

EN.553.420 Introduction to Probability

Daniel Yao

Spring 2024

Contents

0	Introduction.	3
0.1	Introduction.	3
0.2	Remark.	3
1	Probability Spaces.	4
1.1	Introduction	4
1.2	Remark.	4
1.3	Remark.	4
1.4	Definition.	4
1.5	Definition.	4
1.6	Definition.	4
1.7	Definition.	4
1.8	Theorem.	4
1.9	Proof.	5
1.10	Theorem.	5
1.11	Proof.	5
2	Combinatorics.	6
2.1	Combinatorics.	6
2.2	Definition.	6
2.3	Definition.	6
2.4	Definition.	6
2.5	Theorem.	6
2.6	Proof.	6
2.7	Theorem.	6
2.8	Corollary.	7
2.9	Example.	7
2.10	Definition.	7
2.11	Example.	7
2.12	Definition.	7
2.13	Theorem.	7
2.14	Proof.	8
2.15	Corollary.	8
2.16	Corollary.	8
2.17	Lemma.	8
3	Classical Probability.	9
3.1	Introduction.	9
3.2	Definition.	9
3.3	Remark.	9
3.4	Example.	9
4	Conditional Probability.	10
4.0	Introduction.	10
4.1	Definition.	10
4.2	Theorem.	10
4.3	Proof.	10
4.4	Example.	10
5	Random Variables.	12
5.0	Introduction.	12
5.1	Definition.	12
5.2	Definition.	12
5.3	Definition.	12
5.4	Definition.	12
5.5	Definition.	12
5.6	Definition.	12

5.7	Definition.	12
5.8	Remark.	13
5.9	Definition.	13
5.10	Theorem.	13
5.11	Theorem.	13
5.12	Definition.	13
5.13	Definition.	13
5.14	Definition.	13
5.15	Definition.	14
5.16	Definition.	14
5.17	Definition.	14
5.18	Theorem.	14
5.19	Lemma.	14
5.20	Proof.	14
5.21	Definition.	14
5.22	Remark.	14
5.23	Remark.	15
6	Joint Distributions.	16
6.0	Introduction.	16
7	Transformations of Random Variables.	17
7.0	Introduction.	17
8	Order Statistics.	18
8.0	Introduction.	18
9	Convergences and Inequalities.	19
9.0	Introduction.	19

0 Introduction.

0.1 Introduction.

Introduction.

0.2 Remark.

These notes follow the material presented in EN.553.420 Introduction to Probability 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 Probability Spaces.

1.1 Introduction

Probability Spaces.

1.2 Remark.

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

1.3 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.

1.4 Definition.

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

1.5 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.)

1.6 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

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

(Countable additivity.)

1.7 Definition.

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

1.8 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 $A, B \subseteq \Omega$,

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

1.9 Proof.

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

1.10 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

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

1.11 Proof.

(Continued.) The proof is by induction.

2 Combinatorics.

2.1 Combinatorics.

Combinatorics.

2.2 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

$$0! = 1.$$

2.3 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

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

2.4 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

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

2.5 Theorem.

(Pascal's identity.)

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

2.6 Proof.

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

2.7 Theorem.

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

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

2.8 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.

2.9 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.

2.10 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 $k_1 + \dots + k_K = n$.

2.11 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.

2.12 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.,

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

2.13 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

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

2.14 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 $r-1$ 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.

2.15 Corollary.

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

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

2.16 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

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

2.17 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 Classical Probability.

3.1 Introduction.

Classical Probability.

3.2 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 .

3.3 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.

3.4 Example.

Suppose that we have a standard 52-card deck from which we select a hand of five cards. The probability that we have a full house is

$$(13)_2 \binom{4}{3} \binom{4}{2} / \binom{52}{5} \approx 0.001441$$

because we select an ordered pair of ranks (one rank for the three-of-a-kind and one rank for the pair) and then select the suits for the three-of-a-kind and the pair.

4.0 Introduction.

Conditional Probability.

4.1 Definition.

For events $A, B \subseteq \Omega$, the conditional probability of A given B is

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

provided that $P(B) > 0$.

Theorem. (Law of Total Probability) Let B_1, B_2, \dots be a partition of Ω . For any event $A \subseteq \Omega$, we have that

$$P(A) = \sum_{i=1}^{\infty} P(A | B_i)P(B_i).$$

Example. Suppose that we have two coins, one fair and one with only heads. We select a coin at random and flip it n times. Find the probability that we observe all heads.

Let A be the event that we select all heads and let B, B^c be the events that we select the unfair coin and the fair coin, respectively. We have that

$$\begin{aligned} P(A) &= P(A | B)P(B) + P(A | B^c)P(B^c) \\ &= 1/2 + (1/2)^{n+1}. \end{aligned}$$

4.2 Theorem.

(Bayes's Rule) For events $A, B \subseteq \Omega$, the conditional probability of A given B is

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

4.3 Proof.

(Continued.) Observe that

$$\begin{aligned} P(A | B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(B | A)P(A)}{P(B)}. \end{aligned}$$

4.4 Example.

Suppose that we have two coins, one fair and one with only heads. We select a coin at random and flip it n times, observing all heads. Find the minimum number of heads to conclude that the coin is the unfair coin with probability greater than p .

Let A be the event that we select the unfair coin and B be the event that we observe all heads. By Bayes's rule, we have that

$$\begin{aligned} P(A | B) &= \frac{P(B | A)P(A)}{P(B)} \\ &= \frac{P(B | A)P(A)}{P(B | A)P(A) + P(B | A^c | A^c)} \\ &= \frac{1/2}{1/2 + (1/2)^{n+1}} \\ &= \frac{1}{1 + (1/2)^n}. \end{aligned}$$

If we want to solve for n , then we have that

$$p < \frac{1}{1 + (1/2)^n}$$

$$2^{-n} < \frac{1}{p} - 1,$$

which yields our final answer of

$$n > \log_2 \left(\frac{p}{1-p} \right).$$

Definition. Two events $A, B \subseteq \Omega$ are independent if

$$P(A) = P(A \mid B).$$

Equivalently,

$$P(A \cap B) = P(A)P(B).$$

Of course, A is independent of B if and only if B is independent of A .

Definition. Let $A_1, A_2, \dots \subseteq \Omega$ be a sequence of events. The events are independent if for any index set $I \subseteq \mathbb{N}$, we have that

$$P \left(\bigcap_{i \in I} A_i \right) = \prod_{i \in I} P(A_i).$$

Theorem. If $A_1, A_2, \dots \subseteq \Omega$ are independent, then replacing any event A_i with its complement A_i^c yields a sequence of independent events.

5.0 Introduction.

Random Variables.

5.1 Definition.

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

5.2 Definition.

The cumulative distribution function (cdf) of a random variable X is a function $F : \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$F(x) = P(X \leq x).$$

A cdf has the properties

- (1) $\lim_{x \rightarrow -\infty} F(x) = 0$.
- (2) $\lim_{x \rightarrow \infty} F(x) = 1$.
- (3) F is nondecreasing.
- (4) F is right-continuous.

5.3 Definition.

A discrete random variable X is a (measurable) function $X : \Omega \rightarrow \text{supp}(X)$ where $\text{supp}(X)$ is a countable set. We shall furthermore take $\text{supp}(X) \subseteq \mathbb{Z}$.

5.4 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 the probability of event $A \subseteq \Omega$ is

$$P(X \in A) = \sum_{x \in A} p(x).$$

5.5 Definition.

A continuous random variable X is a (measurable) function $X : \Omega \rightarrow \mathbb{R}$ with a continuous cdf $F : \mathbb{R} \rightarrow \mathbb{R}$.

5.6 Definition.

If X is absolutely continuous, then there exists a probability density function (pdf) $f : \mathbb{R} \rightarrow \mathbb{R}$. In particular if a random variable X possesses pdf f , then the probability of event $A \subseteq \Omega$ is

$$P(X \in A) = \int_A f(x) dx.$$

5.7 Definition.

The support of a discrete rv X with pmf p is

$$\text{supp}(p) = \{x \in \mathbb{Z} : p(x) > 0\}.$$

The support of a continuous rv X with pdf f is

$$\text{supp}(f) = \{x \in \mathbb{R} : f(x) > 0\}.$$

An equivalent notation for either is $\text{supp}(X)$.

5.8 Remark.

The following definitions sum over \mathbb{Z} or integrate over \mathbb{R} . Since the pmf or pdf is zero off of the support, then we may equivalently sum or integrate over the support.

5.9 Definition.

The expectation of a discrete rv X with pmf p

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

The expectation of a continuous rv X with pdf f

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

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

5.10 Theorem.

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

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

If instead X is a continuous rv with pdf f , then the expectation of $f(X)$ is

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

5.11 Theorem.

Expectation is a linear operator, i.e., for two random variables X, Y and $a, b \in \mathbb{R}$, we have that

$$E(aX + bY) = aE(X) + bE(Y).$$

5.12 Definition.

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

$$E(X^k).$$

5.13 Definition.

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

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

5.14 Definition.

The variance of an rv X is the second central moment

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

5.15 Definition.

The skewness of an rv X is the third central moment

$$E((X - \mu)^3).$$

5.16 Definition.

The kurtosis ('tailedness') of an 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).

5.17 Definition.

The moment generating function of an rv X is

$$M(\theta) = E(e^{\theta X})$$

whenever the expectation exists on an open interval of θ around 0. If the mgf exists, then the k th moment of X is equal to

$$E(X^k) = M^{(k)}(0).$$

5.18 Theorem.

If the moment generating function of two rvs X and Y are equal, then X and Y have the same distribution.

5.19 Lemma.

Suppose that X has mgf $M_X(\theta)$. Then the mgf of $Y = aX + b$ is

$$M_Y(\theta) = e^{b\theta} M_X(a\theta).$$

5.20 Proof.

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

5.21 Definition.

Common distributions are found in the formula sheet [here](#).

5.22 Remark.

Important interpretations of common discrete distributions include the following.

- (1) The Bernoulli distribution models the success of a single trial.
- (2) The binomial distribution models the number of successes in n independent Bernoulli trials.
- (3) The Poisson distribution models the number of events in a fixed interval of time subject to the assumptions of a Poisson process: (1) events occur independently, (2) events occur at a constant rate.
- (4) The geometric distribution models the number of independent Bernoulli trials until the first success.
- (5) The negative binomial distribution models the number of independent Bernoulli trials until the r -th success.
- (6) The hypergeometric distribution models the number of successes in n draws without replacement from a finite population.
- (7) The discrete uniform distribution models the outcomes of a finite set of equally likely outcomes.
- (8) The multinomial distribution models the number of successes in n independent trials with K possible outcomes.

5.23 Remark.

Important interpretations of common continuous distributions include the following.

- (1) The uniform distribution models the outcomes of a continuous set of equally likely outcomes.
- (2) The exponential distribution models the time until the first event in a Poisson process.
- (3) The Gamma distribution models the time until the r -th event in a Poisson process.
- (4) The normal distribution models the sum of a large number of independent and identically distributed random variables.
- (5) The log-normal distribution models the exponential of a normal random variable.
- (7) The beta distribution models r th largest of n independent uniform random variables.

6.0 Introduction.

Joint Distributions.

7.0 Introduction.

Transformations of Random Variables.

8.0 Introduction.

Order Statistics.

9.0 Introduction.

Convergences and Inequalities.