INTRODUCTION to PROBABILITY

Mark Daniel Ward
Ellen Gundlach

Introduction to Probability

Mark Daniel Ward
Associate Professor of Statistics
Purdue University

Ellen Gundlach
Education Specialist and Continuing Lecturer in Statistics
Purdue University

W. H. Freeman and Company
A Macmillan Education Imprint

Publisher: Terri Ward
Senior Acquisitions Editor: Karen Carson
Marketing Manager: Cara LeClair
Developmental Editor: Katrina Mangold
Editorial Assistant: Victoria Garvey
Photo Editor: Cecilia Varas
Cover Designer: Diana Blume
Text Designer: Mark Daniel Ward
Project Editor: Liz Geller
Copyeditor: Elizabeth Marraffino
Illustrations: Mark Daniel Ward; Eli Ensor
Illustration Coordinator: Matthew McAdams
Production Supervisor: Susan Wein
Printing and Binding: Courier Printing
Cover Photo: Matthias Pahl/Shutterstock

Library of Congress Control Number: 2015935799

ISBN-13: 978-0-7167-7109-8
ISBN-10: 0-7167-7109-8

Printed in the United States of America

First printing

W. H. Freeman and Company
41 Madison Avenue
New York, NY 10010
www.macmillanhighered.com

To my wife, Laura,
and to our children:
Bruce, Audrey, Mary, Luke, and Dean.
—MDW

To my boys, Callum and Philip;
to my parents, John and Dianne;
and to Judd.
—EG

this page left intentionally blank

Contents

this page left intentionally blank

Preface

We want to briefly justify why there should be another probability book, when so many others are available.

Motivation: Students from majors in the mathematical sciences and in other areas will be more engaged with the material if they are studying problems that are relevant to them. While testing drafts of the book in the classroom, the students who used this book were asked to contribute questions. As a result, many of the exercises in this text began with questions motivated by the students' own interests.

Example- and exercise-oriented approach: Our book serves as a student's first introduction to probability theory, so we devote significant attention to a wealth of exercises and examples. We encourage students to practice their skills by solving lots of questions. Our exercises are split into practice, extensions, and advanced types of questions. We recommend assigning a small number of questions to students on a daily basis. This promotes more interactive discussion in class between the students and the instructor. It consistently empowers the students to try their hand at some problems of their own. It reduces stress and "cramming" at exam time, as the students consistently develop their understanding during the course. It also provides a firm foundation for the students' long term understanding of probability. The exercises, theorems, definitions, and remarks are all numbered in one list (instead of numbered separately) because we believe that they all should be used in tandem to understand the chapter material.

Relationship between events and random variables: The jump from events to random variables is often a "leap" in other texts. In the present book, we devote significant attention to outcomes, events, sample spaces, and probabilities, before moving on to random variables. Random variables are introduced explicitly as real-valued functions on the sample space, i.e., as functions that depend on the outcome. The notation $X(\omega)$ is used to introduce a random variable at first (where ω is an outcome), so that students can more easily make a transition from studying outcomes to studying random variables that depend on such outcomes.

Jointly distributed random variables: Many probability texts first emphasize properties of one discrete random variable, followed by properties of one continuous random variable, and finally return to jointly distributed random variables. We believe, in contrast, that a firm understanding of jointly distributed random variables, *from the very beginning*, is most helpful for the students' comprehensive understanding of the material. Using jointly distributed random variables at an early stage of the book also allows for more intuitive definitions of some of the concepts. For instance, many probability texts introduce Binomial random variables by explaining the mass, which requires a good grasp of Binomial coefficients, i.e., of $\binom{n}{k}$. We believe, however, that Binomial random variables are introduced more intuitively by $X = X_1 + \cdots + X_n$, where the X_j's are indicator random variables. This requires the students to have familiarity with more than one random variable at a time (i.e., to understand joint distributions), but it allows the students to be more versatile in their thinking. It promotes the understanding of "big picture" kinds of insights. Students begin to think, from the very start, about the ways in which random variables are related, and the ways in which collections of random variables can give rise to new random variables.

Counting: Many probability books start with counting, which means this material is taught during the first two or three weeks of a course. Unfortunately, this means that a student who is weighing her/his interest in a course will not even begin to grasp the concepts of randomness until the registration period is over. This leads to attrition. Moreover, some questions in counting are best understood from a probabilistic point of view, using (for instance) the linearity of expectation, which is not available to the students at the start of the course. As an example, consider how many couples are expected to sit together, when people sit uniformly at random in a circle. This question can be answered very succinctly, using indicator functions and the linearity of expectation. (The probability mass function, in contrast, is cumbersome to compute.) In general, we believe that our approach to counting is significantly enhanced by the use of sums of indicator random variables. Therefore, we focus on counting after we finish a thorough treatment of discrete random variables, but before moving onwards to continuous random variables. This allows the students to feel confident in their understanding of the discrete world before they tackle difficult counting questions. We emphasize to students that combinatorics is a deep and beautiful subject (much of the Ward's research is motivated by problems in applied discrete mathematics). We also try to emphasize the connections between discrete random variables and counting. We firmly believe that this is best accomplished when the students already understand discrete random variables.

Comparison/summary chapters: A first course in probability theory can feel like a whirlwind tour. Thus, we have checkpoints throughout the book, where material is summarized and reviewed. This helps to ground the reader and build confidence. It also helps the students discriminate between the commonly confused distributions and counting techniques, mass functions vs. CDFs, etc. These summaries are useful while reviewing for examinations, e.g., the Actuarial P/1 exam given by the Society of Actuaries and the Casualty Actuarial Society. Passing the P/1 exam requires knowledge of all the material in this text. We also guide the students through ways to tell which kind of distribution they are working with. We give suggestions about how to grasp nuances in a problem that make it a Binomial or Geometric or Negative Binomial situation. These guides help students home in on what separates these distributions. We summarize each distribution for quick reference, but we also go into details to explain each discrete distribution's mass, expected value, variance, etc.

Our students have had an excellent experience using this probability book. Several of our own students have already passed the SOA/CAS P/1 exam after having learned probability using only the early drafts of this book. Our students seem to enjoy the many examples and friendly tone. We hope that you and your students also find our book approachable and thorough. We are delighted by the kind reception that our students and colleagues have given to the book during its pilot tests.

We have divided our book into seven main parts:

Part I: Randomness. We introduce outcomes, events, sample spaces, basic probability rules, independence, conditional probabilities, and Bayes' Theorem.

Part II: Discrete Random Variables. We discuss the difference between discrete and continuous random variables and introduce probability mass function, cumulative distribution function, expected value, variance, and joint distributions for discrete random variables.

Part III: Named Discrete Random Variables. We consider ways to distinguish between—and perform calculations with—the most common discrete random variables: Bernoulli, Binomial, Geometric, Negative Binomial, Poisson, Hypergeometric, and Discrete Uniform. We include a review chapter to help students see the similarities and differences between all of these distributions.

Part IV: Counting. We use indicator variables and the linearity of expectation as tools to help tackle several different types of counting problems: sampling with and without replacement; when order matters and doesn't matter; and rearrangement problems. We have case studies on poker and Yahtzee, two popular games many students will recognize.

Part V: Continuous Random Variables. We reinforce the difference between discrete and continuous random variables. Then we introduce the probability density function, cumulative distribution function, expected value, variance, and joint and conditional distributions for continuous random variables.

Part VI: Named Continuous Random Variables. We show ways to utilize (and quickly make distinctions between) the most common continuous random variables: Continuous Uniform, Exponential, Gamma, Beta and Normal. We show how the Central Limit Theorems and Laws of Large Numbers work. We include a review chapter to help students see the similarities and differences between all of these continuous distributions and between some of the continuous and discrete distributions.

Part VII: Additional Topics. Here we cover more advanced topics that could be optional, depending on how much time an instructor has in a semester or quarter. We treat the distribution of a function of one continuous random variable, the variance of sums of random variables, correlation, conditional expectation, Markov and Chebyshev Inequalities, order statistics, moment generating functions, and the joint density of two random variables that are functions of another pair of random variables.

Acknowledgements

We thank our students for their feedback, as well as their contributions to the exercises in this text. These students are:

Probability: The Science Of Uncertainty (HONR 399, Fall 2010) J. Blair, J. Covalt, S. Fancher, C. Fleming, B. Goosman, W. Hess, E. Hoffman, C. Holcomb, A. Hurlock, E. Jenkins, J. Ling, D. Mo, B. Morgan, C. Mullen, S. Mussmann, D. Rouleau, C. Sanor, V. Savikhin, S. Sheafer, S. Spence, R. Stevick, J. Wu, S. Yap, M. Zachman, B. Zhou.

Probability (MA/STAT 416, Fall 2011) K. Amstutz, C. Ben, M. Boing, K. Breneman, P. Coghlan, B. Copeland, O. El-Ghirani, M. Fronek, Z. Gao, A. Gerardi, T. He, X. He, K. Hopkins, J. Jaagosild, N. Johnson, K. Kemmerling, J. Kwong, N. Lawrence, K. Leow, B. Lewis, K. Li, S. Li, Y. Liang, E. Luo Cao, M. Mazlan, J. Milligan, E. Myers, Y. Nazarbekov, C. Ng, R. Schwartz, S. Scott, S. Seffrin, J. Sheng, D. Snyder, T. Sun, B. Wilson, E. Winkowska, R. Xue, Y. Yang, G. Zhao.

We heartily thank our colleague Dr. Frederi Viens (Purdue University) for piloting the book in his STAT/MA 416 course in Fall 2011 and again in Spring 2013. We extend our thanks also to Dr. David Galvin (University of Notre Dame), Dr. Patricia Humphrey (Georgia Southern University), Dr. Hosam Mahmoud (The George Washington University), and Dr. Meike Neiderhausen (University of Portland), for their willingness to test our book in Fall 2012 in several different types of courses at their own universities. These colleagues and their students provided very valuable feedback about early drafts of the book. We benefited a great deal from their guidance, insight, and suggestions.

This book was genuinely a team effort. We are very grateful to Deborah Sutton for preparing the answers to the odd-numbered exercises. Dr. Patricia Humphrey provided numerous insights about all aspects of the content and checked the accuracy of the entire book. Pat also produced the solutions in the student and instructor manuals. We appreciate Jackie Miller's insights and ideas at the beginning of this project. We are thankful for the support of Terri Ward, our publisher at W. H. Freeman, and for the contributions of the entire W. H. Freeman team. In particular, Liz Geller and Elizabeth Marraffino provided hundreds of helpful suggestions for improvement. We are also especially thankful for the marketing expertise of Karen Carson, Cara LeClair, and Katrina Mangold. We are grateful to Diana Blume and Matt McAdams for their help with the design, layout, and art, and to Susan Wein for making sure our files were technically sound and printer-ready.

We are also thankful for the feedback of the following reviewers:

David Anderson, *University of Wisconsin*
Alireza Arasteh, *Western New Mexico University*
G. Jogesh Babu, *Pennsylvania State University*
Christian Beneš, *Brooklyn College of the City University of New York*
Kiran R. Bhutani, *The Catholic University of America*
Ken Bosworth, *Idaho State University*
Daniel Conus, *Lehigh University*
Michelle Cook, *Stephen F. Austin State University*
Dana Draghicescu, *City University of New York, Hunter College*
Randy Eubank, *Arizona State University*
Marian Frazier, *Gustavus Adolphus College*
Rohitha Goonatilake, *Texas A&M University*
Patrick Gorman, *Kutztown University*
Ross Gosky, *Appalachian State University*
Marc Goulet, *University of Wisconsin–Eau Claire*
Susan Herring, *Sonoma State University*
Christopher Hoffman, *University of Washington*
Patricia Humphrey, *Georgia Southern University*
Ahmad Kamalvand, *Huston-Tillotson University*
Judy Kasabian, *El Camino College*
Syed Kirmani, *University of Northern Iowa*
Charles Lindsey, *Florida Gulf Coast University*
Susan Martonosi, *Harvey Mudd College*
Ronald Mellado Miller, *Brigham Young University–Hawaii*
Joseph Mitchell, *Stony Brook University*
Sumona Mondal, *Clarkson University*
Cindy Moss, *Skyline College*
Meike Niederhausen, *University of Portland*
Will Perkins, *Georgia Institute of Technology*
Rebecca L. Pierce, *Ball State University*
David J. Rader, Jr., *Rose-Hulman Institute of Technology*
Aaron Robertson, *Colgate University*
Charles A. Rohde, *Johns Hopkins University*
Seyed Roosta, *Albany State University*
Juana Sanchez, *University of California, Los Angeles*
Yiyuan She, *Florida State University*
Therese Shelton, *Southwestern University*
A. Robert Sinn, *University of North Georgia*
Clifton D. Sutton, *George Mason University*
Mahbobeh Vezvaei, *Kent State University*

Frederi Viens, *Purdue University*
P. Patrick Wang, *University of Alabama*
Daniel Weiner, *Boston University*
Alison Weir, *University of Toronto, Mississauga*
Reem Yassawi, *Trent University*
Norbert Youmbi, *Saint Francis University*

M. D. Ward's work is supported by the National Science Foundation grants #0939370, #1140489, and #1246818.

We welcome readers' feedback. Please contact us at `wgbook@purdue.edu`. We will maintain a list of all the corrections and suggestions sent to us, at `http://www.stat.purdue.edu/~mdw/book.html`.

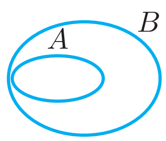
Notation Review

Notation for Named Sets of Numbers:

$\mathbb{Z}^{\geq 0}$	the set of nonnegative integers $\mathbb{Z}^{\geq 0} = \{0, 1, 2, \ldots\}$
$\mathbb{Z}$	the set of integers, $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$
$\mathbb{N}$	the natural (a.k.a. counting) numbers, $\mathbb{N} = \{1, 2, \ldots\}$
$\mathbb{R}^{>0}$	the set of positive real numbers
$\mathbb{R}^{\geq 0}$	the set of nonnegative real numbers
$\mathbb{R}$	the set of real numbers, including positive and negative numbers, i.e., whole numbers, fractions, decimals, roots (although not imaginary), and transcendental numbers like π and e.

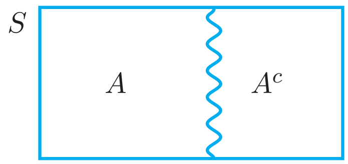
Notation for Events:

sets:	{things in the set | conditions on those things}; e.g., $S = \{(x, y) \mid x + y = 2\}$.
$\emptyset$	the empty set, i.e., event with no outcomes e.g., $A \cap B = \emptyset$ if A, B have no outcomes in common; see the definition of $\cap$ below to clarify further
S	the sample space, i.e., event with all outcomes
ω	an outcome in the sample space
$\in$	inclusion in an event (i.e., $x \in A$ if outcome x is in event A)
$\subset$	subset (i.e., $A \subset B$ if every outcome of A is also in B)
$\lvert A \rvert$	the number of outcomes (also called the size) of event A



Notation for Building New Events Using Known Ones:

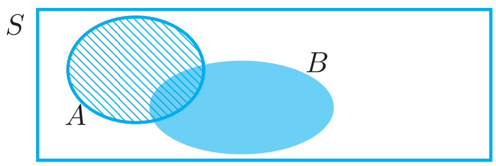
^{c} complement of an event, which corresponds with the word "not"
i.e., x is in A^c exactly when x is not in A



$\setminus$ setminus (i.e., $B \setminus A = B \cap A^c$;
the event containing outcomes in B that are not in A)

$\cup$ union of events, which corresponds with the word "or,"
i.e., $A \cup B$ is the event containing outcomes in A or B or both.
The set $A \cup B$ corresponds to the region containing
shading, lines, or both, in the figure below.

$\cap$ intersection of events, which corresponds with the word "and,"
i.e., $A \cap B$ is the event containing outcomes in A *and* B.
The set $A \cap B$ is the region that contains lines
overlapping the shading in the figure below.



Notation for Random Variables:

X	a random variable (always written in capital letters)
x	a value that a random variable might take on
$P(A)$	probability that event A occurs
$P(X = x)$	is a shorthand notation for $P(\{\omega \mid X(\omega) = x\})$
$\mathbb{E}(X) = \mu$	expected value of X
$\mathrm{Var}(X) = \sigma^2$	variance of X

Math Review

Geometric Sums

For $-1 < a < 1$, recall these two finite summations of geometric terms,

$$1 + a + a^2 + a^3 + \cdots + a^r = \sum_{j=0}^{r} a^j = \frac{1 - a^{r+1}}{1 - a}$$

and

$$a + a^2 + a^3 + a^4 + \cdots + a^r = \sum_{j=1}^{r} a^j = \frac{a - a^{r+1}}{1 - a}.$$

For $-1 < a < 1$, these yield two infinite summations of geometric terms,

$$1 + a + a^2 + a^3 + \cdots = \sum_{j=0}^{\infty} a^j = \frac{1}{1 - a}$$

and

$$a + a^2 + a^3 + a^4 + \cdots = \sum_{j=1}^{\infty} a^j = \frac{a}{1 - a}.$$

Exponential Function

For any real-valued x, the power series definition of the exponential function evaluated at x is

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!},$$

where $n! := (1)(2) \cdots (n)$.

Sum of Integers, Sum of Squares

It is also helpful to know that, for positive integers n,

$$1 + 2 + \cdots + n = \frac{(n)(n+1)}{2}$$

and

$$1^2 + 2^2 + \cdots + n^2 = \frac{(n)(n+1)(2n+1)}{6}$$

Floor and Ceiling Functions

$\lfloor x \rfloor$ round x down to the closest, lower integer, e.g., $\lfloor 14.37 \rfloor = 14$

$\lceil x \rceil$ round x up to the closest, higher integer, e.g., $\lfloor 14.37 \rfloor = 15$

Binomial Coefficient

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$ number of ways to choose k out of n objects, when the order of selection does not matter, e.g., $\binom{5}{3} = 10$ since there are 10 ways to choose 3 out of 5 objects:

1	**2**	**3**	4	5
1	**2**	3	**4**	5
1	**2**	3	4	**5**

1	2	**3**	**4**	5
1	2	**3**	4	**5**
1	2	3	**4**	**5**

1	**2**	**3**	**4**	5
1	**2**	**3**	4	**5**
1	**2**	3	**4**	**5**
1	2	**3**	**4**	**5**

Gamma Function

$\Gamma(n) = (n-1)!$ The gamma function extends the notion of factorials beyond the set of nonnegative integers. We will only use this one fact about the gamma function in this text.

Double Integration

$\int_a^b \int_c^d f(x,y)\,dy\,dx$ This denotes the double integral of $f(x,y)$ over the range where $a \leq x \leq b$ and $c \leq y \leq d$. Under most conditions used in this book, the order of integration can often be switched. (In more advanced courses, this must be done with caution, and suitable convergence theorems must hold—but we do not cover such topics in this text.) If the order of integration is switched, we have instead $\int_c^d \int_a^b f(x,y)\,dx\,dy$. We urge students to make sure that the outer integral corresponds to the outer variable, and the inner integral corresponds to the inner variable.

Dice

We assume that all dice in this text are six-sided, numbered 1, . . . , 6, unless stated otherwise.

Cards

We assume that all decks of playing cards have 52 cards, consisting of four suits (spades ♠, hearts ♡, diamonds ♢, clubs ♣), 13 values each: A, 2, 3, . . . , 10, J, Q, K.

Part I

Randomness

At the beginning of this course in probability, we consider the basic aspects of randomness. We first discuss the outcomes that are possible when something random occurs, with an emphasis on how these outcomes can be collected into events. Then we give the basic, fundamental notions of probability theory. These are very simple to state, but they constitute the groundwork on which the rest of our study of probability theory is based. We also consider independence of events, as well as the way that the occurrence of one event will affect the probability of occurrence of another event. The first part of the book concludes with the introduction of Bayes' Theorem, which allows us to manipulate probabilities and conditional probabilities.

Even small children learn basic probability ideas simply from observing the world around them. We will formalize these probability ideas and introduce mathematical calculations to go with them. Drawing pictures to help visualize the information in the stories is strongly recommended.

By the end of this part of the book, you should be able to:

1. Define basic terms related to probability and events.
2. Use proper set notation for events.
3. Characterize the possible outcomes, when something random occurs.
4. Describe the events into which outcomes can be grouped.
5. Assign probabilities to events, and perform calculations using probability rules.
6. Calculate whether two or more events are independent.
7. Calculate the probability of an event occurring, given that another event occurred.
8. Calculate the conditional probability of an event using Bayes' Theorem.

Math skills you will need: basic understanding of set notation, unions, intersections, and summation $\sum$ notation.

Additional resources: Calculators may be used to assist in the calculations. Colored pencils may be helpful for drawing Venn diagrams clearly.

Chapter 1

Outcomes, Events, and Sample Spaces

On Monday in math class, Mrs. Fibonacci says, "You know, you can think of almost everything as a math problem." On Tuesday I start having problems.
—*Math Curse* by Jon Scieszka and Lane Smith (Viking, 1995)

In a National Public Radio story from November 30, 2012, "That's So Random: The Evolution of an Odd Word," Neda Ulaby writes about the many misuses of the word "random" in our modern culture, including snippets from the comedian Spencer Thompson's routine, "I Hate When People Misuse the Word Random." For example, Thompson explains that if your friends talk about a "random party" they went to, it probably wasn't as random is they think since it was likely to be held within a reasonably small community and planned with some people that your friends already knew. What do mathematicians and statisticians mean by the word "random"?

1.1 Introduction

Probability theory is the study of randomness and all things associated with randomness. Examples abound everywhere. From the time that we are children, we play guessing games, roll dice, and flip coins. We frequently encounter the unknown and the uncertain. We turn on an mp3 player in a "shuffle" mode, or listen to the radio, eagerly waiting to see what song will come on next. The time until an something happens is often random, e.g., until a traffic light turns green, an email arrives, the telephone rings, or a text message buzzes. The sex of a baby remains unknown until birth (or an ultrasound). An athlete runs a race, but the exact finishing time is unknown beforehand. Millions of people play lotteries and other games of chance, often wagering large amounts of money. Throughout this book, we study probability using examples that will be familiar to the reader.

Definition 1.1. When something happens at **random** there are several potential **outcomes**. *Exactly one* of these outcomes occurs. An **event** is defined to be a collection of some outcomes.

Even though the empty set never happens, we will need it to understand **disjoint** events, i.e., events that have no outcome in common.

Two extreme events have names: The **empty set** $\emptyset$ consists of no outcomes (so the empty set never happens). The **sample space** S consists of all outcomes (so the sample space always happens).

Example 1.2. You roll a 6-sided die.

The sample space is $S = \{1, 2, 3, 4, 5, 6\}$. Only one of these six outcomes actually occurs; for instance, 2 is a possible outcome, or 5 is a possible outcome, etc. We cannot "solve" for which outcome occurs because, as we know from practical experience, we do not know (in advance) which outcome will occur. The outcome is random.

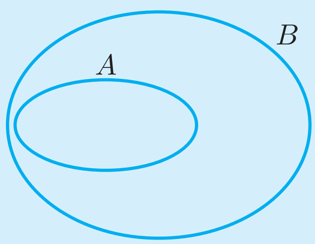
(Q: How many events are there altogether? Hint: It's a power of 2.)

One event is $\{1, 3, 5\}$, i.e., the event that the outcome is odd. The event that the roll is 2 or higher is $\{2, 3, 4, 5, 6\}$. The event that 4 does not appear is $\{1, 2, 3, 5, 6\}$. The event $\{3\}$ consists of only one outcome. Event $\{1, 6\}$ has the smallest and largest possible outcomes.

One event is a **subset** of another if every outcome from the first event is contained in the second event too. Subsets are denoted with the "$\subset$" symbol. For instance, an event with one outcome (such as 5) is a subset of a larger event (such as $\{1, 2, 5\}$), which is a subset of the sample space:

$$\{5\} \subset \{1, 2, 5\} \subset \{1, 2, 3, 4, 5, 6\}.$$

Definition 1.3. Event A is a **subset** of event B, written $A \subset B$, if every outcome in A is also an outcome in B.



Example 1.4. A student buys a book and opens it to a random page. He notes the number of typographical errors on the page.

The sample space is $S = \mathbb{Z}^{\geq 0}$, i.e., the set of nonnegative integers.

The event that the page contains at most 2 errors is $\{0, 1, 2\}$.

Example 1.5. A new mother delivers one baby.

The sample space is $S = \{\text{boy}, \text{girl}\}$. Although there is just one baby, we can describe four events:

$$\emptyset, \qquad \{\text{boy}\}, \qquad \{\text{girl}\}, \qquad \{\text{boy}, \text{girl}\} = S.$$

Example 1.6. A new mother delivers at least one baby.

One possible outcome is that the the mother has triplets, which are all girls; we denote this outcome as (g, g, g). If she delivers a boy and then a girl, the outcome is (b, g). So the sample space is

$$\begin{aligned} S = \{&(b), (g), \\ &(b, b), (b, g), (g, b), (g, g), \\ &(b, b, b), (b, b, g), (b, g, b), (b, g, g), (g, b, b), (g, b, g), (g, g, b), (g, g, g), \ldots\}. \end{aligned}$$

Note: A new mother may have a single baby, twins, triplets, octuplets or any other (relatively small) number of babies at one time. We only listed the possibilities up to triplets explicitly, but the other possibilities are included in S too; hence, the "…" at the end of S.

A set of octuplets (8 babies) was born in 1998 and also in 2009 in the United States.

Let A be the event that the mother has at least one boy and at least one girl. So A does not contain the outcomes (b) or (b, b) or (b, b, b) etc., and does not contain the outcomes (g) or (g, g) or (g, g, g) etc. Thus

$$A = \{(b, g), (g, b), (b, b, g), (b, g, b), (b, g, g), (g, b, b), (g, b, g), (g, g, b), \ldots\}.$$

Example 1.7. You wait at a red traffic light and record the time (in seconds) until the light turns green.

The sample space is the set of all positive real numbers, $\mathbb{R}^{>0}$. One event is $[5, 10]$, the event consisting of all outcomes between 5 and 10 seconds (inclusive). Another event is $(12.7, \infty)$, i.e., the waiting time is strictly more than 12.7 seconds. Another event is $\{32.7\}$ seconds, the event consisting of only the outcome 32.7 seconds. Events can be built using unions and intersections, e.g., $(0, 60) \cup (120, 180)$ is the event consisting of all outcomes less than 1 minute and also consisting of all outcomes of 2 to 3 minutes.

Example 1.8. You notice the color of the next car to pass on the street.

The sample space is the set of all possible colors in the scheme used to classify this car's color, for instance, perhaps it is classified according to the sample space

$$S = \{\text{red}, \text{yellow}, \text{green}, \text{blue}, \text{orange}, \text{silver}, \text{brown}, \text{black}, \text{white}, \text{other}\}.$$

As we see in the examples with the baby's sex or car's color, outcomes do not have to be numbers.

At the most fundamental level, it is essential to consider how we classify the outcomes. There are often several valid viewpoints. As an example:

Example 1.9. We hit or miss the bullseye with a dart (two possible outcomes).

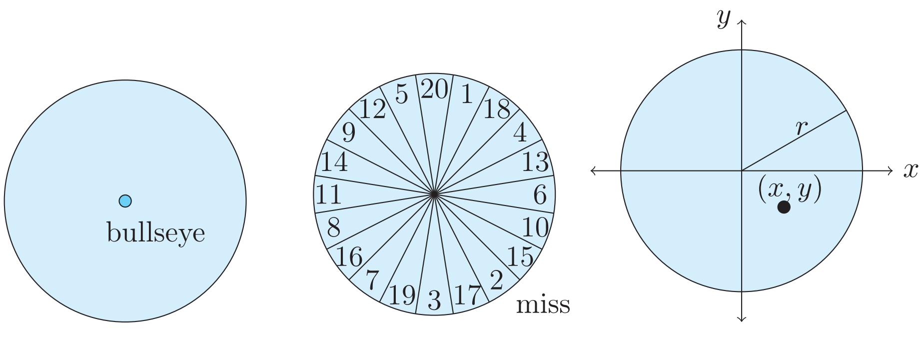


Figure 1.1: Different sample spaces for a dart throw. Left: Two outcomes in the sample space. Middle: Twenty-one outcomes in the sample space (the 21st outcome denotes missing the board altogether). Right: Sample space consists of the outcomes, according to location, given as coordinates.

The sample space is $S = \{\text{hit}, \text{miss}\}$. This is depicted on the left side of Figure 1.1. There are four events:

$$\emptyset, \qquad \{\text{hit}\}, \qquad \{\text{miss}\}, \qquad \{\text{hit}, \text{miss}\} = S.$$

(The empty set never happens because it has no outcomes. Sometimes event $\{\text{hit}\}$ happens; sometimes event $\{\text{miss}\}$ happens. Event $\{\text{hit}, \text{miss}\} = S$ always happens.)

Example 1.9 (continued) When throwing a dart, we hit one of twenty regions, or we miss the entire board (twenty-one possible outcomes). Notice: this classification of the outcomes is very different than the "hit" or "miss" setup.

The sample space is $S = \{\text{miss}, 1, 2, 3, \ldots, 20\}$, consisting of the twenty-one possible outcomes: either we "miss" the board altogether or we hit one of the 20 specified regions. This is depicted in the middle of Figure 1.1. (The board has metal ridges between the regions, so that a dart cannot land exactly on the boundary of two regions.)

Example 1.9 (continued) When throwing a dart, we note the exact location where the dart lands.

The sample space is

$$S = \{(x, y) \mid x, y \in \mathbb{R}\},$$

consisting of the outcome listed according to the x (horizontal) and y (vertical) locations where it landed (using the origin at the center of the dartboard as a reference point). This is depicted on the right side of Figure 1.1.

The set notation we used for S is nice, because we are unable to list *all* of the possible points, in this last version of Example 1.9; in fact, there are infinitely many points in the sample space $S = \{(x, y) \mid x, y \in \mathbb{R}\}$. Set notation is also nice because we can add other conditions to the allowable points. For instance, if we want to consider only the situation in which the dart hits the dartboard, then the sample space could be narrowed to

$$\{(x, y) \mid x^2 + y^2 \leq r^2\},$$

where r is the radius of the dartboard. (In this case, we have not handled darts that miss the board entirely.) For instance, if $r = 9$ inches, then the sample space includes outcomes such as $(x, y) = (3.6, -1.35)$, etc.

Set notation is a way to describe a collection of things, sometimes using an annotation about conditions on these things (maybe it should be called "set annotation"). The things that are inside the set are written on the left, and any conditions about these things go on the right.

Notation 1.10. The notation for a set uses braces, with the contents of the set, often followed by a line and then any conditions on the contents of the set.

$$\left\{ \begin{matrix} \text{things} \\ \text{in the set} \end{matrix} \;\middle|\; \begin{matrix} \text{conditions on} \\ \text{these things} \end{matrix} \right\}$$

The dartboard example should illustrate that it is really important to understand what outcomes are possible when something random happens. This takes some practice at the outset. Sometimes it is helpful to write a list—even an incomplete list—of the different outcomes that are possible from a random phenomenon. (As a rule of thumb, we often encourage students to write five different possible outcomes, if the problem is complicated, just to develop some intuition.) With the darts in Example 1.9, we can certainly write down both outcomes in the first scenario, i.e., "bullseye" or "not bullseye." In the second scenario, the list of all possible outcomes would be "miss," 1, 2, 3, ..., 20. In the third scenario, as soon as we begin to try to write down all of the possible locations on the board by their (x, y) coordinates, we quickly realize that this is a hopeless task. It will not be possible for us to write down every potential outcome, so the concise set notation is crucial to use.

Definition 1.11. We use the **union** notation "$\cup$" when a new set is formed that contains each outcome found in any of the component events. E.g., $A \cup B$ contains each outcome that is found in A, or in B, or in both.

Definition 1.12. We use the **intersection** notation "$\cap$" to construct a new event that contains only the outcomes found in all of the components. E.g., $A \cap B$ contains each outcome that is found in both A and B; it is insufficient to be in just one of these sets.

Example 1.13. A student shuffles a deck of cards thoroughly (one time) and then selects cards from the deck *without replacement* until the ace of spades appears.

"**Without replacement**" means that the cards are not put back into the deck after they are drawn. So on the first draw there are 52 cards available, but on the second draw there are only 51 cards available, and 50 cards available on the third draw, etc. So the ace of spades is certain to appear sometime during the 52 draws. Also, because they are selected without replacement, the chosen cards will be distinct.

The event that exactly three draws are needed to see the ace of spades is

$$\{(x_1, x_2, x_3) \mid x_3 = \mathbf{A}\spadesuit, \text{ and the } x_j\text{'s are distinct}\}.$$

The sample space S consists of all possible draws of distinct cards that end with the ace of spades:

$$\begin{aligned} S = \{(\mathbf{A}\spadesuit)\} &\cup \{(x_1, x_2) \mid x_2 = \mathbf{A}\spadesuit, \text{ and the } x_j\text{'s are distinct}\} \\ &\cup \{(x_1, x_2, x_3) \mid x_3 = \mathbf{A}\spadesuit, \text{ and the } x_j\text{'s are distinct}\} \\ &\cup \{(x_1, x_2, x_3, x_4) \mid x_4 = \mathbf{A}\spadesuit, \text{ and the } x_j\text{'s are distinct}\} \\ &\vdots \\ &\cup \{(x_1, x_2, \ldots, x_{52}) \mid x_{52} = \mathbf{A}\spadesuit, \text{ and the } x_j\text{'s are distinct}\}. \end{aligned}$$

Equivalently, if $B_k = \{(x_1, x_2, \ldots, x_k) \mid x_k = \mathbf{A}\spadesuit, \text{ for distinct } x_j\text{'s}\}$, then the sample space is $S = \bigcup_{k=1}^{52} B_k$.

Example 1.14. A student draws cards from a standard deck of playing cards until the ace of spades appears. After every unsuccessful draw, the student replaces the card and shuffles the deck thoroughly before selecting a new card.

The set of outcomes in which the ace of spades *first appears* on the kth draw is

$$B_k = \{(x_1, \ldots, x_k) \mid \text{only } x_k \text{ is } \mathbf{A}\spadesuit\}$$

Notice that we dropped the condition about the cards being distinct.

The set of all possibilities in which the student actually finds the ace of spades is $\bigcup_{j=1}^{\infty} B_k$. The astute reader will notice that we did not yet mention the possibility that the aces of spades never appears. We write this event as

Since the cards are replaced after each draw, this scenario is quite different from Example 1.13.

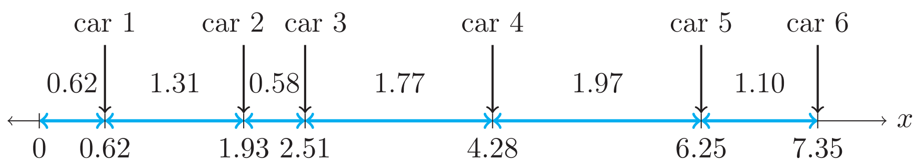
$$C = \{(x_1, x_2, x_3, \ldots) \mid \text{none of the } x_k\text{'s is } \mathbf{A}\spadesuit\}.$$

So the entire sample space is

$$S = \Big(\bigcup_{k\geq 1} B_k\Big) \cup C.$$

Example 1.15. A traffic engineer records times (in seconds) between the next six cars that pass.

For example, consider when the next six cars arrive:



The sample space is

$$S = \{(x_1, \ldots, x_6) \mid x_j \in \mathbb{R}^{>0} \text{ for each } j\}.$$

The engineer uses x_1 for the time until the first car passes, and x_2 is the time between the first and second cars, and in general, x_j is the time between $(j-1)$st and jth cars.

The event where there are at least 3 seconds between all pairs of consecutive cars is

$$\{(x_1, \ldots, x_6) \mid x_j \geq 3 \text{ for each } j\}.$$

The event in which the cars have consecutively longer and longer inter-arrival times (i.e., the distance between cars 1 and 2 is shorter than the distance between cars 2 and 3, which is shorter than the distance between cars 3 and 4, etc.), is

$$\{(x_1, \ldots, x_6) \mid x_j < x_{j+1} \text{ for each } j\}.$$

Many other possible events can be written. The possibilities are endless.