Mathematical Probability and Statistics Supplementary Notes

Fall 2018

David Lyons
Mathematical Sciences
Lebanon Valley College

Mathematical Probability and Statistics
Supplementary Notes

Fall 2018

David Lyons
Mathematical Sciences
Lebanon Valley College
Copyright ©2018

Contents

1 Probability

1.1 Algebras of Sets

(1.1.1) **Some notation from set theory.** Given a set S, we write $\mathcal{P}(S)$ to denote the ***power set*** of S, that is, the set of all subsets of S.

$$\mathcal{P}(S) = \{A \colon A \subseteq S\}$$

Given two sets A, B, we write $A \setminus B$ to denote the set

$$A \setminus B = \{x \in A \colon x \notin B\}.$$

In a context in which a set A is understood to be an element of a power set $\mathcal{P}(\Omega)$ of a particular set Ω, we write A^c to denote the set $\Omega \setminus A$, also called the ***complement*** of A relative to Ω.

(1.1.2) **Definition.** Let Ω be a set and let $\mathcal{A}$ be a nonempty subset of the power set of Ω, that is, the elements of $\mathcal{A}$ are subsets of Ω. We say $\mathcal{A}$ is an ***algebra of subsets of*** Ω if the following hold.

1. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
2. $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$

(1.1.3) **Examples.** *Examples of set algebras.*

(1.1.4) **Proposition.** *Let $\mathcal{A}$ be an algebra of subsets of a set Ω. Then we have*

1. $\emptyset, \Omega \in \mathcal{A}$, *and*
2. $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$.

Proof.

(1.1.5) **Definition.** Let Ω be a set and let $\mathcal{A}$ be a nonempty subset of the power set of Ω, that is, the elements of $\mathcal{A}$ are subsets of Ω. We say $\mathcal{A}$ is a ***σ-algebra*** (pronounced "sigma algebra") of subsets of Ω if the following hold.

1. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
2. $A_n \in \mathcal{A}$ for all $n \in \mathbf{N} \Rightarrow \displaystyle\bigcup_{n \in \mathbf{N}} A_n \in \mathcal{A}$

(Condition 2 can be expressed by saying $\mathcal{A}$ is "closed under countable unions".)

(1.1.6) **Propolsition.** *All σ-algebras are set algebras. Not all set algebras are σ-alebras.*

(1.1.7) **Proposition.** *Let $\mathcal{A}$ be a σ-algebra of subsets of a set Ω. Then we have*

$$\{A_n \colon n \in \mathbf{N}\} \in \mathcal{A} \Rightarrow \bigcap_{n \in \mathbf{N}} A_n \in \mathcal{A},$$

that is, σ-algebras are closed under countable intersections.

(1.1.8) **Examples.** *Examples of σ-algebras: the whole power set, Borel sets of the real line, several non-examples.*

1.2 Probability Spaces and Probability Measures

(1.2.1) **Definition.** A ***probability space*** or ***probability measure*** is a triple $(\Omega, \mathcal{A}, P)$, where Ω is a set, $\mathcal{A}$ is a σ-algebra of subsets of Ω, and $P\colon \mathcal{A} \to [0,1]$ satisfies the following.

1. $P(\Omega) = 1$
2. for every countable collection $A_1, A_2, \ldots$ of disjoint elements of $\mathcal{A}$ we have

$$P\left(\bigcup_{n\in\mathbf{N}} A_n\right) = \sum_{n\in\mathbf{N}} P(A_n)$$

(Here, the term ***disjoint*** means $A_i \cap A_j = \emptyset$ for all i, j in $\mathbf{N}$.)

(1.2.2) **Terminology and conventions.** It is common to use the phrase "the probability space Ω" or "the probability measure P" to refer to the triple $(\Omega, \mathcal{A}, P)$. Another term for Ω is ***sample space***. Elements of Ω are called ***outcomes*** and elements of $\mathcal{A}$ are called ***events***. If Ω is countable (that is, finite or countably infinite), it is common to assume that the σ-algebra of events is the entire power set $\mathcal{P}(\Omega)$ of Ω. When $\mathcal{A} = \mathcal{P}(\Omega)$, then $\mathcal{A}$ is called the ***discrete*** algebra of events and Ω is called a ***discrete*** sample space. For an event A, the symbols $P(A)$ are pronounced "the probability of A". Given an element $\omega \in \Omega$, we usually write $P(\omega)$ instead of the more cumbersome $P(\{\omega\})$. Disjoint events (that is, events A, B such that $A \cap B = \emptyset$) are called ***mutually exclusive***. The complement A^c of an event A is also called the ***opposite*** of A.

(1.2.3) **Remark.** It is a remarkable fact that when Ω is uncountable, it is impossible to satisfy the axioms of a probability space for the discrete σ-algebra $\mathcal{A} = \mathcal{P}(\Omega)$.

(1.2.4) **Propositions.**

1. $P(\emptyset) = 0$
2. $A \subseteq B \Rightarrow P(A) \leq P(B)$
3. the addition rule
4. the opposite rule

(1.2.5) **Proposition (construction of discrete sample spaces).** Let Ω be a countable set and let $p\colon \Omega \to [0,1]$ satisfy the condition that $\sum_{\omega\in\Omega} p(\omega) = 1$. Let $P\colon \mathcal{P}(\Omega) \to [0,1]$ be given by

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

for $A \in \mathcal{P}(\Omega)$. Then the triple $(\Omega, \mathcal{P}(\Omega), P)$ is a (discrete) probability space.

(1.2.6) **Comment.** There is an issue concerning well-definedness of the countable sums $\sum_{\omega\in\Omega} p(\omega)$ and $\sum_{\omega\in A} p(\omega)$ in the proposition above. In general, the value of a series depends on the ordering of the summands. However, the series in question are absolutely convergent, so the value is independent of any ordering of the summands.

(1.2.7) **Exercises.**

1. explore numerous discrete and continuous examples
2. show that axiom 2 of the definition of probability measure implies a finite sum version, show that a finite sum version does *not* imply axiom 2
3. show that a version of axiom 2 for only two sets implies a general finite sum version
4. show that the set of points $\omega \in \Omega$ such that $P(\omega) \neq 0$ is countable
5. prove the properties above

1.3 Independence and conditional probability

(1.3.1) **Definitions.** Events A, B are called ***independent*** if $P(A \cap B) = P(A)P(B)$, and otherwise A, B are said to be ***dependent***. If $P(B) \neq 0$, the ***conditional probability of*** A ***given*** B, denoted $P(A|B)$, is defined to be

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

The rearrangement $P(A \cap B) = P(A|B)P(B)$ is sometimes called the ***multiplication rule***.

(1.3.2) **Propositions.**

1. Let $(\Omega, \mathcal{A}, P)$ be a probability space, let B be an element of $\mathcal{A}$ with $P(B) > 0$, and let $P_B : \mathcal{A} \to \mathbf{R}$ be defined by $P_B(A) = P(A|B)$ for all $A \in \mathcal{A}$. Then $(\Omega, \mathcal{A}, P_B)$ is a probability space.
2. if $P(B) \neq 0$ then A, B are independent if and only if $P(A|B) = P(A)$.
3. Bayes' theorem

(1.3.3) **Exercises.**

1. explore more examples and nonexamples
2. prove the propositions above

2 Random Variables

2.1 Random variables: definitions and properties

(2.1.1) **Definition.** A ***random variable*** is a function $X\colon \Omega \to \mathbf{R}$, where $(\Omega, \mathcal{A}, P)$ is a probability space, such that $X^{-1}(B)$ is an element of $\mathcal{A}$ for all sets B in the Borel σ-algebra of subsets of the real numbers.

(2.1.2) **Notation and conventions.** Given a subset S of the real line, we write the symbols '$X \in S$' to denote the set $X^{-1}(S) = \{\omega \in \Omega\colon X(\omega) \in S\}$. In the special case that S is an interval of the form $S = (-\infty, x]$, we write '$X \le x$' to denote the event $X \in S$. Similarly, we write '$a < X \le b$' to denote $X \in (a, b] = \{\omega \in \Omega\colon a < X(\omega) \le b\}$, and we write '$X = a$' to denote $X \in \{a\} = \{\omega \in \Omega\colon X(\omega) = a\}$.

(2.1.3) **Definition.** The ***(cumulative) distribution function***, or ***cdf***, of the random variable X is the function $F_X\colon \mathbf{R} \to \mathbf{R}$ given by

$$F_X(x) = P(X \le x).$$

(2.1.4) **Proposition.** A function $F\colon \mathbf{R} \to \mathbf{R}$ is the cdf of a random variable X if and only if the following properties hold.

1. $\lim\limits_{x\to-\infty} F(x) = 0$ and $\lim\limits_{x\to\infty} F(x) = 1$
2. F is nondecreasing, that is $a \le b \Rightarrow F(a) \le F(b)$ for all $a, b \in \mathbf{R}$
3. F is ***continuous from the right***, that is, $\lim\limits_{x\to a^+} F(x)$ exists and equals $F(a)$ for all $a \in \mathbf{R}$

(2.1.5) **Examples.** finite, countable, uncountable examples

(2.1.6) **Definitions.** Let X be a random variable and let S be the set

$$S = \{x \in \mathbf{R}\colon P(X = x) > 0\}.$$

The random variable X is called ***discrete***, ***continuous***, or ***mixed***, according to whether $P(X \in S) = 1$, $P(X \in S) = 0$, or $0 < P(X \in S) < 1$, respectively. A continuous random variable X is called ***absolutely continuous*** if its distribution function F_X is differentiable at all but countably many points.

(2.1.7) **Observations.**

1. If Ω is countable, then any random variable on Ω is discrete.
2. The set S is countable.

(2.1.8) **Definitions.** For a discrete random variable X, the ***probability function for*** X, also called the ***pf for*** X, is the function $p_X\colon \mathbf{R} \to \mathbf{R}$ given by

$$p_X(x) = P(X = x).$$

For an absolutely continuous random variable X, the ***probability density function for*** X, also called the ***pdf for*** X, is $f_X = F_X'$.

(2.1.9) **Examples.** pfs, pdfs, a mixed example

(2.1.10) **Proposition (Characterization of probability functions).** A function $p\colon \mathbf{R} \to \mathbf{R}$ is the probability function for a discrete random variable X if and only if the following properties hold.

- p is nonnegative
- the set $S = \{x\colon p(x) > 0\}$ where p is positive is countable
- $\displaystyle\sum_{x\in S} p(x) = 1$

(2.1.11) **Remark.** There is a subtle issue involved in writing the expression $\sum_{x\in S} p(x) = 1$. The notation implies that the sum is abolutely convergent, and therefore well-defined, independent of any ordering of S.

(2.1.12) **Proposition (Characterization of probability density functions).** A function $f\colon \mathbf{R} \to \mathbf{R}$ is the probability density function for an absolutely continuous random variable X if and only if the following properties hold.

- f is nonnegative
- f is piecewise continuous with at most countably many discontinuities
- $\displaystyle\int_{-\infty}^{\infty} f(x)\, dx = 1.$