

EN.553.420 Introduction to Probability

Daniel Yao

Spring 2024

Contents

0	Introduction.	2
0.0	Introduction.	2
0.1	Remark.	2
1	Probability Spaces.	3
1.0	Introduction	3
1.1	Remark.	3
1.2	Remark.	3
1.3	Definition.	3
1.4	Definition.	3
1.5	Definition.	3
1.6	Definition.	3
1.7	Theorem.	3
1.8	Proof.	4
1.9	Theorem.	4
1.10	Proof.	4
2	Combinatorics.	5
2.0	Combinatorics.	5
2.1	Definition.	5
2.2	Definition.	5
2.3	Definition.	5
2.4	Theorem.	5
2.5	Proof.	5
2.6	Theorem.	5
2.7	Corollary.	5
2.8	Example.	6
2.9	Definition.	6
2.10	Example.	6
2.11	Definition.	6
2.12	Theorem.	6
2.13	Proof.	6
2.14	Corollary.	7
2.15	Corollary.	7
2.16	Lemma.	7
3	Classical Probability.	8
3.0	Introduction.	8
3.1	Definition.	8
3.2	Remark.	8
4	Discrete Probability.	9
4.0	Introduction.	9
4.1	Definition.	9
4.2	Definition.	9
4.3	Definition.	9
4.4	Theorem.	9
4.5	Definition.	9
4.6	Definition.	9
4.7	Definition.	9
4.8	Definition.	9
4.9	Definition.	9

0.0 Introduction.

Introduction.

0.1 Remark.

These notes follow the material presented in EN.553.420 taught by Professor Fred Torcaso during the semester of Spring 2024 at The Johns Hopkins University. These notes shall be brief as they are meant to supplement the lectures, homeworks, and notes provided by Professor Torcaso.

1.0 Introduction

Probability Spaces.

1.1 Remark.

The following definitions shall be informal. For a more formal treatment of probability, please refer to the EN.553.430 Mathematical Statistics.

1.2 Remark.

In this class, we shall investigate three notions of probability:

- (1) Classical probability.
- (2) Discrete probability.
- (3) Continuous probability.

We shall see that each of these frameworks satisfies the below axioms.

1.3 Definition.

A sample space is a set denoted by the symbol Ω .

1.4 Definition.

The event space \mathcal{F} is a collection of subsets of Ω . An event is denoted by the symbol A . (Formally, the event space \mathcal{F} is a sigma-algebra.)

1.5 Definition.

A probability measure is a function $P : \mathcal{F} \rightarrow \mathbb{R}$ subject to the three Komolgorov Axioms:

- (1) $\forall A \in \mathcal{F}, P(A) \geq 0$. (Nonnegativity.)
- (2) $P(\Omega) = 1$. (Normalization.)
- (3) If $A_1, A_2, \dots \subseteq \Omega$ is a countable sequence of disjoint events, i.e, for any $i \neq j, A_i \cap A_j = \emptyset$, then

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

(Countable additivity.)

1.6 Definition.

A probability space is an ordered triple (Ω, \mathcal{F}, P) .

1.7 Theorem.

From the Komolgorov axioms, we find that the probability of the empty set is

$$P(\emptyset) = 0$$

and that if $A_1, \dots, A_n \subseteq \Omega$ is a finite sequence of disjoint events, then

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

and that for any two events $A, B \subseteq \Omega$,

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

1.8 Proof.

(Continued.) The proof is presented in lecture. The reproduction of the proof is left as an exercise to the student.

1.9 Theorem.

(Inclusion-exclusion principle.) Suppose that $A_1, \dots, A_n \subseteq \Omega$ is a finite sequence of events. The probability of the union is the sum of all of the probabilities i -intersections with alternating signs. The notation is cumbersome, so the example for $n = 3$ is

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

1.10 Proof.

(Continued.) The proof is by induction.

2.0 Combinatorics.

Combinatorics.

2.1 Definition.

If $n \in \mathbb{N}$, then the factorial is defined as

$$n! = n(n-1) \dots 1.$$

If $n = 0$, then the empty product is defined as

$$0! = 1.$$

2.2 Definition.

Suppose that we have n objects from which we want to choose an ordered permutation of k objects. Then, the number of permutations is given by the falling factorial

$$(n)_k = \frac{n!}{(n-k)!}.$$

2.3 Definition.

Suppose that we have n objects from which we want to choose an unordered subset of k objects. Then, the number of subsets is given by the binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

which is pronounced ' n choose k '. Observe that

$$\binom{n}{k} = \binom{n}{n-k}.$$

2.4 Theorem.

(Pascal's identity.)

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

2.5 Proof.

(Continued.) The proof is left as an exercise to the student.

2.6 Theorem.

(Binomial theorem.) If $a, b \in \mathbb{R}$ and $n = 1, 2, \dots$, then

$$(a+b)^n = \sum_{k=1}^n \binom{n}{k} a^k b^{n-k}.$$

2.7 Corollary.

(Continued.)

$$\sum_{k=1}^n \binom{n}{k} = 2^n.$$

We may interpret this corollary as that if A is a set with n elements, then the power set, i.e., the set of all subsets, of A has 2^n elements.

2.8 Example.

The number of anagrams of *PEEPPEE* is

$$\binom{6}{2} = \frac{6!}{2!4!}$$

because we choose 2 of the positions to be *P* and the rest to be *E*. Alternatively, we may realize that there are 6! ways to arrange the six letters and that within each group, there are 2! and 4! ways, respectively, to arrange the letters within a group such that the result is identical.

2.9 Definition.

If $n, k_1, \dots, k_K \in \mathbb{N}$ and $k_1 + \dots + k_K = n$, then the multinomial coefficient is defined

$$\binom{n}{k_1, \dots, k_K} = \frac{n!}{k_1! \dots k_K!}.$$

If $k_1 + \dots + k_K < n$, then we take that

$$\binom{n}{k_1, \dots, k_K} = \binom{n}{k_1, \dots, k_K, n - k_1 - \dots - k_K}.$$

To be clear, however, it is good practice to ensure that $k_1 + \dots + k_K = n$.

2.10 Example.

The number of anagrams of *PEEPPEEPOOPOO* is

$$\binom{12}{4, 4, 4} = \frac{12!}{4!4!4!}$$

with similar reasoning as with example 2.10.

2.11 Definition.

A multiset is a collection of objects where objects may be repeated, i.e., the objects have a multiplicity. A multiset is denoted with curly braces, e.g.,

$$\{A, B, B, \dots\}.$$

2.12 Theorem.

(Stars and bars.) Suppose that we have r distinguishable bins into which we want to place n indistinguishable objects. Then, the number of ways to do so is given by

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

2.13 Proof.

(Continued.) We enumerate the bins $i = 1, \dots, r$. To each bin i , we want to assign n_i balls such that $n_1 + \dots + n_r = n$. But the ways to do so are in bijection with the anagrams of n stars \star and k bars $|$ where the stars denote the balls and the bars denote the boundaries between the bins. We realize that there are

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

such anagrams.

2.14 Corollary.

(Continued.) The number of multisets of n elements with r distinct labels allowed is

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

2.15 Corollary.

(Continued.) The number of integer r -partitions of n , i.e., the number of integer solutions to $n_1 + \dots + n_r = n$ is

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

2.16 Lemma.

Suppose that we have a two-step experiment. There are m ways to perform the first step. For each of those ways, there are n ways to perform the second step. Then, there are mn ways to perform the experiment in total.

3.0 Introduction.

Classical Probability.

3.1 Definition.

Suppose that Ω is a finite set and $A \subseteq \Omega$. Furthermore, suppose that every singleton subset of Ω is equiprobable. Then,

$$P(A) = \frac{|\Omega|}{|A|}$$

where $|\cdot|$ denotes the cardinality of the set \cdot .

3.2 Remark.

Since every singleton event is equiprobable, then we may determine the probability of a composite event by counting the cardinality of that event. Classical probability is thus combinatorial.

4.0 Introduction.

Discrete Probability.

4.1 Definition.

A discrete random variable is a (measurable) function $X : \Omega \rightarrow \mathbb{Z}$.

4.2 Definition.

The probability mass function of a discrete rv X is a probability measure $p : \mathbb{Z} \rightarrow \mathbb{R}$. In particular if a random variable X possesses pmf p , then

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

4.3 Definition.

The expectation of a discrete rv X with pmf p

$$E(X) = \sum_{i=-\infty}^{\infty} xp(x).$$

We sometimes denote $\mu = E(X)$.

4.4 Theorem.

(Law of the unconscious statistician.) Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a function and let X be a discrete rv with pmf p . Then, the expectation of $f(X)$ is

$$E(f(X)) = \sum_{i=-\infty}^{\infty} f(x)p(x).$$

4.5 Definition.

Let $k = 1, 2, \dots$. The k -th moment of a discrete rv X is

$$E(X^k) = \sum_{i=-\infty}^{\infty} x^k p(x).$$

4.6 Definition.

Let $k = 1, 2, \dots$. The k -th moment central moment of a discrete rv X is

$$E((X - \mu)^k).$$

4.7 Definition.

The variance of a discrete rv X is the second central moment

$$\begin{aligned} \text{Var}(X) &= E((X - \mu)^2) \\ &= E(X^2) - E(X)^2. \end{aligned}$$

4.8 Definition.

The skewness of a discrete rv X is the third central moment $E((X - \mu)^3)$.

4.9 Definition.

The kurtosis ('tailedness') of a discrete rv X is the fourth central moment $E((X - \mu)^4)$. The excess kurtosis is $E((X - \mu)^4) - 3$ (because the kurtosis of the normal distribution is 3).