STAT-2920 Intro to Probability Review

Aidan Richer

November 2025

This paper contains a cumulative review of all definitions, theorems, and proofs within the lecture notes, excluding any examples.

Contents

1   Chapter 1: Combinatorial Analysis

1.1   The Basic Principle of Counting

Theorem. *Suppose that two experiments are to be performed. Then if experiment 1 can result in any one of m possible outcomes and if, for each outcome of experiment 1, there are n possible outcomes of experiment 2, then together there are $m \times n$ possible outcomes of the two experiments.*

1.2   Permutations

Frequently, we are interested in situations where the outcomes are the different ways in which a group of objects can be ordered or arranged. Different arrangements like these are called permutations.

Definition 1. A permutation is a distince arrangement of n different elements of a set.

Theorem. *The number of permutations of n distinct objects taken r at a time is*

$$\mathrm{P}^n_r = \frac{n!}{(n-r)!}$$

for $r = 1, 2, \ldots, n$

Theorem. *The number of permutations of n objects of which n_1 are of one kind, n_2 are of a second kind, $\ldots, n_k$ are of a kth kind, and $n_1 + n_2 + \cdots + n_k = n$ is*

$$\frac{n!}{n_1! \times n_2! \times \cdots \times n_k!}$$

1.3   Combinations

Definition 2. A combination is a selection og r objects taken from n distincy objects without regard to the order of selection.

Theorem. *The number of combinations of n distinct objects taken r at a time is*

$$\mathrm{C}^n_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

for $r = 0, 1, 2, \ldots, n$

Theorem. *The number of ways in which a set of n distinct objects can be partitioned into k subsets with n_1 objects in the first subset, n_2 objects in the second subset, $\ldots$, and n_k objects in the kth subset is*

$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! \times n_2! \times \cdots \times n_k!}, \quad n = n_1 + n_2 + \cdots + n_k$$

1.4   Binomial Coefficients

Definition 3. The coefficient of $x^{n-r}y^r$ in the binomial expansion of $(x+y)^n$ is called the binomial coefficient $\binom{n}{r}$.

Theorem.

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^{n-r} y^r$$

for any positive integer n.

The calculation of binomial coefficients can often be simplified by making use of the three theorems that follow.

Theorem. *For any positive integers n and $r = 0, 1, 2, \ldots, n$,*

$$\binom{n}{r} = \binom{n}{n-r}$$

Theorem. *For any positive integer n and $r = 1, 2, \ldots, n-1$, we have*

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$$

Theorem.

$$\sum_{r=0}^{k} \binom{m}{r}\binom{n}{k-r} = \binom{m+n}{k}$$

2   Chapter 2: Probability

2.1   Review of Set Theory

Definition 4. A <u>set</u> is an unordered collection of objects called elements of the set.

Definition 5. Two sets are equal if they contain exactly the same elements. We say

$$a \in \mathcal{A} \text{ means that element } a \text{ belongs to the set } \mathcal{A}$$

$$a \notin \mathcal{A} \text{ means the element } a \text{ does not belong to the set } \mathcal{A}$$

$$|\mathcal{A}| \text{ means the size, or the cardinality of a set } \mathcal{A}$$

Definition 6. A set A is a subset of a set B if every element of A is also an element of B. We write

$$A \subseteq B$$

If A is not a subset of B, we write

$$A \nsubseteq B$$

Definition 7. $S = \Omega$ is the universal set i.e. the collection of all points of interest for the present situation. The empty set, denoted $\emptyset$ is the set with no elements.

Law (Commutative Laws).
$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

Law (Associative Laws).
$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$

Law (Distributive Laws).
$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$
$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$
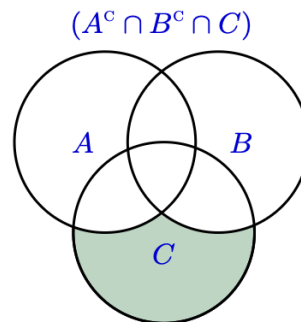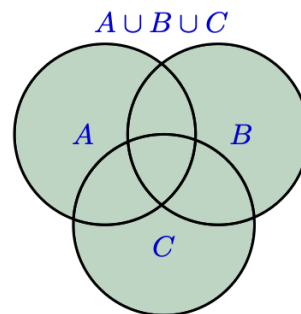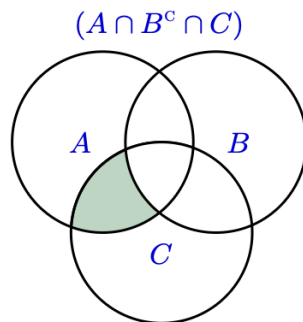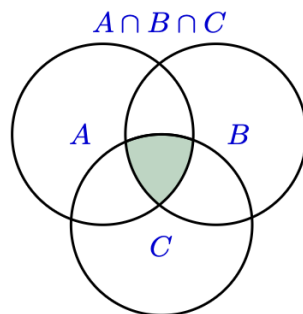
Law (Basic Sample Space Laws).
$$A \cap A^C = \emptyset$$
$$A \cup A^C = S$$

Law (De Morgan's Laws).
$$(A \cup B)^C = A^C \cap B^C$$
$$(A \cap B)^C = A^C \cup B^C$$

2.2   The Probability of an Event

Theorem. *If A is an event in a discrete sample space S, then $\mathbb{P}(A)$ equals the sum of the probabilities of the individual outcomes comprising A.*

Theorem. *If an experiment can result in any one of N different equally likely outcomes, and if n of these outcomes together constitute event A, then the probability of event A is*

$$\mathbb{P}(A) = \frac{n}{N}$$

2.3   Properties of Probability in Sets

Theorem (Properties)**.** *A probability $\mathbb{P}$ satisfies the following properties:*

1. *For each event A, $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$*

2. *The probability of the null set is zero; that is, $\mathbb{P}(\emptyset) = 0$*

3. *$\mathbb{P}(A) \leq \mathbb{P}(B)$ whenever $A \subset B$*

4. *For each A, $0 \leq \mathbb{P}(A) \leq 1$*

Theorem. *For any events A and B, we have*

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

Theorem. *If A, B and C are any three events in a sample space S, then*

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)$$

2.4   Conditional Probability

Definition 8. If A and B are any two events in a sample space S and $\mathbb{P}(A) > 0$, the conditional probability of B given A is

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Theorem. *If A and B are any two events in a sample space S and $\mathbb{P}(A) \neq 0$, then*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B \mid A)$$

Theorem. *If A, B, and C are any three event in a sample space S such that $\mathbb{P}(A \cap B) \neq 0$, then*

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B \mid A) \cdot \mathbb{P}(C \mid A \cap B)$$

2.5   Independent Events

Definition 9. Two events A and B are <u>independent</u> if and only if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Theorem. *If A and B are independent, then A and B^C are also independent.*

Definition 10. Events $A_1, A_2, \ldots, A_k$ are independent if and only if the probability of the intersections of any $2, 3, \ldots$, or k of these events equals the product of their respective probabilities. For three events, A, B, and C, for example, independence requires that

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

$$\mathbb{P}(A \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(C)$$

$$\mathbb{P}(B \cap C) = \mathbb{P}(B) \cdot \mathbb{P}(C)$$

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)$$

2.6   Law of Total Probability

Theorem. *If the events $B_1, B_2, \ldots$, and B_k constitute a partition of the sample space S and $\mathbb{P}(B_i) \neq 0$ for $i = 1, 2, \ldots, k$, then for any event A in S*

$$\mathbb{P}(A) = \sum_{i=1}^{k} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{k} \mathbb{P}(B_i) \cdot \mathbb{P}(A \mid B_i)$$

2.7   Bayes' Theorem

Theorem. *If the events $B_1, B_2, \ldots$, and B_k constitute a partition of the sample space S and $\mathbb{P}(B_i) \neq 0$ for $i = 1, 2, \ldots, k$, then for any event A in S, such that $\mathbb{P}(A) \neq 0$*
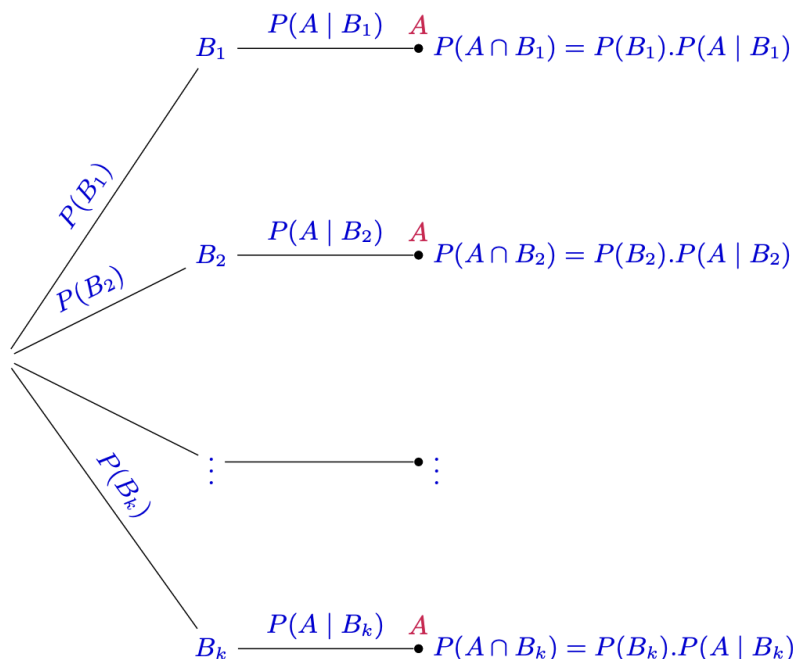
$$\mathbb{P}(B_r \mid A) = \frac{\mathbb{P}(B_r) \cdot \mathbb{P}(A \mid B_r)}{\displaystyle\sum_{i=1}^{k} \mathbb{P}(B_i) \cdot \mathbb{P}(A \mid B_i)}$$

for $r = 1, 2, \ldots, k$

3   Chapter 3: Probability Distribution and Probability Density

3.1   Random Variables

In this chapter, random variables are denoted by capital letters and their values by the corresponding lowercase letter. For instance, we shall write x to denote a value of the random variable X.

Definition 11. If S is a sample space with a probability measure and X is a real-valued function defined over the elements of S, then X is called a <u>random variable</u>.

Definition 12. X is a discrete random variable if its range is finite or countable infinite.

3.2   Probability Mass Function

Definition 13 (Probability Distribution)**.** If X is a discrete random variable, the function given by $f(x) = P(X = x)$ for each x within the range of X is called the <u>probability distribution</u> of X. $f(x)$ is also called the probability mass function (pmf).

Theorem. *A function can serve as the probability distribution of a discrete random variable X if and only if its values, $f(x)$, satisfy the conditions:*

1. *$f(x) \geq 0$ for each value within its domain.*

2. *$\sum_{x} f(x) = 1$, where the summation extends over all the values within its domain.*

3.3   Cumulative Distribution Function

Definition 14. If X is a discrete random variable, the function given by

$$F(x) = \mathbb{P}(X \leq x) = \sum_{t \leq x} f(t) \quad \text{for } -\infty < x < \infty$$

where $f(t)$ is the value of the probability distribution of X at t, is called the <u>distribution function</u>, or the <u>cumulative distribution function (cdf)</u> of X.

Theorem. *The values $F(x)$ of the distribution function of a discrete random variable X satisfy the conditions*

1. *$F(-\infty) = 0$ and $F(\infty) = 1$*

2. *if $a < b$, then $F(a) \leq F(b)$ for any real numbers a and b.*

Theorem. *If the range of a random variable X consists of the values $x_1 < x_2 < x_3 < \cdots < x_n$, then $f(x_1) = F(x_1)$ and*

$$f(x_i) = F(x_i) - F(x_{i-1}) \quad \text{for } i = 2, 3, 4, \ldots, n$$

3.4   Probability Density Functions

Definition 15. A function with values $f(x)$, defined over the set of all real numbers, is called a probability density function (pdf) of the continuous random variable X if and only if

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx$$

for any real constants a and b with $a \leq b$.

Theorem. *If X is a continuous random variable and a and b are real constants with $a \leq b$, then*

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b)$$

Theorem. *A function can serve as a probability density of a continuous random variable X if its values, $f(x)$, satisfy the conditions:*

1. *$f(x) \geq 0$, for $-\infty < x < \infty$*

2. *$\int_{-\infty}^{\infty} f(x)dx = 1$*

3.5   Distribution Functions

Definition 16. If X is a continuous random variable and the value of its probability density at t is $f(t)$, then the function given by

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)dt \quad \text{for } -\infty < x < \infty$$

Theorem. *If $f(x)$ and $F(x)$ are the values of the probability density and the distribution function of X at x, then*

$$\mathbb{P}(a \leq X \leq b) = F(b) - F(a)$$

for any real constants a and b with $a \leq b$ and

$$f(x) = \frac{dF(x)}{dx}$$

where the derivative exists.

4   Chapter 4: Mathematical Expectation

5   Chapter 5: Special Probability Distributions

6   Chapter 6: Special Probability Densities