Stochastic Processes and Financial Mathematics (37363)

Chapter 1

Introduction to axiomatic approach, probability basics

Alex Novikov and Scott Alexander

School of Mathematical and Physical Sciences, UTS

Autumn 2025

Recommended texts

Books on analysis

- *Principles of Mathematical Analysis* [Rudin, 1976]
- *Real and Complex Analysis* [Rudin, 1987]

Books on probability and statistics

- *Mathematical Statistics and Data Analysis* [Rice, 2007] (basic – assumed knowledge)
- *Mathematical Statistics with Applications* [Wackerly et al., 2008] (basic – assumed knowledge)
- *A Course in Probability Theory* [Chung, 2001] (advanced)

Recommended texts

Books on stochastic processes and mathematical finance

- *Elements of Stochastic Modelling* [Borokov, 2014]
- *A Benchmark Approach to Quantitative Finance* [Platen and Heath, 2006]
- *Stochastic Calculus for Finance I* [Shreve, 2005]
- *Stochastic Calculus for Finance II* [Shreve, 2004] (advanced)
- *Financial Modelling with Jump Processes* [Cont and Tankov, 2004] (advanced)

Chapter outline

Topics:

- History of probability theory
- Frequency approach to probability
- Introduction to Kolmogorov's axiomatic approach
- Conditional probability and independent events
- Scalar RVs
  - distribution (discrete case)
  - examples (discrete case)
  - distribution (continuous case)
  - examples (continuous case)
  - expectation of functions
  - expected value
  - MGF
  - CF

Chapter outline

Topics:

- Vector RVs
  - expected value
  - joint distribution
  - marginal distributions
  - expectation of functions
  - MGF and CF
  - covariance
  - autocovariance
- Independent RVs
  - joint distribution and expectation
- Conditional expectation

History of probability theory

17th century

- classical probability, combinatorial methods
- Pascal B. (1623-1662), Fermat P. (1607-1665)

18th century

- geometric probabilities, law of large numbers (1713)
- Bernoulli J. (1654-1705), Euler L. (1707-1783)

19th century

- central limit theorem, analytical methods
- Gauss C. (1777-1855), Poisson S.D. (1781-1840)

20th century

- axiomatic approach, stochastic processes
- Markov A.A. (1856-1922), Kolmogorov A.N. (1903-1987), Wiener N. (1894-1964)

Frequency approach to probability

Outcomes of **random experiments** are not predictable.

What are the probabilities of different outcomes, or **events**, from random experiments?

The probability of an event can be measured by the relative frequency of the event.

Let n be the total number of independent trials of a random experiment.

Let $n(A)$ be the number of these trials in which the event A occurred.

The **relative frequency** of the event A is given by the ratio

$$\frac{n(A)}{n}.$$

Frequency approach to probability

The **frequency approach** defines the probability of the event A as

$$P(A) = \lim_{n \to \infty} \frac{n(A)}{n}$$

or, for large n

$$P(A) \approx \frac{n(A)}{n}.$$

So, if you know $n(A)$ you have the approximation $\frac{n(A)}{n}$.

Often it is used in the opposite direction: if you know $P(A)$ then you can predict $n(A)$. For large n it approximately equals $P(A)n$.

A general question: how to formally define $P(A)$?

All mathematicians and statisticians now use **Kolmogorov's axiomatic approach**.

Introduction to Kolmogorov's axiomatic approach

Modern theory on probability and stochastic processes employs an axiomatic approach.

From these axioms, assertions are formulated, proved and an increasingly detailed theory constructed.

The power of this method lies in the avoidance of mathematical inconsistencies, fallacies and paradoxes.

However, a need for advanced concepts and techniques requires of students a solid training in mathematics.

For our purposes these technical details are not required, but we will provide a (very) brief introduction to some of these concepts.

Introduction to Kolmogorov's axiomatic approach

The axiomatic approach revolves around the triple

$$(\Omega, \mathcal{F}, P)$$

called a **probability space**.

The reasons for considering such an object are technical but essentially allow

- random variables (and stochastic processes) to be constructed as **measurable functions** on $(\Omega, \mathcal{F}, P)$
- expectations of functions of these random variables (and stochastic processes) to be defined as **Lebesgue integrals** on $(\Omega, \mathcal{F}, P)$.

We won't go into all the details, but will describe the main features and introduce others as needed in later chapters.

See [Chung, 2001] for a comprehensive treatment of random variables.

Introduction to Kolmogorov's axiomatic approach

Definition 1 (sample space Ω and events)

The sample space Ω is the set of all (mutually exclusive) elementary events $\omega \in \Omega$.

Any $E \subseteq \Omega$ is called an event, which we write in shorthand as

$$E = \{\omega | \text{“inclusion criterion”}\},$$

where the “inclusion criterion” is a rule that determines whether $\omega \in E$ or $\omega \in E^c$.

So any event is a union of elementary events.

Note that the definition above includes

- the null event $\emptyset \subseteq \Omega$
- the sure event $\Omega \subseteq \Omega$.

Introduction to Kolmogorov's axiomatic approach

Set notation
Consider the events $E, E_1, E_2 \subseteq \Omega$.

Then

- $E_1 \cap E_2 = \{\omega | \omega \in E_1 \text{ and } \omega \in E_2\}$ (intersection)
- $E_1 \cup E_2 = \{\omega | \omega \in E_1 \text{ or } \omega \in E_2\}$ (union)
- $E^c = \Omega \setminus E$ (compliment of E)
- $E_1 \setminus E_2 = E_1 \cap E_2^c$ (compliment of E_2 in E_1)

$\subseteq \Omega$ are also events.

If $E_1 \cap E_2 = \emptyset$ then the events E_1 and E_2 are said to be disjointed.

Introduction to Kolmogorov's axiomatic approach

Definition 2 (sigma-algebra $\mathcal{F}$)

A class $\mathcal{F}$ of events is called as sigma-algebra (or sigma-field) if

1. $\Omega \in \mathcal{F}$
2. if $E \in \mathcal{F}$ then $E^c \in \mathcal{F}$.
3. if $E_1 \in \mathcal{F}$ and $E_2 \in \mathcal{F}$ then $E_1 \cup E_2 \in \mathcal{F}$
4. if $E_i \in \mathcal{F}$ for all $i = 1, 2, \ldots$ then $\cup_{i=1}^{\infty} E_i \in \mathcal{F}$.

Note that

- $\emptyset \in \mathcal{F}$ by (1) and (2)
- $E_1 \cap E_2 = (E_1^c \cup E_2^c)^c \in \mathcal{F}$ by (2) and (3)
- $\cap_{i=1}^{\infty} E_i = (\cup_{i=1}^{\infty} E_i^c)^c \in \mathcal{F}$ by (2) and (4)

and De Morgan's laws.

Examples include

- $\mathcal{F} = \{\emptyset, \Omega\}$ (trivial sigma-algebra)
- $\mathcal{F} = \{\emptyset, E, E^c, \Omega\}$ (sigma-algebra generated by event E).

Introduction to Kolmogorov's axiomatic approach

Definition 3 (probability measure P)

The set function P is called a probability measure if

1. if $E \in \mathcal{F}$ then $0 \leq P(E) \leq 1$
2. $P(\emptyset) = 0$ and $P(\Omega) = 1$
3. if $E_1, E_2 \in \mathcal{F}$ and $E_1 \cap E_2 = \emptyset$ then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$
4. if $E_i \in \mathcal{F}$ for any $i = 1, 2, \ldots$ and $E_i \cap E_j = \emptyset$ when $i \neq j$ then $P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$.

That is, a probability measure can be considered as the mapping $P : \mathcal{F} \to [0, 1]$.

It can be shown that if $E_1, E_2 \in \mathcal{F}$ then

- $P(E_1 \setminus E_2) = P(E_1) - P(E_1 \cap E_2)$
- $P(E_1) \leq P(E_2)$ if $E_1 \subset E_2$
- $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.

Conditional probability and independent events

In applications one often works with the notion of conditional probability.

Definition 4 (conditional probability)

Let $E_1, E_2 \subseteq \Omega$. If $P(E_2) > 0$ then the ratio

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

is called the conditional probability that E_1 will occur given E_2 has occurred.

A conditional probability is a probability and so satisfies the conditions of Definition 3.

Conditional probability and independent events

Conditional probability simplifies in the case of independent variables.

Definition 5 (independent event)

Events $E_1, E_2 \subseteq \Omega$ are independent if

$$P(E_1 \cap E_2) = P(E_1)P(E_2).$$

If $P(E_1), P(E_2) > 0$ then by Definition 4

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = \frac{P(E_1)P(E_2)}{P(E_2)} = P(E_1)$$

and similarly

$$P(E_2|E_1) = \frac{P(E_2 \cap E_1)}{P(E_1)} = \frac{P(E_2)P(E_1)}{P(E_1)} = P(E_2).$$

Conditional probability and independent events

Finally for this section, an important result that we will use with Markov chains.

Theorem 1 (total law of probability)

Let events E_i, $i \geq 1$, form a partition of Ω, i.e. $\cup_i E_i = \Omega$ and $E_i \cap E_j = \emptyset$ for $i \neq j$, and suppose $P(E_i) > 0$ for all i.

Then for any event A

$$P(A) = \sum_{i \geq 1} P(A \cap E_i) = \sum_{i \geq 1} P(A|E_i)P(E_i).$$

Scalar RVs

A scalar random variable (RV) can be used to model a numerical outcome of some random experiment or phenomenon.

Mathematically, a RV X on the probability space $(\Omega, \mathcal{F}, P)$ is a function

$$X : \Omega \to R$$

where $R \subseteq \mathbb{R}$, i.e. some subset of the set of real numbers $\mathbb{R}$.

The RV X is called

- **discrete** if R is countable (e.g. the set of integers $\mathbb{Z}$, natural numbers $\mathbb{N}$, some countable subset of $\mathbb{R}$ etc.)
- **continuous** if R is uncountable (e.g. the set of real numbers $\mathbb{R}$, non-negative real numbers $\mathbb{R}_{\geq 0}$, some interval of $\mathbb{R}$ etc.).

Scalar RVs

A RV can be more precisely defined as a function that maps between two measurable spaces.

Definition 6 (random variable)

A RV X on $(\Omega, \mathcal{F}, P)$ is an $\mathcal{F}$-measurable function

$$X : (\Omega, \mathcal{F}) \to (R, \mathcal{G})$$

where $R \subseteq \mathbb{R}$ and $\mathcal{G}$ a sigma-algebra over R.

By $\mathcal{F}$-measurable we mean that for each $G \in \mathcal{G}$, the inverse image (or pre-image)

$$X^{-1}(G) = \{\omega | X(\omega) \in G\} \equiv \{X \in G\} \in \mathcal{F}.$$

The nature of $\mathcal{G}$ depends on whether X is discrete or continuous and details of its construction are omitted.

Scalar RVs – distribution (discrete case)

The distribution of a RV is its most important property and describes how probability is distributed about outcomes of the RV.

Definition 7 (probability functions of discrete scalar RV)

If X is a discrete RV then it has a probability mass function (PMF) $p_X \in [0, 1]$ defined as

$$p_X(b) = P(X = b) \equiv P\big(\{\omega | X(\omega) = b\}\big)$$

with the property

$$\sum_x p_X(x) = 1.$$

The cumulative distribution function (CDF) $F_X \in [0, 1]$ is defined as

$$F_X(b) = P(X \leq b) \equiv P\big(\{\omega | X(\omega) \leq b\}\big) = \sum_{x \leq b} p_X(x).$$

Note the last follows from $\{X \leq b\} = \cup_{u \leq b}\{X = u\}$.

Scalar RVs – examples (discrete case)

Example (binomial distribution). Let $0 \leq p \leq 1$ and the RV X take the values $x \in \{0, 1, \ldots, n\}$. The PMF of X is given by

$$p_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where the binomial coefficients

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

In this case we write $X \sim B(n, p)$ to indicate X has the binomial distribution with parameters n, p.

Application. Consider a random experiment with probability of success p (a "Bernoulli trial"). If n replications are performed the probability that x will be successful is $p_X(x)$.

Scalar RVs – examples (discrete case)

Example (geometric distribution). Let $0 \leq p \leq 1$ and the RV X take the values $x \in \{1, 2, \ldots\}$. The PMF of X is given by

$$p_X(x) = P(X = x) = p(1-p)^{x-1}.$$

In this case we write $X \sim \text{Geo}(p)$.

Application. The probability that the first success of a Bernoulli trial (see above) occurs on the x-th replication.

Scalar RVs – examples (discrete case)

Example (Poisson distribution). Let $\lambda > 0$ and the RV X take the values $x \in \{0, 1, \ldots\}$. The PMF of X is given by

$$p_X(x) = P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}.$$

In this case we write $X \sim \text{Poisson}(\lambda)$.

Application. Queueing theory, Poisson process (more later).

Scalar RVs – distribution (continuous case)

Definition 8 (probability functions of absolutely continuous scalar RV)

If X is an absolutely continuous RV then it has a probability density function (PDF) $f_X \in \mathbb{R}_{\geq 0}$ satisfying

$$P(a \leq X \leq b) \equiv P\big(\{\omega | a \leq X(\omega) \leq b\}\big) = \int_a^b f_X(x)dx$$

with the property

$$\int_{-\infty}^{\infty} f_X(x)dx = 1.$$

The cumulative distribution function (CDF) $F_X \in [0, 1]$ is defined as

$$F_X(b) = P(X \leq b) \equiv P\big(\{\omega | X(\omega) \leq b\}\big) = \int_{-\infty}^{b} f_X(x)dx.$$

Absolutely continuous means $\frac{d}{dx}F_X(x) = f_X(x)$ almost everywhere.

Scalar RVs – examples (continuous case)

Example (uniform distribution). The RV X has the uniform distribution on the interval $[a, b]$ if its PDF is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b], \\ 0 & \text{if } x \notin [a, b]. \end{cases}$$

In this case we write $X \sim U(a, b)$.

Application. Numerical integration (more later).

Scalar RVs – examples (continuous case)

Example (exponential distribution). The RV X has the exponential distribution with parameter $\lambda > 0$ if its PDF is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$$

In this case we write $X \sim \text{Exp}(\lambda)$.

Application. Waiting times, compound Poisson process (more later).

Scalar RVs – examples (continuous case)

Example (normal distribution). The RV X has the normal (Gaussian) distribution with parameters μ and $\sigma^2 > 0$ if its PDF is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

for all $x \in \mathbb{R}$.

In this case we write $X \sim N(\mu, \sigma^2)$.

Application. Central limit theorem, option pricing (more later).

Scalar RVs – examples (continuous case)

Example (gamma distribution). The RV X has gamma distribution with parameters $\alpha, \beta > 0$ if its PDF is given by

$$f(x) = \begin{cases} \frac{\beta^{-\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & \text{if } x > 0, \\ 0 & \text{if } x \leq 0 \end{cases}$$

where

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

In this case we write $X \sim \mathit{Gamma}(\alpha, \beta)$.

Application. Climatology, queuing models.

Scalar RVs – expectation of functions

Definition 9 (expectation of functions of scalar RV)

Let X be a scalar RV and $g : \mathbb{R} \to \mathbb{R}$ (or subsets of this mapping). Then

$$E[g(X)] = \int_\Omega g(X(\omega))dP(\omega) = \int_{-\infty}^{\infty} g(x)dF_X(x)$$
$$= \begin{cases} \sum_x g(x)p_X(x), & \text{if } X \text{ discrete} \\ \int_{-\infty}^{\infty} g(x)f_X(x)dx, & \text{if } X \text{ absolutely continuous} \end{cases}$$

where $g(x)$ is such that

$$E[|g(X)|] = \int_\Omega |g(X(\omega))|dP(\omega) = \int_{-\infty}^{\infty} |g(x)|dF_X(x)$$
$$= \begin{cases} \sum_x |g(x)|p_X(x) < \infty, \\ \int_{-\infty}^{\infty} |g(x)|f_X(x)dx < \infty. \end{cases}$$

If the last does not hold then $E[g(X)]$ does not exist.

Scalar RVs – expected value

The previous definition is used to obtain the **expected value** defined as

$$E[X] = \int_{\Omega} X(\omega) dP(\omega) = \int_{-\infty}^{\infty} x dF_X(x).$$

The following properties follow directly from the definition above.

Property 1.
Let X_1 and X_2 be integrable RVs. Then if

$$E[|c_1 X_1 + c_2 X_2|] < \infty$$

for any $c_1, c_2 \in \mathbb{R}$ then

$$E[c_1 X_1 + c_2 X_2] = c_1 E[X_1] + c_2 E[X_2].$$

Property 2.
If $E(|X|) < \infty$ then X is integrable and

$$|E[X]| \leq E[|X|].$$

Scalar RVs – MGF

Definition 10 (moment generating function – scalar RV)

The moment generating function (MGF) $M_X(u)$, $u \in R \subseteq \mathbb{R}$ (including 0), of a scalar RV X is defined as

$$M_X(u) = E[e^{uX}] = \int_{-\infty}^{\infty} e^{ux} dF_X(x)$$

under the assumption that this expectation exists.

Note that moments can be calculated as

$$E[X^k] = E[X^k e^{uX}]|_{u=0} = \frac{d^k}{du^k} E[e^{uX}]|_{u=0} = \frac{d^k}{du^k} M_X(u)|_{u=0}$$

provided the derivatives exist for all $u \in (u_0, u_1)$ where $u_0 < 0 < u_1$.

Scalar RVs – CF

Recall Euler's formula $e^{iux} = \cos(ux) + i\sin(ux)$.

Definition 11 (characteristic function – scalar RV)

The characteristic function (CF) $\varphi_X(u)$, $u \in \mathbb{R}$, of a RV variable X is defined as

$$\varphi_X(u) = E[e^{iuX}] = \int_{-\infty}^{\infty} e^{iux} dF_X(x)$$

which always exists.

Note that moments can be calculated as

$$E[X^k] = E[X^k e^{iuX}]|_{u=0} = \frac{1}{i^k}\frac{d^k}{du^k}E[e^{iuX}]|_{u=0} = \frac{1}{i^k}\frac{d^k}{du^k}\varphi_X(u)|_{u=0},$$

where the derivatives exist as long as $E[|X^k|] < \infty$.

Vector RVs

A **random vector** $\boldsymbol{X}$ on $(\Omega, \mathcal{F}, P)$ is denoted

$$\boldsymbol{X} := (X_1, \ldots, X_n)^T = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

and a **random matrix** $\boldsymbol{X}$ on $(\Omega, \mathcal{F}, P)$

$$\boldsymbol{X} := (X_{i,j})_{1 \leq i \leq m, 1 \leq j \leq n} = \begin{pmatrix} X_{1,1} & \cdots & X_{1,n} \\ \vdots & \ddots & \vdots \\ X_{m,1} & \cdots & X_{m,n} \end{pmatrix}.$$

Each component of $\boldsymbol{X}$ is a scalar RV as previously described.

The mapping and $\mathcal{F}$-measurability criteria is a generalisation of that for a scalar RV described in Definition 6 (details omitted).

Vector RVs – expected value

For random vectors and random matrices, expectation is taken component wise, i.e. for random vectors

$$E[\boldsymbol{X}] := \big(E[X_1], \ldots, E[X_n]\big)^T = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{pmatrix}$$

and for random matrices

$$E[\boldsymbol{X}] := \big(E[X_{i,j}]\big)_{1 \le i \le m, 1 \le j \le n} = \begin{pmatrix} E[X_{1,1}] & \cdots & E[X_{1,n}] \\ \vdots & \ddots & \vdots \\ E[X_{m,1}] & \cdots & E[X_{m,n}] \end{pmatrix}.$$

The component-wise expectation has been previously defined.

Vector RVs – joint distribution

Definition 12 (joint probability functions of vector RV)

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ be a vector RV. Then the joint CDF $F_{\boldsymbol{X}} \in [0, 1]$ is defined as

$$\begin{aligned} F_{\boldsymbol{X}}(\boldsymbol{b}) &= P(\boldsymbol{X} \leq \boldsymbol{b}) = P(X_1 \leq b_1, \ldots, X_n \leq b_n) \\ &= \sum_{x_1 \leq b_1} \cdots \sum_{x_n \leq b_n} p_{\boldsymbol{X}}(x_1, \ldots, x_n) \\ &\quad \text{if } \boldsymbol{X} \text{ discrete} \\ &= \int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_n} f_{\boldsymbol{X}}(x_1, \ldots, x_n) dx_n \cdots dx_1 \\ &\quad \text{if } \boldsymbol{X} \text{ absolutely continuous.} \end{aligned}$$

The functions $p_{\boldsymbol{X}} \in [0, 1]$ and $f_{\boldsymbol{X}} \in \mathbb{R}_{\geq 0}$ are called the joint PMF and PDF respectively.

The functions $p_{\boldsymbol{X}}$ and $f_{\boldsymbol{X}}$ sum and integrate to 1 respectively.

Vector RVs – marginal distributions

Joint probability functions can be used to obtain the marginal functions.

Corollary 1 (marginal probability functions of discrete vector RV)

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ be discrete vector RV with joint PMF $p_{\boldsymbol{X}}$ and joint CDF $F_{\boldsymbol{X}}$. The marginal CDF of X_1 can be obtained as

$$\begin{aligned} F_{X_1}(b) = P(X_1 \leq b) &= \lim_{x_2,\ldots,x_n \to \infty} F_{\boldsymbol{X}}(b, x_2, \ldots, x_n) \\ &= \sum_{x_1 \leq b} \sum_{x_2} \cdots \sum_{x_n} p_{\boldsymbol{X}}(x_1, x_2, \ldots, x_n) \\ &= \sum_{x_1 \leq b} p_{X_1}(x_1) \end{aligned}$$

where the marginal PMF of X_1 is

$$p_{X_1}(x_1) = \sum_{x_2} \cdots \sum_{x_n} p_{\boldsymbol{X}}(x_1, x_2, \ldots, x_n).$$

Vector RVs – marginal distributions

Now the absolutely continuous case.

Corollary 2 (marginal prob. funcs of absolutely continuous vector RV)

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ be an absolutely continuous vector RV with joint PDF $f_{\boldsymbol{X}}$ and joint CDF $F_{\boldsymbol{X}}$. The marginal CDF of X_1 can be obtained as

$$\begin{aligned} F_{X_1}(b) = P(X_1 \leq b) &= \lim_{x_2,\ldots,x_n \to \infty} F_{\boldsymbol{X}}(b, x_2, \ldots, x_n) \\ &= \int_{-\infty}^{b} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{X}}(x_1, x_2, \ldots, x_n) dx_n \cdots dx_2 dx_1 \\ &= \int_{-\infty}^{b} f_{X_1}(x_1) dx_1 \end{aligned}$$

where the marginal PDF of X_1

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{X}}(x_1, x_2, \ldots, x_n) dx_n \cdots dx_2.$$

Vector RVs – marginal distributions

The previous definition can be modified to obtain the marginal probability functions of the other components $X_2, \ldots X_n$ of the random vector $\boldsymbol{X}$.

Note how the marginals are obtained by summing/integrating over the components being excluded.

Generalising this procedure allows one to obtain the marginal joint probability functions of any subset of the components $X_1, \ldots X_n$ of the random vector $\boldsymbol{X}$.

For instance, the marginal joint PDF of components X_1, X_n of absolutely continuous RV $\boldsymbol{X}$ can be obtained as

$$f_{X_1,X_n}(x_1, x_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{X}}(x_1, x_2, \ldots, x_{n-1}, x_n) dx_{n-1} \cdots dx_2.$$

Vector RVs – expectation of functions

Definition 13 (expectation of functions of vector RV)

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ be a vector RV and $g : \mathbb{R}^n \to \mathbb{R}$ (or subsets of this mapping). Then

$$
\begin{aligned}
E[g(\boldsymbol{X})] &= \int_{\Omega} g(\boldsymbol{X}(\omega)) dP(\omega) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) dF_{\boldsymbol{X}}(x_1, \ldots, x_n) \\
&= \sum_{x_1} \cdots \sum_{x_n} g\left(x_1, \ldots, x_n\right) p_{\boldsymbol{X}}\left(x_1, \ldots, x_n\right) \\
&\quad \text{if } \boldsymbol{X} \text{ discrete} \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f_{\boldsymbol{X}}(x_1, \ldots, x_n) dx_n \cdots dx_1 \\
&\quad \text{if } \boldsymbol{X} \text{ absolutely continuous.}
\end{aligned}
$$

The expectation exists only if the sum/integral converges absolutely.

Vector RVs – MGF and CF

Definition 14 (moment generating function – vector RV)

The moment generating function (MGF) $M_{\boldsymbol{X}}(\boldsymbol{u})$, $\boldsymbol{u} \in R \subseteq \mathbb{R}^n$ (including $\mathbf{0}$), of a vector RV $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ is defined as

$$M_{\boldsymbol{X}}(\boldsymbol{u}) = E[e^{\boldsymbol{u} \cdot \boldsymbol{X}}]$$

under the assumption that this expectation exists.

Definition 15 (characteristic function – vector RV)

The characteristic function (CF) $\varphi_{\boldsymbol{X}}(\boldsymbol{u})$, $\boldsymbol{u} \in \mathbb{R}^n$, of a vector RV $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ is defined as

$$\varphi_{\boldsymbol{X}}(\boldsymbol{u}) = E[e^{i\boldsymbol{u} \cdot \boldsymbol{X}}]$$

which always exists.

There is a one-to-one correspondence between CFs (and MGFs where these exist) and distributions.

Vector RVs – covariance

Covariance is the expectation a particular function of random variables and is a measure of linear association between RVs.

Definition 16 (covariance matrix)

Let $\boldsymbol{X} = (X_1, \ldots, X_m)^T$, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ be vector RVs. Then

$$\begin{aligned}
\operatorname{cov}(\boldsymbol{X}, \boldsymbol{Y}) &= E\left[(\boldsymbol{X} - E[\boldsymbol{X}])(\boldsymbol{Y} - E[\boldsymbol{Y}])^T\right] \\
&= E[\boldsymbol{X}\boldsymbol{Y}^T] - E[\boldsymbol{X}]E[\boldsymbol{Y}]^T \\
&= \left(\operatorname{cov}(X_i, Y_j)\right)_{1 \le i \le m, 1 \le j \le n} \\
&= \begin{pmatrix} \operatorname{cov}(X_1, Y_1) & \cdots & \operatorname{cov}(X_1, Y_n) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(X_m, Y_1) & \cdots & \operatorname{cov}(X_m, Y_n) \end{pmatrix}.
\end{aligned}$$

Vector RVs – autocovariance

A particular application is the autocovariance of $\boldsymbol{X}$, defined as the square symmetric matrix

$$\begin{aligned}\operatorname{cov}(\boldsymbol{X},\boldsymbol{X}) &= E\left[(\boldsymbol{X}-E[\boldsymbol{X}])(\boldsymbol{X}-E[\boldsymbol{X}])^T\right] \\ &= \left(\operatorname{cov}(X_i,X_j)\right)_{1\le i,j\le n} \\ &= \begin{pmatrix} \operatorname{cov}(X_1,X_1) & \cdots & \operatorname{cov}(X_1,X_n) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(X_n,X_1) & \cdots & \operatorname{cov}(X_n,X_n) \end{pmatrix}.\end{aligned}$$

Note that since $\operatorname{cov}(X_i,X_j)=\operatorname{cov}(X_j,X_i)$ we have the symmetry property

$$\left(\operatorname{cov}(\boldsymbol{X},\boldsymbol{X})\right)^T = \operatorname{cov}(\boldsymbol{X},\boldsymbol{X}).$$

The diagonal components of $\operatorname{cov}(\boldsymbol{X},\boldsymbol{X})$ are the variances, e.g.

$$\operatorname{var}(X_1)=\operatorname{cov}(X_1,X_1)=E\left[(X_1-E[X_1])^2\right].$$

Vector RVs – autocovariance

Note that in addition to being symmetric, the autocovariance matrix $\mathrm{cov}(\boldsymbol{X}, \boldsymbol{X})$ is nonnegative-definite (or positive semi-definite).

To see this, observe for any $\boldsymbol{u} \in \mathbb{R}^n$ we have the quadratic form

$$\begin{aligned}
\boldsymbol{u} \cdot \mathrm{cov}(\boldsymbol{X}, \boldsymbol{X})\boldsymbol{u} &= \boldsymbol{u}^T E\big[(\boldsymbol{X} - E[\boldsymbol{X}])(\boldsymbol{X} - E[\boldsymbol{X}])^T\big]\boldsymbol{u} \\
&= E\big[\boldsymbol{u}^T(\boldsymbol{X} - E[\boldsymbol{X}])(\boldsymbol{X} - E[\boldsymbol{X}])^T\boldsymbol{u}\big] \\
&= E\big[((\boldsymbol{X} - E[\boldsymbol{X}])^T\boldsymbol{u})^T(\boldsymbol{X} - E[\boldsymbol{X}])^T\boldsymbol{u}\big] \\
&= E\big[((\boldsymbol{X} - E[\boldsymbol{X}])^T\boldsymbol{u}) \cdot ((\boldsymbol{X} - E[\boldsymbol{X}])^T\boldsymbol{u})\big] \\
&= E\big[|(\boldsymbol{X} - E[\boldsymbol{X}])^T\boldsymbol{u}|^2\big] \\
&\geq 0.
\end{aligned}$$

Independent RVs – joint distribution and expectation

Definition 17 (joint distribution of independent RVs)

Let X and Y be independent RVs. Then the joint CDF is obtained as

$$\begin{aligned} F_{X,Y}(a,b) &= P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b) \\ &= F_X(a)F_Y(b) \\ &= \sum_{x \leq a} p_X(x) \sum_{y \leq b} p_Y(y) \\ &\quad \text{if } X, Y \text{ discrete} \\ &= \int_{-\infty}^{a} f_X(x)dx \int_{-\infty}^{b} f_Y(y)dy \\ &\quad \text{if } X, Y \text{ absolutely continuous.} \end{aligned}$$

This leads to the following result.

Independent RVs – joint distribution and expectation

Corollary 3 (joint distribution and expectation of independent RVs)

Let X and Y be independent RVs.

If the RVs are discrete then the joint PMF is given by

$$p_{X,Y}(x, y) = p_X(x)p_Y(y).$$

If the RVs are absolutely continuous then the joint PDF is given by

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

Moreover, for functions $g, h : \mathbb{R} \to \mathbb{R}$ (or subsets of this mapping) the expectation of the product

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

as long as the RHS expectations are properly defined.

These results can be extended to larger collections of independent RVs and also to independent vectors RVs.

Conditional expectation

In applications of probability and stochastic processes conditional expectation plays important roles. First the discrete scalar case.

Definition 18 (conditional expectation – discrete RVs)

If X and Y are discrete RVs we define the conditional PMF of X given $Y = y$ with $p_Y(y) > 0$ by

$$p_{X|Y}(x, y) = P(X = x|Y = y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}.$$

The conditional expectation of X given $Y = y$ with $p_Y(y) > 0$ is defined by

$$E[X|Y = y] = \sum_x x p_{X|Y}(x, y).$$

The expectation is properly defined only if the sum converges absolutely.

Conditional expectation

We also have the continuous scalar case.

Definition 19 (conditional expectation – absolutely continuous RVs)

If $(X, Y)^T$ is absolutely-continuous we define the conditional PDF of X given $Y = y$ with $f_Y(y) > 0$ by

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

The conditional expectation of X given that $Y = y$ with $f_Y(y) > 0$ is defined by

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x, y) dx.$$

The expectation is properly defined only if integral converges absolutely.

Generalises to functions $g(X)$ and different conditioning events.

Conditional expectation

Note that in the previous two definitions the conditional expectation is non-random as the conditioning variable has been fixed.

If the conditioning variable is not fixed then the conditional expectation is a random function of the conditioning variable.

Definition 20 (conditional expectation as random function)

Let X and Y be scalar RVs. The RV

$$g(Y) = E[X|Y]$$

is said to be the conditional expectation X given Y.

Discussion of what is meant by conditioning on Y rather than some event $\{Y = y\}$ is deferred until necessary.

These last three definitions can be extended to the case where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are vector RVs.

Conditional expectation

Conditional expectation can be technical to deal with, so where possible calculations are simplified using the results in the following proposition.

Proposition 1 (properties of conditional expectation)

Under the condition of the existence of the expectations

1. $E[c_1\boldsymbol{X}_1 + c_2\boldsymbol{X}_2|\boldsymbol{Y}] = c_1E[\boldsymbol{X}_1|\boldsymbol{Y}] + c_2E[\boldsymbol{X}_2|\boldsymbol{Y}]$,
2. $E\big[E[\boldsymbol{X}|\boldsymbol{Y}]\big] = E[\boldsymbol{X}]$,
3. $E[g(\boldsymbol{Y})\boldsymbol{X}|\boldsymbol{Y}] = g(\boldsymbol{Y})E[\boldsymbol{X}|\boldsymbol{Y}]$,
4. $E[\boldsymbol{X}|\boldsymbol{Y}] = E[\boldsymbol{X}]$ if $\boldsymbol{X}, \boldsymbol{Y}$ are independent.

Note that Parts 2 and 3 imply that for any bounded function $g(\boldsymbol{y})$

$$E[g(\boldsymbol{Y})\boldsymbol{X}] = E[E[g(\boldsymbol{Y})\boldsymbol{X}|\boldsymbol{Y}]] = E\big[g(\boldsymbol{Y})E[\boldsymbol{X}|\boldsymbol{Y}]\big].$$

The second property above is known as the "law of iterated conditioning" or the "tower law".

Conditional expectation

Conditional expectation possesses nice properties, as the next theorem demonstrates.

As stated, the theorem is for scalar RVs X and Y but generalises to the vector case.

Theorem 2 (cond. expectation as best-mean-square predictor)

Let $E[X^2] < \infty$ and Y be any random vector. The conditional expectation

$$Z := E[X|Y],$$

as a function of Y, minimises the distance $E[(X - g(Y))^2]$ amongst all functions $g(Y)$ such that $E[g^2(Y)] < \infty$.

Interpretation.
$E[X|Y]$ is the best predictor of X based on Y in a mean-square sense.

Conditional expectation

Proof.
Let $g(y)$ be an arbitrary function such that $E[g^2(Y)] < \infty$.

Then

$$\begin{aligned} E[(X - g(Y))^2] &= E[(X - Z + Z - g(Y))^2] \\ &= E[(X - Z)^2] + 2E[(X - Z)(Z - g(Y))] + E[(Z - g(Y))^2]. \end{aligned}$$

By properties of conditional expectation

$$\begin{aligned} &E[(X - Z)(Z - g(Y))] \\ &\quad = E\big[E[(X - Z)(Z - g(Y))|Y]\big] \quad \text{(by Part 2 of Prop. 1)} \\ &\quad = E\big[(Z - g(Y))E[X - Z|Y]\big] \quad \text{(by Part 3 of Prop. 1).} \end{aligned}$$

Conditional expectation

But

$$\begin{aligned} E[X - Z|Y] &= E\big[X - E[X|Y]|Y\big] \quad (Z = E[X|Y]) \\ &= E[X|Y] - E\big[E[X|Y]|Y\big] \\ &= E[X|Y] - E[X|Y] \quad (E[X|Y] \text{ is a function of } Y) \\ &= 0 \end{aligned}$$

hence

$$E[(X - g(Y))^2] = E[(X - Z)^2] + E[(Z - g(Y))^2].$$

So it is clear that the minimum of $E[(X - g(Y))^2]$ is achieved when $g(Y) = Z = E[X|Y]$!

Conditional expectation

Exercise.
Let $X = cY + Z$ where the RVs Y and Z are independent and $c \in \mathbb{R}$.

Show that

$$E[X|Y] = cY + E[Z].$$

Solution. Using Proposition 1 we have

$$\begin{aligned} E[X|Y] &= E\big[cY + Z|Y\big] \\ &= cE[Y|Y] + E[Z|Y] \quad \text{(by Prop. 1 Part 1)} \\ &= cYE[1|Y] + E[Z|Y] \quad \text{(by Prop. 1 Part 3)} \\ &= cYE[1|Y] + E[Z] \quad \text{(by Prop. 1 Part 4)} \\ &= cY + E[Z]. \end{aligned}$$

As we will see in Chapter 2, for normal (Gaussian) RVs there is a well developed theory of conditional expectation.

References I


Borokov, K. (2014).
Elements of Stochastic Modelling.
World Scientific, ?, 2nd edition.


Chung, K. (2001).
A Course in Probability Theory.
Academic Press, San Diego, CA, 3rd edition.


Cont, R. and Tankov, P. (2004).
Financial Modelling with Jump Processes.
Chapman and Hall, Boca Raton, FL.


Platen, E. and Heath, D. (2006).
A Benchmark Approach to Quantitative Finance.
Springer-Verlag, Berlin Heidelberg.


Rice, J. (2007).
Mathematical Statistics and Data Analysis.
Thomson Brooks/Cole, Belmont, CA, 3rd edition.

References II


Rudin, W. (1976).
Principles of Mathematical Analysis.
McGraw-Hill, New York City, NY, 3rd edition.


Rudin, W. (1987).
Real and Complex Analysis.
McGraw-Hill, New York City, NY, 3rd edition.


Shreve, S. (2004).
Stochastic Calculus for Finance II.
Springer, New York City, NY.


Shreve, S. (2005).
Stochastic Calculus for Finance I.
Springer, New York City, NY.


Wackerly, D., Mendenhall, W., and Scheaffer, R. (2008).
Mathematical Statistics with Applications.
Thomson Brooks/Cole, Belmont, CA, 7th edition.