

MATH 323: Probability

Anna Brandenberger

September 25, 2018

This is a transcript of the lectures given by Prof. David Stephens¹ during the fall semester of the 2018-2019 academic year (09-12 2018) for the Probability class (MATH 240). **Subjects covered** are: sample space, events, conditional probability, independence of events, Bayes' Theorem; basic combinatorial probability, random variables, discrete and continuous univariate and multivariate distributions; independence of random variables; inequalities, weak law of large numbers, central limit theorem.

¹ david.stephens@mcgill.ca, BH 1225

1 Basics of Probability

1.1 Review of Set Theory

Definition 1.1. A **set** S is a collection of elements $s \in S$:

- **FINITE:** finite number of elements
- **COUNTABLE:** countably infinite number of elements
- **UNCOUNTABLE:** uncountably infinite number of elements

Definition 1.2. The **power set** $\mathcal{P}(A)$ of A is the set of all subsets of A : $\mathcal{P}(A) = \{B \mid B \subseteq A\}$.

SET OPERATIONS

1. **INTERSECTION:** $s \in A \cap B \iff s \in A \text{ and } s \in B$
extends to A_1, A_2, \dots, A_K (finite): $s \in \bigcap_{k=1}^K A_k \iff s \in A_k \forall k$
2. **UNION:** $s \in A \cup B \iff s \in A \text{ or } s \in B$
extends to A_1, A_2, \dots, A_K (finite²): $s \in \bigcup_{k=1}^K A_k \iff \exists k \text{ s.t. } s \in A_k$
3. **COMPLEMENT:** for $A \subseteq S$, $s \in A' \iff s \in S \text{ but } s \notin A$
4. **SET DIFFERENCE** $A - B \equiv A \setminus B \equiv A \cap B'$
5. **EXCLUSIVE OR (XOR):** $A \oplus B \equiv A \cup B - A \cap B$

Definition 1.3. A_1, A_2, \dots, A_K is called a **partition** of S if these sets are pairwise disjoint $A_j \cap A_k = \emptyset \forall j \neq k$ and exhaustive $\bigcup_{k=1}^K A_k = S$.

Theorem 1.1 (De Morgan's Laws).

$$A' \cap B' = (A \cup B)' \quad A \cup B' = (A \cap B)'$$

Probability is a numerical value representing the chance of a particular event occurring given a particular set of circumstances.

Let $A \subseteq S$:

$$\begin{aligned} A \cap \emptyset &= \emptyset & A \cup \emptyset &= A \\ A \cap S &= A & A \cup S &= S \end{aligned}$$

Intersection \cap and union \cup are commutative and associative.

² technically also extends to a countably infinite number of sets.

Proof. Let $A_1 = A \cup B$, $A_2 = A' \cap B'$. To show that $A'_1 = A_2$, one must show $A_1 \cup A_2 = \emptyset$ and $A_1 \cup A_2 = S$.

$$\begin{aligned} A_1 \cap A_2 &= (A \cup B) \cap A_2 = (A \cap A_2) \cup (B \cap A_2) \\ &= (A \cap A' \cap B') \cup (B \cap A' \cap B') \\ &= \emptyset \cup \emptyset = \emptyset \end{aligned}$$

$$\begin{aligned} A_1 \cup A_2 &= A_1 \cup (A' \cap B') = (A_1 \cup A') \cap (A_1 \cup B') \\ &= (A \cup B \cup A') \cap (A \cup B \cup B') \\ &= S \cap S = S \quad \square \end{aligned}$$

1.2 Sample Spaces and Events

Definition 1.4. The set S of possible outcomes of an experiment is the **sample space** of an experiment, and the individual elements in S are **sample points** or **outcomes**.

Definition 1.5. An **event** A is a collection of sample outcomes $A \subseteq S$. Individual sample outcomes $E_1, E_2, \dots, E_K, \dots \in A$ are termed **simple (elementary) events**. We say that A **occurs** if the outcome $s \in A$. S is the **certain event**; \emptyset the **impossible event**.

Definition 1.6. The **probability function** $P(\cdot)$ assigns numerical values to events: $P : \mathcal{A} \rightarrow \mathbb{R}$, $A \rightarrow p$ (i.e. $P(A) = p$, $A = \mathcal{P}(S)$).

1.3 Axioms and Consequences

Theorem 1.2. Basic probability axioms:

1. $P(A) \geq 0$
2. $P(S) = 1$
3. P is countably additive: if A_1, A_2, \dots s.t. $A_j \cap A_k = \emptyset \forall j \neq k$:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Corollary 1.3. $\forall A \ P'(A) = 1 - P(A)$

Corollary 1.4. $P(\emptyset) = 0$

Corollary 1.5. $\forall A \ P(A) \leq 1$

Corollary 1.6. $\forall A \subseteq B, P(A) \leq P(B)$

Corollary 1.7 (General Addition Rule).

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Using inductive arguments, one can construct a formula for $P(\bigcup_{i=1}^n A_i)$ (c.f. MATH 240) but we will mostly use:

Theorem 1.8 (Boole's Inequality). $P(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$

As a method, one can lay out probabilities in table form such as in Table 1, ensuring that $0 \leq p_{\star} \leq 1$ and probabilities along rows and column add up correctly.

Proof. $S = A \cup A'$. By Theorem 1.2(3),
 $P(S) = P(A) + P(A') = 1$
 $\implies P'(A) = 1 - P(A)$ \square

Proof. Simple from Corollary 1.3. \square

Proof. By Theorem 1.2 and 1.3,
 $1 = P(S) = P(A) + P(A') \geq P(A)$ \square

Proof. Take $B = A \cup (A' \cap B)$. Then,
 $P(B) = P(A) + P(A' \cap B) \geq P(A)$ \square

Proof. First, $(A \cup B) = A \cup (A' \cap B)$
 $\implies P(A \cup B) = P(A) + P(A' \cap B)$ [1]

Then, take $B = (A \cap B) \cup (A' \cap B)$
 $\implies P(B) = P(A \cap B) + P(A' \cap B)$ [2]

So taking [1] - [2], we get:
 $P(A \cup B) - P(B) = P(A) - P(A \cap B)$
 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ \square

	A	A'	Total
B	$p_{A \cap B}$	$p_{A' \cap B}$	p_B
B'	$p_{A \cap B'}$	$p_{A' \cap B'}$	$p_{B'}$
Total	p_A	$p_{A'}$	p_S

Table 1: Probabilities laid out in a table.

1.4 Specifying probabilities

Definition 1.7 (Equally likely sample outcomes). Suppose S finite, with sample outcomes E_1, \dots, E_N , then $P(E_i) = \frac{1}{N} \forall i = 1, \dots, N$. For an event $A \in S$, $\exists n \leq N$ s.t. $A = \bigcup_{i=1}^n E_{i_A}$, $i_A \in \{1, \dots, n\}$
 $\implies P(A) = \frac{n}{N} = \frac{\# \text{ sample outcomes in } A}{\# \text{ sample outcomes in } S}$

Examples of experiments with equally likely sample outcomes include fair coins, dice and cards.

Definition 1.8 (Relative frequencies). For a given experiment with a sample space S and interest event A , consider a finite sequence of N repeat experiments and n the number of times that A occurs. The **frequentist definition** of probability, generalizes the previous definition to $P(A) = \lim_{N \rightarrow \infty} \frac{n}{N}$, the relative frequency of appearance of A .

Definition 1.9 (Subjective assessment). For a given experiment with sample space S , the probability of event A is further generalized to a numerical representation of one's own **personal degree of belief** that the actual outcome lies in A , assuming one is **internally consistent**, i.e. rational and coherent.

1.5 Combinatorial Rules

1. **MULTIPLICATION PRINCIPLE**: a sequence of k operations i , each with n_i possible outcomes, results in $n_1 \times n_2 \times \dots \times n_k$ possible sequences of outcomes.
2. **SELECTION PRINCIPLES**: selecting from a finite set can be done:
 - **WITH REPLACEMENT**: each selection is independent;
 - **WITHOUT REPLACEMENT**: set is depleted by each selection.
3. **ORDERING**: the order of a sequence of selections can be:
 - **IMPORTANT** (312 distinct from 123);
 - **UNIMPORTANT** (312 identical to 123).
4. **DISTINGUISHABLE ITEMS**: objects being selected can be:
 - **DISTINGUISHABLE**: individually labelled (e.g. lottery balls);
 - **INDISTINGUISHABLE**: labelled according to a type (e.g. colors).

c.f. MATH 240 notes for examples.

Definition 1.10. A **permutation** is an ordered arrangement of r distinct objects. The number of possible permutations for a set r is $P_r^n = \frac{n!}{(n-r)!}$.

Definition 1.11. The number of ways of partitioning n distinct elements into k disjoint subsets of sizes n_1, \dots, n_k is the **multinomial coefficient** $N = \binom{n}{n_1, \dots, n_k} := \frac{n!}{n_1! \times n_2! \times \dots \times n_k!}$. This is a generalization of the **binomial coeff.** $\binom{n}{j} = \frac{n!}{j!(n-j)!} := \binom{n}{j, n-j}$.

Definition 1.12. The number of **combinations** C_r^n is the number of subsets of size r that can be selected from n objects $C_r^n = \binom{n}{r} = n! \cdot P_r^n$.

1.6 Conditional Probability and Independence

Definition 1.13. The **conditional probability** of A given B is $P(A | B) = \frac{P(A \cap B)}{P(B)}$ when $P(B) > 0$.

Corollary 1.9. $P(A | S) = P(A) \quad P(A | A) = 1 \quad P(A | B) \leq 1$

Warning 1. Important distinction between $P(A \cap B)$ and $P(A | B)$:

- $P(A \cap B)$: chance of A and B occurring **RELATIVE TO S**
- $P(A | B)$: chance of A and B occurring **RELATIVE TO B** — non-symmetric: in general, $P(A | B) \neq P(B | A)$.

Theorem 1.10. The conditional probability function satisfies the probability axioms (Theorem 1.2).

Proof. 1. $P(A | B) = \frac{P(A \cap B)}{P(B)} \geq 0$

2. $P(S | B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

3. $P(\bigcup_{i=1}^{\infty} (A_i | B)) = \frac{P(\bigcup_{i=1}^{\infty} A_i \cap B)}{P(B)} = \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} P(A_i | B)$
and since A_i disjoint $\forall i$, (A_i, B) also disjoint. \square

Definition 1.14. Two events $A, B \in S$ are **independent** if $P(A | B) = P(A) \iff P(A \cap B) = P(A)P(B)$.

Remark 2. Notice then that $P(B | A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B)$.

Warning 3. Not the same as mutual exclusivity where $P(A \cap B) = 0$!

Warning 4. For multiple events $A_1, \dots, A_n \in S$, we can talk about pairwise independence between any A_i, A_j with $i \neq j$, but this does not automatically generalize (c.f. Example 1.2).

Definition 1.15. Events A_1, \dots, A_K are **mutually independent** if $P(\bigcap_{k \in \mathcal{I}} A_k) = \prod_{k \in \mathcal{I}} P(A_k)$ for all subsets \mathcal{I} of $\{1, 2, \dots, K\}$.

Given partial knowledge (B happened) i.e. knowing $s \in B \subseteq S$, one can make a more accurate probability assessment of some other event A , as $s \in A \cap B$.

Example 1.1. Given three two-sided cards, one RR, one RB and one BB, one card is selected randomly and one side displayed: it is red (R). Find the probability that the other side is red.

Answer. $S = \{R_1, R_2, B, R, B_1, B_2\}$
Card 1 (RR) selected $A = \{R_1, R_2\} \implies P(A) = 2/6$
Red side exposed: $B = \{R_1, R_2, R\} \implies P(B) = 3/6$
 $P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{2/6}{3/6} = 2/3$

Example 1.2. Rolling two dice with independent outcomes A_1 : first roll outcome is odd; A_2 : second roll outcome is odd and A_3 : total outcome is odd. Then $P(A_1) = P(A_2) = \frac{1}{2}$ and $P(A_1 | A_3) = P(A_2 | A_3) = \frac{1}{2}$. But $P(A_1 \cap A_2 \cap A_3) = 0$, $P(A_1 | A_2 \cap A_3) = 0$, etc.

Definition 1.16. Consider events $A_1, A_2, B \in S$ with $P(B) > 0$. A_1 and A_2 are **conditionally independent** given B if $P(A_1 | A_2 \cap B) = P(A_1 | B) \iff P(A_1 \cap A_2 | B) = P(A_1 | B)P(A_2 | B)$

Theorem 1.11 (General Multiplication Rule).

$$P(A_1, A_2, \dots, A_K) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) \cdots \\ \cdots P(A_K | A_1 \cap A_2 \cap \cdots \cap A_{K-1})$$

Proof. Trivial, recursively (Chain Rule for probabilities). \square

We can use **probability trees** to display joint probabilities of multiples, where the junctions are the events and the branches correspond to possible sequences of choices of events. We then *multiply along a branch* to get the joint probability.

1.7 Important Theorems

Theorem 1.12 (Theorem of Total Probability). Given a partition B_1, \dots, B_n of the sample space S into n subsets s.t. $P(B_i) > 0$, then for any event $A \in S$, $P(A) = \sum_{i=1}^n P(A | B_i)P(B_i)$.

Proof.

$$\begin{aligned} A &= (A \cap B_1) \cup \cdots \cup (A \cap B_n) \\ P(A) &= P(A \cap B_1) + \cdots + P(A \cap B_n) \\ &= P(A | B_1)P(B_1) + \cdots \\ &\quad + P(A | B_n)P(B_n) \\ &= \sum_{i=1}^n P(A | B_i)P(B_i) \end{aligned} \quad \square$$