

THE ART OF PROBABILITY

*for Scientists
and Engineers*

RICHARD W. HAMMING



CRC Press
Taylor & Francis Group

A CHAPMAN & HALL BOOK

ADVANCED BOOK

Classics

The Art of Probability

FOR SCIENTISTS AND ENGINEERS

Richard W. Hamming

U. S. Naval Postgraduate School



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A CHAPMAN & HALL BOOK

ABP

Advanced Book Program

Text Design: *Peter Vacek*
Cover Design: *Iva Frank*

First published 1991 by Westview Press

Published 2018 by CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

CRC Press is an imprint of the Taylor & Francis Group, an informa business

Copyright © 1991 by Taylor & Francis Group LLC

No claim to original U.S. Government works

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

Library of Congress Cataloging-in-Publication Data

Hamming, R. W. (Richard Wesley), 1915–

The art of probability—for scientists and engineers/Richard W. Hamming.

p. cm.

Includes index.

1. Probabilities. I. Title.

QA273.H3544 1991

519.2–dc20 90-42240

ISBN 0-201-51058-8 (H) CIP

ISBN 0-201-40686-1 (P)

This book was typeset using the \TeX typesetting language on IBM Compatible computer

ISBN 13: 978-0-201-40686-3 (pbk)

Table of Contents

Preface	v
Table of Contents	xi
References	339
Chapter 1 Probability	1
1.1 Introduction	1
1.2 Models in General	4
1.3 The Frequency Approach Rejected	6
1.4 The Single Event Model	7
1.5 Symmetry as the Measure of Probability	9
1.5-1 Selected faces of a die	10
1.5-2 A card with the face 7	10
1.5-3 Probability of a spade	11
1.6 Independence	12
1.6-1 The sample space of two dice	13
1.6-2 The sample space of three coins	13
1.6-3 The number of paths	13
1.7 Subsets of a Sample Space	14
1.7-1 Exactly one head in three tosses of three coins	15
1.7-2 Sum of two dice	15
Aside: Probability scales	16
1.8 Conditional Probability	18
1.8-1 At least two heads in ten tosses	18
1.8-2 An even sum on two dice	19
1.8-3 Probability of exactly three heads given that there are at least two heads	19

1.9 Randomness	20
1.9-1 The two gold coin problem	23
1.9-2 The two children problem	24
1.9-3 The four card deck	25
1.9-4 The two headed coin	26
1.9-5 No information	27
1.9-6 The birthday problem	27
1.9-7 The general case of coincidences	30
1.9-8 The information depends on your state of knowledge	31
1.10 Critique of the Model	34
1.A Bounds on Sums	35
1.B A Useful Bound	40
Chapter 2 Some Mathematical Tools	41
2.1 Introduction	41
2.2 Permutations	43
2.2-1 Arrange 3 books on a shelf	45
2.2-2 Arrange 6 books on a shelf	45
2.2-3 Arrange 10 books on a shelf	45
2.2-4 Two out of three items identical	45
2.2-5 Another case of identical items	46
2.3 Combinations	48
2.3-1 Sums of binomial coefficients	51
2.3-2 Bridge hands	51
2.3-3 3 out of 7 books	52
2.3-4 Probability of n heads in $2n$ tosses of a coin	52
2.3-5 Probability of a void in bridge	53
2.3-6 Similar items in bins	53
2.3-7 With at least 1 in each bin	53
2.4 The Binomial Distribution—Bernoulli Trials	55
2.4-1 Inspection of parts	58
2.4-2 Continued	59
2.4-3 Floppy discs	59
2.4-4 Distribution of the sum of three dice	60
2.4-5 Bose-Einstein statistics	62
2.4-6 Fermi-Dirac statistics	62
2.5 Random Variables, Mean and the Expected Value	64
2.5-1 Expected value of a roll of a die	65
2.5-2 Expected value of a coin toss	65
2.5-3 Sum of two dice	66
2.5-4 Gambling on an unknown bias	70
Mathematical aside: Sums of powers of the integers	70

2.6 The Variance	72
2.6-1 Variance of a die	75
2.6-2 Variance of M equally likely outcomes	76
2.6-3 Variance of the sum of n dice	76
2.7 The Generating Function	78
2.7-1 Mean and variance of the binomial distribution	80
2.7-2 The binomial distribution again	82
2.7-3 The distribution of the sum of three dice	82
2.8 The Weak Law of Large Numbers	84
2.9 The Statistical Assignment of Probability	88
2.9-1 Test of coin and Monte Carlo methods	89
2.9-2 Chevalier de Mere problem	90
2.10 The Representation of Information	91
2.11 Summary	93
2.A Derivation of the Weak Law of Large Numbers	94
2.B Useful Binomial Identities	95
Chapter 3 Methods for Solving Problems	97
3.1 The Five Methods	97
3.1-1 The elevator problem (five ways)	98
3.2 The Total Sample Space and Fair Games	103
3.2-1 Matching pennies (coins)	103
3.2-2 Biased coins	104
3.2-3 Raffles and lotteries	105
3.2-4 How many tickets to buy in a raffle	106
3.3 Enumeration	108
3.3-1 The game of six dice	108
3.3-2 The sum of three dice	111
3.4 Historical Approach	112
3.4-1 The problem of points	113
3.4-2 Three point game with bias	113
3.4-3 Estimating the bias	114
3.5 Recursive Approach	115
3.5-1 Permutations	115
3.5-2 Runs of heads	116
3.5-3 N biased coins	116
3.5-4 Random distribution of chips	117
3.5-5 The gambler's ruin	119
3.6 The Method of Random Variables	121
3.6-1 Pairs of socks	121
3.6-2 Problem de recontre	123
3.6-3 Selecting N items from a set of N	124
3.7 Critique of the Notion of a Fair Game	125

3.8 Bernoulli Evaluation	126
Mathematical aside: Log Expansions	126
3.8-1 Coin toss	127
3.8-2 Symmetric payoff	127
3.8-3 Fair games	128
3.8-4 Insurance (ideal—no extra costs)	129
3.9 Robustness	130
3.9-1 The robust birthday problem	131
3.9-2 The robust elevator problem	132
3.9-3 A variant on the birthday problem	134
3.9-4 A simulation of the variant birthday problem	135
3.10 Inclusion–Exclusion Principle	136
3.10-1 Misprints	137
3.10-2 Animal populations	138
3.10-3 Robust multiple sampling	138
3.10-4 Divisibility of numbers	139
3.11 Summary	140
Chapter 4 Countably Infinite Sample Spaces	141
4.1 Introduction	141
Mathematical aside: Infinite Sums	141
4.2 Bernoulli Trials	145
4.2-1 First occurrence	145
4.2-2 Monte Carlo test of a binomial choice	148
4.2-3 All six faces of a die	149
4.2-4 Monte Carlo estimate	151
4.2-5 Card collecting	151
4.2-6 Invariance principle of geometric distributions	152
4.2-7 First and second failures	153
4.2-8 Number of boys in a family	154
4.2-9 Probability that event A precedes event B	156
4.2-10 Craps	156
4.3 On the Strategy to be Adopted	158
4.4 State Diagrams	159
4.4-1 Two heads in a row	160
4.4-2 Three heads in a row	165
4.5 Generating Functions of State Diagrams	172
4.5-1 Six faces of a die again	174
4.6 Expanding a Rational Generating Function	177
4.7 Checking the Solution	179
4.7-1 Three heads in a row	179

4.8 Paradoxes	180
4.8-1 The St. Petersburg paradox	180
4.8-2 The Castle Point paradox	181
4.8-3 Urn with black and white balls	182
4.9 Summary	184
4.A Linear Difference Equations	185
Chapter 5 Continuous Sample Spaces	189
5.1 A Philosophy of the Real Number System	189
5.2 Some First Examples	191
5.2-1 Distance from a river	192
5.2-2 Second random choice of angle	193
5.2-3 Broken stick	194
5.2-4 The Buffon needle	195
5.2-5 Another Monte Carlo estimate of π	196
5.2-6 Mean and variance of the unit interval	198
5.3 Some Paradoxes	199
5.3-1 Bertrand's paradox	200
5.3-2 Obtuse random triangles	203
5.3-3 A random game	205
5.4 The Normal Distribution	208
5.4-1 Hershel's derivation of the normal distribution	209
5.4-2 Distance to a random point	211
5.4-3 Two random samples from a normal distribution	212
5.5 The Distribution of Numbers	214
5.5-1 The general product	216
5.5-2 Persistence of the reciprocal distribution	218
5.5-3 Probability of shifting	219
5.5-4 The general quotient	220
5.6 Convergence to the reciprocal distribution	222
5.6-1 Product of two numbers from a flat distribution	222
5.6-2 Approach to the reciprocal distribution in general	222
5.6-3 Approach from the flat distribution	224
5.6-4 A computation of products from a flat distribution	224
5.7 Random Times	226
5.7-1 Sinusoidal notion	226
5.7-2 Random events in time	227
5.7-3 Mixtures	230
5.8 Dead Times	230
5.9 Poisson Distributions in time	231
5.9-1 Shape of the state distributions	232
5.10 Queuing Theory	233
5.11 Birth and Death Systems	237
5.12 Summary	238

Chapter 6	Uniform Probability Assignments	239
6.1	Mechanical Probability	239
6.2	Experimental Results	243
6.3	Mathematical Probability	245
6.4	Logical Probability	247
6.5	Robustness	248
6.6	Natural Variables	250
6.7	Summary	251
Chapter 7	Maximum Entropy	253
7.1	What is Entropy?	253
7.2	Shannon's Entropy	256
7.2-1	An entropy computation	257
7.2-2	The entropy of uniform distributions	261
7.3	Some Mathematical Properties of the Entropy Function	262
7.3-1	The log inequality	262
7.3-2	Gibbs' inequality	263
7.3-3	The entropy of independent random variables	263
7.3-4	The entropy decreases when you combine items	264
7.4	Some Simple Applications	265
7.4-1	Maximum entropy distributions	265
7.4-2	The entropy of a single binary choice	267
7.4-3	The entropy of repeated binary trials	267
7.4-4	The entropy of the time to first failure	268
7.4-5	A discrete distribution with infinite entropy	269
7.5	The Maximum Entropy Principle	270
7.5-1	Marginal distributions	271
7.5-2	Jaynes' example	272
7.5-3	Extensions of Jaynes' example	275
7.6	Summary	276
Chapter 8	Models of Probability	279
8.1	General Remarks	279
8.2	Review	280
8.3	Maximum Likelihood	282
8.3-1	Maximum likelihood in a binary choice	283
8.3-2	Least squares	284
8.3-3	Scale free	285
8.4	von Mises' Probability	285
8.5	The Mathematical Approach	287
8.6	The Statistical Approach	289

8.7 When the Mean Does Not Exist	289
8.7-1 The Cauchy distribution	290
8.8 Probability as an Extension of Logic	294
8.8-1 The four liars	295
8.9 di Finetti	296
8.10 Subjective Probability	297
8.11 Fuzzy Probability	299
8.12 Probability in Science	300
8.13 Complex Probability	302
8.14 Summary	304
Chapter 9 Some Limit Theorems	307
9.1 Introduction	307
9.2 The Normal Distribution	309
9.2-1 The approximation of a unimodal distribution	311
9.3 The Binomial Approximation for the Case $p = 1/2$	313
9.3-1 Binomial sums	315
9.3-2 Binomial sums for nonsymmetric ranges	317
9.4 Approximation by the Normal Distribution	317
9.4-1 A normal approximation to a skewed distribution	318
9.4-2 Approximation to a binomial distribution	319
9.5 Another Derivation of the Normal Distribution	320
9.6 Random Times	323
9.7 The Zipf Distribution	324
9.8 Summary	326
Chapter 10 An Essay On Simulation	327
10.1 Introduction	327
10.2 Simulations for checking purposes	328
10.3 When you cannot compute the result	328
10.3-1 Random triangles in a circle	329
10.4 Direct simulations	330
10.5 The use of some modeling	331
10.6 Thought simulations	333
10.7 Monte Carlo methods	333
10.8 Some simple distributions	334
10.8-1 The exponential distribution	335
10.8-2 The normal distribution	335
10.8-3 The reciprocal distribution	336
10.9 Notes on programming many simulations	336
10.10 Summary	337
Index	341



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

1

Probability

“Probability is too important to be left to the experts.”

1.1 Introduction

Who has not watched the toss of a coin, the roll of dice, or the draw of a card from a well shuffled deck? In each case, although the initial conditions appear to be much the same, the specific outcome is not knowable. These situations, and many equivalent ones, occur constantly in our society, and we need a theory to enable us to deal with them on a rational, effective basis. This theory is known as *probability theory* and is very widely applied in our society—in science, in engineering, and in government, as well as in sociology, medicine, politics, ecology, and economics.

How can there be laws of probability? Is it not a contradiction? Laws imply regularity, while probable events imply irregularity. We shall see that many times amid apparent irregularity some regularity can be found, and furthermore this limited regularity often has significant consequences.

How is it that from initial ignorance we can later deduce knowledge? How can it be that although we state that we know nothing about a single toss of a coin yet we make definite statements about the result of many tosses, results that are often closely realized in practice?

Probability theory provides a way, indeed a *style*, of thinking about such problems and situations. The classical theory has proved to be very useful even in domains that are far removed from gambling (which is where it arose and is still, probably, the best initial approach). *This style of thinking is an art* and is not easy to master; both the historical evidence based on its late development, and the experience of current teaching, show that much *careful thinking* on

the student's part is necessary before probability becomes a mental habit. The sophisticated approach of beginning with abstract postulates is favored by mathematicians who are interested in covering as rapidly as possible the material and techniques that have been developed in the past. This is not an effective way of teaching the understanding and the use of probability in new situations, though it clearly accelerates the formal manipulation of the symbols (see the quotation at the top of the Preface). We adopt the slow, cautious approach of carefully introducing the assumptions of the models, and then examining them, together with some of their consequences, before plunging into the formal development of the corresponding theory. We also show the relationship between the various models of probability that are of use in practice; there is not a single model of probability, but many and of differing reliabilities.

Mathematics is not just a collection of results, often called theorems; it is a style of thinking. Computing is also basically a style of thinking. Similarly, probability is a style of thinking. And each field is different from the others. For example, I doubt that mathematics can be reduced to button pushing on a computer and still retain its style of thinking. Similarly, I doubt that the theory of probability can be reduced to either mathematics or computing, though there are people who claim to have done so.

In normal life we use the word "probable" in many different ways. We speak of the probability of a head turning up on the toss of a coin, the probability that the next item coming down a production line will be faulty, the probability that it will rain tomorrow, the probability that someone is telling a lie, the probability of dying from some specific disease, and even the probability that some theory, say evolution, special relativity, or the "big bang," is correct.

In science (and engineering) we also use probability in many ways. In its early days science assumed there was an exact measurement of something, but there was also some small amount of "noise" which contaminated the signal and prevented us from getting the exact measurement. The noise was modeled by probability. We now have theories, such as *information theory* and *coding theory*, which assume from the beginning that there is an irreducible noise in the system. These theories are designed to take account of the noise rather than to initially avoid it and then later, at the last moment, graft on noise. And there are some theories, such as *the Copenhagen interpretation of quantum mechanics*, which say that probability lies at the foundations of physics, that it is basic to the very nature of the world.

Thus there are many different kinds of probability, and any attempt to give only one model of probability will not meet today's needs, let alone tomorrow's. Some probability theories are "subjective" and have a large degree of "personal probability" (belief) as an essential part. Some theories try to be more scientific (objective), and that is the path we will mainly follow—without prejudice to the personal (more subjective) probability approaches which we cannot escape using in our private lives. [Kr, both volumes].

The belief that there is not a single model of probability is rarely held, but it is not unique to the author. The following paraphrase shows an alternate opinion (ignore the jargon for the moment) [G, p.xi,xii]:

In quantum probability theory the state of the system is determined by a complex-valued function A (called the amplitude function) on the outcome space (or sample space). The outcomes of an event $E = \{x_1, x_2, \dots\}$ have the probability $P(E)$ of E which is computed by

$$P(E) = \left| \sum [A(x_i)] \right|^2$$

This allows for interference as happens in optics and sound. If the outcomes do not interfere, then of course

$$P(E) = \sum [|A(x_i)|]$$

This last is the classical probability theory. In the interference case $P(E)$ decomposes into the sum of two parts. The real part is the classical counterpart and the imaginary part leads to the constructive or destructive interference which is characteristic of quantum mechanics and much of physics.

The basic axiom of the path integral formalism is that the state of a quantum mechanical system is determined by the amplitude function A and that the probability of a set of interfering outcomes $\{x_1, x_2, \dots\}$ is the first equation. This point has been missed by the axiomatic approaches and in this sense they have missed the essence of quantum mechanics. In principle this essence can be regained by adjoining an amplitude function axiom to the other axioms of the system, but then the axiomatic system is stronger than necessary and may well exclude important cases.

Since we are going to use a scientific approach to probability it is necessary to say what science is and is not. In the past science has tried to be objective and to insist on the property that different people doing the same experiment would get essentially the same results. But we must not assume that science is completely objective, that this repeatability property is perfectly attained or is even essential (for example, consider carefully observational astronomy). Again, if I were to ask two different people to measure the width of a table, one of them might by mistake measure the length, and hence the measurements might be quite different. There is an unstated amount of "understanding what is meant" that is implied in every experimental description. We can only hope to control this assumed background (that is supposed to produce absolute consistency) so that there is little room for error, but we do not believe that we can ever completely eliminate some personal judgment

in science. In science we try to identify clearly where judgment comes in and to control the amount of it as best we can.

The statement at the head of this chapter,

Probability is too important to be left to the experts,

is a particular case of a very general observation that the experts, by their very expert training and practice, often miss the obvious and distort reality seriously. The classical statement of this general principle is, "War is too important to be left to the generals." We daily see doctors who treat cases but not patients, mathematicians who give us exact solutions to the wrong problems, and statisticians who give us verifiably wrong predictions. In our Anglo-Saxon legal system we have clearly adopted the principle that in guilt vs. innocence "twelve tried and true men" are preferable to an experienced judge. The author clearly believes that the same is true in the use of probability; the desire of the experts to publish and gain credit in the eyes of their peers has distorted the development of probability theory from the needs of the average user.

The comparatively late rise of the theory of probability shows how hard it is to grasp, and the many paradoxes show clearly that we, as humans, lack a well grounded intuition in this matter. Neither the intuition of the man in the street, nor the sophisticated results of the experts provides a safe basis for important actions in the world we live in. The failure to produce a probability theory that fits the real world as it is, more or less, can be very costly to our society which uses it constantly.

In the past science has developed mainly from attempts to "measure how much," to be more precise than general statements such as larger, heavier, faster, more likely, etc. Probability tries to measure more precisely than "more likely" or "less likely." In order to use the scientific approach we will start (Section 1.5) with the *measurement* of probability in an objective way. But first some detours and formalities are needed.

1.2 Models in General

It is reasonably evident to most people that all thought occurs internally, and it is widely believed that thinking occurs in the head. We receive stimuli from the external world, (although some sophists have been known to deny the existence of the external world, we shall adopt a form of the "naive realism" position that the world exists and is to some extent "knowable"). We organize our incoming stimuli according to some ill-defined model of reality that we have in our heads. As a result we do not actually think about the real external

world but rather we “think” in terms of our impressions of the stimuli. Thus all thought is, in some sense, *modeling*.

A model can not be proved to be correct; at best it can only be found to be reasonably consistent and not to contradict some of our beliefs of what reality is. Karl Popper has popularized the idea (which goes back at least to Francis Bacon) that for a theory to be scientific it must be potentially *falsifiable*, but in practice we do not always obey his criterion, plausible as it may sound at first. From Popper’s point of view a scientific theory can only be disproved, but never proved. Supporting evidence, through the application of finite induction, can increase our faith in a theory but can never prove its absolute truth.

In actual practice many other aspects of a model are involved. Occam’s razor, which requires that minimal assumptions be used and extra ones removed, is often invoked as a criterion—we do not want redundant, possibly contradictory assumptions. Yet Occam’s razor is not an absolute, final guide. Another criterion is aesthetic taste, and that depends, among other things, on the particular age and culture you happen to live in. We feel that some theories are nicer than others, that some are beautiful and some are ugly, and we tend to choose the loveliest one in practice. Another very important criterion is “fruitfulness”—does the model suggest many new things to try?

A model is often judged by how well it “explains” some observations. There need not be a unique model for a particular situation, nor need a model cover every possible special case. A model is not reality, it merely helps to explain some of our impressions of reality. For example, for many purposes we assume that a table is “solid,” yet in physics we often assume that it is made of molecules and is mainly “empty space.” Different models may thus seem to contradict each other, yet we may use both in their appropriate places.

In view of the wide variety of applications of probability, past, present, and future, we will develop a sequence of models of probability, generally progressing from the simpler to the more complex as we assume more and more about the model. At each stage we will give a careful discussion of the assumptions, and illustrate some features of them by giving consequences you are not likely to have thought were included in the model. Thus we will regularly discuss examples (paradoxes) whose purpose is to show that you can get rather peculiar results from a model if you are not careful, even though the assumptions (postulates) on the surface seem to be quite bland, [St]. It is irresponsible to teach a probability course as if the contents were entirely safe to use in any situation. As Abelard (1079–1142) said,

“I expose these contradictions so that they may excite the susceptible minds of my readers to the search for truth, and that these contradictions may render their minds more penetrating as the effect of that search.”

Since the applications of probability theory are constantly expanding in range and depth, it is unlikely that anyone can supply models for all the situations that will arise in the future. Hence we are reduced to presenting a series of models, and to discussing some of their relationships, in the hopes that you will be able to create the model you need when the time comes to develop a new one. Each application you make requires you to decide which probability model to adopt.

1.3 The Frequency Approach Rejected

Most people have at least two different models of probability in their minds. One model is that there is a probability to be associated with a single event, the other is that probability is the ratio of the number of successes to the total number of trials in a potentially infinite sequence of similar trials. We need to assume one of these and deduce, in so far as we can, the other from the first—otherwise we risk assuming a contradiction! It will turn out that *the two models are not completely equivalent*.

A third kind of probability that most people are familiar with is that which arises in betting between friends on sporting events. The “odds” are simply a version of probability; the odds of a to b are the probabilities of $a/(a+b)$ and $b/(a+b)$. There is often no serious thought of repetitions of the event, or a “random selection from an ensemble of similar events,” rather it is the feeling that your “hunches” or “insights” are better than those of the opponent. Both sides recognize the *subjective* aspect, that two people with apparently the same information can have different probability assignments for the same event. The argument that they have different amounts of information is unconvincing; it is often merely a temporary mood that produces the difference.

The *frequency approach* (the ratio of the number of successes to the total number of equally likely possible outcomes) seems to be favored by most statistically inclined people, but it has severe scientific difficulties. The main difficulty with the frequency approach is that no one can say how many (necessarily finite) number of trials to make, and it is not practical science to define the probability as the limit of the ratio as the number of trials approaches infinity. Furthermore, the statement of the limiting frequency, in a careful form, will often include the words that you will for any long, finite sequence be *only probably* close! It is circular to say the least! And you can't *know* whether you are looking at an exceptional case or a likely case. It is also true that to get even reasonable accuracy usually requires a discouragingly large number of trials. Finally, there are models, as we shall show in Section 8.7 for which the single event probability has meaning and at the same time the average of n samples has more variability than the original distribution!

For a long time there has been in science a strong trend towards accepting in our theories only objectively measureable quantities and of excluding purely conceptual ones that are only props (like the “ether” in classical physics). The extreme of allowing *only* measureable quantities is perhaps too rigid an attitude (though a whole school of philosophy maintains this position); yet it is a useful criterion in doing science.

Similarly, there are many mathematical theorems in the literature whose hypotheses cannot be known, even in an intuitive sense, to be close let alone be either exactly true or false within the framework of the intended application; hence these theorems are not likely to be very useful in these practical applications of the theory. Such theorems may be interesting and beautiful pure mathematics but are of dubious applicability.

This “practical use” attitude tends to push one towards “constructive mathematics” [B] and away from the conventional mathematics. The increasing use of computers tends to make many people favor “computable” numbers as a basis for the development of probability theory. Without careful attention to the problem for which it is being used it is not a priori obvious which approach to the real number system we should adopt. And there are some people who regard the continuous mathematics as merely a useful approximation to the discrete since usually the equivalent operations are easier in the continuous model than they are in the discrete. Furthermore, quantum mechanics suggests, at least to some people, that ultimately the universe is discrete rather than continuous.

Whenever a new field of application of mathematics arises it is always an open question as to the applicability of the mathematics which was developed earlier; if we are to avoid serious blunders it must be carefully rethought. This point will be raised repeatedly in this book.

For these and other reasons we will not begin with the frequency approach, but rather deduce (Sections 2.8 and Appendix 2.A) the frequency model along with its limitations. For this present Section see [Ke], especially Chapter 8.

1.4 The Single Event Model

In most fields of knowledge it is necessary, sooner or later, to introduce some technical notation and jargon so that what is being talked about is fairly precise and we can get rid of the vagueness of words as used in normal language. We have used some of these words already!

At first a *trial* is a single unit, like the toss of a coin, the roll of a die (singular of dice), or the draw of a single card from a well shuffled deck. The *elementary events* of a trial are the possible outcomes of the trial. For example, for the toss of a coin the elementary events are “head” and “tail.”

For the roll of a die the elementary events are the names of the six faces, 1, 2, 3, 4, 5, 6. The elementary events are considered to be indivisible, atomic, events from which more complex events (say an even numbered face of a die) may be built. We will label the possible outcomes of a trial by the symbols x_i , where i runs from 1 to n (or possibly 0 to n). Thus as a first approximation, to be modified later, we have the trial labeled X whose outcome is some particular x_i . The capital letter X represents the trial, and the set $\{x_i\}$ represents the possible outcomes of the trial; some particular x_i will be the actual outcome of the trial X being conducted. In some circumstances the symbol X may be thought of as being the whole set $\{x_i\}$ just as the function $\sin x$ may be thought of as the entire function. This list of possible outcomes is an idealization of reality. Initially we assume that this list is finite in length.

Later on a trial may be a sequence of these simple trials; the roll of, say, four dice at one time will be a single trial. Still later we will continue for an indefinite number of times until some condition is reached such as tossing a coin until three heads occur in a row. In such cases the entire run will be considered a single trial, and the number of possible events will not be finite but rather countably infinite. Still later, we will examine a continuum of events, a classically non-countable number of events.

In principle the first step in a discrete probability problem is, for the trial (named) X , to get the corresponding list (or set) X of all the possible outcomes x_i . Thus X has two meanings; it is the name of the trial and it is also the list of possible the outcomes of the trial. The name of the outcome should not be logically confused with any value associated with the outcome. Thus logically the face "the number of spots," say 4, on a die is a label and should not be confused with the name "four," nor with the value "4" that is usually associated with this outcome. In practice these are often carelessly treated as being the same thing.

The beginner is often confused between the model and reality. When tossing a coin the model will generally not include the result that the coin ends up on its edge, nor that it is lost when it rolls off the table, nor that the experimenter drops dead in the middle of the trial. We could allow for such outcomes if we wished, but they are generally excluded from simple problems.

In a typical mathematical fashion these events (outcomes) x_i are considered to be points in the *sample space* X ; each elementary event x_i is a point in the (at present) finite, discrete sample space. To get a particular realization of a trial X we "sample" the sample space and select "at random" (to be explained in more detail in Section 1.9) an event x_i . The space is also called a *list* (as was stated above) and we choose (at random) one member (some x_i) of the list X as the outcome of the trial.

1.5 Symmetry as the Measure of Probability

With this notation we now turn to the assignment of a *measure* (the numerical value) of the *probability* p_i to be associated with (assigned to) the event x_i of the trial.

Our first tool is *symmetry*. For the *ideal die* that we have in our mind we see that (except for the labels) the six possible faces are all *the same*, that they are *equivalent*, that they are *symmetric*, that they are *interchangeable*. Only one of the elementary events x_i can occur on one (elementary) trial and it will have the probability p_i . If any one of several different outcomes x_i is allowed on a trial (say either a 4 or a 6 on a roll of a die) then we must have the “additivity” of probabilities p_i . For the original sample space we will get the same probability, due to the assumed symmetry, for the occurrences of each of the elementary events. We will scale this probability measure so that the sum of all the probabilities over the sample space of the elementary events must add up to 1.

If you do not make equal probability assignments to the interchangeable outcomes then you will be embarrassed by interchanging a pair of the nonequal values. Since the six faces of an ideal die are all symmetric, interchangeable if you prefer, then each face must have the probability of $1/6$ (so that the total probability is 1). For a well shuffled *ideal deck* of cards we believe that each of the 52 cards are interchangeable with every other card so each card must have a probability of $1/52$. For a well balanced coin having two symmetric faces each side must have probability $1/2$.

Let us stop and examine this crucial step of assigning a measure (number) p_i called “the probability” to each possible outcome x_i of a trial X . These numbers are the probabilities of the elementary events. From the perceived symmetry we deduced interchangeability, hence the equality of the measure of probability of each of the possible events. The reasonableness of this clearly reinforces our belief in this approach to assigning a measure of probability to elementary events in the sample space.

This perception of symmetry is subjective; either you see that it is there or you do not. But at least we have isolated any differences that may arise so that the difference can be debated sensibly. We have proceeded in as objective a fashion as we can in this matter.

However, we should first note that, “Symmetric problems always have symmetric solutions.” is false [W]. Second, when we say symmetric, equivalent, or interchangeable, is it based on perfect knowledge of all the relevant facts, or is it merely saying that we are not aware of any lack thereof? The first case being clearly impossible, we must settle for the second; but since the second case is vague as to how much we have considered, this leaves a lot to be desired. We will always have to admit that if we examined things more closely we might find reason to doubt the assumed symmetry.

We now have two general rules: (1) if there are exactly n different inter-

changeable (symmetric in some sense) elementary events x_i , then each event x_i has probability $p_i = 1/n$ since (2) the probabilities of the elementary events in the sample space are additive with a total probability of 1.

Again, from the additivity of the probabilities of the elementary events it follows that for more complex situations when there is more than one point in the sample space that satisfies the condition then we add the individual probabilities.

For the moment we will use the words “at random” to mean equally likely choices in the sample space; but even a beginner can see that this is circular—what can “equally likely” mean except equally probable, (which is what we are trying to define)? Section 1.9 will discuss the matter more carefully; for the moment the intuitive notion will suffice to do a few simple problems.

Example 1.5-1 Selected Faces of a Die

Suppose we have a well balanced die and ask for the probability that on a random throw the upper face is less than or equal to 4.

The admissible set of events (outcomes) is $\{x_i\} = \{1, 2, 3, 4\}$. Since each has probability $p_i = 1/6$, we have

$$\text{Prob } \{x_i \leq 4\} = 1/6 + 1/6 + 1/6 + 1/6 = 4/6 = 2/3$$

If instead we ask for the probability that the face is an even number, then we have the acceptable set $\{2, 4, 6\}$, and hence

$$\text{Prob } \{x_i \text{ even}\} = 1/6 + 1/6 + 1/6 = 3/6 = 1/2$$

Example 1.5-2 A Card With the Face 7

What is the probability that on the random draw of a card from the standard card deck of 52 cards (of 4 suits of 13 cards each) that the card drawn has the number 7 on it?

We argue that there are exactly 4 cards with a 7 on their face, one card from each of the four suits, hence since the probability of a single card is $1/52$ we must have the probability

$$\text{Prob } \{\text{face} = 7\} = 4/52 = 1/13.$$

There are other simple arguments that will lead to the same result.

Example 1.5-3 Probability of a Spade

What is the probability of drawing a spade on a random draw from a deck of cards?

We observe that the probability of any card drawn at random is $1/52$. Since there are 13 spades in the deck we have the combined probability of

$$P(\text{spade}) = (13)(1/52) = 1/4$$

Exercises 1.5

1.5-1. If there are 26 chips each labeled with a different letter of the alphabet what is the probability of choosing the chip labeled Q when a chip is drawn at random? Of drawing a vowel (A, E, I, O, U)? Of not drawing a vowel? Ans. $1/26, 5/26, 21/26$.

1.5-2. If 49 chips are labeled $1, 2, \dots, 49$, what is the probability of drawing the 7 at random? A number greater or equal to 7? Ans. $1/49, 43/49$.

1.5-3. If 49 chips are labeled $2, 4, \dots, 98$, what is the probability of drawing the chip 14 at random? Of a number greater than 20?

1.5-4 A plane lattice of points in the shape of a square of size $n > 2$ has n^2 points. What is the probability of picking a corner point? The probability of picking an inner point? Ans. $4/n^2, [(n-2)/n]^2$

1.5-5 A day of the week is chosen at random, what is the probability that it is Tuesday? A "weekday"?

1.5-6 We have n^3 items arranged in the form of a lattice in a cube, and we pick at random an item from an edge (not on the bottom). What is the probability that it is a corner item? Ans. $1/(2n-3)$.

1.5-7 In a set of 100 chips 10 are red; what is the probability of picking a red chip at random? Ans. $1/10$.

1.5-8 If there are r red balls, w white balls, and b blue balls in an urn and a ball is drawn at random, what is the probability that the ball is red? Is white? Is blue? Ans. $r/(r+w+b), w/(r+w+b), b/(r+w+b)$.

1.5-9 There are 38 similar holes in a roulette wheel. What is the probability of the ball falling in any given hole?

1.5-10 In a deck of 52 cards what is the probability of a random drawing being the 1 (ace) of spades? Of drawing at random a heart? Ans. $1/52, 1/4$

1.5-11 There are 7 bus routes stopping at a given point in the center of town. If you pick a bus at random what is the probability that it will be the one to your house?

1.5-12 In a box of 100 items there are 4 defective ones. If you pick one at random what is the probability that it is defective? Ans. $1/25$

1.5–13 If 3% of a population has a certain disease what is the probability that the next person you pass on the street has it? Ans. 0.03

1.5–14 Due to a maladjustment a certain machine produces every fifth item defective. What is the probability that an item taken at random is defective?

1.5–15 In a certain class of 47 students there are 7 with last names beginning with the letter S. What is the probability that a random student's name will begin with the letter S? Ans. $7/47$

1.6 Independence

The next idea we need is *independence*. If we toss a coin and then roll a die we feel that the two outcomes are *independent*, that the outcome of the coin has no influence on the outcome of the die. The sample space is now

	1	2	3	4	5	6
H	$H1$	$H2$	$H3$	$H4$	$H5$	$H6$
T	$T1$	$T2$	$T3$	$T4$	$T5$	$T6$

We regard this sample space as the *product* of the two elementary sample spaces. This is often called the Cartesian (or direct) product by analogy with cartesian coordinates of the two original sample spaces, $\{H, T\}$ for the coin, and $\{1, 2, 3, 4, 5, 6\}$ for the die. Since we believe that the outcomes of the die and coin are independent we believe (from symmetry) that the events in the product sample space are all equally likely, hence each has probability $1/(2 \times 6) = 1/12$. We see that the probability of the compound events in the product sample space are (for independent events) just the product of the corresponding probabilities of the separate original events, that

$$1/12 = (1/2)(1/6).$$

Clearly the product space (meaning the product of the sample spaces) of the coin with the die is the same as the product space of the die with the coin.

We immediately *generalize* to any two independent events, each of these being from an equally likely sample space. If there are n_1 of the first kind of event and n_2 of the second, then the product space of the combined trial has $n_1 n_2$ events each of the same probability $1/n_1 n_2 = (1/n_1)(1/n_2)$.

Example 1.6-1 *The Sample Space of Two Dice*

Consider the toss of two independently rolled dice. The sample space has $6 \times 6 = 36$ events each of probability $1/36$, and the sample space is:

1,1	1,2	1,3	...	1,6
2,1	2,2	2,3	...	2,6
...
6,1	6,2	6,3	...	6,6

Clearly the idea of a product space may be extended to more than two independent things.

Example 1.6-2 *The Sample Space of Three Coins*

Consider the toss of three coins. We have $2 \times 2 \times 2 = 8$ equally likely outcomes. It is awkward to draw 3 dimensional sample spaces, and it is not feasible to draw higher dimensional sample spaces, so we resort to listing the elementary events. For the three coins we have a list (equivalent to the product sample space)

HHH HHT HTH THH HTT THT TTH TTT

each of probability $(1/2)(1/2)(1/2) = 1/8$.

The extension to m different independent trials each having $n_i (i = 1, 2, \dots, m)$ equally likely outcomes leads to a product space of size equal to the product $n_1 n_2 \dots n_m$ of the dimensions of the original spaces. The probabilities of the events in the product space are equal to the corresponding products of the probabilities of the original events making up the original spaces, hence the product space events are also equally likely (since the original ones were equally likely).

Example 1.6-3 *The Number of Paths*

If you can go from A to B in 6 ways, from B to C in 7 ways, and from C to D in 5 ways, then there are $6 \times 7 \times 5 = 210$ ways of going from A to D , Figure 1.6-1. If each of the original ways were equally likely and independent then each of the 210 ways is also equally likely and has a probability of $1/210$.

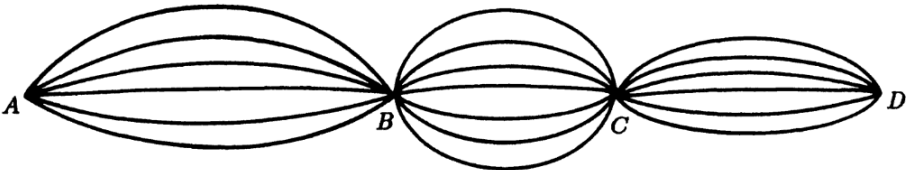


FIGURE 1.6-1

Notice again, the probabilities of the equally likely events in the product sample space can be found by either: (1) computing the number of events x_i in the product space and assigning the same probability $p_i = 1/(\text{the number of events})$ to each event in the product space, or (2) assign the product of the probabilities of the separate parts to each compound event in the product space.

Notice also that we are always *assuming* that the elementary points in the sample space are independent of each other, that one outcome is not influenced by earlier choices. However, in this universe in which we live *apparently* all things are interrelated (“intertwined” is the current jargon in quantum mechanics); if so then independence is an idealization of reality, and is not reality.

It is easy to make up simple examples illustrating this kind of probability problem, and working out a large number of them is not likely to teach you as much as carefully reviewing in your mind *why* they work out as they do.

Exercises 1.6

1.6-1 What is the probability of HHH on the toss of three coins when the order of the outcomes is important? Of HTH ? Of TTH ? Ans. $1/8$.

1.6-2 Describe the product space of the draw of a card and the roll of a die.

1.6-3 What is the size of the product space of the roll of three dice? Ans. $6^3 = 216$.

1.6-4 What is the size of the product space of the toss of n coins? Ans. 2^n .

1.6-5 What is the size of the product space of the toss of n dice?

1.6-6 What is the product space of a draw from each of two decks of cards? Ans. Size = $(52)^2$.

1.6-7 What is the size of the sample space of independent toss of a coin, roll of a die, and the draw of a card?

1.7 Subsets of a Sample Space

We are often interested not in all the individual outcomes in a product space but rather only those outcomes which have some specific property. We have already done a few such cases.

Example 1.7-1 *Exactly One Head in Three Tosses*

Suppose we ask in how many ways can exactly one head turn up in the toss of three coins, or equivalently, what is the probability of the event. We see that in the complete sample space of 8 possible outcomes we are only interested in the three events

$$HTT \quad THT \quad TTH$$

Each of these has a probability of $1/8$, so that the sum of these will be $3/8$ which is the corresponding probability.

Clearly we are using the additivity of the probabilities of the events in the sample space. We need to know the size of the sample space, but we need not list the whole original sample space, we need only count the number of successes. Listing things is merely one way of counting them.

Example 1.7-2 *Sum of Two Dice*

What is the probability that the sum of the faces of two independently thrown dice total 7? We have only the six successful possibilities

$$1 + 6, \quad 2 + 5, \quad 3 + 4, \quad 4 + 3, \quad 5 + 2, \quad 6 + 1$$

in the total sample space of $6^2 = 36$ possible outcomes. See Figure 1.7-1. Each possible outcome has probability of $1/36$, hence the probability of a total of 7 is $6/36 = 1/6$.

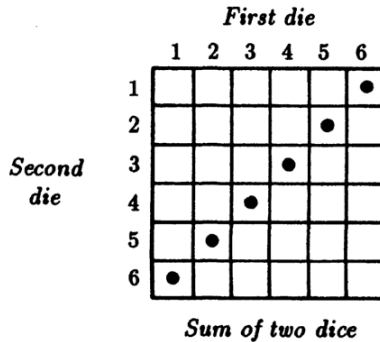


FIGURE 1.7-1

We can look at this in another way. Regardless of what turns up on the first die, there is exactly one face of the second that will make the total equal to 7, hence the probability of this unique face is $1/6$.

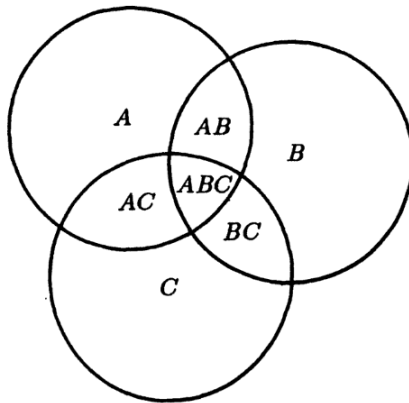
If we were to ask in how many ways, (or equivalently what is the probability), can the sum of three or four, or more dice add up to a given number

then we would have a good deal of trouble in constructing the complete sample space and of picking out the cases which are of interest. It is the purpose of the next two chapters to develop some of the mathematical tools to carry out this work; the purpose of the present chapter is to introduce the main ideas of probability theory and not to confuse you by technical details, so we will not now pursue the matter much further.

We now have the simple rule: count the number of different ways the compound event can occur in the original sample space of equally likely events (outcomes), then the probability of this compound event is the ratio of the number of successes in the sample space to the total number of events in the original sample space.

Aside: Probability Scales

We have evidently chosen to measure probability on a scale from 0 to 1, with 0 being impossibility and 1 being certainty. A scale that might occur to you is from 1 to infinity, which is merely the reciprocal of the first range, (you might also consider 0 to infinity). In these cases you will find that the addition of probabilities is not simple. There are other objections that you can find if you choose to explore these scales.



Venn Diagram

FIGURE 1.7-2

A *Venn diagram* is mainly a symbolic representation of the subsets of a sample space. Typically circles are used to represent the subsets, see Figure 1.7-2. The subsets represented are the insides of the circles A , B and C . Now consider those events which have both properties A and B , shown by the region AB . Similarly for the sets AC and BC . Finally, consider the events which have the properties of A , B and C at the same time; this is the inner piece in the diagram, ABC . In this simple case the representation is very illuminating. But if you try to go to very many subsets then the problem of

drawing a diagram which will show clearly what you are doing is often difficult; circles are not, of course, necessary, but when you are forced to draw very snake-like regions then the diagram is of little help in visualizing the situation. In the Venn diagram the members of a set are often scattered around (in any natural display of the elements of the sample space), hence we will not use Venn diagrams in this book.

Exercises 1.7

- 1.7-1 What is the probability of a die showing an odd digit?
- 1.7-2 What is the probability of drawing a "diamond" when drawing a card from a deck of 52 cards? Ans. $1/4$.
- 1.7-3 On the roll of two dice what is the probability of getting at least one face as a 4? Ans. $11/36$.
- 1.7-4 What is the probability of two H's in two tosses of a coin?
- 1.7-5 On the roll of two independent dice what is the probability of a total of 6? Ans. $5/36$.
- 1.7-6 Make a Venn diagram for four sets.
- 1.7-7 In a toss of three coins what are the probabilities of 0, 1, 2, or 3 heads? Ans. $1/8, 3/8, 3/8, 1/8$.
- 1.7-8 What is the probability of drawing either the 10 of diamonds or the jack of spades?
- 1.7-9 What is the probability of drawing an "ace" (ace = 1), jack, queen or king from a deck of 52 cards? Ans. $4/13$.
- 1.7-10 What is the probability of drawing a black card from the deck?
- 1.7-11 If two dice are rolled what is the probability both faces are the same? Ans. $1/6$.
- 1.7-12 If three dice are rolled what is the probability that all three faces are the same? Ans. $1/36$.
- 1.7-13 If a coin is tossed n times what is the probability that all the faces are the same? Ans. $1/2^{n-1}$
- 1.7-14 If one card is drawn at random from each of two decks of cards what is the probability that the two cards are the same? Ans. $1/52$.
- 1.7-15 Odd man out. Three people each toss a coin. What is the probability of some person being the "odd man out"? Ans. $3/4$.
- 1.7-16 If 4 people play odd man out what is the probability that on a round of tosses one person will be eliminated? Ans. $1/2$.
- 1.7-17 On the toss of two dice what is the probability that at least one face has a 4 or else that the sum of the faces is 4? Ans. $1/18$.
- 1.7-18 Same as 1.7-17 except the number is 6 (not 4). Ans. $4/9$.

1.8 Conditional Probability

Frequently the probability we want to know is *conditional* on some event. The effect of the condition is to remove some of the events in the sample space (list), or equivalently to confine the admissible items of the sample space to a limited region.

Example 1.8-1 At Least Two Heads in Ten Tosses

What is the probability of at least two heads in the toss of 10 coins, given that we know that there is at least one head.

The initial sample space of $2^{10} = 1024$ points is uncomfortably large for us to list all the sample points, even when some are excluded; and we do not care to list even the successes. Instead of counting we will calculate the number of successes and the size of the entire sample space.

We may compute this probability in two different ways. First we may reason that the sample space now excludes the case of all tails, which can occur in only one way in the sample space of 2^{10} points; there are now only $2^{10} - 1$ equally likely events in the sample space. Of these there are 10 ways in which exactly one head can occur, namely in any one of the ten possible positions, first, second, . . . , tenth. To count the number of events with two or more heads we remove from the counting (but not from the sample space) these 10 cases of a single head. The probability of at least two heads is therefore the ratio of the number of successes (at least 2 heads) to the total number of events (at least one head)

$$(2^{10} - 1 - 10)/(2^{10} - 1) = 1013/1023 = 0.9902\dots$$

Secondly, we could use a standard method of computing not what is wanted but the opposite. Since this method is so convenient, we need to introduce a suitable notation. If p is the probability of some simple event then we write $q = 1 - p$ as the probability of it not happening, and call it the *complement probability*. Given the above ten coin problem we can compute the complement event, the probability Q that there was one head, and then subtract this from 1 to get

$$P = 1 - Q = 1 - 10/(2^{10} - 1)$$

As a matter of convenient notation we will often use lower case letters for the sample space probabilities and upper case letters for the probabilities of the compound events.

Example 1.8-2 *An Even Sum on Two Dice*

On the roll of two independent dice, if at least one face is known to be an even number, what is the probability of a total of 8?

We reason as follows. One die being even and the sum is to be 8 hence the second die must also be even. Hence the three equally likely successes (the sum is 8) in the sample space are

$$2, 6 \quad 4, 4 \quad 6, 2$$

out of a total sample space of $\{36 - (\text{both faces odd})\} = 36 - 3 \times 3 = 27$. Hence the probability is $3/27 = 1/9$.

Example 1.8-3 *Probability of Exactly Three Heads Given that there are at Least Two Heads*

What is the probability of exactly three heads knowing that there are at least two heads in the toss of 4 coins?

We reason as follows. First, how many cases are removed from the original complete sample space $2^4 = 16$ points? There is the single case of all tails, and the four cases of exactly 1 head, a total of 5 cases are to be excluded from the original 16 equally likely possible cases; this leaves 11 still equally likely events in the sample space. Second, for the three heads there are exactly 4 ways that the corresponding 1 tail can arise, ($HHHT$, $HHTH$, $HTHH$, $THHH$), so there are exactly 4 successes. The ratio is therefore $4/11 =$ the probability of exactly three heads given that there are at least two heads.

Exercises 1.8

1.8-1 If a fair coin is tossed 10 times what is the probability that the first five are the same side? Ans. $1/16 = 6.25\%$.

1.8-2 Given that the sum of the two faces of a pair of dice is greater than 10, what is the probability that the sum is 12? Ans. $1/3$.

1.8-3 In drawing two cards from a deck (without returning the first card) what is the probability of two aces when you get at least one ace?

1.8-4 What is the probability of different faces turning up on two tosses of a die? Ans. $5/6$.

1.8-5 What is the probability of no two faces being the same on three tosses of a die? Ans. $5/9$.

1.8-6 What is the probability that at least one face is a 6 on the toss of a pair of dice, given that the sum is ≥ 10 ? Ans. $5/6$.

1.8-7 Given that in four tosses of a coin there are at least two heads, what is the probability that there are at least three heads?

1.8-9 Given that both faces on the toss of two dice are odd, what is the probability of at least one face being a 5? Ans. $5/9$.

- 1.8-10 In n tosses of a coin, given that there are at least two heads, what is the probability of n or $n - 1$ heads?
- 1.8-11 Given that two cards are honor cards (10, J, Q, K, A) in spades what is the probability that exactly one of them is the ace? Ans. $8/25$.
- 1.8-12 Given that a randomly drawn card is black, what is the probability that it is an honor card? Ans. $10/26 = 5/13$.
- 1.8-13 Given that an even number of heads show on the toss of 4 coins (0 is an even number), what is the probability of exactly 2 heads? Ans. $3/4$.
- 1.8-14 On the toss of three coins what is the probability of an even number of heads? Ans. $1/2$.
- 1.8-15 On the toss of 4 coins what is the probability of an even number of heads?
- 1.8-16 From the previous two problems what is the probability of an even number of heads on the toss of n coins? Ans. $1/2$.
- 1.8-19 Given that the sum of the faces on two dice is 6, what is the probability that one face is a 4? Ans. $2/5$.
- 1.8-20 Given that each of the faces on three dice show even number, what is the probability of at least one face being a 6?

1.9 Randomness

Randomness is a negative property; it is the absence of any pattern (we reject the idea that the absence of a pattern is a pattern and for set theory formalists this should be considered carefully). Randomness can never be proved, only the lack of it can be shown. We use the absence of a pattern to assert that the past is of no help in predicting the future, the outcome of the next trial. A random sequence of trials is the realization of the assumption of independence—they are the same thing in different forms. Randomness is “a priori” (before) and not “a posteriori” (after).

We have not defined “pattern.” At present it is the intuitive idea that if there is a pattern then it must have a simple description—at least simpler than listing every element of the pattern.

Randomness is a mathematical concept, not a physical one. Mathematically we think of random numbers as coming from a *random source*. Any particular sequence of numbers, once known, is then predictable and hence cannot be random. A reviewer of the famous RAND [R] *Table of a Million Random Numbers* caught himself in mid review with the observation that now that they had been published the numbers were perfectly predictable (if you

had the table) and therefore could not be random! Thus the *abstract mathematical concept* of a random sequence of numbers needs to be significantly modified when we actually try to handle particular realizations of randomness.

In practice, of course, you can have only a finite sequence of numbers, and again, of course, the sequence will have a pattern, if only itself! There is a built in contradiction in the words, "I picked a random number between 1 and 10 and got 7." Once selected (a posteriori) the 7 is definite and is not random. Thus the mathematical idea of random (which is a priori) does not match closely what we do in practice where we say that we have a random sample since once obtained the random sample is now definite and is not mathematically random. Before the roll of a die the outcome is any of the 6 possible faces; after the roll it is exactly one of them. Similarly, in quantum mechanics before a measurement on a particle there is the set of possible outcomes (states); after the measurement the particle is (usually) in exactly one state.

On most computers we have simple programs that generate a sequence of "pseudo random numbers" although the word "pseudo" is often omitted. If you do not know that they are being generated by a simple formula then you are apt to think that the numbers are from a random source since they will pass many reasonable tests of randomness; each new pseudo random number appears to be unpredictable from the previous ones and there appears to be no pattern (except that for many random number generators they are all odd integers!), but if you know the kind of generating formula being used, and have a few numbers, then all the rest are perfectly predictable. Thus whether or not a particular stream of numbers is to be regarded as coming from a random source or not depends on your state of knowledge—an unsatisfactory state of affairs!

For practical purposes we are forced to accept the awkward concept of "relatively random" meaning that with regard to the proposed use we can see no reason why they will not perform as if they were random (as the theory usually requires). This is highly subjective and is not very palatable to purists, but it is what statisticians regularly appeal to when they take "a random sample" which once chosen is finite and definite, and is then not random—they hope that any results they use will have approximately the same properties as a complete counting of the whole sample space that occurs in their theory.

There has arisen in computing circles the idea that the measure of randomness (or if you wish, nonrandomness) should be via the shortest program that will generate the set of numbers (without having carefully specified the language used to program the machine!). This tends to agree with many people's intuitive feelings - it should not be easy to describe the lack of a pattern in any simple fashion, it ought to be about as hard as listing all the random numbers. Thus the sequence

0 1 0 1 0 1 0 1 0...

is "more random" than is the sequence

0 0 0 0 0 0 0 0 0...

but is less random than the sequence

0 1 0 0 0 1 1 0 1 1 0 0 0 0 0 1...

In this approach pseudo random numbers are only slightly random since the program that generates them is usually quite short! It would appear that a genuinely mathematically random number generator cannot be written in any finite number of symbols!

We will use the convenient expression "chosen at random" to mean that the probabilities of the events in the sample space are all the same *unless* some modifying words are near to the words "at random." Usually we will compute the probability of the outcome based on the uniform probability model since that is very common in modeling simple situations. However, a uniform distribution does not imply that it comes from a random source; the numbers $1, 2, 3, \dots, 6n$ when divided in this order by 6 and the remainders $(0, 1, 2, 3, 4, 5)$ tabulated, gives a uniform distribution but the remainders are not random, they are highly regular!

These various ideas of randomness (likeliness) are not all in agreement with our intuitive ideas, nor with each other. From the sample space approach we see that any hand of 13 cards dealt from a deck of 52 cards is as likely as any other hand (as probable—as random). But the description of a hand as being 13 spades is much shorter than that of the typical hand that is dealt. Furthermore, the shortness of the description must depend on the vocabulary available—of which the game of bridge provides many special words. We will stick to the sample space approach. Note that the probability of getting a particular hand is not connected with the "randomness of the hand" so there is no fundamental conflict between the two ideas.

If this idea of a random choice seems confusing then you can take comfort in this observation: while we will often speak of "picking at random" we will always end up averaging over the whole sample space, or a part of it; we do not actually compute with a single "random sample." This is quite different from what is done in statistics where we usually take a small "sample" from the sample space and *hope* that the results we compute from the sample are close to the ideal of computing over the complete sample space.