Probability Basics

Liming Feng

Dept. of Industrial & Enterprise Systems Engineering
University of Illinois at Urbana-Champaign

Probabilistic modeling

- Probabilistic models in finance
  - Characterize randomness observed in financial markets
  - Brownian motion is used to model asset returns in the **Black-Scholes-Merton model**
  - It characterizes future asset price movements
- Probability vs statistics
  - Probability is **deductive**: given a distribution, compute the probability of an event
  - Statistics is **inductive**: given data, find the distribution that fits data
  - Probability theory provides the foundation for statistical analysis

Probability brainteasers in interviews

- Paul Wilmott, 2009, *Frequently Asked Questions in Quantitative Finance*
- Mark Joshi, Nick Denson, Andrew Downes, 2013, *Quant Job Interview Questions and Answers*
- Timothy Crack, 2013, *Heard on the Street: Quantitative Questions from Wall Street Job Interviews*
- Xinfeng Zhou, 2008, *A Practical Guide to Quantitative Finance Interviews*
- Dan Stefanica, Rados Radoicic and Tai-Ho Wang, 2013, *150 Most Frequently Asked Questions on Quant Interviews*
- John Hull, 2014, *Options, Futures, and other Derivatives*
- Search *Interview* on Amazon; Numerous internet resources for interview preparation: Start today! Attend fall/spring career fairs prepared

1. Probability space

Reference: Jacod and Protter 2004, Chapters 1, 2
Ross 2010, §1.1 - §1.3

Probability space

- Basic ingredients in a probabilistic model
    - **Sample space** Ω: set of all possible outcomes of a random experiment
    - Event: a subset of the sample space
    - $\sigma-$**algebra** $\mathcal{F}$: a collection of events
    - **Probability measure** $\mathbb{P}$: assigns probabilities to events
- $(\Omega, \mathcal{F})$ is called a measurable space
- **Probability space**: the triplet $(\Omega, \mathcal{F}, \mathbb{P})$

Example 0.1

Flip a fair coin for three times. Each time one gets a head (**H**) or tail (**T**).

- **Sample space**

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

- **Event**: a subset of Ω

$$A = \{HHT, HTH, THH\} = \{\text{there are two heads}\}$$

$$B = \{HHH, HHT\} = \{\text{the first two flips are heads}\}$$

$\sigma-$algebra

- $\sigma-$**algebra** $\mathcal{F}$: a collection of events such that
  - $\emptyset \in \mathcal{F}$ (empty-set)
  - $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ (complement)
  - $A_i \in \mathcal{F} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$ (countable union)
- It follows that
  - $\Omega \in \mathcal{F}$
  - $A_i \in \mathcal{F} \Rightarrow \cap_{i=1}^{\infty} A_i \in \mathcal{F}$
- $\mathcal{F}$ is thus closed under **complement, finite or countable union, intersection** operations
- Examples

$$\mathcal{F} = \{\emptyset, \Omega\}$$

$$\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$$

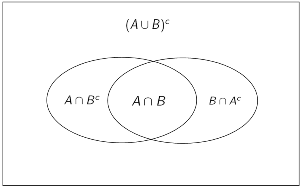
$$\mathcal{F} = 2^{\Omega} = \text{the set of all subsets of } \Omega$$

$\sigma-$algebra generated by $\mathcal{C}$

- For $\mathcal{C} \subset 2^{\Omega}$, the **$\sigma-$algebra generated by $\mathcal{C}$**, denoted by $\sigma(\mathcal{C})$, is the smallest $\sigma-$algebra that contains $\mathcal{C}$. E.g.,

$$\mathcal{C} = \{A, B\}; \quad \sigma(\mathcal{C}) = \{\emptyset, \Omega, A, B, A^c, B^c, A \cap B, A \cup B, A^c \cap B, \cdots\}$$

- $\sigma(\mathcal{C})$ contains all possible combinations of $A \cap B^c$, $B \cap A^c$, $A \cap B$, $(A \cup B)^c$ (assuming that they are non-empty; see the **venn-diagram** below)



- How many events there are in $\sigma(\mathcal{C})$ in the above? $2^4 = 16$

Borel $\sigma-$algebra

- Let $\Omega = \mathbb{R}$, $\mathcal{C} = \{(-\infty, a], a \in \mathbb{R}\}$
- **Borel $\sigma-$algebra** on $\mathbb{R}$ is the $\sigma-$algebra generated by $\mathcal{C}$

$$\mathcal{B}(\mathbb{R}) := \sigma(\mathcal{C})$$

- $\mathcal{B}(\mathbb{R})$ contains all open intervals, open sets, closed sets, single points, etc.

$$(a, b] = (-\infty, b] \cap (-\infty, a]^c, \quad (a, b) = \cup_{n=1}^{\infty}(a, b - \frac{1}{n}]$$

$$\{a\} = \bigcap_{n=1}^{\infty}(a - \frac{1}{n}, a + \frac{1}{n}), \quad [a, b] = (a, b] \cup \{a\}$$

- Any $B \in \mathcal{B}(\mathbb{R})$ is called a **Borel set**

Probability measure

- **Probability measure** $\mathbb{P}$: a map from $\mathcal{F}$ to $[0,1]$
  - $\mathbb{P}(\Omega) = 1$
  - $\sigma-$**additivity** (or countable additivity): for **mutually exclusive** events $A_i \in \mathcal{F}$, $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$
- It follows that for any $A, B, A_n \in \mathcal{F}$,
  - $\mathbb{P}(\emptyset)=0$
  - $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
  - If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
  - $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
  - If $A_n \to A$, then $\mathbb{P}(A_n) \to \mathbb{P}(A)$

2. Classical probability, counting techniques

Reference: Casella and Berger 2004, §1.2.3, §1.2.4

- In **Example** 0.1, let $\mathcal{F} = 2^{\Omega}$; we believe that all outcomes are equally likely and assign probability to each element in Ω as below
$$\mathbb{P}(\{HHH\}) = \cdots = \mathbb{P}(\{TTT\}) = \frac{1}{8}$$
- For any event A
$$\mathbb{P}(A) = \frac{\#(A)}{\#(\Omega)}$$
where $\#(A)$ is the number of elements in A. E.g.,
$$\mathbb{P}(\{HHH, HHT\}) = \frac{2}{8} = \mathbb{P}(\{HHH\}) + \mathbb{P}(\{HHT\})$$

Interpretation of probability

- **Classical probability**: Suppose a random experiment has n different outcomes that are equally likely to occur. Then the probability of an event A containing m outcomes is $\mathbb{P}(A) = m/n$
- **Frequentist probability**: if the above experiment is repeated many times, about $100m/n\%$ of the time A occurs (Jon Kerrich obtained 5067 heads out of 10,000 coin flips); easily extends to problems with no symmetry
- When calculating a classical probability, counting techniques are often used to determine $n = \#(\Omega)$ and $m = \#(A)$

Counting techniques

- **Multiplication rule**: A job consists of k different tasks. The ith task can be done in n_i ways. The total number of ways to finish the job is

$$n_1 n_2 \cdots n_k$$

- Number of **permutations**: number of ways to take r objects out of n distinct ones and form a sequence (order matters!)

$$n \cdot (n-1) \cdot \cdots \cdot (n-r+1) = \frac{n!}{(n-r)!}$$

It follows from the multiplication rule directly

Number of combinations

- Number of **combinations**: number of ways to take r objects out of n distinct ones (order is irrelevant!)

$$\binom{n}{r} = \frac{n!}{(n-r)!r!}$$

  (The number of combinations $\times r! =$ the number of permutations)

- $\binom{n}{r}$ is the coefficient of the term $x^r y^{n-r}$ in $(x+y)^n$, known as binomial coefficient

With identical objects

- Suppose $n = n_1 + \cdots + n_k$ objects belong to k different types. n_i of them are of type i, $1 \leq i \leq k$. Number of ways to form a sequence of these n objects:

$$\frac{n!}{n_1! \cdots n_k!}$$

(Number of ways to place the n_1 objects of type 1: $\left(\begin{array}{c} n \\ n_1 \end{array} \right)$; number of ways to place the n_2 objects of type 2: $\left(\begin{array}{c} n - n_1 \\ n_2 \end{array} \right)$, etc. Then apply the multiplication rule)

- How many ways are there to divide n distinct objects into k ordered piles so that the ith pile has n_i objects, $1 \leq i \leq k$?

- What is the coefficient of $x_1^{n_1} \cdots x_k^{n_k}$ in $(x_1 + \cdots + x_k)^n$?

- In a room of n people, what is the probability that no one shares a birthday (assuming 365 days a year)?

3. Probability measure on general spaces

Reference: Jacod and Protter 2004, Chapters 4, 6, 7

Finite or countable sample spaces

- Classical probability based on counting techniques assumes finite sample space with equally likely outcomes:

$$\mathbb{P}(\{\omega\}) = \frac{1}{\#(\Omega)}, \forall \omega \in \Omega; \quad \mathbb{P}(A) = \frac{\#(A)}{\#(\Omega)}$$

- More generally, for a finite or countable sample space Ω and $\mathcal{F} = 2^{\Omega}$, it suffices to specify

$$\mathbb{P}(\{\omega\}), \ \forall \omega \in \Omega$$

  as long as $\mathbb{P}(\{\omega\}) \in [0,1]$ and $\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1$
- For any $A \in \mathcal{F}$,

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$$

Uncountable sample spaces

Example 0.2

Randomly choose a number in the interval $[0, 1]$. Any number may be chosen equally likely. Sample space $\Omega = [0, 1]$

- **Cannot** construct probability measure by defining probabilities for individual elements in Ω
- Define probabilities for "representative" events directly; the probability measure can then be extended to the sigma algebra generated by these events

- To construct a desired probability measure: for any $a \in [0,1]$, let

$$\mathbb{P}([0,a]) = a$$

- $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\{0\}) = 0$. For any $0 \le a \le b \le 1$

$$\mathbb{P}((a,b]) = \mathbb{P}([0,b]) - \mathbb{P}([0,a]) = b - a$$

- The above probability is simply given by the usual length. Define the probability (i.e., the length) of any finite disjoint union of $\{0\}$ and intervals of the form $(a,b]$ in the usual way

$$\mathbb{P}((\frac{1}{4},\frac{1}{2}] \cup (\frac{3}{4},1]) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$\mathbb{P}([0,\frac{1}{4}]) = P(\{0\} \cup (0,\frac{1}{4}]) = 0 + \frac{1}{4} = \frac{1}{4}$$

- $\mathbb{P}$ constructed this way is consistent with "any number may be chosen equally likely". E.g.,

$$\mathbb{P}((0, \frac{1}{2}]) = \mathbb{P}((\frac{1}{2}, 1]) = \frac{1}{2}$$

$$\mathbb{P}((0, \frac{1}{3}]) = \mathbb{P}((\frac{1}{3}, \frac{2}{3}]) = \mathbb{P}((\frac{2}{3}, 1]) = \frac{1}{3}$$

- Let $\mathcal{C}$ be the collection of finite disjoint unions of $\{0\}$ and intervals of the form $(a, b]$
- $\mathbb{P}$ can then be extended to $\mathcal{F} = \sigma(\mathcal{C}) = \mathcal{B}([0, 1])$, the **Borel $\sigma-$algebra** on $[0, 1]$
- $\mathcal{F} = 2^{\Omega}$ too large for the above probability measure to be well defined

Constructing probability measures in general

- In general, it is often easier to specify probabilities for events in some **algebra** $\mathcal{C}$
- An algebra is a collection of events that contains $\emptyset$ and is closed under **complement, finite union** operations
- A probability measure $\mathbb{P}$ defined on an algebra $\mathcal{C}$ can be **uniquely extended** to $\mathcal{F} = \sigma(\mathcal{C})$
- The collection of all finite disjoint unions of $\{0\}$ and intervals of the form $(a, b]$ in the previous example is in fact an algebra

Zero probability

- Probability of any single point is zero

$$\mathbb{P}(\{a\}) = \lim_{n\to+\infty} \mathbb{P}((a-\frac{1}{n}, a]) = 0$$

  $\mathbb{P}$ couldn't have been constructed by specifying probabilities for all possible outcomes
- $\mathbb{P}(A) = 0$ does not mean A will never occur; but it will not occur **almost surely** (a.s.)
- $\mathbb{P}(A) = 1$ does not mean A will definitely occur: e.g., $A = \{\frac{1}{2}\}^c$; but it will occur a.s.
- Let Q be the set of rational numbers on $[0,1]$. What is $\mathbb{P}(Q)$?

Constructing probability measures on $\mathbb{R}$

- For $\Omega = \mathbb{R}$, it suffices to specify probabilities $\mathbb{P}((-\infty, x])$ for $-\infty \leq x \leq +\infty$
- For any $-\infty \leq x \leq y \leq +\infty$

$$\mathbb{P}((x, y]) = \mathbb{P}((-\infty, y]) - \mathbb{P}((-\infty, x])$$

- The probability on $\mathcal{C} = \{$finite disjoint unions of intervals of the form $(x, y]\}$ could then be specified naturally
- $\mathcal{C}$ is an algebra; the probability on $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$ is hence uniquely determined

4. Conditional probability, independence

Reference: Jacod and Protter 2004, Chapter 3
Ross 2010, §1.4 - §1.6

Conditional probability

- Conditional on that B occurs (assuming $P(\mathbb{B}) > 0$), the probability that A occurs

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

- In **Example** 0.1, $A = \{HHT, HTH, THH\}$, $B = \{HHH, HHT\}$. Repeat the random experiment n times, approximately $\frac{2}{8}n$ of the outcomes are either HHH or HHT (i.e., B occurs). Among these $\frac{2}{8}n$ experiments, approximately $\frac{1}{8}n$ of the outcomes are HHT (i.e., A also occurs)

$$\mathbb{P}(A|B) = \frac{1/8}{2/8} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

- A family has two children of different ages. Conditional on that one of them is a boy, what is the probability that the other is a girl?

Independence

- A and B are **independent** if
$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$
- Suppose $\mathbb{P}(B) > 0$. A and B are independent iff
$$\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) = \mathbb{P}(A)$$
- $\{A_i, i \geq 1\}$ are **mutually independent** if $\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})$ for any $1 \leq i_1 < i_2 < \cdots < i_k$
- Mutually exclusive $\neq$ independent: flip a coin, $A = \{H\}$, $B = \{T\}$, A and B are exclusive of each other but dependent

A useful formula

- Let $A_1, \cdots, A_n$ be such that $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$

$$
\begin{aligned}
\mathbb{P}(A_1 \cap \cdots \cap A_n) &= \mathbb{P}((A_1 \cap \cdots A_{n-1}) \cap A_n) \\
&= \mathbb{P}(A_n | A_1 \cap \cdots A_{n-1}) P(A_1 \cap \cdots A_{n-1}) \\
&\vdots \\
&= \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \\
&\quad \cdots \mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1})
\end{aligned}
$$

Total probability rule

- Let $B_1, \cdots, B_n$ be a partition of Ω (i.e., B_i's are mutually exclusive and $\cup_{i=1}^n B_i = \Omega$) with positive probabilities

$$\mathbb{P}(A) = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \cdots + \mathbb{P}(A|B_n)\mathbb{P}(B_n)$$

- Open surgery has success rate of 93% for removing small stones and 73% for larger stones. 75% of patients treated by open surgery have large stones and 25% of them have small stones. What is the overall success rate of open surgery?

Bayes' formula

- Let $B_1, \cdots, B_n$ be a partition of Ω

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\mathbb{P}(A|B_1)\mathbb{P}(B_1) + \cdots + \mathbb{P}(A|B_n)\mathbb{P}(B_n)}$$

- There are 9 fair coins and 1 double-headed coin. You randomly pick 1 and flip for three times. Each time you get a head. What is the probability that the coin you take is double-headed?

5. Random variables

Reference: Jacod and Protter 2004, Chapters 5, 8
Ross 2010, §2.1 - §2.3

Random variable

- Interested in real valued functions defined on Ω

$$X: \quad \Omega \to \mathbb{R}$$

whose value depends on the outcome of the random experiment

- Interested in probabilities of the form

$$\begin{aligned} \mathbb{P}(X \in B) &= \mathbb{P}(\{\omega \in \Omega : \; X(\omega) \in B\}) \\ &= \mathbb{P}(X^{-1}(B)), \quad B = [a, b], (-\infty, x], \{a\}, etc. \end{aligned}$$

- A real valued **random variable (r.v.)** X is a **measurable function**: $(\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$

$$\{X \in B\} \in \mathcal{F}, \quad \forall B \in \mathcal{B}(\mathbb{R})$$

Example 0.3

In a 3-day period, stock price either increases by a factor of 2 or decreases by a factor of 1/2 during each day. All price paths are equally likely. Current stock price is 8.

- Define $(\Omega, \mathcal{F}, \mathbb{P})$ as in **Example** 0.1, $\mathcal{F} = 2^{\Omega}$, stock price increases if H, decreases if T
- Denote day-0,1,2,3 stock prices by S_0, S_1, S_2, S_3

$$S_0(\omega) = 8, \forall \omega \in \Omega, \text{ a constant r.v.}$$

$$S_1(\omega) = \begin{cases} 16 & \omega \in \{HHH, HHT, HTH, HTT\} \\ 4 & \omega \in \{THH, THT, TTH, TTT\} \end{cases}$$

Example 0.4

Given $([0,1], \mathcal{B}([0,1]), \mathbb{P})$ in **Example** 0.2, define

$$X(\omega) = \omega,$$

$$Y(\omega) = \begin{cases} -1, & \omega \in [0, 1/2) \\ 1, & \omega \in [1/2, 1] \end{cases}$$

where $\omega \in \Omega$ is the outcome of the random experiment

$\forall B \in \mathcal{B}(\mathbb{R})$, $\{Y \in B\} = \emptyset$ or $[0,1]$ or $[0,1/2)$ or $[1/2,1]$; all of them are in $\mathcal{F} = \mathcal{B}([0,1])$

Distribution

- The **distribution** P_X of a r.v. X

$$P_X(B) = \mathbb{P}(X \in B), \forall B \in \mathcal{B}(\mathbb{R})$$

  Probability distribution describes values a r.v. can take and the likelihoods

- P_X is a probability measure defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and is uniquely determined as long as one knows

$$P_X((-\infty, x]) = \mathbb{P}(X \leq x)$$

Cumulative distribution function

- Distribution of X uniquely determined by its **cumulative distribution function (cdf)**

$$F(x) = \mathbb{P}(X \le x)$$

- cdf of S_1 in **Example** 0.3

$$F(x) = \begin{cases} 0, & x \in (-\infty, 4) \\ 1/2, & x \in [4, 16) \\ 1, & x \in [16, \infty) \end{cases}$$

- cdf of X in **Example** 0.4: $F(x) = x, \ x \in [0, 1]$
- $F(x)$ must be (1) non-decreasing; (2) right continuous; (3) $F(-\infty) := \lim_{x \to -\infty} F(x) = 0, F(+\infty) := \lim_{x \to +\infty} F(x) = 1$

Discrete distributions

- **Discrete random variable**: X only takes values in $\{x_1, x_2, \cdots\}$; **probability mass function (pmf)**:

$$p_i = \mathbb{P}(X = x_i)$$

  pmf of S_1 in **Example** 0.3: $x_1 = 16, x_2 = 4, p_1 = \frac{1}{2}, p_2 = \frac{1}{2}$
- cdf of a discrete r.v.

$$F(x) = \mathbb{P}(X \leq x) = \sum_{i:\, x_i \leq x} p_i$$

  It is a **step function** with jumps at x_i, jump size p_i
- When X only takes values in $\{x_1, \cdots, x_n\}$, and $p_i = 1/n$, $1 \leq i \leq n$, X has a **discrete uniform** distribution

Continuous distributions

- The cdf of a **continuous random variable** admits the following form

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} p(y)dy$$

where the **probability density function (pdf)** p satisfies

$$p(x) \geq 0, \int_{\mathbb{R}} p(x)dx = 1$$

- X in **Example** 0.4 is **uniformly distributed** on $[0, 1]$

$$X \sim \text{Uniform}[0, 1], \quad p(x) = 1, x \in [0, 1]$$

Neither discrete nor continuous

- **Mixed distribution**: a r.v. may have probability mass at several points and density elsewhere
- Payoff of a call option $X = (S_T - K)^+$: X has a probability mass at zero and **partial density** on $(0, \infty)$

$$\mathbb{P}(X = 0) = \mathbb{P}(S_T \leq K)$$

For any $0 \leq a < b$,

$$\mathbb{P}(X \in (a, b]) = \mathbb{P}(a < S_T - K \leq b) = \int_a^b p(x)dx$$

where p is the density of $S_T - K$ and the partial density of X
- There exist distributions with **neither density nor probability mass**

6. Expectation

Reference: Jacod and Protter 2004, Chapter 9
Ross 2010, §2.4

Expectation

- **Expectation** of a r.v.: **probability weighted average**
- For countable Ω

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega)$$

  when the above summation is well defined
- Expectation of S_1 in **Example** 0.3

$$\begin{aligned} \mathbb{E}[S_1] &= 16(\mathbb{P}(\{HHH\}) + \mathbb{P}(\{HHT\}) + \mathbb{P}(\{HTH\}) + \mathbb{P}(\{HTT\})) \\ &\quad +4(\mathbb{P}(\{THH\}) + \mathbb{P}(\{THT\}) + \mathbb{P}(\{TTH\}) + \mathbb{P}(\{TTT\})) \\ &= 10 \end{aligned}$$

Expectation for simple r.v.'s

- $\sum_{\omega \in \Omega}$ not well defined for uncountable Ω (e.g., X, Y in **Example** 0.4)
- Expectation for **simple r.v.'s**: $X(\omega) = \sum_{i=1}^{n} x_i \mathbf{1}_{A_i}(\omega)$, where $A_i = \{\omega \in \Omega : X(\omega) = x_i\}$

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i \mathbb{P}(A_i)$$

Note that $Y(\omega) = -\mathbf{1}_{[0,1/2)}(\omega) + \mathbf{1}_{[1/2,1]}(\omega)$

$$\mathbb{E}[Y] = -\mathbb{P}([0, 1/2)) + \mathbb{P}([1/2, 1]) = 0$$

- For any event $A \in \mathcal{F}$,

$$\mathbb{P}(A) = \mathbb{E}[\mathbf{1}_A]$$

Expectations for positive r.v.'s

- Want to define expectations for general r.v.'s (including r.v.'s that are neither continuous nor discrete)
- For any $X \geq 0$, there exist simple r.v.'s $X_n \geq 0$ of the following form

$$X_n = \sum_{k=1}^{k_n} x_{n,k-1} \mathbf{1}_{\{x_{n,k-1} \leq X < x_{n,k}\}} + x_{n,k_n} \mathbf{1}_{\{X \geq x_{n,k_n}\}}$$

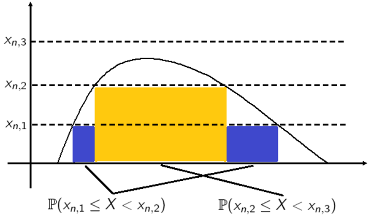
  such that $0 \leq X_n \uparrow X$. Define

$$\mathbb{E}[X] = \lim_{n \to +\infty} \mathbb{E}[X_n]$$

- $\mathbb{E}[X]$ is well defined, but could be $+\infty$

- More specifically,

$$\mathbb{E}[X] = \lim_{n\to+\infty}\left(\sum_{k=1}^{k_n} x_{n,k-1}\mathbb{P}(x_{n,k-1} \le X < x_{n,k}) + x_{n,k_n}\mathbb{P}(X \ge x_{n,k_n})\right)$$



- Integration of this type is called **Lebesgue integration** (partition is along the range of the integrand X)

$$\mathbb{E}[X] = \int_\Omega X(\omega)\mathbb{P}(d\omega) = \int_\Omega X d\mathbb{P}$$

Expectations for general r.v.'s

- For arbitrary X, define

  $$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-], \;\; X^+ = \max\{X, 0\}, \;\; X^- = \max\{-X, 0\}$$

  if the above is well defined (only $\infty - \infty$ is not well defined)
- For a discrete r.v. X with at most countably many possible values $\{x_1, x_2, \cdots\}$

  $$\mathbb{E}[X] = \int_\Omega X d\mathbb{P} = \sum_{i \geq 1} x_i \mathbb{P}(X = x_i)$$

  if the above summation is well defined

- $\mathbb{E}[X]$ in **Example** 0.4: partition $0 = x_{n,0} < \cdots < x_{n,k_n} = 1$

$$
\begin{aligned}
\mathbb{E}[X] &= \int_\Omega X d\mathbb{P} \\
&= \lim_{n\to+\infty} \sum_{k=1}^{k_n} x_{n,k-1}\mathbb{P}(x_{n,k-1} \le X < x_{n,k}) \\
&= \lim_{n\to+\infty} \sum_{k=1}^{k_n} x_{n,k-1}(x_{n,k} - x_{n,k-1}) \\
&= \int_0^1 x dx \;=\; \frac{1}{2}
\end{aligned}
$$

- For a continuous r.v. with pdf $p(x)$, $\mathbb{E}[X] = \int_{\mathbb{R}} xp(x)dx$

Properties of expectation

- X is **integrable** if $\mathbb{E}[|X|] < \infty$: i.e., $\mathbb{E}[X^+] < \infty, \mathbb{E}[X^-] < \infty$
- If X, Y are integrable and $X \leq Y$ a.s., $\mathbb{E}[X] \leq \mathbb{E}[Y]$
- If X is integrable, $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$
- **Linearity**: if X, Y are integrable and $a, b \in \mathbb{R}$

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

- **Monotone convergence**: if $0 \leq X_n \uparrow X$ a.s.

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[\lim_{n\to\infty} X_n] = \mathbb{E}[X]$$

Dominated convergence

- **Dominated convergence**: if $X_n \to X$ a.s. and $|X_n| < Y$ with $\mathbb{E}[Y] < \infty$, then $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$
- Let X_n be a sequence of r.v.'s. If $\sum_{n=1}^{\infty} \mathbb{E}[|X_n|] < \infty$, then

$$\mathbb{E}\Big[\sum_{n=1}^{\infty} X_n\Big] = \sum_{n=1}^{\infty} \mathbb{E}[X_n]$$

  That is, one can **interchange the order of summation and expectation**

7. Function of r.v.'s, moments

Reference: Ross 2010, §2.6

Function of r.v.'s

- For any r.v.'s $X_1, \cdots, X_n$ and a **measurable function** f from $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $f(X_1, \cdots, X_n)$ is also a r.v. (e.g., $f(x) = x^2, \exp(x), \ln(x), |x|$, all continuous functions, $f(x, y) = ax + by, xy, \max(x, y)$, etc.)
- All functions we encounter in this course are measurable
- Expectation of $f(X)$

$$X \text{ discrete: } \mathbb{E}[f(X)] = \sum_i f(x_i)\mathbb{P}(X = x_i)$$

$$X \text{ continuous with density } p(x): \mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)p(x)dx$$

Moments

- **Moments** of X

$$\mathbb{E}[X^m], \ m = 1, 2, \cdots$$

$\mu := \mathbb{E}[X]$ is the **mean** (or expectation, expected value); measures the location of the distribution of X

- **Central moments** of X

$$\mathbb{E}[(X - \mu)^m], \ m = 1, 2, \cdots$$

$\sigma^2 := \mathrm{var}(X) = \mathbb{E}[(X - \mu)^2]$ is the **variance**; measures the dispersion of the distribution of X; $\mathrm{stdev}(X) = \sigma$ is the **standard deviation** (same unit as X)

- For any $a, b \in \mathbb{R}$

$$\mathrm{var}(aX + b) = a^2 \mathrm{var}(X), \ \mathrm{stdev}(aX + b) = |a| \cdot \mathrm{stdev}(X)$$

Jensen's inequality

- σ^2 can also be computed with

$$\sigma^2 = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

  In particular, $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$

- For a convex function $f(x)$ (e.g., $f(x) = x^2$, $f(x) = e^x$), we have **Jensen's inequality**

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$$

  Generally,

$$\mathbb{E}[f(X)] \neq f(\mathbb{E}[X])$$

Skewness and kurtosis

- **Skewness** measures the asymmetry of the distribution

$$\frac{\mathbb{E}[(X-\mu)^3]}{\sigma^3}$$

  If the skewness is negative, the distribution is negatively skewed (left skewed)

- **Kurtosis** measures tail heaviness (how likely are extreme values)

$$\frac{\mathbb{E}[(X-\mu)^4]}{\sigma^4}$$

- Stock returns are often **negatively skewed** and have **fat tails**: longer tails on the left, kurtoses greater than 3 (3 is the kurtosis of a normal distribution)

Moment generating function

- **Moment generating function** of a r.v. X

$$M_X(t) = \mathbb{E}[e^{tX}]$$

  if the above expectation exists
- For any $n \geq 1$, $M_X^{(n)}(0) = \mathbb{E}[X^n]$, where $M_X^{(n)}(t)$ is the nth order derivative of $M_X(t)$
- It uniquely determines the distribution

Tails of a distribution

- **Markov's inequality**: if X is non-negative, then for any $x > 0$,

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}[X]}{x}$$

which follows from

$$x\mathbb{P}(X \geq x) = \mathbb{E}[x\mathbf{1}_{\{X \geq x\}}] \leq \mathbb{E}[X\mathbf{1}_{\{X \geq x\}}] \leq \mathbb{E}[X]$$

- **Chebyshev's inequality**: Suppose $\mathbb{E}[X] = \mu, \text{var}(X) = \sigma^2$. Then for any $x > 0$,

$$\mathbb{P}(|X - \mu| \geq x) \leq \frac{\sigma^2}{x^2}$$

8. Common discrete distributions

Reference: Jacod and Protter 2004, Chapter 5
Ross 2010, §2.2

Binomial distribution

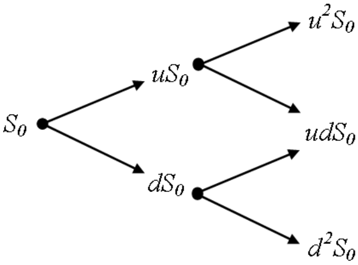
- **Bernoulli trial**: a random experiment with two outcomes, success with probability p and failure with probability $1-p$
- Bernoulli distribution: $\mathbb{P}(X=1)=p, \mathbb{P}(X=0)=1-p$
- **Binomial distribution** $B(n,p)$ models the number of successes out of n independent Bernoulli trials

$$\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\ 0\le k\le n,\ \mathbb{E}[X]=np$$

- Among N balls, M are black, the remaining are white. Randomly pick n (with replacement). $X=\#$ of black balls

$$\mathbb{P}(X=k)=\frac{\binom{n}{k}M^k(N-M)^{n-k}}{N^n}=\binom{n}{k}p^k(1-p)^{n-k}, p=\frac{M}{N}$$

- **Binomial tree:** Consider the price of an asset in $[0, T]$. There are n periods, each with length $\delta = T/n$. Over the period $[j\delta, (j+1)\delta]$, the asset price goes to $S_{(j+1)\delta} = uS_{j\delta}$ with probability p and $S_{(j+1)\delta} = dS_{j\delta}$ with probability $1-p$. Price movements over different periods are independent
- Illustration with $n = 2$



- Converges to a geometric Brownian motion; **No arbitrage pricing**/**risk neutral pricing**; **American style** contracts
- There are $\binom{n}{k}$ paths that lead to the price $S_0 u^k d^{n-k}$ at time T

$$\mathbb{P}(S_T = S_0 u^k d^{n-k}) = \binom{n}{k} p^k (1-p)^{n-k}, \; 0 \le k \le n$$

- Pricing a **European call option** with strike price K reduces to computing

$$\mathbb{E}[\max(0, S_T - K)] = \sum_{k=0}^{n} \max(0, S_0 u^k d^{n-k} - K) \binom{n}{k} p^k (1-p)^{n-k}$$

Hypergeometric distribution

- Among N balls, M are black, the remaining are white. Randomly pick n (without replacement). $X = \#$ of black balls

$$\mathbb{P}(X = k) = \frac{\#(A)}{\#(\Omega)} = \frac{\left(\begin{array}{c} M \\ k \end{array}\right)\left(\begin{array}{c} N - M \\ n - k \end{array}\right)}{\left(\begin{array}{c} N \\ n \end{array}\right)}$$

$$\max(0, n + M - N) \leq k \leq \min(n, M)$$

where $\#(\Omega) =$ possible ways of selecting n balls from N, $\#(A) =$ possible ways to get m black balls among n randomly selected balls

Geometric and negative binomial distributions

- **Geometric distribution** models the number of independent Bernoulli trials needed until the first success

$$\mathbb{P}(X = n) = (1-p)^{n-1}p, \quad n \geq 1, \ \mathbb{E}[X] = 1/p$$

Let Y be the number of failures until the first success

$$\mathbb{P}(Y = n) = (1-p)^{n}p, \quad n \geq 0$$

- **Negative binomial distribution** models the number of failures until the kth success occurs ($k \geq 1$)

$$\mathbb{P}(X = n) = \binom{n+k-1}{k-1} p^k (1-p)^n, \ n \geq 0$$

Poisson distribution

- Models the **number of arrivals** of a certain object (insurance claims, asset price jumps, etc.) in a given period: λ arrivals on average; arrivals occur independently
- Divide the period into n subintervals for a large n so that, in each subinterval, there is one arrival with probability p, and no arrival with probability $1-p$, where $p = \lambda/n$. Let N be the number of total arrivals

$$\mathbb{P}(N = k) = \left(\begin{array}{c} n \\ k \end{array} \right) (\frac{\lambda}{n})^k (1 - \frac{\lambda}{n})^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}, \ n \to +\infty$$

- **Poisson distribution** with arrival rate (intensity) λ

$$\mathbb{P}(N = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \ k \geq 0, \ \mathbb{E}[N] = \mathrm{var}(N) = \lambda$$

9. Common continuous distributions

Reference: Jacod and Protter 2004, Chapter 7
Ross 2010, §2.3

Normal distribution

- The pdf of a normal r.v. $X \sim N(\mu, \sigma^2)$:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

  In particular,

$$\mathbb{E}[X] = \mu$$

$$\mathrm{var}(X) = \mathbb{E}[(X-\mu)^2] = \sigma^2$$

- Normal distributions are also called **Gaussian** distributions
- Suppose $X \sim N(\mu, \sigma^2)$, then

$$aX + b \sim N(a\mu + b, a^2\sigma^2)$$

Standard normal distribution

- **Standard normal distribution**: $\mu = 0$, $\sigma = 1$
- **Standardization**: if $X \sim N(\mu, \sigma^2)$, then

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

- cdf of a standard normal r.v. $Z \sim N(0, 1)$:

$$\Phi(x) = \mathbb{P}(Z \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$$

- No analytical expression for $\Phi(x)$; numerically implemented in statistical software

Uniform and exponential distributions

- **Uniform distribution** $U[0,1]$: the pdf and cdf of $X \sim U[0,1]$ are

$$p(x) = 1, \; F(x) = \mathbb{P}(X \leq x) = x, \; \forall x \in [0,1]$$

- If the number of arrivals is Poisson with λ arrivals per unit time on average, the inter-arrival time has an **exponential distribution** $Exp(\lambda)$: the pdf and cdf of $X \sim Exp(\lambda)$:

$$p(x) = \lambda e^{-\lambda x}, \; F(x) = 1 - e^{-\lambda x}, \; x > 0, \; \mathbb{E}[X] = 1/\lambda$$

- Uniform distribution is essential for **Monte Carlo simulation**; uniform and exponential distributions can be used to simulate normal r.v.'s

Memorylessness

- Exponential distribution is memoryless: $X \sim Exp(\lambda)$, $t, s \geq 0$

$$\begin{aligned} \mathbb{P}(X \geq t+s | X \geq t) &= \frac{\mathbb{P}(X \geq t+s)}{\mathbb{P}(X \geq t)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\ &= e^{-\lambda s} \\ &= \mathbb{P}(X \geq s) \end{aligned}$$

Conditional on $X \geq t$, the probability that $X \geq t+s$ is the same as the probability of $X \geq s$

Density of $f(X)$

- X is a continuous r.v. with density p_X. f is continuously differentiable strictly monotone with inverse f^{-1}. Then the density of $Y = f(X)$ is

$$p_Y(y) = p_X(f^{-1}(y)) \left| \frac{df^{-1}(y)}{dy} \right|$$

- **Lognormal distribution**: suppose $X \sim N(\mu, \sigma^2)$. Then $Y = \exp(X)$ is lognormally distributed
- Stock price is assumed to have a lognormal distribution in the **Black-Scholes-Merton** model

Lognormal distribution

- Probability density: let $p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$. Then

$$p_Y(y) = \frac{1}{y} p_X(\ln(y)), \;\; y > 0$$

- Expected value: $\mathbb{E}[Y] = \mathbb{E}[\exp(X)] \neq \exp(\mu)$

$$\begin{aligned}
\mathbb{E}[Y] &= \int_{\mathbb{R}} \exp(x) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} \exp\left(\mu + \frac{1}{2}\sigma^2\right) \exp\left(-\frac{(x-\mu-\sigma^2)^2}{2\sigma^2}\right) dx \\
&= \exp\left(\mu + \frac{1}{2}\sigma^2\right)
\end{aligned}$$

Heavy-tailed distributions

- Financial data exhibit distributions with heavy (fat) tails: extreme movements are common
- Exponential tails are heavier than tails of normal distributions: e.g., **Laplace distribution** ($b > 0$)

$$p(x) = \frac{1}{2b} e^{-|x-\mu|/b}, \quad x \in \mathbb{R}$$

- Polynomial tails are heavy: e.g., **Cauchy distribution**

$$p(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}$$

Mean does not exist. **Pareto distribution**: for $a > 0, c > 0$

$$p(x) = ac^a/x^{a+1}, \quad x > c$$

Gamma distribution

- Gamma distribution with parameters $\alpha > 0, \beta > 0$ has density

$$p(y) = \frac{\beta^\alpha y^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta y), y > 0$$

  where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ is the Gamma function ($\Gamma(1/2) = \sqrt{\pi}$; for positive integer n, $\Gamma(n) = (n-1)!$)
- Exponential distribution a special case with $\alpha = 1$
- α shape parameter, $1/\beta$ scale parameter, β rate parameter

Location, scale parameters

- **Location parameter** horizontally shifts a distribution without changing its shape

$$X \sim p(x) \Leftrightarrow X + a \sim p(x - a)$$

A location parameter a only appears in the form of $x - a$ in the pdf; e.g., μ in $N(\mu, \sigma^2)$

- **Scale parameter** changes the shape of a distribution via stretching

$$X \sim p(x) \Leftrightarrow bX \sim \frac{1}{|b|} p\left(\frac{x}{b}\right), b \neq 0$$

A scale parameter b only appears in the form of $\frac{1}{|b|} p(\frac{x-a}{b})$; e.g., σ in $N(\mu, \sigma^2)$

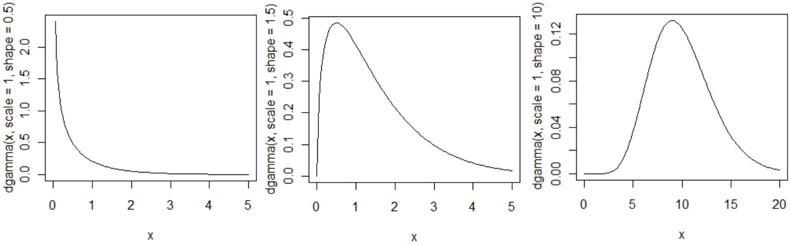
Shape parameter

- Suppose $X \sim p(x)$, $a, b \in \mathbb{R}$, $b \neq 0$

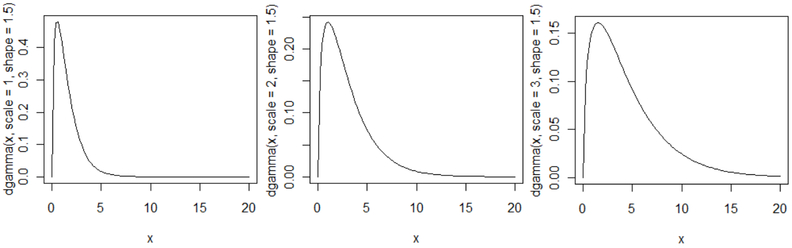
$$bX + a \sim \frac{1}{|b|} p\left(\frac{x-a}{b}\right)$$

- **Shape parameter** changes the shape of a distribution not simply by stretching
- Gamma distribution with parameters $\alpha > 0, \beta > 0$
  - $b = 1/\beta$ is a scale parameter
  - α is a shape parameter

- Gamma distribution with $b = 1$ and $\alpha = 0.5, 1.5, 10$ (varying shape parameter)



- Gamma distribution with $b = 1, 2, 3$ (varying scale parameter) and $\alpha = 1.5$



10. Characteristic function

Reference: Jacod and Protter 2004, Chapters 13, 14

Characteristic function

- The **characteristic function** of a r.v. X

$$\phi_X(\xi) = \mathbb{E}[e^{i\xi X}]$$

- $\phi_X(\xi)$ always exists for $\xi \in \mathbb{R}$
- $|\phi_X(\xi)| \leq \phi_X(0) = 1$ for any $\xi \in \mathbb{R}$
- $\phi_X(\xi)$ is uniformly continuous on $\mathbb{R}$
- It uniquely determines the distribution (one to one correspondence)
- For any $a, b \in \mathbb{R}$,

$$\phi_{aX+b}(\xi) = e^{i\xi b}\phi_X(a\xi)$$

Computing moments

- If $\mathbb{E}[|X|^n] < \infty$, then $\phi_X(\xi)$ has a continuous nth order derivative
$$\phi_X^{(n)}(\xi) = i^n \mathbb{E}[X^n e^{i\xi X}]$$
In particular, $\mathbb{E}[X^n] = (-i)^n \phi_X^{(n)}(0)$
- For $Z \sim N(0,1)$,
$$\phi_Z(\xi) = e^{-\frac{1}{2}\xi^2}$$
$$\phi_Z^{(3)}(\xi) = (3\xi - \xi^3)e^{-\frac{1}{2}\xi^2}$$
$$\phi_Z^{(4)}(\xi) = (3 - 6\xi^2 + \xi^4)e^{-\frac{1}{2}\xi^2}$$
- For any $X \sim N(\mu, \sigma^2)$, let $Z = (X-\mu)/\sigma \sim N(0,1)$. The **skewness** of X equals $\mathbb{E}[Z^3] = 0$, the **kurtosis** of X equals $\mathbb{E}[Z^4] = 3$

Fourier transform

- Fourier transform of $f \in L^1(\mathbb{R})$

$$\hat{f}(\xi) = \mathcal{F}f(\xi) = \int_{\mathbb{R}} e^{i\xi x} f(x)dx$$

- Characteristic function of a continuous r.v. X with pdf $p(x)$ is simply the Fourier transform of the pdf

$$\phi(\xi) = \mathbb{E}[e^{i\xi X}] = \int_{\mathbb{R}} e^{i\xi x} p(x)dx = \mathcal{F}p(\xi)$$

- **Caution**: there are several slightly different definitions of Fourier transform

Inversion formulas

- **Inverse Fourier transform**: if $\hat{f} \in L^1(\mathbb{R})$

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\xi x} \hat{f}(\xi) d\xi$$

- Recover pdf of a r.v. when $|\phi|$ is integrable

$$p(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\xi x} \phi(\xi) d\xi$$

- **Gil-Pelaez formula** for the cdf

$$F(x) = \frac{1}{2} + \frac{1}{2\pi} \int_0^\infty \frac{e^{i\xi x}\phi(-\xi) - e^{-i\xi x}\phi(\xi)}{i\xi} d\xi$$

- In many financial models, only know c.f., but not pdf: stochastic volatility, jump diffusion, Lévy models
- Asset price is modeled by the following

$$S_t = S_0 e^{X_t}, t \geq 0$$

- Pricing a **cash-or-nothing put** option with strike price K and maturity T (such an option pays one unit of cash if the asset price at maturity does not exceed K)

$$V = \mathbb{E}[e^{-rT}\mathbf{1}_{\{S_T \leq K\}}] = e^{-rT}\mathbb{P}\left(X_T \leq \ln(\frac{K}{S_0})\right)$$

- The **Gil-Pelaez formula** could be used to price the above option

11. Joint distributions

Reference: Jacod and Protter 2004, Chapters 12
Ross 2010, §2.5

Joint distribution

- **Joint cdf** of r.v.'s X and Y

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y)$$

- **Joint pdf** $p(x, y)$ of continuous r.v.'s X and Y

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(u, v) dv du$$

- **Marginal distribution** and marginal density

$$F_X(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} p(u, v) dv du, \quad p_X(u) = \int_{-\infty}^{\infty} p(u, v) dv$$

- Multi-dimensional/discrete cases can be handled similarly

Conditional distribution

- Given joint pdf $p(x, y)$, expectation of $f(X, Y)$

$$\mathbb{E}[f(X, Y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y)p(x, y)dydx$$

  E.g.,

$$\mathbb{E}[Y] = \int_{\mathbb{R}} \int_{\mathbb{R}} yp(x, y)dydx$$

- Given discrete r.v.'s X, Y. **Conditional pmf** of Y given $X = x_i$

$$P(Y = y_j | X = x_i) = \frac{\mathbb{P}(X = x_i, Y = Y_j)}{\mathbb{P}(X = x_i)}$$

- Given continuous r.v.'s X, Y. **Conditional pdf** of Y given $X = x$

$$p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)}, \quad p(x,y) = p_X(x) p_{Y|X}(y|x)$$

- Intuitively, for small $\Delta x, \Delta y$

$$\mathbb{P}(X \in [x, x+\Delta x]) = \int_x^{x+\Delta x} p_X(y) dy \approx p_X(x) \Delta x$$

$$\mathbb{P}(X \in [x, x+\Delta x], Y \in [y, y+\Delta y]) \approx p(x,y) \Delta x \Delta y$$

$$p_{Y|X}(y|x) \Delta y \approx \frac{\mathbb{P}(X \in [x, x+\Delta x], Y \in [y, y+\Delta y])}{\mathbb{P}(X \in [x, x+\Delta x])} = \frac{p(x,y)\Delta x \Delta y}{p_X(x) \Delta x}$$

- The joint pdf p of X, Y, Z satisfies

$$p(x,y,z) = p_X(x) \cdot p_{Y|X}(y|x) \cdot p_{Z|X,Y}(z|x,y)$$

Conditional expectation

- Suppose X, Y are continuous. **Conditional expectation** of Y given that $X = x$

$$\mathbb{E}[Y|X = x] = \int_{\mathbb{R}} y p_{Y|X}(y|x) dy$$

Conditional variance

$$\text{var}(Y|X = x) = \mathbb{E}[Y^2|X = x] - (\mathbb{E}[Y|X = x])^2$$

Useful properties

- Let $\mathbb{E}[Y|X]$ denote the function of X that takes value $\mathbb{E}[Y|X=x]$ when $X=x$. $\mathbb{E}[Y|X]$ is a random variable
- **Iterated conditioning**: $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$. E.g., if X, Y are discrete,

$$\begin{aligned}
\sum_i \mathbb{E}[Y|X=x_i]\mathbb{P}(X=x_i) &= \sum_i \sum_j y_j \mathbb{P}(Y=y_j|X=x_i)\mathbb{P}(X=x_i) \\
&= \sum_i \sum_j y_j \mathbb{P}(X=x_i, Y=y_j)
\end{aligned}$$

- **Take out what is known**: if $Z=f(X)$, $\mathbb{E}[ZY|X] = Z\mathbb{E}[Y|X]$

- X and Y are **independent** if

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B), \quad \forall A, B \in \mathcal{B}(\mathbb{R})$$

  or equivalently $F(x, y) = F_X(x)F_Y(y)$, or for continuous r.v.'s, $p(x, y) = p_X(x)p_Y(y)$ (i.e., $p_{Y|X}(y|x) = p_Y(y)$)
- Suppose the joint pdf of X and Y is $p(x, y) = 8xy$, $0 \leq x \leq 1$, $0 \leq y \leq x$. Are X and Y independent?

- If X and Y are independent, then $\mathbb{E}[Y|X] = \mathbb{E}[Y]$

Example 0.5

In a jump diffusion model, the asset return contains a term $\sum_{n=1}^{N_t} X_n$, where N_t is the number of market shocks up to time t, X_n is the jump size in the return when the nth shock occurs. $\{X_1, X_2, \cdots\}$ are i.i.d. and independent of N_t. What is $\mathbb{E}\left[\sum_{n=1}^{N_t} X_n\right]$?

$$
\begin{aligned}
\mathbb{E}\left[\sum_{n=1}^{N_t} X_n\right] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{n=1}^{N_t} X_n \middle| N_t\right]\right] \\
&= \sum_{N=0}^{\infty} \mathbb{E}\left[\sum_{n=1}^{N} X_n \middle| N_t = N\right] \mathbb{P}(N_t = N) \\
&= \mathbb{E}[X_1] \sum_{N=0}^{\infty} N\mathbb{P}(N_t = N) \;=\; \mathbb{E}[X_1]\mathbb{E}[N_t]
\end{aligned}
$$

Sum of two independent r.v.'s

- Two independent continuous r.v.'s X and Y with pdf's $p_X(x)$ and $p_Y(y)$, respectively
- The cdf of $X + Y$

$$\mathbb{P}(X + Y \le z) = \int_{\mathbb{R}} \int_{-\infty}^{z-y} p_X(x) p_Y(y) dx dy$$

- Take derivative w.r.t. z, one obtains the density of $X + Y$:

$$\int_{\mathbb{R}} p_X(z - y) p_Y(y) dy$$

which is the **convolution** of p_X and p_Y

Covariance and correlation

- Suppose X and Y are independent, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y], \quad \mathbb{E}[Y|X] = \mathbb{E}[Y]$$

$f(X), g(Y)$ are independent,

$$M_{X+Y}(t) = M_X(t)M_Y(t), \; \phi_{X+Y}(\xi) = \phi_X(\xi) \cdot \phi_Y(\xi)$$

- $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ does not imply independence
- **Covariance** and **correlation coefficient**

$$\sigma_{XY} = \text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y], \; \rho = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

where σ_X, σ_Y are the standard deviations of X and Y. They measure the linear relationship between X and Y

- Covariance is bilinear:

$$\text{cov}(aX + bY, Z) = a \cdot \text{cov}(X, Z) + b \cdot \text{cov}(Y, Z)$$

$$\text{cov}(X, aY + bZ) = a \cdot \text{cov}(X, Y) + b \cdot \text{cov}(X, Z)$$

- Simple facts: $\text{cov}(X, X) = \text{var}(X)$

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$$

$$\begin{aligned} \text{var}\left(\sum_{i=1}^{n} w_i X_i\right) &= \sum_{i,j=1}^{n} w_i w_j \text{cov}(X_i, X_j) = \sum_{i=1}^{n} w_i^2 \text{var}(X_i) \\ &\quad + 2 \sum_{1 \le i < j \le n} w_i w_j \text{cov}(X_i, X_j) \end{aligned}$$

12. Multivariate normal distribution

Reference: Jacod and Protter 2004, Chapter 16

Multivariate normal distribution

- Characteristic function of $X = (X_1, \cdots, X_n)^\top$

$$\phi(\xi) = \mathbb{E}[e^{i\xi^\top X}] = \mathbb{E}[e^{i\xi_1 X_1 + \cdots i\xi_n X_n}]$$

- X has a **multivariate normal distribution** $N(\mu, \Sigma)$ if

$$\phi(\xi) = \exp\left(i\mu^\top \xi - \frac{1}{2}\xi^\top \Sigma \xi\right)$$

  for $\mu \in \mathbb{R}^n$ and **symmetric positive semi-definite** matrix Σ
  - Σ is the **covariance matrix** with entries $\Sigma_{ij} = \mathrm{cov}(X_i, X_j)$
  - Positive semi-definite: $\forall a \in \mathbb{R}^n$,

$$a^\top \Sigma a = \mathrm{cov}(a_1 X_1 + \cdots + a_n X_n, a_1 X_1 + \cdots + a_n X_n) \geq 0$$

  - Positive definite if strict inequality in the above (Σ^{-1} exists)

- Marginal distributions: $X_i \sim N(\mu_i, \sigma_i^2)$ where $\sigma_i^2 = \Sigma_{ii}$
- X has joint pdf (if Σ^{-1} exists)

$$p(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)$$

- X is **standard multivariate normal** if $X \sim N(0, I_n)$ where I_n is an $n \times n$ identity matrix
  - $X \sim N(0, I_n)$ iff $X_i \sim N(0, 1)$ and are i.i.d
- Any linear combination $a_1X_1 + \cdots a_nX_n$ is normal. More generally, for any $m \times n$ matrix A:

$$AX \sim N(A\mu, A\Sigma A^\top)$$

- Components of a multivariate normal r.v. are normal
- **The converse is not true**: given arbitrary normal r.v.'s X and Y, (X, Y) is not necessarily bivariate normal, $X + Y$ is not necessarily normal
- But if X and Y are independent, $X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$, then (X, Y) is bivariate normal,

$$X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$

Example 0.6

The daily log returns of a certain stock in a week are denoted by r_1, r_2, r_3, r_4. The weekly return is given by

$$r = r_1 + \cdots + r_4.$$

Suppose the daily returns are independent and normally distributed with mean $\mu = 0.001$ and standard deviation $\sigma = 0.02$. Which of the following is the distribution of the weekly return r?

$$4N(\mu, \sigma^2) = N(4\mu, 16\sigma^2)$$

$$N(4\mu, 4\sigma^2)$$

Cholesky decomposition

- If $Z \sim N(0, I_n)$, then

$$X = \mu + AZ \sim N(\mu, AA^\top) \text{ for any } n \times n \text{ matrix } A$$

- For $X \sim N(\mu, \Sigma)$, find A such that $AA^\top = \Sigma$
- **Cholesky decomposition** (when Σ is positive definite)

$$\Sigma = \begin{pmatrix} a_{11} & & & \\ a_{21} & a_{22} & & \\ \vdots & \vdots & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ & a_{22} & \cdots & a_{n2} \\ & & \ddots & \vdots \\ & & & a_{nn} \end{pmatrix}$$

- Bivariate normal r.v.'s

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

- Cholesky decomposition

$$AA^\top = \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} \\ 0 & a_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Equations to solve:

$$a_{11}^2 = \sigma_1^2, \quad a_{11}a_{21} = \rho\sigma_1\sigma_2, \quad a_{21}^2 + a_{22}^2 = \sigma_2^2$$

- One solution to the above

$$A = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1-\rho^2}\ \sigma_2 \end{pmatrix}$$

- *Simulate a bivariate normal r.v.* $X = (X_1, X_2)^\top$
  1. *Simulate two independent standard normal r.v.'s* $Z_1, Z_2 \sim N(0,1)$
  2. $X_1 = \mu_1 + \sigma_1 Z_1$
  3. $X_2 = \mu_2 + \rho\sigma_2 Z_1 + \sqrt{1-\rho^2}\ \sigma_2 Z_2$