Stochastic Process

Introduction and review of probability

Lecturer: Xiao Ma

Office: 631G, Tel: 3994 3112

maxiao@mail.sysu.edu.cn

School of Data and Computer Science
Sun Yat-sen University

GraduateCourse, Autumn2018

Chapter 1. Introduction and review of probability

§ 1.1 Probability models

1. Introduction
2. The sample space of a probability model
3. Assigning probabilities for finite sample spaces

1. Introduction

- Probability model

  ◇ *sample space*: a *set* whose elements are called *sample points* or *outcomes*.

  ◇ a class of *events*: the class of all subsets of the sample space.

  ◇ *probability measure*: the assignment of a non-negative number to each outcome, with the restriction that these numbers must sum to 1 over the sample space.

1. Introduction

Example

The standard probability model for rolling a die uses $\{1, 2, 3, 4, 5, 6\}$ *as the sample space, with each possible outcome having probability* $1/6$*. An odd result, i.e., the subset* $\{1, 3, 5\}$*, is an example of an event in this sample space, and this event has probability* $1/2$*.*

2. The sample space of a probability model

An *outcome* or *sample point* in a probability model corresponds to a complete result (with all detail specified) of the experiment being modeled.

2. The sample space of a probability model

An *outcome* or *sample point* in a probability model corresponds to a complete result (with all detail specified) of the experiment being modeled.

An outcome $\{\omega\}$ is often called a *finest grain* result of the model in the sense that a singleton event $\{\omega\}$ containing only $\{\omega\}$ clearly contains no proper subsets. Thus events (other than singleton events) typically give only partial information about the result of the experiment, whereas an outcome fully specifies the result.

2. The sample space of a probability model

An *outcome* or *sample point* in a probability model corresponds to a complete result (with all detail specified) of the experiment being modeled.

An outcome $\{\omega\}$ is often called a *finest grain* result of the model in the sense that a singleton event $\{\omega\}$ containing only $\{\omega\}$ clearly contains no proper subsets. Thus events (other than singleton events) typically give only partial information about the result of the experiment, whereas an outcome fully specifies the result.

In choosing the sample space for a probability model of an experiment, we often *omit details* that appear irrelevant for the purpose at hand.

3. Assigning probabilities for finite sample spaces

The word *probability* is widely used in everyday language, and most of us attach various intuitive meanings to the word.

3. Assigning probabilities for finite sample spaces

The word *probability* is widely used in everyday language, and most of us attach various intuitive meanings to the word.

Between virtual impossibility and certainty, if one outcome appears to be closer to certainty than another, its probability should be correspondingly greater.

3. Assigning probabilities for finite sample spaces

Symmetry can often provide a better rationale for choosing probabilities. For example, the symmetry between H and T for a coin, or the symmetry between the six faces of a die, motivates assigning equal probabilities.

3. Assigning probabilities for finite sample spaces

Symmetry can often provide a better rationale for choosing probabilities. For example, the symmetry between H and T for a coin, or the symmetry between the six faces of a die, motivates assigning equal probabilities.

Another approach is to perform the experiment many times and choose the probability of each outcome as the *relative frequency* of that outcome (i.e., the number of occurrences of that outcome divided by the total number of trials).

Chapter 1. Introduction and review of probability

§ 1.2 The axioms of probability theory

1. Axioms for events
2. Axioms of probability

1. Axioms for events

- Given a sample space Ω, the class of subsets of Ω that constitute the set of events satisfies the following axioms:
  - ◇ Ω is an event.
  - ◇ For every sequence of events $A_1, A_2, \ldots$, the union $\bigcup_{n=1}^{\infty} A_n$ is an event.
  - ◇ For every event A, the complement A^c is an event.

1. Axioms for events

Remark. Not all subsets need be events. Usually each sample point is taken to be a singleton event.

1. Axioms for events

Remark. Not all subsets need be events. Usually each sample point is taken to be a singleton event.

The empty set $\varnothing$ is Ω^c, so is an event. If all sample points are singleton events, then all finite and countable sets are events (i.e., they are finite and countable unions of singleton sets).

1. Axioms for events

Remark. Not all subsets need be events. Usually each sample point is taken to be a singleton event.

The empty set $\varnothing$ is Ω^c, so is an event. If all sample points are singleton events, then all finite and countable sets are events (i.e., they are finite and countable unions of singleton sets).

From de Morgan's law,

$$\left[\bigcup_{i=1}^{\infty} A_i\right]^c = \bigcap_{i=1}^{\infty} A_i^c.$$

so countable intersections of events are events. All combinations of intersections and unions of events are also events.

2. Axioms of probability

- Given any sample space Ω and any class of events $\mathcal{E}$ satisfying the axioms of events, a probability rule is a function $\Pr\{\cdot\}$ mapping each $A \in \mathcal{E}$ to a (finite) real number in such a way that the following three probability axioms hold

2. Axioms of probability

- Given any sample space Ω and any class of events $\mathcal{E}$ satisfying the axioms of events, a probability rule is a function $\Pr\{\cdot\}$ mapping each $A \in \mathcal{E}$ to a (finite) real number in such a way that the following three probability axioms hold
  - $\diamond$ $\Pr\{\Omega\} = 1$.

2. Axioms of probability

- Given any sample space Ω and any class of events $\mathcal{E}$ satisfying the axioms of events, a probability rule is a function $\Pr\{\cdot\}$ mapping each $A \in \mathcal{E}$ to a (finite) real number in such a way that the following three probability axioms hold
  - ◇ $\Pr\{\Omega\} = 1$.
  - ◇ For every event A, $\Pr\{A\} \geq 0$.

2. Axioms of probability

- Given any sample space Ω and any class of events $\mathcal{E}$ satisfying the axioms of events, a probability rule is a function $\Pr\{\cdot\}$ mapping each $A \in \mathcal{E}$ to a (finite) real number in such a way that the following three probability axioms hold
  - ◇ $\Pr\{\Omega\} = 1$.
  - ◇ For every event A, $\Pr\{A\} \geq 0$.
  - ◇ The probability of the union of any sequence $A_1, A_2, \ldots$ of disjoint events is given by

$$\Pr\{\bigcup_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} \Pr\{A_n\}\}$$

where $\sum_{n=1}^{\infty} \Pr\{A_n\}$ is shorthand for $\lim_{n\to\infty} \sum_{n=1}^{\infty} \Pr\{A_n\}$.

2. Axioms of probability

- It is surprising that this is all that is needed to avoid paradoxes. A few simple consequences are:

2. Axioms of probability

- It is surprising that this is all that is needed to avoid paradoxes. A few simple consequences are:
  - ◇ $\Pr\{\varnothing\} = 0$.

2. Axioms of probability

- It is surprising that this is all that is needed to avoid paradoxes. A few simple consequences are:
  - ◇ $\Pr\{\varnothing\} = 0$.
  - ◇ $\Pr\{A^c\} = 1 - \Pr\{A\}$.

2. Axioms of probability

- It is surprising that this is all that is needed to avoid paradoxes. A few simple consequences are:
  - ◇ $\Pr\{\varnothing\} = 0$.
  - ◇ $\Pr\{A^c\} = 1 - \Pr\{A\}$.
  - ◇ $\Pr\{A\} \leq Pr\{B\} \leq 1$, for $A \subseteq B$.

2. Axioms of probability

- It is surprising that this is all that is needed to avoid paradoxes. A few simple consequences are:
  - ◇ $\Pr\{\varnothing\} = 0$.
  - ◇ $\Pr\{A^c\} = 1 - \Pr\{A\}$.
  - ◇ $\Pr\{A\} \leq Pr\{B\} \leq 1$, for $A \subseteq B$.
  - ◇ $\sum_n \Pr\{A_n\} \leq 1$, for $A_1, A_2, ...$disjoint.

2. Axioms of probability

- It is surprising that this is all that is needed to avoid paradoxes. A few simple consequences are:
  - ◇ $\Pr\{\varnothing\} = 0$.
  - ◇ $\Pr\{A^c\} = 1 - \Pr\{A\}$.
  - ◇ $\Pr\{A\} \leq Pr\{B\} \leq 1$, for $A \subseteq B$.
  - ◇ $\sum_n \Pr\{A_n\} \leq 1$, for $A_1, A_2, \ldots$disjoint.
  - ◇ $\Pr\{\bigcup_n A_n\} \leq \sum_n \Pr\{A_n\}$. (union bound)
  - ◇ ...

Chapter 1. Introduction and review of probability

§ 1.3 Probability review

1. Conditional probabilities and statistical independence
2. Repeated idealized experiments
3. Random variables
4. Multiple random variables and conditional probabilities

1. Conditional probabilities and statistical independence

Definition

For any two events A and B in a probability model, the **conditional probability** *of A, conditional on B, is defined if* $Pr\{B\} > 0$ *by*

$$Pr\{A|B\} = Pr\{AB\}/Pr\{B\}$$

Remark

◇ *Bayes' law:*

$$Pr\{A|B\}Pr\{B\} = Pr\{B|A\}Pr\{A\}$$

1. Conditional probabilities and statistical independence

Example

Drug Test: *Suppose a drug test is* 99% *sensitive and* 99%*specific. That is, the test will produce* 99% *true positive results for drug users and* 99% *true negative results for non-drug users. Suppose that* 0.5% *of people are users of the drug. What is the probability that a randomly selected individual with a positive test is a user?*

1. Conditional probabilities and statistical independence

Example

$$
\begin{aligned}
&\Pr(\text{User} \mid +) \\
=&\frac{\Pr(+ \mid \text{User})\Pr(\text{User})}{\Pr(+)} \\
=&\frac{\Pr(+ \mid \text{User})\Pr(\text{User})}{\Pr(+ \mid \text{User})\Pr(\text{User}) + \Pr(+ \mid \text{Non} - \text{User})\Pr(\text{Non} - \text{User})} \\
=&\frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.01 \times 0.995} \\
\approx&33.2\%
\end{aligned}
$$

1. Conditional probabilities and statistical independence

Definition

Two events, A and B, are **statistically independent** *(or, more briefly, independent) if*

$$Pr\{AB\} = Pr\{A\}Pr\{B\}$$

Remark

◇ *For $Pr\{B\} > 0$, this is equivalent to $Pr\{A|B\} = Pr\{A\}$.*

1. Conditional probabilities and statistical independence

Definition

The events $A_1, ..., A_n, n > 2$ *are* **statistically independent** *if for each collection S of two or more of the integers 1 to n.*

$$Pr\{\bigcap_{i \in S} A_i\} = \prod_{i \in S} Pr\{A_i\}$$

2. Repeated idealized experiments

Much of our intuitive understanding of probability comes from the notion of repeating the same idealized experiment many times.

2. Repeated idealized experiments

Much of our intuitive understanding of probability comes from the notion of repeating the same idealized experiment many times.

Given an original sample space Ω, the sample space of an n-repetition model is the Cartesian product:

$$\Omega^n = \{(\omega_1, \omega_2, ..., \omega_n) : \omega_i \in \Omega \text{ for each } i, 1 \le i \le n\},$$

2. Repeated idealized experiments

Much of our intuitive understanding of probability comes from the notion of repeating the same idealized experiment many times.

Given an original sample space Ω, the sample space of an n-repetition model is the Cartesian product:

$$\Omega^n = \{(\omega_1, \omega_2, ..., \omega_n) : \omega_i \in \Omega \text{ for each } i, 1 \leq i \leq n\},$$

For each extended event $\{A_1 \times A_2 \times \cdots \times A_n\}$ contained in Ω^n:

$$\Pr\{A_1 \times A_2 \times \cdots \times A_n\} = \prod_{i=1}^{n} \Pr\{A_i\},$$

3. Random variables

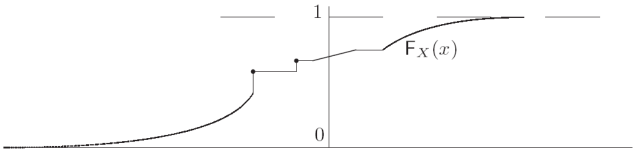
Definition

A **random variable** *(rv) is essentially a function X from the sample space Ω of a probability model to the set of real numbers $\mathbb{R}$. Three modifications are needed to make this precise. First, X might be undefined or infinite for a subset of Ω that has 0 probability. Second, the mapping $X(\omega)$ must have the property that $\{\omega \in \Omega : X(\omega) \leq x\}$ is an event for each $x \in \mathbb{R}$. Third, every finite set of rv s $X_1, \cdots, X_n$ has the property that for each $x_1 \in \mathbb{R}, ..., x_n \in \mathbb{R}$, the set $\{\omega : X_1(\omega) \leq x_1, \cdots, X_n(\omega) \leq x_n\}$ is an event.*

3. Random variables

Definition

The **cumulative distribution function** *(CDF) of a rv X is a function $F_X(x)$ mapping each $x \in \mathbb{R}$ into $F_X(x) = Pr\{\omega \in \Omega : X(\omega) \leq x\}$. The argument ω is usually omitted for brevity, so $F_X(x) = Pr\{X \leq x\}$.*



3. Random variables

Remark

◇ *If X has only a finite or countable number of possible sample values, say $x_1, x_2, \cdots$,the probability $Pr\{X = x_i\}$ of each sample value x_i is called the* **probability mass function** *(PMF) at x_i and denoted by $p_X(x_i)$; such a rv is called discrete.*

3. Random variables

Remark

◇ *If* X *has only a finite or countable number of possible sample values, say* $x_1, x_2, \cdots$ *,the probability* $Pr\{X = x_i\}$ *of each sample value* x_i *is called the* **probability mass function** *(PMF) at* x_i *and denoted by* $p_X(x_i)$*; such a rv is called discrete.*

◇ *If the CDF* $F_X(x)$ *of a rv* X *has a (finite) derivative at* x*, the derivative is called the density, or more precisely the* **probability density function** *(PDF) of* X *at* x *and denoted by* $f_X(x)$*.*

4. Multiple random variables and conditional probabilities

- **joint CDF**: If $X_1, X_2, \ldots, X_n$ are rv.s or the components of a vector rv, their joint CDF is defined by

$$
\begin{aligned}
& F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) \\
=& \Pr\{\omega \in \Omega : X_1(\omega) \leqslant x_1, X_2(\omega) \leqslant x_2, \ldots, X_n(\omega) \leqslant x_n\}.
\end{aligned}
$$

The r.v.s $X_1, \ldots, X_n$ are independent if

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

4. Multiple random variables and conditional probabilities

- **marginal CDF**: A marginal CDF if a given joint CDF is given by

$$F_{X_i}(x_i) = F_{X_1,\ldots,X_{i-1}X_iX_{i+1}\ldots X_n}(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty)$$

4. Multiple random variables and conditional probabilities

- **marginal CDF**: A marginal CDF if a given joint CDF is given by

$$F_{X_i}(x_i) = F_{X_1,\ldots,X_{i-1}X_iX_{i+1}\ldots X_n}(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty)$$

- **joint PMF**: If the rv s are all discrete, there is a joint PMF given by

$$p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \Pr\{X_1 = x_1, \ldots, X_n = x_n\}.$$

4. Multiple random variables and conditional probabilities

- **marginal CDF**: A marginal CDF if a given joint CDF is given by

$$F_{X_i}(x_i) = F_{X_1,...,X_{i-1}X_iX_{i+1}...X_n}(\infty, ..., \infty, x_i, \infty, ..., \infty)$$

- **joint PMF**: If the rv s are all discrete, there is a joint PMF given by

$$p_{X_1,..,X_n}(x_1, ..., x_n) = \Pr\{X_1 = x_1, ..., X_n = x_n\}.$$

- **joint PDF**: If the joint CDF can be differentiated , then the joint PDF is given by,

$$f_{X_1,..,X_n}(x_1, ..., x_n) = \frac{\partial^n F(x_1, ..., x_n)}{\partial x_1 \partial x_2 ... \partial x_n}$$

4. Multiple random variables and conditional probabilities

- Two rv s, say X and Y, are **statistically independent** (or, more briefly, independent) if

$$F_{XY}(x,y) = F_X(x)F_Y(y) \text{ for each } x \in \mathbb{R}, y \in \mathbb{R}.$$

4. Multiple random variables and conditional probabilities

- Two rv s, say X and Y, are **statistically independent** (or, more briefly, independent) if

$$F_{XY}(x,y) = F_X(x)F_Y(y) \text{ for each } x \in \mathbb{R}, y \in \mathbb{R}.$$

Remark

◇ *If X and Y are discrete rv s, then the definition of independence is equivalent to the corresponding statement for PMFs*

$$p_{XY}(x_i, y_i) = F_X(x_i)F_Y(y_i) \;\; \textit{for each value } x_i \textit{ of } X \textit{and } y_i \textit{ of } \; Y.$$

4. Multiple random variables and conditional probabilities

Remark

◇ *If X and Y have a joint density, then the definition of independence is equivalent to the corresponding statement for PDFs*

$$f_{XY}(x,y) = F_X(x)F_Y(y) \quad \text{for each } x \in \mathbb{R}, y \in \mathbb{R}.$$

4. Multiple random variables and conditional probabilities

- **Independent identically distributed (IID)** case:
  - $X_1, ..., X_n$ are independent.
  - They all have the same PDF.
  - $F_X(x_1, ..., x_n) = \prod_{i=1}^{n} F_X(x_i)$.

Assignment: Exercises after 1.1, No. 7(a)(b), 9(c)(d), 27(a)(b)(c).

Thank you for your attention!