STATISTICAL METHODS WITH APPLICATION TO FINANCE

a.y. 2025-2026

Behaviour of Returns: review of statistical concepts

R. PAPPADÀ
rpappada@units.it

Department of Economics, Business, Mathematics and Statistics "B. de Finetti"
University of Trieste

March 11, 2026

A first look at risk and return

What can we say about returns?

- Returns cannot be perfectly predicted, but rather they are random: this means that investing involves **risk**
- Without any assumptions, at time t, P_{t+1} (or R_{t+1}) would not only be unknown, but we would not know their probability distributions.
- If we are willing to make the assumption that future returns will be similar to past returns (*stationarity*), the probability distribution of R_t can be *estimated* from past data

The distribution of returns

We focus on returns, $\{R_1, R_2, \ldots, R_T\}$, on a single asset at times $t = 1, 2, \ldots, T$ or the corresponding log returns $r_t = \log(1 + R_t)$.

One of the major issues in finance is how we should model the probability distribution of returns. Show more

The common assumption that the returns from an asset are normally distributed is not supported by empirical studies that showed that the distribution of returns can exhibit asymmetries and heavy tails

Conditional and Unconditional Distribution

When modeling returns or other quantities of interest, we consider a sequence of random variables, say $X_1, X_2, \ldots$.

- If we assume that the series $\{X_t\}$ forms a *stationary* process (the distribution of $\{X_t\}_{t\in\mathbb{N}}$ is invariant under shifts of time), then we consider the random variable (r.v.) X with the same distribution as $X_1, \ldots, X_t$, denoted with F_X (unconditional distribution)

- In most financial time series, the conditional distribution of X_{t+1} given current information $\mathcal{F}_t$

$$X_{t+1}|\mathcal{F}_t$$

  is not equal to the stationary distribution F_X, hence specific models are required to capture the dynamics of the time series $\{X_t\}_{t\in\mathbb{N}}$

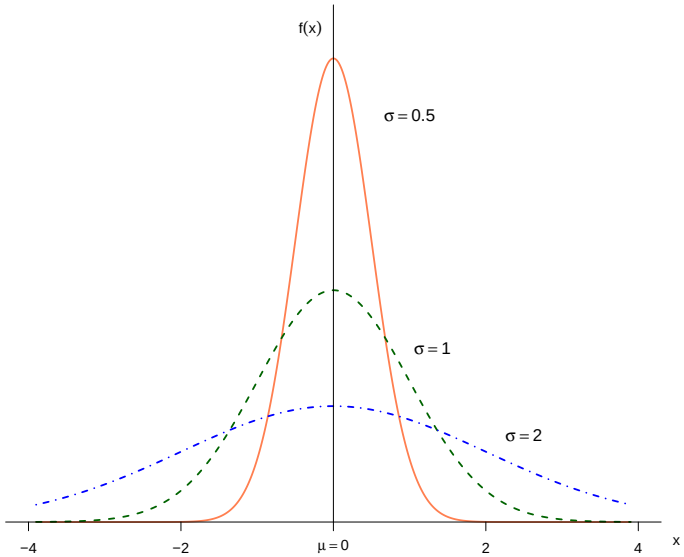
Table of Contents

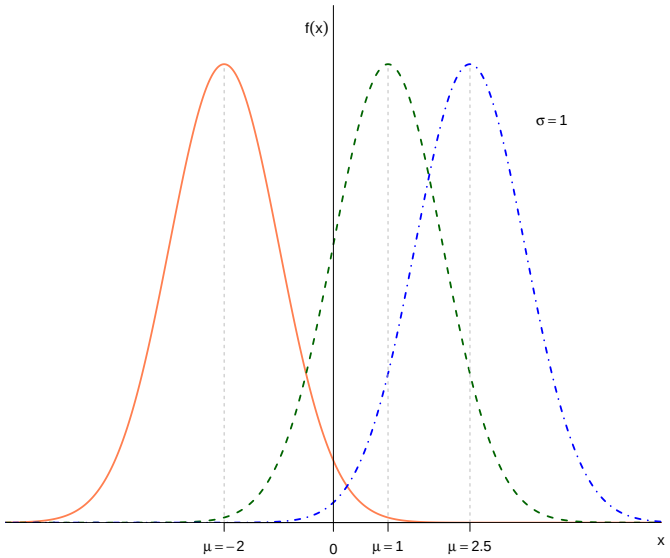
The role of the Normal distribution

A central role in economics and financial applications is played by the normal distribution. This is due to its mathematical tractability and several desirable characteristics:

- It is symmetric and completely specified by its mean and variance
- Any linear transformation of a normally distributed random variable will still be normally distributed
- The normal distribution may be suitable for approximating many phenomena given a sufficiently "large" sample size (*Central Limit Theorem*)

Normal distribution





$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Density $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$

$-\infty < x < \infty$

$f(x)$ maximal for $x = \mu$;

$f(x)$ symmetric around μ

parameters: $\mu \in \mathbb{R}$ (location), $\sigma > 0$ (scale)

moments: $\mathrm{E}(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$

Standard Normal: $\mu = 0, \sigma = 1$

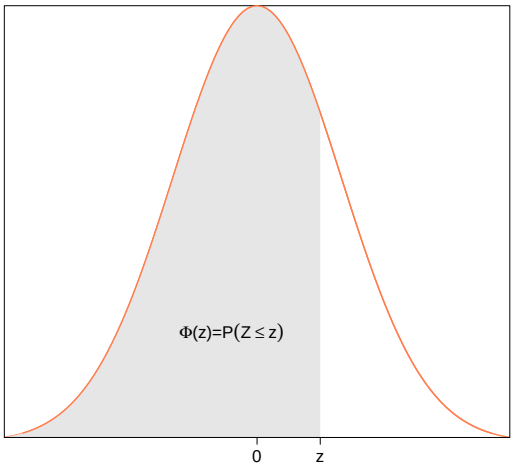
$$X \sim \mathcal{N}(\mu, \sigma^2) \rightarrow Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Show more

Standard Normal and normal quantiles

$$Z \sim \mathcal{N}(0,1)$$

PDF: $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$; CDF: $\Phi(z) = Pr(Z \leq z),\ -\infty < z < \infty$



probability (α)	*quantile* (z_α)
0.01	−2.326
0.05	−1.645
0.5	0.00
0.95	1.645
0.99	2.326

The probability that Z is below its α-quantile is precisely α:

$$Pr(Z \leq z_\alpha) = \Phi(z_\alpha) = \alpha$$

- quantiles can be obtained from the normal probability table
- $z_{1-\alpha} = -z_\alpha$
- $P(z_{\frac{0.05}{2}} \leq Z \leq z_{1-\frac{0.05}{2}}) = 0.95,\ z_{0.975} = 1.96$
- $X = \sigma Z + \mu \sim \mathcal{N}(\mu, \sigma^2)$
  $F(x) = P(\sigma Z + \mu \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

  Quantile of order q of X: $x_q = \mu + \sigma z_q$

The Chi-Square distribution

If X is a $\mathcal{N}(0,1)$ random variable, then the distribution of X^2 is called the *chi-squared distribution* with 1 degree of freedom.

Now, if $Z_1, Z_2, \ldots, Z_N$ are $\mathcal{N}(0,1)$, and independent, then the distribution of $X = Z_1^2 + Z_2^2 + \cdots + Z_N^2$ is the χ^2 distribution with N *degrees of freedom*:

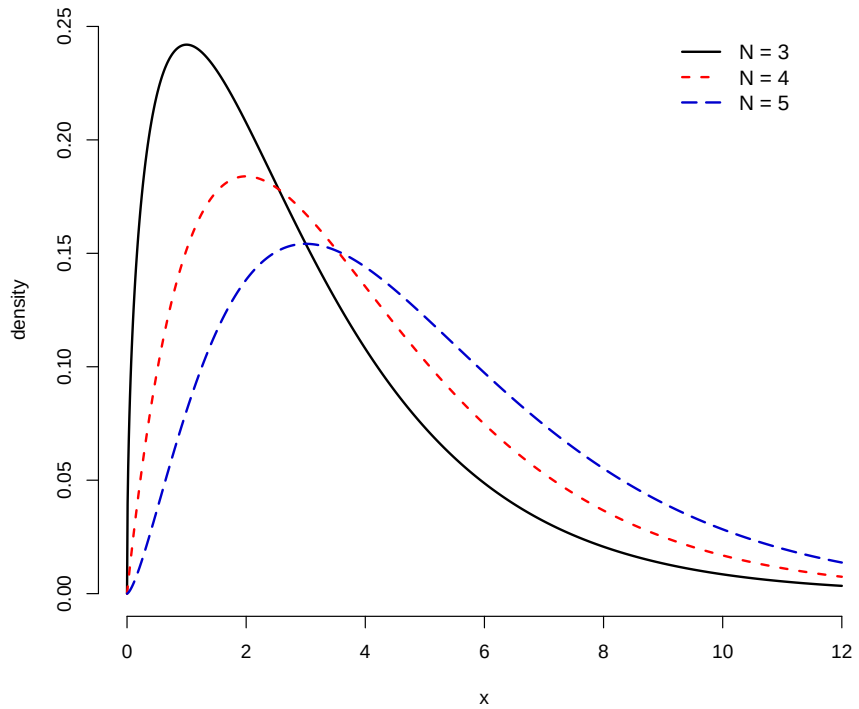
$$X \sim \chi^2_N$$

This distribution has an expected value of N and a variance of $2N$.

The χ^2 distribution is a skewed distribution, and has support $[0, \infty)$. It depends on a single parameter, the number of degrees of freedom. The skewness decreases as the degrees of freedom increase.

The Chi-Square distribution (cont)

Chi-Square distribution for some values of the degrees of freedom N.



The $(1-\alpha)$-quantile is denoted by $\chi^2_{1-\alpha,N}$ and is such that:

$$Pr(X \geq \chi^2_{1-\alpha,N}) = \alpha$$

Example: the 95% quantile of χ^2 with 10 degree of freedom is $\chi^2_{0.95,10} = 18.307$.

Central Limit Theorem

If $X_1, \ldots, X_n$ are independent r.v.'s with mean μ and finite variance σ^2, then the distribution of $S_n = X_1 + \cdots + X_n$ gets closer to a normal distribution as n converges to ∞. In particular,

$$Pr\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right) \to \Phi(z) \quad \text{as } n \to \infty \text{ for all } x \in \mathbb{R}$$

(recall that $\mathrm{E}(S_n) = n\mu$ and $\mathrm{Var}(S_n) = n\sigma^2$).

Stated differently, for large n, the sample mean has approximate distribution

$$\bar{X}_n = \frac{S_n}{n} \dot{\sim} \mathcal{N}(\mu, \sigma^2/n)$$

or

$$\frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \dot{\sim} \mathcal{N}(0, 1)$$

I.i.d. normal returns

A traditional assumption made in financial study is that the simple returns R_t $(t = 1, \ldots, T)$ are **independent and identically distributed** (iid) as normal with fixed mean and finite variance

$$R_t \sim iid\ \mathcal{N}(\mu, \sigma^2)$$

Problems with this assumption

- Because a normally distributed random variable can take any value between $-\infty$ and $+\infty$, the model implies the possibility of unlimited losses, the lower bound of a simple return is -1
- If R_t is normally distributed, then the multiperiod net return **is not** normally distributed, because it is a product of one-period returns
- the normality assumption is not supported by many empirical asset returns

Table of Contents

The Lognormal distribution

Another commonly used assumption is that the **log returns** r_t of an asset are i.i.d. normal with mean μ and variance σ^2. The simple net returns are then iid lognormal random variables.

The lognormal distribution.

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then the continuous variable

$$Y = e^X$$

is said to have a *Lognormal*(μ, σ^2) distribution. That is, Y is *lognormal* if

$$\log(Y) = X \sim \mathcal{N}(\mu, \sigma^2)$$

μ: log-mean (scale parameter);
σ^2 : log-variance (shape parameter)

Lognormal distributions

The lognormal r.v. can take on only nonnegative values.

$$f_Y(y) = \frac{1}{y\sqrt{2\pi}\sigma} e^{-(\log(y)-\mu)^2/(2\sigma^2)},$$

con $y > 0$.

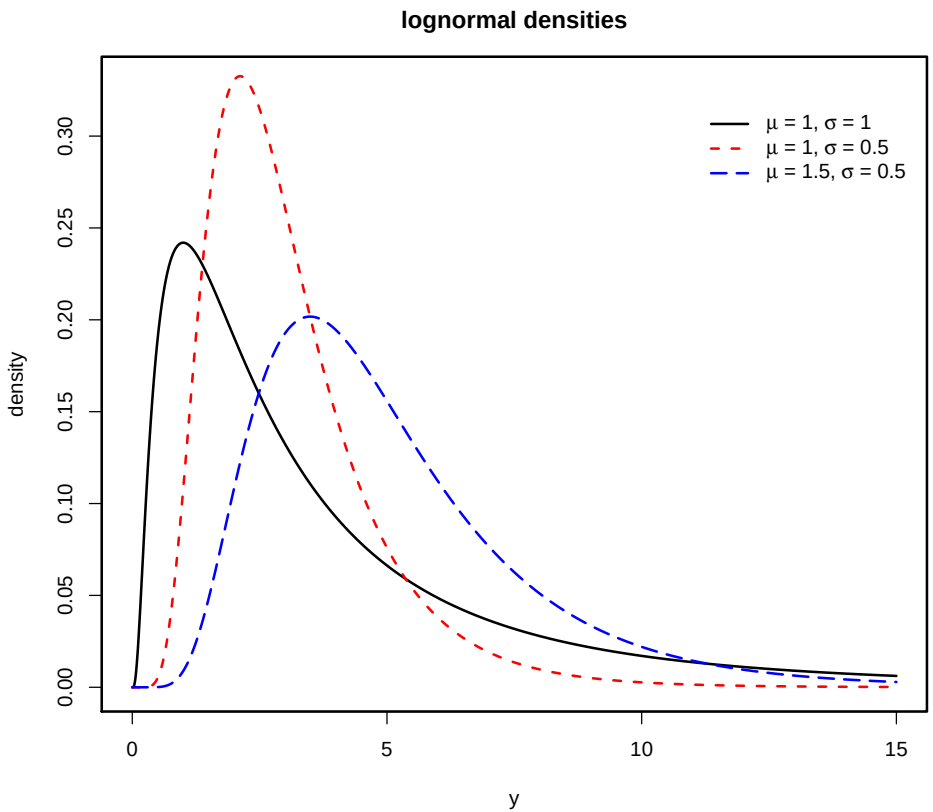
Positively skewed (right-skewed) distribution: long right tail compared to the left.

The larger the variance σ^2 of the associated normal distribution, the more skewed the lognormal distribution is.

Median $\text{Me}(Y) = e^{\mu}$

Mean $\text{E}(Y) = e^{(\mu+\sigma^2/2)}$

Variance $\text{Var}(Y) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$



The lognormal density

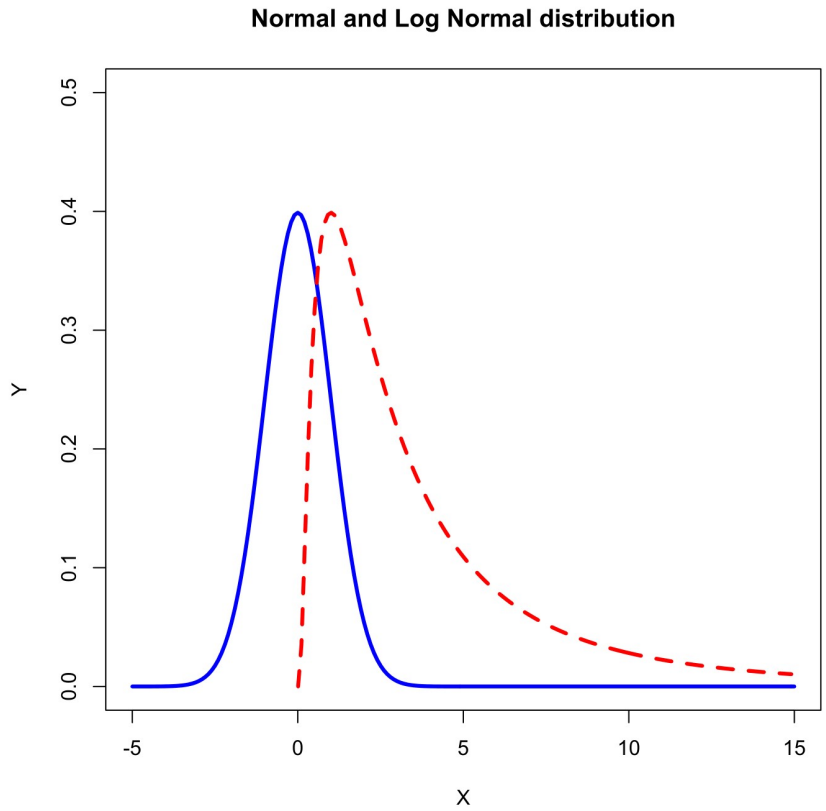


Figure 1: Density comparison of lognormally and normally distributed random variables.

Table of Contents

Heavy tails

The *tails* of a distribution are the regions far from the center.

- Heavy tails imply that there is a higher probability of extreme outcomes than one would get from the normal distribution with the same mean and variance
- Also implies that there is a lower probability of non-extreme outcomes
- Heavy-tailed distributions are of great interest because of the possibility of an extremely large negative return

The Student-t distribution

If $Z \sim \mathcal{N}(0,1)$, $U \sim \chi^2_\nu$, and Z, U are independent, then

$$T = \frac{Z}{\sqrt{U/\nu}} \sim t_\nu$$

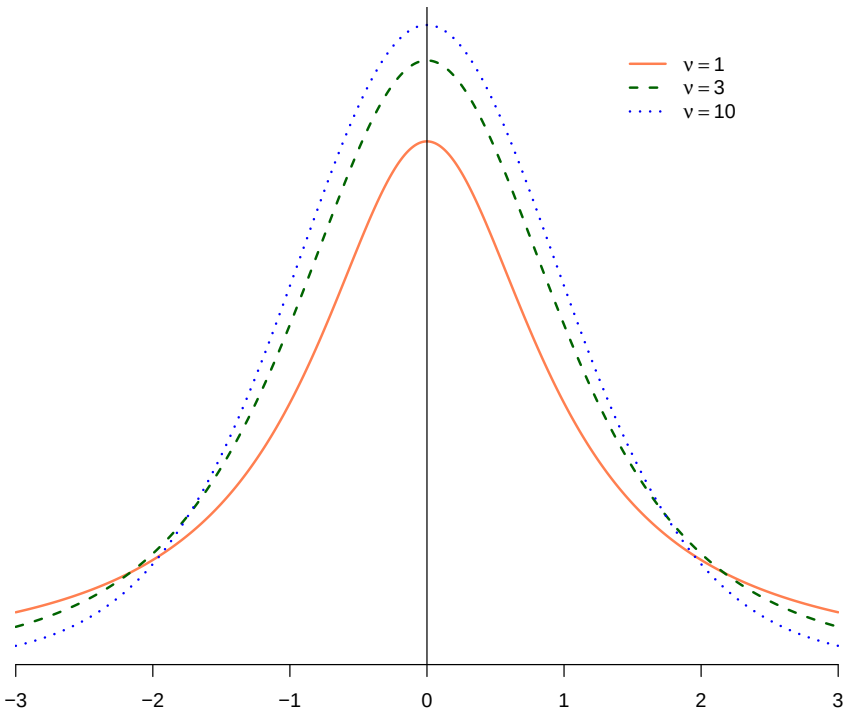
T has a Student-t distribution with $\nu > 0$ df.

$\mathrm{E}(T) = 0$ for $\nu > 1$

$\mathrm{Var}(T) = \nu/(\nu-2)$, for $\nu > 2$.

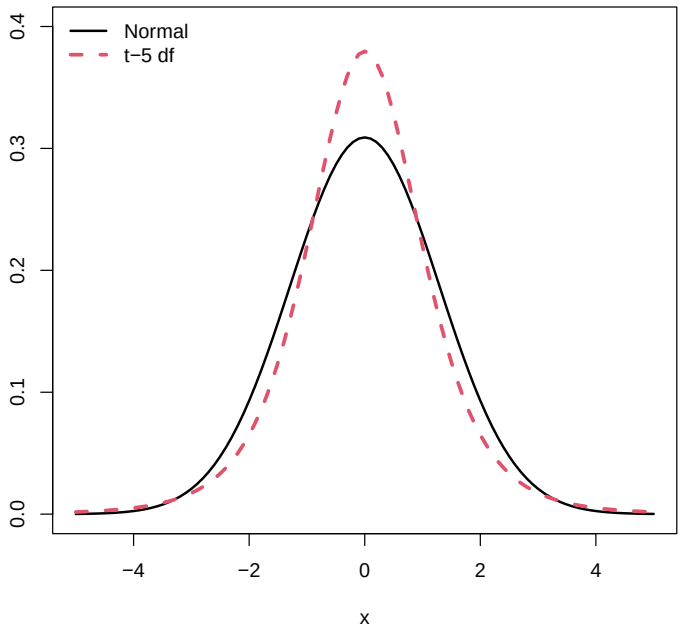
If $T \sim t_\nu$ ($\nu > 4$), then

$$Kur(T) = 3 + \frac{6}{\nu - 4}$$



The t distribution: properties

- the degrees of freedom of the Student-t distribution indicate how fat the tails are
- the kurtosis becomes smaller if ν increases
- As the degrees of freedom ν tend to infinity the t_ν distribution approximates the standard normal distribution
- in market risk, typically $\nu \leq 5$



Probability of extreme outcomes

October 19, 1987 saw global stock markets drop around 23%.

- If S&P 500 returns were normally distributed, the probability of a one-day drop of 23% would be 1.72×10^{-87}!
- The table below gives probabilities of different returns assuming normality and standard deviation 0.0116 (using daily returns from 1928 to 2009)

Returns above or below	Two tailed probability
1%	0.3886
2%	0.0847
3%	0.0097
5%	1.63×10^{-5}
15%	$< 10^{-35}$
23%	$< 10^{-80}$

Iid normal log-returns

Recall that $r_t = \log(1 + R_t)$. Hence,

$$1 + R_t = e^{r_t}.$$

If $r_t \sim \mathcal{N}(\mu, \sigma^2)$ then

$$1 + R_t = \exp(r_t) \sim Lognormal(\mu, \sigma^2)$$

The multiple-period return is:

$$\begin{aligned}\frac{P_t}{P_{t-k}} = 1 + R_t(k) &= (1 + R_t)(1 + R_{t-1}) \dots (1 + R_{t-k+1}) \\ &= \exp(r_t) \cdots \exp(r_{t-k+1}) \\ &= \exp(r_t + \cdots + r_{t-k+1})\end{aligned}$$

Therefore

$$\log(1 + R_t(k)) = r_t + \cdots + r_{t-k+1} \tag{1}$$

The lognormal model

If the log returns are assumed to be

- all normally distributed with mean μ and s.dev $\sigma > 0$
- mutually independent

then

$$\log(1 + R_t(k)) = \sum_{i=1}^{k} r_i \sim \mathcal{N}(k\mu, k\sigma^2)$$

that is, under the lognormal model assumptions, normality of single-period log returns implies normality of multiple-period log returns.

Remark Recall that if $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, and X and Y are independent, then

$$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

The lognormal model

From $\log(1 + R_t(k)) \sim \mathcal{N}(k\mu, k\sigma^2)$ we have

$$(1 + R_t(k)) \sim Lognormal(k\mu, k\sigma^2)$$

Then

$$\begin{aligned} Pr(1 + R_t(k) < x) &= Pr(\log(1 + R_t(k)) < \log(x)) \\ &= Pr\left(\frac{\log(1 + R_t(k)) - k\mu}{\sqrt{k}\sigma} < \frac{\log(x) - k\mu}{\sqrt{k}\sigma}\right) \\ &= \Phi\left(\frac{\log(x) - k\mu}{\sqrt{k}\sigma}\right) \end{aligned}$$

where $\Phi(\cdot)$ is the CDF of the standard normal.

The lognormal model: Example

Example: *A simple gross return,* $S = 1 + R$*, is* $Lognormal(0, 0.01)$*. What is the probability* $Pr(S < 0.9)$*?*

$\log(S)$ is $\mathcal{N}(0, 0.01)$. Since $\log(0.9) = -0.105$,

$$\begin{aligned} Pr(S < 0.9) &= Pr(\log(S) < -0.105) \\ &= \Phi\left(\frac{-0.105 - 0}{\sqrt{0.01}}\right) \\ &= \Phi(-1.05) = 0.1469 \end{aligned}$$

The lognormal model: Example (cont)

Now, assume that $1 + R \sim Lognormal(0, 0.01)$. *Also, assume that the returns are iid.*
What is the probability that a simple gross two-period return is less than 0.9?

The two-period gross return is $Lognormal$ with log-mean 0 and log-variance 2×0.01, so we find

$$\begin{aligned} Pr(1 + R(2) < 0.9) &= \Phi(\log(0.9)/\sqrt{0.02}) \\ &= \Phi(-0.75) = 1 - \Phi(0.75) \\ &= 1 - 0.773 \\ &= 0.2266 \end{aligned}$$

Geometric Random Walks

Recall that, from Eq.(1), we have

$$\log(P_t/P_{t-k}) = \log(1 + R_t(k)) = r_t + \cdots + r_{t-k+1}$$

Taking $k = t$,

$$\log(P_t/P_0) = r_t + r_{t-1} + \cdots + r_1$$

and the price of the asset at time t is given by

$$P_t = P_0\, e^{(r_t + r_{t-1} + \cdots + r_1)} \tag{2}$$

If the log returns $r_1, r_2, \ldots$ are iid $\mathcal{N}(\mu, \sigma^2)$, then P_t is lognormal for all t. In this case, the development of prices is said to follow a lognormal geometric random walk with parameters (μ, σ^2).

Geometric Random Walks

Eq. (2) implies that the model for the log price is

$$p_t = p_0 + r_t + r_{t-1} + \cdots + r_1 \tag{3}$$

where $p_t = \log(P_t)$ and p_0 denotes the initial log price.
The expectation and variance of p_t, conditional given p_0, are

$$\mathrm{E}(p_t|p_0) = p_0 + t\mu$$

and

$$\mathrm{Var}(p_t|p_0) = t\sigma^2$$

The constant term μ represents the time trend of the log price p_t (also called the *drift*).

Example

The log returns r_t on a stock have a mean of about 10% and a variance of about 20%. Assume a geometric random walk with $\mu = 0.1$ and $\sigma^2 = 0.2$ for the stock prices.

- The log return $r_t \sim \mathcal{N}(0.1, 0.2)$, then the return $1 + R_t$ has a lognormal distribution with log-mean $\mu = 0.1$ and log-variance $\sigma^2 = 0.2$;
- The expected log return on each step is $\mathsf{E}(r_t) = 0.1$, hence after k periods the mean and the median are

$$\mathsf{E}(r_t(k)) = \mathsf{Me}(r_t(k)) = 0.1k$$

- The median of the k-period gross return is

$$\mathsf{Me}(1 + R_t(k)) = e^{0.1k}$$

since $P(1 + R_t(k) \leq e^{0.1k}) = P(r_t(k) \leq 0.1k) = 0.5$.

- the median price after k years is

$$P_0 e^{0.1k}$$

where P_0 is the price at time 0.

Summing-up

The *random walk hypothesis* states that the single-period log returns are independent, that is future returns are independent of the past.

If the log returns are iid and normally distributed, that is $r_1, r_2, \ldots$ are *iid* $\mathcal{N}(\mu, \sigma^2)$, than the model for the price P_t is a *lognormal geometric random walk* with parameters μ and σ^2.

The parameter μ is called the *drift* and determines the general direction of the random walk.

Are prices a Lognormal Geometric Random walk?

The lognormal geometric random walk makes two assumptions:
(1) the log returns are normally distributed
(2) the log returns are mutually independent.

Empirical evidence shows at least some deviation from random walk. The distribution of log returns has generally heavier tails than normal tails: many stock returns exhibit a positive **excess kurtosis**

The independence assumption is also violated

- there is some correlation between returns
- returns exhibit *volatility clustering*, which means that if we see high volatility in current returns then we can expect this higher volatility to continue, at least for a while

Table of Contents

Probability Distributions

In the unconditional framework, we first suppose that returns all have the same distribution F_X and we are interested in methods for modeling and estimating the cumulative distribution function (df)

$$F_X(x) = Pr(X \leq x)$$

that is, the probability that, by the end of the period under consideration, the value of X is less than or equal to a given number x. If X is a **continuous** r.v. with df F_X, then

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du \tag{4}$$

for some non-negative function f, which is then known as its probability density function

- $\int_{-\infty}^{\infty} f_X(x)dx = 1$;
- $F_X(\cdot)$ is nondecreasing and $0 \leq F_X(x) \leq 1$;
- $Pr(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx$

Moments

The expectation, variance, skewness coefficient, and kurtosis of a random variable are all special cases of moments.

Let X be a continuous random variable and k denote a positive integer. The k-th moment of X is $E(X^k)$. The first moment is the expectation of X

$$E(X) = \mu = \int_{-\infty}^{\infty} xf(x)dx$$

The k-th central moment is

$$m_k = E[(X - E(X))^k]$$

so, for instance, m_2 is the variance of X:

$$V(X) = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$$

The positive square root of variance is the standard deviation of X: $\sigma = +\sqrt{V(X)}$

Skewness and Kurtosis

The third and fourth moments of X quantify its skewness and kurtosis, respectively.

- The skewness of X is the third moment of the X standardized

$$Sk(X) = \mathsf{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

  The distribution of X is said to be positively skewed, negatively skewed or symmetric depending on whether $Sk(X) > 0$, $Sk(X) < 0$ or $Sk(X) = 0$

- The kurtosis of X is

$$Kur(X) = \mathsf{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$$

The skewness measures the symmetry in the data, while the kurtosis measures the *peakedness* in the data.

Table of Contents

Joint and marginal distributions

Consider two continuous r.v's X and Y. The behavior of X and Y is characterized by the joint df $F_{X,Y}$:

$$F_{X,Y}(x,y) = Pr(X \leq x, Y \leq y) \tag{5}$$

If the joint probability density function $f_{X,Y}$ of X and Y exists, then

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)dudv \tag{6}$$

The *marginal* distribution of X, F_X, is obtained by integrating out Y (similarly for Y). The *marginal density* functions of X and Y are given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

Conditional distributions

We can make conditional probability statements about the probability that a variable takes certain values given that the other takes other values.

The conditional distribution of X given $Y \leq y$ is

$$F_{X|Y}(x|y) = \frac{Pr(X \leq x, Y \leq y)}{Pr(Y \leq y)} = \frac{F_{X,Y}(x,y)}{F_Y(y)} \tag{7}$$

The conditional density function of X given y is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \tag{8}$$

for $f_Y(y) > 0$. (Similar expressions can be derived for Y).

Independence

Suppose (X, Y) has a continuous joint density $f_{X,Y}(x, y)$, and the marginal densities are $f_X(x)$ and $f_Y(y)$, respectively.

Then X and Y are *independent* if and only if

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

or, equivalently,

$$F_{X,Y}(x, y) = F_X(x) F_Y(y)$$

In this case, we have

$$f_{X|Y}(x|y) = f_X(x), \quad f_{Y|X}(y|x) = f_Y(y),$$

that is, the conditional distribution and density of X (Y) given Y (X) are identical to the marginal distribution and density of X (Y).

Independence - n random variables

We say $X_1, X_2, \ldots, X_n$ are independent if only if

$$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

IID RVs: If the variables $X_1, X_2, \ldots, X_n$ are independent and have the same distribution, that is, they have common law F, then we say $X_1, \ldots, X_n$ are independent and identically distributed (iid) random variables.

Table of Contents

Covariance

The **covariance** between two variables X and Y is a measure of their linear relationship. If μ_X and μ_Y are the mean of X and Y, respectively, then

$$\text{Cov}(X, Y) = \text{E}[(X - \mu_X)(Y - \mu_Y)] = \text{E}(XY) - \mu_X \mu_Y$$

where the variances are assumed to be finite.

X and Y are said to be uncorrelated if $\text{Cov}(X, Y) = 0$.

$$X \text{ and } Y \textit{ independent} \rightarrow X \text{ and } Y \textit{ uncorrelated}$$

(in such a case $\text{E}(XY) = \mu_X \mu_Y$). The converse is not true in general!

Covariance: properties

The covariance between two random variables depends on their variances as well as the strength of the linear relationship between them.

The properties of the covariance are

- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(a + X, b + Y) = \text{Cov}(X, Y)$
- $\text{Cov}(aX + bY, cW + dZ) =$
  $ac \cdot \text{Cov}(X, W) + ad \cdot \text{Cov}(X, Z) + bc \cdot \text{Cov}(Y, W) + bd \cdot \text{Cov}(Y, Z)$

where a, b, c, d are constants and X, Y, W, Z are random variables.

Correlation

Given a bivariate sample $\{(X_i, Y_i)\}_{i=1}^n$, the *sample covariance* is

$$s_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1} \tag{9}$$

where $\bar{x}$ and $\bar{y}$ are the sample means.

Linear correlation is a pure measure of how closely two random variables are linearly related. The **Bravais-Pearson correlation coefficient** between X and Y is

$$Corr(X, Y) = \rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \tag{10}$$

and lies in $[-1, 1]$, where a value of ρ_{XY} close to 1 indicates a *positive* linear relationship, a value of -1 implies perfect negative dependence.

Correlation

The Bravais-Pearson *sample correlation* coefficient is computed as

$$\hat{\rho}_{XY} = \frac{s_{XY}}{s_X s_Y} \tag{11}$$

where s_X and s_Y are the sample standard deviations. This coefficient measures the strength of linear dependence between X and Y.

For two independent random variables, $\rho_{XY} = 0$. Note that $\rho_{XY} = 0$ does not imply independence. Indeed, when $\rho_{XY} = 0$, there is no linear relationship between X and Y–but not necessarily a lack of relationship.

Example: Let X is a uniform r.v. on $[-1, 1]$ and $Y = X^2$, then $\text{Cov}(X, Y) = 0$ and therefore $\rho_{XY} = 0$, but the two random variables are not independent.

Mean and variance of linear functions of rv's

First, we look at a linear function of a single random variable. If Y is a random variable and a and b are constants, then

$$\mathsf{E}(aY+b) = a\mathsf{E}(Y)+b; \qquad \mathsf{Var}(aY+b) = a^2\mathsf{Var}(Y)$$

Consider a linear combination of X and Y, $aX+bY$, we have

$$\mathsf{E}(aX+bY) = a\mathsf{E}(X)+b\mathsf{E}(Y)$$

and

$$\mathsf{Var}(aX+bY) = a^2\mathsf{Var}(X)+b^2\mathsf{Var}(Y)+2ab\mathsf{Cov}(X,Y)$$

If X and Y are uncorrelated, then $\mathsf{Var}(aX+bY) = a^2\mathsf{Var}(X)+b^2\mathsf{Var}(Y)$.

Remark if X and Y are independent, then they are uncorrelated and

$$\mathsf{Var}(X+Y) = \mathsf{Var}(X)+\mathsf{Var}(Y); \qquad \mathsf{Var}(X-Y) = \mathsf{Var}(X)+\mathsf{Var}(Y)$$

Sums of random variables

The following result is a generalization to n variables:

Let $X_1, X_2, \ldots, X_n$ be n random variables with common mean μ and variance σ^2. If the X_i $(i = 1, \ldots, n)$ are independent, or at least uncorrelated, then

$$\mathsf{E}(X_1 + X_2 + \cdots + X_n) = n\mu; \qquad \mathsf{Var}(X_1 + X_2 + \cdots + X_n) = n\sigma^2$$

In particular, it follows that if $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$ is the sample mean, then

$$\mathsf{E}(\bar{X}) = \mu; \qquad \mathsf{Var}(\bar{X}) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$