Introduction: Background Review

Prof. Dr. Svetlozar (Zari) Rachev

Frey Family Foundation Chair-Professor,
Applied Mathematics and Statistics, Stony Brook University
Chief Scientist, FinAnalytica Inc.
Outline: Background Review (10am-12pm)

Introduction

Discrete Probability Distributions

Continuous Probability Distributions

Statistical Moments and Quantiles

Joint Probability Distributions
Introduction

Discrete Probability Distributions

Continuous Probability Distributions

Statistical Moments and Quantiles

Joint Probability Distributions
Most of the concepts in theoretical and empirical finance that have been developed over the last 50 years rest on the assumption that the financial assets’ returns follow a normal distribution.

Yet, studies have overwhelmingly failed to support the validity of this assumption. Instead, returns are heavy-tailed and, generally, skewed.

The adoption of a non-normal framework requires modification/generalization of financial theories for risk-measurement and management, as well as portfolio selection.

In our sessions today and tomorrow, we will discuss both the mainstream knowledge and the cutting-edge developments in financial risk and portfolio modeling.
Basic Probability Concepts

- A *random variable* is a variable that exhibits randomness.\(^1\)
- *Outcomes* of a random variable are mutually-exclusive, potential results that could occur.
- A *sample space* is the collection of all possible outcomes.
- An *event* is a subset of the sample space and a collection of several outcomes.

**Example:** Microsoft’s stock return over the next year

- Sample space: outcomes ranging from 100% loss to an extremely high positive return.
- The sample space could be partitioned into two subsets: a subset where return is less than 10% and a subset where return exceeds 10%. Then, return greater than 10% is an event.

\(^1\) A more general concept is that of a probability space. Random variables are measurable functions from a probability space into the set of real numbers. See Billingsley (1995) and Shiryaev (1996).
Introduction

Discrete Probability Distributions

Continuous Probability Distributions

Statistical Moments and Quantiles

Joint Probability Distributions
A discrete random variable limits the outcomes to where the variable can take on only discrete values.

Example: Option contract on S&P500 index

Consider an option whose payoff depends on the realized return of the S&P500, as follows:

<table>
<thead>
<tr>
<th>If S&amp;P return is</th>
<th>Payment received</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than or equal to zero</td>
<td>$0</td>
</tr>
<tr>
<td>Greater than zero but less than 5%</td>
<td>$10,000</td>
</tr>
<tr>
<td>Greater than 5% but less than 10%</td>
<td>$20,000</td>
</tr>
<tr>
<td>Greater than or equal to 10%</td>
<td>$100,000</td>
</tr>
</tbody>
</table>

Next, assign a probability to each outcome and calculate the probability of an arbitrary event. For e.g., if outcomes are equally-likely, what is the probability of losing money on the long option contract?

On the following slides, we provide a short introduction to several of the most important discrete probability distributions.
A random variable has the *Bernoulli distribution* with parameter $p$ if it has only two possible outcomes, usually encoded as “1” (which might represent “success”) and “0” (which might represent “failure”).

**Example: Default event of a company**

- Observe a company $C$ during a specified time period $T$ to $T + 1$.
- Define:
  \[
  X = \begin{cases} 
  1 & \text{if } C \text{ defaults in } [T, T + 1], \\
  0 & \text{else}.
  \end{cases}
  \]
- The parameter $p$ is in this case the probability of default of company $C$. 
In practical applications, we consider a whole portfolio of companies $C_1, C_2, \ldots, C_n$. Assume all these $n$ companies have the same annualized probability of default, $p$.

The Bernoulli distribution generalizes to the *binomial distribution*.\(^1\) A binomial random variable $Y$, with parameters $p$ and $n$, is obtained as the sum of $n$ independent and identically Bernoulli-distributed random variables.

**Example: Number of defaults in a credit portfolio**

Given the two parameters, the probability of observing $k$ defaults ($0 \leq k \leq n$) during the period $[T, T + 1]$ can be computed as

$$P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k},$$

where $\binom{n}{k} = \frac{n!}{(n-k)!k!}$.

\(^1\)We’ll see it in various applications—from VaR backtesting to option pricing.
The *Poisson distribution* is used to describe the random number of events occurring over a certain time interval.

The difference with the binomial distribution is that the number of events that could occur is unbounded (at least theoretically).

The Poisson parameter $\lambda$ indicates the *rate of occurrence* of the random events, i.e., it tells how many events on average occur per unit of time.

The probability distribution of a Poisson-distributed random variable $N$ is described by the following equation:

$$P(N = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots$$

In financial applications, the Poisson distribution arises as the distribution of a stochastic process, called the *Poisson process*, which is used especially in credit-risk modeling. We discuss stochastic processes in the 3pm-6pm session today.
Outline: Background Review (10am-12pm)

Introduction

Discrete Probability Distributions

Continuous Probability Distributions

Statistical Moments and Quantiles

Joint Probability Distributions
A *continuous random variable* can take on any possible value within the range of outcomes.

For a continuous random variable, the calculation of probabilities is substantially different from the discrete case. The probability of a realization of the random variable within some range cannot be found by simply adding the probabilities of the individual outcomes, since the number of values in the interval is very large.

The *probability density function* (PDF) is a mathematical function representing the continuous probability distribution. The PDF of the random variable $X$ is denoted by $f_X(x)$.

The density function is always non-negative. Large values of it at some point $x$ implies a relatively high probability of a realization in the neighborhood of $x$. 
Continuous Probability Distributions

Probability Density Function

Probabilities are represented by areas under the PDF.

- The shaded area is the probability of realizing a return greater than \( a \) and less than \( b \).
- The probability of each individual outcome of a continuous random variable is zero.

Calculating the probability of the shaded region involves computing the area under the curve, i.e., integration according to the formula:

\[
P(a \leq X \leq b) = \int_{a}^{b} f_X(x) \, dx.
\]
Continuous Probability Distributions

Probability Distribution Function

- The mathematical function that provides the cumulative probability, i.e., probability of getting an outcome less than $x$, is called the *cumulative distribution function* (CDF) and is denoted by $F_X(x)$.

- A CDF is always non-negative, non-decreasing, and takes values between 0 and 1, since it represents probabilities.

- The relation between CDF and PDF is given by the expression:

$$ P(X \leq t) = F_X(t) = \int_{-\infty}^{t} f_X(x)dx. $$

- The probability that the random variable is in the range between $a$ and $b$ is expressed through the CDF as

$$ P(a \leq X \leq b) = F_X(b) - F_X(a). $$
The normal (Gaussian) distribution is one of the most important probability distributions in statistics and one of the most widely-used ones in finance.

The random variable $X$ is said to be normally-distributed with parameters $\mu$ and $\sigma$, denoted by $X \sim \mathcal{N}(\mu, \sigma^2)$, if its PDF is given by the formula

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$ 

The parameter $\mu$ is called the location parameter, while $\sigma$ is the scale (or shape) parameter. If $\mu = 0$ and $\sigma = 1$, $X$ is said to have the standard normal distribution.
The plots below illustrate the effect $\mu$ and $\sigma$ have on the normal density function.
Location-Scale Invariance

★ Suppose $X \sim N(\mu, \sigma^2)$ and consider $Y = aX + b$.

★ In general, $Y$’s distribution differs from $X$’s. When $X$ is normal, $Y \sim N(a\mu + b, a^2\sigma^2)$.

Summation Stability

★ Let $X_1, \ldots, X_n$ be independent and normally-distributed with $N(\mu_i, \sigma_i^2)$.

★ The sum $\sum_{i=1}^n X_i$ is normally-distributed with

\[
\mu = \mu_1 + \cdots + \mu_n \\
\sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2.
\]
Continuous Probability Distributions
Properties of the Normal Distribution: Domain of Attraction

- A mathematical result, called the *Central Limit Theorem*, states that under certain conditions, the distribution a large sum of random variables behaves necessarily like a normal distribution.\(^1\)

- The normal distribution is not the unique class with this property, as many believe.

- It is the class of stable distributions (containing the normal distribution) which is unique with this property:

  A *large sum of linearly normalized, independent, identically-distributed random variables can only converge to a stable distribution. That is, the stable distribution has domain of attraction.*

- We will discuss stable distributions in the 1pm-3pm session today.

\(^1\)See, for example, Billingsley (1995), Samorodnitsky and Taqqu (1994), and Rachev, Hoechstoetter, Fabozzi, and Focardi (2010).
Continuous Probability Distributions

Exponential Distribution

- Used to model waiting times, i.e., time we need to wait until the next event takes place.
- An exponentially-distributed random variable $T$ takes on positive real values and its density function has the form

\[ f_T(t) = \frac{1}{\beta} e^{-\frac{t}{\beta}}, \quad t > 0. \]

- In credit risk modeling, the parameter $\lambda = 1/\beta$ has an interpretation as the hazard rate or default intensity.
Consider a sequence of independent and identically distributed random variables $T_1, T_2, \ldots$ with the exponential distribution.

We can think of $T_1$ as the time until a firm in a high-yield bond portfolio defaults, $T_2$ the time between the first and the second default, and so on. These waiting times are sometimes called *interarrival times*.

Let $N_t$ denote the number of defaults which have occurred till time $t \geq 0$.

The random variable $N_t$ is distributed with the Poisson distribution with parameter $\lambda = 1/\beta$. 
Continuous Probability Distributions

Student’s t Distribution

- The *Student's t distribution* is used in finance as a probabilistic model of asset returns.

- Its density is symmetric and mound-shaped, like the normal distribution’s. However, it has fatter tails. This makes it more suitable for return modeling than the Gaussian distribution.

- The density function of the $t$ distribution is given by the expression

$$f_X(x) = \frac{1}{\sigma \sqrt{\nu \pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{1}{\nu} \left(\frac{x - \mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}} \quad , \quad x \in \mathbb{R},$$

where $\nu$ is an integer-valued parameter called *degree of freedom*, $\mu$ is a real-valued location, and $\sigma$ is a positive scale parameter. Denote the distribution by $t_\nu(\mu, \sigma)$.

- For values of $\nu$ larger than 30, the $t$-distribution is considered as equal to the normal distribution. Small values of $\nu$ imply heavy tails.
Continuous Probability Distributions
Student’s $t$ Distribution

The plot visualizes the comparison between the Student’s $t$ and the normal distributions.

The version of the Student’s $t$ distribution considered here is also known as the “classical” $t$-distribution to distinguish it from the asymmetric versions of it, which we cover in our discussion on univariate fat-tailed models in the next session.
Outline: Background Review (10am-12pm)

Introduction

Discrete Probability Distributions

Continuous Probability Distributions

Statistical Moments and Quantiles

Joint Probability Distributions
It is common to summarize the characteristics of distributions with the help of various measures. The five most commonly used ones are:

- Location
- Dispersion
- Asymmetry
- Concentration in tails
- Quantiles

We describe these measures next.
Location

- Represents the central value of the distribution.
- The mean (average) value, median, and mode are used as measures. The relationship among these three measures depends on the skewness of the distribution, which we describe below.
- In a symmetric distribution, they all coincide.
- The most commonly-used one is the mean and denoted by $\mu$ or $E(X)$.

Dispersion

- Indicates how spread out the values of the random variable could be.
- Measures of dispersion are the range, variance, and mean absolute deviation.
- The most commonly-used one is the variance, which measures the dispersion of values relative to the mean. The square root of it is the standard deviation, usually denoted by $\sigma$. 
A probability distribution may be symmetric or asymmetric around its mean. A measure of asymmetry is called *skewness*.

Negative skewness indicates that the distribution is skewed to the left. That is, compared with its right tail, the left tail is elongated.

Positive skewness indicates that the distribution is skewed to the right. That is, its right tail is longer than the left one.
The concentration (mass) of potential outcomes in the tails provides additional information about a distribution.

The tails contain the extreme values and—in financial applications—indicate the potential for a financial fiasco or ruin.

The fatness of the tails is related to the peakedness of the distribution around its mean or center.

The joint measure of peakedness and tail-fatness is called kurtosis.
In statistical language, the four measures above are called *statistical moments*, or simply *moments*.

Generally, the *moment of order n* of a random variable is denoted by $\mu_n$. It is defined and computed (for a continuous variable) as

$$\mu_n = E(X^n), \quad n = 1, 2, \ldots$$

$$\mu_n = \int_{-\infty}^{\infty} x^n f_X(x) \, dx.$$ 

The *centered moment of order n* is denoted by $m_n$ and computed (for a continuous variable) as

$$m_n = E(X - E(X))^n, \quad n = 1, 2, \ldots$$

$$m_n = \int_{-\infty}^{\infty} (x - E(X))^2 f_X(x) \, dx.$$ 

---

1 See, for example, Larson and Marx (1996) and Rachev, et.al. (2010).
How are the measures we discussed related to the statistical moments?

- The mean is the first moment and is also called the expected value:
  \[ \mu = E(X). \]

- The variance is the second central moment:
  \[ \sigma^2 = E(X - E(X))^2. \]

- The skewness is a rescaled third central moment:
  \[ \varsigma = \frac{E(X - E(X))^3}{(\sigma^2)^{3/2}}. \]

- The kurtosis is a rescaled fourth central moment:
  \[ \kappa = \frac{E(X - E(X))^4}{(\sigma^2)^4}. \]
A concept, called $\alpha$-quantile, is used in addition to the moments to summarize a probability distribution.

The $\alpha$-quantile gives information about where the first $\alpha\%$ of the distribution are located. Denote it by $q_\alpha$.

An arbitrary observation falls below $q_\alpha$ with probability $\alpha\%$ and above $q_\alpha$ with probability $(100-\alpha)\%$.

Some quantiles have special names. The 25%--, 50%--, and 75%-quantile are referred to as the first, second, and third quartile, respectively.

The 1%--, 2%--, . . . , 98%--, and 99%-quantiles are called percentiles.

The concept of $\alpha$-quantile is closely related to the value-at-risk ($\text{VaR}_\alpha(X)$) measure commonly used in risk measurement. We will discuss risk measures in tomorrow’s morning session.
In practical situations, we observe realizations from some probability distribution (for e.g., daily return of the S&P 500 index over the last one year) but we don’t know the distribution that generated these returns.

With a sample of observations, \( r_1, r_2, \ldots, r_n \), we can try to estimate the “true moments”. The estimates are sometimes called “sample moments”.

The empirical analogue of an expectation of a random variable is the average of the observations:

\[
E(X) \approx \frac{1}{n} \sum_{i=1}^{n} r_i.
\]

The **Law of Large Numbers** says that for large \( n \), the average of the observations will not be far away from the mean of the distribution.

The sample mean and variance, for e.g., can be computed, respectively, as

\[
\bar{r} = \frac{1}{n} \sum_{i=1}^{n} r_i \quad \text{and} \quad s^2 = \frac{1}{n} \sum_{i=1}^{n} (r_i - \bar{r})^2.
\]
Introduction

Discrete Probability Distributions

Continuous Probability Distributions

Statistical Moments and Quantiles

Joint Probability Distributions
Above, we explained the properties of a distribution of a single random variable, that is, a univariate distribution. Knowledge of univariate distributions is necessary to analyze the time series characteristics of individual assets.

Understanding multivariate distributions (joint distributions of multiple random variables) is important because financial theories such as portfolio selection theory and asset-pricing theory involve distributional properties of sets of investment opportunities (i.e., multiple variables).

We describe the basics of multivariate distributions next, starting with conditional probability.
Consider the returns on stocks of companies $X$ and $Y$ belonging to the same industry. They are not unrelated, since they are subject to some extent to common industry factors.

A reasonable question may be, for e.g.: what is the probability that return on $X$ is smaller than -2%, on condition that $Y$ realizes a huge loss of 10%?

The *conditional probability* calculates the probability of an event, given that another event happens. Denote the first event by $A$ and the second by $B$. The conditional probability of $A$ given $B$ is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)},$$

which is also known as *Bayes formula*.

In the numerator is the probability that $A$ and $B$ occur simultaneously, denoted by $A \cap B$, that is, the joint probability of the two events.
Joint Probability Distributions: Defined

- Suppose that a portfolio consists of a position in two assets, asset 1 and asset 2. Then, there will be a probability distribution for (1) asset 1, (2) asset 2, and (3) asset 1 and asset 2.

- The first two distributions are referred to as marginal distributions. The distribution for asset 1 and asset 2 is the joint distribution.¹

- The relation between the CDF $F$ and the PDF $f$ of a multivariate random variable (random vector) $X = (X_1, \ldots, X_n)$ is given by

$$P (X_1 \leq t_1, \ldots, X_n \leq t_n) = F_X (t_1, \ldots, t_n) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_n} f_X (x_1, \ldots, x_n) \, dx_1 \cdots dx_n.$$  

- Interpretation: the joint probability that the first variable has a value less than or equal to $t_1$, the second variable—less than or equal to $t_2$, and so on is given by the CDF $F$. The probability value can be obtained by calculating the volume under the PDF $f$.

¹For more on multivariate distributions, see Kotz, Balakrishnan, and Johnson (2000).
When considering multivariate distributions, we are faced with inference among the distributions: large values of one variable imply large values of another variable or small values of a third one.

This property is known as *dependence of random variables*. The inverse of dependence is *stochastic independence*.

Two variables are *independently distributed* if and only if their joint distribution, given in terms of the CDF or PDF, equals the product of their marginal distributions:

\[
F_X(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)
\]
\[
f_X(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).
\]

In the earlier example, if we assume that the events \( X \leq -2\% \) and \( Y \leq -10\% \) are independent, the conditional probability is:

\[
P(X \leq -2\% \mid Y \leq -10\%) = \frac{P(X \leq -2\%)P(Y \leq -10\%)}{P(Y \leq -10\%)} = P(Y \leq -10\%).
\]
Two strongly related measures commonly used to measure dependence between two random variables are the covariance and correlation.

Notation:

\[ \sigma_X = \text{standard deviation of } X \quad \sigma_{XY} = \text{covariance between } X \text{ and } Y \]
\[ \sigma_Y = \text{standard deviation of } Y \quad \rho_{XY} = \text{correlation between } X \text{ and } Y. \]

The covariance is defined as

\[
\sigma_{XY} = E(X - E(X))(Y - E(Y)) \\
= E(XY) - E(X)E(Y).
\]

The correlation is a rescaled covariance:

\[
\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.
\]
The correlation takes on values between -1 and +1.

When the correlation is zero, the two variables are \textit{uncorrelated}. Notice that uncorrelatedness is not the same as independence, in general. Two variables can be uncorrelated but still dependent, since correlation is only a measure of the linear dependence.

The variance of the sum of two variables is:

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}.$$ 

In tomorrow’s discussion of mean-variance portfolio optimization, we will see that the variance of portfolio return is expressed by means of the variances of the assets’ returns and the covariances between them.

We will have more to say about dependence in the 1pm-3pm session today, when we discuss copulas—a powerful concept for dependence modeling.
In finance, it is common to assume that the random variables are normally-distributed. The joint distribution is referred to as a *multivariate normal distribution*.

Consider first \( n \) independent standard normal variables \( X_1, \ldots, X_n \). Their joint density is written as the product of the marginal ones:

\[
f_X(x_1, \ldots, x_n) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{x'x}{2}},
\]

where \( x'x = \sum_{i=1}^{n} x_i^2 \).

Now, consider the random variable

\[
Y = AX + \mu,
\]

where \( X = (X_1, \ldots, X_n) \), \( A \) is a \( n \times n \) matrix (\( n \) vectors with \( n \) components) and \( \mu \) is a vector of constants.
Y has a general multivariate normal distribution. Its density function can be expressed as

\[ f_Y(y_1, \ldots, y_n) = \frac{1}{(\pi |\Sigma|)^{n/2}} e^{-\frac{(y-\mu)'\Sigma^{-1}(y-\mu)}{2}}, \]

where \(|\Sigma|\) denotes the determinant of the matrix \(\Sigma\) and \(\Sigma^{-1}\) denotes the inverse of \(\Sigma\).

The matrix \(\Sigma\) can be calculated from \(A\): \(\Sigma = AA'\). That is, \(A\) is the “square root” (Cholesky factor) of \(\Sigma\).

\(\Sigma\) is the covariance matrix of the vector \(Y\), while \(\mu\) is the vector of component means.

The elements of \(\Sigma = \{\sigma_{ij}\}_{i,j=1}^n\) are the covariances between the components of \(Y\),

\[ \sigma_{ij} = \text{cov}(Y_i, Y_j). \]
Multivariate Normal Distribution

- Two-dimensional normal distribution
- Covariance matrix:
  \[ \Sigma = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix} \]
- Mean vector: \( \mu = (0, 0) \).
- \( \rho = 0.8 \): strong positive correlation.
- Realizations of \( Y \) cluster along the diagonal.
- Upper plot: two-dimensional PDF
- Lower plot: contour lines of PDF have elliptical shape
A generalization of the multivariate normal distribution is given by the class of elliptical distributions.

Elliptical distributions offer desirable properties in the context of portfolio selection. For this class, the correlation is the right dependence measure. When distributions do not belong to that class, alternative dependence concepts are necessary.

An \( n \)-dimensional random vector \( X \) is called spherically-distributed if all contours (set of points where the PDF has a constant value) have the form of a sphere. In the 2-dimensional case, the contours are circles.

Similarly, \( X \) is elliptically-distributed if the contours have the form of ellipses.

Elliptical distributions are symmetric. Some examples are multivariate normal, multivariate \( t \) distributions, and a part of the multivariate stable distributions.\(^1\)

\(^1\)See Fang, Kotz, and Ng (1994).
Elliptical distributions (whose density exists) can be described by a triple \((\mu, \Sigma, g)\), where \(\mu\) and \(\Sigma\) play roles similar to the mean vector and the covariance matrix in the multivariate normal setting.

The function \(g\) is the so-called *density generator*.

All three elements together define the density function of the distribution as:

\[
f_X(x) = \frac{c}{\sqrt{|\Sigma|}} g \left( (x - \mu)'\Sigma^{-1}(x - \mu) \right).
\]
References