGAGDR) is:

$$LGAGDR_{0\to T} = \ln\left(\left(\prod_{t=1}^{T}\left(\frac{p_t}{p_{t-1}}\right)\right)^{\frac{1}{T}}\right)$$

$$= \frac{1}{T}.\ln\left(\frac{p_T}{p_0}\right) = \frac{1}{T}.\ln(GDR_{0\to T})$$

Log Gross Discrete Return Probability Density Function



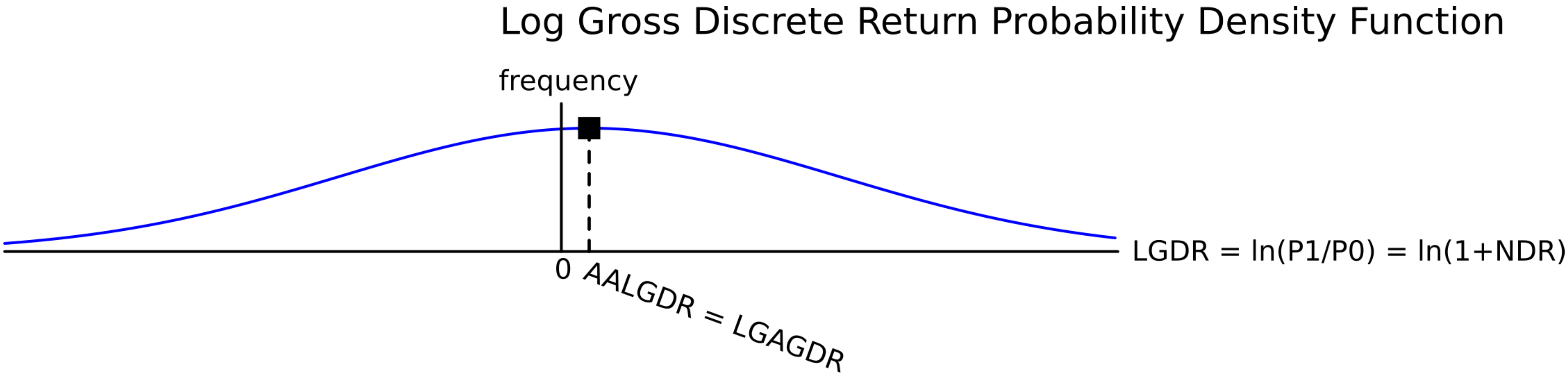
The log of the geometric average of the gross discrete returns (LGAGDR) equals the arithmetic average of the log gross discrete returns (AALGDR):

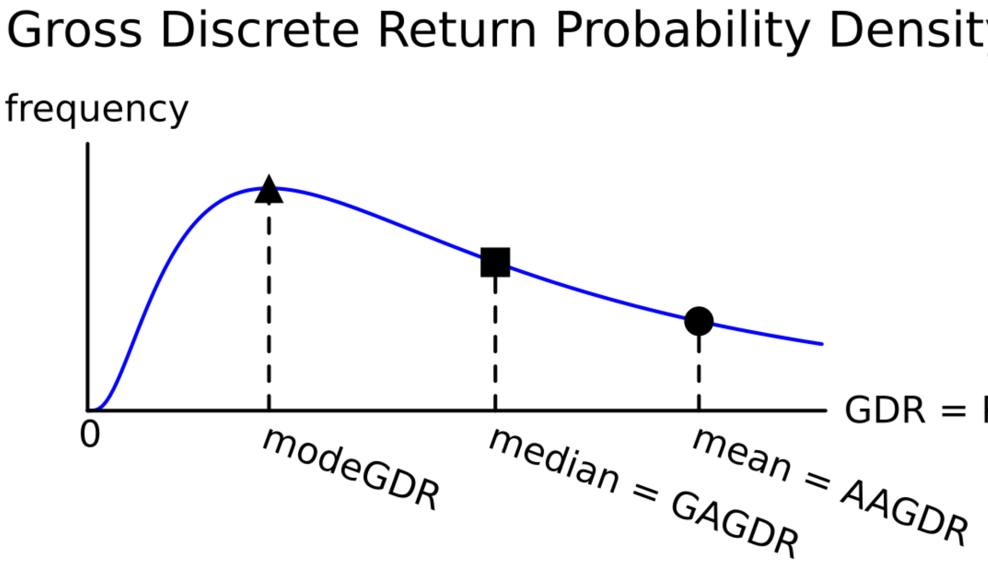
$$LGAGDR_{0\to T} = \ln\left(\left(\prod_{t=1}^{T}\left(\frac{p_t}{p_{t-1}}\right)\right)^{\frac{1}{T}}\right)$$

$$= \frac{1}{T}.\ln\left(\frac{p_1}{p_0}\times\frac{p_2}{p_1}\times\frac{p_3}{p_2}\times\ldots\times\frac{p_T}{p_{T-1}}\right)$$

$$= \frac{1}{T}.\left(\ln\left(\frac{p_1}{p_0}\right)+\ln\left(\frac{p_2}{p_1}\right)+\ln\left(\frac{p_3}{p_2}\right)+\cdots+\ln\left(\frac{p_T}{p_{T-1}}\right)\right)$$

$$= \frac{1}{T}.\sum_{t=1}^{T}\left(\ln\left(\frac{p_t}{p_{t-1}}\right)\right) = AALGDR_{0\to T}$$


Log Gross Discrete Return Probability Density Function
frequency
LGDR = ln(P1/P0) = ln(1+NDR)
0
AALGDR = LGAGDR


Gross Discrete Return Probability Density Function
frequency
GDR = P1/P0 = 1+NDR
0
modeGDR
median = GAGDR
mean = AAGDR

SDLGDR

The ***arithmetic*** standard deviation of the log gross discrete returns (SDLGDR) is defined as:

$$SDLGDR = \sigma = \sqrt{\frac{1}{T}.\sum_{t=1}^{T}\left(\left(\ln\left(\frac{p_t}{p_{t-1}}\right) - AALGDR_{0\to T}\right)^2\right)}$$

Since $AALGDR_{0\to T} = LGAGDR_{0\to T} = \ln(GAGDR_{0\to T})$, then:

$$SDLGDR = \sqrt{\frac{1}{T}.\sum_{t=1}^{T}\left(\left(\ln\left(\frac{GDR_{t-1\to t}}{GAGDR_{0\to T}}\right)\right)^2\right)}$$

The ***geometric*** standard deviation of the gross discrete returns is defined as the exponential of the ***arithmetic*** standard deviation of the ***log*** gross discrete returns (SDLGDR).

$$\text{GeometricSDGDR} = \exp(\text{SDLGDR})$$

$$= \exp\left(\sqrt{\frac{1}{T} . \sum_{t=1}^{T} \left(\left(ln\left(\frac{GDR_{t-1\to t}}{GAGDR_{0\to T}}\right)\right)^2\right)}\right)$$

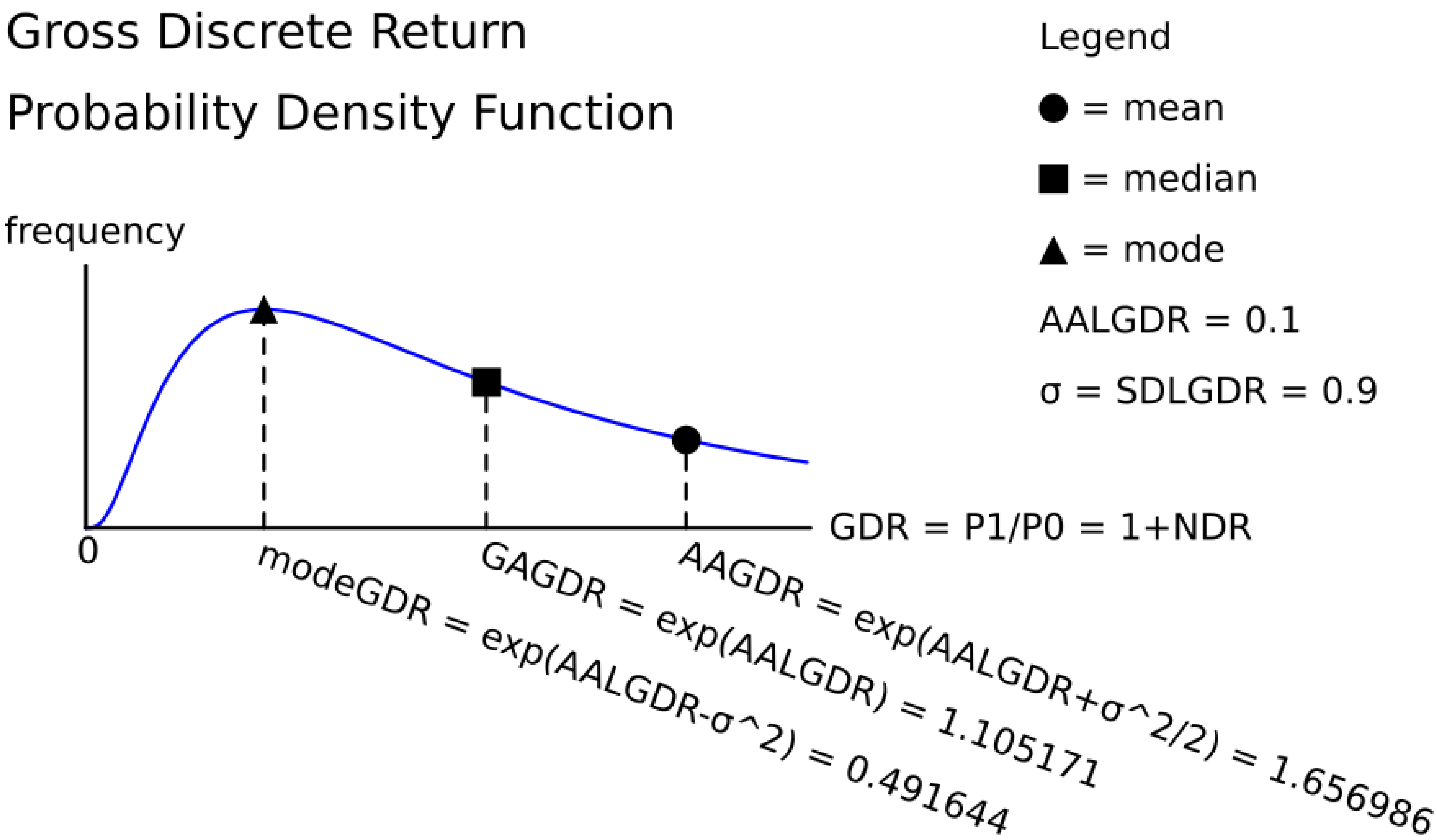
AAGDR = exp(AALGDR + SDLGDR²/2)

Another interesting relationship between the arithmetic average gross discrete return (AAGDR) and the arithmetic average of the log gross discrete return (AALGDR or continuously compounded return).

$$AAGDR = e^{AALGDR + \frac{SDLGDR^2}{2}}$$

$$= \exp(LGAGDR) \times \exp\left(\frac{SDLGDR^2}{2}\right)$$

$$= GAGDR \times \exp\left(\frac{SDLGDR^2}{2}\right)$$



Gross Discrete Return
Probability Density Function
Legend
= mean
= median
= mode
AALGDR = 0.1
σ = SDLGDR = 0.9
frequency
0
GDR = P1/P0 = 1+NDR
modeGDR = exp(AALGDR-σ^2) = 0.491644
GAGDR = exp(AALGDR) = 1.105171
AAGDR = exp(AALGDR+σ^2/2) = 1.656986

Some people prefer to write the expression without the exponential function by taking the logarithm of both sides:

$$AAGDR = e^{AALGDR + \frac{SDLGDR^2}{2}}$$

$$LAAGDR = AALGDR + \frac{SDLGDR^2}{2}$$

There is a difference between how the LAAGDR and AALGDR are calculated from prices:

$$LAAGDR = \ln\left(\frac{1}{T}.\sum_{t=1}^{T}\left(\frac{p_t}{p_{t-1}}\right)\right)$$

$$AALGDR = \frac{1}{T}.\sum_{t=1}^{T}\left(\ln\left(\frac{p_t}{p_{t-1}}\right)\right) = LGAGDR$$

Necessary assumptions:

- LGDR's (also called continuously compounded returns) must be normally distributed and therefore the GDR's are log-normally distributed.
- If the returns (LGDR's) are of a portfolio, this equation only holds if the weights are continuously rebalanced.

Mean GDR ≥ Median GDR

$$AAGDR = GAGDR \times \exp\left(\frac{SDLGDR^2}{2}\right)$$

This is equivalent to:

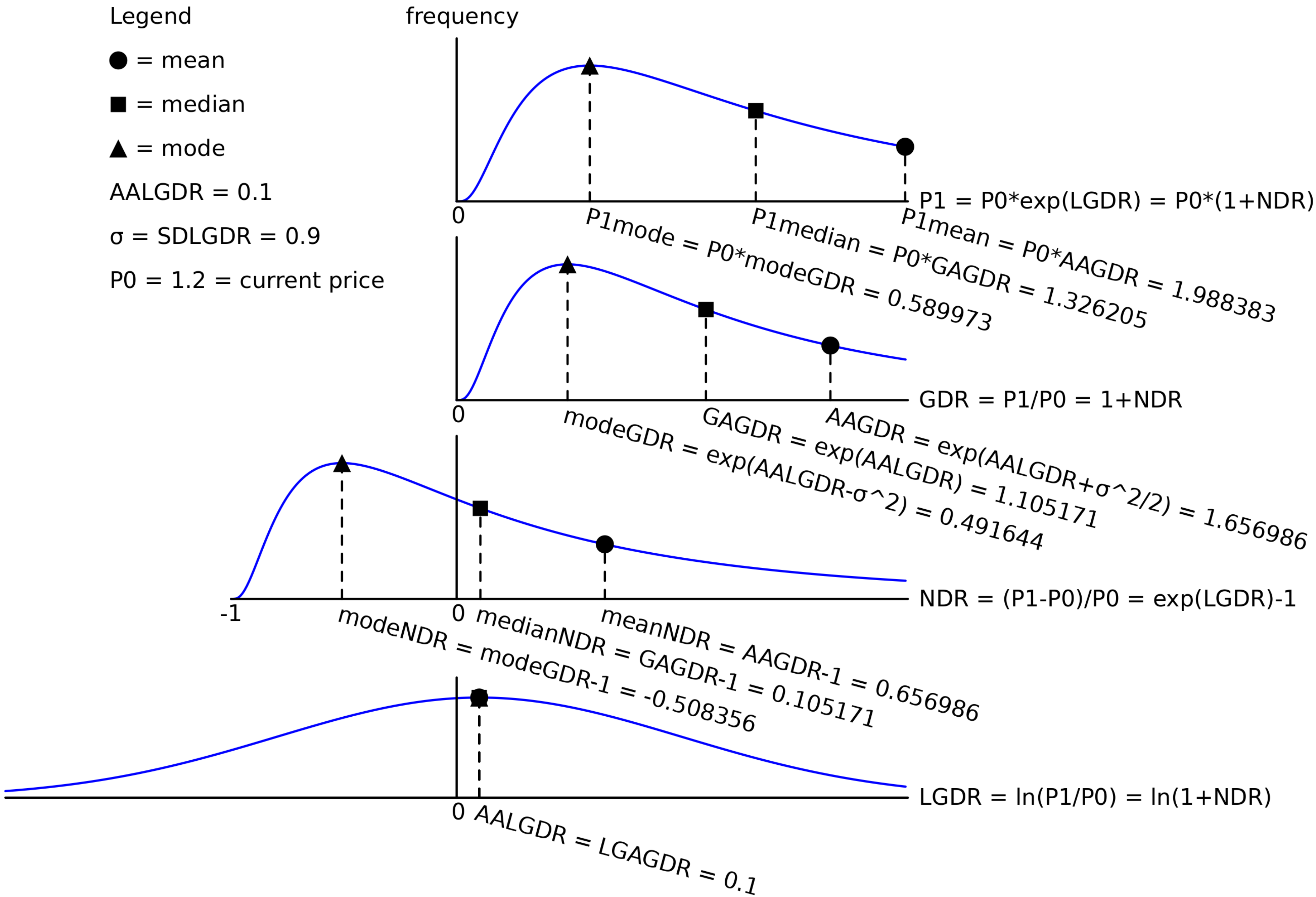
$$\text{Mean GDR} \ = \ \text{Median GDR} \times \exp\left(\frac{\text{SDLGDR}^2}{2}\right)$$

So the Mean GDR (or AAGDR) is always greater than or equal to the Median GDR (or GAGDR) because $\exp\left(\frac{\text{SDLGDR}^2}{2}\right)$ is always greater than or equal to one.

Mean GDR ≥ Median GDR or **AAGDR ≥ GAGDR**

This can be shown intuitively using a graph.

Price and Return Probability Density Functions

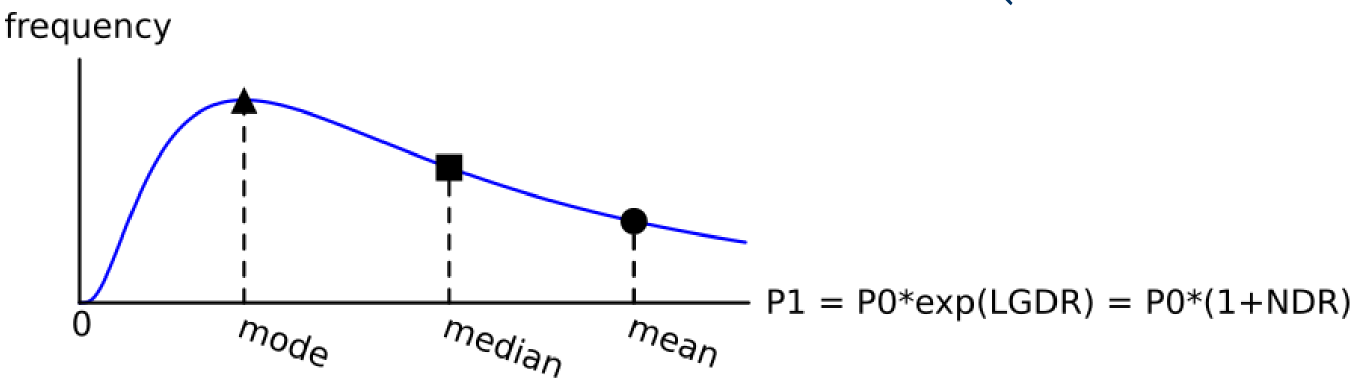


Why is this interesting?

If future prices are log-normally distributed, the mean and median future prices will be different!

$$MeanP_t = P_0.e^{\left(AALGDR+\frac{\mathrm{SDLGDR}^2}{2}\right).t} = P_0.AA\boldsymbol{G}DR^t$$
$$= P_0.(1 + AA\boldsymbol{N}DR)^t$$

$$MedianP_t = P_0.e^{AALGDR.t} = P_0.AA\boldsymbol{G}DR^t.e^{-\frac{SDLGDR^2}{2}.t}$$
$$= P_0.(1 + AA\boldsymbol{N}DR)^t / \exp\left(\frac{SDLGDR^2.t}{2}\right)$$

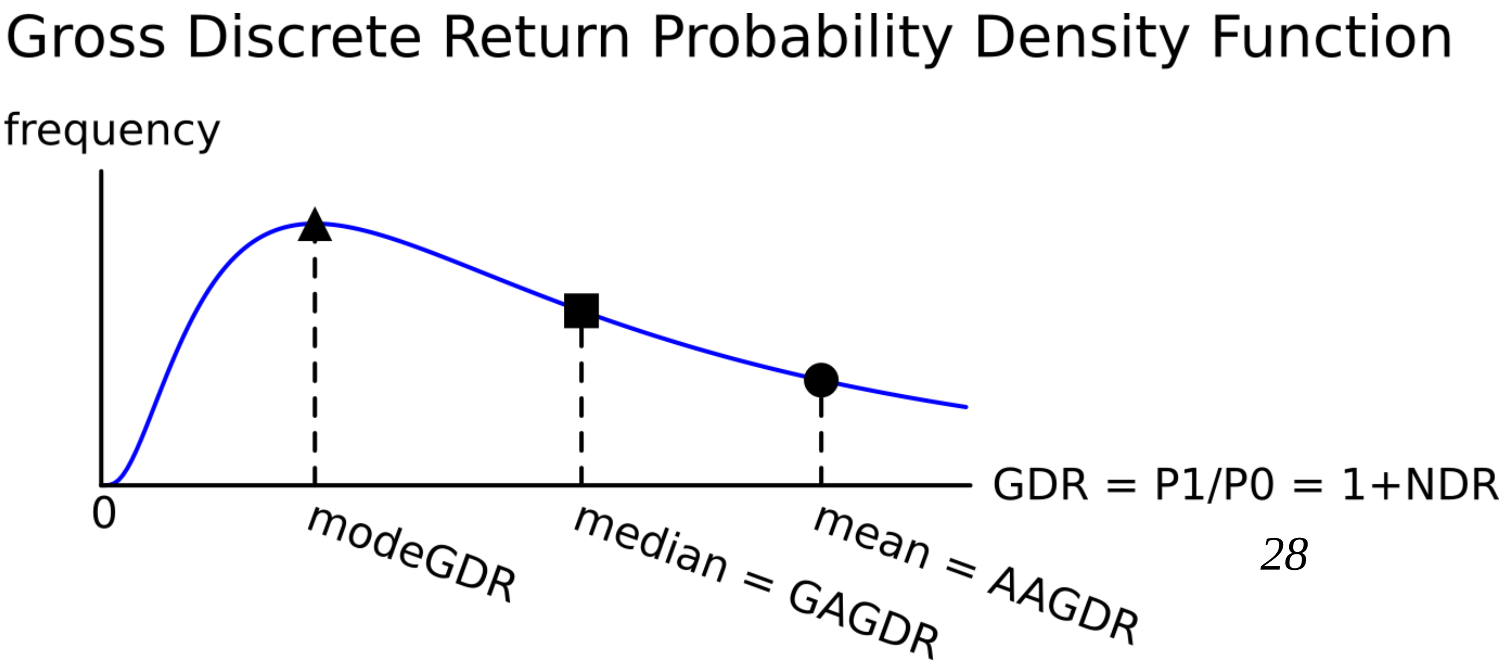


Mean versus Median of a Log-normally distributed variable

Due to the long right tail in the log-normal distribution, there are some very high values which increase the mean but have less effect on the median.

The chance of achieving a GDR, NDR or price above the mean is less than 50%.

The chance of achieving a return above the median is exactly 50% since the median is the 50th percentile.



Time Weighted Average = GAGDR = Median GDR

The 'time-weighted average' GDR or NDR is synonymous with the geometric average GDR or NDR.

Over time, the chance of achieving or exceeding the median price, GDR or NDR will always be 50%.

Ensemble Average = AAGDR = Mean GDR

Over time, the chance of achieving or exceeding the **mean** price, GDR or NDR will get smaller and smaller. As time approaches infinity, the chance will approach zero.

This makes sense because there are a small number of shares that will do exceptionally well. Think of Berkshire Hathaway, Google or Amazon which have had incredibly high returns. There is no limit to how high prices, GDR's or NDR's can go.

However, the lowest that prices and GDR's can go is zero and NDR's can't get below minus 1 which is losing 100%.

The arithmetic average is pulled up by the small number of very high returns, while the median or geometric average, the ‘middle value’, is not affected as much.

Calculation Example: Mean and Median Prices

Question 1: A stock has an arithmetic average continuously compounded return (AALGDR) of **10**% pa, a standard deviation of continuously compounded returns (SDLGDR) of **16.45**% pa and current stock price of $**1**.

In **20** years, what do you expect its **mean** and **median** prices to be?

Assume that stock prices are log-normally distributed.

Answer: The median price can be found with the formula.

$$MedianP_t = P_0 . e^{AALGDR.t}$$

$$MedianP_{20} = 1 \times e^{0.1 \times 20} = 7.389056099$$

The mean price is a little harder.

$$MeanP_t = P_0 . e^{\left(AALGDR + \frac{\text{SDLGDR}^2}{2}\right).t}$$

$$MeanP_{20} = 1 \times e^{\left(0.1 + \frac{0.1645^2}{2}\right) \times 20} = 9.68523441$$

Calculation Example: Probabilities

Question 2: A stock has an arithmetic average continuously compounded return (AALGDR) of **10**% pa, a standard deviation of continuously compounded returns (SDLGDR) of **46.52**% pa and current stock price of $**1**.

In **50** years, what do you expect its **mean** and **median** prices to be?

Assume that stock prices are log-normally distributed.

Answer: The median price can be found with the formula.

$$MedianP_t = P_0.e^{AALGDR.t}$$

$$MedianP_{50} = 1 \times e^{0.1 \times 50} = 148.41$$

The mean price is a little harder.

$$MeanP_t = P_0 . e^{\left(AALGDR + \frac{\mathrm{SDLGDR}^2}{2}\right).t}$$

$$MeanP_{50} = 1 \times e^{\left(0.1 + \frac{0.4652^2}{2}\right) \times 50} = 33199.03$$

The mean is a lot bigger than the median.

Question 2: What is the probability of the share price exceeding the **median** price of $148.41 in 50 years?

Answer: The median price is the 'middle price' when all possible prices are arranged from smallest to biggest. So the chance of achieving a price higher than the median is always 50%.

Question 3: What is the probability of the share price exceeding the **mean** price of $33,199.03 in 50 years? Note that the mean is the arithmetic mean.

Answer: The chance of achieving a price more than the mean will be less than 50%.

To find the exact probability, convert the **log-normally** distributed prices into **normally** distributed continuously compounded returns (LGDR's).

$$LAAGDR_{50yr} = \ln\left(\frac{MeanP_{50}}{P_0}\right)$$

$$= \ln\left(\frac{33199.03}{1}\right) = 10.410276 \approx 1041\%$$

$$AALGDR_{50yr} = AALGDR_{1yr} \times 50 = 0.1 \times 50 = 5 \approx 500\%$$

$$SDLGDR_{50yr} = SDLGDR_{1yr} \times \sqrt{50} = 0.4652 \times \sqrt{50}$$

$$= 3.289460746 \approx 329\%$$

Note that these returns and the standard deviation are all measured over the 50 year period, they're not per annum.

By standardising these continuously compounded returns into Z scores we can then find the probability of exceeding the mean price.

$$Z = \frac{x - \mu}{\sigma}$$

$$= \frac{\ln(MeanP_{50}/1) - \ln(\mathrm{MedianP}_{50}/1)}{SDLGDR \times \sqrt{50}}$$

$$= \frac{10.410276 - 5}{3.289460746} = 1.644730373$$

$$Prob(P_{50} < MeanP_{50}) = N(1.64473) = 0.95 = 95\%$$

$$Prob(P_{50} > MeanP_{50}) = 1 - Prob(P_{50} < MeanP_{50})$$
$$= 1 - N(1.64473) = 1 - 0.95 = 5\%$$

Therefore there's only a 5% chance of the stock price exceeding the mean expected future price of $33,199.03 in 50 years!

However, there's a 50% chance of the stock price exceeding the median expected future price of $148.41 in 50 years.

As you can see, the longer the time, the larger the difference between the mean and median, and the smaller the probability of exceeding the mean return.