FOR REVIEW
Single Spaced, 10pt Type,
Figures and Tables in Text

An Introduction to Econometric Theory:

Measure Theoretic Probability and Statistics with Applications to Economics

A. Ronald Gallant
Department of Economics
University of North Carolina
Chapel Hill NC 27599-3305 USA

June 1995
Last Revised: December 1996

Contents

Chapter 1

Probability

1.1 Examples

This chapter introduces the basic ideas of probability theory. Four examples are used throughout to motivate the theoretical constructs. The first two, craps and keno, are games of chance, the third is a coin tossing experiment, and the fourth is the triangle map, which generates deterministic chaos. The theory that we shall develop is applicable to each of these examples.

1.1.1 Craps

Craps is a dice game that has one feature that makes it especially interesting as an example in our study of probability: The number of rolls of the dice required to decide the outcome of a bet is indeterminate. Conceptually the dice might have to be rolled forever and the bet never decided. Dealing with this contingency forces us to develop a theory of probability that turns out to be rich enough to support the study of econometrics in general and the study of time series phenomena, such as weekly interest rates, in particular. Moreover, there is nothing abstract about craps. The game is real; it is tangible; you can play it yourself in Las Vegas, Reno, Atlantic City, and elsewhere; and people have been playing it in something like its present form since at least the time of the Crusades.

The game is played at a table laid out as shown in Figure 1.1. The casino crew consists of a boxman, a stickman, and two dealers who occupy positions as shown in the figure. Dealers manage betting at their end of the table. The stickman manages the dice and the bets at the center of the table, which are hardway and one roll bets. The boxman makes change and supervises the game. The players crowd around the table at either side of the stickman. Play is noisy with players and crew announcing, deciding, and paying bets, encouraging the dice, talking to one another, etc. It is great fun.

The flow of the game is determined by the *pass line* bet. Custom and social pressure require the shooter, who is the player throwing the dice, to place a bet

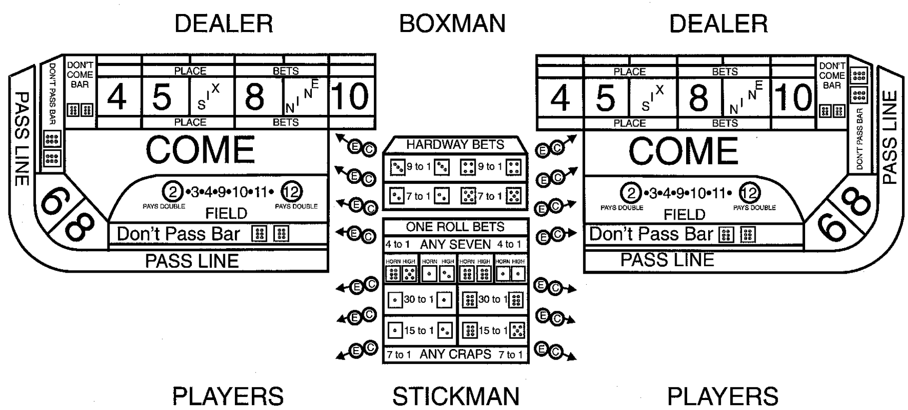

Figure 1.1. The Layout of a Crap Table and the Positions of Players and Crew.

on the pass line before the first roll, which is called a *come out* roll, although the rules do permit a bet on the don't pass bar instead. A come out roll occurs immediately after the previous pass line bet has been decided. If the previous pass line bet lost on the roll of a seven, then the losing shooter also loses the dice and they are offered to players to the shooter's left, in turn, until a player accepts them to become the new shooter. The payoff on the pass line is stated as either "1 to 1" or "2 for 1". Each means that a winning player who bet \$1 gets that \$1 back plus \$1.

Figure 1.2 shows the 36 possible positions in which the dice may land when thrown. The stickman will disallow rolls that bounce off the table, land on a pile of chips, or in the dice bin, and will scold a shooter who does not throw hard enough to hit the opposite end of the table or players who get their hands in the way of the dice.

If the sum of the dice on the come out roll is *craps*, which is a 2, 3, or 12, then the roll is called a miss and the pass line loses. If the come out roll is a 7 or 11, it wins. Otherwise, a 4, 5, 6, 8, 9, or 10 has been thrown. Whichever it is becomes the *point*. The shooter then continues to roll the dice until either the point recurs, in which case the pass line wins, or a 7 occurs, in which case the pass line loses and the dice pass leftward. It is this indeterminate number of rolls after the point is established that makes the game of craps interesting to us as an example.

If the come out roll is 4, 5, 6, 8, 9, or 10, then players who have bet the pass line are offered *free odds*. They can make a fair bet – called odds, taking the odds, or a right bet – that wins if the point recurs before a 7 is rolled. Minimally, one can bet \$1 for every \$1 bet on the pass line, but some casinos (e.g., Binion's Horseshoe, Downtown Las Vegas) have allowed as much as 100 times the pass line bet. This is a form of price competition among casinos. Free odds pay 6 to 5 if the point is 6 or 8, 3 to 2 if the point is 5 or 9, and 2 to 1 if

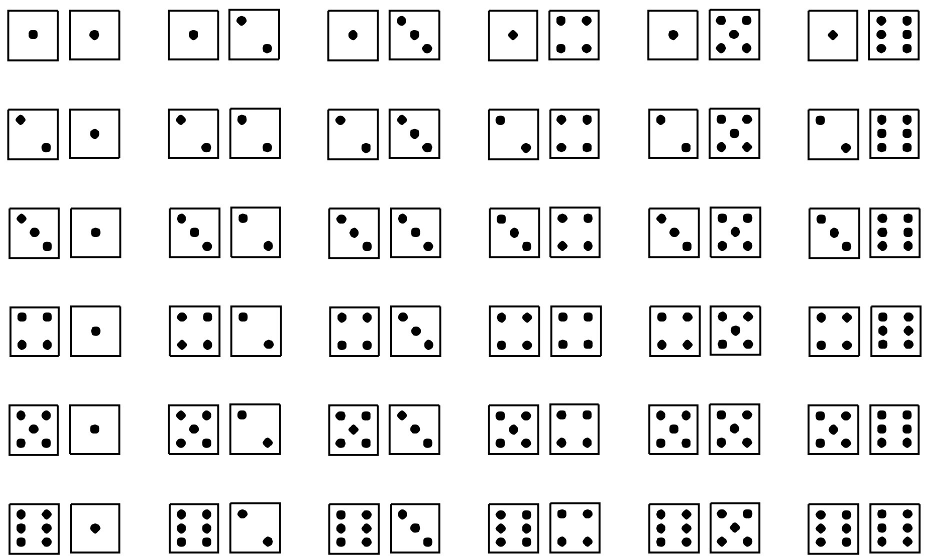

Figure 1.2. The Possible Outcomes of a Single Roll of a Pair of Dice.

the point is 4 or 10.

The *don't pass* bar bet is the opposite of the pass line bet, in the sense that the don't pass bet wins when the pass line bet loses and conversely, except that the don't pass bet neither wins nor loses if a 12 is thrown on the come out roll. Of course, the free odds bet is also reversed, it wins if a 7 is thrown before the point is made. Don't pass free odds pay 5 to 6 if the point is 6 or 8, 2 to 3 if the point is 5 or 9, and 1 to 2 if the point is 4 or 10.

The *come* and *don't come* bets are the same as the pass and don't pass bets except that a player may place that bet before any roll except the come out roll.

If you play craps, and want to keep the house advantage to a percentage that is nearly irrelevant, then play the pass, don't pass, come, don't come, and always take maximum odds. Stay away from all other bets. Admittedly, this strategy takes much of the entertainment value out of the game.

A *place* bet to win is the same as a pass line bet without the initial skirmish of the come out roll. The bettor chooses a point, a 4, 5, 6, 8, 9, or 10, and the bet wins if the point is rolled before a 7. The bet is usually off on any come out roll. Similarly, a place bet to lose is the same as a don't pass bar bet without the initial skirmish. Payoffs vary somewhat from casino to casino on place bets. Typical payoffs are shown in Table 1.1. Casinos that are more generous with free odds are often more generous with place bet payoffs as well.

A *hardway* bet on the 8 wins if a hard 8 is rolled before either an easy 8 or a 7. A hard 8 occurs when (4,4) is thrown; an easy 8 occurs when either (2,6) or (3,5) is thrown. The other hardway bets are 4, 6, and 10. Typical payoffs are

Bet	True Odds	Payoff Odds	% Casino Advantage
		Multiple Roll Bets	
Pass or Come	251 to 244	1 to 1	1.414
with free odds			0.848
with double odds			0.606
Don't Pass or Don't Come	976 to 949	1 to 1	1.402
with free odds			0.832
with double odds			0.591
Place 4 or 10 to win	2 to 1	9 to 5	6.666
5 or 9	3 to 2	7 to 5	4.000
6 or 8	6 to 5	7 to 6	1.515
Place 4 or 10 to lose	1 to 2	5 to 11	3.030
5 or 9	2 to 3	5 to 8	2.500
6 or 8	5 to 6	4 to 5	1.818
Hardway 4 or 10	8 to 1	7 to 1	11.111
6 or 8	10 to 1	9 to 1	9.090
Big 6 or Big 8	6 to 5	1 to 1	9.090
Buy 4 or 10	2 to 1	True odds less 5% of bet	4.761
5 or 9	3 to 2	True odds less 5% of bet	4.761
6 or 8	6 to 5	True odds less 5% of bet	4.761
Lay 4 or 10	1 to 2	True odds less 5% of payoff	2.439
5 or 9	2 to 3	True odds less 5% of payoff	3.225
6 or 8	5 to 6	True odds less 5% of payoff	4.000
		Single Roll Bets	
Field	5 to 4	1 to 1, 2 to 1 on 2 and 12	5.556
Any 7	5 to 1	4 to 1	16.666
2 or 12	35 to 1	30 to 1	13.890
3 or 11	17 to 1	15 to 1	11.111
Any craps	8 to 1	7 to 1	11.111

Table 1.1. True Odds, Payoff Odds, and Casino Advantage at Craps. Source: Patterson and Jaye 1982 and Dunes Hotel 1984.

shown in Table 1.1.

Other bets, whose definitions are plainly marked in Figure 1.1, are the field and the single roll bets at the center of the table. The single roll bets at the center are 2 (snake eyes), 3, 12 (box cars), any craps, 11 (the yo), and any 7. Typical payoffs are shown in Table 1.1.

These are all the bets we shall need as examples. For the remainder, see a casino brochure or Patterson and Jaye 1982.

1.1.2 Keno

Keno has the appeal of a state lottery: for a small wager you can win a lot of money. Unlike a state lottery, one does not have to wait days to learn the outcome of the bet. A new game is played every half hour or so, twenty-four hours a day. Moreover, the game comes to you, you do not have to go to it. Keno runners are all over the casino, in the restaurant, bars, they are ubiquitous. Or, there is a keno parlor set aside for the game where you can place wagers directly and watch the numbers be drawn.

The game is played by marking a ticket such as is shown in the top panel of Figure 1.3. One can mark any number of spots from 1 to 15. The ticket in the figure has eight spots marked.

Write the amount of the wager at the top of the ticket and hand it in with the wager to a writer at the keno parlor or to a keno runner. You receive back an authorized game ticket marked as shown in the second panel of Figure 1.3. The authorized game ticket is a full receipt that shows the amount wagered, the number of the game, the number of spots marked, and the marked spots.

Twenty numbered balls are drawn from 80. The mechanism that draws the balls is usually made of clear plastic and sounds, acts, and functions much as a hot air corn popper; one sees similar machines in bingo parlors. These numbers are displayed on electronic boards throughout the casino. The boards look much like a huge ticket with the draws lit up. The *catch* is the number of *draws* that match those marked on the game ticket.

If you are in the keno parlor you can pick up a draw card such as is shown in the bottom of Figure 1.3. This card has the game number on it with the draws indicated by punched holes. The catch can be determined quickly by laying the draw card over the game card. The draw card is a convenience to keno runners and to players who play multiple tickets per game.

Payoffs vary somewhat from casino to casino. A typical set of payoffs is shown in Table 1.2.

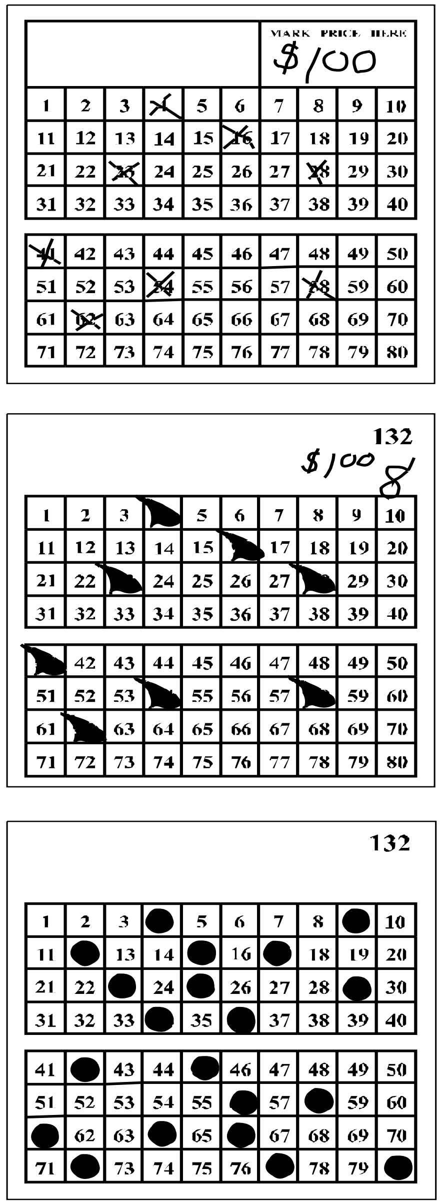

Figure 1.3. Keno Player's Ticket, Authorized Game Ticket, and Draw Card. The top panel is a blank keno ticket as marked by a player. The middle panel is the authorized game ticket issued by the casino for a wager as specified by the player's ticket. The game ticket shows the spots marked, the amount wagered, and the game number, which is 132 in this instance. The bottom panel gives the twenty numbers that were drawn on game number 132. The catch is the number of draws that match those marked on the game ticket, which is three in this instance.

Wager	\$1.00	\$5.00	\$10.00
Catch	Win	Win	Win
	Mark 1 Spot		
1	3.00	15.00	30.00
	Mark 2 Spots		
2	12.00	60.00	120.00
	Mark 3 Spots		
2	1.00	5.00	10.00
3	40.00	200.00	400.00
	Mark 4 Spots		
2	1.00	5.00	10.00
3	3.00	15.00	30.00
4	112.00	560.00	1,120.00
	Mark 5 Spots		
3	1.00	5.00	10.00
4	22.00	110.00	220.00
5	500.00	2,500.00	5,000.00
	Mark 6 Spots		
3	1.00	5.00	10.00
4	3.00	15.00	30.00
5	85.00	425.00	850.00
6	1,500.00	7,500.00	15,000.00
	Mark 7 Spots		
4	2.00	10.00	20.00
5	23.00	115.00	230.00
6	350.00	1,750.00	3,500.00
7	5,000.00	25,000.00	50,000.00
	Mark 8 Spots		
5	9.00	45.00	90.00
6	85.00	425.00	850.00
7	1,500.00	7,500.00	15,000.00
8	18,000.00	50,000.00	50,000.00
	Mark 9 Spots		
5	4.00	20.00	40.00
6	40.00	200.00	400.00
7	300.00	1,500.00	3,000.00
8	4,000.00	20,000.00	40,000.00
9	18,000.00	50,000.00	50,000.00
	Mark 10 Spots		
5	2.00	10.00	20.00
6	20.00	100.00	200.00
7	126.00	630.00	1,260.00
8	950.00	4,750.00	9,500.00
9	4,000.00	20,000.00	40,000.00
10	18,000.00	50,000.00	50,000.00

Wager	\$1.00	\$5.00	\$10.00
Catch	Win	Win	Win
	Mark 11 Spots		
6	9.00	45.00	90.00
7	75.00	375.00	750.00
8	380.00	1,900.00	3,800.00
9	2,000.00	10,000.00	20,000.00
10	12,000.00	50,000.00	50,000.00
11	25,000.00	50,000.00	50,000.00
	Mark 12 Spots		
6	5.00	25.00	50.00
7	30.00	150.00	300.00
8	240.00	1,200.00	2,400.00
9	600.00	3,000.00	6,000.00
10	1,500.00	7,500.00	15,000.00
11	8,000.00	40,000.00	50,000.00
12	25,000.00	50,000.00	50,000.00
	Mark 13 Spots		
6	1.00	5.00	10.00
7	16.00	80.00	160.00
8	80.00	400.00	800.00
9	700.00	3,500.00	7,000.00
10	2,000.00	20,000.00	40,000.00
11	8,000.00	40,000.00	50,000.00
12	20,000.00	50,000.00	50,000.00
13	25,000.00	50,000.00	50,000.00
	Mark 14 Spots		
6	1.00	5.00	10.00
7	10.00	50.00	100.00
8	40.00	200.00	400.00
9	300.00	1,500.00	3,000.00
10	1,000.00	5,000.00	10,000.00
11	3,500.00	17,500.00	35,000.00
12	12,000.00	50,000.00	50,000.00
13	25,000.00	50,000.00	50,000.00
14	36,000.00	50,000.00	50,000.00
	Mark 15 Spots		
7	8.00	40.00	80.00
8	25.00	125.00	250.00
9	130.00	650.00	1,300.00
10	300.00	1,500.00	3,000.00
11	2,600.00	13,000.00	26,000.00
12	8,000.00	40,000.00	50,000.00
13	25,000.00	50,000.00	50,000.00
14	32,000.00	50,000.00	50,000.00
15	50,000.00	50,000.00	50,000.00

Table 1.2. Keno Payoffs. No limit to betting. \$50,000.00 aggregate payoff limit to all players per game. From MGM Grand Hotel 1984.

1.1.3 Coin Tossing

Consider $x \in [0,1]$ written as a decimal (or base 10) number. For example,

$$.625 = 6\frac{1}{10} + 2\frac{1}{100} + 5\frac{1}{1000}.$$

This number also has a binary (or base 2) form

$$.625 = .101_2 = 1\frac{1}{2} + 0\frac{1}{4} + 1\frac{1}{8}.$$

Similarly to decimals, every number $x \in [0,1]$ has a binary representation and, conversely, every sequence of 0's and 1's represents a number in [0,1].

If the sequence of 0's and 1's is repetitive, then the formula

$$\frac{1}{1 \Leftrightarrow r} = 1 + r + r^2 + r^3 + \cdots \qquad 0 \le r < 1$$

for the sum of a geometric progression (Abramowitz and Stegun 1964) may be used to determine which $x \in [0,1]$ the sequence represents. For example

$$.010101\cdots_2 = 0\frac{1}{2} + 1\frac{1}{4} + 0\frac{1}{8} + 1\frac{1}{16} + \cdots = \frac{1}{4}\Big(1 + \frac{1}{4} + \frac{1}{4^2} + \cdots\Big) = \frac{1}{3}.$$

Moreover, one can observe from this that every repetitive sequence of 0's and 1's must be a ratio of positive integers and therefore a rational number. A sequence of 0's and 1's that terminates is a rational number also. A sequence that terminates has more than one representation. For example, $.1_2 = .011\cdots_2$.

Suppose that we select the 0's and 1's by tossing a fair coin. We can determine to which $x \in [0,1]$ the sequence corresponds to any desired accuracy by tossing the coin long enough. We can also determine whether or not $x \in [a,b]$ for $0 \le a < b \le 1$. There is a problem with endpoints. For example, if $b = 1/2$, then in principle one would have to toss the coin an infinite number of times to determine if x were either of the two sequences $.011\cdots_2$ or $.100\cdots_2$ that represent $1/2$. As a practical matter, the chance of this occurring is $0 = \lim_{n\to\infty}(1/2)^n$. Thus, endpoints can be disregarded and the chance that $x \in (a,b)$ or $x \in (a,b]$ or $x \in [a,b)$ or $x \in [a,b]$ is the same.

Disregarding endpoints, what are the chances of getting a sequence that represents $x \in (0,1/2]$? Observe that each $x \in (0,1/2]$ has first digit 0 and has an exact counterpart in (1/2,1] obtained by putting that 0 to a 1. Therefore, all that matters is the first toss. The chance of a 0 is 1/2 so the chance of $x \in (0,1/2]$ is 1/2.

By similar logic one concludes that the chance of x being in (0,1/4] or (1/4,1/2] or (1/2,3/4] or (3/4,1] is 1/4. The terminus of this reasoning is that the chance that x is in $(a,b]$ where $0 \le a < b \le 1$ is the length $b \Leftrightarrow a$ of the interval.

To summarize, we see that it is quite possible to describe a physical mechanism that generates numbers x in [0,1] for which it is reasonable to state that the chance that x is in some subinterval (a,b] is the length $b \Leftrightarrow a$ of that subinterval.

1.1.4 Triangular Map

The next example is interesting because there is nothing random to it at all, it is completely deterministic. Yet probability theory can be used to describe the salient characteristics of sequences $\{x_t\}_{t=0}^{\infty}$ generated according to this deterministic recipe.

The triangular map is

$$T(x) \;=\; 1 \Leftrightarrow 2\left|\frac{1}{2} \Leftrightarrow x\right| \;=\; \begin{cases} 2x & x \in [0, \frac{1}{2}] \\ 2 \Leftrightarrow 2x & x \in (\frac{1}{2}, 1] \end{cases} .$$

Consider a sequence $\{x_t\}_{t=1}^{\infty}$ generated by starting with some point $x_0 \in [0, 1]$ and using the recursion

$$x_{t+1} = T(x_t)$$

for $t = 1, 2, \ldots$ to generate the rest of the sequence. If we let x have binary representation (see Section 1.1.3), then we see that the action of the triangular map is to discard the leading 0 or 1 and occasionally flip digits. For example,

$$T(.0101_2) = .101_2$$

$$T(.101_2) = 10.0_2 \Leftrightarrow 1.01_2 = 1.111..._2 \Leftrightarrow 1.01_2 = .10111..._2 = .11_2$$

$$T(.110\,110\ldots_2) = 1.111\,111\,111\ldots_2 \Leftrightarrow 1.101\,101\,101\ldots_2 = .010\,010\,010\ldots_2.$$

Notice if we start the recursion with an x_0 whose binary representation terminates, then from some point on the sequence $\{x_t\}_{t=0}^{\infty}$ has $x_t \equiv 0$. If we start with x_0 whose digits repeat then the sequence $\{x_t\}_{t=0}^{\infty}$ cycles among some finite set of points. Starting the recursion with a rational number leads to uninteresting sequences.

However, if we start with an irrational number, the sequence $\{x_t\}_{t=0}^{\infty}$ is interesting. It is an example of what is known as a chaotic process. The fact that it is chaotic and various properties of the process are discussed in Chapter 1 of Schuster 1988. Of these properties, one is of special interest to us: For any a and b with $0 \le a < b \le 1$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} I_{(a,b]}(x_t) = b \Leftrightarrow a,$$

where $I_{(a,b]}(x_t)$ denotes the *indicator function* of the set $(a, b]$. That is, for a set A

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} .$$

What this means is that the proportion of the sequence $\{x_t\}_{t=0}^{\infty}$ that is in the interval $(a, b]$ is given by the interval's length $b \Leftrightarrow a$.

1.2 Sample Space

A useful first step in building a mathematical model with which to analyze data that might arise from the examples of Section 1.1 is to list all possible outcomes. The set of all possible outcomes is called the *sample space*, which, following custom, we shall denote by Ω. This listing of all possible outcomes is context dependent and is not unique. There may be many acceptable listings for a given application. What is important is that the listing be exhaustive; that is, there is a sample point to represent every outcome. We now illustrate.

Suppose one should like to analyze the single roll bets in the game of craps (described in Section 1.1.1). The set of ordered pairs of the numbers from 1 to 6,

$$\Omega_p = \left\{ \begin{array}{cccccc} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6) \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6) \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6) \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6) \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6) \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array} \right\}$$

consisting of 36 points would be an adequate sample space. It exhausts the possibilities (see Figure 1.2). Since order is not important in deciding the outcome of any single roll bet, the sample space

$$\Omega_u = \left\{ \begin{array}{cccccc} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6) \\ (2,2), & (2,3), & (2,4), & (2,5), & (2,6), & (3,3) \\ (3,4), & (3,5), & (3,6), & (4,4), & (4,5), & (4,6) \\ (5,5), & (5,6), & (6,6) & & & \end{array} \right\}$$

consisting of 21 points would also be adequate. As a practical matter, it is usually easier to work with Ω_p.

An analysis of the field bet requires only knowledge of the sum of the spots showing on the thrown dice; therefore the sample space

$$\Omega_s = \{2,3,4,5,6,7,8,9,10,11,12\}$$

consisting of 11 points would suffice.

Bets on the pass line require an indeterminate number of rolls to decide. In principle, it is possible that neither the shooter's point nor a seven will ever be rolled and the game could continue indefinitely. One possible choice of Ω is all possible infinite sequences of ordered pairs (n_1, n_2) of the numbers from 1 to 6. If we let $\Omega_{p,i}$ be a copy of Ω_p above, then all possible sequences of ordered pairs can be written as the Cartesian product of the $\Omega_{p,i}$

$$\Omega_p^\infty = \mathsf{X}_{i=1}^\infty \Omega_{p,i}.$$

This choice of Ω is rich enough to support the analysis of any bet on the table. For example, consider the point

$$\omega = [(1,6),\,(1,3),\,(6,1),\,(4,3),\,\ldots]$$

from Ω_p^∞. If the first throw is a come out roll, then bets on the pass line would win, lose, win in the first four rolls. Field bets would win once and lose thrice in the first four rolls. Hardway bets on the four would lose on each of the first four rolls.

If one only wanted to analyze place bets, come bets, and bets on the pass line, it would be enough to keep track of the sum of the dice on each roll. In this case the sample space could consist of all infinite sequences of the numbers from 2 to 12, namely,

$$\Omega_s^\infty = \times_{i=1}^\infty \Omega_{s,i},$$

where each $\Omega_{s,i}$ is a copy of Ω_s above.

For the game of keno (Section 1.1.2), the sample space comprised of all sequences of length 20 made up of the numbers from 1 to 80 where no number is repeated within the sequence is adequate. We could let order be important so that $(3, 5, 6, \ldots)$ and $(6, 3, 5, \ldots)$ count as different sequences, or we could let order be unimportant so they count as the same sequence. The order of draws is not important in determining whether or not a keno bet wins, so either is acceptable.

For the coin tossing example (Section 1.1.3), one could put $\Omega = (0, 1]$ or take Ω to be all possible sequences of 0's and 1's. Recall that endpoints do not matter so that $\Omega = (0, 1), [0, 1)$, or [0,1] are also acceptable choices.

For the triangular map example (Section 1.1.4), one could put $\Omega = [0, 1]$ or take Ω to be all the irrational numbers in [0,1].

1.3 Events

An *event* E is a subset of the sample space Ω. It may be empty, a proper subset of the sample space, or the sample space itself. Situations such as Section 1.1 describe are often called *experiments*. An event *occurs* if the experiment is performed, ω is the outcome, and $\omega \in E$. We illustrate using the game of craps, which is described in Section 1.1.1.

The event "snake eyes" is a single roll bet that pays 30 to 1. Relative to the sample space Ω_p consisting of all pairs (n_1, n_2) of the numbers 1 through 6, which is displayed in Section 1.2 and again later in this section, snake eyes is the singleton set

$$E = \{(1, 1)\}.$$

The shooter rolls. If the dice land $\omega = (1, 1)$, then snake eyes occurs. The event "any seven" is another single roll bet. It pays 4 to 1. It is

$$E = \{(1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3)\}.$$

As noted in Section 1.2, multiple roll bets such as a place bet on the eight to win require a more complicated sample space. For place bets, the sample space Ω_s^∞ consisting of all infinite sequences of the numbers from 2 to 12 is adequate.

With this sample space, one wins a place bet on the eight if the event

$$\begin{aligned} E &= \{\text{“an eight before a seven”}\} \\ &= \Big\{\omega : \omega = (\omega_1, \omega_2, \ldots),\ \min\{i : \omega_i = 8\} < \min\{i : \omega_i = 7\}\Big\} \end{aligned}$$

occurs. For instance, if the shooter rolls the sequence

$$\omega = (\omega_1, \omega_2, \omega_3, \ldots) = (5, 6, 8, \ldots),$$

then a place bet on the eight wins and E occurs. If the shooter rolls

$$\omega = (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \ldots) = (6, 6, 9, 4, 7, \ldots),$$

then it loses and E does not occur.

Subsetting, or *containment*, is indicated by $A \subset B$, which means that every ω that is in A is also in B. The definition can be written symbolically as

$$A \subset B \quad \overset{\text{def}}{\Leftrightarrow} \quad (\omega \in A \Rightarrow \omega \in B),$$

which reads

A is a subset of B if and only if ω in A implies that ω is in B.

As an example, consider $A = \{x : x < 5\}$ and $B = \{x : x < 8\}$:

$$x \in A \Rightarrow x < 5$$

$$x < 5 \Rightarrow x < 8$$

$$x < 8 \Rightarrow x \in B,$$

therefore

$$x \in A \Rightarrow x \in B.$$

By the definition,

$$A \subset B.$$

Two events A and B are *equal*, written $A = B$, if they contain the same elements. This can be written symbolically as

$$A = B \quad \overset{\text{def}}{\Leftrightarrow} \quad (A \subset B \text{ and } B \subset A).$$

To prove equality, one must take an arbitrary element ω from A and show that it is in B and then take an arbitrary element ω from B and show that it is in A.

The *union* of A and B, written $A \cup B$, is the set of elements that belong to either A or B, or both,

$$A \cup B = \{\omega : \omega \in A \text{ or } \omega \in B\}.$$

The *intersection* of A and B, written $A \cap B$, is the set of elements that belong to both A and B,

$$A \cap B = \{\omega : \omega \in A \text{ and } \omega \in B\}.$$

The *complement* of A, written as $\tilde{A}$ or $\sim A$, is the set of elements in Ω that are not in A,

$$\tilde{A} = \{\omega \in \Omega : \omega \notin A\}.$$

We illustrate complement, union, and intersection with some single roll bets from craps.

$$\Omega_p = \left\{ \begin{array}{cccccc} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6) \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6) \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6) \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6) \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6) \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array} \right\} \text{ “sample space”}$$

$$F = \left\{ \begin{array}{cccccc} (1,1), & (1,2), & (1,3), & (2,1), & (2,2), & (3,1), \\ (3,6), & (4,5), & (4,6), & (5,4), & (5,5), & (5,6), \\ (6,3), & (6,4), & (6,5), & (6,6) & & \end{array} \right\} \text{ “field”}$$

$$\tilde{F} = \left\{ \begin{array}{cccccc} (1,4), & (1,5), & (1,6), & (2,3), & (2,4), & (2,5), \\ (2,6), & (3,2), & (3,3), & (3,4), & (3,5), & (4,1), \\ (4,2), & (4,3), & (4,4), & (5,1), & (5,2), & (5,3), \\ (6,1), & (6,2) & & & & \end{array} \right\} \text{ “no field”}$$

$$H = \{(4,4)\} \quad \text{“hard eight”}$$

$$E = \{(2,6),(6,2),(3,5),(5,3)\} \quad \text{“easy eight”}$$

$$H \cup E = \{(4,4),(2,6),(6,2),(3,5),(5,3)\} \quad \text{“any eight”}$$

$$H \cap E = \{\} = \emptyset \quad \text{“empty set.”}$$

The union and intersection operations are *commutative*, *associative*, and *distributive*. Specifically, if A, B, and C are subsets of Ω, then union and intersection commute

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A,$$

associate

$$A \cup (B \cup C) = (A \cup B) \cup C$$

$$A \cap (B \cap C) = (A \cap B) \cap C,$$

and distribute

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

Also useful are *DeMorgan's laws*

$$\sim(A \cup B) = \tilde{A} \cap \tilde{B}$$

$$\sim(A \cap B) = \tilde{A} \cup \tilde{B}.$$

The first is proved as follows:

$$\begin{aligned} \omega^o \in (\widetilde{A \cup B}) &\Rightarrow \omega^o \notin (A \cup B) \\ &\Rightarrow \omega^o \notin \{\omega : \omega \in A \text{ or } \omega \in B\} \\ &\Rightarrow \omega^o \notin A \text{ and } \omega \notin B \\ &\Rightarrow \omega^o \in \tilde{A} \text{ and } \omega \in \tilde{B} \\ &\Rightarrow \omega^o \in \{\omega : \omega \in \tilde{A} \text{ and } \omega \in \tilde{B}\} \\ &\Rightarrow \omega^o \in (\tilde{A} \cap \tilde{B}). \end{aligned}$$

Thus

$$(\widetilde{A \cup B}) \subset \tilde{A} \cap \tilde{B}.$$

A similar argument yields

$$\tilde{A} \cap \tilde{B} \subset (\widetilde{A \cup B}),$$

which proves the result.

It is possible to take the union or intersection of a countable number of sets. A point is in $\bigcup_{i=1}^{\infty} A_i$ if it is in at least one of the A_i; that is,

$$\begin{aligned} \bigcup_{i=1}^{\infty} A_i &= \{\omega : \exists\, i \text{ in } 1 \leq i < \infty \ni \omega \in A_i\} \\ &= \{\omega : \omega \in A_i \text{ for some } i \text{ in } 1 \leq i < \infty\}. \end{aligned}$$

A point is in $\bigcap_{i=1}^{\infty} A_i$ if it is in every one of the A_i; that is,

$$\begin{aligned} \bigcap_{i=1}^{\infty} A_i &= \{\omega : 1 \leq i < \infty \Rightarrow \omega \in A_i\} \\ &= \{\omega : \omega \in A_i \text{ for every } i \text{ in } 1 \leq i < \infty\}. \end{aligned}$$

To illustrate, if the game is craps, the sample space is Ω_s^{∞}, and E_i is the event "the shooter rolls 8 on roll i", then $\bigcap_{i=1}^{\infty} E_i$ contains the single point $\omega = (8, 8, \ldots)$ and $\bigcup_{i=1}^{\infty} E_i$ is the set of all sequences that have at least one 8 as an element. Other examples are

$$\bigcup_{i=1}^{\infty} [\frac{1}{i}, 1] = \{x : \frac{1}{i} \leq x \leq 1 \text{ for some } 1 \leq i < \infty\} = (0, 1]$$

and

$$\bigcap_{i=1}^{\infty} [\frac{1}{i}, 1] = \{x : \frac{1}{i} \leq x \leq 1 \text{ for every } 1 \leq i < \infty\} = [1].$$

DeMorgan's Laws apply to countable intersections and unions:

$$\sim \bigcup_{i=1}^{\infty} A_i = \bigcap_{i=1}^{\infty} \tilde{A}_i,$$

$$\sim \bigcap_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} \tilde{A}_i.$$

Consider the event E consisting of those sample points

$$\omega = (\omega_1, \omega_2, \omega_3, \ldots) \in \Omega_s^{\infty}$$

for which an infinite number of the ω_i are equal to 8. Let E_i be the event "8 on roll i" as above. If $\omega \in E$, then ω is in infinitely many of the events in the sequence $E_1, E_2, E_3, \ldots$; that is, events in the sequence occur infinitely often. For this reason, the event E is called "E_i *infinitely often*" and is written $[E_i \text{ i.o.}]$. We can characterize $[E_i \text{ i.o.}]$ in terms of countable unions and intersections of the E_i as follows:

$$\omega \in [E_i \text{ i.o.}] \quad \Leftrightarrow \quad \text{for every } I \geq 1 \text{ there is an } i \geq I \text{ such that } \omega \in E_i$$

$$\omega \in [E_i \text{ i.o.}] \quad \Leftrightarrow \quad \text{for every } I \geq 1 \text{ we have } \omega \in \bigcup_{i=I}^{\infty} E_i$$

$$\omega \in [E_i \text{ i.o.}] \quad \Leftrightarrow \quad \omega \in \bigcap_{I=1}^{\infty} \bigcup_{i=I}^{\infty} E_i.$$

Therefore

$$[E_i \text{ i.o.}] = \bigcap_{I=1}^{\infty} \bigcup_{i=I}^{\infty} E_i.$$

Another event of interest is E_i occurs all but a finite number of times. By a similar logic this event is $\bigcup_{I=1}^{\infty} \bigcap_{i=I}^{\infty} E_i$. One sometimes sees the notation $\liminf E_i$ to mean $\bigcup_{I=1}^{\infty} \bigcap_{i=I}^{\infty} E_i$ and $\limsup E_i$ to mean $\bigcap_{I=1}^{\infty} \bigcup_{i=I}^{\infty} E_i$.

Two events A and B are *disjoint* (or *mutually exclusive*) if $A \cap B = \emptyset$. A sequence of events $A_1, A_2, \ldots$ is disjoint (or mutually exclusive) if the sets A_i are pairwise disjoint; that is, if $A_i \cap A_j = \emptyset$ for all $i \neq j$. A sequence of events is called *exhaustive* if $\Omega = \bigcup_{i=1}^{\infty} A_i$. In the coin tossing experiment (Section 1.1.3) with $\Omega = [0, 1]$, the sets

$$A_0 = [0], \quad A_i = (\frac{1}{i+1}, \frac{1}{i}] \quad i = 1, 2, \ldots$$

are mutually exclusive and exhaustive.

1.4 Probability Spaces

1.4.1 Coin Tossing: One Dimension

Probability theory is designed to permit a mathematical analysis of the situations described in the examples of Section 1.1 and similar situations. The coin tossing example serves best to illustrate the ideas. Recall that the salient feature of that example is that it is a physical mechanism that generates numbers ω in (0,1] for which it is reasonable to state that the chance that ω is in some subinterval $(a, b]$ is the length $b - a$ of that subinterval. A sample space for the coin tossing example is $\Omega = (0, 1]$. With this choice of sample space, an event A is a subset of (0,1]. To determine if an event occurs, one tosses the coin long enough to determine if

$$\omega = t_1 \frac{1}{2} + t_2 \frac{1}{4} + t_3 \frac{1}{8} + \cdots$$

is in A, where t_i is 1 if the coin lands heads on toss i and is 0 if it lands tails.

The probability function P assigns to an event A the chance $P(A)$ that it will occur. If the event A is a subinterval of $\Omega = (0, 1]$, then its probability is its length, viz.

$$P\Big\{(a, b]\Big\} = b - a.$$

If A is the union of two disjoint subintervals, then its probability is the sum of the lengths of the two subintervals,

$$P\Big\{(a, b] \cup (c, d]\Big\} = b - a + d - c.$$

From this together with $P(\Omega) = P\{(0, 1]\} = 1$, one can infer that

$$P\Big\{\sim (a, b]\Big\} = 1 - P\Big\{(a, b]\Big\}.$$

If A is the finite union of disjoint subintervals, that is, $A = \bigcup_{i=1}^{n} (a_i, b_i]$ and $(a_i, b_i] \cap (a_j, b_j] = \emptyset$ for $i \neq j$, then

$$P(A) = \sum_{i=1}^{n} (b_i - a_i).$$

The empty set has no length and cannot occur so it is natural to put

$$P(\emptyset) = 0.$$

If we let $\mathcal{A}$ denote the collection of sets of the form $(a, b]$ with $0 \leq a < b \leq 1$, finite unions of such sets, plus the empty set $\emptyset$, then, at present, we have P defined over $\mathcal{A}$. Note that (i) the empty set is in $\mathcal{A}$, (ii) $\tilde{A} \in \mathcal{A}$ whenever A is, and (iii) $A \cup B \in \mathcal{A}$ whenever A and B are. A collection of sets with these three properties is called an *algebra* of sets. This is as far as the notion of length will take us. Unfortunately it is not far enough. We will need P to be defined for a larger class of events than the algebra $\mathcal{A}$.

A *σ-algebra* (or *σ-field* or *Borel field*) is a collection of sets $\mathcal{B}$ that satisfy the following three properties: (i) $\emptyset \in \mathcal{B}$ (the empty set is a member of $\mathcal{B}$), (ii) if $B \in \mathcal{B}$, then $\tilde{B} \in \mathcal{B}$ ($\mathcal{B}$ is closed under complementation), and (iii) if $B_1, B_2, \ldots \in \mathcal{B}$, then $\bigcup_{i=1}^{\infty} B_i \in \mathcal{B}$ ($\mathcal{B}$ is closed under countable union).

Let $\mathcal{F}$ denote the smallest σ-algebra that contains $\mathcal{A}$ (Problem 11). Note that $\mathcal{F}$ will contain intervals of the form (a, b), $[a, b)$, and $[a, b]$ because they can be constructed from countable unions and intersections of sets of the form $(a, b] \in \mathcal{A}$. For instance, $(a, b) = \bigcup_{i=1}^{\infty}(a, b - 1/i]$. We can extend the definition of P to $\mathcal{F}$.

Before doing so, let us introduce or recall, as the case may be, the definitions of the *supremum* and *infimum* of a subset B of the real line, denoted $\sup B$ and $\inf B$, respectively. If B has an upper bound, that is, there is some real number b such that $x \le b$ for all $x \in B$, then $\sup B$ is the smallest such b. If B does not have an upper bound, then $\sup B = \infty$. Define $\sup \emptyset = -\infty$. For example, $\sup(0, 1] = \sup(0, 1) = 1$, $\sup(-\infty, \infty) = \infty$, and $\sup\{x : x = 1 - 1/i,\ i = 1, 2, \ldots\} = 1$. Inf B is defined analogously: If B has a lower bound, then $\inf B$ is the largest lower bound. If B does not have a lower bound, then $\inf B = -\infty$. $\inf \emptyset = \infty$. As examples, $\inf[0, 1) = \inf(0, 1) = 0$, $\inf(-\infty, \infty) = -\infty$, and $\inf\{x : x = 1/i,\ i = 1, 2, \ldots\} = 0$. Supremum and infimum are related by $\inf B = -\sup\{-x : x \in B\}$. If $A \subset B$, then $\sup A \le \sup B$ and $\inf B \le \inf A$.

Returning to the extension of P to $\mathcal{F}$, for $F \in \mathcal{F}$ define

$$P(F) = \inf \sum_{i=1}^{\infty} P(A_i),$$

where $\{A_i\}$ ranges over all sequences $A_1, A_2, \ldots$ from $\mathcal{A}$ such that $F \subset \bigcup_{i=1}^{\infty} A_i$. That is,

$$P(F) = \inf\Big\{p : p = \sum_{i=1}^{\infty} P(A_i),\ F \subset \bigcup_{i=1}^{\infty} A_i,\ A_i \in \mathcal{A}\Big\}.$$

The probability function so defined will satisfy three properties: (i) $P(F) \ge 0$ for all $F \in \mathcal{F}$ (P is *positive*), (ii) $P(\Omega) = 1$, and (iii) if $F_1, F_2, \ldots \in \mathcal{F}$ are disjoint, then $P(\bigcup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} P(F_i)$ (P is *countably additive*). Because the function P takes as its argument the elements of $\mathcal{F}$, which are sets, P is called a *set function*. Properties (i) through (iii) are called the *axioms of probability*. Thus, a *probability function* P is a positive, countably additive, set function that is defined over a σ-algebra $\mathcal{F}$ of subsets of a sample space Ω and satisfies $P(\Omega) = 1$. A *probability space* is the triplet $(\Omega, \mathcal{F}, P)$.

For later reference, we note that if P is countably additive, then it must also be *finitely additive*; that is, if $F_1, F_2, \ldots, F_n \in \mathcal{F}$ are disjoint, then $P(F) = \sum_{i=1}^{n} P(F_i)$. This is proved by noting that sequence $F_1, F_2, \ldots, F_n, \emptyset, \emptyset, \ldots$ is disjoint and that for this sequence $\sum_{i=1}^{n} P(F_i) = \sum_{i=1}^{\infty} P(F_i)$ and $\bigcup_{i=1}^{n} F_i = \bigcup_{i=1}^{\infty} F_i$.

1.4.2 Coin Tossing: Two Dimensions

As the coin tossing example suggests, probability is akin to the notions of length, area, and volume and we will make use of this analogy frequently in the sequel. Because area lends itself better to graphical illustration than length or volume, we shall extend the coin tossing example to two dimensions.

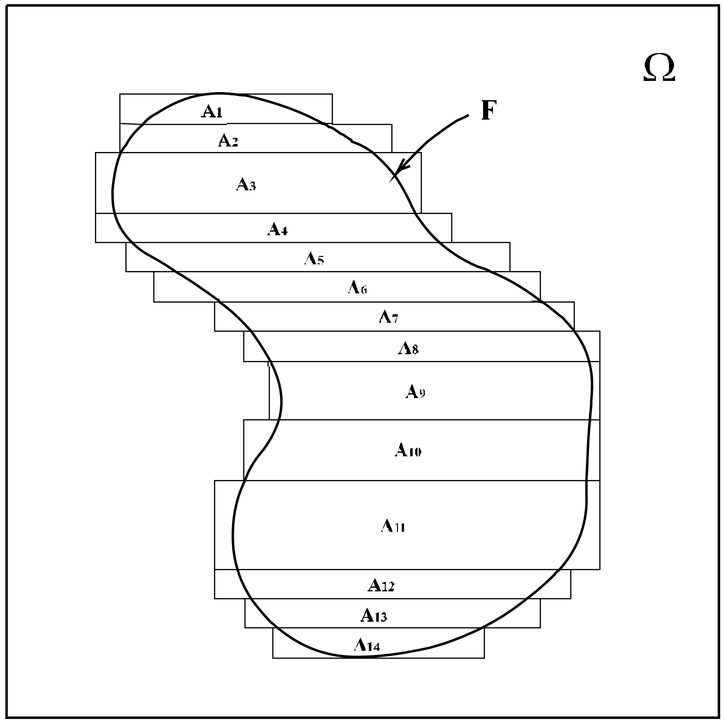

Figure 1.4. A Covering of an Irregularly Shaped Set F by Disjoint Rectangles. The probability of F is approximated by the smallest value of $\sum_{t=1}^{n} \text{Area}(A_i)$ that can be achieved by rectangles such as those shown. The approximation converges to $P(F)$ as n tends to infinity.

Consider performing the coin tossing experiment of Section 1.1.3 twice with two different coins and letting the outcome be recorded as the two-dimensional point (x, y) where x corresponds to the tosses of the first coin and y to the tosses of the second. The relevant sample space is $\Omega^2 = (0, 1] \times (0, 1]$. The probability of a rectangle is its area $P\{(a, b] \times (c, d]\} = \text{Area}\{(a, b] \times (c, d]\} = (b \Leftrightarrow a) \times (d \Leftrightarrow c)$. The probability of the union $\bigcup_{i=1}^{n} A_i$ of disjoint rectangles of the form $A_i = (a_i, b_i] \times (c_i, d_i]$ is the sum of the areas of the rectangles: $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \text{Area}(A_i)$. The collection $\mathcal{A}$ consisting of $\emptyset$ and all finite unions of rectangles of the form $(a, b] \times (c, d]$ is an algebra. Let $\mathcal{F}$ be the smallest σ-algebra that contains $\mathcal{A}$. A covering $\bigcup_{i=1}^{n} A_i$ of $F \in \mathcal{F}$ by a union of disjoint rectangles $A_i \in \mathcal{A}$ is shown in Figure 1.4. The probability of F is approximated by the smallest value of $\sum_{t=1}^{n} \text{Area}(A_i)$ that can be achieved by rectangles such as those shown. Indeed, this is how the area of irregular objects is computed in practice. The approximation converges to $P(F)$ as n tends to infinity. More generally, to accommodate sets less regularly shaped than shown in Figure 1.4, the probability of $F \in \mathcal{F}$ is

$$P(F) = \inf\Big\{p : p = \sum_{i=1}^{\infty} \mathrm{Area}(A_i),\ F \subset \bigcup_{i=1}^{\infty} A_i,\ A_i \in \mathcal{A}\Big\}.$$

1.4.3 Craps: Single Roll Bets

To illustrate how these ideas can be extended beyond the coin tossing example, we shall apply them to the game of craps described in Subsection 1.1.1. For a given bet, our goal shall be to describe an appropriate sample space Ω, a σ-algebra $\mathcal{F}$, and a probability function P.

To analyze the single roll bets, we take Ω_p, defined in Section 1.2, as the sample space. The σ-algebra $\mathcal{F}_p$ over which P is defined is the collection of all possible subsets of Ω_p. $\mathcal{F}_p$ contains $\emptyset$, all singleton sets, of which there are 36, all sets containing two elements, of which there are $1260 = 36 \times 35$, and so on. Every outcome in Ω_p is equally likely – presumably state gaming commissions make sure that this is true – so the probability assigned to any singleton set is $P(\{\omega\}) = 1/36$.

We extend the definition beyond singleton sets by making P be finitely additive. Thus, the probability assigned to an event with two elements is $P(\{w_1, w_2\}) = P(\{w_1\}) + P(\{w_2\}) = 1/36 + 1/36 = 1/18$, and, in general, the probability of any event is the number of points in it divided by 36. For example, the probability that a place bet on the 4 wins on the first roll is $P[\{(1,3),(2,2),(3,1)\}] = 1/12$.

1.4.4 Craps: Multiple Roll Bets

If we consider two tosses of the dice, we would let $\Omega_{p,1}$ and $\Omega_{p,2}$ each be copies of Ω_p, and let the sample space be the Cartesian product

$$\begin{aligned}
\Omega_p^2 &= \Omega_{p,1} \times \Omega_{p,2} \\
&= \{(\omega_1, \omega_2) : \omega_1 \in \Omega_{p,1}, \omega_2 \in \Omega_{p,2}\} \\
&= \{[(n_1, n_2), (n_3, n_4)] : (n_1, n_2) \in \Omega_{p,1}, (n_3, n_4) \in \Omega_{p,2}\}.
\end{aligned}$$

Let $\mathcal{F}_{p,1}$ and $\mathcal{F}_{p,2}$ each be copies of $\mathcal{F}_p$. The σ-algebra $\mathcal{F}_p^2$ over which P is defined is the smallest σ-algebra that contains

$$\mathcal{F}_{p,1} \times \mathcal{F}_{p,2} = \{E_1 \times E_2 : E_1 \in \mathcal{F}_{p,1}, E_2 \in \mathcal{F}_{p,2}\},$$

where

$$\begin{aligned}
E_1 \times E_2 &= \{(\omega_1, \omega_2) : \omega_1 \in E_1, \omega_2 \in E_2\} \\
&= \{[(n_1, n_2), (n_3, n_4)] : (n_1, n_2) \in E_1, (n_3, n_4) \in E_2\}.
\end{aligned}$$

The operation of taking the smallest σ-algebra containing some class of sets $\mathcal{A}$ is often written $\sigma(\mathcal{A})$ so that

$$\mathcal{F}_p^2 = \sigma(\mathcal{F}_{p,1} \times \mathcal{F}_{p,2}).$$

As above, each of the outcomes $(\omega_1, \omega_2) \in \Omega_p^2$ is equally likely and there are $36 \times 36 = 1,296$ such points so that, as above, to assign a probability $P(E)$ to an event E is a matter of counting up the number of points in E and dividing by 1,296.

Let us consider the probability that a place bet on the 4 wins on the second roll. This is the event

$$E = \left\{ \begin{array}{llllll} (1,1), & (1,2), & (1,4), & (1,5), & (2,1), & (2,3), \\ (2,4), & (2,6), & (3,2), & (3,3), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,4), & (4,5), & (4,6), & (5,1), \\ (5,3), & (5,4), & (5,5), & (5,6), & (6,2), & (6,3), \\ (6,4), & (6,5), & (6,6) & & & \end{array} \right\} \times \left\{ \begin{array}{c} (1,3), \\ (2,2), \\ (3,1) \end{array} \right\}.$$

There are 27×3 points in this event so that

$$P(E) = \frac{27 \times 3}{36 \times 36} = \frac{3}{4} \times \frac{1}{12}.$$

By continuing along these lines, one can determine the probability space $(\Omega_p^n, \mathcal{F}_p^n, P)$ for n rolls. One would conclude that the probability of the event $E_i =$ "a place bet on the 4 wins on roll i" is

$$P(E_i) = \frac{1}{12} \times \left(\frac{3}{4}\right)^{i-1}.$$

Note that these probability spaces are consistent in that if one computed the probability that a place bet wins on roll i in any of them for which $n \geq i$, one would get the same answer. That is,

$$P(E_i) = P(E_i \times \Omega_{p,i+1} \times \cdots \times \Omega_{p,n}).$$

For multiple roll bets the sample space is

$$\Omega_p^\infty = \mathsf{X}_{i=1}^\infty \Omega_{p,i},$$

as described in Section 1.2, where each $\Omega_{p,i}$ is a copy of Ω_p. The σ-algebra on which P is defined is constructed as follows. Let $\mathcal{A}$ be the collection of sets formed by taking events E^n from $\mathcal{F}_p^n$ and appending an infinite number of copies of Ω_p for $n = 1, 2, \ldots$. That is,

$$\mathcal{A} = \bigcup_{n=1}^\infty \{A : A = E_n \times \Omega_{p,n+1} \times \Omega_{p,n+2} \times \cdots, \; E_n \in \mathcal{F}_p^n, \; \Omega_{p,i} = \Omega_p\}.$$

Probabilities are assigned to $A \in \mathcal{A}$ according to

$$P(E_n \times \Omega_{p,n+1} \times \Omega_{p,n+2} \times \cdots) = P(E_n).$$

Put $\mathcal{F}_p^\infty = \sigma(\mathcal{A})$. The definition is extended to $F \in \mathcal{F}_p^\infty$ by putting $P(F) = \inf \sum_{i=1}^\infty P(A_i)$, where the sequence $A_1, A_2, \ldots$ ranges over all disjoint sequences of sets from $\mathcal{A}$ whose union contains F. The triple $(\Omega_p^\infty, \mathcal{F}_p^\infty, P)$ so constructed is a probability space.

Notations such as

$$E = \bigcup_{n=1}^{\infty} E_n \times \Omega_{p,n+1} \times \Omega_{p,n+2} \times \cdots \quad \text{and} \quad P(E_n \times \Omega_{p,n+1} \times \Omega_{p,n+2} \times \cdots)$$

are cumbersome. Henceforth, we will let the fact that copies of Ω_p must be appended to $E_n \in \mathcal{F}_p^n$ in order to get membership in $\mathcal{F}_p^\infty$ be understood and we will write

$$E = \bigcup_{n=1}^{\infty} E_n \quad \text{and} \quad P(E_n)$$

instead.

In applications, the only probabilities one actually needs to compute are probabilities for sets from $\mathcal{A}$, which is a matter of counting as we have seen. For example, to compute the probability of the event $E =$ "a place bet on the 4 wins," one notes that the events $E_i =$ "a place bet on the 4 wins on roll i" are disjoint and that

$$E = \bigcup_{i=1}^{\infty} E_i.$$

By countable additivity and the definition of P, the probability that a place bet on the 4 wins is

$$P(E) = \sum_{i=1}^{\infty} P(E_i) = \frac{1}{12}\sum_{i=1}^{\infty}\left(\frac{3}{4}\right)^{i-1} = \frac{1}{12} \times 4 = \frac{1}{3}.$$

This can be compared to the true odds of 3 for 1 — 3 for 1 is the same as 2 to 1 — in Table 1.1. Our computation agrees!

By the same logic used to work out the place bet, $A_i =$ "a place bet on the 4 is decided on roll i" occurs with probability $P(A_i) = (1/4)(3/4)^{i-1}$ so that $P(\bigcup_{i=1}^{\infty} A_i) = 1$. Since $\omega \in \bigcup_{i=1}^{\infty} A_i \Leftrightarrow \exists\, i \ni \omega \in A_i$, the bet is decided in a finite number of rolls with probability 1. The probability of other multiple roll bets will be easier to work out after some more ideas from probability theory are in place, in particular, the notion of conditional probability.

1.4.5 Coin Tossing: Countable Dimensions

We extended the coin tossing example from one dimension to two. The same construction can be used to extend it to n dimensions. Having extended to n dimensions, the extension to $\Omega^\infty = \times_{i=1}^{\infty}(0,1]$ is the same as for multiple roll bets in craps: $\mathcal{A}$ is the collection of all events from the finite dimensional spaces with an infinite number of copies of $(0,1]$ appended. Probabilities $P(A)$ are assigned to events $A \in \mathcal{A}$ using probabilities from the finite dimensional spaces. $\mathcal{F} = \sigma(\mathcal{A})$. For $F \in \mathcal{F}$, $P(F) = \inf \sum_{i=1}^{\infty} P(A_i)$, where the sequence $A_1, A_2, \ldots$ ranges over all disjoint sequence of sets from $\mathcal{A}$ whose union contains F. The reader who would like to pursue the ideas behind these constructions in more depth should see Royden 1988, Chapter 11.

1.5 Properties of Probability Spaces

We can summarize the previous section in the following definition.

DEFINITION 1.1 A probability space is the triple $(\Omega, \mathcal{F}, P)$ consisting of a sample space Ω, a σ-algebra $\mathcal{F}$ of subsets of Ω, and a function P defined on $\mathcal{F}$ that satisfies the axioms of probability:

1. $P(A) \geq 0$ for all $A \in \mathcal{F}$.
2. $P(\Omega) = 1$.
3. If $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

In this section we derive some useful properties of probability spaces that follow from the definition, which we can summarize as follows.

PROPOSITION 1.1 Let $(\Omega, \mathcal{F}, P)$ be a probability space and let A, B, and $A_1, A_2, \ldots$ be sets in $\mathcal{F}$. Then

1. $P(\emptyset) = 0$.
2. $P(A) \leq 1$.
3. $P(A) + P(\tilde{A}) = 1$.
4. $P(A \cap B) + P(A \cap \tilde{B}) = P(A)$.
5. $P(A \cup B) = P(A) + P(B) \Leftrightarrow P(A \cap B)$.
6. If $A \subset B$, then $P(A) \leq P(B)$.
7. If A_1, A_2, $\ldots$ are mutually exclusive and exhaustive, then $P(A) = \sum_{i=1}^{\infty} P(A \cap A_i)$.
8. $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$ (countable subadditivity).

As pointed out in the previous section, probability is akin to area. More precisely, the properties of the probability function are similar to those of area if all sets under consideration are confined to some bounded region that has total area 1, such as $\Omega = (0,1] \times (0,1]$. As may be seen from inspection of Figure 1.5, the properties listed in Proposition 1.1 are properties of area. We shall verify a few of them rigorously in the remainder of this section.

To show Property 4, note that $\Omega = B \cup \tilde{B}$. Therefore, by the distributive laws,

$$A = A \cap \Omega = A \cap (B \cup \tilde{B}) = (A \cap B) \cup (A \cap \tilde{B}).$$

Moreover, by the commutative and associative laws,

$$(A \cap B) \cap (A \cap \tilde{B}) = A \cap (B \cap \tilde{B}) = A \cap \emptyset = \emptyset,$$

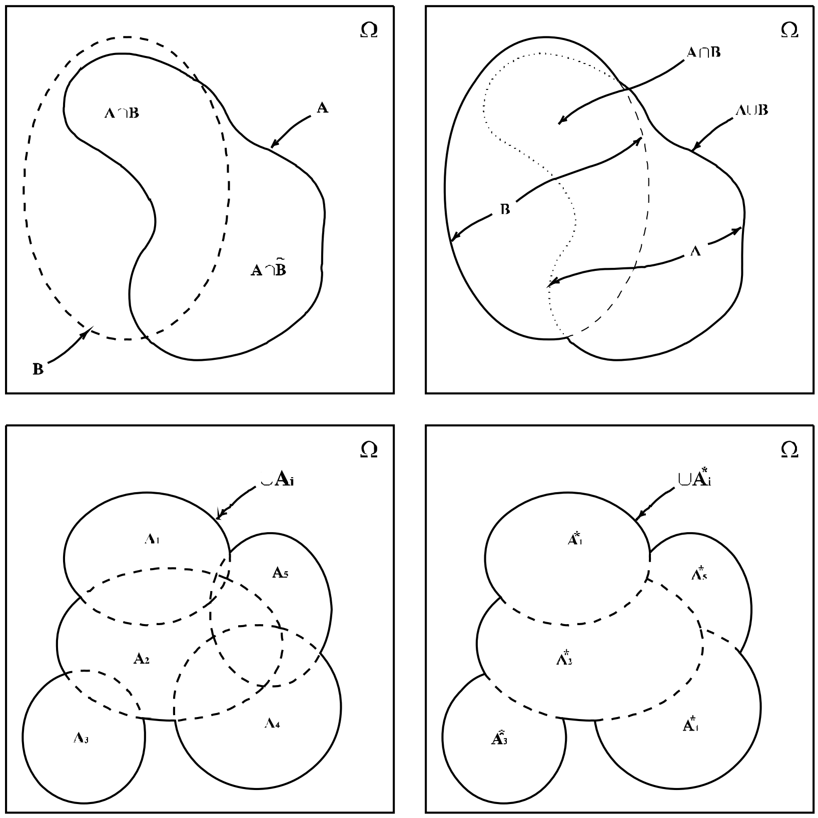

Figure. 1.5. Illustration of Proposition 1.1. The sample space is $\Omega = (0,1] \times (0,1]$. The area of a set A is equal to its probability $P(A)$. The upper left panel shows $P(A \cap B) + P(A \cap \tilde{B}) = P(A)$. The upper right panel shows $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. The lower two panels show that $\bigcup_{i=1}^{5} A_i = \bigcup_{i=1}^{5} A_i^*$ where $A_1^* = A_1$ and $A_i^* = A_i \cap [\sim (\bigcup_{j=1}^{i-1} A_j)]$. Therefore, $P(\bigcup_{i=1}^{5} A_i) = P(\bigcup_{i=1}^{5} A_i^*) = \sum_{i=1}^{5} P(A_i^*) \leq \sum_{i=1}^{5} P(A_i)$.

which shows that $(A \cap B)$ and $(A \cap \tilde{B})$ are disjoint. By finite additivity, which is a consequence of countable additivity as verified in Section 1.4, we have

$$P(A) = P(A \cap B) + P(A \cap \tilde{B}).$$

This proves Property 4.

To show Property 5, we apply the distributive law to get

$$(A \cap \tilde{B}) \cup B = (A \cup B) \cap (\tilde{B} \cup B) = (A \cup B) \cap \Omega = A \cup B.$$

The sets $(A \cap \tilde{B})$ and B are disjoint because, by the associative law,

$$(A \cap \tilde{B}) \cap B = A \cap (\tilde{B} \cap B) = A \cap \emptyset = \emptyset.$$

Therefore, by finite additivity,

$$P(A \cup B) = P(A \cap \tilde{B}) + P(B).$$

Using $P(A) = P(A \cap B) + P(A \cap \tilde{B})$ from above,

$$P(A \cup B) = P(A) \Leftrightarrow P(A \cap B) + P(B),$$

which proves Property 5.

To show Property 7, we apply the distributive law to get

$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{\infty} A_i \right) = \bigcup_{i=1}^{\infty} (A \cap A_i).$$

The sets $(A \cap A_1), (A \cap A_2), \ldots$ are disjoint because the associative law implies that for $i \neq j$ we have

$$(A \cap A_i) \cap (A \cap A_j) = (A \cap A) \cap (A_i \cap A_j) = A \cap \emptyset = \emptyset.$$

Countable additivity implies

$$P(A) = \sum_{i=1}^{\infty} P(A \cap A_i),$$

which proves Property 7.

Lastly, we shall verify Property 8. As indicated in the bottom two panels of Figure 1.5, the idea is to show that $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} A_i^*$, where $A_1^*, A_2^*, \ldots$ are disjoint and $A_i^* \subset A_i$ so that $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i^*) \leq \sum_{i=1}^{\infty} P(A_i)$ by countable additivity and Property 6.

The A_i^* are defined by

$$A_1^* = A_1, \quad A_i^* = A_i \cap \left[\sim \left(\bigcup_{k=1}^{i-1} A_k \right) \right] = A_i \cap \left(\bigcap_{k=1}^{i-1} \tilde{A}_k \right)$$

for $i = 1, 2, \ldots$, where the last equality is due to DeMorgan's laws.

To see that $A_1^*, A_2^*, \ldots$ are disjoint, let $i < j$ and apply the commutative and associative laws repeatedly to get

$$\begin{aligned}
A_i^* \cap A_j^* &= A_i \cap \left(\bigcap_{k=1}^{i-1} \tilde{A}_k \right) \cap A_j \cap \left(\bigcap_{k=1}^{j-1} \tilde{A}_k \right) \\
&= A_i \cap \left[\left(\bigcap_{k=1}^{i-1} \tilde{A}_k \right) \cap \tilde{A}_1 \right] \cap A_j \cap \left(\bigcap_{k=2}^{j-1} \tilde{A}_k \right) \\
&= A_i \cap \left(\bigcap_{k=1}^{i-1} \tilde{A}_k \right) \cap A_j \cap \left(\bigcap_{k=2}^{j-1} \tilde{A}_k \right) \\
&\vdots \\
&= A_j \cap \left(\bigcap_{k=1}^{i-1} \tilde{A}_k \right) \cap \left(\tilde{A}_i \cap A_i \right) \cap \left(\bigcap_{k=i+1}^{j-1} \tilde{A}_k \right).
\end{aligned}$$

But $\tilde{A}_i \cap A_i = \emptyset$ which implies $A_i^* \cap A_j^* = \emptyset$.

To see that $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} A_i^*$, let $\omega \in \bigcup_{i=1}^{\infty} A_i$. Then ω is in one or more A_i; let A_j be the first such A_i. Then $\omega \in A_j^*$ by the definition of A_j^*. We have $\bigcup_{i=1}^{\infty} A_i \subset \bigcup_{i=1}^{\infty} A_i^*$. Conversely, let $\omega \in \bigcup_{i=1}^{\infty} A_i^*$. Since $A_1^*, A_2^*, \ldots$ are disjoint, ω is in exactly one A_i^*; denote it by A_j^*. Then $\omega \in A_j$ by the definition of A_j^*. We have $\bigcup_{i=1}^{\infty} A_i \supset \bigcup_{i=1}^{\infty} A_i^*$. Because both $\bigcup_{i=1}^{\infty} A_i \subset \bigcup_{i=1}^{\infty} A_i^*$ and $\bigcup_{i=1}^{\infty} A_i \supset \bigcup_{i=1}^{\infty} A_i^*$ hold, it follows that $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} A_i^*$.

We now have $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i^*)$ and $A_i^* \subset A_i$ which proves Property 8 as remarked above.

1.6 Combinatorial Results

A sample space Ω for the game of keno, described in Section 1.1.2, consists of all sequences of length 20 made up of the numbers 1 to 80 with no number repeated within the sequence. As we have seen in Section 1.4, if N denotes the total number of such sequences, then the probability function will assign probability $1/N$ to each singleton set; that is, $P(\{\omega\}) = 1/N$. In order to do this, we need to be able to compute N.

The σ-algebra $\mathcal{F}$ for keno is the set of all possible subsets of Ω. For each event F in $\mathcal{F}$ the probability function will assign the value $P(F) = C/N$, where C is the number of points in F. For example, to determine the probability of catching three on an 8-spot ticket we need to determine the number C of sequences of length 20 made up of the numbers 1 to 80 with no number repeated within the sequence that have exactly three numbers from our specified list of eight.

These are the sorts of questions that this section addresses. Specifically, we seek the answers to four questions:

1. *Ordered samples with replacement.* How many different sequences of r numbers can be formed from the numbers $1, 2, \ldots, n$ if numbers can be repeated within the sequence and the order in which the numbers appear matters?

2. *Ordered samples without replacement.* How many different sequences of r numbers can be formed from the numbers $1, 2, \ldots, n$ if numbers cannot be repeated within the sequence and the order in which the numbers appear matters?

3. *Unordered samples without replacement.* How many different sequences of r numbers can be formed from the numbers $1, 2, \ldots, n$ if numbers cannot be repeated within the sequence and the order in which the numbers appear does not matter?

4. *Unordered samples with replacement.* How many different sequences of r numbers can be formed from the numbers $1, 2, \ldots, n$ if numbers can be repeated within the sequence and the order in which the numbers appear does not matter?

We will answer each of these questions in turn.

Question 1. Consider the sequence $(n_1, n_2, n_3, n_4, \ldots, n_r)$. There are n choices for n_1 and n choices for n_2, making $n \times n = n^2$ choices for the first two entries. Continuing thus, there are $n^2 \times n = n^3$ choices for the first three entries, $n^3 \times n = n^4$ for the first four, and so on up to n^r, which is the answer.

Question 2. Again consider $(n_1, n_2, \ldots, n_r)$. There are n choices for n_1, there are $n \Leftrightarrow 1$ choices for n_2, making $n \times (n \Leftrightarrow 1)$ choices for the first two entries. Continuing thus, there are

$$P_r^n = n \times (n \Leftrightarrow 1) \times \cdots \times (n \Leftrightarrow r + 1)$$

choices for a sequence of length r, which is the answer.

For a positive integer n, define

$$n! = n \times (n \Leftrightarrow 1) \times (n \Leftrightarrow 2) \times \cdots 3 \times 2 \times 1$$

and define $0! = 1$. Read n *factorial* for $n!$. In factorial notation

$$P_r^n = \frac{n!}{(n \Leftrightarrow r)!}\,.$$

By the logic of Question 2, $n!$ is the number of permutations $(n_1, n_2, \ldots, n_n)$ of the numbers 1 through n.

Question 3. There are at least three ways of looking at this problem:

Answer 1. Suppose we denote the answer by $\binom{n}{r}$. If we took this answer, and multiplied it by the number of permutations of r objects, then we would have the answer to Question 2. Thus $r!\binom{n}{r} = P_r^n$ or

$$\binom{n}{r} = \frac{n!}{r!(n \Leftrightarrow r)!}.$$

Read n *choose* r for $\binom{n}{r}$.

Answer 2. We know that the answer is P_r^n when order is important. What we need to do is divide out the redundant permutations of $(n_1, n_2, \ldots, n_r)$ because they are no longer regarded as important. Thus, the answer is

$$\binom{n}{r} = \frac{P_r^n}{r!} = \frac{n!}{r!(n \Leftrightarrow r!)}.$$

Answer 3. Consider the permutations of n objects where we have added some grouping

$$n! = [n \times (n \Leftrightarrow 1) \times \cdots \times (n \Leftrightarrow r + 1)] \times [(n \Leftrightarrow r) \times \cdots \times 3 \times 2 \times 1].$$

As the grouping suggests, this is the number of ways of dividing n objects into two groups, the first of size r and the second of size $n \Leftrightarrow r$, where the order of the objects within each group matters. What we want to do is disregard the permutations within each group. Therefore we must divide them out to get

$$\binom{n}{r} = \frac{[n \times (n \Leftrightarrow 1) \times \cdots \times (n \Leftrightarrow r + 1)]}{r!} \times \frac{[(n \Leftrightarrow r) \times \cdots \times 3 \times 2 \times 1]}{(n \Leftrightarrow r)!}.$$

Other interpretations of $\binom{n}{r}$ are

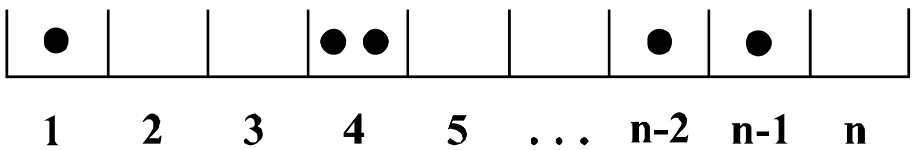

Figure. 1.6. Unordered Samples with Replacement. The number of different sequences of r numbers that can be formed from the numbers $1, 2, \ldots, n$ when numbers can be repeated within the sequence and the order in which the numbers appear does not matter is given by the number of ways that r balls can be placed in n bins. From Cassela and Berger 1990.

$\binom{n}{r}$ is the number of permutations of n objects of which r are alike and of one kind and $n \Leftrightarrow r$ are alike and of another kind; and

$\binom{n}{r}$ is the number of ways n distinct objects can be put in two boxes, r in the first box and $n \Leftrightarrow r$ in the second box.

Some extensions are

$n!/(n_1! \times n_2! \times \cdots \times n_k!)$, where $n_1 + n_2 + \cdots + n_k = n$, is the number of permutations of n objects of which n_1 are alike and of one kind, n_2 are alike and of another, and so on; and

$n!/(n_1! \times n_2! \times \cdots \times n_k!)$, where $n_1 + n_2 + \cdots + n_k = n$, is the number of ways n distinct objects can be put in k boxes, n_1 in the first box, n_2 in the second box, and so on up to the kth box.

Question 4. As seen from Figure 1.6, the number of ways a sequence $(n_1, n_2, \ldots, n_r)$ can be formed from the numbers $1, 2, \ldots, n$ when numbers can be repeated is the number of ways that r balls can be placed in n bins. This number can be obtained by dropping the two outer bin partitions and considering the number of permutations of $n \Leftrightarrow 1 + r$ objects of which $n \Leftrightarrow 1$ are alike and of one kind and r are alike and of another. This number is $\binom{n-1+r}{r}$, which is the answer.

The game of keno is covered by Question 3. Consider the probability C/N of catching three on an 8-spot ticket. There are 80 numbers of which 20 are chosen; order is not important. Thus, $N = \binom{80}{20}$. Of the 80 numbers, eight are marked on the player's ticket. The number of ways of choosing three from eight is $\binom{8}{3}$. The number of ways of choosing 17 misses from the 72 unmarked numbers is $\binom{72}{17}$. Therefore, $C/N = \binom{8}{3}\binom{72}{17}/\binom{80}{20}$. This analysis will work for any number of catches i and spots S, where $0 \le i \le S$ and $1 \le S \le 20$, so we have

$$P(\text{catch } i \text{ of } S \text{ spots}) = \frac{\binom{S}{i}\binom{80-S}{20-i}}{\binom{80}{20}}.$$

These probabilities for $S = 8$ are shown in Table 1.3.

i	$P(F_i)$
0	0.0882662377
1	0.2664641139
2	0.3281456217
3	0.2147862251
4	0.0815037015
5	0.0183025856
6	0.0023667137
7	0.0001604552
8	0.0000043457

Table 1.3. Probabilities for an Eight Spot Keno Ticket. F_i is the event "catch i spots on an 8 spot ticket".

If our logic is correct, then it must be true that

$$\sum_{i=0}^{S} \frac{\binom{S}{i}\binom{80-S}{20-i}}{\binom{80}{20}} = 1$$

because the events

$$F_i = \{\text{catch } i \text{ of } S \text{ spots}\}, \qquad i = 1, 2, \ldots, S,$$

are mutually exclusive and exhaustive. More generally, the following is true:

$$\sum_{i=\max(0,n+D-N)}^{\min(n,D)} \binom{D}{i}\binom{N \Leftrightarrow D}{n \Leftrightarrow i} = \binom{N}{n}$$

for any $n, N, D \geq 0$ such that $n \leq N$ and $D \leq N$. We shall need this fact several times in Chapter 2.

1.7 Conditional Probability

Consider Figure 1.7, which displays the sample space Ω for a roll of a pair of dice and the events A and B. Given that B has occurred, the relevant sample space becomes

$$\Omega_0 = B = \left\{ \begin{array}{cccccc} (1,4), & (1,5), & (2,4), & (2,5), & (3,4), & (3,5) \\ (4,4), & (4,5), & (5,4), & (5,5), & (6,4), & (6,5) \end{array} \right\}.$$

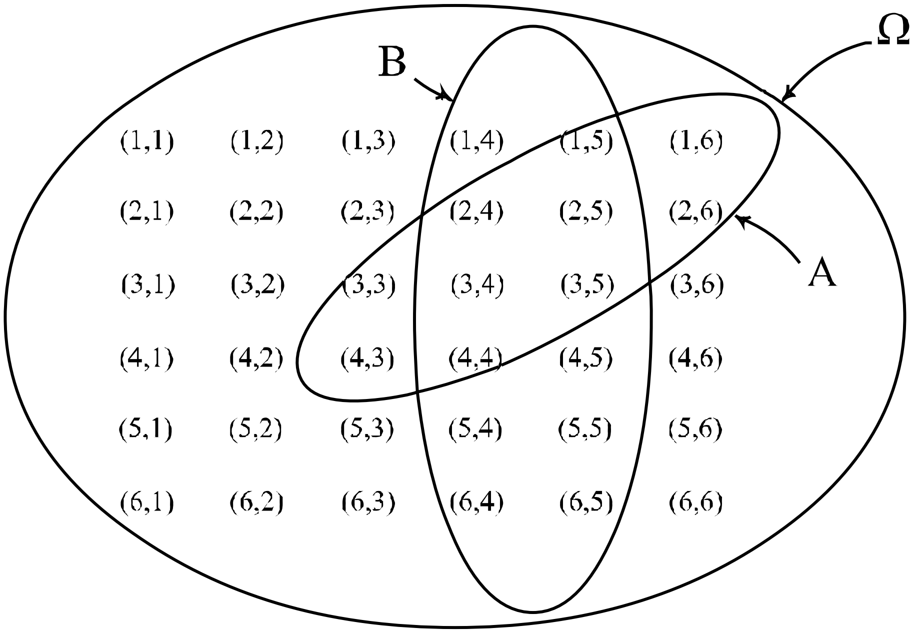

Figure 1.7. Conditional Probability. The unconditional probability of A is 10/36=0.28. The conditional probability of A is 6/12=0.50.

The other points in Ω are now irrelevant. Furthermore, the only points in A that are now relevant are

$$A \cap \Omega_0 = A \cap B = \left\{ \begin{array}{ccc} (1,5), & (2,4), & (2,5) \\ (3,4), & (3,5), & (4,4) \end{array} \right\}.$$

Because there is no information available that suggests otherwise, it seems appropriate to assume that the points in Ω_0 bear the same relative probability to one another as they did in Ω. That is, it seems appropriate that the new probability function P_0 defined on $(\Omega_0, \mathcal{F}_0)$ satisfies

$$\frac{P_0(\{\omega_i\})}{P_0(\{\omega_j\})} = \frac{P(\{\omega_i\})}{P(\{\omega_j\})},$$

where $\omega_i, \omega_j \in \Omega_0$ and P is the probability function defined on $(\Omega, \mathcal{F})$. As in Section 1.4, $\mathcal{F}_0$ consists of all possible subsets of Ω_0 and $\mathcal{F}$ all possible subsets of Ω. Because $P_0(\Omega_0) = 1$, we can recover the constant of proportionality from

$$\frac{1}{P_0(\{\omega_j\})} = \sum_{\omega_i \in \Omega_0} \frac{P_0(\{\omega_i\})}{P_0(\{\omega_j\})} = \sum_{\omega_i \in B} \frac{P(\{\omega_i\})}{P(\{\omega_j\})} = \frac{P(B)}{P(\{\omega_j\})},$$

giving

$$P_0(\{\omega_j\}) = \frac{P(\{\omega_j\})}{P(B)}$$

for $\omega_j \in \Omega_0$. From this it follows that for $A_0 \in \mathcal{F}_0$,

$$P_0(A_0) = \sum_{\omega_j \in A_0} \frac{P(\{\omega_j\})}{P(B)} = \frac{P(A \cap B)}{P(B)}.$$

This seems the obvious way to proceed, and motivates the following definition.

DEFINITION 1.2 If A and B are events in $\mathcal{F}$, then the conditional probability of A given B, denoted $P(A|B)$, is

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

If $P(B) = 0$, then define $P(A|B) = 0$ for every $A \in \mathcal{F}$.

It is nearly obvious at sight that $\big(\Omega \cap B, \mathcal{F} \cap B, P(\cdot|B)\big)$ is a probability space, where $\mathcal{F} \cap B = \{F \cap B : F \in \mathcal{F}\}$ (Problem 12). The connection with the example above is $\Omega \cap B = \Omega_0$, $\mathcal{F} \cap B = \mathcal{F}_0$, and $P(\cdot|B) = P_0(\cdot)$.

Conditional probability makes it easy to work out the probabilities of some of the multiple roll bets in craps. Consider a place bet on the 4. Conditional on the game terminating, the last roll must be a 7 or a 4. Thus, the relevant sample space is

$$\Omega_0 = \left\{ \begin{array}{lll} (1,3), & (1,6), & (2,2) \\ (2,5), & (3,1), & (3,4) \\ (4,3), & (5,2), & (6,1) \end{array} \right\}.$$

The bet wins if the event $A_0 = \{(1,3),(2,2),(3,1)\}$ occurs. Thus, the conditional probability that a place bet on the 4 wins is $P_0(A_0) = 3/9 = 1/3$.

If we are convinced that the game terminates with unconditional probability 1, then the conditional probability is the unconditional probability, which may be verified as follows. If $P(B) = 1$, then $P(\tilde{B}) = 1 \Leftrightarrow P(B) = 0$ so that $P(A \cap B) = P(A) \Leftrightarrow P(A \cap \tilde{B}) = P(A)$, because $A \cap \tilde{B} \subset \tilde{B}$ implies $0 \leq P(A \cap \tilde{B}) \leq P(\tilde{B}) = 0$. Then $P(A|B) = P(A \cap B)/P(B) = P(A)/1$.

The conditional argument above is a little slippery because, for instance, it represents all infinite sequences $\omega \in \Omega_p^\infty$ that have (1,3) before (1,6), (2,5), (3,4), (4,3), (5,2), or (6,1) by the single point (1,3) (Ω_p^∞ is defined in Section 1.2). Perhaps it could be made rigorous. It is certainly intuitively obvious and does give the correct answer with a lot less bother than we were put to in Section 1.4.

Referring to Figure 1.8, we can use the conditional argument to deduce quickly that a place bet on the 10 wins with probability 3/(3+6)=1/3; the 5 and 9 win with probability 4/(4+6)=2/5; and the 6 and 8 win with probability 5/11. These computations agree with the true odds from Table 1.1 of 3 for 1, 5 for 2, and 11 for 5, respectively.

Some useful relationships that follow from

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

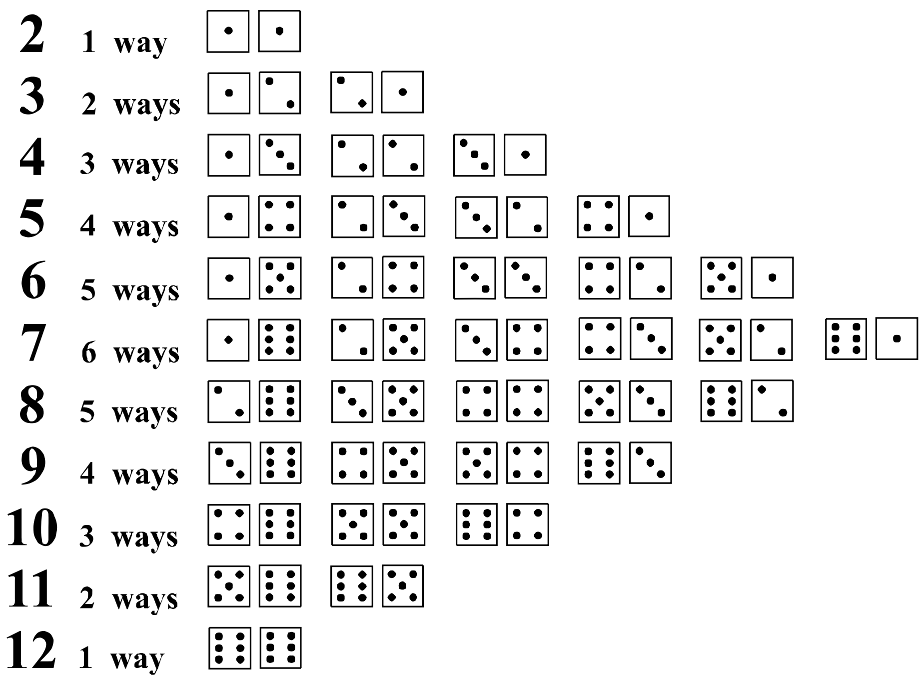

Figure 1.8. The Possible Outcomes of a Single Roll of a Pair of Dice.

are the following:

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

If $B_1, B_2, \ldots$ are mutually exclusive and exhaustive, then

$$\begin{aligned} P(A) &= \sum_{t=1}^{\infty} P(A \cap B_i) \\ &= \sum_{t=1}^{\infty} P(A|B_i)P(B_i). \end{aligned}$$

Obviously the relationship holds for a finite sequence of events $B_1, \ldots, B_n$ as well.

We can use it to work out the probability of a win on the pass line in craps. The conditioning events are $B_i =$ "i on the come out roll," for $i = 2, 3, \ldots, 12$. The conditional probability that the pass line wins given a 4 on the first roll is the same as the unconditional probability that a place bet on the 4 wins. Similarly for points 5, 6, 8, 9, and 10. Thus

$$P(\text{pass line wins}) \quad = \quad 0 \times [P(B_2) + P(B_3) + P(B_{12})] + 1 \times [P(B_7) + P(B_{11})]$$

$$
\begin{aligned}
& + \frac{1}{3}P(B_4) + \frac{2}{5}P(B_5) + \frac{5}{11}P(B_6) \\
& + \frac{1}{3}P(B_{10}) + \frac{2}{5}P(B_9) + \frac{5}{11}P(B_8) \\
= \; & \frac{6}{36} + \frac{2}{36} \\
& + \frac{1}{3} \times \frac{3}{36} + \frac{2}{5} \times \frac{4}{36} + \frac{5}{11} \times \frac{5}{36} \\
& + \frac{1}{3} \times \frac{3}{36} + \frac{2}{5} \times \frac{4}{36} + \frac{5}{11} \times \frac{5}{36} \\
= \; & 488/(2 \times 3 \times 3 \times 5 \times 11) = 244/495 \approx 0.492929.
\end{aligned}
$$

1.7.1 A Digression

The theory of probability, which is a mathematical model, has a variety of applications. When applied to a game of chance such as craps, most people, if asked, would say that the statement "$P\{(3,3),(4,4)\} = 2/36$" means is that in many tosses of a pair of dice the fraction that will land hard six (3,3) or hard eight (4,4) is approximately 2/36. They would expect the approximation to improve as the number of tosses increases.

Restated in terms of our probability model $(\Omega_p^\infty, \mathcal{F}_p^\infty, P)$ for multiple roll bets, described in Subsection 1.4.4, what this means that for each outcome

$$\omega = (\omega_1, \omega_2, \ldots, \omega_n, \ldots) \in \Omega_p^\infty$$

we expect that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I_H(\omega_i) = \frac{2}{36},$$

where $H = \{(3,3),(4,4)\}$ and $I_H(\omega_i)$ denotes the indicator function, which is the function that has the value 1 if ω_i is in the set H and has the value 0 if it is not. Similarly, if $A = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$, which is the event "any eight," we expect that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I_A(\omega_i) = \frac{5}{36}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I_{H\cap A}(\omega_i) = \frac{1}{36}.$$

If we have set up both our single roll model $(\Omega_p, \mathcal{F}_p, P)$ and multiple roll model $(\Omega_p^\infty, \mathcal{F}_p^\infty, P)$ correctly, this is how we expect them to relate to each other.

Jumping ahead, Theorem 4.1 implies that every outcome ω in Ω_p^∞ exhibits the desired behavior except for outcomes in events $E_H, E_A, E_{A\cap H} \subset \Omega_p^\infty$ that occur with probability zero. This result, coupled with the fact that there are only a finite number of events in the single roll σ-algebra $\mathcal{F}_p$ and therefore only a finite number of events that can cause trouble, would allow us, if desired, to modify the multiple roll probability space by deleting from Ω_p^∞ all outcomes in the union of these troublesome sets so that every outcome in the sample space has the requisite behavior (Problem 31).

The interpretation of our probability model just described is a bit odd because the meaning attached to the probability of an event in the single roll probability space $(\Omega_p, \mathcal{F}_p, P)$ is derived from behavior that the single roll space induces in the multiple roll space $(\Omega_p^\infty, \mathcal{F}_p^\infty, P)$. Nonetheless, this is the most common interpretation of probability when applied to games of chance and our model is consistent with this interpretation.

What most people would say that the statement "$P(H|A) = 1/5$" means in the single roll model is that in many tosses of a pair of dice the fraction that will land hard eight (4,4) of those that land eight $\{(2,6), (3,5), (4,4), (5,3), (6,2)\}$ is approximately $1/5$. They would expect the approximation to improve as the number of tosses increases.

In terms of our multiple roll model $(\Omega_p^\infty, \mathcal{F}_p^\infty, P)$, this means that for each outcome

$$\omega = (\omega_1, \omega_2, \ldots, \omega_n, \ldots) \in \Omega_p^\infty$$

we expect that

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n I_{H\cap A}(\omega_i)}{\sum_{i=1}^n I_A(\omega_i)} = \frac{1}{5}.$$

However,

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n I_{H\cap A}(\omega_i)}{\sum_{i=1}^n I_A(\omega_i)} = \frac{\lim_{n\to\infty}(1/n)\sum_{i=1}^n I_{H\cap A}(\omega_i)}{\lim_{n\to\infty}(1/n)\sum_{i=1}^n I_A(\omega_i)} = \frac{1/36}{5/36} = \frac{1}{5}.$$

Therefore, the multiple roll probability model $(\Omega_p^\infty, \mathcal{F}_p^\infty, P)$ does exhibit the desired behavior and the formula $P(H|A) = P(H \cap A)/P(A)$ does give the desired answer for the single roll model $(\Omega_p, \mathcal{F}_p, P)$.

This digression provides another justification for the formula for conditional probability introduced in Definition 1.2. The formula produces answers that are consistent with most people's interpretation of probability when applied to games of chance.

1.8 Independence

Suppose that $P(A) > 0$ and $P(B) > 0$. If $P(A|B) = P(A)$, then we learn nothing about A from observing B. Figure 1.9 is an illustration: We learn nothing about the first toss by observing that the second is a four or a five. Not only that, if $P(A|B) = P(A)$, then $P(A \cap B) = P(A)P(B)$, which implies $P(B|A) = P(B)$. Therefore, $P(A|B) = P(A)$ not only implies that we learn nothing about A from observing B but also that we learn nothing about B from knowing A. Actually, this argument has shown that if $P(A) > 0$ and $P(B) > 0$, then

$$P(A|B) = P(A) \quad \Leftrightarrow \quad P(B|A) = P(B) \quad \Leftrightarrow \quad P(A \cap B) = P(A)P(B).$$

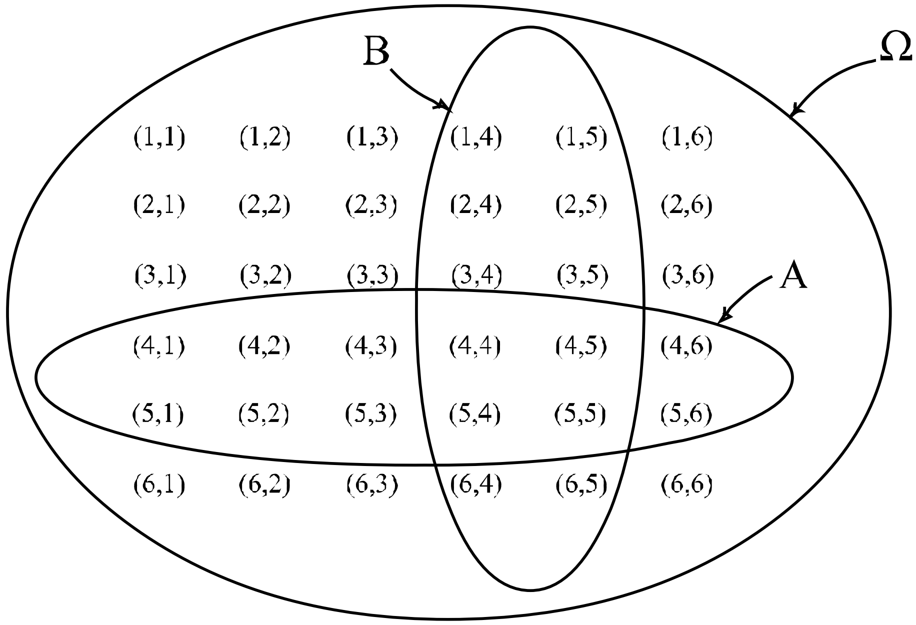

Figure 1.9. Independence. The unconditional probability of A is 12/36=0.33. The conditional probability of A is 4/12=0.33.

In this situation, the events A and B are called *independent*. We shall adopt the last of the three equivalent statements above as the definition because it also covers the case when $P(A) = 0$ or $P(B) = 0$.

DEFINITION 1.3 Two events A and B are independent if $P(A \cap B) = P(A)P(B)$.

If the events A and B are independent, then the events A and $\tilde{B}$ are independent and the events $\tilde{A}$ and $\tilde{B}$ are independent (Problem 18).

The requirements for more than two events are more stringent than mere pairwise independence.

DEFINITION 1.4 A sequence $A_1, A_2, \ldots$ of events from $\mathcal{F}$ are mutually independent if, for any subsequence $A_{i_1}, \ldots, A_{i_k}$, we have

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

1.9 Problems

1. For each of the following experiments, describe the sample space. (i) Toss a coin five times. (ii) Count fish in a pond. (iii) Measure time to failure

of a memory chip. (iv) Observe the number of defectives in a shipment. (v) Observe the proportions of defectives in a shipment.

2. Prove DeMorgan's laws for countable unions and intersections.

3. Prove that union and intersection are commutative, associative, and distributive.

4. Let F_i where $i = 1, 2, \ldots$ be an infinite sequence of events from the sample space Ω. Let F be the set of points that are in all but a finite number of the events F_i. Prove that $F = \bigcup_{k=1}^{\infty} \bigcap_{i=k}^{\infty} F_i$. Make sure that the proof is done carefully: First, take a point ω from F and show that it is in $\bigcup_{k=1}^{\infty} \bigcap_{i=k}^{\infty} F_i$. Second, take a point ω from $\bigcup_{k=1}^{\infty} \bigcap_{i=k}^{\infty} F_i$ and show that it is in F.

5. Find the supremum and infimum of the following sets: $\emptyset$, $(-10, 10)$, $(-\infty, \infty)$, $\bigcap_{i=1}^{\infty}\{1/i\}$, $\bigcup_{i=1}^{\infty}\{1/i\}$, $\bigcap_{i=1}^{\infty}[1/i, 1]$, $\bigcup_{i=1}^{\infty}[1/i, 1]$, $\{x : x = 1/i,\ i = 1, 2, \ldots\}$, $\{x : x = -1/i,\ i = 1, 2, \ldots\}$, $\{x : x = i,\ i = 1, 2, \ldots\}$, and $\{x : x = -i,\ i = 1, 2, \ldots\}$.

6. Show that $\mathcal{F} = \{\emptyset, \Omega\}$ is a σ-algebra.

7. Show the collection of all subsets of Ω is a σ-algebra.

8. Show that if $F_1, F_2, \ldots, F_N$ are mutually exclusive and exhaustive, then the collection of all finite unions plus the empty set is an algebra.

9. Show that if $F_1, F_2, \ldots$ are mutually exclusive and exhaustive, then the collection of all countable unions plus the empty set is a σ-algebra.

10. Show that the intersection of two σ-algebras is a σ-algebra.

11. Let $\mathcal{A}$ be some collection of subsets of Ω. Problem 7 implies that there exists at least one σ-algebra that contains $\mathcal{A}$ (Why?). Let $\mathcal{F}$ be the intersection of all σ-algebras that contain $\mathcal{A}$. Show that $\mathcal{F}$ is a σ-algebra. Show that $\mathcal{F}$ is not empty. Why is $\mathcal{F}$ the smallest σ-algebra that contains $\mathcal{A}$?

12. Show that if $\mathcal{F}$ is a σ-algebra, then $\mathcal{F} \cap B = \{F \cap B : F \in \mathcal{F}\}$ is a σ-algebra.

13. Show that $P(\emptyset) = 0$, $P(A) \leq 1$, $P(A) + P(\tilde{A}) = 1$, and that if $A \subset B$, then $P(A) \leq P(B)$.

14. If $P(A) = \frac{1}{3}$ and $P(\tilde{B}) = \frac{1}{4}$, can A and B be disjoint?

15. Find formulas for the probabilities of the following events: (i) either A or B or both, (ii) either A or B but not both.

16. A pair of dice are thrown and the sum is noted. The throws are repeated until either a sum of 6 or a sum of 7 occurs. What is the sample space for this experiment? What is the probability that the sequence of throws terminates in a 7? Be sure to include an explanation of the logic that you used to reach your answer.

17. In a shipment of 1,000 transistors, 100 are defective. If 50 transistors are inspected, what is the probability that five of them will be defective. Be sure to include an explanation of the logic that you used to reach your answer.

18. Show that if two events A and B are independent, then so are A and $\tilde{B}$ and $\tilde{A}$ and $\tilde{B}$.

19. Assume that $P(A) > 0$ and $P(B) > 0$. Prove that if $P(B) = 1$, then $P(A|B) = P(A)$ for any A. Prove that if $A \subset B$, then $P(B|A) = 1$.

20. Assume that $P(A) > 0$ and $P(B) > 0$. Prove that if A and B are mutually exclusive, then they cannot be independent. Prove that if A and B are independent, then they cannot be mutually exclusive.

21. Prove that if $P(\cdot)$ is a legitimate probability function and B is a set with $P(B) > 0$, then $P(\cdot|B)$ also satisfies the axioms of probability.

22. Compute the probability of a win for each of the one roll bets in craps.

23. Compute the probability of a win for each of the place bets in craps. Work the problem two ways: (i) Compute the probability of the union of the events "win on roll i." (ii) Compute the probability of a win conditional on termination.

24. How many different sets of initials can be formed if every person has one surname and (i) exactly two given names; (ii) either one or two given names; (iii) either one, two, or three given names?

25. If n balls are placed at random into n cells, what is the probability that exactly one cell remains empty?

26. If a multivariate function has continuous partial derivatives, the order in which the derivatives are calculated does not matter (Green's theorem). For example, $(\partial^3/\partial x^2 \partial y)f(x, y) = (\partial^3/\partial y \partial x^2)f(x, y)$. (i) How many third partial derivatives does a function of two variables have. (ii) Show that a function of n variables has $\binom{n+r-1}{r}$ rth partial derivatives.

27. Suppose that an urn contains n balls all of which are white except one which is red. The urn is thoroughly mixed and all the balls are drawn from the urn without replacement by a blindfolded individual. Show that the probability that the red ball will be drawn on the kth draw is $1/n$.

28. Two people each toss a coin n times that lands heads with probability $1/3$. What is the probability that they will each have the same number of heads? What is the probability if the coin lands heads with probability $1/4$?

29. For the game of craps, compute the probability that a shooter coming out will roll the first 7 on the kth roll.

30. For the two dimensional coin tossing experiment described in Subsection 1.4.2, let $A = (0, 1/2] \times (0, 1]$, $B = (0, 1] \times (0, 1/2]$, and $C = \{(x, y) : x < y\}$. Show that $P(C) = 1/2$ and that $P(A \cap C) = 1/8$. Are A and B independent? Are A and C independent? Compute $P(A)$ and $P(A|C)$.

31. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Show that if $P(E) = 0$, then $P(F \cap \tilde{E}) = P(F)$ for every F in $\mathcal{F}$. Why are the two probability spaces $(\Omega, \mathcal{F}, P)$ and $(\Omega \cap \tilde{E}, \mathcal{F} \cap \tilde{E}, P)$ equivalent? See Problem 12 for the definition of $\mathcal{F} \cap \tilde{E}$.

References

Abramowitz, Milton, and Irene A. Stegun. 1964. *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables.* Washington, DC: U.S. Government Printing Office.

Bansal, Ravi, A. Ronald Gallant, Robert Hussey, and George Tauchen. 1993. "Computational Aspects of Nonparametric Simulation Estimation." In *Computational Techniques for Econometrics and Economic Analysis* ed. David A. Belsley, 3-22. Boston: Kluwer Academic Publishers.

Bartle, Robert G. 1976. *The Elements of Real Analysis, Second Edition.* New York: Wiley.

Billingsly, Patrick. 1995. *Probability and Measure, Third Edition.* New York: Wiley.

Casella, George, and Roger L. Berger. 1987. "Reconciling Bayesian and Frequentist Evidence in the One-Sided Testing Problem." *Journal of the American Statistical Association* 82, 106–111.

Casella, George, and Roger L. Berger. 1990. *Statistical Inference.* Pacific Grove, CA: Wadsworth & Brooks-Cole.

Chung, Kai Lai. 1974. *A Course in Probability Theory, Second Edition.* New York: Academic Press.

Cox, David R., and David V. Hinkley. 1974. *Theoretical Statistics.* London: Chapman and Hall.

Dunes Hotel. 1984. "Gaming Guide". Brochure. Las Vegas, NV: Dunes Hotel and Country Club.

Gallant, A. Ronald. 1980. "Explicit Estimators of Parametric Functions in Nonlinear Regression." *Journal of the American Statistical Association* 75, 182–193.

Gallant, A. Ronald. 1987. *Nonlinear Statistical Models.* New York: Wiley.

Gallant, A. Ronald, and Jonathan R. Long. 1997. "Estimating Stochastic Differential Equations Efficiently by Minimum Chi Square." *Biometrika*, forthcoming.

Gallant, A. Ronald, and Douglas W. Nychka. 1987. "Seminonparametric Maximum Likelihood Estimation." *Econometrica* 55, 363–390.

Gallant, A. Ronald, and George Tauchen. 1996. "Which Moments to Match?" *Econometric Theory* 12, 657–681.

Gourieroux, C., A. Monfort, and E. Renault. 1993. "Indirect Inference." *Journal of Applied Econometrics* 8, S85–S118.

Hansen, Lars Peter. 1982. "Large Sample Properties of Generalized Methods of Moments Estimators." *Econometrica* 50, 1029–1054.

Hansen, Lars Peter, and Kenneth J. Singleton. 1982. Generalized Instrumental Variables Estimators of Nonlinear Rational Expectations Models. *Econometrica* 50, 1269–1286.

Johnson, Norman L., and Samuel Kotz. 1969. *Distributions in Statistics: Discrete Distributions*. New York: Wiley.

Johnson, Norman L., and Samuel Kotz. 1970a. *Distributions in Statistics: Continuous Univariate Distributions-1*. New York: Wiley.

Johnson, Norman L., and Samuel Kotz. 1970b. *Distributions in Statistics: Continuous Univariate Distributions-2*. New York: Wiley.

Johnson, Norman L., and Samuel Kotz. 1972. *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley.

Lehmann, Erich L. 1983. *Theory of Point Estimation*. New York: Wiley.

Lehmann, Erich L. 1986. *Testing Statistical Hypotheses, Second Edition*. New York: Wiley.

MGM Grand Hotel. 1984. "Keno". Brochure. Las Vegas, NV: MGM Grand Hotel.

Newey, Whitney K. 1990). "Efficient Instrumental Variables Estimation of Nonlinear Models." *Econometrica* 58, 809–837.

Patterson, Jerry L., and Walter Jaye. 1982. *Casino Gambling*. New York: Putnam Publishing Group.

Pollard, David. 1984. *Convergence of Stochastic Processes*, New York: Springer-Verlag.

Pratt, John W. 1961. "Length of Confidence Intervals." *Journal of the American Statistical Association* 56, 549–567.

Royden, H. L. 1988. *Real Analysis, Third Edition.* New York: Macmillan.

Schuster, Heinz Georg. 1988. *Deterministic Chaos.* New York: VCH Publishers.

Theil, Henri. 1961. *Economic Forecasts and Policy, Second Edition.* Amsterdam: North Holland.

Tierney, Luke. 1994. "Markov Chains for Exploring Posterior Distributions." *The Annals of Statistics* 22, 1701–1727.

Wald, Abraham. 1949. "Note on the Consistency of the Maximum Likelihood Estimate." *Annals of Mathematical Statistics* 20, 595–601.

Zellner, Arnold, and Henri Theil. 1962. "Three-stage Least Squares: Simultaneous Estimation of Simultaneous Equations." *Econometrica* 30, 54–78.

Index