Last modified November 1, 2019

STA257: Probability and Statistics I

University of Toronto — Fall 2019

Jeff Shen

Contents

1 Probability and Counting

1.1 Introduction

Experiments: situations wherre outcome is random. eg. flipping a coin is an experiment.
Sample space: set of all possible outcomes, denoted S or Ω. The number of elements in the sample space (cardinality) is denoted $|\Omega|$.
Event: a subset of a sample space.
Outcome: a particular element of a sample space $s_1 \in \Omega$.

1.2 Set Theory

1.2.1 Definitions

- **union**: $A \cup B$. Elements in either A or B.
- **intersection**: $A \cap B$. Elements in both A and B.
- **complement** of A: A^c. Elements not in A.
- **empty set**: $\varnothing$. Set with no elements in it.
- A and B are **disjoint**: $A \cap B = \varnothing$. There are no elements in the intersection of A and B.

1.2.2 Laws of Set Theory

1. **commutativity**: $A \cup B = B \cup A$, $A \cap B = B \cap A$
2. **associativity**: $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap B)$
3. **distributivity**: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

1.3 Probability Measures

Probability measure: a function which maps subsets of Ω, which can be defined on any space, to real numbers $\mathbb{R}$.

1.3.1 Axioms of Probability Measures

- $P(\Omega) = 1$
- $\forall A \in \Omega, P(A) \geq 0$
- if $A_1, A_2, \dots A_n, \dots$ are mutually disjoint, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

1.3.2 Properties of Probability Measures

- $\forall A \in \Omega, P(A^c) = 1 - P(A)$

  Proof:

  $$\begin{aligned} 1 &= P(\Omega) && \text{by Axiom 1} \\ &= P(A \cup A^c) && \text{by definition of complement} \\ &= P(A) + P(A^c) && \text{by Axiom 3 (since } A, A^c \text{ are disjoint)} \end{aligned}$$

  Rearrange this to see that $P(A^c) = 1 - P(A)$.

- $P(\varnothing) = 0$

  Proof:

  $$\begin{aligned} P(\Omega) &= P(\Omega \cup \varnothing) && \text{since } \Omega \cup \varnothing = \Omega \\ &= P(\Omega) + P(\varnothing) && \text{by Axiom 3 (since } \Omega, \varnothing \text{ are disjoint)} \end{aligned}$$

  So $P(\varnothing) = 0$.

- For $A, B \subseteq \Omega, A \subseteq B \implies P(A) \leq P(B)$

  Proof:

$$\begin{aligned} P(B) &= P(A \cup (B \cap A^c)) \\ &= P(A) + P(B \cap A^c) && \text{by Axiom 3 (since } A, A^c \text{ are disjoint)} \end{aligned}$$

  But note that $P(B \cap A^c) \geq 0$ by Axiom 2.

  Then $P(B) = P(A) + P(B \cap A^c) \geq P(A)$.

- For $A, B \subseteq \Omega, P(A \cup B) = P(A) + P(B) - P(A \cap B)$

  Proof:

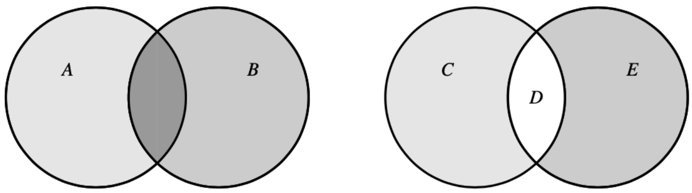
  Case 1: A, B are disjoint. Then $A \cap B = \varnothing \implies P(A \cap B) = 0$.

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) && \text{by Axiom 3 (since } A, B \text{ are disjoint)} \\ &= P(A) + P(B) + P(A \cap B) && \text{since we can add 0 wherever we want} \end{aligned}$$

  Case 2: A, B not disjoint. Then $A \cap B \neq \varnothing$.

  Let $C = A \cap B^c$, $D = A \cap B$, $E = A^c \cap B$.

  Then C, D, E are disjoint, and $A = C \cup D$, $B = D \cup E$, and $A \cup B = C \cup D \cup E$.



$$\begin{aligned} P(A) + P(B) - P(A \cap B) &= P(C \cup D) + P(D \cup E) - P(D) && \text{by how we defined } C, D, E \\ &= P(C) + P(D) + P(D) + P(E) - P(D) && \text{by Axiom 3 and disjointness of } C, D, E \\ &= P(C) + P(D) + P(E) \\ &= P(C \cup D \cup E) && \text{by Axiom 3 and disjointness of } C, D, E \\ &= P(A \cup B) \end{aligned}$$

1.4 Permutations and Combinations

1.5 Conditional Probability

1.6 Independence, Law of Total Probability

2 Random Variables

2.1 Discrete Random Variables

2.1.1 Bernoulli

2.1.2 Binomial

2.1.3 Geometric

2.1.4 Negative Binomial

2.1.5 Hypergeometric

2.1.6 Poisson

2.2 Continuous Random Variables

2.2.1 Uniform

2.2.2 Exponential

2.2.3 Gamma

2.2.4 Beta

2.2.5 Uniform

2.2.6 Standard Normal

2.2.7 General Normal

2.3 Transformations of Random Variables

3 Expected Values

3.1 Mean and Variance

3.1.1 LOTUS

3.1.2 Inequalities

3.2 Moment Generating Functions

4 Joint Distributions

4.1 Joint and Marginal Distributions

4.1.1 Discrete

4.1.2 Continuous

4.2 Independence in Joint Distributions

4.3 Conditional Distributions

4.3.1 Discrete

4.3.2 Continuous

4.4 Functions of Joint Distributions

4.5 Order Statistics