Lecture 4: Ideal probability metrics

Prof. Dr. Svetlozar Rachev

Institute for Statistics and Mathematical Economics
University of Karlsruhe

Portfolio and Asset Liability Management

Summer Semester 2008

Copyright

The material is based on the text-book:

Svetlozar T. Rachev, Stoyan Stoyanov, and Frank J. Fabozzi

Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization: The Ideal Risk, Uncertainty, and Performance Measures

John Wiley, Finance, 2007

Prof. Svetlozar (Zari) T. Rachev
Chair of Econometrics, Statistics
and Mathematical Finance
School of Economics and Business Engineering
University of Karlsruhe
Kollegium am Schloss, Bau II, 20.12, R210
Postfach 6980, D-76128, Karlsruhe, Germany
Tel. +49-721-608-7535, +49-721-608-2042(s)
Fax: +49-721-608-3811
http://www.statistik.uni-karslruhe.de

Introduction

Limit theorems in probability theory have a long and interesting history:

- In 1713, the Swiss mathematician Jacob Bernoulli gave a rigorous proof that the average number of heads resulting from many tosses of a coin converges to the probability of having a head.
- In 1835, the French mathematician Simeon-Denis Poisson described this result as “The Law of Large Numbers” and formulated an approximation valid in the case of rare events. Nowadays, this result is known as the approximation of Poisson to the binomial distribution.

Introduction

- In 1733, the English mathematician Abraham de Moivre published an article in which he calculated approximately the probability of the number of heads resulting from many independent tosses of a fair coin. In this calculation, he used the normal distribution as approximation.
- His discovery was rediscovered and extended in 1812 by the French mathematician Pierre-Simon Laplace. Nowadays, this is known as the theorem of de Moivre-Laplace of the normal approximation to the binomial distribution.
- The theorem of de Moivre-Laplace is a special case of the Central Limit Theorem (CLT) but it was not until the beginning of the twentieth century.

Introduction

- In 1901, the Russian mathematician Aleksandr Lyapunov gave a more abstract formulation and showed that the limit result holds under certain very general conditions known as Lyapunov's conditions. Later, other conditions were established which generalized Lyapunov's conditions.
- A final solution to the problem was given by Bernstein, Lindeberg, and Feller who specified necessary and sufficient conditions for the CLT.

In the past century, the limit theory has been widely extended.

- The abstract ideas behind the CLT were applied to stochastic processes and it was shown that Brownian motion is the limit process in the so-called Functional Limit Theorem or Invariance Principle.
- Brownian motion is the basic ingredient of the subsequently developed theory of Ito processes which has huge application in finance. The celebrated Black-Scholes equation and, in general, derivative pricing is based on it.

- <u>The idea of Generalized CLT is as follows:</u> When summing i.i.d. *infinite* variance random variables we do not obtain the normal distribution at the limit but so called the Lévy alpha-stable distributions because of the fundamental work of the French mathematician Paul Lévy. The normal distribution is only a special case of the stable distributions.
- The limit theory of max-stable distributions was developed which studies the limit distribution with respect to maxima of random variables.

From the standpoint of the applications, the limit theorems are appealing because the limit law can be regarded as an approximate model of the phenomenon under study.

- For example, the limiting normal distribution (the result of de Moivre-Laplace) can be accepted as an approximate model for calculation of the number of heads in many tosses of a fair coin.
- In the max-stable scheme, the limiting max-stable distribution can be regarded as an approximate model for the maximum loss a financial institution may face in a given period of time.
- Similarly, in modeling returns for financial assets, the alpha-stable distributions can be used as an approximate model as they generalize the widely applied normal distribution and are the limiting distribution in the Generalized CLT.

Introduction

How close is the limiting distribution to the considered phenomenon?

The only way to answer this question is to employ the theory of probability metrics. In technical terms, we are looking for a way to estimate the rate of convergence to the limit distribution.

The following concepts will be considered:

- The classical CLT
- The Berry-Esseen result (the first attempt to estimate the rate of convergence to the normal distribution)
- The convergence rate estimation in the Generalized CLT based on probability metrics called ideal metrics

- The goal is to illustrate the classical CLT.
- We consider the simple experiment of flipping an unfair coin.
- We are interested in calculating the probability that the number of heads resulting from a large number of independent trials belongs to a certain interval, i.e. if we toss a coin 10,000 times, then what is the probability that the number of heads is between 6,600 and 7,200 provided that the probability of a head is equal to 2/3?

The binomial approximation to the normal distribution

- Let us first derive a simple formula which gives the probability that we obtain exactly a given number of heads.
- Consider a small number of independent tosses, for example four.
- Denote by p the probability that a head occurs in a single experiment, by $q = 1 - p$ the probability that a tail occurs in a single experiment, and by X the random variable indicating the number of heads resulting from the experiment.
- Then the probability that no head occurs is given by

$$P(X = 0) = q \cdot q \cdot q \cdot q = q^4$$

because we multiply the probabilities of the outcomes since we assume independent trials.

The binomial approximation to the normal distribution

The probability that exactly one head occurs is a little more difficult to calculate.

- If the head occurs on the very first toss, then the probability of the event "exactly one head occurs" is equal to $p \cdot q \cdot q \cdot q = pq^3$.
- If the head occurs on the second trial, the corresponding probability is $q \cdot p \cdot q \cdot q = pq^3$.
- Similarly, if the head occurs on the very last trial, the probability is $q \cdot q \cdot q \cdot p = pq^3$.
- The probability of the event "exactly one head occurs in a sequence of four independent tosses" is equal to the sum of the probabilities of the events in which we fix the trial when the head occurs,

$$P(X = 1) = 4pq^3.$$

The binomial approximation to the normal distribution

- Similar reasoning shows that the probability of the event "exactly two heads occur in a sequence of four independent tosses" equals

$$P(X = 2) = 6p^2q^2$$

  as there are six ways to obtain two heads in a row of four experiments.

- For the other two events, that the heads are exactly three and four, we obtain

$$P(X = 3) = 4p^3q \qquad \text{and} \qquad P(X = 4) = p^4.$$

The binomial approximation to the normal distribution

- The power of p coincides with the number of heads, the power of q coincides with the number of tails, and the coefficient is the number of ways the given number of heads may occur in the experiment.
- The coefficient can be calculated by means of a formula known as the binomial coefficient.
- It computes the coefficients which appear in front of the unknowns when expanding the expression $(x+y)^n$ in which x and y are the unknowns. For example,

$$\begin{aligned}(x+y)^4 &= (x+y)^2 \cdot (x+y)^2 \\ &= (x^2+2xy+y^2)\cdot(x^2+2xy+y^2) \\ &= x^4+4x^3y+6x^2y^2+4xy^3+y^4.\end{aligned}$$

The binomial approximation to the normal distribution

- The general formula is given by the equation

$$(x+y)^n = \binom{n}{0}x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \dots + \binom{n}{n-1}xy^{n-1} + \binom{n}{n}y^n \tag{1}$$

in which n is a positive integer and the coefficients $\binom{n}{k}$ and $k = 0, \dots, n$ are the binomial coefficients.

- They are calculated through the formula,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{2}$$

where the notation $n!$ stands for the product of all positive integers smaller or equal to n, $n! = 1 \cdot 2 \dots (n-1) \cdot n$.

The binomial approximation to the normal distribution

- In the context of the independent tosses of a coin, n stands for the total number of tosses and k denotes the number of heads.
- Thus, the probability of the event "exactly two heads occur in a sequence of four independent tosses" can be written as

$$P(X = 2) = \binom{4}{2} p^2 q^2 = \frac{1 \cdot 2 \cdot 3 \cdot 4}{(1 \cdot 2) \cdot (1 \cdot 2)} p^2 q^2 = 6p^2q^2.$$

- Denote the random variable by X_n, where n stands for the number of trials, we obtain

$$P(X_n = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1 \ldots n. \tag{3}$$

The binomial approximation to the normal distribution

- The probability distribution defined in equation (3) is known as the binomial distribution.
- Replace the tossing of a coin by an experiment in which we identify a certain event as "success". All other events do not lead to "success" and we say that "failure" occurs. Thus, the binomial distribution gives the probability that exactly k "successes" occur on condition that we carry out n experiments.
- The mean value of the binomial distribution equals $EX = np$ and the variance equals $DX = npq$.

The binomial approximation to the normal distribution

- Let us consider the experiment of tossing a fair coin, i.e. $p = q = 1/2$, and fix the number of tosses to 20.
- What is the probability that exactly 4 heads occur? We can easily calculate this by means of equation (3),

$$P(X_{20} = 4) = \binom{20}{4} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{16} \approx 0.46\%$$

  Table below gives the corresponding probabilities for other choices of the number of heads.

Number of heads, k	4	7	10	13	16
Probability, $P(X = k)$	0.46%	7.39%	17.62%	7.39%	0.46%

Table: The probability that exactly k heads occur resulting from 20 independent tosses of a fair coin.

The binomial approximation to the normal distribution

Figure below graphically displays all probabilities, when the number of heads range from zero to 20.

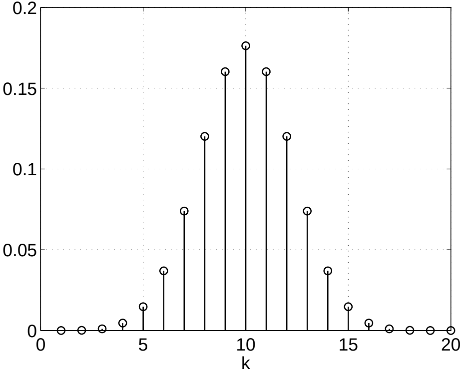

Figure: The probabilities that exactly k heads occur in 20 independent tosses of a fair coin.

The binomial approximation to the normal distribution

- Note that the probabilities change in a symmetric way around the value mean value $k = 10$, which very much resembles the density of the normal distribution. This similarity is by no means random. As the number of experiments, n, increases, the similarity becomes more and more evident.
- The limit theorem which proves this fact is known as the theorem of de Moivre-Laplace. It states that, for large values of n, the probability that k heads occur equals approximately the density function of a normal distribution evaluated at the value k.
- The mean value of the normal distribution is np and the standard deviation is $\sqrt{npq}$, in short-hand notation $N(np, npq)$. The density of a normal distribution with mean m and variance σ^2 is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

The binomial approximation to the normal distribution

- Therefore, the limit result states that

$$P(X_n = k) = \binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k-np)^2}{2npq}\right) \qquad (4)$$

  for large values of n.

- We can say that the normal distribution can be adopted as an approximate model because the probabilistic properties of the binomial distribution for large values of n are "close" to the probabilistic properties of the normal distribution.

The binomial approximation to the normal distribution

- Are there other indications that the normal distribution can be adopted as an approximate model? The answer is affirmative.
- Consider the question we asked at the beginning of the lecture on slide 10. Provided that n is large, what is the probability that the number of heads resulting from independent tosses of an unfair coin is between two numbers a and b?
- For example, suppose that we toss an unfair coin 10,000 times. Then, what is the probability that the number of heads is between $a = 6{,}600$ and $b = 7{,}200$?

The binomial approximation to the normal distribution

- Suppose that we independently toss an unfair coin twenty times. The probability of the event that no more than three heads occur can be computed by summing the probabilities $P(X_{20} = 0)$, $P(X_{20} = 1)$, $P(X_{20} = 2)$, and $P(X_{20} = 3)$.
- Similarly, in order to calculate the the probability that the number of heads is between a = 6,600 and b = 7,200 in 10,000 tosses, we have to sum up the probabilities $P(X_{10000} = k)$ where $6{,}600 \leq k \leq 7{,}200$. This is not a simple thing to do.
- The limit result in the theorem of de Moivre-Laplace can be adapted to calculate such probabilities. We can use the limiting normal distribution in order to calculate them,

$$P(a \leq X_n \leq b) \approx \int_a^b \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(x-np)^2}{2npq}\right) dx \qquad (5)$$

which means that instead of summing up the binomial probabilities, we are summing up the normal probabilities.

The binomial approximation to the normal distribution

- The calculation of the right hand-side of (5) is easier because it can be represented through the cumulative distribution function (c.d.f.) of the normal distribution,

$$P(a \leq X_n \leq b) \approx F(b) - F(a)$$

  where $F(x)$ is the c.d.f. of the normal distribution with mean np and variance npq.
- The c.d.f. of the normal distribution is tabulated and is also available in software packages.
- In fact, if we assume that $p = 2/3$, then the actual probability, $P(6{,}600 \leq X_{10{,}000} \leq 7{,}200) = 0.9196144$ and through the corresponding normal distribution we obtain $F(7{,}200) - F(6{,}600) = 0.92135$ which means that we make an error of about 0.17%.

The binomial approximation to the normal distribution

- Another implication of (5) means that the c.d.f. of the binomial distribution is approximated by the c.d.f. of the corresponding normal distribution,

$$P(X_n \leq b) \approx F(b).$$

- It virtually means that the probabilistic properties of the binomial distribution are approximately the same as the ones of the normal distribution.
- This is illustrated on the next slide, where we plot the c.d.f. of a binomial distribution resulting from 200 independent tosses of a fair coin and the corresponding normal approximation.

The binomial approximation to the normal distribution

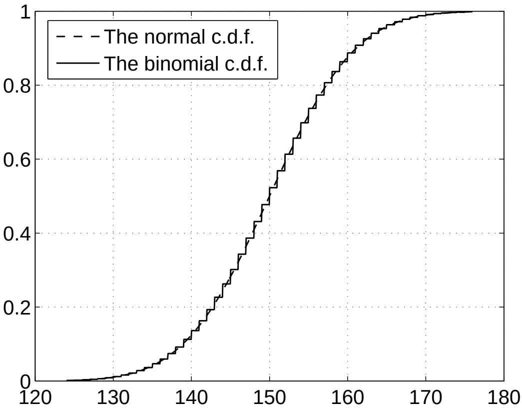

Figure: The binomial c.d.f. resulting from 200 independent tosses of a fair coin and the normal approximation.

The binomial approximation to the normal distribution

- Generalizations of approximations (4) and (5) are used in pricing options and other derivatives, the price of which depends on another instrument called underlying instrument.
- The binomial distribution is behind the construction of binomial trees employed to evolve the price of the underlying into the future. The basic principle is that, as the steps in the tree increase (the number of trials), the binomial path becomes closer to a sample path of the price process of the underlying instrument.
- Therefore, this technique provides a powerful numerical way to pricing path-dependent derivatives.

The general case

- Usually, the convergence to the normal distribution is derived by means of centered and normalized binomial distributions.
- We considered the binomial distribution directly, while in this setting, it is not possible to obtain a non-degenerate limit as the number of trials approaches infinity.

  This is because the mean value of np and the variance npq explode and the normal approximation $N(np, npq)$, which is well-defined for any finite n, stops making any sense.

The general case

- The procedure of centering and normalizing a random variable means that we subtract the mean of the random variable and divide the difference by its standard deviation so that the new random quantity has a zero mean and a unit variance.
- For instance, in the case of the binomial distribution, the random quantity

$$Y_n = \frac{X_n - np}{\sqrt{npq}}$$

has a zero mean and unit variance, $EY_n = 0$ and $DY_n = 1$.
- Therefore, it makes more sense to consider the limit distribution of Y_n as n approaches infinity because it may converge to a non-degenerate limit distribution as its mean and variance do not depend on the number of trials.

The general case

- In fact, the approximation in equation (5) is an illustration of the limit result

$$\lim_{n\to\infty} P\left(u \le \frac{X_n - np}{\sqrt{npq}} \le v\right) = \frac{1}{\sqrt{2\pi}} \int_u^v e^{-x^2/2} dx \qquad (6)$$

which means that as the number of trials approaches infinity, the centered and normalized binomial distribution approaches the standard normal distribution $N(0, 1)$.

- By observing the centered binomial distributions, we can visually compare the improvement in the approximation as the scale is not influenced by n. This is shown on the next slide.

The general case

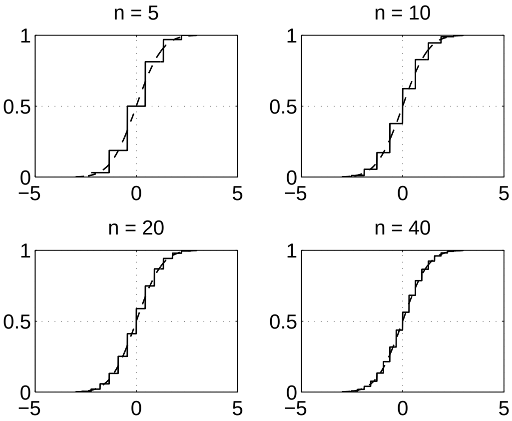

Figure: The centered and normalized binomial c.d.f.s resulting from 5, 10, 20, and 40 independent tosses of a fair coin and the normal approximation.

The general case

- Suppose now that in the *n*-tosses experiment we look at each toss separately. That is, each toss is a random variable that can take only two values — zero with probability q (if a tail occurs), and one with probability p (if a head occurs).
- Since each toss is a new experiment in itself, we denote these random variables by δ_i, i is the number of the corresponding toss.
- If δ_2 takes the value zero, it means that on the second toss, a tail has occurred.
- In this setting, the random variable X_n describing the number of heads resulting from n independent tosses of a coin can be represented as a sum of the corresponding single-toss experiments,

$$X_n = \delta_1 + \delta_2 + \ldots + \delta_n \tag{7}$$

where the random variables δ_i, $i = 1, \ldots, n$ are i.i.d.

The general case

- It appears that the limit relation (6) concerns a sum of i.i.d. random variables in which the number of summands approaches infinity.
- It turns out that (6) holds true for sums of i.i.d. random variables, just as in (7), the distribution of which may be quite arbitrary.
- This result is the celebrated *Central Limit Theorem*.
- There are several sets of regularity conditions, only two of them will be described below, as they have vast implications.

The meaning of summation in financial variables

Before proceeding with the regularity conditions, let us discuss briefly why summing random variables is important in the context of finance.

- A huge topic in finance is imposing a proper distributional assumption for the returns of a variable such as stock returns, exchange rate returns, changes in interest rates, and the like.
- Usually, the distributional hypothesis concerns the logarithmic returns in particular or the changes in the values which are also known as the increments.

The meaning of summation in financial variables

- Let us consider the price P_t of a common stock. The logarithmic return, or simply the log-return, for a given period (t, T) is defined as

$$r_{(t,T)} = \log \frac{P_T}{P_t}.$$

- If the period (t, T) is one month, then $r_{(t,T)}$ is the monthly log-return.
- We split this period into two smaller periods (t, t_1) and (t_1, T). The log-return of the longer period is actually the sum of the log-returns of the shorter periods,

$$r_{(t,T)} = \log \frac{P_{t_1}}{P_t} + \log \frac{P_T}{P_{t_1}} = r_{(t,t_1)} + r_{(t_1,T)}.$$

- Then we can split further the time interval and we obtain that the log-return of the longer period is the sum of the log-returns of the shorter periods.
- Thus, the monthly log-return is the sum of the daily log-returns.
- The daily log-returns are the sum of the ten-minute log-returns in one day, etc.

$\Rightarrow$ The general rule is that the lower frequency log-returns accumulate the corresponding higher frequency log-returns.

The meaning of summation in financial variables

Exactly the same conclusion holds with respect to the increments.

- Consider an interest rate in a period (t, T). The increments are defined as,

$$\Delta IR_{(t,T)} = IR_T - IR_t$$

  which is simply the difference between the interest rate at moment t and T.

- Splitting the interval into two smaller intervals results in

$$\Delta IR_{(t,T)} = IR_{t_1} - IR_t + IR_T - IR_{t_1} = \Delta IR_{(t,t_1)} + IR_{(t_1,T)}$$

  meaning that the increment in the longer period equals the sum of the increments in the smaller period.

The meaning of summation in financial variables

- The concept that a variable accumulates the efects of other variables is natural in finance.
- This observation makes the limit theorems in probability theory appealing because they show the limit distribution of sums of random variables without the complete knowledge of the distributions of the summands.
- Nevertheless, there are certain regularity conditions which the summands should satisfy in order for the sum to converge to a particular limit distribution.

Two regularity conditions

- Suppose that the random variables $X_1, X_2, \ldots, X_n, \ldots$ are independent and share a common distribution with mean μ and variance σ^2. Consider their sum

$$S_n = X_1 + X_2 + \ldots + X_n. \tag{8}$$

- The CLT states that the centered and normalized sequence of S_n converges to the standard normal distribution as n approaches infinity on condition that the variance σ^2 is finite.
- The mean of the sum equals the sum of the means of the summands,

$$ES_n = EX_1 + EX_2 + \ldots + EX_n = n\mu.$$

Two regularity conditions

- The same conclusion holds for the variance because the summands are assumed to be independent,

$$DS_n = DX_1 + DX_2 + \ldots + DX_n = n\sigma^2.$$

- Thus, subtracting the mean and dividing by the standard deviation, we obtain the statement of the CLT,

$$\lim_{n\to\infty} P\left(u \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq v\right) = \frac{1}{\sqrt{2\pi}} \int_u^v e^{-x^2/2} dx. \qquad (9)$$

Two regularity conditions

- The truly striking implication of the CLT is that the result is, to a large extent, invariant on the distributions of the summands.
- The distributions only need to be i.i.d. and their variance needs to be finite, $\sigma^2 < \infty$.
- The common distribution of the summands may be discrete, see equation (6) for the binomial distribution, or skewed, or it may have point masses.

$\Rightarrow$ The limit distribution is the standard normal law.

Two regularity conditions

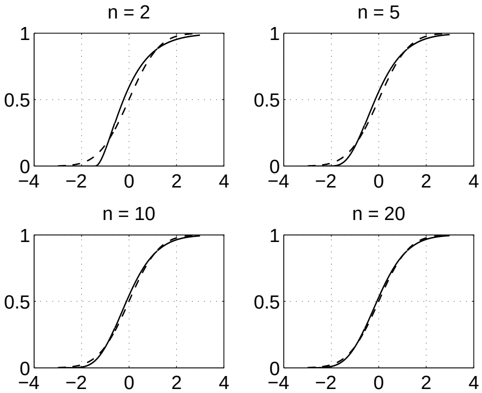

Figure: The c.d.f.s of the centered and normalized sum of exponential distributions (solid line) resulting from 2, 5, 10, and 20 summands and the normal approximation (dashed line). The exponential distribution by definition takes only positive values, which means that it is also asymmetric.

Two regularity conditions

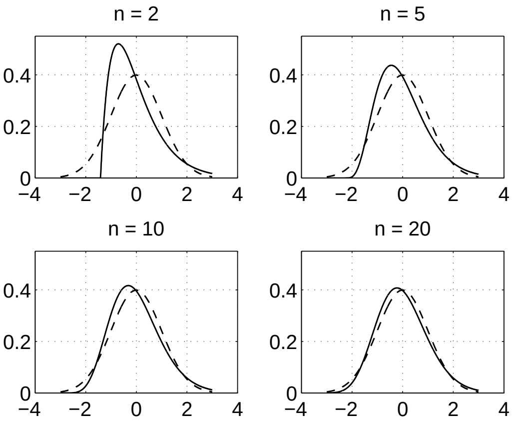

Figure: The density functions of the centered and normalized sum of exponential distributions (solid line) resulting from 2, 5, 10, and 20 summands and the normal approximation (dashed line).

Two regularity conditions

- The CLT states that the distribution function of the centered and normalized sum converges to the distribution function of the standard normal distribution.
- Thus, for a large number of summands,

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq v\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{v} e^{-x^2/2} dx.$$

  which is the same as saying that the c.d.f of a normal distribution with mean $n\mu$ and variance $n\sigma^2$ is close to the distribution function of the sum S_n.
- When the number of summands is large, the normal distribution can be accepted as an approximate model because the probabilistic properties of the $N(n\mu, n\sigma^2)$ are “close” to the probabilistic properties of the corresponding sum.

Two regularity conditions

- The fact that the CLT holds when the summands in (8) are i.i.d. with finite variance is already a very strong result with far-reaching consequences.
- The assumption of common distribution can be replaced by a different property. It states that as n grows to infinity, the summands should become negligible with respect to the total sum. That is, the contribution of each summand to the sum should become more and more negligible as their number increases. This property is called asymptotic negligibility.
- The summands need not have a common distribution. Some of them may be discrete random variables, some may have symmetric distribution, others asymmetric.

$\Rightarrow$ The only conditions the summands have to satisfy in order for the CLT to hold is, first, they have to be independent and, second, they have to be asymptotically negligible.

Application of the CLT in modeling financial assets

Let us go back to the discussion of the behavior of financial variables.

- If the daily log-returns appear as a sum of so many short period log-returns, can we safely assume, on the basis of the CLT, that the distribution of the daily log-return is approximately normal?
- Such a direct application of the limit result is not acceptable because there are certain conditions which need to be satisfied before we can say that the limit result holds.

We have to answer two questions:

1. Is it true that the shorter period log-returns are independent?
2. Are they asymptotically negligible? Is it true that if we sum them up, the total sum is not dominated by any of the summands?

- The answer to the first question is negative because of the empirically observed clustering of the volatility effect and the autocorrelations existing in the high-frequency time series.
- The answer to the second question is also negative. Usually, there are very large log-returns in absolute value which dominate the sum and dictate its behavior. They translate into what is known as the heavy-tailed behavior of the log-returns time series of stock prices.

While the autocorrelations and the clustering of the volatility can be taken care of by advanced time-series models, the outliers available in the data creep into the residual and, very often, can only be modeled by a non-normal, heavy-tailed distribution.

$\Rightarrow$ As a result, we can reject the normal distribution as a realistic approximate model of the log-returns of stock prices.

Estimating the distance from the limit distribution

- Under mentioned 2 sets of conditions, we can adopt the normal distribution as an approximate model for the sum of random variables (8) when the number of summands is large.
- As was explained, the rationale is that the distribution function of the sum with the number of summands fixed is “close” to the distribution function of the corresponding normal distribution.
- We say “the corresponding normal distribution” because its mean and variance should equal the mean and variance of the sum.

Estimating the distance from the limit distribution

- We would like to quantify the phenomenon that the two c.d.f.s do not deviate too much. For this purpose, we take advantage of a probability metric which computes the distance between the two c.d.f.s. and is, therefore, a simple metric.
- For example, suppose that we would like to fix the number of summands to 20. If the distribution of the summands is symmetric, then we may expect that the sum of 20 terms could be closer to the normal distribution compared to a sum of 20 asymmetric terms.
- Therefore, we need a way to estimate the error of adopting the limit distribution as a model which is not influenced by the particular distribution of the summands.

Estimating the distance from the limit distribution

- In the classical setting, when the sum (8) consists of i.i.d. summands, there is a result[1] which states how quickly the distance between the c.d.f. of the centered and normalized sum and the c.d.f. of the standard normal distribution decays to zero in terms of the Kolmogorov metric.
- Denote by $\widetilde{S}_n$ the centered and normalized sum,

$$\widetilde{S}_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}.$$

[1] In the theory of probability, this result is known as the Berry-Esseen theorem.

Estimating the distance from the limit distribution

- The result states that if $E|X_1|^3 < \infty$, then in terms of the Kolmogorov metric the distance between the two c.d.f.s can be bounded by,

$$\rho(\widetilde{S}_n, Z) \leq \frac{C \cdot E|X_1|^3}{\sigma^3 \sqrt{n}} \tag{10}$$

  in which C is an absolute constant which does not depend on the distribution of X_1, $Z \in N(0,1)$, and the Kolmogorov metric ρ is defined as

$$\rho(\widetilde{S}_n, Z) = \sup_{x \in \mathbb{R}} |F_{\widetilde{S}_n}(x) - F_Z(x)|.$$

- The only term which depends on n is $\sqrt{n}$ in the denominator. The only facts about the common distribution of the summands we have to know are the standard deviation σ and the moment $E|X_1|^3$.

Estimating the distance from the limit distribution

- As a result, the "speed" with which the c.d.f. $F_{\widetilde{S}_n}(x)$ approaches $F_Z(x)$ as the number of summands increases, or the convergence rate, is completely characterized by $n^{-1/2}$.
- We also need the value of the constant C. Currently, its exact value is unknown but it should be in the interval $(2\pi)^{-1/2} \leq C < 0.8$.
- At any rate, an implication of the inequality (10) is that the convergence of the c.d.f. of $\widetilde{S}_n$ to the c.d.f. of the standard normal distribution may be quite slow.

- What happens if the condition in the classical CLT is relaxed; if the summands are so erratic that one of them can actually dominate the others and thus influence the behavior of the entire sum?
- The normal distribution is not the limit law under these conditions but still there are non-degenerate limit distributions - stable distributions. The limit theorem is a generalization of the CLT and is known as the Generalized CLT.

The Generalized Central Limit Theorem

- Any properly centered and normalized sum of i.i.d. random variables converges at the limit to a stable distribution. This means that the stable distributions are the *only* distributions which can arise as limits of sums of i.i.d. random variables.
- This feature makes the stable distributions very attractive for the modeling of financial assets because only they can be used as an approximate model for sums of i.i.d. random variables.

$\Rightarrow$ The normal distribution is a special case of the stable distributions, just as the CLT is a special case of the Generalized CLT.

- In contrast to the normal distribution, the class of non-normal stable distributions has skewed and heavy-tailed representatives. Because of these differences, stable non-normal laws are also called stable Paretian or Lévy stable.

Stable distributions

- A random variable X is said to have a stable distribution if there are parameters $0 < \alpha \leq 2$, $\sigma > 0$, $-1 \leq \beta \leq 1$, $\mu \in \mathbb{R}$ such that its characteristic function $\varphi_X(t) = Ee^{itX}$ has the following form

$$\varphi_X(t) = \begin{cases} \exp\{-\sigma^\alpha |t|^\alpha (1 - i\beta \frac{t}{|t|} \tan(\frac{\pi\alpha}{2})) + i\mu t\}, & \alpha \neq 1 \\ \exp\{-\sigma |t| (1 + i\beta \frac{2}{\pi} \frac{t}{|t|} \ln(|t|)) + i\mu t\}, & \alpha = 1 \end{cases} \tag{11}$$

where $\frac{t}{|t|} = 0$ if $t = 0$ and

α	is called the *index of stability* or the *tail exponent*
β	is a skewness parameter
σ	is a scale parameter
μ	is a location parameter

- Since stable distributions are uniquely determined by the four parameters, the common notation is $S_\alpha(\sigma, \beta, \mu)$.

Stable distributions

- The parameter α determines how heavy the tails of the distribution are. That is why it is also called the tail exponent.
- The lower the tail exponent, the heavier the tails.
- If $\alpha = 2$, then we obtain the normal distribution.
- Thicker tails indicate that the extreme events become more frequent.
- Due to the important effect of the parameter α on the properties of the stable distributions, they are often called α-stable or alpha stable.

Stable distributions

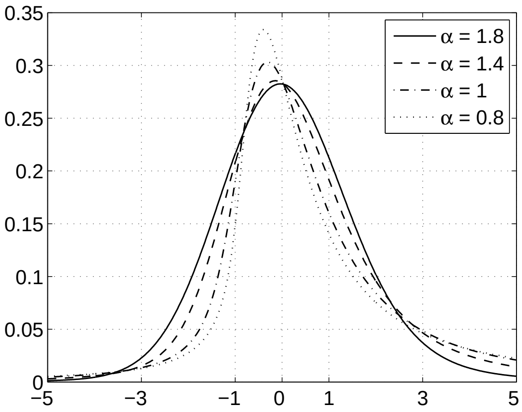

Figure: The density functions of stable laws with parameters $\alpha = 1.8,\ 1.4,\ 1,$ and 0.8, $\beta = 0.6$, $\sigma = 1$, $\mu = 0$. All densities are asymmetric but the skewness is more pronounced when the tail exponent is lower.

Stable distributions

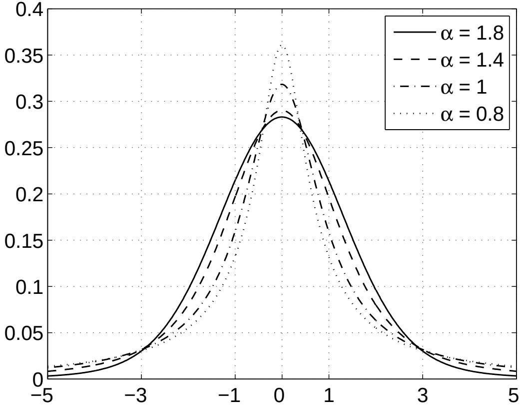

Figure: The density functions of stable laws with parameters $\alpha = 1.8,\ 1.4,\ 1,$ and 0.8, $\beta = 0$, $\sigma = 1$, $\mu = 0$. All densities are symmetric.

Stable distributions

- There is another important characteristic which is the stability property.
- According to the stability property, appropriately centered and normalized sums of i.i.d. α-stable random variables is again α-stable.
- This property is unique to the class of stable laws.

Modeling financial assets with stable distributions

- We noted that the outliers in the high-frequency data cannot be modeled by the normal distribution.
- But the stable Paretian distributions are heavy-tailed and they have the potential to describe the heavy-tails and the asymmetry of the empirical data.
- The stable Paretian distributions arise as limit distributions of sums of i.i.d. random variables with infinite variance and their variance is also unbounded.
- If $X \in S_\alpha(\sigma, \beta, \mu)$, then the moment $E|X|^p < \infty$ only if $p < \alpha \leq 2$. So if we assume a stable Paretian distribution as a model for the log-returns of a price time series, then we assume that the variance of the log-returns is infinite.
  From a practical viewpoint, this is not a desirable consequence.

- A large number of empirical studies have shown that the stable distributions provide a very good fit to the observed daily log-returns for common stocks in different countries and, thus, the overall idea of using the limit distributions in the Generalized CLT as a probabilistic model has empirical support.

$\Rightarrow$ As a result, the probabilistic properties of the daily log-returns for common stocks seem to be well approximated by those of the stable distributions.

- The infinite variance of the stable hypothesis appear as an undesirable consequence. Therefore, it is reasonable to search for distributions close to the stable distributions and, at the same time, have finite variance.

The methods to obtain such distributions concern truncating the tail of the stable law very far away from the center of the distribution:

- A straightforward approach is to cut the tails of the distribution and to make the random variable defined not on the entire real line but on the interval defined by the two truncation points.
- Another, more sophisticated approach involves replacing the stable tails very far away from the center of the distribution by the tails of another distribution so that the variance becomes finite.

  This is the method behind the smoothly truncated stable distributions which have been very successfully used in option pricing.

The tail truncation method is reasonable from a practical viewpoint:

- On every stock exchange, there are certain regulations according to which trading stops if the market index loses more than a given percentage.
- In fact, this is a practical implementation of tail truncation because huge losses (very small negative log-returns) usually happen when there is a crisis and in market crashes the market index plunges.
- Thus, astronomical losses (incredibly small negative log-returns) are not possible in practice.

From the point of view of the limit theorems, the tail truncation results in finite variance. Don't we actually assume that it is the normal distribution which drives the properties of the monthly log-returns if the daily log-returns are assumed to follow the truncated stable distribution?

- This is not the case because the truncated stable distributions converge very slowly to the normal distribution.
- The c.d.f. of the sum will begin to resemble the normal c.d.f. only when the number of summands becomes really huge.
- For small and medium number of summands, the density of the sum is actually closer to the density of the corresponding stable distribution. This fact has been established using the theory of probability metrics and is also known as a pre-limit theorem.

Fixing the number of summands, how close is the sum of i.i.d. variables to the limit distribution? What is the convergence rate?

- There are many results which state the convergence rate in terms of different simple probability metrics, such as the Kolmogorov metric, the total variation metric, the uniform metric between densities, the Kantorovich metric, etc.
- It turned out that probability metrics with special structure have to be introduced - ideal metrics. Their special structure is dictated by the particular problem under study — different additional axioms are added depending on the limit problem.
- They are called ideal because they solve the problem in the best possible way due to their special structure.

- From Lecture 3, a probability metric $\mu(X, Y)$ is a functional which measures the "closeness" between the random variables X and Y, satisfying the following three properties:

Property 1. $\mu(X, Y) \geq 0$ for any X, Y and $\mu(X, X) = 0$

Property 2. $\mu(X, Y) = \mu(Y, X)$ for any X, Y

Property 3. $\mu(X, Y) \leq \mu(X, Z) + \mu(Z, Y)$ for any X, Y, Z

- The three properties are called the identity axiom, the symmetry axiom and the triangle inequality, respectively.
- The ideal probability metrics are probability metrics which satisfy 2 additional properties: the homogeneity property and the regularity property.

Homogeneity property

- The homogeneity property is

Property 4. $\mu(cX, cY) = |c|^r \mu(X, Y)$ for any X, Y and constants $c \in \mathbb{R}$ and $r \in \mathbb{R}$.

Basically, the homogeneity property states that if we scale the two random variables by one and the same constant, the distance between the scaled quantities ($\mu(cX, cY)$) is proportional to the initial distance ($\mu(X, Y)$) by $|c|^r$.

In particular, if $r = 1$, then the distance between the scaled quantities changes linearly with c.

Homogeneity property

The homogeneity property has a financial interpretation.

- We briefly remark that if X and Y are random variables describing the random return of two portfolios, then converting proportionally into cash, for example, 30% of the two portfolios results in returns scaled down to $0.3X$ and $0.3Y$.
- Since the returns of the two portfolios appear scaled by the same factor, it is reasonable to assume that the distance between the two scales down proportionally.

Regularity property

- The regularity property is

  Property 5. $\mu(X + Z, Y + Z) \leq \mu(Y, X)$ for any X, Y and Z independent of X and Y.

  The regularity property states that if we add to the initial random variables X and Y one and the same random variable Z independent of X and Y, then the distance decreases.

Regularity property

The regularity property has a financial interpretation.

- Suppose that X and Y are random variables describing the random values of two portfolios and Z describes the random price of a common stock.
- Then buying one share of stock Z per portfolio results in two new portfolios with random wealth $X + Z$ and $Y + Z$. Because of the common factor in the two new portfolios, we can expect that the distance between $X + Z$ and $Y + Z$ is smaller than the one between X and Y.

$\Rightarrow$ Any functional satisfying Property 1, 2, 3, 4, and 5 is called an ideal probability metric of order r.

Examples

There are examples of both compound and simple ideal probability metrics.

- For instance, the p-average compound metric $\mathcal{L}_p(X, Y)$ defined in equation (20) and the Birnbaum-Orlicz metric $\Theta_p(X, Y)$ defined in equation (22) in Lecture 3 are ideal compound probability metrics of order one and $1/p$ respectively.
- Almost all known examples of ideal probability metrics of order $r > 1$ are simple metrics.

Examples

Almost all of the simple metrics defined in Lecture 3 are ideal:

1. The uniform metric between densities $\ell(X, Y)$ defined in equation (16) is an ideal metric of order -1.
2. The L_p-metrics between distribution functions $\theta_p(X, Y)$ defined in equation (13) is an ideal probability metric of order $1/p$, $p \geq 1$.
3. The Kolmogorov metric $\rho(X, Y)$ defined in equation (9) is an ideal metric of order 0. This can also be inferred from the relationship $\rho(X, Y) = \theta_\infty(X, Y)$.
4. The L_p-metrics between inverse distribution functions $\ell_p(X, Y)$ defined in equation (15) is an ideal metric of order 1.

5. The Kantorovich metric $\kappa(X, Y)$ defined in equation (12) is an ideal metric of order 1. This can also be inferred from the relationship $\kappa(X, Y) = \ell_1(X, Y)$.
6. The total variation metric $\sigma(X, Y)$ defined in equation (17) is an ideal probability metric of order 0.
7. The uniform metric between inverse c.d.f.s $\mathbf{W}(X, Y)$ defined in equation (14) is an ideal metric of order 1.

Examples

Let us illustrate the order of ideality, or the homogeneity order, by the ideal metrics $\ell(X, Y)$ and $\sigma(X, Y)$ which are both based on measuring distances between density functions.

- The left part of Figure on the next slide shows the densities $f_X(x)$ and $f_Y(x)$ of two random variables X and Y.
- At the bottom of the figure, we can see the absolute difference between the two densities $|f_X(x) - f_Y(x)|$ as a function of x.
- The upper right plot shows the densities of the scaled random variables $0.5X$ and $0.5Y$. Note that they are more peaked at the means of X and Y.
- The lower right plot shows the absolute difference $|f_{X/2}(x) - f_{Y/2}(x)|$ as a function of x.

Examples

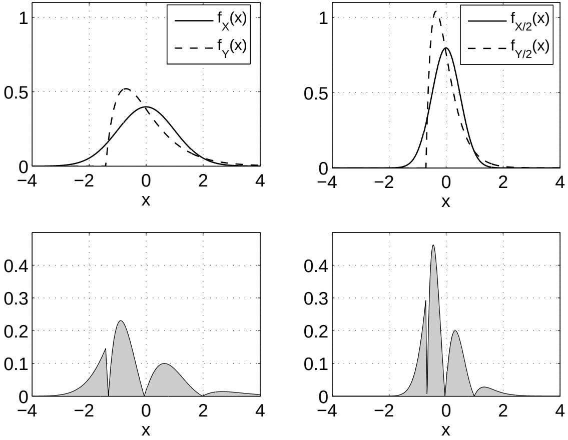

Figure: The left part shows the densities of X and Y and the absolute difference between them. The right part shows the same information but for the scaled random variables $0.5X$ and $0.5Y$.

Examples

- Recall that

$$\ell(X, Y) = \max_{x \in \mathbb{R}} |f_X(x) - f_Y(x)|.$$

which means that the uniform distance between the two densities is equal to the maximum absolute difference.

- On the figure above we can see that the maximum between the densities of the scaled random variables is clearly larger than the maximum of the non-scaled counterparts. Actually, it is exactly twice as large,

$$\ell(X/2, Y/2) = 2\ell(X, Y)$$

because of the metric $\ell(X, Y)$ is ideal of order -1.

Examples

- In Lecture 3 we explained that the total variation metric $\sigma(X, Y)$ can be expressed as one half the area closed between the graphs of the two densities.
- Since the total variation metric is ideal of order zero,

$$\sigma(X/2, Y/2) = \sigma(X, Y),$$

  then it follows that the surface closed between the two graphs is not changed by the scaling.
- Therefore, the shaded areas on the figure on slide 75 are exactly the same.

Examples

- Suppose that X and Y are random variables describing the return of two portfolios.
- While interpretation of the homogeneity property, if we start converting those portfolios into cash, then their returns appear scaled by a smaller and smaller factor.
- Our expectations are that the portfolios should appear more and more alike; that is, when decreasing the scaling factor, the ideal metric should indicate that the distance between the two portfolios decreases.
- We verified that the metrics $\ell(X, Y)$ and $\sigma(X, Y)$ indicate otherwise. Therefore, in the problem of benchmark tracking it makes more sense to consider ideal metrics of order greater than zero, $r > 0$.

Examples

Besides the ideal metrics we have listed above, there are others which allow for interesting interpretations.

1. The Zolotarev ideal metric
   Zolotarev's family of ideal metrics is very large. Here we state only one example.

$$\zeta_2(X, Y) = \int_{-\infty}^{\infty} \left| \int_{-\infty}^{x} F_X(t)dt - \int_{-\infty}^{x} F_Y(t)dt \right| dx \qquad (12)$$

   where X and Y are random variables with equal means, $EX = EY$, and they have finite variances. The metric $\zeta_2(X, Y)$ is ideal of order 2.

Examples

- The Zolotarev ideal metric $\zeta_2(X, Y)$ can be related to the theory of preference relations of risk-averse investors, who are characterized by the their concave utility functions.
- Suppose that X and Y are random variables describing the returns on two investments.
- Rothschild and Stiglitz (1970) showed that investment "X" is preferred to investment "Y" by all risk-averse investors if and only if $EX = EY$ and

$$\int_{-\infty}^{t} F_X(x)dx \leq \int_{-\infty}^{t} F_Y(x)dx, \quad \forall t \in \mathbb{R}. \tag{13}$$

This relation is known as *Rothschild-Stiglitz dominance*.

Examples

- The Zolotarev ideal metric $\zeta_2(X, Y)$ sums up the absolute deviations between the two quantities in inequality (13) for all values of t.
- Therefore, it measures the distance between the investments returns X and Y directly in terms of the quantities defining the preference relation of all risk-averse investors. As a result, we can use it to quantify the preference order.
- For example, if we know that investment "X" is preferred to investment "Y" by all risk-averse investors, we can answer the question of whether X is preferred to Y only to a small degree (if $\zeta_2(X, Y)$ is a small number), or whether X dominates Y significantly (if $\zeta_2(X, Y)$ is a large number).

2. The Rachev family of ideal metrics is also very large. Consider only one example.

$$\zeta_{s,p}(X,Y) = C_s \left(\int_{-\infty}^{\infty} |E(t-X)_+^s - E(t-Y)_+^s|^p dt \right)^{1/p}, \quad (14)$$

where

C_s	is a constant, $C_s = 1/(s-1)!$
p	is a power parameter, $p \geq 1$
s	takes integer values, $s = 1, 2, \ldots, n, \ldots$
$(t-x)_+^s$	is a notation meaning the larger quantity between $t-x$ and zero raised to the power s, $(t-x)_+^s = (\max(t-x, 0))^s$
$X,\ Y$	are random variables with finite moments $E\lvert X\rvert^s < \infty$ and $E\lvert Y\rvert^s < \infty$.

- The quantity $E(t-X)_+^s$ appearing in the definition of the metric is also known as the lower partial moment of order s. The simple metric $\zeta_{s,p}(X,Y)$ is ideal with order $r = s + 1/p - 1$.
- Suppose that X and Y are random variables describing the return distribution of two common-stocks. The quantity $E(t-X)_+$ calculates the average loss of X provided that the loss is larger than the performance level t.
- Likewise, $E(t-Y)_+$ calculates the average loss of Y larger than t. The absolute difference $|E(t-X)_+ - E(t-Y)_+|$ calculates the deviation between the average loss of X and the average loss of Y for one and the same performance level t.

The Rachev ideal metric

- In the case $p = 1$, the metric

$$\zeta_{1,1}(X, Y) = \int_{-\infty}^{\infty} |E(t - X)_+ - E(t - Y)_+| dt$$

  sums up the absolute deviations for all possible performance levels.
- In this respect, it is an aggregate measure of the deviations between the average losses above a threshold. If $s > 1$, then the metric $\zeta_{s,1}(X, Y)$ sums up the deviations between the lower partial moments for all possible performance levels.

- As the power p increases, it makes the smaller contributors to the total sum in $\zeta_{1,1}(X,Y)$ become even smaller in the Rachev ideal metric $\zeta_{1,p}(X,Y)$ defined in (14).
- Thus, as p grows, only the largest absolute differences $|E(t-X)_+ - E(t-Y)_+|$ start to matter. At the limit, as p approaches infinity, only the largest difference $|E(t-X)_+ - E(t-Y)_+|$ becomes significant and the metric $\zeta_{1,p}(X,Y)$ turns into

$$\zeta_{1,\infty}(X,Y) = \sup_{t\in\mathbb{R}} |E(t-X)_+ - E(t-Y)_+|. \qquad (15)$$

The Rachev ideal metric

- Note that the Rachev ideal metric given in equation (15) is entirely concentrated on the largest absolute difference between the average loss of X and Y for a common performance level t.
- Similarly, the Rachev ideal metric $\zeta_{s,\infty}(X, Y)$ is calculated to be represented by the expression

$$\zeta_{s,\infty}(X, Y) = C_s \sup_{t \in \mathbb{R}} |E(t - X)_+^s - E(t - Y)_+^s|.$$

- It is entirely concentrated on the largest absolute difference between the lower partial moments of order s of the two random variables.

$\Rightarrow$ The Zolotarev ideal metric defined in equation (12) appears as a special case of the Rachev ideal metric.

The Rachev ideal metric

- In financial theory, the lower partial moments are used to characterize preferences of difference classes of investors.
- For example, the lower partial moment of order 2 characterizes the investors preferences who are non-satiable, risk-averse, and prefer positively skewed distributions.
- Suppose that X and Y describe the return distribution of two portfolios. X is preferred to Y by this class of investors if $EX = EY$ and

$$E(t-X)_+^2 \le E(t-Y)_+^2, \quad \forall t \in \mathbb{R}.$$

- The Rachev ideal metric $\zeta_{2,p}(X,Y)$ quantifies such a preference order in a natural way — if X is preferred to Y, then we can calculate the distance by $\zeta_{2,p}(X,Y)$ and check whether X significantly dominates Y.

Svetlozar T. Rachev, Stoyan Stoyanov, and Frank J. Fabozzi
Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization: The Ideal Risk, Uncertainty, and Performance Measures
John Wiley, Finance, 2007.

Chapter 4.