

N. Richard Werthamer

Risk and Reward

The Science of Casino Blackjack

$$\left\langle R_f(d) \right\rangle_\gamma = \exp \left[\sum_\lambda \left(\hat{\gamma}_\lambda \hat{\alpha}_\lambda \cdot \hat{\Delta} + \frac{1}{2 \hat{\Delta}^2} (\hat{\alpha}_\lambda \cdot \hat{\Delta})^2 \right) \right] R_0(d_0).$$



Springer

Risk and Reward

N. Richard Werthamer

Risk and Reward

The Science of Casino Blackjack



Springer

Dr. N. Richard Werthamer
43 West 16th Street
New York NY 10011
USA
rwerthamer@nyc.rr.com

ISBN 978-1-4419-0252-8 e-ISBN 978-1-4419-0253-5
DOI 10.1007/978-1-4419-0253-5

Library of Congress Control Number: 2009926050

Mathematics Subject Classification (2000): (Primary) 91A60 (Secondary) 60J60, 60-02

© 2009 Springer Science+Business Media, LLC

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

How did a physicist like me get involved with blackjack? I skied in. As a young doctoral candidate at the University of California, Berkeley, I would occasionally join my friends on ski trips to the Lake Tahoe area. Always on a student's tight budget, we found that we could stay inexpensively in Reno motels and that the cheapest meals were at the coffee shops astutely placed at the back of the town's casinos. I found it hard to pass all the gambling frenzy without joining it. Fortunately for that budget and my financial peace of mind, I was usually lucky in winning a few dollars. Once, I paid for my entire trip with an unlikely hit at roulette.

When I learned that an optimized method for playing blackjack (what is now called Basic Strategy) had been published in a scholarly statistics journal, I quickly looked it up in the university library. Blackjack became my game of choice for subsequent Tahoe excursions and, in later years after skiing and I had parted company, for trips to other destinations with casinos. But my real recreation of choice became blackjack analysis!

My original motive for exploring its mathematics was to see if the articles and books I had chanced upon were correct, or if I could find a better way to win. Visions of big money danced in my head! Later, as I came to appreciate the realities involved, my interest shifted to the mathematics for its own sake, complex enough to be a challenge even for a professional scientist yet easy enough to yield to a sustained effort on almost every issue. Also, I wanted to verify (or, as sometimes happened, falsify) claims about the best way to play, asserted in the literature with little or no proof.

When I had finally addressed all the questions that occurred to me, my first instinct as a scientist was to write up the results and publish them in a suitable scholarly journal: an archival record for other practitioners of applied math. Yet I came to realize that the results were sufficiently insightful about playing actual blackjack in the real world that they would be even more valuable to the wider casino-going public. To reconcile my internal debate between these two readerships, I decided to serve both at once. The outcome is this single two-part book, which first describes all its conclusions in completely non-mathematical terms for the many that enjoy gambling, and then also provides detailed derivations of those conclusions so that

they can be followed and checked by the more mathematically inclined. Some of the derivations rely only on algebra; others take varying levels of calculus and/or probability theory.

When talking to friends about my blackjack project, they invariably ask whether I have “tested my theories in a casino.” I respond that the conclusions drawn from a mathematical proof are intrinsically true, and do not need field testing to be valid. I suggest that they should rather be asking whether I have “applied” the theories; and I add that analyzing blackjack is at least as entertaining for me as playing it. I also point out that the cash outcome of playing several hundred hands at an actual blackjack table, which would occupy a full evening, is little more than an unpredictable statistical fluctuation rather than a test of the methods. Many millions of hands, simulated on a computer, are needed to approach significant conclusions. I am happy to have examined the game in comprehensive detail, without having risked a single dime - even though I have not won one, either!

Several leading authorities on blackjack have been generous in sharing their time and knowledge with me. Stanford Wong and Anthony Curtis were encouraging, and introduced me to several other prominent figures. One of them, Don Schlesinger, intensively reviewed more than one earlier draft of my manuscript, thereby prompting major revisions and expansions. More recently, Stewart Ethier graciously did the same, with similar results. It is a pleasure to acknowledge their help and support. Additionally, I have benefited from communications with Sergei Maslov, Kim Lee, Steve Jacobs, Michael Canjar, and Norm Wattenberger.

I am also indebted to staff members of Wolfram Research for their active help and guidance with *Mathematica*, the software tool for much of the numerical computation and all of the figure production in this book. I am similarly grateful for the dedicated efforts of editorial and production personnel at Springer.

New York, NY

N. Richard Werthamer

Introduction

Considering the large number of blackjack books published since that original statistics paper, more than 50 years ago, it may well be wondered what I could possibly add. Basic Strategy, after all, has been known and unchallenged for decades. Card counting methods, as a complement to Basic Strategy, have also been fully developed. But most of the previous publications have shortcomings and missing pieces, which I fill in and clarify. The result is a systematic description of each major aspect of optimal play: how best to play the cards in your hand, how best to assess the odds expected for the next round, and how best to adjust the amount you bet on that round based on its odds.

Furthermore, I emphasize the trade-off between simplicity and performance: I discuss how easy each aspect of optimal play is to actually use in a casino, and I recommend simplified versions, where desirable, that give nearly optimal performance. A serious player should ideally be carrying out a number of distinct mental processes simultaneously, rapidly, and accurately; he may well wish to drop the least valuable ones, to ensure his fidelity to those more worthwhile, as well as to lower his stress level.

I contribute much new material as well. The most significant is to that third major aspect – how you, as a player, should manage your cash. (I will usually refer to you in the third person, as Player, with masculine gender assumed; and the dealer, correspondingly, as Dealer, feminine gender.) If you are advised to increase your bet when you have detected favorable odds, then it is likely you will immediately ask, “By how much?” Closely related is how you can control the risk of a losing streak that wipes out the total capital you brought to the casino. This risk depends critically on the scale of your betting relative to your capital: most players probably lose more money from inadequate capitalization than from an imperfect playing strategy.

Almost all previous authors on blackjack either avoid the capital management issue altogether or give vague, hand-waving impressions. Almost no one supplies an answer that seems convincing, let alone thoroughly justified; and the better work is largely in restricted Web chat rooms. In contrast, I demonstrate the optimal relationship between the amount bet on a hand and its indicated odds, and I link it to your risk tolerance.

I also fully analyze several under-appreciated refinements to betting strategy. Two of these, referred to in the literature as back-counting and Kelly betting, are usually mentioned only in passing. A third is betting on multiple simultaneous hands. I show that these can each significantly enhance performance. Their use, in fact, is a key to the success of blackjack teams.

Part I of *Risk and Reward* gives a straightforward, self-contained guide to blackjack for a general reader, even a beginning or occasional player. It recommends how best to play the game, and what results to expect.

Chapter 1 reviews the rules of blackjack, along with definitions of the key terms specific to the game. Chapter 2 then covers the simplest and most-used technique for playing a hand, widely known as Basic Strategy, in which Player bases his actions on just the cards in his own hand and the one card visible in Dealer's hand. Limited to only this information, Player should always bet the same amount on each round.

Chapter 3 goes on to describe the more sophisticated group of strategies known as card counting. In these, Player tracks the cards dealt on the previous rounds in order to estimate the odds on the next round and decide how much to bet. Of a number of variant counting schemes proposed over the years, several of the more accurate and currently popular are compared and evaluated.

Chapter 4 is in many ways the most significant (and original) in the book. Section 4.1 relates the scale of Player's bet size, the rate at which he raises and lowers his bet dependent on the count, and his tolerance for the risk of losing his entire capital. The way in which these factors and others influence whether or not Player has an edge over the casino is carefully explained. Section 4.2 examines a modified betting scheme that accelerates the rate of winning. It also points out the advantages of playing multiple hands simultaneously. Section 4.3 shows how Player further benefits from properly choosing the moments to begin betting at a table and to leave it.

Chapter 5 revisits play of the hand, to examine ways in which it interacts with card counting. Section 5.1 considers modifications to the counting technique that best support a play strategy generalized to become count-dependent. Section 5.2, in contrast, identifies the best count-independent play strategy for a Player who systematically varies his bets.

Chapter 6 collects the key points of Optimal Strategy from earlier chapters. It summarizes the techniques of play strategy, of card counting, and of bet sizing; and it reviews the complex of factors, including casino countermeasures, determining the amount of money Player might win or lose. It presents the case that casino blackjack for a single Player is better thought of as an entertainment, which can be enjoyed with only moderate skill at relatively tolerable expense, rather than as a consistent moneymaker. On the other hand, a blackjack team can use techniques not available to an individual that make it a true business, with attractive rates of return. Chapter 6 should be scanned by every reader, whether or not he has dipped into the other chapters.

Part II then details the analysis justifying those recommendations, intended for the subset of readers who wish to understand the game more deeply, and/or follow its mathematics. Part II gives detailed, self-contained derivations of the assertions

and qualitative descriptions of Part I, with Chaps. 7–10 corresponding, respectively, to Chaps. 2–5.

The tools used in Chap. 7 are almost entirely algebra and basic probability, so more advanced mathematics should not be needed to follow it. Chapters 8–10, however, bring in the machinery of vector calculus; college training equivalent to a science or engineering degree should be sufficient to understand some of the derivations, while the more sophisticated steps of the others are detailed in step-by-step appendices. For those readers who desire an added boost, I recommend Morse & Feshbach (1953), my own guidebook to applied mathematics and even today a leading authority. For a recent text focused on matrix algebra I suggest Shores (2007).

Some further comments are in order on the style of the mathematics. First of all, I have avoided the use of theorems, formally stated and rigorously proved as in typical mathematics publications. Rather, I take the approach of most physicists, which is to narrate a derivation assuming, without explicit formal proof, that each manipulation is valid unless there is some evident reason to question it.

Furthermore, at many points in Chaps. 8–10 where a quantity is strictly speaking a rational number (ratio of integers), I approach its analysis by approximating it by a decimal one, justified because the number of cards shuffled into the pack is large. (Illustration: the fraction $17/52$ is a rational, while 0.327 is a close approximation to it.) The move from the exact discrete math of earlier blackjack analysts, primarily based on probability and statistics, to approximate continuum math opens the door to the powerful tools of calculus, and facilitates new results difficult to obtain in the traditional way.

Supplementing the text are indices of terms and of notation. Each entry cites the page where the term or symbol is first introduced.

Contents

Preface	v
Introduction	vii
 Part I Strategy	
1 The Game	3
1.1 History of Casino Blackjack and Its Analysis	3
1.2 Rules, Procedures, and Terminology	5
2 Playing the Hand	9
2.1 Basic Strategy	9
2.1.1 Expected Return with Variant Rules and Procedures	13
2.1.2 Expected Return vs. Return on Investment	13
2.2 Composition-Dependent Play	14
3 Tracking the Cards	15
3.1 Linear Counts	16
3.2 Choosing a Counting Vector	17
3.3 Unbalanced Counting Vectors	19
3.4 Relating the True Count to the Expected Return	20
4 Betting	21
4.1 Yield, Risk, and Optimal Bet Strategies	21
4.1.1 Camouflage	26
4.2 Betting Proportional to Current Capital	27
4.2.1 Multiple Hands with Rescaling	28
4.3 Back-Counting and Table-Hopping	29
5 Playing the Hand When the Count and Bet Vary	31
5.1 Play Strategies that Vary with the Count	31
5.1.1 Reconsidering the Counting Vector	31
5.1.2 Count Dependence of the Play Parameters	33

5.1.3	The Insurance Bet	34
5.2	Counter Basic Strategy for the Variable Bettor.....	35
6	Synthesis and Observations	37
6.1	A Practical, Nearly Optimal Strategy	37
6.2	Blackjack as a Recreation vs. a Profession	43
Part II Analysis		
7	Play Strategies	47
7.1	Basic Strategy, Large Number of Decks.....	47
7.1.1	Analytical Framework	47
7.1.2	Expected Player Return	49
7.1.3	Frequency of Tied Hands	52
7.1.4	Multiple Simultaneous Hands: Return, Variance, and Covariance	53
7.1.5	Expected Number of Cards Used per Round	56
7.2	Basic Strategy, Small Number of Decks.....	58
7.2.1	Analytical Framework	58
7.2.2	Expected Return, and Optimal Basic Strategy, vs. Number of Decks.....	61
7.2.3	Surrender; Insurance	62
7.3	Play Parameters Dependent on Identities of Initial Cards	63
7.3.1	Comparison with Previous Authorities.....	65
8	Card Counting	67
8.1	Analytical Framework	67
8.1.1	Asymptotic Distribution	67
8.1.2	Expected Return; Invariance Theorem	70
8.1.3	Hermite Series	72
8.2	Expected Return at Nonzero Depth	73
8.3	Optimizing the Counting Vectors	75
8.4	Optimizing the Counting Vectors: Many-Cards Limit	76
8.5	Computation of the Derivatives of the Expected Return.....	78
8.6	Unbalanced Counts	78
Appendix 1 Asymptotic Distribution of Card Likelihoods		79
Appendix 2 Eigenmodes		80
9	Bet Strategies.....	81
9.1	Risk and Capitalization.....	81
9.1.1	Risk in a Game with Fixed Return	81
9.1.2	Optimal Betting When Return Fluctuates: Expected Capital and Risk	85
9.1.3	Connections with Finance	89
9.1.4	Properties of the Risk and Expected Capital Expressions	90

9.1.5	Optimal Betting When Return Fluctuates: Bet Strategy.....	93
9.1.6	Yield When the Bet Size Is Discrete; Wong Benchmark Betting.....	98
9.2	Betting Proportional to Current Capital	100
9.2.1	One Hand per Round.....	100
9.2.2	Multiple Simultaneous Hands	103
9.3	Back-Counting and Table-Hopping.....	105
9.3.1	Entry	105
9.3.2	Entry and Exit	107
9.3.3	Entry and Departure.....	110
9.3.4	Entry, Exit, and Departure	112
Appendix 1	The Risk of Ruin Formula	113
Appendix 2	Chain Rule Convolution	114
10	Play Strategies with Card Counting.....	115
10.1	Count-Dependent Playing Strategy	115
10.1.1	Counting Vector Optimal for Play Variation Alone	117
10.1.2	Single Counting Vector Optimal for Bet and Play Together ..	119
10.1.3	Two Distinct Counting Vectors	121
10.1.4	Insurance with Variable Betting	122
10.2	Counter Basic Strategy	124
	References.....	127
	Index.....	129

List of Figures

2.1	Expected return, reshuffle round, vs. inverse number of decks (those in Table 2.2, plus 3/2 and 5/4)	11
4.1	Schematic representation of optimal betting strategy: ramp (<i>thick line</i> , the <i>dashed diagonal</i> constrained by lower and upper bet sizes) and a two-step staircase approximation	23
4.2	Risk of ruin (<i>curve with a minimum</i>) and effective yield (<i>curve with a maximum</i>) vs. ramp steepness	24
4.3	Effective yield vs. risk of ruin, with points MM (<i>solid squares</i>) and HJY (<i>open circles</i>): <i>upper curve</i> for 1,000,000 rounds; <i>lower curve</i> for 1,000 rounds	24
6.1	Distribution of outcomes: <i>upper plot (a)</i> after 400 rounds, <i>lower plot (b)</i> after 10,000 rounds. For each plot, the left-most curve represents the noncounter, the right-most the counter	41
8.1	Expected return vs. true count, reshuffle round, six decks: <i>upper pair of curves</i> for count-dependent play, the <i>topmost</i> with DAS, the <i>middle</i> without; the <i>lowest curve</i> is for count-independent Optimal Basic play	73
8.2	Expected return vs. true count as in Fig. 8.1, but superimposing onto the six-deck results those for single deck as well, the latter set with a slight upward shift vs. the former	74
8.3	Count-dependent (<i>kinky curve</i>) and Optimal Basic (<i>smooth curve</i>) first differentials at small values of true count; six decks.....	75
9.1	Distribution of return, scaled to its magnitude at zero count, for various depths: <i>upper plot</i> for six decks, <i>lower plot</i> for one deck. At small depths the distributions are narrowly peaked around -1 , while as the depth grows they spread further out and have increasing weight at positive returns. The spreading is more pronounced for one deck than for six	86

9.2	Survival probability (a) and yield reduction factor (b) as functions of the parameters q and N , log–log scale	91
9.3	Representation of the distribution of capital, for several values of the number of rounds N . As N increases, the distribution shifts, widens, and flattens: its mean and its variance both increase	92
9.4	Distributions of expected returns: at depth 0.4 (<i>Gaussian curve</i>) and averaged over depths to a penetration 0.8 (<i>peaked curve</i>), for representative game parameters	95
9.5	Contours of minimum risk (<i>solid curves</i>) and maximum risk (<i>dashed curves</i>) vs. b_{\pm}	96
9.6	Contours of effective yield vs. b_{\pm} , bounded between curves of minimum risk (on the <i>right</i>) and maximum risk (on the <i>left</i>). The curve of HJY points is <i>dashed</i>	97
9.7	Wong’s “Benchmark” bet staircase (<i>solid</i>) and ramp modification (<i>dashed</i>)	99
9.8	Distributions of true counts, subsequent to entry at true counts 0, 1, 2	108
9.9	Distributions of true counts: with neither entry nor exit, with entry at +1 but without exit, and with entry at +1 and exit at -1	109
10.1	The function $\Psi(F)$	117
10.2	Graphical solution to eigenvalue equation	120

List of Tables

2.1	Generic Strategy (Optimal Basic Strategy for three to six decks)	10
2.2	Optimal Basic Strategy expected return vs. number of decks.....	11
2.3	Differences in Optimal Basic from Generic, one and two decks.....	12
2.4	Generic Strategy applied to other numbers of decks	12
2.5	Increase in expected return from Composition-Dependent play vs. Optimal Basic Strategy (resplits, no DAS)	14
3.1	Most-recommended counting vectors	17
3.2	Other counting vectors	18
3.3	Unbalanced counting vectors	19
3.4	Relationship of true count to expected return	20
4.1	Operating parameters for the MM and HJY betting points, Lifetimer and Weekender examples	25
4.2	Operating parameters vs. penetration, Lifetimer and Weekender examples	26
4.3	Operating parameters vs. spread, Lifetimer and Weekender examples	27
4.4	Factor of bet proportion and of median capital growth rate, vs. number of seats	28
4.5	Optimal true count (tc) thresholds and yield ratio for table-hopping	30
5.1	Counting vector optimal for play variation	32
5.2	Ranges of true count, between -12 and $+20$, within which standing on a particular hard hand value, between 17 and 12 , is optimal, depending on Dealer upcard; six decks	33
5.3	Separate counting vector for insurance	34
5.4	Yield ratios for Counter Basic, Optimal Basic and Count-dependent play strategies.....	35
5.5	Counter Basic Strategy, single deck	36

6.1	Hi-Lo counting vector	38
6.2	Differences in one-deck Optimal Strategy from Generic	40
7.1	Ties, variance, and covariance vs. number of decks	61
7.2	Composition-dependent play strategy	64
8.1	Parameters in polynomial fits to expected return vs. true count, count-dependent (C-D) and Optimal Basic (OB) play strategies	74
8.2	Gradient of return and optimal counting vector	78
9.1	Performance of a coarse staircase vs. a ramp	99
9.2	Performance of Wong bet function	100
10.1	Optimal counting vector for play variation	118

Part I

Strategy

Chapter 1

The Game

1.1 History of Casino Blackjack and Its Analysis

Casino gambling has been booming steadily for decades throughout the world. In the USA, from a time before the 1930s when casinos were illegal everywhere, they have grown in number and in popularity. Legalized initially in Nevada, they took off after World War II, particularly in Las Vegas and Reno, then around Lake Tahoe. Regulated casinos later mushroomed in Puerto Rico and, beginning in the 1970s, in Atlantic City, New Jersey. Casinos then sprang up on river boats in parts of the Midwest and South. More recently, American Indian tribes have been using their distinctive legal status to develop casinos on their reservations, especially in Connecticut (the largest anywhere in the world) but also in Illinois, California, Washington, and elsewhere. Older casinos in Las Vegas have even been bulldozed and replaced by much larger and grander ones. In almost every instance, each new casino has stimulated another increment in the public's interest. Total casino gaming revenue has risen year after year, into the tens of billions of dollars in the USA alone, with little evidence of saturation.

By far the most popular game in practically every casino is the slot machine. Slots require no skill – in fact almost no thought – and treat the occasional winner to a celebration of sound and light that cannot escape the attention (and envy) of nearby patrons. Engineered with odds that heavily favor the casino (usually 8% or more on each spin of the wheels, several spins per minute), they are rapid and inexorable money-winners for their operators.

The table games have also expanded in popularity, proportional to the growth of casinos. These games, chiefly craps and blackjack but also roulette and baccarat, vary both in the range of options they afford Player – and hence in the skills required to best take advantage of those options – and in their odds. Roulette is clearly the most straightforward: no strategy decisions are called for, and the odds on each spin of the wheel are uniformly 5.26% against Player (on US tables with both zero and double zero; 2.70% on European tables with just the zero). Craps offers a number of different bets from which Player can choose, each with its own odds, but all favor the casino and none offer strategy decisions. Even the best choice, on the Pass Line, favors the casino by about 1.4%. The casino edge in baccarat is only slightly less unfavorable, at about 1.06% when betting on Bank.

On the contrary, blackjack requires Player decisions that have a great influence on the expected return. Apart from the added entertainment value of a game of strategy vs. a game of mere chance, blackjack with optimal play has the potential not only to reduce the casino advantage virtually to zero but also in fact to become favorable to an astute Player. In blackjack (and possibly also poker, alone among casino games), skill creates the possibility, at least in the long run, to systematically win money. Nevertheless, achieving that edge is difficult: not only does Player have to apply concentrated intellectual effort to approach optimal performance, but also the casino has several legitimate tools available to counter him.

The published literature about blackjack is considerable. Prior to the mid-1950s the game had not been carefully analyzed. The popular books about casino gambling (as one example, Scarne (1961)) made assertions that were not much more than guesswork; some in fact were quite wrong. A major breakthrough was the work of Baldwin et al. (1956) published in a scholarly statistics journal, which for the first time presented a mathematical approach for analyzing blackjack and an optimized playing technique which (with later extensions and correction of their minor errors) has become known as Basic Strategy. Baldwin et al. were also the first to apply computational methods in their analysis, although their “computers” were of the mechanical adding-machine kind. However, Baldwin et al. were explicit, and correct, that Basic Strategy still has a negative expected return: the odds favor the casino for a game with multiple decks.

A few years later, Edward Thorp (1966) took this approach further, by realizing that the expected return for a given hand varied with the composition of the as-yet-undealt cards, and that a careful scan of the cards as they were dealt contained useful clues to that composition. Thorp proposed a specific scanning scheme, or count, which he claimed could identify instances when the odds were actually in favor of Player. Thorp concluded that Player, by increasing his bet when these instances were detected, could achieve a positive overall edge.

Thorp’s work, published originally in 1962 as a book titled “Beat the Dealer,” created a sensation in both the gambling public and the casino industry. Gamblers rushed to apply his methods, while casinos quickly made rule and procedure changes to counter what they felt was a serious threat to their viability. Some players, based on casino suspicions that they were “counting,” were politely but firmly asked to leave. Soon, however, the casinos’ panic subsided and the changes were reversed.

Thorp had also written a program for electronic computer that improved on the work of Baldwin et al. Later, Julian Braun, an IBM staffer, wrote a more extensive and precise routine. The use of electronic computers, even though primitive then relative to today’s, gave results within hours rather than months. Over the next decade, Braun refined his methods, supplied results to other blackjack authors, and eventually published his own book. Braun and others developed a different form of counting scheme than Thorp’s, one that was both simpler and more informative. Several variants of this form of counting, particularly one ascribed to Harvey Dubner, have been in wide use ever since. An extensive history of blackjack and developments in its analysis is given by Snyder (2006).

In the several decades since the work of Thorp, Braun, and others, dozens of books about blackjack strategy have appeared, mostly repackaging, extending, and popularizing the conclusions of the pioneers. The more quantitative of the subsequent authors have included Peter Griffin and Stanford Wong. Griffin (1999) went beyond Thorp and Braun to analyze many other aspects of blackjack strategy. On the contrary, Wong (1994) undertook an exhaustive study of one favored card-counting method, using a large-scale computer simulation. Wong programmed a computer to “play” millions of hands of blackjack for each of many possible play strategies within that counting method. By comparing outcomes, he identified the most favorable of those strategies and the overall average return to Player.

More recently, several Internet “chat-rooms” on blackjack have been set up. Those operating like a club, accessed by a password requiring a fee or member recommendation, became especially popular with blackjack experts. Some sophisticated analyses have been contributed to these Web sites and discussed by their participants, although these have not yet been assembled into the sort of coherent and widely available description that a book represents. I have declined to reference derivations posted only to closed sites, and in fact all my results have been obtained independently.

1.2 Rules, Procedures, and Terminology

Blackjack is a table card game. In a casino, the cards are dealt by a casino employee, here called Dealer. The casino and its staff are also referred to collectively as “the house.” Players, numbering from one to six or seven, occupy positions opposite Dealer and each plays individually against her, not against the others. (In blackjack tournaments, conversely, the contestants play against each other, not the house; tournaments require very different strategies than the casino game.) Any Player may play more than one position simultaneously, as long as the table is not full.

Each table uses one or more standard 52-card decks, shuffled together. In many casinos, such as those in the Northeast US, tables use either six decks or eight. In other casinos, where regulation permits more flexibility, some tables use one or two decks while the rest use either four or six. Dealers at one-deck and two-deck tables hold and deal the cards with their hands. They reshuffle when they wish. At tables with more than two decks, the cards, after shuffling and cutting, have a distinctive reshuffle marker card inserted somewhere halfway or more through, and are placed in a wooden box called a “shoe.” The cards are dealt, one by one, out of a slot at one end of the shoe. I also use the term shoe for the totality of cards shuffled together. I will always refer to the cards undealt at any moment of time, either in the shoe or held by Dealer, as the “pack.”

Each card has an associated value. The value of spot cards (except the ace) equals the number of spots; e.g., a three-spot has the value three. Face cards (i.e., jacks, queens, and kings) have value ten, just like the ten-spot. I will refer generically to all ten-value cards as 10s. The ace takes either of two values, 1 or 11, at Player’s

option, although usually the choice is obvious; I will discuss shortly the criterion for choosing one value or the other. The suit of a card is irrelevant to its value.

In a casino, money for blackjack and the other table games takes the form of chips, in various denominations, redeemed for cash at a cashier's window and purchased either there or at the tables. At the start of every round, Player selects the amount of his bet on that hand and places the corresponding chips on the table in front of him. The amount of his bet (in chip units) can be varied from round to round, as desired, except that the casino establishes a minimum and a maximum bet size for each table.

When all Players have placed their bets, Dealer deals one card to each Player (beginning at her left) followed by one to herself, face up (the "upcard"); then a second card in turn to each Player and one to herself, face down (the "hole card"). Typically, Player's cards are dealt face down in hand-dealt games and face up in shoe games. If Player's cards are face down, he may pick them up and hold them; he must not touch face-up cards.

Generally speaking, each Player in turn has the option of receiving additional cards, one at a time, until Player is satisfied with his hand. Player bases those decisions primarily (but not necessarily exclusively, as discussed in Chap. 2) on the total value of all cards he has received in that hand. Player's objective is simultaneously to maximize the total value of his hand and yet avoid the value exceeding 21. Player's judgment on how to balance these conflicting tasks should also make use of his knowledge of Dealer's upcard.

A two-card hand (held by either Player or Dealer) made up of an ace and a 10 is termed a "blackjack" (sometimes also called a "natural"). Because the ace is valued in this case at 11, the hand's total value is 21. If, after dealing the initial two rounds, Dealer's upcard is an ace, her next action is to offer Player a side bet, called "insurance," on whether or not her two cards form a blackjack. A Player electing insurance bets an additional amount equal to half his original bet, and is paid off at 2 to 1. Thus, if Dealer does prove to have a blackjack, Player without his own blackjack loses his original bet but wins the same amount back through the insurance. In the opposite outcome, Player loses his insurance bet but continues to play his hand against Dealer's.

In most casinos, Dealer with an ace or 10 upcard then "peeks," i.e., checks her hole card, to see if it forms a blackjack. If so, Dealer turns it up to show the Players, pays off any insurance bets, and play of that round is over. Any Players who have also drawn a blackjack tie Dealer; those without blackjacks lose their bets. (Significant for Chaps. 2 and 7 is that, for those hands whose play proceeds beyond this point, information has been gained about Dealer's hole card: when combined with the upcard it does not form a blackjack. For example, with an ace up, the hole card is deduced not to be a 10.) Whether or not Dealer peeks, any Player blackjack – should Dealer not also draw a blackjack and thereby tie – wins and is paid off at 3 to 2. For example, if Player bets \$20 on a hand and draws a blackjack, while Dealer does not, then he gets his bet back plus another \$30.

If neither Dealer nor at least one Player has drawn blackjack (or in casinos where Dealer does not peek), play proceeds. The first Player without blackjack now selects one of his several options:

1. He may receive no further cards, termed “standing” or also “sticking.”
2. He may double his bet and receive one and only one further card, termed “doubling down.”
3. If he has a pair of equal-valued cards, he may “split” them. Each of the two cards becomes a separate hand and Player doubles his bet: he places his original bet on the first hand and adds an equal amount on the second. Each hand wins or loses its bet on its own merits, irrespective of the outcome of the other. Provided the split pair is not aces, Player then plays each hand in turn: after each receives a second card, Player selects from the various play options (including, in many casinos, doubling) except:
 - (a) A two-card ace plus 10, although having value 21, is not a blackjack and is not paid 3 to 2
 - (b) If the second card to a split hand is equal in value to the first, again forming a pair, then in many casinos that pair can be resplit and option three repeats
 - (c) Split aces each receive only a second card and no more, and they cannot be doubled or resplit
4. If the hand is not a pair, or he elects not to split them, Player may receive another card, called “hitting” (or “drawing”). At this point his total hand value (i.e., the sum of the values of its component cards, recognizing the ace’s dual value) may or may not exceed 21. If the value is not over 21, then he may either stand (option 1 again) or receive a further card (option 4 again). If at any point the value exceeds 21 without an ace valued at 11, a condition termed “busting” or “breaking,” Player loses, irrespective of the pending outcome of Dealer’s hand: she immediately takes Player’s bet and puts his cards in a discard tray to await reshuffle.

A hand with an ace valued at 11 is called “soft.” Conversely, a hand without an ace valued at 11 is called “hard.” A soft hand with a value exceeding 21 can revalue the ace back to 1 and becomes hard; the hand value is now less than 21 and Player may either stand or hit. Any further aces drawn are valued at 1 and do not push the hand over 21.

When each Player in turn has finished, Dealer turns up her hole card and plays out her hand. Dealer, in contrast to the Players, has no options whatever. She cannot double down or split pairs, and must proceed according to fixed rules: if her hand value initially is 17 or more (with the first of any aces valued at 11) she must stand, whereas if it is 16 or less she must draw cards until the value exceeds 16. If a soft hand comes to exceed 21, the ace is revalued to 1 and play of the hand continues until its value exceeds 16.

When Dealer has reached either a stand or a bust, the bets are settled. If Dealer has busted, each Player who reached a stand is paid 1 for 1: he gets his bet back plus an equal amount. The Player who also busted has already lost; he does not tie.

If Dealer stands, then each Player's hand value is compared with that of Dealer: if Player is closer to 21 than Dealer, then he wins and is paid 1 for 1; if Player is farther from 21, then he loses his bet; if the values are equal, then the two hands tie (sometimes called a "push") and Player recovers his bet. Dealer places the used cards in a discard tray and the next round begins.

At the conclusion of the round during which Dealer has encountered the marker, at shoe-dealt tables, she combines all cards (both the discards and those still in the shoe), shuffles them together, cuts (or invites a Player to cut), replaces the marker card, and reloads the shoe. The fraction of the shoe dealt out before reshuffling is called the "penetration." Some casinos now use a shuffling machine to speed up the action; it randomizes a second shoe while the first is being dealt, after which the two shoes are switched. The most highly automated machines accept the discards after each round and continuously shuffle them into the cards remaining in the shoe; the penetration is always zero.

In most casinos, Players may join a table, or leave it, at the conclusion of any round. As shown in Sect. 4.3, there are strategy and performance implications in the decision to join or leave a table.

Although these rules are generally in effect at many casinos, variants are prevalent. Among the more frequently encountered departures are

- Dealer is required to hit a soft 17, rather than stand
- Soft hands may not be doubled
- Doubling a split hand after a second card is drawn (usually abbreviated "DAS") is not allowed
- Resplits are not allowed
- Dealer showing an ace or a 10 does not check her hole card for blackjack
- Player blackjack at single-deck games pays off at 6 to 5, rather than 3 to 2
- Player can enter a game only during a shuffle

Also, some casinos offer additional options to Player, the most common of which is called "surrender": when Dealer has not drawn blackjack, Player may elect to abandon any two-card hand, at the cost of losing half his bet.

Chapter 2

Playing the Hand

2.1 Basic Strategy

Even the most occasional Player wants to avoid egregious mistakes. He is looking forward to some fun, and losing heavily is not. He would like some guidance on how best to make all those choices available and, of course, what the odds are when he does make the best choices.

Every Player has information that is actually quite pertinent to his choices: he knows the identity of the cards in his hand and he knows the value of Dealer's upcard. Even this small amount of data is enough to provide a straightforward but powerful strategy for playing each hand, one that reduces the house advantage to much less than that of any other game in the casino. That recipe for play is usually called "Basic Strategy."

Basic Strategy is defined as the class of play guidelines with the best odds, given only the values of Dealer's upcard and Player's hand; the values of his individual cards here serve only to distinguish hard hands from soft and to identify pairs for possible splitting. Basic Strategy also assumes that each split hand, after the second card is dealt to it, is played with the same guidelines as for unsplit hands. Optimal Basic Strategy is that Basic Strategy with the best odds for the specific number of decks in the shoe, here labeled D . Discussed later are small improvements in play from use of the individual values of Player's first two cards, as well as that of the second card dealt to a split hand. Chapter 3 lays out the further improvements in optimal play enabled by information on cards dealt in previous rounds.

Optimal Basic Strategy is the same for all numbers of decks between three and six; because of its independence from deck number in this range, it is frequently referred to as Generic Strategy (to follow Wong (1994, pp. 26–27) with DAS, and Vancura and Fuchs (1998, pp. 28–29) without DAS). The guidelines for Generic Strategy are shown in Table 2.1.

The prescriptions of Generic Strategy strongly reflect one overriding characteristic: Dealer's hand is weak when showing less than 7, especially 5 or 6; and is strong when showing more than 6, especially ace or 10. When Dealer is weak, Player should draw cautiously, standing on as little as 12 or 13. When Dealer is strong, Player should draw aggressively, hitting until reaching at least 17. Furthermore,

Table 2.1 Generic Strategy (Optimal Basic Strategy for three to six decks)

Action	Hand value	When Dealer shows	
Double	11	2 through 10	
	10	2 through 9	
	9, soft 17 and 18	3 through 6	
	Soft 15 and 16	4 through 6	
	Soft 13 and 14	5 or 6	
Split		DAS allowed	DAS not allowed
	Aces and 8 + 8	Any	Any
	2 + 2 and 3 + 3	2 through 7	4 through 7
	4 + 4	5 or 6	—
	6 + 6	2 through 6	3 through 6
	7 + 7	2 through 7	2 through 7
	9 + 9	2 through 9, but not 7	2 through 9, but not 7
Stand on	13 and higher	2 or 3	
	12 and higher	4 through 6	
	17 and higehr should be higher	Ace, or 7 through 10	
	Soft 18 and higher	2,7, or 8	
	Soft 18, 3 or more cards	3 through 6	
	Soft 19 and higher	9,10, or ace	
Surrender, when available	15	10	
	16, but split 8 + 8	9, 10, or ace	

The final element of Generic Strategy is to always refuse the insurance bet

Player should double or split a number of two-card hands when Dealer is weak, while doubling or splitting very little when Dealer is strong.

The indication of strength or weakness from Dealer’s upcard reflects, in turn, the fact that she is required to draw to 16 or less. A two-card hand showing a 6 is more likely than not to have a value in the range 13–16, requiring her to draw with a greater than even chance of a bust on the next card. In contrast, a hand showing a 10 (even without a blackjack) is more likely than not to have a value of 17–20, requiring her to stand and posing a strong challenge to Player.

Table 2.2 lists the first round (or “reshuffle”) expected return with Optimal Basic Strategy, for four variants of the rules for pair splitting (although both DAS and resplits are allowed in most casinos). The returns, when plotted vs. the inverse number of decks, $1/D$, in Fig. 2.1, are virtually indistinguishable from straight lines. Table 2.2 also lists the slopes of the lines, recurring in Sect. 7.2.

As seen in Fig. 2.1 for the game with pair resplits and DAS, the return for an eight-deck shoe is -0.0043 ; the reshuffle return improves to $+0.0014$ for a single-deck shoe ($1/D = 1$). Without resplits or DAS, the single-deck return is -0.0001 .

Thus, Optimal Basic Strategy in a one-deck game has nearly even odds, slightly positive or negative depending on the rules for split pairs. For the other size shoes

Table 2.2 Optimal Basic Strategy expected return vs. number of decks

D	No DAS, no resplit	No DAS, resplit,	DAS, no resplit	DAS, resplit	Surrender increment
∞	-0.6902	-0.6510	-0.5704	-0.5904	+0.0932
8	-0.6901	-0.5730	-0.4877	-0.4310	+0.0825
7	-0.5974	-0.5617	-0.4758	-0.4197	+0.0810
6	-0.5819	-0.5468	-0.4599	-0.4047	+0.0790
5	-0.5601	-0.5258	-0.4378	-0.3837	+0.0764
4	-0.5274	-0.4943	-0.4046	-0.3523	+0.0726
3	-0.4731	-0.4420	-0.3495	-0.3000	+0.0662
2	-0.3621	-0.3349	-0.2368	-0.1930	+0.0553
1	-0.0147	+0.0018	+0.1143	+0.1419	+0.0236
Slope	0.336	0.323	0.344	0.325	

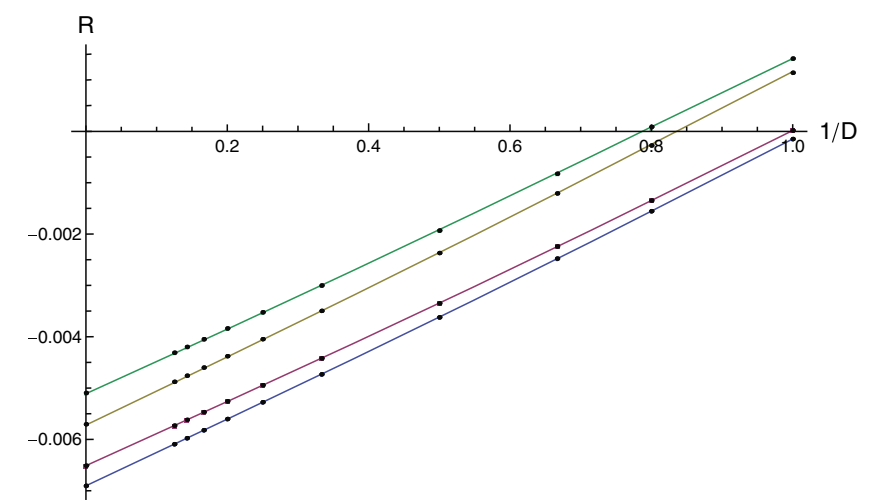


Fig. 2.1 Expected return, reshuffle round, vs. inverse number of decks (those in Table 2.2, plus 3/2 and 5/4)

typically encountered, the reshuffle return with resplits and DAS is -0.0040 for six decks and -0.0035 for four decks; without resplits or DAS the respective returns are -0.0058 and -0.0053 . The return improvement from surrender, also listed in Table 2.2, decreases with decreasing deck number, from the 0.0008 for eight decks down to just 0.0002 for a single deck.

The specifics of optimal card play also vary with number of decks. Table 2.3 lists the changes in play from Generic Strategy of Table 2.1, for one and two decks. Optimal Strategy for more decks than six differs in that $4 + 4$ is not split vs. Dealer's 5; soft 13 is not doubled vs. upcard 5 for more than eight decks; and, for more than 26 decks (!), soft 15 is not doubled vs. 4 (as first noted by Griffin (1999, p. 176)).

Table 2.3 Differences in Optimal Basic from Generic, one and two decks

Action	Hand value	When Dealer shows			
		<i>D</i> = 2		<i>D</i> = 1	
Double	11	Ace		Ace	
	9	2		2	
	8			5,6	
	Soft 19			6	
	Soft 17			2	
	Soft 13 and 14			4	
Split		DAS	No DAS	DAS	No DAS
	2 + 2				3
	3 + 3			8	
	4 + 4			4	
	6 + 6		2	7	2
	7 + 7		8	8	
Stand on	Soft 18 and higher		Ace		Ace
Surrender when available	15	Not 10		Not 10	
	16	Not 9		Not 9	

Table 2.4 Generic Strategy applied to other numbers of decks

<i>D</i>	Reduction in expected return × 100, Generic vs. Optimal			
	No DAS, no resplit	No DAS, resplit	DAS, no resplit	DAS, resplit
∞	0.0008	0.0008	0.0008	0.0008
3 through 6	0	0	0	0
2	0.0026	0.0026	0.0031	0.0031
1	0.0271	0.0272	0.0306	0.0310

The availability of DAS significantly increases the advantage of splitting pairs, enough so that more pairs should optimally be split than without it. Those added (at least in Generic Strategy) are 2 + 2 and 3 + 3 against Dealer upcards 2 and 3, 6 + 6 against upcard 2, and 4 + 4 against upcards 5 and 6; still more pairs are split for one and two decks. With DAS allowed, Player’s expected return increases by about 0.0014 with resplits and 0.0012 without.

Conversely, applying Generic Strategy to one or two decks is only slightly suboptimal: as seen in Table 2.4, the expected return degrades by at most about −0.0003 for one deck, depending on the rules for splitting. These reductions are so small, at least for two decks, that they are entirely outweighed by the simplicity of using just a single uniform strategy for any number of decks. With one deck, however, the return improvement from Optimal rather than Generic might be enough to prompt consideration of switching; but a very serious Player may elect instead to adopt

either the more complex, Composition-Dependent (but count independent) Strategy discussed below and/or a count-dependent strategy as per Chap. 5.

2.1.1 Expected Return with Variant Rules and Procedures

The rules variation most significant for Player is when Dealer does not stand on soft 17, but rather hits soft 17 and stands on soft 18 or more. Player's expected return worsens: with resplits but not DAS, by -0.0022 , at least for six and eight decks. The playing strategy remains mostly the same, but with some added situations where doubling is recommended: double 11 against Dealer ace, double soft 18 against Dealer 2, double soft 15 against Dealer 4, and double soft 19 (!) against Dealer 6. These additional doubles arise from Dealer's acquiring a stronger position, so that more aggressive play is required.

The variation where soft hands cannot be doubled is more benign; here the expected return worsens by -0.0008 . Not being able to resplit a split hand worsens the expected return by -0.0004 .

The variation in which Dealer does not peek (i.e., does not check her hole card for possible blackjack when showing an ace or 10) requires Player to make decisions on whether to double or split without the knowledge that she does not have blackjack. As a result, optimal play changes slightly for those upcards: do not double 11 or split $8 + 8$ against Dealer 10, and split no pairs (i.e., do not split aces or $8 + 8$) against Dealer ace. The expected return worsens by -0.0011 .

2.1.2 Expected Return vs. Return on Investment

Player's expected return is not the same as his "return on investment" (ROI). The former is defined as Player's average cash win or loss per hand, per unit base bet. The ROI, in contrast, is defined as the same cash increment but instead per typical amount of cash risked in playing the hand. Since hands that are doubled or split require a total bet of twice the initial bet (or more, in the event of DAS or resplitting), the expected return is clearly larger in magnitude than the ROI. A computation gives the ratio as about 1.12 for Basic Strategy play.

Although this distinction is slight for blackjack, it is highly significant in craps, where the best choice of bet has ROI of about -0.0084 . But this is the result of combining an initial unit bet on the "Pass Line," having expected return of -0.014 , with an equal bet on "Free Odds," allowed on an average of $2/3$ of all Pass Line bets, having an expected return of zero (i.e., no house advantage in its payoff). This ROI is only a bit more unfavorable for Player than Basic Strategy in blackjack. But Player nonetheless loses an average of 0.014 units per wager, much worse than Basic Strategy.

Table 2.5 Increase in expected return from Composition-Dependent play vs. Optimal Basic Strategy (resplits, no DAS)

D	∞	8	6	4	2	1
Return increase $\times 100$	0	0.0017	0.0028	0.0052	0.0132	0.0366

The concept of ROI will appear again in Sect. 9.1, where it plays a significant role in optimal betting.

2.2 Composition-Dependent Play

The focus thus far has been on Basic play strategies, using just the total value of Player's hand and Dealer's upcard. But in fact Player also knows the specific cards in his hand, not just their combined value. A reasonable question is whether, for example, play might be different for a hand whose first two cards are 2 and 6 than for a hand with 3 and 5, even though both have value 8. The more general question is whether the expected return can be further improved through "Composition-Dependent" play that takes account of the composition of those first two cards, not just their total value. The improvement, as seen in Table 2.5, is only slight.

Four, six, and eight-deck games have a best play strategy with only a tiny amount of composition dependence. The composition dependence of a two-deck game is greater, but still improves the return only slightly. A one-deck game benefits from playing with a Composition-Dependent Strategy, but the return increase, about 0.00037, is still of interest only to the very serious Player; his return would then range from +0.0002 to +0.0018, depending on the splitting rules. The strategy details are laid out in Sect. 7.3. Basic Strategy, of course, is defined without composition dependence; and composition *independence* is in fact optimal with a large number of decks, nearly so with as few as four decks.

Chapter 3

Tracking the Cards

So far, several items of information have been pointed out that help Player make decisions. The total value of the cards in his hand and the value of Dealer's upcard together lead to Basic Strategy. The specific composition of Player's hand is additional information of potential use but, as seen, provides at best only a slight improvement in return.

The most important information not yet used is the values of the cards dealt in previous rounds. If these could be tracked, Player would gain insight into the composition of the remaining pack and whether it is more or less favorable to him than immediately after the shuffle. Exploiting the information contained in the cards dealt during prior rounds is the basis for so-called "card counting."

The effect underlying the importance of card counting is that the composition of the pack fluctuates as cards are dealt. Although the *average* likelihood (averaged over many shuffles) of drawing a specific value for the next card is always the same, 1/13 in the instance of a five, the likelihood in any given pack can vary considerably from its average. If a shuffle has clustered the fives mostly in the first half of the shoe, the second half will offer a lower than average likelihood of drawing another. Furthermore, the deviations of the likelihoods from average will increase as more cards are dealt.

The success of card counting in blackjack stems from the fact that Player's expected return on a given hand is sensitively dependent on the likelihoods for that hand. (This is in contrast to most other card games; in baccarat, for example, the expected return varies negligibly with pack composition, and card counting is essentially useless.) In blackjack it is possible, and not uncommon, for the return to shift from slightly unfavorable at the start of a shoe, as seen in Chap. 2, to significantly favorable later in the pack. A Player who can gauge the shifts in return can then adjust the size of his bet accordingly. In those situations where the return has swung in Player's favor, he can increase his bet. If the favorable periods are sufficiently common and if during them the bet size can be increased enough, then Player can develop a net advantage over the house.

There are two important issues for developing a practical card counting method. The first is to decide what indicators of pack composition are meaningful and yet not beyond Player's capabilities. Clearly, if Player could count separately and simultaneously each of the ten different card values as they are dealt, then he would have

the maximum amount of information to predict the return on the next round. But rapidly updating ten distinct numerical registers in one's head, and arithmetically processing them to arrive at the current expected return, is more than almost any human can handle. Thus, the challenge is to find a simpler scheme that still retains as much predictive power as possible. Fortunately, to a very good approximation, only a single combined register is needed. Several variants of such a combination are detailed below, along with the objective measures by which they can be evaluated and compared.

The second issue is to identify a rule for adjusting bets according to the indicated return advantage. Although bets need to be increased during periods of favorable return, the bet on the next round should not jump by a large multiple just because the expected return has moved to barely positive; intuitively that seems to escalate risk unduly. The key to finding an optimal strategy for bet sizing is to balance reward with risk. Chapter 4 uses this key to develop a class of optimal betting strategies and the risk each entails.

3.1 Linear Counts

Even if no one is able to count each card value separately and simultaneously, it is quite feasible to track a single linear combination of those numbers. We introduce what we call the “counting vector,” a set of ten numbers indexed from 1 to 10. We call each number in the set an “element,” and its index corresponds to a card value. The ace is here always indexed as 1 (although some authors consider the ace to follow 10, i.e., 11). Typically, the elements of the counting vector are chosen to be integers, most simply to be either $+1$, -1 , or 0 ; in general, though, they could be any positive or negative numbers, not necessarily integers.

Next, we introduce the “running count.” At each shuffle, the running count resets to 0. If the first card dealt thereafter has value j , the running count increments by the j th element of the vector. If the second card dealt has value k , the running count increments again by the k th element. With every succeeding card, the running count further increments by the element of the counting vector corresponding to the value of the card dealt. (An example is shown in the next section.) Maintaining a running count throughout the dealing of a shoe, although at first seeming like a difficult task, can be mastered with a moderate amount of practice and concentration; it is, after all, only a single number, not ten of them, and usually an integer at that. A blackjack game program on a personal computer, or other electronic device, is useful for card counting practice.

The running count may also be recognized as a linear combination of the numbers of cards dealt of each value. Thus, if there actually were an individual who could keep ten separate registers in his head, then the best possible use he could make of this information for blackjack would be to form just that one linear combination corresponding to the running count!

Finally, we introduce what most authorities call the “true count,” or count per deck. The true count is the running count divided by the number of decks remaining in the pack, i.e., the number of remaining cards divided by 52. Although estimating this ratio in one’s head, on the fly, is obviously an added complication and a nuisance, it is not that difficult and is imperative for gauging the expected return. Running count by itself is not a quantitative measure of pack favorability, whereas true count is.

Fortunately, the number of decks remaining does not have to be known accurately; a rough estimate is sufficient. Useful is the fact that Player, on average, draws about 2.7 cards in playing his hand while Dealer draws about 2.8; thus a table with just one Player consumes 5.5 cards per average round, or about 9.5 rounds per each 52-card deck in the shoe. Player can simply track the number of rounds played since the last shuffle to approximate the number of decks remaining undealt. For example, a six-deck shoe contains cards sufficient on average for about 57 rounds, so 19 rounds after a shuffle correspond to a pack with about four decks remaining. Even more simply, some experienced Players estimate the decks remaining just from glancing at the height of the discard pile.

3.2 Choosing a Counting Vector

The majority of popular books on blackjack that discuss card counting advocate one or the other of two simple counting vectors, known as “Hi-Lo” and “Hi-Opt” and shown in Table 3.1.

As an example of keeping the running count, consider observing the cards 3, 10, 5, and 7 dealt in succession following a shuffle: the running count (with either Hi-Lo or Hi-Opt) would then be $+1 - 1 + 1 + 0 = +1$. Both schemes call for assigning $+1$ to card values 3 through 6, and -1 to value 10; values 7 through 9 are ignored. The schemes differ only in whether or not to pay attention to values 1 (i.e., ace) and 2. Since the Hi-Opt scheme allows Player to ignore more values in keeping his running count, it is clearly simpler than Hi-Lo. The latter, however, turns out to be a more accurate approximation for gauging expected return.

These, and a number of other schemes that have been proposed, can be assessed quantitatively as to merit. There exists an *optimal* counting vector, in the sense of giving a nearly exact measure of the expected return of a round based on the pack

Table 3.1 Most-recommended counting vectors

Card value	1	2	3	4	5	6	7	8	9	10
Vector element										
Hi-Lo	-1	+1	+1	+1	+1	+1	0	0	0	-1
Hi-Opt	0	0	+1	+1	+1	+1	0	0	0	-1

Table 3.2 Other counting vectors

Card value	1	2	3	4	5	6	7	8	9	10	<i>BC</i>
Vector element											
Optimum (<i>vb</i>)	-1.28	+0.82	+0.94	+1.21	+1.52	+0.98	+0.57	-0.06	-0.42	-1.07	1.000
Ultimate	-9	+5	+6	+8	+11	+6	+4	0	-3	-7	0.998
Halves I	-1.5	+1	+1	+1	+1.5	+1	+0.5	0	-0.5	-1	0.994
Halves II	-1	+0.5	+1	+1	+1.5	+1	+0.5	0	-0.5	-1	0.992
Revere	-1	+0.5	+1	+1	+1	+1	+0.5	0	0	-1	0.975
Hi-Lo											0.967
Hi-Opt											0.874

composition at its start. The difficulty with using this optimum in actual play is that its elements are not integers; the task of rapidly adding and subtracting a series of decimal numbers together in one's head is too hard. Hence an approximation to the optimum is needed that involves only simple integers, especially just the very simplest, +1, -1, and 0.

The appropriate figure of merit for an approximate counting vector has been called its betting correlation by Griffin (1999, p. 44) and here labeled *BC*. (It is defined precisely in Equation (8.30), with a slightly different label.) The optimal counting vector has a *BC* of exactly one, and every approximate counting vector has a *BC* less than one. The closer *BC* is to one, the better is the approximation. Hi-Opt has a *BC* of 0.874, while Hi-Lo has a *BC* of 0.967. The fact that Hi-Lo's *BC* is significantly closer to unity demonstrates its superior accuracy relative to Hi-Opt.

Table 3.2 displays a handful of other approximate counting vectors proposed in the literature, alongside the optimum. (For some of these vectors I have divided all elements as usually quoted by a factor of two, for ease of intercomparison; the scale change has no effect on performance.) They are listed in inverse order of their *BC*, showing that slight performance improvements can be achieved at the cost of considerably increased complexity, as decimal numbers are approximated ever more closely by integer or half-integer ones.

In my opinion, Hi-Lo strikes a good balance between simplicity and closeness to the optimum. Either one of the Halves gives an excellent approximation, with complexity only somewhat greater than Hi-Lo and much less than Ultimate. A convenient way to count the half integers, popularized by Snyder, is to assign 1 to the red cards, 0 to the black; reversing the colors, of course, works just as well. We drop Ultimate, whose tiny performance increment over the Halves hardly warrants the extra effort. An informal survey of Players by Wong in 2003 indicated that Hi-Lo was favored by more than a third of respondents who count, far more than any other method and preferred by 4 to 1 over Hi-Opt. Unbalanced counts, (see below) were used by just over a quarter, with the K-O method predominating.

3.3 Unbalanced Counting Vectors

The counting vectors listed above all have the characteristic that they are “balanced”: the sum of the elements (but weighting the tenth element by a factor of 4, since four distinct card faces have value 10) is constrained to equal 0. As a consequence, the count immediately after a shuffle resets to 0, as stated above. Several authorities (see particularly Snyder (2005), and Vancura and Fuchs (1998)) advocate unbalanced counting vectors, in which the constraint is relaxed: the reshuffle count is nonzero and depends on the number of decks. The more widely used unbalanced counting vectors, along with their betting correlation, are shown in Table 3.3.

But in fact any fixed constant can be added onto each and every element of a balanced counting vector (such as those listed in Tables 3.1 and 3.2), thereby unbalancing it, without changing its betting correlation, BC . In other words, every unbalanced counting vector has a balanced counterpart that performs identically. The main advantage, then, of unbalanced counting vectors is that ones with simple integer elements can be constructed whose balanced counterpart, necessarily with noninteger elements, has a BC superior to the favored all-integer approximation, Hi-Lo.

The disadvantage, though, of an unbalanced vector is that the running count must be adjusted, to subtract off its mean value, prior to converting it to a true count and using it for sizing the next bet. Advocates of unbalanced counts instead recommend using the running count directly, without any conversion to true count. Although this generally forgoes the close link between true count and expected return, the unbalance can be exploited to find that single specific running count value (called the “pivot”) where it always equals the true count. Then an empirical strategy is suggested for betting and card playing that interpolates between the pivot and the “key count,” that count at which the expected return is 0.

In my view, unbalanced counts only shift, rather than avoid, the mental arithmetic necessary to accurately assess the current expected return; instead of making the conversion from running count to true count, Player is required to carry out the pivot/key interpolation recipe. And the indicated bet size that results can only be approximate, not optimal. Nevertheless, despite this skepticism, the underlying algebra is detailed at the end of Chap. 8.

Table 3.3 Unbalanced counting vectors

Card value	1	2	3	4	5	6	7	8	9	10	BC
Vector element											
K-O	-1	+1	+1	+1	+1	+1	+1	0	0	-1	0.973
Red 7	-1	+1	+1	+1	+1	+1	+0.5	0	0	-1	0.969
Zen	-0.5	+0.5	+1	+1	+1	+1	+0.5	0	0	-1	0.963

Table 3.4 Relationship of true count to expected return

	Optimum	Halves I	Halves II	Hi-Lo	Hi-Opt
Coefficient	0.00459	0.00520	0.00580	0.00506	0.00511
Cross-over true count					
6 decks	1.184	1.045	0.936	1.074	1.063
4 decks	1.066	0.941	0.843	0.967	0.957

3.4 Relating the True Count to the Expected Return

Assuming that Player has selected a counting vector, he must be able to translate the true count, based on that vector, into the current expected return; once the return is indicated, a bet size can be chosen. But, at least for smaller true count magnitudes, the return and the true count are related linearly: the change in return from its initial, or reshuffle, value is roughly proportional to the true count. The choice of counting vector affects only the constant of proportionality: if the scale of a counting vector is modified (e.g., a Halves vector using 1, 2 and 3 rather than 0.5, 1 and 1.5), then the constant of proportionality is modified inversely by the same scale factor (in the example, reducing it by a factor of 2).

Table 3.4 lists the coefficients of proportionality for several of the more practical counting vectors. The threshold true count at which the expected return crosses over from unfavorable to favorable is also listed.

Remarkably, for all of these counting schemes a true count of about +1 is the threshold for a favorable expected return, which increases by about another 0.005 for each further increment in true count by +1. For example, a true count of +5 indicates an expected return of about +0.02.

For some of the computations to be reported later, it suffices to model multideck blackjack simply as a game with six decks, a reshuffle return of -0.005 and variance of 1.26, penetration of 0.8, crossover true count of +1, and a linear relationship between expected return and true count – independent of such rule details as the treatment of split pairs or of the particular play strategy used. This model is similar to the “Benchmark game” of Wong (1994, p. 18). For other computations, the deviation from linearity of the exact return/count relationship becomes significant.

Chapter 4

Betting

4.1 Yield, Risk, and Optimal Bet Strategies

The crux of Chap. 3 is that the true count is a valuable clue to the expected return on the next round. But Player is concerned operationally only with how much to bet, not with the expected return: he really needs a rule for converting true count directly into bet size. Furthermore, the mental exertion of counting cards and adjusting bet size is only justified to the extent that it gives Player an overall advantage over the casino. An appropriate quantity by which to measure that advantage is the average, over all rounds between shuffles, of the product of the expected return and the bet size, weighted by their probability of occurrence; this is Player's "yield," or his average cash flow per round. The average over the shoe is necessary because the first (and probably several) rounds after a shuffle have a negative expected return (except for the single-deck game, as seen in Chap. 2). Only when enough of the shoe is dealt to permit sufficiently likely fluctuations in composition that result in a positive return, and the bet is increased correspondingly, can Player make up for his expected losses early in the shoe and achieve a positive yield.

However, yield as defined here depends on the currency denomination of the bet, for example its size in dollars. More useful would be a denomination-independent quantity, the ratio of yield to a measure of bet size. One such measure is the "base" bet, the bet Player makes on the first round after a shuffle and on other rounds with a negative expected return. But since bets vary from round to round and are sometimes multiplied because of doubling or pair splitting, another measure is the typical bet (defined more precisely in Sect. 9.1), averaged over the shoe; the typical bet can be several times the base bet. For now, define the "yield ratio" as the yield per unit base bet; later, the yield per typical bet (the same as the return on investment discussed in Chap. 2) becomes significant.

Intimately tied to the issue of bet size is that of risk. A Player who makes very large bets relative to his current cash (his "capital," sometimes called his "wealth" or "fortune"), especially at rounds where his edge over the house is only slight, obviously has an increased risk of substantial losses – even of losing all his money. On the contrary, a Player who keeps his bets very small has little or no likelihood of gaining a positive overall advantage. The overriding determinants in specifying a

bet size are Player's capital as he starts his session, usually called his "trip capital" or "stake" (sometimes his "bankroll"), and his tolerance for the risk of losing it all, a distressing event traditionally called "ruin."

The occurrence of ruin does not usually imply that Player loses his total assets and can never gamble again. Instead, the stake he is lost is probably only a fraction of those assets; he is then able to tap into the remainder and return on a later occasion with a fresh stake. Over a single playing session without ruin, Player's expected capital would grow at a rate per round given by his yield; but when averaged over all sessions including those that end in ruin, the long-term expected capital growth rate is reduced. That latter rate is here called the "effective yield," and is lower than its no-ruin value by a "yield reduction factor"; most of our subsequent references to yield reduction factor will abbreviate it as YRF. Plausibly, the YRF and the risk of ruin are complementary: the YRF behaves qualitatively (to within a factor of two or less) like the probability of "survival," i.e., one minus the probability of ruin.

In sum, an optimal bet strategy strikes the best balance of risk vs. reward, specifically by maximizing Player's effective yield while fixing his ruin probability at a predetermined value. This criterion leads to a family of betting strategies, depending on the risk value selected.

Several consequences follow from the criterion. The first is that the ideal bet strategy is to size the bet on a hand in direct proportion to (i.e., linearly with) its expected return. But the bet obviously cannot be less than zero, even if the expected return is negative; and a zero bet violates the casino rule that it cannot be less than the table minimum. Rather, Player makes a base bet on negative return rounds; it may, though, be more than the table minimum. (Discussed in a later section is the technique of table-hopping, which avoids betting on most rounds with negative expected return.) Also, bets cannot exceed the maximum set by the casino for that table; but Player may wish to cap his bet size below that maximum, at a multiple, called the "spread," of his base bet. The ideal strategy, constrained in this way, is illustrated in Fig. 4.1, a relationship here called a "ramp." Furthermore, expected return and true count are also related roughly linearly, as discussed in Chap. 3; deviations from linear are taken up in Chap. 5. Thus, the ramp strategy sizes the bet, between the base and its spread multiple, roughly in proportion to the amount by which the true count exceeds the crossover threshold.

The second consequence of the optimization criterion is that the constant of proportionality between bet size and expected return (i.e., the steepness of the ramp) is tied to Player's risk of ruin. A steep ramp, as noted above, forces Player into large bets even with only slightly favorable odds, and hence magnifies the chance of ruin during a losing streak. On the contrary, a shallow ramp, although less risky, dictates small bets even with distinctly favorable odds and so hinders or even prevents Player's expected capital from growing.

The ramp in Fig. 4.1 further violates the casino rule that bet amounts must be in discrete multiples of the table minimum rather than smooth like the ramp. To accommodate this added reality while still approximating the ideal ramp as closely as possible, bets should instead be sized according to a jagged "staircase" in true count, as schematized in Fig. 4.1. Since a staircased bet has performance only modestly

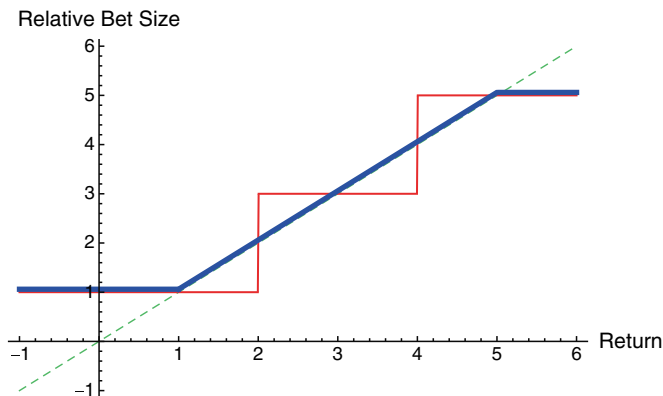


Fig. 4.1 Schematic representation of optimal betting strategy: ramp (thick line, the dashed diagonal constrained by lower and upper bet sizes) and a two-step staircase approximation

reduced from that of the corresponding ramp, as investigated in Chap. 8, most of the discussion here still focuses on the idealized ramp.

Although Player would naturally like to achieve the simultaneous goals of both high yield and low risk, instead the two are mutually exclusive and require a tradeoff. To pare down the multiplicity of variables and factors influencing the tradeoff, concentrate on just two examples of Player. One typifies a dedicated, intensive professional (call him the “Lifetimer”), spanning his playing career; the other, the “Weekender,” typifies a more casual player, although still a card counter. For the former, set his number of rounds at a million (about 10,000 hours at a typical 100 rounds per hour) and his trip capital at a thousand base bets; for the latter, set his number of rounds at a thousand (about 10 hours) and his trip capital at a hundred base bets. Assume for both examples the model Benchmark game, along with a bet spread of 10.

Computations for these examples find qualitative similarities. Risk plotted vs. steepness, in Fig. 4.2, shows a minimum for a small steepness value and a rise for larger ones. Effective yield also rises with steepness, but with a maximum at a larger value. Thus, effective yield increases with increasing risk, as plotted in Fig. 4.3, but only between the points where risk is minimum (at a low steepness) and where effective yield is maximum (at a high steepness). All points in between are legitimately optimal, according to the criterion of maximal effective yield for a given risk; and Player must then address his yield vs. risk tradeoff.

The Lifetimer, since he is very well capitalized, might at first choose the point that maximizes his effective yield, irrespective of risk; while the Weekender, undercapitalized, might instead choose to minimize his risk at a sacrifice of yield. But neither endpoint on the yield vs. risk curve is a sensible strategy: moving a bit away from either endpoint provides a much better yield/risk tradeoff. Instead, intermediate points along the curve seem more reasonable bet strategy targets; we will take two particular ones as examples for closer study.

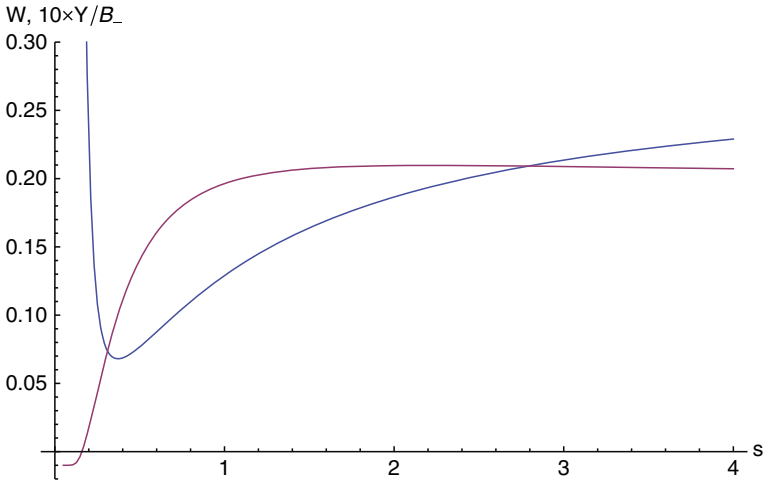


Fig. 4.2 Risk of ruin (curve with a minimum) and effective yield (curve with a maximum) vs. ramp steepness

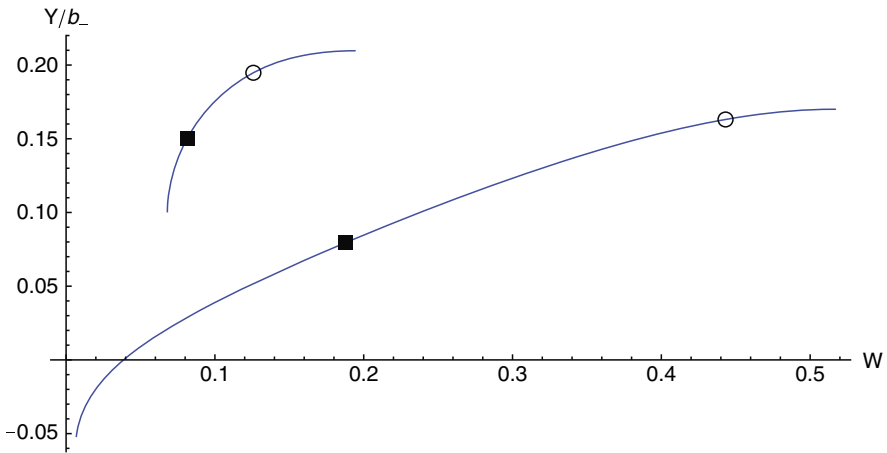


Fig. 4.3 Effective yield vs. risk of ruin, with points MM (solid squares) and HJY (open circles): upper curve for 1,000,000 rounds; lower curve for 1,000 rounds

Call one of these the “minmax (or MM)” point. It confronts the yield/risk tradeoff directly by maximizing the effective yield *per unit* of risk. Call the other the “HJY” point, originally identified by Harris, Janecek, and Yamashita (1997, HJY) through the quite different criterion noted below. For now, the two points are marked on each curve in Fig. 4.3, and Table 4.1 lists the corresponding operating parameters: steepness of the bet ramp, risk of ruin, and ratio of effective yield to base bet.

For the Lifetimer the HJY point seems to be a reasonable bet choice: its effective yield ratio is near the maximum attainable, yet with noticeably lower risk of ruin. For

Table 4.1 Operating parameters for the MM and HJY betting points, Lifetimer and Weekender examples

Player	Point	Steepness	Risk of ruin	Effective yield ratio
Lifetimer	MM	0.52	0.074	+0.013
	HJY	0.94	0.118	+0.017
	Max	2.22	0.193	+0.019
Weekender	Min	0.12	0.007	−0.005
	MM	2.80	0.136	+0.006
	HJY	9.37	0.409	+0.015

the Weekender, on the other hand, the MM point seems more reasonable: its risk is near the minimum allowed, yet with noticeably increased – and positive – effective yield ratio. The HJY point seems too aggressive and dangerous for the Weekender; the MM point seems too timid and unrewarding for the Lifetimer. Although any Player can choose another point from the continuum of optimal curves of Fig. 4.3, the Weekender should be careful to avoid too shallow a ramp: at less than half of the MM steepness the yield is negative!

Generalizing beyond these two examples, the risk and the YRF are determined by several factors. One is Player’s trip capital. A Player with a large stake relative to his typical bet, a ratio here called “coverage,” can easily afford to raise his bets when signaled by an above-threshold true count, while also riding out most losing streaks and averting ruin. A Player with low coverage, on the other hand, is vulnerable even to relatively brief (and hence frequently encountered) losing streaks.

A second factor is the number of rounds he plays: the more rounds, the greater his risk of ruin. A third factor is the yield, most usefully when measured relative to his typical bet rather than his base. As noted above, this ratio is Player’s return on investment, or ROI. Obviously, the greater the ROI the less the risk, when the other factors are the same. Maximizing the ROI for fixed spread is equivalent to the bet optimality criterion used by Harris (1997).

An important influence on the ROI is the penetration, the fraction of all cards that have been dealt when Dealer reshuffles. Immediately following the shuffle, the pack’s composition and expected return are known. The bet ramp must overcome a reshuffle return that, although dependent on the number of decks as seen in Sect. 4.2.1, is negative for multideck shoes. But as additional rounds are dealt, the pack’s composition can increasingly depart from that at the shuffle, so the distribution of returns on the next round can also vary from its reshuffle value: positive expected returns are encountered with increasing likelihood. The less the penetration, the less the overall proportion of rounds with positive expected return and opportunity for increased bet size, and so the less the capital growth rate for a bet ramp of given steepness.

Table 4.2 shows the operating parameters for the examples of Lifetimer (at the HJY point) and Weekender (at the MM point) vs. penetration, confirming the performance degradation as the penetration decreases. Note that the effective yield ratio is always positive, even for relatively low penetrations; the bet ramp steepens to

Table 4.2 Operating parameters vs. penetration, Lifetimer and Weekender examples

Player	Penetration	Steepness	Risk of ruin	Effective yield ratio
Lifetimer	0.8	0.94	0.118	+0.017
	0.7	1.28	0.209	+0.012
	0.6	1.75	0.319	+0.008
	0.5	2.53	0.449	+0.005
Weekender	0.8	2.80	0.136	+0.006
	0.7	4.70	0.223	+0.007
	0.6	7.90	0.308	+0.006
	0.5	13.28	0.374	+0.004

achieve this. But at high steepness the risk becomes quite large, signaling Player to compensate by increasing his coverage.

Although the several factors discussed above are each influential, the risk and the YRF are determined by just two parameters that combine them. One variable is the ratio of the number of rounds to the square of the coverage. The other is the product of the ROI and the coverage; call it the “ruin protection” since the risk varies oppositely to it. When the ruin protection is 1 or more, risk is tiny for all but extremely large values of the ratio of rounds to the square of the coverage. But for small ruin protection, and as that ratio exceeds unity, the YRF and survival probability fall below 0.85 and decline rapidly. To forestall ruin for a given number of rounds, Player’s coverage must be increased: his trip capital must be a larger multiple of his typical bet. With a YRF of 0.85, for example, a capitalization of 20 base bets (e.g., a \$200 stake at a \$10 minimum table) is barely adequate for 400 rounds: roughly one evening’s session at about 100 rounds per hour. A capitalization of 50–100 base bets is not too large for an avid player who might spend a dozen hours at the table over the course of several days (like the Weekender example). The requirement on capital is even more severe if Player aims for a ruin likelihood of as low as a few percent. This reality of blackjack puts extra pressure on Player: to limit risk and capture the reward promised by a counting strategy, he must be prepared to bring a large stake to the game. Conversely, undercapitalization can sabotage the performance of an otherwise optimized strategy (as well as prematurely end the gaming trip!).

4.1.1 Camouflage

Another reality of actual casino play is also significant, but only subjective. Virtually every blackjack authority advises that casinos try to protect against card counters by closely scrutinizing Player behavior and requiring such Players when detected to leave the blackjack tables (or possibly even the casino itself – permanently). To reduce the likelihood of detection, most authorities recommend that Players apply caution to their bet strategy: they advise restricting bets to no more than a moderate

Table 4.3 Operating parameters vs. spread, Lifetimer and Weekender examples

Player	Spread	Steepness	Risk of ruin	Effective yield ratio
Lifetimer (HJY)	20	1.20	0.188	+0.032
	10	0.94	0.118	+0.017
	5	1.03	0.143	+0.006
Weekender (MM)	20	2.82	0.147	+0.007
	10	2.80	0.136	+0.006
	5	3.17	0.104	+0.004

spread over the base bet, typically around 5–10; or, alternatively, restricting staircases to only a small number of steps (4 or 5, for example); the table maximum, by comparison, might be 50 or more times the minimum. Some authorities also suggest not increasing the bet, even if so indicated by the true count, unless the previous hand was won; and, conversely, not decreasing the bet unless the previous hand was lost. Tactics of these kinds (usually called “camouflage”) all lower the yield and increase the risk. Yet they depend on Player’s assessment of the casino’s oversight and enforcement practices, and of the balance he wishes to strike between aggressiveness and discretion. Such decisions are primarily judgmental, not amenable to optimization.

The degradation in performance from lowering the bet spread, however, can be computed. Table 4.3 shows this for the HJY Lifetimer and the MM Weekender. Dropping the spread much below 10 sacrifices a great deal of yield advantage.

4.2 Betting Proportional to Current Capital

One of the earliest advances in blackjack analysis, contemporaneous with the Basic Strategy computation of Baldwin et al. (1956), was an insight by Kelly (1956) applicable to betting. He pointed out the benefits of bet sizes that are always proportional to Player’s current capital. An underlying rationale is that a Player presently with a few hundred dollars should be betting less than if he had many thousands, irrespective of the capital he started with. Or alternatively, a Player who happens to be winning (or losing) big becomes inclined, quite rationally, to rescale his bet sizes upward (or downward) because his risk has changed.

This concept is mentioned, cursorily, in much of the subsequent blackjack literature. It is typically stated, following Kelly, that with this betting approach the expected capital increases exponentially with the number of rounds, rather than linearly as in the previous section, and that Player notionally can never be ruined. But he avoids ruin only by decreasing his bet size when/as his capital diminishes; he thereby comes to violate the casino’s table minimum rule. As a consequence, Kelly betting has been widely regarded as an impractical curiosity (in gambling, but decidedly not in financial investing – see Poundstone (2005)).

What seems like a fatal flaw, though, stems from a misinterpretation of the ruin concept. If Player is indeed experiencing a losing streak, with shrinking capital, and reaches the situation where his next bet, as dictated by the Kelly fraction of that capital, would drop below the table minimum, then his betting scheme dictates that he stop the play even though he still has some capital left. This event, rather than zero capital, is the more appropriate definition of ruin.

With this modified concept of ruin, Kelly betting becomes very much like the approach of the previous section. In both cases, the optimal bet size turns out to be directly proportional to the expected return (when positive) on the next round, although the factor of proportionality differs somewhat. Previously, the factor was adjusted to control risk and is proportional to Player's initial trip capital; here instead, the factor is just his *current* capital. Kelly betting is like risk-controlled betting where Player's capital has increased to the point that he can rescale his bets upward in keeping with his newly won wealth – or downward during losing streaks.

Rescaling of bet sizes can be frequent when Player is so well capitalized that he is betting large multiples of the table minimum and the bet sizes he is permitted are relatively close together. Alternatively, bets can be rescaled occasionally when Player's current capital has diverged significantly from its starting point. Either way, Kelly betting is a natural extension of risk-controlled betting.

4.2.1 *Multiple Hands with Rescaling*

Kelly betting becomes even more advantageous when playing multiple hands simultaneously (permitted as long as there are unoccupied seats at the table). Although multiple hands with additive betting have the same expected return per hand as a single one, as noted in Chap. 3, nevertheless a Kelly bettor who regularly redistributes capital among his seats can measurably increase the growth rate of his median capital and/or reduce its variance.

With one such approach, Player optimally allots for betting on a round a larger proportion of his total capital than he would with just a single seat, by a factor listed in Table 4.4, and distributes that allotment equally among the seats. (As an example, \$30 on each of two seats rather than \$40 on just one.) As a result, the median capital growth rate is greater by that same factor in Table 4.4, typically between 1.5 and 2.

The factor is less than the number of seats, limited by the correlation between the hands resulting from them all challenging the same Dealer hand. The increased capital growth rate occurs because those seats which are losing, and so might not recover enough to contribute significantly to Player's end-of-session capital, instead

Table 4.4 Factor of bet proportion and of median capital growth rate, vs. number of seats

Seats	1	2	3	4	5	6	7
Factor	1.00	1.46	1.72	1.89	2.01	2.09	2.16

are regularly reinforced by the winning seats and restored to a more normal “productivity.” While the capital growth rate is increased in this way, the variance of capital is similarly – and undesirably – increased.

This concern is removed in a second approach: the bet allotment per round is still distributed evenly among multiple seats, but its proportion of total capital is instead just the same as if it were all bet at just a single seat. (As an example, \$20 on each of two seats rather than the \$40 on just one.) The median capital rate of growth is less than that of the previous approach but is still greater than if the bet had been concentrated at one seat. The advantage here is that the capital variance is correspondingly reduced.

These changes are much more pronounced if each of the seats could be played at a separate table, rather than all at the same table: since the tables (and their Dealers) are now uncorrelated, the growth rate of median capital scales directly with the number of seats/tables. Although obviously impossible for a single Player, playing multiple tables simultaneously is easily realized by a team. The team only needs an arrangement to pool and redistribute their capital at appropriate intervals; it need not be nearly as frequent as every round to derive most of the benefit available. Such a team also gains from the size of its capital pool, thereby assuring it a much more steadily expanding stream of winnings than an individual could by himself.

4.3 Back-Counting and Table-Hopping

Performance can be improved even further by a playing technique commonly called “back-counting” (or sometimes “wonging”). The back-counter begins by counting a freshly shuffled shoe while standing alongside the seats and not betting. When and if the true count crosses a certain positive “entry” threshold, he sits down and starts betting.

In addition, he may choose to leave the table when and if the true count crosses certain negative thresholds, either prior to entry (called “departure”) or subsequent to entry but before the next reshuffle (called “exit”). If he does leave before the reshuffle, most likely he will immediately turn his attention to another table and repeat the process. If he is playing at a second table before the first is reshuffled, we will call him a “table-hopper.”

The optimal table-hopping strategy is specified by three true count thresholds: for entry, for exit, and for departure. Results are shown in Table 4.5, for the two Player examples of the Lifetimer and the Weekender, each at the Wong “Benchmark” game; for computational simplicity, risk is neglected.

The first rows give the yield ratio in the absence of any back-counting, i.e., “play-all.” The second rows give the optimal entry true count, without any exit or departure. The third and fourth rows give the jointly optimal entry and exit true counts, without and with table-hopping, respectively. The fifth rows give the optimal exit true count for a table-hopper who enters at a shuffle, without any back-counting. The sixth rows give the jointly optimal entry and departure true counts

Table 4.5 Optimal true count (tc) thresholds and yield ratio for table-hopping

	# Tables	Entry tc	Departure tc	Exit tc	Yield ratio
Lifetimer					
Play-all	1	0	$-\infty$	$-\infty$	0.0180
Entry only	1	2.09	$-\infty$	$-\infty$	0.0238
Entry and exit	1	1.89	$-\infty$	-3.57	0.0242
Entry and exit	2	1.81	$-\infty$	-3.18	0.0243
Exit only	2	0	$-\infty$	-1.91	0.0224
Entry and departure	2	2.08	-3.49	$-\infty$	0.0240
Entry, exit, and departure	2	1.85	-3.29	-3.45	0.0245
Weekender					
Play-all	1	0	$-\infty$	$-\infty$	0.0052
Entry only	1	2.60	$-\infty$	$-\infty$	0.0113
Entry and exit	1	2.21	$-\infty$	-2.72	0.0116
Entry and exit	2	2.15	$-\infty$	-2.57	0.0117
Exit only	2	0	$-\infty$	-1.54	0.0095
Entry and departure	2	2.58	-5.03	$-\infty$	0.0113
Entry, exit, and departure	2	2.14	-5.13	-2.61	0.0117

when table-hopping. The last rows combine all table-hopping procedures into a single jointly optimal set of entry, exit, and departure true counts.

The Lifetimer example is an aggressive one, such that his play-all yield ratio, without any back-counting, is already an attractive +0.018. Entry, at an optimal true count of +2.1, adds another 0.006 to the yield ratio, so is a valuable addition to his repertoire. Exit after entry adds still another 0.00045 to his yield ratio; the optimal exit true count is around -3.6, while the optimal entry shifts slightly downward to around +1.9. Departure is also of some value: the improvement in yield ratio from departure, even without exit, is nearly equal to that from exit without departure; and the combination of both exit and departure (always, of course, with entry and subsequent table-hopping) gives a yield ratio improvement about twice that from either alone.

Similar results also hold for the Weekender, especially considering that the Weekender's play-all yield ratio is so much lower than the Lifetimer's. The yield increments from table-hopping maneuvers are roughly the same for both Players.

Remarkably, the table-hopper who merely exits at a true count of roughly -1.5 to -1.9, and always begins play at a table immediately following a shuffle, accomplishes much of what entry and exit together can. This technique could be particularly difficult for casino surveillance to spot and so provides the table-hopper excellent camouflage with a relatively small sacrifice from ideal performance.

In summary, and irrespective of betting strategy, the simplest rule of thumb for the table-hopper is: enter in the true count range of +2 to +2.5, depart at around -4, and exit in the range -2 to -2.5. Together these improve the yield ratio by nearly 0.0065, a benefit much greater than from complicated variations in playing the hand to be discussed next in Chap. 5.

Chapter 5

Playing the Hand When the Count and Bet Vary

5.1 Play Strategies that Vary with the Count

As discussed in the previous two chapters, most Players count cards to gauge the appropriate amount to bet on the next round. But the best way to play that round also depends to some extent on the undealt cards, for which counting gives a reasonable indicator; some improvement in performance could be expected from count-guided play adjustments.

Yet any count-dependent play is quite complex. Dozens of “play parameters” must be specified, such as on which hands to stand or draw, which initial hands to double or split, etc., all depending on Dealer’s upcard. Each of these many play parameters may have its own, separate dependence on count, a lengthy set of procedures usually called “strategy indices.”

Because “ideal” play – parameters that change with the individual identities of every card as well as with the count – is generally regarded as too difficult (if not in fact super-human), several classes of simplification are possible with varying levels of precision. For a given count value, the composition-dependent class (as per Sect. 2.2), using the identities of just the first two cards plus the upcard, has only slightly reduced performance; the decision rules for subsequent cards are fixed. The value-dependent class, using just the total value of the initial cards rather than their identities (although still recognizing opportunities to double soft hands and split pairs), is much easier still, but with a small further performance penalty. Basic Strategy, whether Generic or Optimal, is the count-independent instance of this class. This chapter works exclusively with value-dependent strategies, while extending them to become count-dependent.

5.1.1 *Reconsidering the Counting Vector*

A count-dependent extension, however, also calls into question the choice of counting vector(s)! Some choices may be superior for gaining play advantage, even though less desirable for betting purposes, or vice versa. Most ideally, of course, Player would simultaneously maintain all nine separate and distinct counting

Table 5.1 Counting vector optimal for play variation

Card value	1	2	3	4	5	6	7	8	9	10
Vector element										
Optimal (vp)	+0.37	+0.33	+0.57	+1.10	+1.52	+0.22	+0.88	+0.66	-0.12	-1.38
Simplified	0	0	0	+1	+1	0	+1	+1	0	-1

registers (in effect, knowing the exact composition of the remaining pack); but this is clearly beyond human capability. Less ideal, but difficult nonetheless, is to keep a second, distinct count on which to base play decisions, a possibility suggested by several authors. Even more of a compromise is to use just a single count for both betting and playing decisions, one that optimizes the yield from both together. As an example, advocates of the Hi-Opt counting vector feel it does a better job than the Hi-Lo vector in gaining play advantage. But any of these techniques need to offer enough yield improvement to justify its added complexity.

The single counting vector that takes optimal advantage of play variation, without any variation in bet, is displayed in Table 5.1.

This vector (labeled vp) is much like that in Table 3.2 optimal for bet variation (and labeled vb), except for the striking reversal of sign for aces (card value 1) and the interchanged importance of values 6 and 8. Although vp has noninteger elements and so is not practical for actual counting, several simplifications with integer or half-integer elements suggest themselves, analogous to those of Tables 3.1 and Tables 3.2. The simplest, analogous to the Hi-Opt scheme but with the switch of $j = 3, 6$ with $j = 7, 8$, is also shown in Table 5.1; it has a play correlation of $PC = 0.95$. The Hi-Opt vector itself has $PC = 0.88$, larger than the Hi-Lo vector's $PC = 0.80$ and even the 0.83 of vb ; the higher play correlation of the Hi-Opt vector confirms the claim of its proponents that it does a better job than Hi-Lo for play variation.

The reversed effect of aces in playing vs. betting is at the root of the notion in the literature to count aces separately and to use the ace count for improved bet and play variation. This “ace adjustment” technique, as discussed for example by Griffin (1999, pp. 56 and 62), advises Player to start with a counting vector that does not recognize aces, i.e., is “ace-neutral,” such as Hi-Opt. At the same time, Player also keeps a separate count of aces – or, more precisely, the difference between the number of aces actually dealt and the number expected based on the depth (a modification necessary to keep the count balanced). Player then subtracts the aces “side-count” from the ace-neutral running count to arrive at the ace-adjusted running count. The latter is then used to guide the bet size, while the ace-neutral count is available (as above) for play variation. The ace-adjusted count has betting correlation $BC = 0.95$, an improvement over the 0.87 of Hi-Opt by itself, although not quite up to the 0.97 of Hi-Lo.

Finally, in the more general circumstance where two separate counting vectors are employed, yield improvement is maximized by using vb for bet variation and vp for play variation.

5.1.2 Count Dependence of the Play Parameters

To exhibit play strategy variation in some detail, Table 5.2 shows how the optimal hard stand decision varies with true count, in this example when counting six decks using *vb*. Optimal play becomes more aggressive (defensive) as the true count is increasingly positive (negative). The other shifts in play strategy, including those for doubles and splits, parallel those shown in Sect. 2.1 for a changing deck number, *D*; the single-deck strategy there corresponds roughly to that for many decks at a true count of about $+2\frac{1}{2}$. As pointed out by previous authorities, the most significant strategy variations are those that take place near zero true count: the decision whether to stand on 16 or 17 vs. 10, near true count $+0.3$; and whether to stand on 12 or 13 vs. 4, near -0.3 . These two shifts are evident in Fig. 8.3. The insurance decision, similarly emphasized by others, is reviewed in the next section. The strategy variations for one deck are qualitatively similar to these for six, typically with a shift of about one true count unit.

Nevertheless, play variation by itself improves yield by only a very small amount, much less than that available from optimal betting and/or from table-hopping. At best, the improvement is comparable to that found in Sect. 2.1 for playing Optimal Strategy with a small number of decks rather than Generic. The actual yield improvement depends sensitively on the penetration of the pack: a smaller penetration provides less improvement. Furthermore, about 2/3 of this yield improvement is accrued if play is varied based on the counting vector *vb*, optimal for betting, rather than the *vp* optimal for play; this improvement is additive to that resulting from any bet variation.

Examining this picture, I feel the improvement from play variation is too small to be worth the trouble of memorizing much if any of the extensive strategy indices, only a part of which are laid out in Table 5.2. For example, the improvement in yield ratio for the full array is less than $+0.001$ for a six-deck pack with *vb* as the counting vector (although it is larger for a single deck). It then follows that a varying play strategy has negligible influence on the evaluation and selection of a counting

Table 5.2 Ranges of true count, between -12 and $+20$, within which standing on a particular hard hand value, between 17 and 12, is optimal, depending on Dealer upcard; six decks

Upcard	Stand on					
	17	16	15	14	13	12
1	-12, +9.2	+9.4, +10.6	+10.8, +13.4	+13.6, +19.4	+19.6, +20	
2	-12, -11.0	-10.8, -7.4	-7.2, -4.4	-4.2, -1.2	-1.0, +3.4	+3.6, +20
3	-12, -11.9	-11.8, -8.6	-8.4, -5.8	-5.6, -2.8	-2.6, +1.4	+1.6, +20
4		-12, -10.0	-9.8, -7.0	-6.8, -4.2	-4.0, -0.4	-0.2, +20
5		-12, -11.6	-11.4, -8.8	-8.6, -6.0	-5.8, -2.4	-2.2, +20
6			-12, -9.4	-9.2, -6.2	-6.0, -1.8	-1.6, +20
7	-12, +9.4	+9.6, +11.4	+11.6, +15.8	+16.0, +20		
8	-12, +8.6	+8.8, +12.2	+12.4, +20			
9	-12, +5.4	+5.6, +9.8	+10.0, +20			
10	-12, +0.2	+0.4, +4.4	+4.6, +11.4	+11.6, +20		

Overall, insurance does not represent much incremental value, and certainly less than usually claimed: the Hi-Lo or Hi-Opt card counter can either disregard the offer entirely with little loss, or accept the bet for true counts above roughly +5.

5.2 Counter Basic Strategy for the Variable Bettor

Blackjack analysis has been absorbed for decades by the issue of how to simplify or truncate the many strategy index changes indicated by a varying card count. An example of a truncated scheme, the “Illustrious 18” (or its extension, the “Catch 22”), appears in Schlesinger (2005). An alternative and much simpler scheme is “Counter Basic Strategy” (here abbreviated CBS), proposed independently and from different lines of reasoning by Werthamer (2007) and Marcus (2007).

To begin, focus on the single counting vector optimized for betting. Rather than each play parameter being individually dependent on true count, instead seek the count-independent Basic Strategy that matches the count-dependent one at a “best” choice of true count. The “best” choice (the “effective true count”) is the one that maximizes Player’s yield, recognizing that he is betting different amounts on different hands. Since he makes higher bets for higher true counts, the distribution of his bet sizes is peaked at a true count that is significantly positive. Play that is optimal at or near that peak – not at zero count as is usually prescribed – is the “best” choice for the otherwise count-independent CBS.

The methodology and results are illustrated by a few representative examples, adopting *vb*, the counting vector optimal for betting. The same two betting examples are considered as previously, the Weekender and the Lifetimer. Results are displayed, in Table 5.4, only for games with a single deck (where yield is the most sensitive to choice of play strategy) and with four decks (representative of six and eight deck games, as well).

Clearly, CBS increases the yield over that from Optimal Basic by a meaningful amount. The improvement is particularly dramatic for an aggressive bettor like the Lifetimer against a single deck.

Table 5.5 details the CBS for a single deck, consistent with an effective true count in the vicinity of +5, and identical to the count-dependent but bet-independent play

Table 5.4 Yield ratios for Counter Basic, Optimal Basic and Count-dependent play strategies

Betting style	Lifetimer		Weekender	
Number of decks	1	4	1	4
Count-independent				
Optimal Basic	+0.076	+0.019	+0.006	+0.008
Counter Basic	+0.091	+0.023	+0.009	+0.011
Count-dependent	+0.083	+0.022	+0.013	+0.011
Effective true count	5.1	3.2	4.6	3.5

Table 5.5 Counter Basic Strategy, single deck

Upcard	Stand on	Double	Split
A	17; s18	<i>10</i> , 11	A, 8, <i>9</i>
2	<i>12</i> ; s18	9–11; s17, <i>18</i>	A, 6–9
3	<i>12</i> ; s18	9–11; s <i>15, 16</i> , 17, 18, <i>19</i>	A, 2, 3, 6–9
4	12; s18	<i>8</i> , 9–11; s13–18, <i>19</i>	A, 2, 3, 6–9
5	12; s18	8–11; s13–18, <i>19, 20</i>	A, 2, 3, 6–9
6	12; s18	<i>8</i> –11; s13–19	A, 2, 3, 6–9
7	17; s18	<i>9</i> , 10, 11	A, 2, 3, 7, 8
8	17; s18	10, 11	A, 8, 9
9	<i>16</i> ; s19	10, 11	A, 8, 9
10	<i>15</i> ; s19	<i>10</i> , 11	A, 8
Never take insurance			

(as per Fig. 5.1) at that true count value. The parameters that differ from those at zero true count (i.e., Optimal Basic) are bolded and italicized.

It is surprising to see in Table 5.4 that the yield ratio for CBS in some cases is actually superior to Count-dependent play. The comparison is even more dramatic vs. Illustrious 18 play, whose yield is typically around 70% of Count-dependent. The seeming paradox is understood by noting that Count-dependent play maximizes the reshuffle return, at zero depth. In contrast, CBS maximizes the yield, the dominant contributions to which come from sizable depths, where the true count is more likely to take on large, positive values and the expected returns and bet sizes are correspondingly larger. In such a regime, the depth dependence of the expected return for a given true count is especially influential in the computational results.

Chapter 6

Synthesis and Observations

6.1 A Practical, Nearly Optimal Strategy

The preceding chapters have amply illustrated their contention that blackjack is a game that gives Player many options and as a result can be optimized exhaustively. The reader's exhaustion may well have begun a number of pages back! But from the body of material presented a synthesis can be extracted. A simple picture for nearly optimal blackjack emerges, one that greatly enhances Player's yield yet avoids elaborate procedures that do not seem to justify their effort. Let's synthesize that picture, and then compare blackjack as a game and as an enterprise.

The most important message is that *how* one should play blackjack depends on *why* one is playing. The great majority of patrons in casinos are there primarily to have a good time. They enjoy the exciting atmosphere, the pleasure of playing an absorbing game, the thrill of risking a modest amount of money. Although they hope to go home financially ahead, they realize that they probably won't; their priority is to minimize losses and avoid ruin, which would abruptly and prematurely end their session. The remaining patrons, a small minority, are playing primarily to win and are prepared for the hard intellectual effort that winning takes. While the game is just as absorbing and exciting for these Players as for the recreational majority – possibly a good deal more so – it is the financial outcome that motivates them. Their competition against the house is uppermost and they search out every available tool to gain the edge.

For the recreational Player, I feel that Basic Strategy without counting is adequate. A particular virtue is simplicity: it is easy to learn (or to relearn, after a long hiatus) and to use at the table. Before my occasional casino visits, I practice typically for about 15 min; a first-timer, with no previous exposure, should be comfortable after about an hour. Making Basic Strategy decisions in real time is not demanding work, even without others at the table to slow down the pace.

But an attribute just as important is that the expected return from Basic Strategy is no worse than roughly -0.006 , even for a shoe with eight decks and unfavorable rules for splitting and surrender; it is better still with fewer decks (recall Table 2.1) and/or more opportunities for splitting and surrender. Although the house has an edge (except for the one-deck game with DAS and/or resplits), it is much less

unfavorable for Player than any other game in the casino. Certainly, anyone tempted to feed the slot machines (odds of -0.080 or worse) or to bet on the roulette wheel (-0.053) would be well advised to head over to a blackjack table!

Generic Strategy, from Table 2.1, uses just the total value of Player's hand and Dealer's upcard. The modifications in play procedure that optimize for the number of decks (Optimal, Table 2.3), or for the specific composition of each hand (Sect. 2.2), improve the expected return so slightly as to not seem worth the recreational Player's effort to memorize. Furthermore, he should bet the same fixed amount on every hand, as he does not have the information needed to vary his bet size objectively.

For the competitive Player, on the other hand, Generic Strategy by itself is not enough: he also must track the cards as dealt and adjust his bet size accordingly. With proper card counting and betting he can swing his yield from negative to positive: he has an edge over the house, at least in a statistical sense. "Proper" here implies more than just tallying each card as it is dealt: it also involves selecting a counting vector that adequately approximates the optimum, estimating the number of decks remaining in the shoe, and choosing a bet size staircase that appropriately balances Player's reward vs. risk, his greed vs. fear.

The card counting procedure, in review, requires a number of steps. Player must make some choices, a few of which are subjective.

- (a) Choose a counting vector. The optimal vector is too difficult for practical use, so an approximation must be adopted. Although several have been discussed, including unbalanced ones, I feel that the Hi-Lo is both simple enough and sufficiently accurate. The Hi-Lo vector from Chap. 3 is reiterated in Table 6.1.
- (b) Form the running count. During each shuffle, reset the running count to zero. As each card is dealt, increment the running count by the corresponding element of the counting vector: if a card of value j is dealt, increment the running count by the j th vector element.
- (c) Convert the running count into true count. First estimate the number of decks remaining undealt in the shoe. One way is to track the number of rounds dealt since the last shuffle. If Player is the only one at the table, the number of decks used is about one for each nine rounds dealt; with two Players it's six rounds, with three it's 5, with four it's 4. Then divide the running count by the *unused* decks to arrive at the true count. Since the conversion need only be approximate, estimating the remaining decks just by glancing at the size of the discard pile is quite adequate; rounding off the mental division is also adequate.
- (d) Choose a bet staircase (schematically, Fig. 4.1) to convert the true count into the bet size. Specifically, choose the size of the stair-tread, the span of true counts over which the bet size is a given integer multiple of the table minimum; and choose the spread, the ratio of maximum to base bets.

Table 6.1 Hi-Lo counting vector

Card value	Ace	2	3	4	5	6	7	8	9	10
Vector element	-1	+1	+1	+1	+1	+1	0	0	0	-1

- The choice, in turn, depends on Player's assessment of his desire for high yield vs. his aversion to ruin. A cautious Player (e.g., the Weekender) might select the MM stair-tread, where his ruin protection is large and his yield is moderately positive. On the contrary, a more aggressive Player (e.g., the Lifetimer) could select the HJY stair-tread, 50% narrower than the MM with a correspondingly steeper staircase, such that his yield is nearly doubled although his ruin protection is rather less. But the Weekender must avoid the temptation to widen his stair-tread much beyond the MM point, i.e., make his staircase even less steep, since the yield then degrades precipitously and soon becomes negative. The effort of card counting is hard to justify if it does not give Player an edge!
- The optimal size of the stair-tread also depends quantitatively on the penetration (as diagramed in Fig. 4.1). The less of the shoe that is dealt before a reshuffle, the more difficult is Player's achievement of a positive yield: the stair-tread must be narrower and the staircase steeper. Yet a bet size that varies substantially from one round to the next can call the house's attention to Player's card counting.
- In any case, bet the base amount for true counts of $+1$ or less, because the expected returns on those hands are negative.

(e) Play the hand using an appropriate Basic Strategy procedure.

- For two or more decks, use Generic Strategy as per Table 2.1. Deviating from Generic play based on the true count does improve the yield still further, but so slightly as to question its substantial extra effort. This recommendation also holds even for the more refined procedures of (1) modifying the counting vector to better accommodate play adjustments or (2) keeping a second true count exclusively to guide adjustments in play.
- Although Generic Strategy is still quite adequate for one deck, some Players who do not vary their bets may instead elect to use Optimal Strategy; a few may include some dependence on composition and/or true count. Table 6.2 recaps the modifications to Generic Strategy that convert it into Optimal for a single deck. The expected return is thereby improved by about $+0.0003$, a very small amount (yet with little or no incremental effort).
- When counting and adjusting bets against a single deck, substitute Counter Basic Strategy, Table 5.4, instead of Generic or Optimal.

The competitive Player can significantly enhance his results in several ways. First of all, he can table-hop, by joining a table only when he has observed its true count reach the entry threshold, $+2$ to $+3$, and/or leaving before the next reshuffle should the true count subsequently drop to the exit/departure thresholds, -2 to -3 . Since the pack after entry is quite favorable and his risk of ruin is otherwise low, he might increase the amount of his base bet and/or, for camouflage, flat bet throughout. Some casinos, however, in an attempt to protect themselves against back-counters, forbid joining a table except during a shuffle. Even in these casinos, though, the exit maneuver is quite valuable all by itself.

Table 6.2 Differences in one-deck Optimal Strategy from Generic

Action	Hand value	When Dealer shows
Double	9 and soft 17	2
	Soft 13 and 14	4
	Soft 19	6
	11	Ace
Split	6 + 6	2
	2 + 2	3
	4 + 4	4 with DAS
	6 + 6	7 with DAS
	3 + 3 and 7 + 7	8 with DAS
Stand	Soft 18	Ace
Do not surrender	16	9
	15	10

A second technique is to employ Kelly betting. Player should scale his bet size to his current (rather than initial) capital, by occasionally adjusting his base bet in step with his gains or losses. And if the casino is not too crowded, he benefits from playing multiple seats, allocating his total bet equally to each. For fastest capital growth the total bet should be a larger proportion of capital than if playing only one place; but if the proportion is just held the same, the capital variance is reduced even while its growth rate is enhanced.

Team play takes advantage of these techniques to their fullest. A group of optimal Kelly Players who pool and periodically redistribute their capital can come closer to reproducing, and neutralizing, some of the casino's own advantages: multiple (and uncorrelated) tables, large capital reserves, and reduced volatility. Table-hopping is also a particularly effective tool for the team that assigns some members as spotters, one to a table; each spotter then signals a "whale" teammate to join and place large bets, when the count at their table has reached the entry threshold. With these procedures, teams can achieve much better results, even per member, than any member could by himself.

The effective yield ratio from counting and optimal betting can be as much as +0.02 or more, even at multideck games; and still more when table-hopping. The precise result, though, (as well as the attendant risk of ruin) depends on the number of decks in the shoe and its penetration – set by the house – and especially on the bet stair-tread chosen by Player. Remarkably, Player can always adopt a stair-tread that gives him a positive effective yield, no matter how small the penetration, even though the resulting staircase may be conspicuously steep or, in an extreme case, bump against the table maximum.

Using a strategy with a positive yield does not necessarily mean, however, that Player will win during any given session. The yield is a measure of the *average* outcome, over a very large number of rounds. Over any moderate number of rounds, deviations from that average can be large. For example, consider playing 400 rounds – a several-hour session – from a six-deck shoe and compare the statistical distribution of outcomes for the recreational Player, whose Generic Strategy

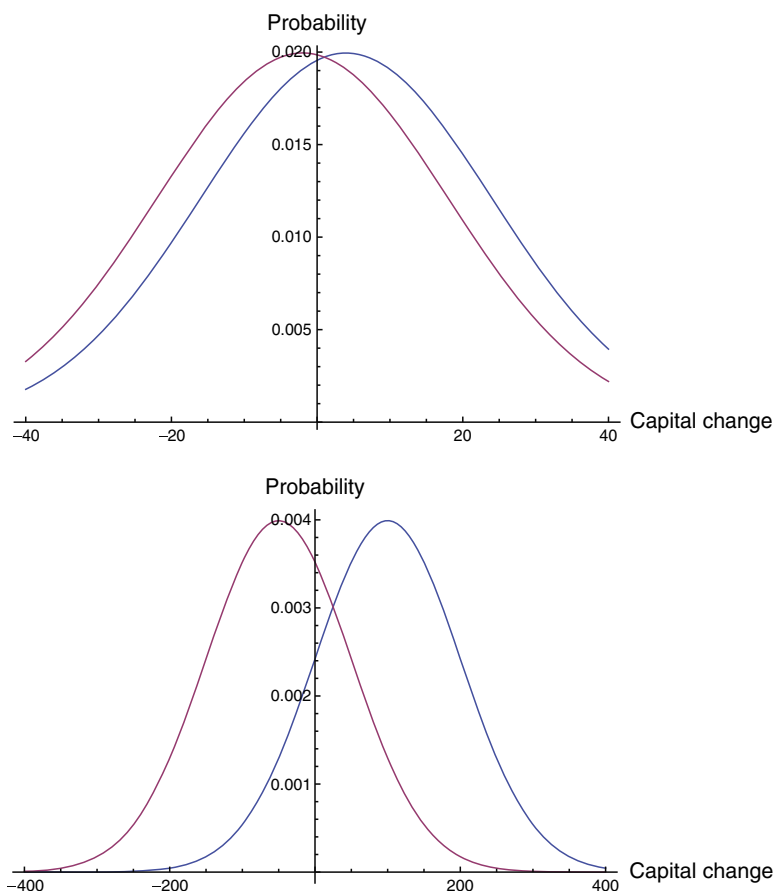


Fig. 6.1 Distribution of outcomes: *upper plot (a)* after 400 rounds, *lower plot (b)* after 10,000 rounds. For each plot, the left-most curve represents the noncounter, the right-most the counter

expected return is about -0.005 , with that of the competitive but camouflaged Player, whose yield ratio might be $+0.01$. *On average* over that session, the recreational Player loses 2 units while the competitive Player wins 4 units.

But the distributions of outcomes in Fig. 6.1a show that each Player is likely to win or lose up to 20 units or more. Even though the peaks (and means) of the distributions are indeed different (-2 vs. $+4$), the distributions themselves are so broad that the distinction between them is barely perceptible. The competitive card counter loses in 44% of such sessions; the recreational noncounter wins in 46% of them.

An observer who tracked the outcomes of the two Players during that session would be hard pressed to tell who was playing which strategy. Picking as the card counter that Player with the greater outcome would be correct only 57% of the time, not much better than an even chance. Moreover, the observer would not be

able to distinguish the two with much certainty until after dozens of such sessions. For example, after 10,000 rounds the distributions are well separated, as shown in Fig. 6.1b; the observer could then assign strategies to the Players – identify as the card-counter that Player with the better performance – with 82% accuracy. Turning the argument on its side, the Player who counts optimally does not have a substantial assurance of actually winning, or even of outperforming a noncounter, until he has played many thousands of rounds; and even then the assurance is far from total.

Yet the more rounds that are played (whether in one session or in multiple sessions) the greater is the possibility of ruin, terminating the session abruptly and sending Player home unhappy. Since Player's initial capital is finite, it can be consumed by a run of more losing than winning rounds, even if a strategy with positive yield is being used. But ruin does not turn a competitive Player's otherwise winning strategy into a losing one. Even though he may have been ruined in a given session, if he persists and plays many sessions (each, for simplicity of argument, with the same stake and counting strategy), he still has a positive but lesser effective yield as a long-term average. For the recreational Player, ruin lowers his negative yield still further.

The yield reduction factor due to ruin depends on three parameters. The first of these is the coverage (i.e., the ratio of capital to typical bet), which Player can most easily control. The probability of ruin can be decreased by increasing the coverage; a sufficiently high stake makes ruin negligible. Conversely, a recreational Player whose limited budget dictates a small stake should scale down his bets, particularly by choosing a table with the lowest possible minimum.

The second parameter is the ruin protection, associated with the bet staircase selected. Once Player has made his choice between a cautious bet strategy such as the MM point, and a more aggressive one such as the HJY point, the ruin protection is established.

The third parameter is the number of rounds played. Over a short session with at least a moderate coverage, ruin is quite unlikely. As more rounds are played, the risk of ruin grows, although it can be slowed (even if not prevented altogether) by a high enough coverage. In almost all realistic circumstances, ruin will be an occasional occurrence that must be accepted. Player should select a coverage that balances his desire to reduce risk of ruin against a prudent allocation to gambling of a sum he can afford to entirely lose, even if only rarely.

Money management is the least discussed but nonetheless critical aspect of optimal blackjack (as well as the most complicated to describe!). My sense is that many Players bring too little cash to the casino relative to the size of their bets, and are ruined unnecessarily often. The casino, on the other hand, wins steadily and inexorably because, relative to its patrons, it has vastly more capital, competes at multiple tables, and plays continuously rather than in limited sessions; thus, it can achieve returns much more consistent with its fundamental odds. The facts that those odds favor the house in every game (except, of course, blackjack against a close to optimal Player), and that most Players do not play close to optimally in any event, only reinforce the casino's profitability.

6.2 Blackjack as a Recreation vs. a Profession

The picture just described points up the issues facing a Player who seeks not merely to be competitive, but in fact to earn a livelihood exclusively from blackjack. Several well-known books on the subject suggest that blackjack can be a “business” (Revere 1980) or a “profession” (Wong 1994). But let’s see what it takes to accomplish this goal. Such a Player is devoting much of his time to working at a blackjack table as well as risking a substantial capital investment. The success of a blackjack professional should be judged on whether his winnings are an adequate return on the combination of his time (his “hourly wage”) and on his capital employed (what many businesses abbreviate ROCE).

Consider as an example a cautious Player with yield ratio of $+0.01$, who plays at a table with a base of \$20 and a typical rate of 100 hands per hour. On average, this Player wins \$20 per hour, a decent if unspectacular wage (it may be more than Dealer’s, though!). But such a Player would probably want coverage of around 500, or initial capital of \$10,000. His wage does not leave much room for a high ROCE. A more aggressive Player, on the other hand, with a yield ratio of $+0.02$, might play at a table with a base of \$50. His winnings average \$100 an hour, within the compensation range of other skilled professions. But his stake should probably be \$25,000, and perhaps even higher for added safety. Furthermore, because the investment is at high risk, it should probably be returning at least 20–40% annually, expensed out of winnings. This requires allocating to ROCE, for instance, at least \$5–10 per hour for the first 1,000 hours per year.

The real issue, though, for a blackjack professional is the degree of scrutiny he is likely to receive from the house. Although clearly not a topic open to objective analysis, it seems to me that any Player who repeatedly shifts his bet size up and down, regularly plays two dozen hours or so per week at a high stakes table (even though he probably changes casinos frequently), and wins more often than not, is going to be observed closely: that’s one reason why casinos have powerful video cameras above every table, and are equipped with sophisticated face-recognition software! If the casino then deduces that Player’s bet size scales with the true count, his patronage is probably not going to be welcome much longer. In trying to avoid this fate, he will no doubt employ several forms of camouflage. Although camouflage is an art, not a science, whatever techniques he adopts will necessarily degrade his performance. He will feel an ever-present tension between maintaining a decent yield and concealing his counting. Professional blackjack is not a day at the beach!

On the other hand, some experienced managers (see, e.g., Zender 2008) point out that many casinos overreact in trying to prevent card counting. For instance, frequent shuffling (low penetration) slows down the game and arguably reduces the casino’s overall profitability, even after allowing for some increased winnings by those customers who count cards. Preventing mid-shoe entry is similarly counterproductive. It is possible that expositions of this kind could lead to some relaxation of anticounter table practices in the future, although casino scrutiny is likely to remain a given in any event.

The competitive nonprofessional, however, faces less difficulty. He can lower his bet ramp somewhat to be less conspicuous. He will probably play occasionally, rather than steadily, and perhaps over a shorter session, so he is less likely to be recognized. He may feel comfortable with less camouflage and so his performance may more closely approach the optimum. His average winnings per session will probably be satisfying, although his total winnings over a year should resemble merely an income supplement or a bonus, not his livelihood. Such a Weekender can not only enjoy his session as much as the noncounting recreational Player, but also earn some compensation for his intellectual skills and effort.

Another contributor to Player's financial performance is casino comps, "gifts" to regulars including a mix of free room, meals, and entertainment (see Scott 2005). These are likely to comprise a more significant percentage of total compensation for the Weekender, who has less money at stake, than for the Professional, although both are otherwise faced with covering living expenses. On the other hand, a Player who accepts comps is immediately recognized by the casino, which can then evaluate his skill and track his activities.

The noncounting Player should think of blackjack as an entertainment. More likely than not he will lose some money, but the amount should typically not be greater than what he had to pay for an evening at the theater or concert hall (or, if his taste runs in a different direction, a pro sports game). He will experience a comparable degree of diversion and, on some occasions, he will have the added thrill of winning. The Player who counts, on the other hand, is working hard but at a game he enjoys. He should think of blackjack more like his bowling or softball league, with a big cash prize pool. He will practice to hone his skills and take satisfaction in how well he is using them. Slightly more often than not he will win some money, and the thrill is more in the looming danger of losing. For each kind of Player, though, there's a strategy optimal for his purpose in playing the game.

Although these realities and choices confront any individual Player, a team is a different story (for particularly vivid ones see Mezrich (2002, 2006)). A group of card-counters, working collectively, can take advantage of several betting techniques not available to the individual. The team can, in effect, table-hop, via multiple spotters who flat bet the table minimum and "whales" who enter on signal with big bets and exit as the count indicates. Team-members' time can be allocated most efficiently with the right ratio of spotters to whales. (The spotters can also depart and switch tables based on the true count threshold, if it would not attract too much attention.) Several whales, always at different tables (or even in different casinos), can occasionally (and out of sight of the casinos, of course) redistribute their capital and Kelly-rescale their bets, gaining almost its full benefit without reduction from any correlation between each other. As long as the team does not encounter casino interference, its win rate should be sufficient to cover both a competitive compensation to its members and a reasonable ROCE. A team is, in both principle and fact, a business enterprise – with all the attendant management issues of recruiting, training, performance assessment, compensation, assignments of responsibility, investor relations, etc. I stop short, though, of laying out a full team business plan here; such groups do not really need my advice!

Part II

Analysis

Chapter 7

Play Strategies

7.1 Basic Strategy, Large Number of Decks

7.1.1 Analytical Framework

Although casinos deal blackjack from a pack containing up to eight decks, analysis of the game assuming a much larger number of decks is instructive, both in the tools needed and in those readily obtainable results that are insensitive to the pack size.

Immediately following a shuffle, the first card drawn from the shoe has a 1 in 13 likelihood of having any specific value between 2 and 9, a 1 in 13 likelihood of being an ace, and a 4 in 13 likelihood of having value 10. In general, the likelihood of a specific value for any subsequently drawn card depends on which cards have previously been drawn; i.e., the composition of the remaining pack fluctuates as cards are dealt from it. However, for a shoe with a large number of decks the likelihood of drawing a specified value varies only negligibly from that of the first card. Thus, the approximation that those likelihoods are fixed and unvarying is the same as if the shoe contained an infinite number of decks. Griffin (1999) has called the shoe in this limit “nondepleting”; Epstein (1995) says that the draw is “with replacement.”

Let $d(i)$ be the likelihood, at any point in the course of dealing the shoe, of drawing the card value i , with $1 \leq i \leq 10$. Also let $d_0(i)$ be the likelihood function at the start of the shoe; clearly $d_0(i) = 1/13$ for $1 \leq i \leq 9$, and $d_0(10) = 4/13$. In the nondepleting shoe approximation, $d(i) = d_0(i)$ for every card drawn.

Next, introduce the pair of integers s and s' to parameterize the hit/stand rules. Let s and s' be the minimum hand values at which Player stands for hard and soft hands, respectively; assume that if a hand should stand on value s , it should also stand on values $> s$. For Dealer, under typical rules, $s = s' = 17$; in some casinos, $s' = 18$.

Now introduce the conditional probability that a hand valued at c , in the hit range, reaches the value j , in the stand range, after receiving one or more cards. Define $P(j|c)$ as the conditional probability for hard hands, with $c < s \leq j \leq 21$, and $P'(j|c)$ for soft hands, with $c < s' \leq j \leq 21$. These conditional probabilities

satisfy interlinked relations dependent on the size of c . Thus, for c in the range $11 \leq c \leq s-1$,

$$P(j|c) = d(j-c) + \sum_{k=1}^{s-c-1} d(k) P(j|c+k). \quad (7.1a)$$

The first term corresponds to standing after drawing one card (valuing an ace as 1) and the second corresponds to drawing the value k and hitting again.

The analogous relation for $P'(j|c)$ depends on the size of j relative to s' . A convenient way to accommodate this dependence is to define $\hat{j} \equiv j$ if $s' \leq j \leq 21$, and $\hat{j} \equiv j + 10$ if $s \leq j \leq s' - 1$. Then

$$\begin{aligned} P'(j|c) &= d(\hat{j}-c) + \sum_{k=1}^{s'-c-1} d(k) P'(j|c+k) \\ &+ \sum_{k=22-c}^{s+9-c} d(k) P(j|c-10+k). \end{aligned} \quad (7.1b)$$

The third term corresponds to drawing a card that would bust the hand except that the ace is revalued from 11 down to 1 and the hand becomes hard.

Finally, returning to $P(j|c)$ in the remaining range of c , namely $2 \leq c \leq 10$,

$$\begin{aligned} P(j|c) &= d(j-c) + d(1) P'(j|c+11) \\ &+ \sum_{k=2}^{s-c-1} d(k) P(j|c+k). \end{aligned} \quad (7.1c)$$

The middle term corresponds to drawing an ace, valued at 11 in this case, resulting in a soft hand; and the third term now excludes drawing an ace. Because the form of these relations depends on c and j relative to s and s' , both P and P' are also implicitly dependent on s and s' .

These relations look at first sight like a large array of simultaneous linear equations for the P and P' , a cumbersome task to disentangle. On closer examination, though, they can be seen to be merely recursive: for a given j , s , and s' , $P(j|c)$ is determined only by the $P(j|c')$ for $c' > c$. Also, $P'(j|c)$ can be expressed explicitly in terms of the P functions. A computer code to generate the P , given the likelihoods d , is straightforward and executes quickly. Still, the code can be streamlined by first placing the recursive aspect into polynomials $K(m)$, $K^*(m)$ generated from

$$\begin{aligned} K(m) &= \delta(m, 0) + \sum_{n=1}^m d(n) K(m-n), \\ K^*(m) &= \delta(m, 0) + \sum_{n=2}^m d(n) K^*(m-n); \end{aligned} \quad (7.2)$$

and then defining auxiliary functions

$$\begin{aligned}\Pi_{\sigma}(j|c) &\equiv \sum_{m=0}^{\sigma-1-c} d(j-c-m) K(m), \\ \Pi^*(j|c) &\equiv \sum_{m=0}^{10-c} d(j-c-m) K^*(m).\end{aligned}\tag{7.3}$$

The Kronecker delta functions in (7.2) are defined as $\delta(i, j) \equiv 1$ for $i = j$, and $\equiv 0$ for $i \neq j$. It is important when interpreting these definitions to recognize that $d(j) = 0$ when j is outside the range $1 \leq j \leq 10$. The expressions (7.1a)–(7.1c) can then be recast, respectively, in the nonrecursive form,

$$\begin{aligned}P(j|c) &= \Pi_s(j|c), \\ P'(j|c) &= \Pi^*(j|c) + d(1) \sum_{m=0}^{10-c} P'(j|c + 11 + m) K^*(m) \\ &\quad + \sum_{k=11}^{s-1} \Pi_s(j|k) \Pi^*(k|c).\end{aligned}\tag{7.4}$$

The conditional probabilities P determine the outcome of Player's contest against Dealer. Whereas for Player the variable c represents the total value of his first two cards (except for split pairs, analyzed later), similar probability functions also apply to Dealer, where the variable instead represents the value of just her upcard; we will use the notation u instead of c when referencing Dealer. Also, while $P(j|c)$ for Player depends implicitly on his stick/hit parameters s, s' , the corresponding Dealer's function $P(j|u)$ implicitly assumes her mandatory $s = s' = 17$ (or sometimes $s' = 18$).

7.1.2 Expected Player Return

For a specific c and u , the expected return to Player, labeled $R(c, u)$, is given by the probability that he wins (excluding blackjack), minus the probability that he loses, plus 1.5 times the probability that he (but not Dealer) gets blackjack. Furthermore, Player wins (excluding blackjack) either when his hand total is greater than Dealer's (but neither busts) or when Dealer busts (but Player does not). Player loses when his hand total is less than Dealer's (but neither busts), or when he busts, or when Dealer (but not Player) gets blackjack. In the case where Player's initial hand is hard (but not blackjack), this translates algebraically into

$$\begin{aligned}
R(c, u) = & \sum_{j=s}^{21} P(j|c) \sum_{k=17}^{j-1} P(k|u) + \sum_{j=s}^{21} P(j|c) \left[1 - \sum_{k=17}^{21} P(k|u) \right] \\
& - \sum_{j=s}^{20} P(j|c) \left[\sum_{k=j+1}^{21} P(k|u) - P(bj|u) \right] \\
& - \left[1 - \sum_{j=s}^{21} P(j|c) \right] [1 - P(bj|u)] - P(bj|u),
\end{aligned} \tag{7.5}$$

where bust hands have been eliminated using conservation of probability. Also, $P(bj|u)$ is the conditional probability of Dealer getting a blackjack when her upcard has value u ; in terms of the conditional probability $P^{(1)}(k|c)$ for reaching value k by drawing only one card to value c , $P(bj|u) = P^{(1)}(21|u)$. The lengthy expression of (7.5) for $R(c, u)$ can be manipulated to just

$$R(c, u) = \sum_{j=s}^{21} P(j|c) G(j|u) - 1, \tag{7.6}$$

where $G(j|u)$ is a linear combination of Dealer's conditional probabilities,

$$\begin{aligned}
G(j|u) \equiv & 2 \left[1 - \sum_{k=j}^{21} P(k|u) \right] \\
& + P(j|u) - \delta(j, 21) P(bj|u);
\end{aligned} \tag{7.7}$$

note for Dealer that $P(k|u) = 0$ if $k < 17$.

If Player's initial hand is soft (but not blackjack), (7.6) for the outcome still applies but with $P(j|c)$ replaced by $P'(j|c)$; the lower limit on the j sum remains s , not s' . Finally, when Player gets blackjack,

$$R(21, u) = 1.5 [1 - P(bj|u)]; \tag{7.8}$$

Player wins provided Dealer does not also draw blackjack.

From $R(c, u)$, four more steps are needed to derive Player's net expected return. First, $R(c, u)$ must be modified to account for his options of doubling and splitting. Second, the modified $R(c, u)$ must be weighted over the likelihood of his having a value c on his initial two cards. Third, the hit/stand parameters s, s' must be determined so as to maximize the weighted returns, for each Dealer upcard u ; the optimal s, s' are then functions of u . Finally, the maximized returns for a given u must be weighted with the likelihood of Dealer having u as her upcard. Let's take each step in turn.

Player *doubles* whenever doubling the bet and drawing only one additional card is more advantageous than his other play options. Thus, replace $R(c, u)$ by $\max\{R(c, u), 2R^{(1)}(c, u)\}$, where $R^{(1)}$ is the return from drawing just one card and

is obtained from the expression for R by replacing P by $P^{(1)}$; and $\max\{A, B, \dots\}$ denotes the largest of the group A, B, \dots . It may be advantageous to double a hand even if, played normally, it would be in the stand range; for example, Basic Strategy calls for doubling soft 18, rather than standing, when Dealer shows 3 through 6. Another consideration is that the doubling becomes an option only *after* Dealer has determined that she does not have blackjack; thus R and $R^{(1)}$ in the doubling expression must be further modified by adding back the term $P(bj|u)$, which had been subtracted out in the general expression.

Player *splits pairs* whenever doubling his bet and playing each card of the pair as a separate hand is more advantageous than his other options. The split option occurs when the two-card hand has value $c = 2i$, where i is the value of each individual card. Furthermore, splitting should also be compared with doubling, although splitting almost always yields a superior return to doubling; the exception, for Basic Strategy, is to double two fives rather than split them. Thus, replace $R(2i, u)$ by $\max\{R(2i, u), 2R(i, u), 2R^{(1)}(2i, u)\}$, at least when resplits are not allowed. As in the doubling analysis, the term $P(bj|u)$ must be added back into the split hand return, taking account of the knowledge that Dealer does not have blackjack.

Deriving an expression for Player return when resplits are allowed seems, at first sight, to be formidable; in fact, the task is straightforward. Let $R_{sp}(i, u)$ be the expected return for each card of the split pair. The second card drawn will either have value $k \neq i$, with probability $d(k)$, giving a hand with value $i + k$ (when k is not an ace, $i + 11$ when it is) which is played out normally; or the second card will again have value i , with probability $d(i)$, which prompts a resplit with return $2R_{sp}(i, u)$. Thus,

$$\begin{aligned} R_{sp}(i, u) &= \sum_{k \neq i} d(k) R(i + k, u) + 2d(i) R_{sp}(i, u) \\ &= \frac{\sum_k d(k) R(i + k, u) - d(i) R(2i, u)}{1 - 2d(i)}. \end{aligned} \quad (7.9)$$

But $\sum_k d(k) R(i + k, u) = R(i, u)$. Thus, the modification for resplits is that $2R(i, u)$ is replaced by $2(R(i, u) - d(i) R(2i, u)) / (1 - 2d(i))$. This modified expression exceeds $R(2i, u)$ (i.e., splitting is favored) precisely when $2R(i, u)$ does itself; hence, the option of splitting a pair is decided independently of whether or not resplits are allowed. Recall, of course, that aces cannot be resplit.

Once $R(c, u)$ has been modified for doubles and splits, it must be weighted with the likelihood of Player's hand having value c . The weighted return, denoted $R(u)$, is

$$R(u) = \sum_{c_1=2}^{11} d(c_1) \sum_{c_2=2}^{11} d(c_2) R(c_1 + c_2, u). \quad (7.10)$$

This sum must then be broken up into groups, each group being specific for whether or not the two cards form a pair without aces, or contain one or two aces without blackjack, or form blackjack. Comparing the $R(u)$ for all possible Player

hit/stand strategies yields the optimized return, $\max_{s,s'} \{R(u)\}$, for the specified Dealer upcard. The optimized return is then weighted over the upcard likelihoods, leading to an expected Player return

$$R = \sum_{u=1}^{10} d(u) \max_{s,s'} \{R(u)\}. \quad (7.11)$$

Combining all four steps into a single expression, lengthy but more explicit,

$$R = \sum_{u=1}^{10} d(u) \max \left\{ \begin{aligned} &2 \sum_{c_1=3}^{10} \sum_{c_2=2}^{c_1-1} d(c_1)d(c_2) \max \{R(c_1 + c_2, u), 2R^{(1)}(c_1 + c_2, u)\} \\ &+ 2 \sum_{c_1=2}^9 d(c_1)d(1) \max \{R'(c_1 + 11, u), 2R'^{(1)}(c_1 + 11, u)\} \\ &+ \sum_{i=2}^{10} (d(i))^2 \max \{R(2i, u), 2R_{sp}(i, u), 2R^{(1)}(2i, u)\} \\ &+ (d(1))^2 \max \{R(12, u), 2R^{(1)}(11, u)\} \\ &-(1 - 2d(1)d(10))P(bj|u) + 2d(1)d(10)R(21, u) \end{aligned} \right\} \quad (7.12)$$

The terms containing $d(c_1)d(c_2)$ and $d(c_1)d(1)$ correspond to hard and soft hands, respectively, evaluated for doubling; the terms with $(d(i))^2$ and $(d(1))^2$ correspond to pairs of nonaces and of aces, respectively, evaluated for splitting and doubling; and the term with $d(1)d(10)$ corresponds to blackjack.

7.1.3 Frequency of Tied Hands

Although most hands either win or lose, some hands tie and no money is exchanged. This occurs on nearly 10% of all rounds. Although this fact may not seem significant, since ties do not affect the player's financial situation, the frequency of ties does enter the picture in Sect. 4.1, which considers optimal betting and money management. Ties also affect the variance of the return, considered in the next section.

In a formalism suitable for the variance as well, begin with a compressed notation. First, let k index the final value of Dealer's hand, itemized as: stand on 17 to 21 but not blackjack, bust on 22 and higher, or blackjack on two-card 21. Further, let $\xi_k(\omega)$ be the probability (also depending implicitly on u) that Player wins/loses ω units against Dealer outcome k ; ω takes on the values (assuming no DAS or resplits) $0, \pm 1, +3/2, \pm 2$. Then the probability of tying is

$$T = \sum_u d(u) \sum_k P(k|u) \xi_k(0). \quad (7.13)$$

Now reexpand the notation to make explicit Player's initial hand value c , with probability $P(c) = \sum_{c_1, c_2} d(c_1)d(c_2)\delta(c_1 + c_2, c)$. Distinguish those values that are doubled, as c_d ; those that are split, as $c_s = 2i$; and those that are blackjack. Denote the probability of Player winning/losing ω units contingent on c and k as $\xi_k(\omega|c)$. Then

$$T = \sum_u d(u) \sum_k P(k|u) \left[\begin{aligned} & \sum_c P(c) \xi_k(0|c) + \sum_{c_d} P(c_d) \xi_k(0|c_d) \\ & + P(bj) \xi_k(0|bj) \\ & + \sum_{c_s} P(c_s) \left(\xi_k(0|i)^2 + 2\xi_k(1|i) \xi_k(-1|i) \right) \end{aligned} \right]. \quad (7.14)$$

Note (third line) that a split pair can tie not just by each of the two resulting hands maxing separately but also by one winning and the other losing.

Finally, substitute for the $\xi_k(\omega|c)$ in terms of the conditional probabilities and use conservation of probability to eliminate bust terms, yielding

$$\begin{aligned} T = \sum_u d(u) & \left[\sum_c P(c) \sum_{\kappa} P(\kappa|u) P(\kappa|c) + \sum_{c_d} P(c_d) \sum_{\kappa} P(\kappa|u) P^{(1)}(\kappa|c_d) \right. \\ & + P(bj|u) P(bj) + \sum_{c_s} P(c_s) \left[P(\kappa|i) (2P(\kappa|i) - 1) P(\kappa|u) \right. \\ & \left. \left. + \sum_{j=s}^{21} P(j|i) \left((1 - P(j|i)) G(j|u) - 2 \sum_{j'=s}^{21} P(j'|i) G(j'|u) \right) \right] \right] \end{aligned} \quad (7.15)$$

where κ runs over those stand values common to both Dealer and Player, $\max\{17, s\} \leq \kappa \leq 21$. Computation of this expression for Basic Strategy yields $T = 0.0982$. For the rounds that are not tied, Player wins about 42.7% of the total and loses about 47.5%; the fact that some winning hands pay more than 1 to 1, particularly blackjack at 3 to 2, substantially reduces the house's edge over Player from the 4.8% indicated by the difference of these two numbers.

It proves convenient for the next section to distinguish notationally the separate contributions to T from doubled, split and blackjack hands:

$$T \equiv \sum_c P(c) t + \sum_{c_d} P(c_d) t_d + \sum_{c_s} P(c_s) t_s + P(bj) t_{bj}. \quad (7.16)$$

7.1.4 Multiple Simultaneous Hands: Return, Variance, and Covariance

Intuition suggests that the expected return from H hands played simultaneously, each with a unit bet, should be identical to the return from a single hand played at

H times the bet. But a mathematical first look is less obvious. Although H unit-bet hands played *sequentially* obviously have the same total return as a single hand with a bet of H units, each of these hands competes against a separate Dealer hand. Multiple hands played *simultaneously* all compete against the *same* Dealer hand, and so might acquire some correlation that would affect the results.

Nevertheless, intuition is correct for the expected return. Carrying over the notation of the previous section, the expected return and the mean square return from a single hand are

$$\begin{aligned}\langle R_1 \rangle &= \sum_u d(u) \sum_k P(k|u) \sum_{\omega} \omega \xi_k(\omega), \\ \langle R_1^2 \rangle &= \sum_u d(u) \sum_k P(k|u) \sum_{\omega} \omega^2 \xi_k(\omega),\end{aligned}\quad (7.17)$$

so that the variance of the return is $\sigma_1^2 = \langle R_1^2 \rangle - \langle R_1 \rangle^2$. The total expected return from H hands against a single Dealer hand is

$$\langle R_H \rangle = \sum_u d(u) \sum_k P(k|u) \left(\prod_{h=1}^H \left(\sum_{\omega_h} \xi_k(\omega_h) \right) \right) \left(\sum_{h=1}^H \omega_h \right), \quad (7.18)$$

which, since $\sum_{\omega} \xi_k(\omega) = 1$ for every k , easily reduces to $\langle R_H \rangle = H \langle R_1 \rangle$ as claimed.

The distinction between simultaneous and sequential play of multiple hands sharpens when turning to the second moment (or mean square deviation, or variance) of the return. Putting pair splitting aside for the moment, the analog to (7.18) is

$$\langle R_H^2 \rangle = \sum_u d(u) \sum_k P(k|u) \left(\prod_{h=1}^H \left(\sum_{\omega_h} \xi_k(\omega_h) \right) \right) \left(\sum_{h=1}^H \omega_h \right)^2, \quad (7.19)$$

which can be manipulated into the form

$$\begin{aligned}\sigma_H^2 &\equiv \langle R_H^2 \rangle - \langle R_H \rangle^2 \\ &= H \left(\langle R_1^2 \rangle - \langle R_1 \rangle^2 \right) \\ &\quad + H(H-1) \left[\sum_u d(u) \sum_k P(k|u) \left(\sum_{\omega} \omega \xi_k(\omega) \right)^2 - \langle R_1 \rangle^2 \right] \\ &\equiv H \sigma_1^2 + H(H-1) \Gamma.\end{aligned}\quad (7.20)$$

The covariance Γ does not vanish because all Player hands are played against the same Dealer hand, and so are correlated. As a result, the variance of the H Player hands is greater than H times that of a single hand.

When pair splitting is reintroduced, the expressions grow longer but the key results,

$$\begin{aligned}\langle R_H \rangle &= H \langle R_1 \rangle, \\ \langle R_H^2 \rangle - \langle R_H \rangle^2 &= H \sigma_1^2 + H(H-1)\Gamma,\end{aligned}\quad (7.21)$$

remain the same. The single hand variance and the covariance generalize to

$$\begin{aligned}\sigma_1^2 &= \sum_u d(u) \sum_k P(k|u) \left[\sum_{\omega} \omega^2 \xi_k(\omega) \right. \\ &\quad \left. + \sum_{\omega_1, \omega_2} (\omega_1 + \omega_2)^2 \xi_k(\omega_1) \xi_k(\omega_2) \right] - \langle R_1 \rangle^2, \\ \Gamma &= \sum_u d(u) \sum_k P(k|u) \left[\sum_{\omega} \omega \xi_k(\omega) \right. \\ &\quad \left. + \sum_{\omega_1, \omega_2} (\omega_1 + \omega_2) \xi_k(\omega_1) \xi_k(\omega_2) \right]^2 - \langle R_1 \rangle^2,\end{aligned}\quad (7.22)$$

where here the ω sums are over nonsplit results and the ω_1, ω_2 sums are over the results of the two split hands.

Substitution for the $\xi_k(\omega)$ and manipulation as before (making use of the notation of (7.16), as predicted) leads to the compact expression

$$\begin{aligned}\sigma_1^2 + \langle R_1 \rangle^2 &= \sum_c P(c) (1-t) + 4 \sum_{c_d} P(c_d) (1-t_d) \\ &\quad + (9/4) P(bj) (1-t_{bj}) + 4 \sum_{c_s} P(c_s) [1-t_s \\ &\quad - (3/2) \sum_u d(u) \sum_{\kappa} P(\kappa|u) P(\kappa|i) (1-P(\kappa|i))].\end{aligned}\quad (7.23)$$

Thus, the probability of tying enters essentially into the variance, which can be characterized as the sum of the squares of the amounts won or lost, each weighted by their probability excluding ties. The computation of the variance is thus an easy extension of that for the probability of ties, resulting in $\sigma_1^2 = 1.261$ and $\sigma_1 = 1.123$.

The covariance expression in (7.22) can also be manipulated into a computation-friendly form, but most compactly with a somewhat different notation from that of the variance. First define the auxiliary quantities

$$\begin{aligned}\wp(j) &\equiv \sum_c P(c) P(j|c) \\ &\quad + 2 \sum_{c_d} P(c_d) P^{(1)}(j|c_d) + 2 \sum_{c_s} P(c_s) P(j|i);\end{aligned}$$

$$\wp \equiv \sum_j \wp(j) = \sum_c P(c) + 2 \sum_{c_d} P(c_d) + 2 \sum_{c_s} P(c_s); \quad (7.24)$$

$$\wp^*(j) \equiv \sum_i \varepsilon_{i,j} \wp(i), \quad \varepsilon_{i,j} \equiv +1, i > j; -1, i < j; 0, i = j.$$

Then in this notation the expected return of (7.17) can be recast as

$$\langle R_1 \rangle = \sum_u d(u) \left[\sum_j \wp(j) G(j|u) - \wp \sum_k P(k|u) \right] + R_{bj}. \quad (7.25)$$

The covariance analogously becomes

$$\Gamma = \sum_u d(u) \left[\sum_j \begin{pmatrix} 2\wp(j) \wp^*(j) G(j|u) \\ -\wp(j)^2 P(j|u) \end{pmatrix} + \wp^2 \sum_k P(k|u) \right] \quad (7.26)$$

$$+ [1.5P(bj) - \langle R_1 \rangle]^2 + P(bj) [1.5P(bj)]^2.$$

Computation of this form gives $\Gamma = 0.472$. Previous authorities have either estimated the covariance as $\Gamma \approx 0.50$ (Epstein 1995, p. 249; Griffin 1999, p. 142) or evaluated it using simulations as 0.47 (Wong 1994, Table 85 (no DAS), p. 203).

7.1.5 Expected Number of Cards Used per Round

The conditional probabilities P can also be used to derive the expected number of cards used in playing a round. Although this aspect is not directly relevant to Player's financial return, as noted before, it is useful in Chap. 3, which discusses card counting methods.

Looking back for a moment at (7.1a)–(7.1c) satisfied by $P(j|c)$ and $P'(j|c)$, note that they could be iterated to express $P(j|c)$ as a power series in d . But for each such term, the number of cards drawn is just the number of factors of d . Thus, expressions for $N(j|c)$ and $N'(j|c)$, the expected number of cards used in moving the hand value from c to $j \leq 21$, can be obtained by replacing in $P(j|c)$ (at least symbolically) every factor of d^n by nd^n . Such a replacement can be accomplished concretely by functionally differentiating the equations for $P(j|c)$, $P'(j|c)$ with respect to d , since

$$d \frac{\delta}{\delta d} d^n = n d^n. \quad (7.27)$$

Functionally differentiating (7.1a), for example, yields the recursion relation

$$\begin{aligned}
 N(j|c) &= d \frac{\delta}{\delta d} P(j|c) \\
 &= P(j|c) + \sum_{k=1}^{s-c-1} d(k) N(j|c+k), \quad (7.28)
 \end{aligned}$$

which is similar to (7.1a) except for the change in the inhomogeneous term. Coding these relations to compute N is only a minor supplement to the code for computing P itself.

However, a consideration not present in analyzing the return is that cards are also used for hands that bust and must be included alongside those used for hands that stand. But (7.1a), and by extension (7.25), are also valid for $21 < j < s + 10$. Furthermore, the quantity of interest is $N(c)$, the number of cards drawn to a given initial hand, irrespective of the final hand value or whether or not the hand busts. Then summing (7.28) over all $s \leq j \leq s + 9$ gives the very simple result,

$$N(c) = 1 + \sum_{k=1}^{s-c-1} d(k) N(c+k) = \sum_{k=1}^{s-c-1} K(m). \quad (7.29)$$

In the remaining ranges of c , $N(c)$ satisfies the corresponding analogs of (7.4). In particular, by introducing analogs to Π , Π' of (7.3),

$$\Omega_{\sigma}(c) \equiv \sum_{m=0}^{\sigma-1-c} K(m), \quad \Omega^*(c) \equiv \sum_{m=0}^{10-c} K^*(m), \quad (7.30)$$

N and N' can be expressed as

$$\begin{aligned}
 N(c) &= \Omega_s(c), \\
 N'(c) &= \Omega_{s'}(c) + \sum_{k=12}^{s-1} \Omega_s(k) \Pi_{s'}(k+10|c), \quad (7.31) \\
 N(c) &= \Omega^*(c) + d(1) \sum_{m=0}^{10-c} N'(c+11+m) K^*(m) \\
 &\quad + \sum_{k=11}^{s-1} \Omega_s(k) \Pi^*(k|c).
 \end{aligned}$$

The next steps in deriving the total number of cards used by Player, N_p , are to modify $N(c)$ and $N'(c)$ for doubles and splits, which depend on Dealer's upcard u ; to weight the result by the probabilities of c and u ; and then to add 2 for the initial two cards in Player's hand. The result is

$$\begin{aligned}
N_p = & 2 + \sum_{u=1}^{10} d(u) \left[\sum_{c_1=3}^{10} \sum_{c_2=2}^{c_1-1} d(c_1) d(c_2) [N(c_1 + c_2), 1] \right. \\
& + \sum_{c_1=2}^9 d(c_1) d(11) [N'(c_1 + 11), 1] \\
& + \sum_{i=2}^{10} d(i)^2 [N(2i), 2 + 2N_{sp}(i)] \\
& \left. + d(1)^2 [N'(12), 2] \right].
\end{aligned} \tag{7.32}$$

The number of cards used in splitting is given by $N_{sp}(i) = N(i)$, without resplits, and $= (N(i) - d(i)N(2i))/(1 - 2d(i))$ with resplits.

A similar analysis holds for Dealer, except that she neither doubles nor splits; her corresponding quantity $N(u)$ is weighted over the probability of drawing the upcard u ; and her initial card number is just 1, not 2. Hence

$$N_d = 1 + \sum_{u=2}^{11} d(u)N(u). \tag{7.33}$$

Computation of these expressions gives $N_p = 2.75$, $N_d = 2.91$.

But the analysis has assumed that Dealer's and Player's actions are uncorrelated. In casino practice, though, there are several interrelationships between the two, particularly when only a single Player is at the table: if either Dealer or Player has blackjack, the other does not draw additional cards; and, if Player busts, Dealer similarly does not draw additional cards. With these modifications, the respective card numbers change to $N_p = 2.72$ and $N_d = 2.76$, with a 5.48 total average number of cards used per round. For multiple Players at the table, the average number of cards used by Dealer and by each Player rises slightly.

7.2 Basic Strategy, Small Number of Decks

7.2.1 Analytical Framework

Now turn to treating the game with a small number of decks. An exact analysis accounting for depletion is daunting: rather than the compact recursion relations of the previous section, now the likelihood of drawing a card at any point in the play depends on precisely which cards were already drawn, both in that round and in all previous rounds of the shoe. As a consequence, every hand allowed by the rules must be catalogued and its probability evaluated; for hands that split pairs, the possibilities reach into the billions. Here we concentrate on just the first round

after a shuffle; subsequent rounds, further into the pack, are considered in the next chapter.

The number of distinct hands needing individual consideration can be reduced considerably by recognizing that:

1. Hands for which Player and/or Dealer bust can be eliminated (using conservation of probability) in favor of a more complex weighting of those that do not, generalizing the result of (7.6)–(7.7)
2. Rounds that tie do not contribute to the expected return
3. All hands that have the same cards (even though in different sequence) with each conforming to the rules, need evaluation only once with an appropriate weighting. This claim requires some justification

Begin by denoting the number of decks used as D , so that immediately after the shuffle the pack contains $52D$ cards. Assume that M cards have been dealt in playing one or more previous rounds; despite claiming that only the first round after a shuffle is being considered here, the extension to $M > 0$ provides a base for an important result proved in Chap. 8. Then $52D - M$ cards remain in the pack at the start of dealing the current hand. Its first card (dealt to Player) has a value c_1 with likelihood $d^{(1)}(c_1) = d(c_1)$; immediately after the shuffle $d(c_1)$ is unchanged from its value $d_0(c_1)$, but this is generalized in Chap. 8. The second card (dealt to Dealer) has value u with likelihood

$$\begin{aligned} d^{(2)}(u; c_1) &= \frac{(52D - M) d(u) - \delta(u, c_1)}{52D - M - 1} \\ &= d(u) + \frac{d(u) - \delta(u, c_1)}{52D - M - 1}. \end{aligned} \quad (7.34)$$

The numerator of the first expression gives the number of cards remaining in the pack with value u , reduced by one if u equals c_1 ; while the denominator is the total number of cards remaining after the first is dealt. The third card (dealt to Player again) has value c_2 with likelihood

$$d^{(3)}(c_2; c_1, u) = d(c_2) + \frac{2d(c_2) - \delta(c_2, u) - \delta(c_2, c_1)}{52D - M - 2}. \quad (7.35)$$

Extending this sequence to the m^{th} card with value j_m gives the likelihood expression

$$d^{(m)}(j_m; j_1, \dots, j_{m-1}) = \frac{d(j_m) - \varepsilon \sum_{k=1}^{m-1} \delta(j_m, j_k)}{1 - \varepsilon(m-1)}, \quad (7.36)$$

where $\varepsilon \equiv 1/(52D - M)$, the inverse number of cards remaining, can be considered a small parameter.

The probability of any given hand is just the product of the likelihoods of each card in the hand, i.e., a product of factors (7.36), noting that each factor depends on the identities of the cards preceding it. But the order of the cards can be interchanged

without affecting the hand's probability. Specifically, from (7.36) it can be shown that

$$\begin{aligned} d^{(m)}(j_m; j_1, \dots, j_{m-1}) d^{(m+1)}(j_{m+1}; j_1, \dots, j_m) \\ = d^{(m)}(j_{m+1}; j_1, \dots, j_{m-1}) d^{(m+1)}(j_m; j_1, \dots, j_{m-1}, j_{m+1}), \end{aligned} \quad (7.37)$$

which demonstrates that interchanging the adjacent pair of cards j_m, j_{m+1} does not alter the probability of the hand they are in. By iteration, any reordering of a hand's cards leaves its probability unchanged. This result is so important for the subsequent analysis that we elevate it to the status of a theorem, calling it the Reordering Theorem.

Reducing the number of distinct hands using the Reordering Theorem is complicated, however, by the need to exclude hands not conforming to the rules. An example shows the nontrivial nature of the symmetry involved. The card sequence 6 followed by ace followed by 7 (here denoted $6 + A + 7$) is a hand with value 14, a legitimate Player stand for Dealer upcards 2 through 6. Player always draws a third card to the soft $6 + A$. However, the rearranged sequence $6 + 7 + A$, with the identical value and probability, is not a legitimate hand because Player stands after the $6 + 7$ and does not draw the third card. More generally, an algorithm to generate the irreducible set of distinct, legitimate hands follows the form of (7.1a)–(7.1c). For my computations, however, I manually enumerated the irreducible set and the reordering weight of each.

Another element in the analysis that warrants an extended description stems from the procedure of Dealer peeking to determine whether she holds blackjack. As discussed in Chap. 1, when Dealer's upcard is 10 or ace she immediately checks her downcard (or "peeks") for blackjack, in which event there is no further play of the hand. Only if the downcard is not ace or 10, respectively, does play proceed. Thus, the analysis is predicated, when $u = 10$ or 11 , on the downcard having value $u_2 \neq u^*$, where $u^* \equiv 21 - u$. Thus with the probability of the downcard u_2 prior to peeking being $d^{(4)}(u_2; k_1, u, k_2)$, the conditional probability after peeking becomes $d^{(4)}(u_2; k_1, u, k_2) / C^{(4)}(k_1, u, k_2)$, where

$$C^{(4)}(k_1, u, k_2) \equiv 1 - (\delta(u, 1) + \delta(u, 10)) d^{(4)}(u^*; k_1, u, k_2). \quad (7.38)$$

This result mirrors that of Epstein (1995, p. 224). Note that the $d^{(4)}$ factor in $C^{(4)}$ anchors the downcard, after peeking, as the fourth card dealt in the round.

The computational program described above is both exact and feasible, but needs some further elaboration when pair splitting is considered. In that case, since the precise probability of the second split hand depends on the cards dealt to the first, the number of distinct combined hands becomes roughly the square of the number without splitting; the computational effort to proceed along this line becomes too vast. Unlike those earlier authors who evaluate split pairs by resorting to simulation, we here adopt the stratagem that Player's return from splitting is just twice that of playing the first of the split hands alone. As an example, when the pair $8 + 8$ is split, instead of playing out each 8 in turn followed by Dealer's play, we take the resulting

return to be the same as twice that of playing out only one of the 8s followed directly by Dealer playing her hand.

This stratagem, although it may seem to be an approximation, is actually exact: a consequence of applying the Reordering Theorem. Rigorously, the return from the split pair is the sum of the returns from each of the split hands individually, played in sequence and followed by Dealer's hand. But rather than calculating the return from the first hand as though it were dealt first, followed by the second, followed by Dealer, the Theorem shows the equivalence with dealing to the first hand, then to Dealer's, and then to the second hand. The return from the second hand, similarly, is equivalent with dealing to it first, followed by Dealer's hand and then to the first hand. The sum resulting from the reordering is just the stratagem described. The equivalence further extends to the case of multiple resplits; and also to a small numbers of decks, where one or more card values may become exhausted before the end of the round.

7.2.2 *Expected Return, and Optimal Basic Strategy, vs. Number of Decks*

The outcome of the computations with Optimal Basic Strategy for the expected return from the first round following a shuffle is listed in Table 2.2 of Part I. The four rule possibilities for split pairs are included: with and without DAS, with and without resplits. The returns are also plotted in Fig. 2.1. In each case, the points are closely linear in the inverse number of decks, with a barely perceptible upward quadratic; this functionality is justified from a more general perspective in Sect. 8.1. The linear dependence is fitted in the form $R_D \approx R_\infty + R'/52D$, where the slope parameters R' are also shown in Table 2.2.

Also, shown here in Table 7.1 are results for the variation with number of decks of the probability of tying and of the return's variance and covariance. These were not only given in the previous section in the nondepleting pack approximation, but they can also be computed for any finite number of decks. If Generic Strategy is assumed throughout, then T changes only slightly with D while σ_1^2 and Γ change even less.

Table 7.1 Ties, variance, and covariance vs. number of decks

D	T	σ^2	Γ
∞	0.0982	1.2611	0.4719
8	0.0975	1.2608	0.4715
6	0.0972	1.2607	0.4713
4	0.0968	1.2605	0.4710
2	0.0954	1.2599	0.4700
1	0.0926	1.2587	0.4681
Linear fit	$0.0982 - 0.292/52D$	$1.2611 - 0.125/52D$	$0.4719 - 0.201/52D$

Computation of the covariance – the correlation of returns from two hands played simultaneously against the same Dealer hand – is feasible only with an approximation: that the probability of the second hand reaching any given final count is the same as that of the first hand. Since the covariance depends only slightly on pack size anyway, this approximation likely has little effect on the results.

Optimal Basic Strategy for various numbers of decks is listed in Table 2.3. There are more changes in strategy for smaller D , which contribute slightly to the dependence of the expected return displayed in Table 2.2. But the optimization has been carried out with the Basic Strategy constraints:

- That the parameters for standing are a function only of Dealer's upcard, not the individual identities of Player's first two cards, other than to distinguish pairs from nonpairs and hard from soft
- That the doubling specifications are a function only of Dealer's upcard *and* the combined value of Player's initial two cards, distinguishing hard from soft
- That each split hand is played with the same strategy as an unsplit one

7.2.3 Surrender; Insurance

Surrender is easily analyzed: any two-card hand whose expected return is less than -0.5 should be surrendered, when that option is available. For a pack with more than two decks, this is the case for hard hands with value 16 (other than the $8 + 8$ pair, which should still be split) against upcards 9, 10 and ace; and also 15 against upcard 10. When surrender is available, the overall expected return with a large number of decks improves by 0.0009, a meaningful amount, although as shown in Table 2.2 the improvement declines substantially with decreasing D . Note that surrender is indicated mostly when Dealer's upcard is 10 or ace and she has not drawn blackjack; proper account of this added information is necessary in computing the expected return from Player's two-card hand.

With respect to insurance, Player has the option of side-betting an additional half his primary bet when Dealer shows an ace. The side bet pays off at 2 to 1, independently of all other aspects of the hand, if Dealer's hole card is a 10. Hence Player's expected return from insurance on the reshuffle round is just

$$R_I = d_0(11)(2d_0(10) - (1 - d_0(10))) / 2, \quad (7.39)$$

which evaluates to approximately -0.003 . Since the return is negative, and the bet is not contingent on any other aspect of play, the insurance offer should be declined. However, when a card count is kept, circumstances do arise, addressed in Chap. 10, when insurance develops a positive return and the option might be accepted.

7.3 Play Parameters Dependent on Identities of Initial Cards

In the previous section, the stand parameters were optimized as a function only of Dealer's upcard. Thus, Player stands on any hard hand value $\geq s$ and draws to any value $< s$, independent of the cards making up that hand, and correspondingly s' for soft hands. But he might also use knowledge of the identities of his first two cards. For example, an initial hand value of 13 might be comprised of either $6 + 7$ or $5 + 8$. With upcards such that hitting the $6 + 7$ is optimal, would Player improve his return by instead standing on $5 + 8$? As another example, can the return be increased by doubling $3 + 5$ but not $2 + 6$? Once these questions are answered – yes in both cases – it is then natural to ask by how much.

Thus, now allow s, s' to depend on the specific identities of the initial two cards, not just Dealer's upcard; also, separately optimize the play of each two-card hand formed from the first of a split pair and the next card drawn to it. (A still further generalization, termed “ideal” in Chap. 10, would be to let s and s' depend on *each and every* card in the hand, not just the first two – in other words, to reoptimize the decision of hitting vs. standing each and every time another card is drawn – but the resulting strategy rules would be overwhelmingly complex and with very little improvement in expected return to show for it.)

The results of this “composition-dependent strategy” analysis are presented in Table 7.2, along with the consequent improvement in expected return (roughly independent of the rules for playing split hands) over Optimal Basic Strategy.

The stand profile vs. upcard 10 is truly remarkable! When no account is taken of Player's initial hand value (i.e., in Optimal Basic Strategy) the stand parameter is $s = 17$ for all D , even $D = 1$. Nevertheless, $s = 16$ is in fact optimal for the majority of initial hands, even for as many as eight decks! This seeming contradiction is resolved by discovering that $s = 16$, when optimal, is favored only very slightly over $s = 17$; and that $s = 17$, when favored over 16, does so by a margin one to two orders of magnitude larger. Thus, $s = 17$ emerges from a weighted average over all hands. Nearly 70% of the improvement is reproduced by a simplified rule: stand on every hard hand with value 16 or more except draw to $6 + 10$ (and split $8 + 8$, of course). Also remarkable against upcard 10, when $D = 1$, is the rule to stand on (or preferably surrender, if possible) $7 + 7$.

Similarly, unexpected is the soft stand parameter $s' = 18$ against Dealer ace for hands $2 + 2$, $3 + 2$, and $3 + 3$, for all $D \leq 8$. Although Optimal Basic Strategy calls for stand on soft 18 against ace for two or fewer decks, the persistence of this effect for small two-card hand values up to as many as eight decks is surprising. Nevertheless, the result is quite plausible: those few hands that begin with $2 + 2$, $2 + 3$ or $3 + 3$ vs. ace, and conclude with soft 18, have so depleted the pack in low-valued cards (ace, 2 or 3) that the chance of drawing another such, to reach 19, 20, or 21, is much reduced.

The improvement in expected return resulting from the composition-dependent strategy is insignificant except for $D = 1$, validating the remark in Chap. 2 that the complexity of a composition-dependent strategy might only be justified for a single-deck game. At the other extreme, for a large number of decks the computation

Table 7.2 Composition-dependent play strategy

D	Dealer upcard	Exceptions to Optimal Basic Strategy, Table 2.3	Increase in $R \times 1,000$
8	10	Stand on 16 ^a	0.017
	Ace	Stand on soft 18 for 2 + 2, 3 + 2, 3 + 3	
6	4	Hit 2 + 10	0.028
	10	Stand on 16 ^a ; do not surrender 8 + 7	
	Ace	Stand on soft 18 for 2 + 2, 3 + 2, 3 + 3	
4	2	Double split 2 + 7, 6 + ace, 7 + 2, 7 + ace	0.052
	4	Hit 2 + 10	
	10	Stand on 16 ^a ; do not surrender 8 + 7	
	Ace	Stand on soft 18 for soft $c < 17$, hard $c < 8$	
2	2	Double split 6 + ace, 7 + ace	0.132
	3	Stand on 12 for split 6, 7, 8	
	4	Hit 2 + 10; double split 3 + ace	
	6	Double split 3 + 5	
	10	Stand on 16 ^a ; surrender 6 + 9 and 5 + 10	
	Ace	Stand on soft 19 for soft $c > 16$, hard $c > 17$	
		Do not double 3 + 8, 2 + 9	
1	2	Hit 3 + 10	0.370
	3	Stand on 12 but hit 3 + 9, 2 + 10	
	4	Hit 2 + 10	
	5	Do not double 2 + 6	
	6	Hit 2 + 10; do not double 2 + 6	
	10	Stand on 14 for 7 + 7, surrender if allowed	
		Stand on 15 for 5 + 4 and 5 + 5	
		Stand on 16 ^a ; surrender 6 + 9 and 5 + 10	
	Ace	Stand on 19 with 6+ ace; do not surrender 7 + 9	

^aBut stand on 17 with 9 + 7 and 10 + 6 if surrender is not available; and also with 8 + 2 and all other hands with a 6, 10 + 3 for $D \geq 2$, 7 + 3 for $D \geq 4$, 10 + 2 with $D \geq 6$, and 3 + 3 with $D \geq 8$

demonstrates the absence of any composition dependence: in that case Basic Strategy is also optimal even over the much broader class of strategies based on the individual identities of the first three cards.

As noted earlier, the value of surrender, and the hands that should be surrendered, shrink from that of Chap. 2 as the number of decks decreases and Player's return correspondingly improves; the surrender increment for $D = 1$ is a mere 0.00024. Composition-dependent changes to Chap. 2 surrender guidelines are that 7 + 8 should not be surrendered against upcard 10 for $D < 7$; nor 16-valued hands vs. 9 for $D < 3$; nor 7 + 9 vs. ace for $D = 1$.

7.3.1 *Comparison with Previous Authorities*

A number of analyses have previously been published on exact strategies for the first round after a shuffle with specific numbers of decks. I am aware of those by: Manson et al. (1975), Griffin (1999), Epstein (1995), Revere (1980), Thorp (1966), and Wong (1994); more recently an exhaustive tabulation by Cacarulo has been reported in Schlesinger (2005, Appendix A). A close comparison with these results is worthwhile.

The work of Manson et al. purports to treat exactly the reshuffle round from a four-deck shoe, although in fact they cope with pair splitting by resorting to computer simulation. On the other hand, Wong (1994) obtains strategy results entirely by simulation. Although such a method can be accurate to within narrow statistical limits, it offers few insights into questions other than the specific ones that have been programmed in. Most of the other studies employ a computational approach apparently originated by Julian Braun. Although no study describes its methodology in detail, Griffin (1999, pp. 154 and 172) and Epstein (1995, pp. 223–224) give the fullest outline. A key idea is to compare, for each possible Player hand, the expected return from standing with that from drawing an additional card, and to select the option whose return is larger. Each hand is linked to those with higher values by a recursion similar to that expressed by (7.1)–(7.4).

The approach here, on the other hand, is to compute the expected return weighted over all hands with a given initial value, and to maximize the return with respect to the stand parameters s , s' . (Recall that these are the values at or above which hard or soft hands, respectively, should stand.) But these two criteria can be shown to be identical in outcome: for example, the hard hand value for which standing gives a larger return than hitting, while hands with value lower by one are better off hitting than standing, is just the value s for which the expected return is a local maximum.

Some overlaps among the parameters in these publications permit direct inter-comparisons. In particular, Manson et al. (1975) and Griffin (1999, pp. 173 ff) give composition-dependent strategy results for four decks that agree completely with each other. Griffin (1999) and Wong (1994) also agree on strategy with a single deck. However, as noted earlier, Manson et al. (1975) investigate pair splitting using a simulation technique; Griffin implies that this part of their work is inaccurate and he provides corrected results (1999, pp. 154–157). Various discrepancies among the other earlier studies can be found.

Relative to them, it is widely agreed that Cacarulo's more recent results in Schlesinger (2005) are the most reliable. He gives strategies and expected returns for one, two, four, six, and eight decks, with both "total-dependent (TD)" and "composition-dependent (CD)" hands. However, he uses slightly different definitions of these than we do: his "TD" includes certain multicard decisions that we consider only as CD, such as hit-or-stand on 16 vs. Dealer's 10; and his "CD" excludes some multicard decisions, such as double-or-hit a two-card split hand distinct from if it were not split. Our strategy results agree exactly with Cacarulo's for two-card hit-or-stand decisions, even though he defines away from "CD" some of the multicard rules shown in Table 7.2 and includes others into "TD." As a

further consequence of these definition differences, the expected returns exhibited in Tables 2.2 and 2.4, although within two to three significant figures of his in Table A1 of Schlesinger (2005, p. 394), are slightly smaller than for “TD” and slightly larger than for “CD.”

Chapter 8

Card Counting

8.1 Analytical Framework

The analysis in Chap. 7 of Optimal Basic Strategy considered only the round immediately following a shuffle, so that the first card drawn had value j with likelihood $d_0(j)$. But for subsequent rounds, deeper into the pack, the distribution of the remaining cards varies and so do the likelihoods $d(j)$.

Begin by assuming a pack of D decks, with $v_j \equiv 52Dd_0(j)$ cards of value j , $1 \leq j \leq 10$. Further assume that m_j cards of value j are drawn at random from the pack, M cards altogether. At this point the distribution of cards $\tilde{m}_j \equiv v_j - m_j$ remaining in the pack, with total $\tilde{M} \equiv 52D - M$, is the multinomial (more strictly, the multivariate hypergeometric)

$$p(\tilde{\mathbf{m}}) = \delta\left(\sum_j \tilde{m}_j, \tilde{M}\right) \frac{\tilde{M}!(52D - \tilde{M})!}{(52D)!} \prod_j \frac{v_j!}{\tilde{m}_j!(v_j - \tilde{m}_j)!}. \quad (8.1)$$

The distribution is normalized, in that $\prod_{j=1}^{10} \sum_{\tilde{m}_j=0}^{v_j} p(\tilde{\mathbf{m}}) = 1$. But it is usually easier to work with the asymptotic limit of this distribution for large v_j .

8.1.1 Asymptotic Distribution

The likelihood of drawing value j on the next card is $d(j) = \tilde{m}_j/\tilde{M}$. Use the notation \mathbf{d} for the 10-dimensional vector of likelihoods, and $\rho(\mathbf{d})$ for its distribution. Also denote by $f \equiv M/52D$ the fraction (or “depth”) of the shoe already dealt, and define the quantity $\Delta \equiv (f/52D(1-f))^{1/2}$. Assume the pack has a sufficiently large number of cards of each value (i.e., $4D \gg 1$), that \mathbf{d} can be closely approximated as a continuous variable, permitting differential and integral operations on functions of it. In fact, the results emerging from this continuous variable approach are good approximations even for small numbers of decks. As derived in the Appendix 1, $\rho(\mathbf{d})$ is then given asymptotically by

$$\rho(\mathbf{d}) = \sqrt{2\pi} \Delta \delta \left(\sum_{j=1}^{10} d(j) - 1 \right) \times \prod_{j=1}^{10} \frac{1}{\sqrt{2\pi d_0(j) \Delta}} \exp \left[-\frac{1}{2} \frac{(d(j) - d_0(j))^2}{d_0(j) \Delta^2} \right]. \quad (8.2)$$

Each of the card values is Gaussian (i.e., bell curve) distributed about its mean, and Δ scales the width of the distribution. The width increases with increasing f and with decreasing D . The distribution is normalized, in the sense that $\int d\mathbf{d} \rho(\mathbf{d}) = 1$, with a mean $\langle \mathbf{d} \rangle \equiv \int d\mathbf{d} \mathbf{d} \rho(\mathbf{d}) = \mathbf{d}_0$.

Chapter 3 discussed using a single counting vector α . Generalize here to multiple counting vectors, designated α_λ , where $1 \leq \lambda \leq \Lambda \leq 9$. This generalization extends to the extreme of counting each card value separately, corresponding to $\Lambda = 9$ (the tenth is fixed by $\sum_j d(j) = 1$), which is the maximum amount of information that can be gained from scanning the dealt cards. Multiple counting vectors lead to multiple true counts γ_λ :

$$\gamma_\lambda \equiv \sum_{i=1}^{10} \frac{\alpha_\lambda(i) m(i)}{D - M/52} = \frac{\alpha_\lambda \cdot \mathbf{m}}{D(1 - f)}. \quad (8.3)$$

On the basis of knowledge of the true counts, the distribution of the likelihoods is narrowed; since $\mathbf{d} = (52D \mathbf{d}_0 - \mathbf{m})/(52D - M)$, the distribution of \mathbf{d} conditional on $\boldsymbol{\gamma}$ is

$$\rho(\mathbf{d}|\boldsymbol{\gamma}) = \frac{\rho(\mathbf{d})}{p(\boldsymbol{\gamma})} \prod_{\lambda=1}^{\Lambda} \delta \left(\gamma_\lambda + \alpha_\lambda \cdot \left[52(\mathbf{d} - \mathbf{d}_0) - \frac{M \mathbf{d}_0}{D(1 - f)} \right] \right), \quad (8.4)$$

where δ is the Dirac delta function such that, for any function $f(x)$, $\int f(x) \delta(x) dx = f(0)$; and $p(\boldsymbol{\gamma})$ is the normalization factor ensuring that $\int d\mathbf{d} \rho(\mathbf{d}|\boldsymbol{\gamma}) = 1$ for every $\boldsymbol{\gamma}$. Furthermore, $p(\boldsymbol{\gamma})$ is the probability distribution of $\boldsymbol{\gamma}$, such that $\int d\boldsymbol{\gamma} p(\boldsymbol{\gamma}) \rho(\mathbf{d}|\boldsymbol{\gamma}) = \rho(\mathbf{d})$.

Moments of the conditional distribution can be extracted by substituting (8.2) into (8.4) and taking the integral over \mathbf{d} ; by then replacing each delta function by its integral representation, $\delta(x) = (2\pi)^{-1} \int_{-\infty}^{\infty} dy \exp(ixy)$; and by performing the various Gaussian integrals, beginning with the \mathbf{d} variables, using as a prototype $\int_{-\infty}^{\infty} dy \exp(ixy - y^2/2) = \sqrt{2\pi} \exp(-x^2/2)$. The manipulations are straightforward except when the matrix of inner products

$$A_{\lambda, \lambda'}^2 \equiv (52)^2 \left[\sum_j d_0(j) \alpha_\lambda(j) \alpha_{\lambda'}(j) - \sum_{j,k} d_0(j) d_0(k) \alpha_\lambda(j) \alpha_{\lambda'}(k) \right] \quad (8.5)$$

is encountered. Although in many situations the counting vectors might be orthogonal, so that \mathbf{A}^2 is a diagonal matrix, it is informative to consider the more general

case of nonorthogonal vectors. Then the zeroth moment of the conditional distribution leads to an expression for $p(\boldsymbol{\gamma})$,

$$p(\boldsymbol{\gamma}) = \frac{1}{\text{Det } \mathbf{A}} \left(\frac{1}{\sqrt{2\pi} \Delta} \right)^\Lambda \exp \left\{ -\frac{1}{2} \frac{1}{\Delta^2} \sum_{\lambda, \lambda'} X_\lambda A_{\lambda, \lambda'}^{-2} X_{\lambda'} \right\}, \quad (8.6)$$

where $X_\lambda \equiv \gamma_\lambda - M \boldsymbol{\alpha}_\lambda \cdot \mathbf{d}_0 / D(1 - f)$ and Det is the determinant operation. Expression (8.6) shows that $\boldsymbol{\gamma}$ is also Gaussian distributed, not surprisingly since $\boldsymbol{\gamma}$ and \mathbf{d} are linearly related. Also, the mean of γ_λ is just $\langle \gamma_\lambda \rangle = M \boldsymbol{\alpha}_\lambda \cdot \mathbf{d}_0 / D(1 - f)$, and its second moments are

$$\begin{aligned} \langle \gamma_\lambda \gamma_{\lambda'} \rangle - \langle \gamma_\lambda \rangle \langle \gamma_{\lambda'} \rangle &\equiv \int d\boldsymbol{\gamma} p(\boldsymbol{\gamma}) (\gamma_\lambda - \langle \gamma_\lambda \rangle) (\gamma_{\lambda'} - \langle \gamma_{\lambda'} \rangle). \\ &= \Delta^2 A_{\lambda, \lambda'}^2. \end{aligned} \quad (8.7)$$

The objective is to derive the expected return as a function of $\boldsymbol{\gamma}$, and to optimize it with respect to the counting vectors. To begin, it proves convenient to make a change of variables to a normalized, scale-invariant basis, designating the new variables with a caret: define

$$\begin{aligned} \hat{d}(i) &\equiv (d_0(i))^{1/2}, \\ \hat{\alpha}_\lambda(i) &\equiv 52 \hat{d}(i) \sum_{\lambda'} A_{\lambda, \lambda'}^{-1} (\alpha_{\lambda'}(i) - \boldsymbol{\alpha}_{\lambda'} \cdot \mathbf{d}_0), \\ \hat{\gamma}_\lambda &\equiv \sum_{\lambda'} A_{\lambda, \lambda'}^{-1} (\gamma_{\lambda'} - \langle \gamma_{\lambda'} \rangle). \end{aligned} \quad (8.8)$$

One of the advantages of this “caret” representation is that the $\Lambda + 1$ variables are now orthonormal:

$$|\hat{\mathbf{d}}|^2 = 1, \quad \hat{\mathbf{d}} \cdot \hat{\boldsymbol{\alpha}}_\lambda = 0, \quad \hat{\boldsymbol{\alpha}}_\lambda \cdot \hat{\boldsymbol{\alpha}}_{\lambda'} = \delta(\lambda, \lambda'). \quad (8.9)$$

Note also that $\hat{\gamma}_\lambda$ is invariant against a scale change in the counting vectors $\boldsymbol{\alpha}_\lambda$. Furthermore, it is convenient to introduce both the derivative operator $\hat{\nabla}$, where

$$\hat{\nabla}_i \equiv \hat{d}(i) \left(\frac{\partial}{\partial d_0(i)} - \sum_j d_0(j) \frac{\partial}{\partial d_0(j)} \right). \quad (8.10)$$

is such that $\hat{\mathbf{d}} \cdot \hat{\nabla} = 0$, as well as the tensor $\hat{\mathbf{D}}$, where

$$\hat{D}_{i,j} \equiv \delta(i, j) - \hat{d}(i) \hat{d}(j) - \sum_{\lambda=1}^{\Lambda} \hat{\alpha}_\lambda(i) \hat{\alpha}_\lambda(j). \quad (8.11)$$

$\hat{\mathbf{D}}$ is a projection matrix: it has $\Lambda + 1$ vanishing eigenvalues, corresponding to its eigenvectors $\hat{\mathbf{d}}, \hat{\alpha}_\lambda$, with the remaining eigenvalues equal to 1. When $\Lambda = 9$, these eigenvectors are also complete, satisfying (8.11) with $\hat{\mathbf{D}} = 0$. Eigenmodes and related concepts in matrix algebra are sketched in Appendix 2; see also Shores (2007).

Returning to the conditional distribution, (8.4), the caret notation allows exhibiting its first two moments in an especially simple form:

$$\begin{aligned} \langle d(i) \rangle_\gamma &\equiv \int d\mathbf{d} d(i) \rho(\mathbf{d}|\gamma) \\ &= d_0(i) - \sum_{\lambda=1}^{\Lambda} \hat{\gamma}_\lambda \hat{d}(i) \hat{\alpha}_\lambda(i), \end{aligned} \quad (8.12)$$

and

$$\langle (d(i) - \langle d(i) \rangle_\gamma)(d(j) - \langle d(j) \rangle_\gamma) \rangle_\gamma = \Delta^2 \hat{d}(i) \hat{d}(j) \hat{D}_{i,j}. \quad (8.13)$$

Since the conditional distribution is Gaussian, all of its higher moments can be expressed in terms of these two.

8.1.2 Expected Return; Invariance Theorem

Turn now to Player's return when counting cards. Section 7.1 derived the expected return $R(\mathbf{d})$ from a reshuffle round with likelihoods \mathbf{d}_0 at its start. Here, knowing the true counts γ , the likelihood \mathbf{d} for the first card departs from \mathbf{d}_0 as per (8.12).

The expected return on a hand, at a depth f and conditional on true counts γ , is given by

$$\begin{aligned} \langle R_f(\mathbf{d}) \rangle_\gamma &= \langle \exp [(\mathbf{d} - \langle \mathbf{d} \rangle_\gamma) \cdot \nabla] \rangle_\gamma R_f(\langle \mathbf{d} \rangle_\gamma) \\ &= \exp((1/2) \Delta^2 \hat{\mathbf{D}} : \hat{\nabla} \hat{\nabla}) R_f(\langle \mathbf{d} \rangle_\gamma) \\ &= \exp \left[\frac{1}{2} \Delta^2 \left(\hat{\nabla}^2 - \sum_{\lambda} (\hat{\alpha}_\lambda \cdot \hat{\nabla})^2 \right) - \sum_{\lambda} \hat{\gamma}_\lambda \hat{\alpha}_\lambda \cdot \hat{\nabla} \right] R_f(\mathbf{d}_0), \end{aligned} \quad (8.14)$$

where the second line uses (8.13) and the third uses (8.11) and (8.12). The exponential of derivatives in (8.14) is symbolic. If $R_f(\mathbf{d}_0)$ had a Fourier transform, the exponential operator on it would become a Gaussian integration over the transform. Here, where $R_f(\mathbf{d}_0)$ is only known numerically, the operator is defined in practice through its series expansion. Note that only the linear combinations of card values that are *not* counted enter the diffusion-like exponential factor, as per (8.11), whereas those that *are* counted enter only in $\langle \mathbf{d} \rangle_\gamma$, as per (8.12).

Expression (8.14) can be substantially simplified, in the absence of counting, via the important result $\langle R_f(\mathbf{d}) \rangle = R_0(\mathbf{d}_0)$. This implies that the effect of fluctuations in the composition of a depleting pack exactly cancels the change in its depth from zero, as alluded to in the previous chapter. A significant further consequence is that with no counting the expected return from a round is totally independent of pack depth, a result termed the Invariance Theorem by Thorp (2000).

The proof begins by observing that the expected return is the return from each specific configuration of cards dealt to Player and Dealer, weighted by the probability of that configuration being dealt, and summed over all allowable configurations. Thus, like the approach of Sect. 7.2,

$$R_f(\mathbf{d}) = \sum_{\mathbf{j}} \Xi(\mathbf{j}) \prod_{m=1}^N d^{(m)}(j_m; j_1, \dots, j_{m-1}), \quad (8.15)$$

where $\Xi(\mathbf{j})$ is the return from the configuration $\mathbf{j} = j_1, \dots, j_N$ of an N card hand, and the product $d^{(1)} \dots d^{(N)}$ is the probability of that configuration.

Next, rearrange the cards of the hand, after M cards have previously been dealt, into a canonical form where the n_1 aces are dealt first, the n_2 twos next, etc., so that $\sum_{k=1}^{10} n_k = N$. The hand can be characterized just by the numbers \mathbf{n} rather than by the full set \mathbf{j} . Then by the Reordering Theorem of Sect. 7.2.1,

$$\begin{aligned} P(\mathbf{n}; M) &\equiv \left\langle \prod_{m=1}^N d^{(m)}(j_m; j_1, \dots, j_{m-1}) \right\rangle \\ &= \left\langle \prod_{k=1}^{10} \prod_{\mu_k=0}^{n_k-1} \frac{v_k - m_k - \mu_k}{\tilde{M} - \sum_{l=1}^{k-1} n_l - \mu_k} \right\rangle. \end{aligned} \quad (8.16)$$

For the expectation in (8.16), fold its argument against the exact multinomial distribution (8.1) rather than the asymptotic limit (8.2). Replace the Kronecker delta in (8.1) by its Fourier representation, leading to

$$\begin{aligned} P(\mathbf{n}; M) &= \frac{M! \tilde{M}!}{(52D)!} \int_0^{2\pi} \frac{d\varphi}{2\pi} \exp(-iM\varphi) \\ &\times \prod_{k=1}^{10} \left[\sum_{m_k=0}^{v_k} \left(\prod_{\mu_k=0}^{n_k-1} \frac{v_k - m_k - \mu_k}{\tilde{M} - \sum_{l=1}^{k-1} n_l - \mu_k} \right) \frac{v_k! \exp(im_k\varphi)}{m_k! (v_k - m_k)!} \right] \\ &= \frac{M! (\tilde{M} - N)!}{(52D)!} \int_0^{2\pi} \frac{d\varphi}{2\pi} \exp(-iM\varphi) \\ &\times \prod_{k=1}^{10} \left[\prod_{\mu_k=0}^{n_k-1} \left(v_k - \frac{1}{i} \frac{\partial}{\partial \varphi} - \mu_k \right) \right] (1 + e^{i\varphi})^{v_k}. \end{aligned} \quad (8.17)$$

The multiple derivative operations can be carried out by iteration of the prototype $(a + i \partial/\partial \varphi)(1 + e^{i\varphi})^a = a(1 + e^{i\varphi})^{a-1}$, such that each iteration reduces the parameter a by one. Then straightforward manipulations give

$$\begin{aligned}
 P(\mathbf{n}; M) &= \frac{M!(\tilde{M} - N)!}{(52D)!} \int_0^{2\pi} \frac{d\varphi}{2\pi} \exp(-i M \varphi) \\
 &\quad \times \prod_{k=1}^{10} \left[\frac{v_k!}{(v_k - n_k)!} (1 + e^{i\varphi})^{v_k - n_k} \right] \\
 &= \frac{(52D - N)!}{(52D)!} \prod_{k=1}^{10} \frac{v_k!}{(v_k - n_k)!} \\
 &= P(\mathbf{n}; 0), \quad QED
 \end{aligned} \tag{8.18}$$

Thus, the probability of drawing any hand is independent of the depth at which it is dealt, and so $\langle R_f(\mathbf{d}) \rangle = R_0(\mathbf{d}_0)$ as claimed! Furthermore, since the proof is valid for any function $\Xi(\mathbf{j})$, it is true not just for blackjack but for a wide variety of other card games, baccarat for example. This important result has previously been obtained by Thorp (2000), using an entirely different line of reasoning.

8.1.3 Hermite Series

Substituting this result back into (8.14) gives the simpler expression,

$$\langle R_f(\mathbf{d}) \rangle_{\gamma} = \exp \left[- \sum_{\lambda} \left(\hat{\gamma}_{\lambda} \hat{\alpha}_{\lambda} \cdot \hat{\nabla} + \frac{1}{2} \Delta^2 (\hat{\alpha}_{\lambda} \cdot \hat{\nabla})^2 \right) \right] R_0(\mathbf{d}_0). \tag{8.19}$$

Furthermore, using (8.7) in the caret form $\langle \hat{\gamma}_{\lambda} \hat{\gamma}_{\lambda'} \rangle = \Delta^2 \delta(\lambda, \lambda')$, the exponential in (8.19) can be expressed as a power series in the small parameter Δ^2 ; in the case of a single counting vector, the coefficients are the Hermite polynomials H_n :

$$\langle R_f(\mathbf{d}) \rangle_{\gamma} = \sum_{n=0}^{\infty} \frac{\Delta^n}{n!} H_n \left(- \sum_{\lambda} \frac{\hat{\gamma}_{\lambda}}{\Delta} \hat{\alpha}_{\lambda} \cdot \hat{\nabla} \right) R_0(\mathbf{d}_0). \tag{8.20}$$

These polynomials are mutually orthogonal against the weighting factor $p(\gamma)$, in that for an arbitrary Λ -dimensional vector \mathbf{X} ,

$$\langle H_n(\hat{\gamma} \cdot \mathbf{X}/\Delta) H_m(\hat{\gamma} \cdot \mathbf{X}/\Delta) \rangle = \delta(n, m) n! |\mathbf{X}|^{2n}. \tag{8.21}$$

Also, using multiple integrations by parts on any differentiable function $J(\hat{\gamma})$, they can be shown to satisfy

$$\langle J(\hat{\gamma}) H_n(\hat{\gamma} \cdot \mathbf{X}/\Delta) \rangle = (\mathbf{X} \cdot)^n \langle \partial^n J(\hat{\gamma}) / \partial \hat{\gamma}^n \rangle, \quad (8.22)$$

as well as the derivative relation

$$\partial H_n(\hat{\gamma} \cdot \mathbf{X}/\Delta) / \partial \mathbf{X} = n (\hat{\gamma}/\Delta) H_{n-1}(\hat{\gamma} \cdot \mathbf{X}/\Delta), \quad (8.23)$$

Some low-order examples in the scalar case are $H_0(z) = 1$, $H_1(z) = z$, $H_2(z) = z^2 - 1$, and $H_3(z) = z^3 - 3z$.

8.2 Expected Return at Nonzero Depth

The Hermite series gives a precise way to extend results away from the zero-depth reshuffle round to finite depths. It is particularly valuable in simplifying computations of yield and related functions, to be pursued in subsequent chapters. Note that the expected return when counting is also dependent on depth, even though the Invariance Theorem proves its depth independence without counting.

In the case of a single counting vector, the reshuffle return $R_0(\langle \mathbf{d} \rangle_\gamma)$ can be computed as a function of $\hat{\gamma}$, starting with (7.12) for R_0 and (8.12) for $\langle \mathbf{d} \rangle_\gamma$. Results for expected return over the true count range -12 to $+20$ are shown in Fig. 8.1, for both Optimal Basic play and count-dependent play, six decks, with and without DAS. The count-independent Optimal Basic curve is convex downward, whereas

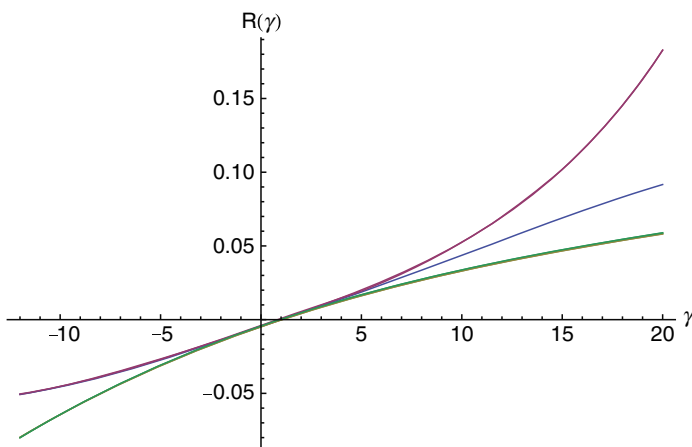


Fig. 8.1 Expected return vs. true count, reshuffle round, six decks: *upper pair of curves* for count-dependent play, the *topmost* with DAS, the *middle* without; the *lowest curve* is for count-independent Optimal Basic play

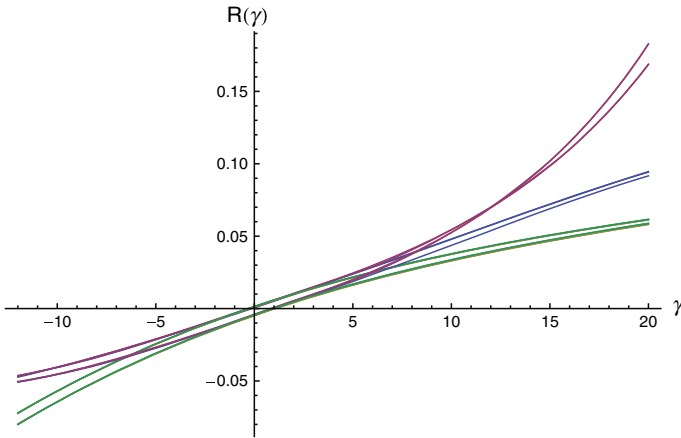


Fig. 8.2 Expected return vs. true count as in Fig. 8.1, but superimposing onto the six-deck results those for single deck as well, the latter set with a slight upward shift vs. the former

Table 8.1 Parameters in polynomial fits to expected return vs. true count, count-dependent (C-D) and Optimal Basic (OB) play strategies

	$10^3 c_0$	$10^3 c_1$	$10^5 c_2$	$10^6 c_3$	$10^7 c_4$
$D = 1, \text{C-D}$	0.747	4.49	3.46	2.53	2.76
$D = 1, \text{OB}$	1.46	4.58	-11.0	1.59	0
$D = 6, \text{C-D}$	-4.69	4.62	4.95	2.75	3.33
$D = 6, \text{OB}$	-4.22	4.75	-11.1	1.59	0

the adjustments of count-dependent play as per Sect. 5.1.2 improve the expected return to become concave upward. Although the curves of Fig. 8.1 are with resplits allowed, the shape is similar without resplits. Figure 8.2 illustrates the similarity of the results for single deck by superimposing them onto those for six decks.

These zero-depth curves can be fitted adequately by a low-order polynomial in true count, $R_0(\langle \mathbf{d} \rangle_\gamma) \cong \sum_n c_n \hat{\gamma}^n$, based on $1/52D$ being a small parameter. A cubic works very well for Optimal Basic play while, in the case of count-dependent play with DAS, accuracy can be improved by adding a quartic term. Figures 8.1 and 8.2 in fact show *both* the data *and* its polynomial fits, a superposition that is undetectable because the differences are much less than the width of the curves! The resulting coefficients c_n are given in Table 8.1. Then a comparable fit to the return at a general depth is just $\langle R_f(\mathbf{d}) \rangle_\gamma \cong \sum_n c_n \Delta^n H_n(\hat{\gamma}/\Delta)$. This simplification is exploited repeatedly in Chaps. 9 and 10.

With the relatively coarse mesh used for plotting Fig. 8.1, all curves appear smooth. In finer detail, though, the count-dependent play curve has subtle kinks arising from the numerous discontinuous changes of the play parameters with true count. This effect is illustrated in Fig. 8.3, comparing the returns for six decks from Optimal Basic and count-dependent play but magnified by plotting the “differential” $[R_0(\langle \mathbf{d} \rangle_\gamma) - R_0(\mathbf{d}_0)]/\gamma$ over a narrow range of small true counts. The kink at small positive true count is due to the change from $s = 17$ to $s = 16$ for upcard 10,

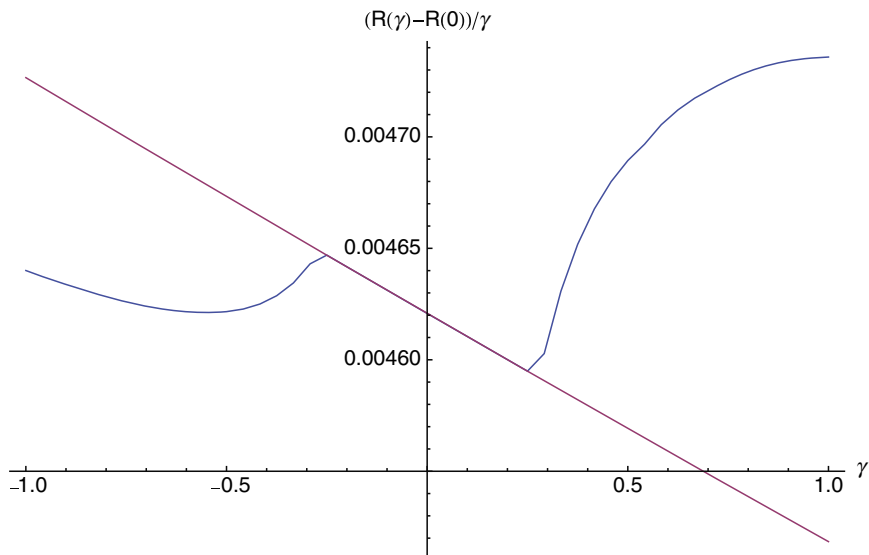


Fig. 8.3 Count-dependent (*kinky curve*) and Optimal Basic (*smooth curve*) first differentials at small values of true count; six decks

while the kink at small negative true count corresponds to the change from $s = 12$ to $s = 13$ for upcard 4; Table 5.2 shows these shifts occurring at comparably small true count values. Similar changes in stand parameter occur for single deck, but there are also a number of less dramatic kinks from shifts in the doubling specifications at even smaller true count values. The visual evidence that the two “differential” curves have the same intercept with the vertical axis confirms that both return functions have the same slope at zero true count.

8.3 Optimizing the Counting Vectors

The task now is to optimize Player’s performance with respect to his choice of counting vectors. But since these are fixed throughout dealing the shoe, the quantity to be maximized is Player’s cash return per round, averaged over the shoe. Call this Player’s “yield” and denote it Y . It is given by averaging the conditional expected return of a hand over the probability distribution of $\hat{\mathbf{y}}$, weighted by the bet size B for that hand, and further averaged over all depths f of the shoe up to the penetration F :

$$\begin{aligned}
 Y &\equiv \frac{1}{F} \int_0^F df \int d\hat{\mathbf{y}} \hat{p}(\hat{\mathbf{y}}) B(\hat{\mathbf{y}}) \langle R_f(\mathbf{d}) \rangle_{\hat{\mathbf{y}}} \\
 &\equiv \left\langle \left\langle B(\hat{\mathbf{y}}) \langle R_f(\mathbf{d}) \rangle_{\hat{\mathbf{y}}} \right\rangle \right\rangle,
 \end{aligned} \tag{8.24}$$

where

$$\hat{p}(\hat{\gamma}) \equiv \left(\frac{1}{\sqrt{2\pi} \Delta} \right)^\Lambda \exp \left(-\frac{|\hat{\gamma}|^2}{2\Delta^2} \right). \quad (8.25)$$

To maximize Y , look for its stationary points, i.e., the zeros of its first derivatives with respect to α_λ . But because Y has been reexpressed entirely in terms of the orthonormal vectors $\hat{\alpha}_\lambda$ which are defined as functions of the α_λ , it suffices to find the stationary points of Y with respect to the $\hat{\alpha}_\lambda$, subject to the orthonormality conditions of (8.9). The constraints are most efficiently handled by use of Lagrange multipliers, here denoted L . The equation for the stationary points is

$$\begin{aligned} 0 &= \frac{\partial}{\partial \hat{\alpha}_\lambda} \left[Y + \sum_\lambda L_\lambda \hat{\alpha}_\lambda \cdot \hat{\mathbf{d}} + \frac{1}{2} \sum_{\lambda, \lambda'} L_{\lambda, \lambda'} (\delta(\lambda, \lambda') - \hat{\alpha}_\lambda \cdot \hat{\alpha}_{\lambda'}) \right] \\ &= \partial Y / \partial \hat{\alpha}_\lambda + L_\lambda \hat{\mathbf{d}} - \sum_{\lambda'} L_{\lambda, \lambda'} \hat{\alpha}_{\lambda'}; \end{aligned} \quad (8.26)$$

while from (8.24), with (8.20) and (8.23),

$$\frac{\partial Y}{\partial \hat{\alpha}_\lambda} = \left\langle \left\langle B(-\hat{\gamma}_\lambda \hat{\nabla} - \Delta^2 \hat{\alpha}_\lambda \cdot \hat{\nabla} \hat{\nabla}) \langle R_f(\mathbf{d}) \rangle_{\hat{\gamma}} \right\rangle \right\rangle. \quad (8.27)$$

Equations (8.26) and (8.27) together (along with (8.9) to eliminate the Lagrange parameters) are sufficient in principle to determine the optimal $\hat{\alpha}_\lambda$. In particular, since $\hat{\mathbf{d}} \cdot \hat{\nabla} = 0$, the inner product of (8.26) with $\hat{\mathbf{d}}$ shows that $L_\lambda = 0$. But since $\langle R_f(\mathbf{d}) \rangle_{\hat{\gamma}}$ is a nonlinear function of $\hat{\alpha}$, no explicit expression for $\hat{\alpha}_\lambda$ is possible in general.

8.4 Optimizing the Counting Vectors: Many-Cards Limit

The analysis to this point has not restricted the size of the pack in any way. But now recognize that the number of cards in the pack is large, so that $\Delta^2 \equiv f/52D(1-f)$ is a small parameter, less than 0.02 for $D \geq 4$ and $f \leq 0.8$. Since the Hermite expansion, (8.19), when substituted in (8.24), represents a de facto series expansion in Δ^2 , the many-card limit corresponds to keeping just the zeroth and first order terms in the expansion. In this case, the yield reduces to just

$$Y = \left\langle \left\langle B(1 - \sum_\lambda \hat{\gamma}_\lambda \hat{\alpha}_\lambda \cdot \hat{\nabla}) R_0(\mathbf{d}_0) \right\rangle \right\rangle, \quad (8.28)$$

so that

$$\frac{\partial Y}{\partial \hat{\alpha}_\lambda} = -\langle \langle B \gamma_\lambda \rangle \rangle \hat{\nabla} R_0(\mathbf{d}_0), \quad (8.29)$$

which can then be substituted into (8.26).

The resulting equation is now linear and can be solved exactly: one counting vector, $\lambda = \lambda^*$, is given by

$$\hat{\alpha}^* \equiv \hat{\alpha}_{\lambda^*} = -\hat{\nabla} R(\mathbf{d}_0) / \left| \hat{\nabla} R(\mathbf{d}_0) \right| \equiv \mathbf{v}_b, \quad (8.30)$$

while the other $\Lambda - 1$ vectors are orthogonal to it. Also, for $\lambda \neq \lambda^*$, the solution requires $\langle \langle B \gamma_\lambda \rangle \rangle = 0$, implying that B cannot depend on any of these other counts. Thus, the optimal counting vector in this many-deck approximation is aligned opposite to the gradient of the expected return function, irrespective of the form of B and for all depths. Also, Y is maximal if and only if the bet size and the conditional expected return are dependent on just the single counting vector \mathbf{v}_b .

In terms of the optimal vector (8.30), the expected return for any other counting vector in the many-card approximation is just

$$\langle R_f(\mathbf{d}) \rangle_\gamma = R_0 + C_b \hat{\gamma} \left| \hat{\nabla} R_0 \right|, \quad (8.31)$$

where C_b is the “betting correlation” between the given counting vector and the optimal one, $C_b \equiv \hat{\alpha} \cdot \mathbf{v}_b$. Clearly $C_b \leq 1$, with $C_b = 1$ for the optimum; the increment to the return from counting is directly proportional to C_b .

With the expected return given by (8.31), the yield expression (8.28) becomes

$$Y = \left\langle \left\langle B \left(R_0 + C_b \hat{\gamma} \left| \hat{\nabla} R_0 \right| \right) \right\rangle \right\rangle \equiv Y_0 + C_b Y_1. \quad (8.32)$$

The first term of (8.32) is clearly the Basic Strategy result (first hand after a shuffle) with a bet equal to $\langle \langle B \rangle \rangle$; as already seen, this term is negative (except for $D = 1$). The second term is the increment from card counting, which is positive when B increases with $\hat{\gamma}$. Besides the factor of C_b , it is also proportional to the *rate* of bet variation: since $\langle \langle B \hat{\gamma} \rangle \rangle = \langle \langle \Delta^2 dB/d\hat{\gamma} \rangle \rangle$ provided $B(\hat{\gamma})$ is a differentiable function, a steep increase in B with $\hat{\gamma}$ enhances this term. Conversely, counting does not improve the yield if the bet size is fixed and not adjusted with the true count. As a consequence, if Dealer shuffles frequently to discourage Player’s card counting, thereby reducing Δ^2 , Player can counterattack only by raising the rate of increase of his bet. This was discussed in Sect. 4.1.

The maximization of yield with respect to the counting vector could also be carried out to the next higher order in $1/52D$; this would indicate, for example, whether the optimal counting vector changes with number of decks. Although the expressions become cumbersome, this change is found to be orthogonal to the lowest order term, so that it makes no contribution to the yield to this order. Thus, any dependence of α^* on number of decks contributes only at order $(1/52D)^2$, and hence is negligible.

Table 8.2 Gradient of return and optimal counting vector

i	$\hat{\nabla}_i R_0$	$v_b(i)$	$\alpha(i) = v_b(i)/\hat{d}(i)$
1	+0.08493	-0.3561	-1.2839
2	-0.05419	+0.2272	+0.8191
3	-0.06202	+0.2600	+0.9376
4	-0.08015	+0.3361	+1.2117
5	-0.10037	+0.4209	+1.5174
6	-0.06452	+0.2705	+0.9753
7	-0.03804	+0.1595	+0.5751
8	+0.00400	-0.0168	-0.0604
9	+0.02747	-0.1152	-0.4153
10	+0.14145	-0.5931	-1.0692

8.5 Computation of the Derivatives of the Expected Return

The optimal counting vector and the maximized expected return and yield depend on the first and second derivatives of R_0 . These cannot be evaluated analytically, but instead require numerical methods based on those of Sect. 8.2: systematically increment the likelihoods \mathbf{d} by small amounts, and extract the first and second finite differences in the expected return. The result for the first derivative is listed in Table 8.2. The (un-normalized) optimal counting vector, in the last column, is just that displayed in Table 3.2; it is consistent with Griffin's result (Griffin 1999, p. 71) for the shifts in return from removing one card from a single deck.

Other useful results are that $|\hat{\nabla} R_0| = .2385$ and $\hat{\nabla}^2 R_0 = -.6527$. Furthermore, it can be shown analytically that $|\hat{\nabla} R_0| = \partial R_0((\mathbf{d})_\gamma)/\partial \gamma|_{\gamma=0}$. The right-hand side of this relationship, from the polynomial fit of Table 8.1, computes to $52 \times .00453 = .2356$, comfortably close to the separately computed result for the left-hand side.

Denote as γ_0 the value of γ for which the return (when $D > 1$) crosses over between negative and positive: to lowest order in Δ^2 ,

$$\langle\langle R_f(\mathbf{d}) \rangle\rangle_\gamma = |R_0|(-1 + \gamma/\gamma_0), \quad \gamma_0 \equiv 52 |R_0| / C_b |\hat{\nabla} R_0|. \quad (8.33)$$

The expected return becomes positive, and to this order increases linearly, as γ exceeds γ_0 . Numerical results for γ_0 are given in Table 3.4.

8.6 Unbalanced Counts

The focus thus far has been on those counting methods, termed “balanced,” for which the count just after a shuffle reverts to zero. An alternative methodology is to use “unbalanced” counting vectors, where the count just after a shuffle and its mean value are nonzero. The so-called Initial Running Count or IRC in the terminology of Vancura and Fuchs (1998) is $52 \mathbf{d}_0 \cdot \boldsymbol{\alpha}(1 - D)$ and the mean running count is

$52 D f \mathbf{d}_0 \cdot \boldsymbol{\alpha} / D(1 - f)$. But the inversion of (8.8) to return from the caret representation to the original shows that any fixed constant can be added onto each and every element of a balanced counting vector, thereby unbalancing it, without changing its correlation C_b or its yield. Thus for each balanced counting vector there exists a continuum of unbalanced ones with identical performance; and, conversely, every unbalanced counting vector performs identically to some balanced one. A reason for considering unbalanced vectors is that one with simple integer elements and a favorable C_b can be chosen, whose balanced counterpart has noninteger elements and so may be impractical to use.

A disadvantage is that immediately following a shuffle the running count for the unbalanced vector (its IRC) is not zero; this quantity would need to be subtracted off from the running count when forming the true count so as to accurately gauge the current expected return. Advocates of unbalanced vectors instead recommend a work-around. They introduce two specific values of the running count: one is the so-called “pivot count,” at which the running and true counts are equal, given by $\gamma_p = 52 \mathbf{d}_0 \cdot \boldsymbol{\alpha}$. The other is the so-called “key count,” at which the expected return is zero. If the cross-over true count is roughly $+1$, independent of the number of decks or the penetration, then some algebra shows that the key count is $\gamma_k = \gamma_p + (1 - \gamma_p)D(1 - f)$. Although strictly the key count is dependent on the depth, mirroring the conversion between running count and true count, the unbalanced count advocates replace f by a constant, typically around 0.6 – 0.75 . (Vancura and Fuchs (1998, p. 75) put forward an empirical key count dependence on deck number of roughly $\gamma_k \approx (20 - 7D)/66$, together with a pivot of $+4$.) Then they recommend linearly interpolating the running count between the pivot and the key to arrive at an approximate expected return, and a bet size.

Two concerns with this procedure can be raised: first, that it does not save any mental effort, since the true count conversion is replaced instead by the running count interpolation; and, second, that the resulting bet size is only approximate (except near the pivot), whereas optimal bet sizing is much more significant for Player’s performance than is a small improvement in expected return via C_b . Although unbalanced counts are of some popularity with serious Players, they cannot be close to optimal.

Appendix 1 Asymptotic Distribution of Card Likelihoods

Equation (8.1), exhibited at the start of this chapter, should be recognizable to those familiar with discrete probabilities; but its asymptotic limit (8.2) may be less familiar, so the derivation is worth outlining. For compactness we use d_j^0 and d_j here rather than $d_0(j)$ and $d(j)$ as in the text body; and we set $d_j \equiv d_j^0(1 + \varepsilon_j)$.

If the number of cards of each value is considered large, $v_j \gg 1$, then each factorial in (8.1) can be replaced by its asymptotic Stirling approximation, $n! \approx \sqrt{2\pi n} (n/e)^n$; Stirling is within 2% of correct even for n as small as 4, and the error diminishes rapidly as n increases. After considerable manipulation, (8.1) can

then be written as the distribution of likelihoods \mathbf{d} ,

$$p(\hat{M}\mathbf{d}) = \delta\left(\sum_j d_j^0 \varepsilon_j, 0\right) \sqrt{\frac{2\pi(52D)f\tilde{f}}{\prod_j \left[2\pi v_j (\tilde{f} + \tilde{f}\varepsilon_j)(f - \tilde{f}\varepsilon_j)\right]}} \quad (8.34)$$

$$\times \exp\left[\sum_j v_j \Phi(\varepsilon_j)\right],$$

with the further notation $\tilde{f} \equiv 1 - f$ and

$$\Phi(\varepsilon) \equiv \ln \left[\frac{f f(\tilde{f}) \tilde{f}}{(\tilde{f} + \tilde{f}\varepsilon) \tilde{f} + \tilde{f}\varepsilon (f - \tilde{f}\varepsilon) f - \tilde{f}\varepsilon} \right]. \quad (8.35)$$

In the asymptotic limit, the otherwise discrete variables \mathbf{d} become quasi-continuous: the expected value of a function of \mathbf{d} , which in the discrete case sums the function over the distribution $p(\hat{M}\mathbf{d})$, becomes an integration over \mathbf{d} . Also, the Kronecker delta converts to a Dirac delta. Furthermore, within that integration the exponential part of (8.34) is mostly very small, since v_j is very large and Φ can be shown to be nonpositive. Thus, it is asymptotically valid for each d_j integral to focus on the integration range near $\varepsilon_j = 0$, where $\Phi(0) = 0$: we replace Φ and the square root in (8.34) by the lowest nonvanishing term in their Taylor series about $\varepsilon_j = 0$. In particular, $\Phi(\varepsilon_j) \approx -\tilde{f} \varepsilon_j^2 / 2f$. The process described here for taking the asymptotic limit is typically called the Method of Stationary Phase.

To complete the derivation, rescaling is needed for the conversion of the ten discrete sums to continuous integrations (over the range extended to $-\infty < \mathbf{d} < +\infty$), and from the Kronecker to Dirac delta functions. Also, bring back the parameter $\Delta \equiv \sqrt{f/52D\tilde{f}}$ and restore the original variables via the reverse substitution $\varepsilon_j = (d_j - d_j^0)/d_j^0$. Then the asymptotic limit of (8.34) becomes just the form (8.2), QED.

Appendix 2 Eigenmodes

In matrix algebra, an $n \times n$ dimensional real symmetric matrix $M_{i,j} = M_{j,i}$ is said to have eigenvectors e_i^μ , where $1 \leq \mu \leq n$, if $\sum_j M_{i,j} e_j^\mu = m_\mu e_i^\mu$ is satisfied for each μ ; m_μ is termed the corresponding eigenvalue. If the eigenvectors are also chosen to be orthonormal, $\sum_{j=1}^n e_j^\mu e_j^\nu = \delta(\mu, \nu)$, then the matrix \mathbf{M} can be represented in terms of its eigenmodes as $M_{i,j} = \sum_{\mu=1}^n m_\mu e_i^\mu e_j^\mu$. For a fuller exposition see Shores (2007).

Chapter 9

Bet Strategies

9.1 Risk and Capitalization

9.1.1 Risk in a Game with Fixed Return

Blackjack is a game in which the expected return on a given round differs from that of the previous round, because the pack composition has changed; as a result of the differing return, the bet should change correspondingly. A proper analysis of risk in such a game must account for these fluctuations. Nevertheless, analysis of a simplified game in which the return and bet stay fixed is a useful warm-up exercise, in terms of both the mathematical tools employed and some characteristics of the results. The reasoning here, as in much of this chapter, follows Werthamer (2005).

Thus, consider playing a model game consisting of a sequence of rounds, on each of which a fixed amount B is bet. Every round has an outcome in which Player wins an amount equal to his bet with probability $\xi(+1)$, loses his bet with probability $\xi(-1)$, and ties (i.e., no money is exchanged) with probability $\xi(0)$. (Ties are included because they occur in blackjack, as seen in Chap. 2, about 10% of the time; but for now overlook the different payoffs from Player blackjack, doubling, and splitting.) Obviously, since these three possibilities exhaust all outcomes, $\sum_{\omega} \xi(\omega) = 1$, where $\omega = \pm 1, 0$. In this game, the expected return is $R = \sum_{\omega} \omega \xi(\omega)$; and the variance of the return is $\sigma^2 = \sum_{\omega} (\omega - R)^2 \xi(\omega) = 1 - \xi(0) - R^2$.

Assume that Player begins this model game with trip capital C_0 , taken for convenience to be an integer multiple of B . The “coverage” is the dimensionless ratio $c_0 \equiv C_0/B$. If Player then plays N rounds where $N < c_0$, he cannot lose all his money even if he loses every round. In that case, it is easy to show (later, from (9.9)) that Player’s expected capital becomes $\langle C_N \rangle = C_0 + NBR$, with variance $\langle C_N^2 \rangle - \langle C_N \rangle^2 = N B^2 \sigma^2$. His expected capital has grown or shrunk, directly proportional to N , depending on whether R is positive or negative.

The more realistic case, $N \geq c_0$, includes situations where Player’s trip capital is not big enough, and/or his bet size is too large, and/or he plays too many rounds, to cover every possible losing streak. He may then, with some nonvanishing probability, lose his entire stake and be forced to stop playing. This event is

traditionally called “ruin.” Deriving how Player’s capital evolves, taking account of the ruin possibility, is mathematically nontrivial.

Begin that analysis with the standard trinomial expression for $p_N(\mathbf{n})$, the probability of N rounds resulting in a total of n_+ wins, n_- losses, and n_0 ties:

$$p_N(\mathbf{n}) = N! \prod_{\omega} (\xi(\omega)^{n_{\omega}} / n_{\omega}!) \quad (9.1)$$

Furthermore, the conditional probability $\hat{p}_N(\mathbf{n})$ of that result occurring without ruin (i.e., with “survival”) is

$$\hat{p}_N(\mathbf{n}) = \left(1 - \prod_{\omega} \frac{n_{\omega}!}{(n_{\omega} + \omega c_0)!} \right) p_N(\mathbf{n}) \quad (9.2)$$

This expression holds for all n_- by interpreting $n_-! / (n_- - c_0)! = 0$ for $n_- < c_0$. Expression (9.2) can be deduced only with difficulty, but it can readily be proved by induction: if true for \hat{p}_N , then it can be shown to be true as well for \hat{p}_{N+1} , due to the basic relationship

$$\hat{p}_{N+1}(\mathbf{n}) = \sum_{\omega} \hat{p}_N(\mathbf{n} - \omega) \xi(\omega) \quad (9.3)$$

More useful for further analysis are the probabilities $\hat{p}_N(C)$ of Player’s capital being C after N rounds, conditional on survival, given by

$$\hat{p}_N(C) \equiv \sum_{\mathbf{n}} \delta \left(N, \sum_{\omega} n_{\omega} \right) \delta \left(C - C_0, B \sum_{\omega} \omega n_{\omega} \right) \hat{p}_N(\mathbf{n}), \quad (9.4)$$

and its ruin-free analog, $p_N(C)$. The former satisfies a recursion like (9.3),

$$\hat{p}_{N+1}(C) = \Theta(C) \sum_{\omega} \hat{p}_N(C - \omega B) \xi(\omega) \quad (9.5)$$

using the step function $\Theta(C) \equiv 1, C > 0; \equiv 0, C \leq 0$. These quantities are convenient when deriving the probability L_N of ruin on the N th round: since ruin occurs when a Player with capital B bets it all and loses, so that $L_N = \hat{p}_{N-1}(B) \xi(-1)$, substitution of (9.1)–(9.4) and some manipulation shows that

$$L_N = \Theta(N - c_0) (c_0 / N) p_N(0) \quad (9.6)$$

Results equivalent to (9.4) are quoted by Epstein (1995, p. 68, Eqs.3–11) and attributed to Lagrange.

Conservation of probability can be confirmed using (9.5): since

$$\sum_C \hat{p}_N(C) = \sum_C \hat{p}_{N-1}(C) - L_N, \quad (9.7)$$

hence iterating downward $N - c_0$ times gives

$$\sum_C \hat{p}_N(C) + \sum_{\mu=c_0}^N L_\mu = 1; \quad (9.8)$$

in words, the probability of survival plus the cumulative probability of ruin is unity. Furthermore, a similar approach shows that the expected capital after N rounds is

$$\begin{aligned} \langle C_N \rangle &= \sum_C C \hat{p}_N(C) = \sum_C C \sum_\omega \hat{p}_{N-1}(C - \omega B) \xi(\omega) \\ &= \sum_{C'} \sum_\omega (C' + \omega B) \hat{p}_{N-1}(C') \xi(\omega) \\ &= \langle C_{N-1} \rangle + B R \left(1 - \sum_{\mu=c_0}^{N-1} L_\mu \right) \\ &= C_0 + B R \left(N - \sum_{\mu=c_0}^N (N - \mu) L_\mu \right). \end{aligned} \quad (9.9)$$

When there is no possibility of ruin, so that $L_\mu = 0$ for all μ , (9.9) proves the assertion regarding the expected capital made earlier.

The results so far have all been exact. But the simplicity of (9.9) is deceptive: after substituting (9.6) and its predecessors (9.4) and (9.1), no further progress can be made toward a simpler closed expression. Instead, resort to taking the limit where the number of wins, losses, and ties becomes large; correspondingly, also regard $N \sim c_0$; the derivation is outlined in Appendix 1. The result is that

$$\begin{aligned} \frac{\langle C_N \rangle - C_0}{N B R} &\approx 1 - \exp(-c_0 \bar{R}) \sum_{\pm} \exp(\pm c_0 \hat{R}) \left(1 - \frac{c_0^2 / \hat{N}}{c_0 \hat{R}} \right) \\ &\quad \times \text{Erf} \left(\frac{c_0^2 / \hat{N} \pm c_0 \hat{R}}{(c_0^2 / \hat{N})^{1/2}} \right), \end{aligned} \quad (9.10)$$

in terms of the error function (sometimes called the complementary normal integral)

$$\text{Erf}(x) \equiv \frac{1}{\sqrt{2\pi}} \int_x^\infty dy \exp\left(-\frac{y^2}{2}\right); \quad (9.11)$$

a direct integration shows that $\text{Erf}(0) = 1/2$, and hence that $\text{Erf}(-\infty) = 1$. The caret variables in (9.10) (distinct from those of (8.7)) are here defined as $\hat{N} \equiv N \sigma^2$, $\hat{R} \equiv R/\sigma^2$, and $\bar{R} \equiv -\hat{R} + \ln((1 + \hat{R})/(1 - \hat{R}))$.

The same analytical techniques can be used to derive the cumulative ruin probability,

$$W_N \equiv \sum_{\mu=c_0}^N L_\mu = \exp(-c_0 \bar{R}) \sum_{\pm} \exp(\pm c_0 \hat{R}) \text{Erf} \left(\frac{c_0^2 / \hat{N} \pm c_0 \hat{R}}{(c_0^2 / \hat{N})^{1/2}} \right) \quad (9.12)$$

as well as a lengthier expression for the variance of the capital. In the very large N limit, $\hat{N} \approx \hat{R}^{-2}$, (9.12) reduces to just $W_N \approx \exp[c_0 \ln((1 - R/\sigma^2)/(1 + R/\sigma^2))]$. The similar but not identical expression, $W_N = \exp[(c_0/\sigma) \ln((1 - R/\sigma)/(1 + R/\sigma))]$, with σ^2 as the exact variance, is quoted without proof both by Carlson (1992, p. 156) and by Schlesinger (2005, p. 112). It appears still earlier as Eq. 9.5 of Sileo (1992), who apparently misuses a model due to Griffin (1999, pp. 141–142). The expression here, however, does agree with Epstein (1995, p. 59). For the remainder of this chapter, assume the return is small, $R \ll 1$, so that this discrepancy disappears and $\bar{R} \cong \hat{R}$.

Comparison of (9.10) and (9.12) shows that $(\langle C_N \rangle - C_0)/NBR$ is just the probability of survival at N but with a correction term that removes random walks into negative capital prior to N . The right-hand side of (9.10), representing the effect of ruin, is straightforward to evaluate numerically and graph, as is (9.12); although the computation is deferred to the end of the next section, the correction term does not have a major qualitative effect.

These results can also be obtained with a different but complementary mathematical approach, one that will prove valuable in the next section. In this approach, introduce the Fourier representation,

$$\delta(m, n) = \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} \exp(i(m - n)\theta), \quad (9.13)$$

for each of the Kronecker delta functions in (9.4). Then the sums over \mathbf{n} can easily be performed, leading to the expression

$$p_N(C) = \frac{B}{2\pi} \int_{-\pi/B}^{\pi/B} d\varphi \exp[-i(C - C_0)\varphi + N \ln v(\varphi)], \quad (9.14)$$

where

$$v(\varphi) \equiv \sum_{\omega} \xi(\omega) \exp(i\omega B\varphi). \quad (9.15)$$

Furthermore, (9.15) can now be generalized to include the differing payoffs of doubles, splits, and blackjack by extending the definition of ω to also include values ± 2 and $3/2$,

$$v(\varphi) \rightarrow \sum_{\omega=0,\pm 1,\pm 2,3/2} \xi(\omega) \exp(i \omega B \varphi). \quad (9.16)$$

The asymptotic limit of $p_N(C)$ for $N \gg c_0$ can be derived by applying the Stationary Phase method to the φ integration. The integrand is asymptotically vanishing except near the zero of $\ln v$, which is at $\varphi = 0$. Thus,

$$\ln v(\varphi) \approx iBR\varphi - B^2\sigma^2\varphi^2/2, \quad (9.17)$$

which in turn, after extending the integration limits to $\varphi = \pm\infty$, leads to

$$p_N(C) \approx (2\pi\hat{N})^{-1/2} \exp\left(-(C - C_0 - NBR)^2 / 2\hat{N}B^2\right); \quad (9.18)$$

the caret variables now are defined to contain the exact variance of the return. Player's capital is Gaussian distributed, with mean $C_0 + NBR$ and width $\hat{N}^{1/2}B$, just as for finite N . Substituting (9.18) back into (9.6) and (9.9) reproduces (9.10) and (9.12).

However, the Fourier representation is unsuitable for the probability $\hat{p}_N(C)$ conditional on survival. For example, (9.10) cannot be reproduced directly without going through (9.6). This obstacle complicates the generalization to nonvanishing risk of ruin. Of value, though, is having found the point of stationary phase in (9.15) to be at $\varphi = 0$.

9.1.2 Optimal Betting When Return Fluctuates: Expected Capital and Risk

The instructive but simplified warm-ups of the previous section facilitate taking on the real game of blackjack, where the return R varies from round to round. Section 8.1 has already demonstrated that R is distributed about a mean, $\langle R \rangle$, with a width that increases with the fraction of the pack dealt out. For a game with multiple decks, $\langle R \rangle$ is negative. That part of the distribution where $R > 0$ offers the opportunity to increase the bet size, so that $B \rightarrow B(R)$. These features are illustrated in Fig. 9.1.

Because each round has a different return and bet size, label the rounds by an index, μ , and generalize the quantities of the previous section to R_μ , B_μ , $v_\mu(\varphi)$. The bet B_μ is some integer multiple of a base bet B_0 . Then the probability distribution $p_N(C)$ is naturally expressed in the Fourier representation, generalizing (9.14), as

$$p_N(C) = \frac{B_0}{2\pi} \int_{-\pi/B_0}^{\pi/B_0} d\varphi \exp\left(-i(C - C_0)\varphi + \sum_{\mu=1}^N \ln v_\mu(\varphi)\right) \quad (9.19)$$

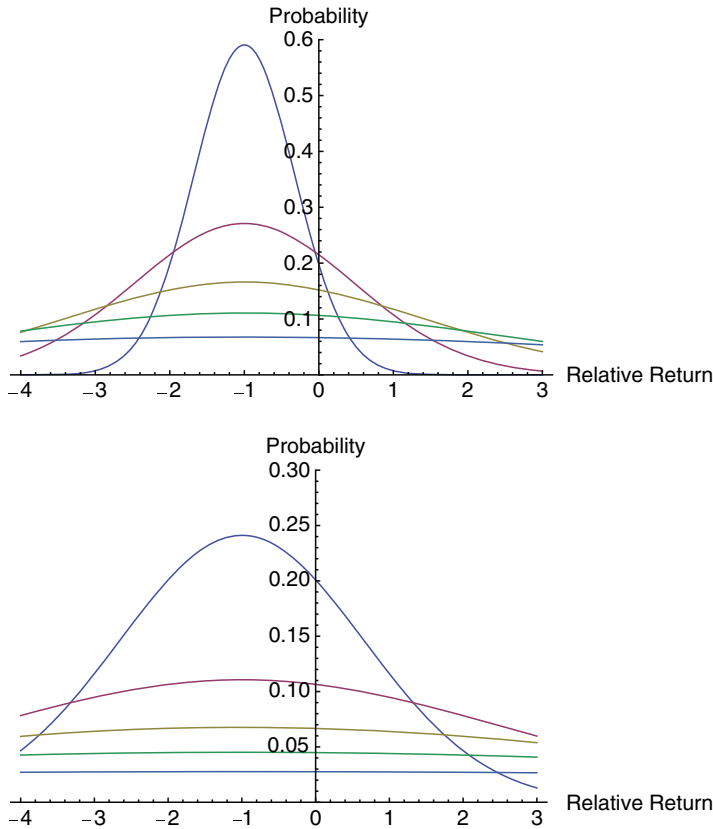


Fig. 9.1 Distribution of return, scaled to its magnitude at zero count, for various depths: *upper* plot for six decks, *lower* plot for one deck. At small depths the distributions are narrowly peaked around -1 , while as the depth grows they spread further out and have increasing weight at positive returns. The spreading is more pronounced for one deck than for six

Once again, temporarily disregard the distinctive payoffs to hands that are doubled, split, or blackjack.

For a sufficiently long sequence of rounds, through multiple reshuffles of the shoe, the sum over μ tends to its ensemble average. Then, in (9.19),

$$\sum_{\mu=1}^N \ln v_{\mu}(\varphi) \approx N \langle \langle \ln v(\varphi) \rangle \rangle \approx \left\langle \left\langle \hat{N} \left(i B \hat{R} \varphi - B^2 \varphi^2 / 2 \right) \right\rangle \right\rangle \quad (9.20)$$

the second equivalence in (9.20) takes the small-argument expansion of $v(\varphi)$, appropriate for the asymptotic, large N limit. The φ integral, just as in the derivation of (9.18), then gives

$$p_N(C) = \frac{B_0}{\sqrt{2\pi \langle \hat{N} B^2 \rangle}} \exp \left[-\frac{(C - C_0 - N \langle \langle BR \rangle \rangle)^2}{2 \langle \hat{N} B^2 \rangle} \right]. \quad (9.21)$$

Comparison of (9.18) and (9.21) shows that when the return and bet size vary, the expected capital after N rounds becomes $\langle C_N \rangle = C_0 + N \langle \langle BR \rangle \rangle$. It grows with N to the extent that $\langle \langle BR \rangle \rangle > 0$, even if $\langle R \rangle < 0$. Note that $\langle \langle BR \rangle \rangle$ is just the (risk-free) yield per round, Y , central to Chaps. 3 and 4. Thus, a desirable goal for bet strategy is to arrive at a functional dependence of B on R for which the yield is positive, so that the expected capital grows rather than shrinks with N .

The analysis of this section has thus far neglected ruin, thus omitting the key source of risk. To include this effect, return to (9.5) and generalize it to

$$\hat{p}_N(C) = \Theta(C) \sum_{\omega} \hat{p}_{N-1}(C - \omega B_N) \xi_N(\omega). \quad (9.22)$$

Also, the ruin probability becomes $L_N = \hat{p}_{N-1}(B_N) \xi_N(-1)$; and $L_N = 0$ if $N < c$, where the coverage c is now defined via

$$C_0 = \sum_{\mu=1}^c B_{\mu} \equiv c \langle \langle B \rangle \rangle. \quad (9.23)$$

A related coverage parameter that will be especially useful later is $\bar{c} \equiv C_0 / \langle \langle B^2 \sigma^2 \rangle \rangle^{1/2}$, the ratio of trip capital to what in Chap. 4 is called the typical bet. Here, “typical” means the root-mean-square total amount wagered on a hand, including the increased initial bets on hands with positive expected return and the additional amounts needed to cover doubles and splits.

The analysis leading to (9.8) and (9.9) proceeds as before: probability is conserved, just as expressed by (9.8), and the expected capital becomes

$$\langle C_N \rangle = C_0 + \sum_{\mu=1}^N B_{\mu} R_{\mu} - \sum_{\mu=1}^N L_{\mu} \sum_{\nu=\mu+1}^N B_{\nu} R_{\nu}. \quad (9.24)$$

For large N , the sum $\sum_{\nu} B_{\nu} R_{\nu}$ includes enough terms that it approaches its ensemble average; furthermore, it is uncorrelated from the outcome of all rounds $\nu \leq \mu$. Hence

$$\langle C_N \rangle = C_0 + \langle \langle BR \rangle \rangle \left(N - \sum_{\mu=c}^N (N - \mu) L_{\mu} \right). \quad (9.25)$$

To proceed further, a generalization of (9.6) is needed; it could then be combined with (9.21) and (9.25) to derive a generalization of (9.10). To accomplish this, first rederive (9.6) by iteration of $L_N = \hat{p}_{N-1}(B) \xi(-1)$, but instead use (9.3) and avoid (9.2). The result, in schematic form, is

$$L_N = \sum_{\text{perms}} \delta \left(-C_0, B \sum_{\omega} \omega \xi(\omega) \right) \prod_{\omega} \xi(\omega)^{n_{\omega}}, \quad (9.26)$$

where the sum is over all permutations of sequences in which the capital goes from C_0 to 0 in N rounds without reaching 0 at any intermediate round. But from (9.6), when the probability factors are the same for every hand,

$$\sum_{\text{perms}} \cdots = \frac{C}{N} N! \sum_n \delta \left(N, \sum_{\omega} n_{\omega} \right) \prod_{\omega} \left(\frac{1}{n_{\omega}!} \right) \cdots \quad (9.27)$$

By introducing the Fourier representation for the delta function, (9.26) can be expressed as

$$L_N = \frac{B}{2\pi} \int_{-\pi B}^{\pi B} d\varphi \exp(i C_0 \varphi) \sum_{\text{perms}} \prod_{\omega} (\xi(\omega) \exp(i \omega B \varphi))^{n_{\omega}}. \quad (9.28)$$

When the returns and bets vary from round to round, the form (9.28) for L_N generalizes to

$$L_N = \frac{B_0}{2\pi} \int_{-\pi/B_0}^{\pi/B_0} d\varphi \exp(i C_0 \varphi) \sum_{\text{perms}} \prod_{\omega} \prod_{\mu_{\omega}} \xi_{\mu_{\omega}}(\omega) \exp(i \omega B_{\mu_{\omega}} \varphi), \quad (9.29)$$

where the products over the μ_{ω} respectively contain n_{ω} terms. Again, as for (9.17), each product contains enough terms, when N is large, to tend to its ensemble average and is uncorrelated from the outcomes of the other sequences. Then

$$\begin{aligned} L_N &\approx \frac{B_0}{2\pi} \int_{-\pi/B_0}^{\pi/B_0} d\varphi \exp(i C_0 \varphi) \sum_{\text{perms}} \prod_{\omega} \langle \xi(\omega) \exp(i \omega B \varphi) \rangle^{n_{\omega}} \\ &\approx \frac{B_0}{2\pi} \frac{C}{N} \int_{-\pi/B_0}^{\pi/B_0} d\varphi \exp(i C_0 \varphi + N \ln \langle \langle v(\varphi) \rangle \rangle). \end{aligned} \quad (9.30)$$

The second line follows by using (9.27) and the argument leading to (9.14).

The last step in the derivation points out that

$$\begin{aligned} N \ln \langle \langle v(\varphi) \rangle \rangle &\approx \left\langle \left\langle \hat{N} \left(i B \hat{R} \varphi - B^2 \varphi^2 / 2 \right) \right\rangle \right\rangle \\ &\approx N \langle \langle \ln v(\varphi) \rangle \rangle \end{aligned} \quad (9.31)$$

in the asymptotic limit, using $|B\varphi| \ll 1$ and the variance of the yield \ll the mean square bet size. Thus (9.30) and (9.31), compared with (9.19)–(9.21), show again

that $L_N \approx \Theta(N - c)(c/N) p_N(0)$ as anticipated. Substituting into (9.25) leads to a form for $\langle C_N \rangle$ just like (9.10), but with the replacements $BR \rightarrow \langle \langle BR \rangle \rangle$, $c_0^2/\hat{N} \rightarrow C_0^2/N \langle \langle B^2 \sigma^2 \rangle \rangle = \bar{c}^2/N$, and $c_0 \hat{R} \rightarrow \langle \langle BR \rangle \rangle C_0 / \langle \langle B^2 \sigma^2 \rangle \rangle \equiv q$: explicitly,

$$\begin{aligned} \frac{\langle C_N \rangle - C_0}{N \langle \langle BR \rangle \rangle} &= 1 - \exp(-q) \sum_{\pm} \exp(\pm q) \left(1 \pm \frac{\bar{c}^2/N}{q} \right) \\ &\quad \times \text{Erf} \left(\frac{(\bar{c}^2/N) \pm q}{(\bar{c}^2/N)^{1/2}} \right) \\ &\equiv \Psi_N. \end{aligned} \quad (9.32)$$

Ψ_N , of course, is just the yield reduction factor (YRF) discussed in Sect. 4.1. Adding plausibility to this result for $\langle C_N \rangle$ is its close parallel to the generalization that takes, in the absence of ruin, the distribution of C from (9.18) for fixed betting to (9.21) for variable betting. Equation (9.12), for the cumulative ruin probability, similarly generalizes to

$$W_N \rightarrow \exp(-q) \sum_{\pm} \exp(\pm q) \text{Erf} \left(\frac{(\bar{c}^2/N) \pm q}{(\bar{c}^2/N)^{1/2}} \right). \quad (9.33)$$

Schlesinger (2005, p. 132) displays, without proof but attributed (private communication) to Chris Cummings, an expression equivalent to (9.33).

9.1.3 Connections with Finance

Although expression (9.33) is relatively unfamiliar in blackjack analysis, it is similar to a celebrated milestone in the mathematics of finance: the Black–Scholes formula for the rational price of an option on an asset, such as a stock, whose price fluctuates. This result, derived by Fischer Black and Myron Scholes (1973), and by Robert Merton (1973), earned the Nobel Prize in Economics for Scholes and Merton (Black had died before the prize was awarded). Finding such a close connection between a game like blackjack on the one hand, where the outcome of each round is statistically independent of all previous rounds (i.e., a Markov stochastic process), and the stock market on the other requires the observation, which Black, Scholes, and Merton (BSM) adopted from earlier work of Samuelson (1965), that fluctuations of the logarithm of stock prices approximately obey Markov statistics.

BSM take a different mathematical path to their result than here, where the rounds of play were first regarded as discrete, arriving at expressions (9.2) and (9.9), and then the asymptotic, or continuum, limit was taken to reach (9.10) and (9.12). BSM instead assume the continuum limit at the start of their derivation, appeal to the stochastic calculus of Ito (1951), and obtain their formula as the solution of a drifting diffusion equation. Their approach is plausible since a Markov process

generates a random walk, whose continuum limit is diffusion. The connection with the approach here can be confirmed directly by verifying that the expected capital, (9.32), satisfies the standard diffusion equation,

$$\left(\frac{\partial}{\partial N} - \langle \langle BR \rangle \rangle \frac{\partial}{\partial C_0} - \frac{1}{2} \langle \langle B^2 \sigma^2 \rangle \rangle \frac{\partial^2}{\partial C_0^2} \right) \langle C_N \rangle = 0 \quad (9.34)$$

– as does the cumulative ruin probability, (9.33), as well – with diffusion constant $\langle \langle B^2 \sigma^2 \rangle \rangle$ and drift, or bias, $\langle \langle BR \rangle \rangle$; the number of rounds is a proxy for time. Skipping the discrete model in favor of the continuum could have led to (9.32), for instance, just by solving the diffusion equation with the conditions $\langle C_{N \rightarrow 0} \rangle \rightarrow C_0$, $\langle C_N \rangle = 0$ for $C_0 = 0$ and $N > 0$. However, since the approach here only requires the Stirling approximation and a steepest descents integration, (9.32)–(9.33) can be conjectured (see below) to already be close approximations for $N \approx 50$, not nearly so large as the true continuum limit.

An approach to deriving the Black–Scholes result based on discrete steps, as here, was later taken by Cox, Ross, and Rubenstein (Cox et al. 1979, CRR). Their rationale was to circumvent the sophisticated continuum mathematics of the stochastic calculus. They start with a simple binary model in which equity prices fluctuate either up or down by a fixed amount (like here, but without ties), such that the log of the price conforms to a Markov process. Then they take the limit of increasingly short time steps and small fluctuation amplitudes. Since this asymptotic limit, applied in their case to the value of an option, involves only the mean and variance of the Markov process, they are able to argue that the result would also hold for any more complex process, such as the actual stock market, provided only that the mean and variance of the model are matched exactly with those of the market. Their argument is an alternative way to derive a diffusion equation like (9.34) whose coefficients are averages over the actual process – blackjack in our application, equity markets in theirs. The CRR approach thus provides a direct proof of (9.32), alternative to the argument given here. Furthermore, subsequent simulations with the CRR model, e.g., Hull (2002), have demonstrated that the continuum limit of the discrete step model is approached quite closely in well under a 100 steps, bolstering the conjecture above.

9.1.4 Properties of the Risk and Expected Capital Expressions

Having confirmed that (9.32)–(9.33) are the appropriate generalization of (9.10) and (9.12) for varying returns and bets, they can now be analyzed in more detail. The expected capital and ruin probability depend on the two parameters, N/\bar{c}^2 and q , each of which can range from 0 to ∞ ; the expressions also depend on the sign of $Y = \langle \langle BR \rangle \rangle$, although here a bet scheme with positive yield is assumed. Numerical evaluations of (9.32)–(9.33) are straightforward. The survival probability, $1 - W_N$, and the YRF, Ψ_N , are plotted in Fig. 9.2 on a log–log scale. Although the survival

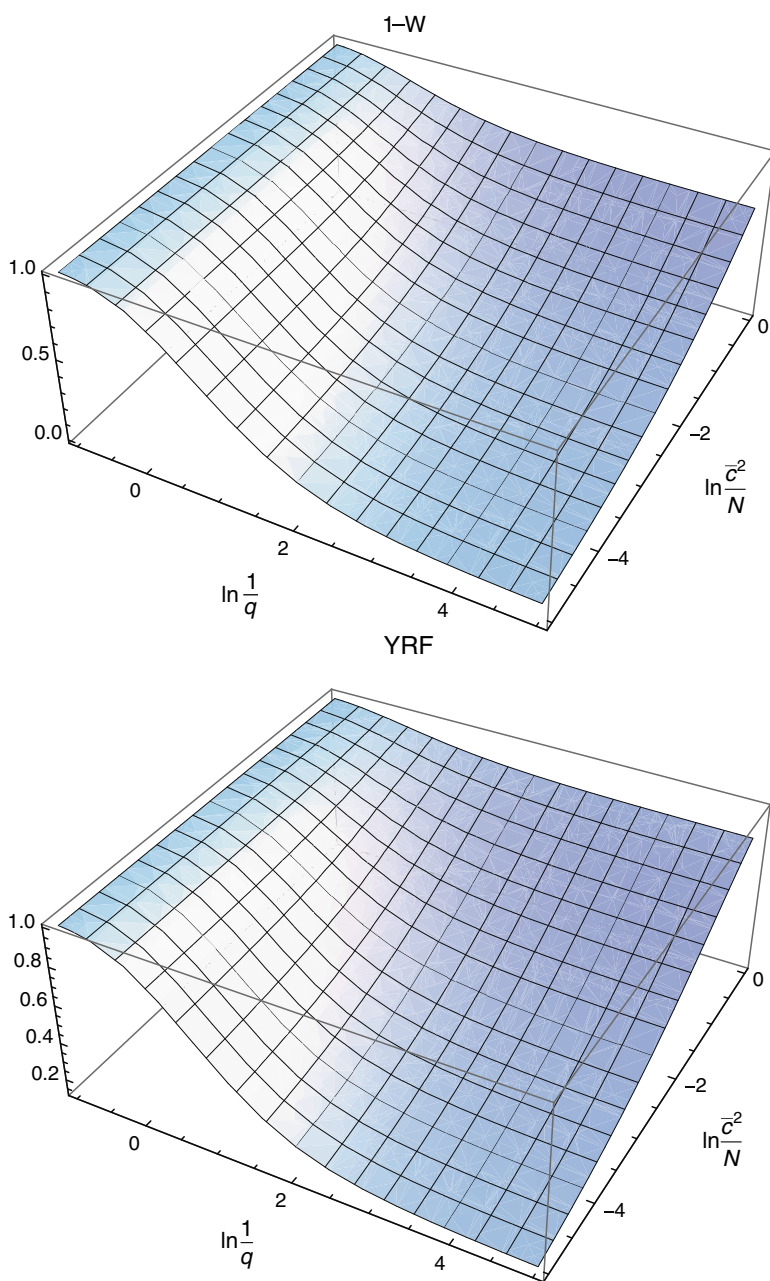


Fig. 9.2 Survival probability (a) and yield reduction factor (b) as functions of the parameters q and N , log-log scale

probability is not precisely identical to the YRF, the two are qualitatively similar and differ by at most a factor of two.

Figure 9.2 illustrates that the YRF and survival probability vary monotonically with q (i.e., always increase as q increases, and vice versa), so q is called the “ruin protection” in Sect. 4.1. When q is large enough (e.g., comparable to or larger than 1), ruin is very unlikely and the YRF approaches unity for all but extraordinarily large values of N/\bar{c}^2 (as shown in the left-hand portion of the plot). But for q significantly less than one, the YRF decreases steeply with increasing N/\bar{c}^2 (right-hand portion), especially after many rounds such that $N \gg \bar{c}^2$. The conclusion is that ruin can be a decided risk if the coverage is insufficient.

A way to understand these results more intuitively is to look back at (9.21) for the probability distribution of capital, C . The peak of the distribution drifts linearly with increasing N to larger positive values, while the width increases as $N^{1/2}$. Ruin becomes likely to the extent that the distribution has any significant weight at $C = 0$, i.e., when $(C_0 + N \langle \langle BR \rangle \rangle)^2 < N \langle \langle B^2 \sigma^2 \rangle \rangle$. This condition, in turn, is satisfied if the two conditions $q < \bar{c}/N^{1/2} < 1$ both hold, much like the conditions (low ruin protection and low capital coverage) of the previous paragraph (the foreground of Fig. 9.2). Ruin occurs when the distribution, despite the steady drift of its peak to more positive values, has diffused outward to overlap $C = 0$ (see Fig. 9.3). Ruin protection affects the relation of the drift to the broadening; large ruin protection q , as well as large coverage \bar{c} , keeps the overlap exponentially small and forestalls ruin.

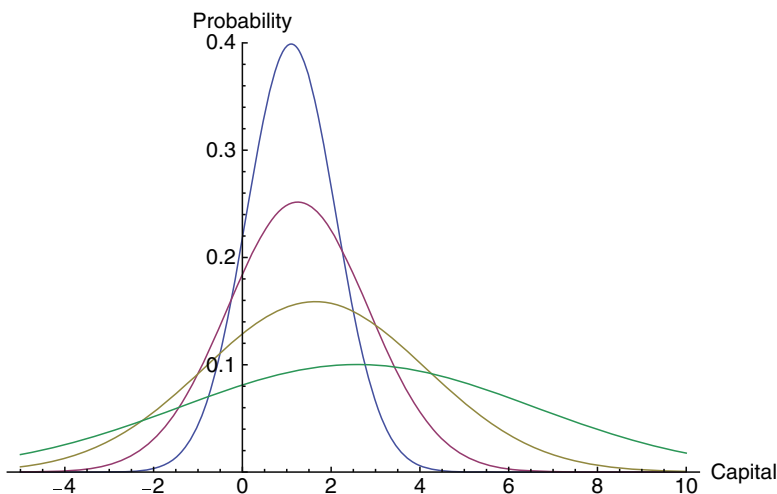


Fig. 9.3 Representation of the distribution of capital, for several values of the number of rounds N . As N increases, the distribution shifts, widens, and flattens: its mean and its variance both increase

9.1.5 Optimal Betting When Return Fluctuates: Bet Strategy

Having derived the evolution of the expected capital with the relative number of rounds played, dependent also on the ruin protection, the bet strategy can now be optimized. The criterion for optimization invoked here is to find the bet function that maximizes the expected capital for a given number of rounds, yet fixes the cumulative ruin probability at a predetermined value. Microeconomists might instead prefer the introduction and maximization of a Utility Function (see, e.g., Ingersoll (1987), especially Chap. 1), but such a criterion is primarily a formal one, difficult to apply for any practical results.

Thus, functionally maximize $\langle C_N \rangle$ – or more conveniently an “effective yield” $y_N \equiv (\langle C_N \rangle - C_0)/NC_0 = Y \Psi_N/C_0$ – with respect to $B(\gamma)$ while holding W_N constant.

However, $B(\gamma)$ is also constrained to lie within the range between the table minimum B_- and maximum B_+ as set by house rules. If Player elects, for whatever reason, to cap his bet size below the table maximum, then B_+ instead represents that cap. Similarly, he might also choose his base bet, B_- , above the table minimum, as long as $B_- \leq B_+$. In any event,

$$\begin{aligned} \langle \langle BR \rangle \rangle = & \left(B_- \int_{-\infty}^{\gamma_-} + B_+ \int_{\gamma_+}^{\infty} \right) d\gamma \rho(\gamma) R(\gamma) \\ & + \int_{\gamma_-}^{\gamma_+} d\gamma \rho(\gamma) B(\gamma) R(\gamma) \end{aligned} \quad (9.35a)$$

and a similar expression for $\langle \langle B^2 \sigma^2 \rangle \rangle$, referenced later as (9.35b). Since $B(\gamma)$ is intuitively expected to increase monotonically with γ , define γ_{\pm} via $B(\gamma_{\pm}) = B_{\pm}$. The “spread,” then, is $\beta \equiv B_+/B_-$.

The maximization of y_N with respect to $B(\gamma)$ in the range $\gamma_- \leq \gamma \leq \gamma_+$ – with the constraint of fixed W_N best handled by use of a Lagrange multiplier, denoted L – leads to the functional stationary equation $\delta(y_N - L W_N)/\delta B(\gamma) = 0$. The differentiation at first looks complicated; but the complexity can be circumvented by recognizing, from (9.32) and (9.33), that the expression depends on $B(\gamma)$ only via $\langle \langle BR \rangle \rangle$, whose functional derivative is proportional to $R(\gamma)$, and on $\langle \langle B^2 \sigma^2 \rangle \rangle$, whose functional derivative is proportional to $B(\gamma) \sigma^2$. Thus, the result of differentiation is just a linear combination of these two simple expressions, and so the stationary equation has the solution $B(\gamma) \propto R(\gamma)/\sigma^2$. As expected, the bet is proportional to R between a lower and upper threshold. Define a constant of proportionality s such that $b(\gamma) \equiv B(\gamma)/C_0 = s R(\gamma)/\sigma^2$ in this range. Then the threshold true counts satisfy $R(\gamma_{\pm}) = b_{\pm} \sigma_{\pm}^2/s$. A bet function of this form is consistent with an existence theorem asserted by Epstein (1995, p. 67), and referenced by Griffin (1998,

p. 139). Substitution of this functional form for $b(\gamma)$ into (9.35a) and (9.35b) leads to the expressions

$$\langle\langle bR \rangle\rangle = A_1 + sA_2, \quad \langle\langle b^2\sigma^2 \rangle\rangle = A_0 + s^2A_2, \quad (9.36)$$

where the A coefficients are defined as weighted integrals over γ .

Although the expected return in general is a nonlinear function of true count – Sect. 8.2 has shown it can be closely approximated by a cubic – for purposes of this section we adopt the rough approximation that the function is linear, and hence is independent of depth. In this event, the distribution of returns is Gaussian just like that of true counts. Then we can convert the $A_{0,1,2}$ quantities into the integrals over R ,

$$\begin{aligned} A_0 &\equiv \left(b_-^2 \int_{-\infty}^{R_-} + b_+^2 \int_{R_+}^{\infty} \right) dR \sigma^2 \rho(R), \\ A_1 &\equiv \left(b_- \int_{-\infty}^{R_-} + b_+ \int_{R_+}^{\infty} \right) dR R \rho(R), \\ A_2 &\equiv \int_{R_-}^{R_+} dR \left(\frac{R}{\sigma} \right)^2 \rho(R), \end{aligned} \quad (9.37)$$

with weight $\rho(R)$ which incorporates the average of that Gaussian distribution over all depths f of the pack up to the reshuffle penetration F :

$$\rho(R) = \frac{1}{F} \int_0^F df \frac{1}{\sqrt{2\pi}\Delta_R} \exp \left[-\frac{1}{2} \left(\frac{R - R_0}{\Delta_R} \right)^2 \right], \quad \Delta_R \equiv 52\Delta R_0. \quad (9.38a)$$

The slope parameter s is set by the specified level of risk, W_N , which in turn is parameterized by q and \bar{c} . In principle, it would be desirable to invert that relationship to find s as a function of W_N , and substitute into y_N . But the inversion is not possible in closed form, instead requiring computation. Needing to specify the two bet threshold parameters b_{\pm} creates additional complexity for a numerical approach.

To gain some partial insights, begin the computational work by choosing the representative set of parameters for a blackjack game adopted in Chap. 4: six decks with penetration at 80% of the shoe before reshuffle; Generic Strategy expected return of -0.005 , with variance of 1.26 independent of R ; and zero expected return at a true count of $+1$. With these parameters, the distribution of returns becomes

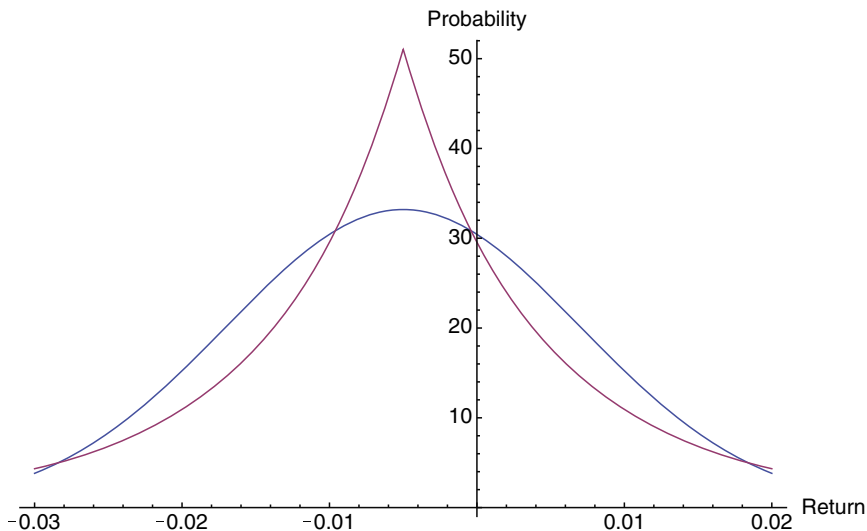


Fig. 9.4 Distributions of expected returns: at depth 0.4 (*Gaussian curve*) and averaged over depths to a penetration 0.8 (*peaked curve*), for representative game parameters

$$\rho(R) = \frac{1}{0.8} \int_0^{0.8} df \frac{1}{\sqrt{2\pi}\Delta_R} \exp\left(-\frac{1}{2} \left(\frac{R + .005}{\Delta_R}\right)^2\right),$$

$$\Delta_R^2 = \frac{52 (.005)^2 f}{6(1-f)}. \quad (9.38b)$$

Figure 9.4 graphs $\rho(R)$, along with its integrand at $f = 0.4$, showing how the averaging over depth distorts the shape away from a Gaussian. Selecting values for b_{\pm} and N permits computation of both y_N and W_N vs. the slope s . An example is shown in Fig. 4.2 with the Lifetimer parameters, $b_- = 0.001$, $b_+ = 0.010$, and $N = 10^6$. Each curve has an extremum: y_N has a shallow maximum while, at a smaller value of slope, W_N has a deep minimum. This pattern is a general characteristic over a wide range of parameters, although the W_N minimum becomes quite shallow for smaller N . Thus, slopes that are either greater than at the maximum, or less than at the minimum, correspond to both lesser yield and greater risk than at the respective extrema; only slopes between these two points correspond to solutions of maximal yield.

With these insights, the slope can be eliminated and y_N obtained directly as a function of W_N , as plotted in Fig. 4.3; the Weekender example, $b_- = 0.01$, $b_+ = 0.10$, $N = 10^3$, is also shown. The latter might be typical of a weekend casino trip, with trip capital of a 100 base bets; the former of a “lifetime” of extensive play, with trip capital of a 1,000 base bets. Also note that the effective yield may in fact be negative for sufficiently low risks, although always greater than that without counting.

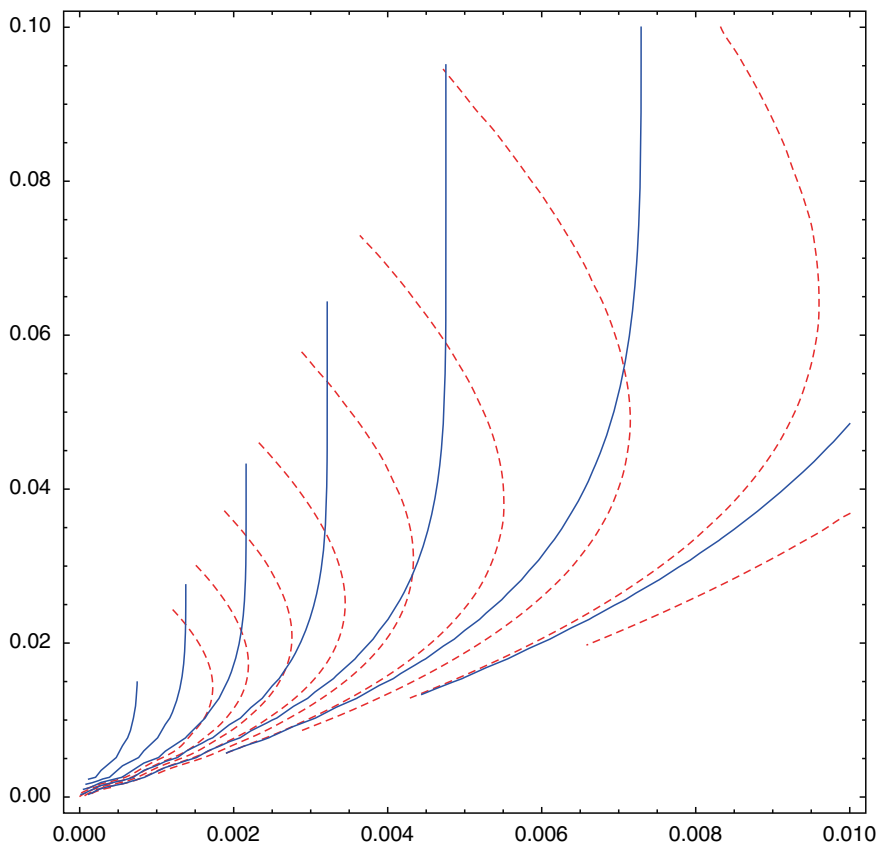


Fig. 9.5 Contours of minimum risk (*solid curves*) and maximum risk (*dashed curves*) vs. b_{\pm}

Details of the effective yield vs. risk curves vary with the choice of parameters, although the family of curves retains the general shape shown here. Displaying these variations in full detail is difficult, since the parameters form a three-dimensional family. Nevertheless, some insights can be gained from contour plots over the two-dimensional b_{\pm} plane. First, Fig. 9.5 shows the end points, of greatest and least risk, for the Lifetimer example. Then, Fig. 9.6 shows the contours of effective yield, with the risk chosen at the so-called Kelly value, $W_N = w_K \equiv e^{-2} = .1353$. These contours are bounded by the curves along which w_K intersects the maximum and minimum risk surfaces.

So far, the upper and lower bet bounds are regarded as chosen independently: the lower bound is held at the table minimum with the upper bound as large as practicable. Indeed, Fig. 9.6 shows for a fixed upper bound that y_N decreases as the lower bound is raised; and that for fixed lower bound y_N increases with increasing upper bound. But an emphasis on camouflage and on maintaining a predefined spread suggests instead fixing the *ratio*, β , of upper to lower bounds. Then the geometric

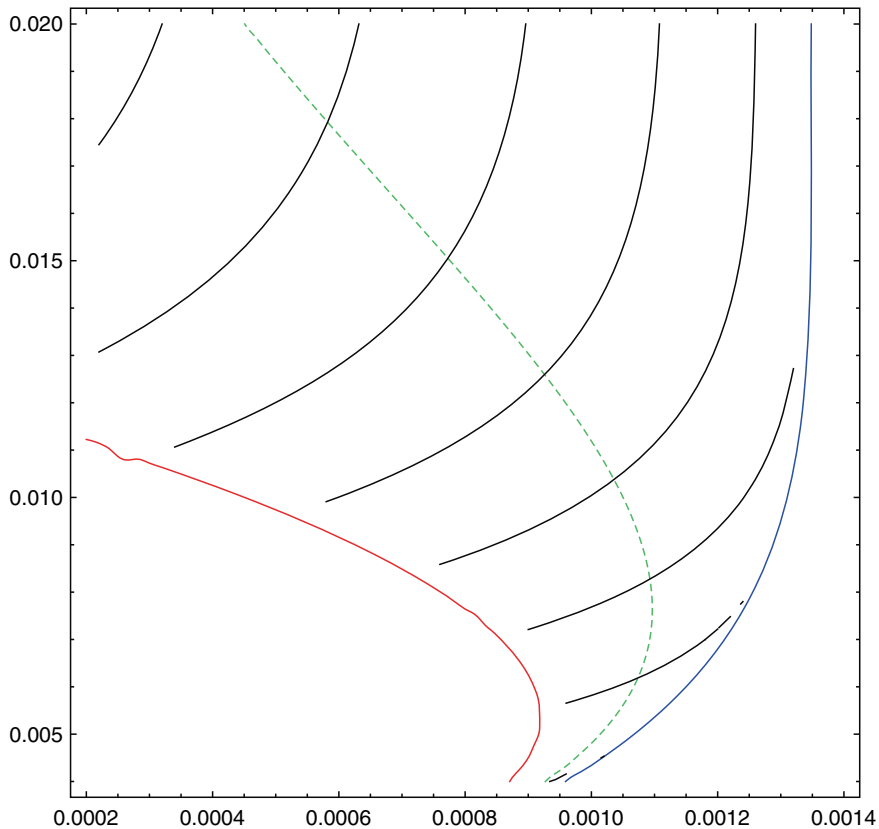


Fig. 9.6 Contours of effective yield vs. b_{\pm} , bounded between curves of minimum risk (on the right) and maximum risk (on the left). The curve of HJY points is *dashed*.

mean of the bounds could be chosen so as to maximize y_N , even if the resulting lower bound exceeds the table minimum. The maximum of y_N under these circumstances occurs for the remarkably simple condition $sq = 1$, which implicitly determines s . Even more remarkably, the maximal condition is independent of the functional forms of y_N and W_N and so, in particular, is independent of N !

This condition was first emphasized by Harris, Janecek, and Yamashita (1997, HJY). But HJY arrive at it by adopting quite different optimization criteria than here. Harris (1997), following Sileo (1992) and Schlesinger (2005), instead posits that a betting optimum maximizes the dimensionless ratio $\langle\langle BR \rangle\rangle^2 / \langle\langle B^2 \sigma^2 \rangle\rangle$ at fixed spread (actually, yet equivalently, Harris (1997) minimizes the inverse ratio); risk is not a determinant in his criterion. The square root of this ratio is just the ROI (the yield per unit of typical bet size) from Sects. 4.1 and 6.1; maximal return on investment is a natural criterion for a betting optimum. Computed solutions of the HJY condition, also plotted in Fig. 9.6, are a curve through the contours of effective yield. The fact that a ray through the origin of Fig. 9.6 (a line of constant spread) is

tangent to a contour at the HJY point provides a visible geometric verification that the point maximizes effective yield for that spread.

But Fig. 9.6 also shows that a continuum of optimal bet patterns, depending on risk, exists in addition to the HJY point. They range from the smallest risk and effective yield, which might be the choice of a casual, lightly capitalized player over a short session, up to the largest risk and effective yield, which could be the choice of a dedicated, well-capitalized player tolerant of occasional ruined sessions to gain greater long-term winnings. But neither end of the continuum is robust: at the lower end, a small increase in risk produces a much larger increase in effective yield; and at the upper end a small decrease in effective yield produces a much larger decrease in risk. A better balance of effective yield and risk occurs at intermediate points.

Since Player would ideally like to both minimize risk and maximize effective yield, one way to make this tradeoff is to maximize the ratio y_N / W_N , the effective yield per unit of risk. Although difficult to pursue analytically, the maximization is straightforward computationally and leads to a well-defined intermediate point, distinct from that of HJY; call it the “minmax,” or “MM,” point. Whereas the HJY point is at a rather high effective yield as well as risk, the MM point typically occurs at much lower values of both. Thus, the former may be an appropriate choice for the dedicated player, the Lifetimer example of Chap. 4, while the latter seems more suitable for the casual player, the Weekender example.

9.1.6 Yield When the Bet Size Is Discrete; Wong Benchmark Betting

The optimal bet strategy derived in the previous section calls for a bet size that is a ramp function of the expected return. Clearly, however, bets in actual practice must be sized as a discrete multiple (an integer, or half-integer if the casino has chips of that denomination) of the minimum bet. Thus, the ramp of (9.35) must be approximated by a staircase, much like that represented in Fig. 4.1.

Although staircases with various numbers of steps and sizes can be used, their results were asserted earlier to be quite close to that of the corresponding ramp. To model this, select the Benchmark game and a bet spread of 10, and then replace the optimal ramp by a coarse staircase with just two intermediate steps: at 4 and at 7 base bets. Computational results of the performance of the staircase vs. the ramp, for both the Lifetimer (HJY point) and the Weekender (MM point), are displayed in Table 9.1 (the ramp results are the same as in Table 4.1). Comparison demonstrates that both yield and risk are degraded by only a few percent. Other staircases, subdivided more finely, perform even more closely to the ramp.

A betting scheme recommended by Wong (1994, p. 18) and shown in Fig. 9.7 is also a staircase, but it lowers the bet another step for negative true counts and so is suboptimal. Furthermore, Wong specifies the slope of his staircase independently of

Table 9.1 Performance of a coarse staircase vs. a ramp

Player	Bet shape	Steepness	Risk of ruin	Effective yield ratio
Lifetimer (HJY)	Ramp	0.94	0.1185	+0.01716
	Staircase	0.94	0.1213	+0.01713
Weekender (MM)	Ramp	2.80	0.1360	+0.00617
	Staircase	2.80	0.1404	+0.00603

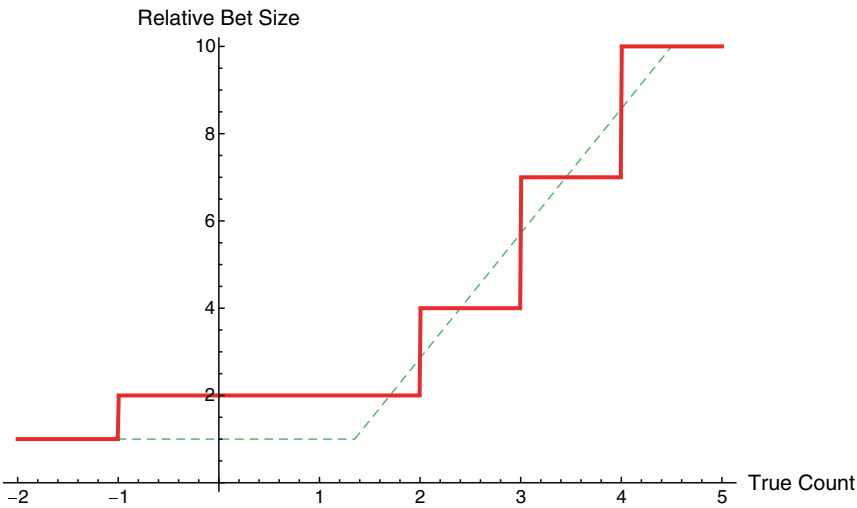


Fig. 9.7 Wong’s “Benchmark” bet staircase (*solid*) and ramp modification (*dashed*)

both the penetration and Player’s risk profile. Figure 9.7 also shows a ramp function (dashed) that gives a good representation of his staircase.

Wong applies this betting scheme in a computer simulation of his Benchmark game. He quotes a yield ratio of 0.016 and a ratio of average to minimum bet of 2.65. Wong does not exhibit a risk result, possibly because his simulation may assume infinite initial capital so that ruin is never possible. Current commercial simulation programs, however, do permit specification of an initial capital and generate ruin probabilities.

Performance measures for Wong’s bet function computed in the Benchmark game are shown in Table 9.2.

The yield ratio is 0.017, with an average bet of 2.70 units, not far from the values he quotes; thus, the simulation and analysis approaches give results reasonably close to each other. Also, with a typical bet of 4.2 units the return on investment is around 0.004. However, modifying Wong’s function so that the minimum is bet for true counts less than +2, rather than below 0 as he does, improves both yield and ROI: a hand with negative expected return should optimally not be bet at above the minimum. In addition, Wong’s yields fall sharply as the penetration is reduced, becoming negative for penetrations less than 0.5; these low penetrations need an even steeper staircase to sustain a positive yield in such unfavorable circumstances.

Table 9.2 Performance of Wong bet function

Betting style	Yield ratio	ROI	Effective yield	Risk
Wong	0.0173	0.0041		
Wong (modified)	0.0177	0.0045		
Lifetimer			0.016	0.10
Weekender			0.014	0.37

The effective yield and risk, with the ramp shown in Fig. 9.7, for both the Lifetimer and the Weekender, are also listed in Table 9.2. The effective yield is below that of the yield ratio because the former averages over sessions ending in ruin, whose risk is particularly high for the Weekender. The ramp is quite steep; it gives performance close to that of the HJY point: fine for the Lifetimer but probably too aggressive for the Weekender.

9.2 Betting Proportional to Current Capital

9.2.1 One Hand per Round

Now consider bet sizes scaled to the current capital, as first suggested by Kelly (1956). Thus, set $B_N = \varphi C_N$ at the N th round and apply a criterion for choosing the scaling factor φ , independent of N . In this case, the capital changes multiplicatively rather than additively: $C_{N+1} = C_N (1 + \varphi \Omega_{N+1})$, depending on the stochastic process Ω whose values give the amount Player wins on a hand. For a simplified game with just a single expected return, $R = \langle \Omega \rangle > 0$ independent of N , it is easy to show that the expected, or mean, value of C_N/C_0 then increases exponentially with N :

$$\langle C_N \rangle / C_0 = (1 + R\varphi)^N, \quad (9.39)$$

at a rate per round

$$\frac{1}{N} \ln \frac{\langle C_N \rangle}{C_0} = \ln(1 + R\varphi). \quad (9.40)$$

Since this rate is positive and increases monotonically with φ , it is maximized by choosing $\varphi = 1$. (When $R < 0$, the rate is negative and is minimized by $\varphi = 0$: do not bet on a losing game!) But this maximum is pathological, because it implies betting one's entire capital at every round. Only a single sequence of outcomes is successful: winning all N rounds in a row so that the capital grows dramatically to $C_N/C_0 = 2^N$, but with the tiny probability $((1 + R)/2)^N$. Every other sequence loses, sooner or later, with $C_N = 0$; so the mean is given by (9.40). Thus, the mean is not a useful measure of the capital distribution from which to select φ in any sensible way.

More promising is to examine (following Maslov et al. (1998), and also discussed by Ethier (2004)) the *median* of the distribution, labeled C_N^* and defined by

$$\left\langle \Theta \left(C_N - C_N^* \right) \right\rangle = 1 / 2. \quad (9.41)$$

To determine C_N^* , differentiate (9.41) with respect to φ , so that

$$\frac{\partial C_N^*}{\partial \varphi} = N C_N^* \frac{\partial}{\partial \varphi} \langle \ln (1 + \varphi \Omega) \rangle; \quad (9.42)$$

reintegrating gives

$$\frac{1}{N} \ln \frac{C_N^*}{C_0} = \langle \ln (1 + \varphi \Omega) \rangle. \quad (9.43)$$

Thus, the median capital also varies exponentially with N ; but now the rate per round, the right-hand side of (9.43), has a vanishing first derivative at an intermediate value of φ . Since its second derivative there is negative definite, the value corresponds to a maximum. For $R^2 \ll 1$ the maximum is found at $\varphi \ll 1$ and so, by expanding (9.43) to second order in φ and R , is given by $\varphi_{\text{opt}} \approx R/\sigma^2 > 0$. The median capital, evaluated at that maximum, grows as $(C_N^*/C_0)_{\text{opt}} \approx \exp(N R^2/2\sigma^2)$.

The approach of Kelly (and of most of the subsequent literature) to selecting φ is instead described as “maximizing the expectation of the rate of capital growth.” Although that terminology is loose mathematically – growth cannot be ascribed to a stochastic variable, only to some measure over its distribution – the usual derivation leads to the same results as here. Our criterion, though, of maximizing the (exponential) growth rate of the median of the distribution, seems more precise and better motivated.

Note, though, that the growth rate of the median turns negative (i.e., the median shrinks) for fractions larger than about $2R$, despite the mean being positive; the rate heads toward $-\infty$ as the fraction tends toward one. The larger bets increase the risk of losing streaks eating into Player’s capital; and as $\varphi \rightarrow 1$ Player is ruined very quickly with almost 100% probability. Also notice that the exponential character of the growth is discernible only for $N > \langle \langle \Theta(R) R^2/\sigma^2 \rangle \rangle^{-1}$, which might be 10^4 or more; for fewer rounds, the increase still looks linear.

So-called Kelly betting (there is little agreement on terminology in the literature: other labels sometimes used include scaled, proportional, multiplicative, and geometric) seems to have the remarkable advantage that Player is never ruined, in the sense of losing all his money. If he experiences a losing streak, so that his capital shrinks, he is directed merely to reduce his bet sizes accordingly and to continue play. But a bet size that varies continuously toward zero is clearly not allowed in a casino: bets can only be integer multiples of a nonzero table minimum, B_- . Rather, “ruin” should be interpreted as capital so low that the prescribed fractional bet cannot be placed.

Once ruin is reconsidered in this way, analysis of optimal Kelly betting can proceed in analogy with Sect. 9.1. Instead of wins/losses being additive/subtractive to capital, here it's to the *log* of capital. And instead of ruin occurring when $C = 0$, here it's when $C = C_- \equiv B_-/\varphi$. With these analogies, a drifting diffusion equation like (9.34) also applies here, but after modification to the drift coefficients, the replacement of the C_0 variable by $\ln C_0$, and the generalized ruin boundary condition. Thus, the risk satisfies

$$\left[\frac{\partial}{\partial N} - V \frac{\partial}{\partial \ln C_0} - \frac{1}{2} \sigma^2 \frac{\partial^2}{\partial (\ln C_0)^2} \right] W_N = 0, \quad (9.44)$$

with drift $V \equiv \langle \ln(1 + \varphi\Omega) \rangle$; the same differential equation determines the median capital. The appropriate solutions, at least in the limit of very large N , become

$$W = \exp\left(-2 \frac{V}{\sigma^2} \ln \frac{C_0}{C_-}\right), \quad \frac{1}{N} \ln \frac{C_N^*}{C_0} = V(1 - W). \quad (9.45)$$

Again maximizing $\ln(C_N^*/C_0)$ with respect to φ , but now subject to a fixed risk, leads to the same optimal Kelly fraction as before. The maximal rate, however, is reduced by the survival factor. With V evaluated at the maximum, the risk becomes

$$W_{\text{opt}} = \exp\left[-\frac{R^2}{\sigma^4} \ln \frac{C_0 R}{B_- \sigma^2}\right] \quad (9.46)$$

which, under typical conditions, is extremely small.

Blackjack, however, is a compound game with a varying expected return dependent on the count. In this case, R becomes a random variable, as in Sect. 9.1. A functional maximization of C_N^* with respect to $\varphi(R)$, with $R > R_-$, still gives $\varphi(R)_{\text{opt}} = R/\sigma^2$ as before; the maximum median capital becomes $C_N^* \cong C_0 \exp(N \langle R^2/\sigma^2 \rangle / 22)$. When R is below a positive threshold the session is terminated in ruin, as discussed above.

Another frequently voiced criticism of Kelly betting is that it leads to wide swings in capital, even though the median is growing at a maximal rate. To reduce the size of the fluctuations about the median, some authors (and a number of practitioners) resort to so-called “fractional Kelly”; here the bet is sized at less than the optimal fraction of the current capital (i.e., “full Kelly”), perhaps half or even a quarter of the optimum.

Just as the median (or equivalently the expected logarithm) of the capital can easily be derived, so the variance of log capital is also readily obtained:

$$\text{var} \equiv \left\langle \left(\ln \frac{C}{C_0} \right)^2 \right\rangle - \left\langle \ln \frac{C}{C_0} \right\rangle^2 \cong N \sigma^2 \left(\frac{1}{2} \ln \frac{1 + \varphi}{1 - \varphi} \right)^2 \quad (9.47)$$

which rises monotonically with φ from the origin, initially as φ^2 . Thus, a lower bet fraction than the optimum does reduce the variance. At the same time, a measure of return on investment, such as $\langle \ln C / C_0 \rangle / \sqrt{\text{var}}$, increases slowly with decreasing φ ; but it does not exhibit a peak, so there is no optimal fraction of Kelly based on such a measure. Kelly bettors are forced to use their own judgment in selecting a reduced Kelly fraction.

On the other hand, criticism of full Kelly based on the size of its capital fluctuations almost always overlooks the concept of ruin in Kelly betting. Should a wide downward swing in capital occur, the player might encounter the ruin threshold and simply start over with a replenished stake, rather than continue to play with a tiny one and even tinier bets. Recovering from a situation of sharply diminished (but not ruined) capital usually takes a dismayingly long playing time, despite the exponentially increasing median.

9.2.2 Multiple Simultaneous Hands

A technique that combats high Kelly variance, while at the same time enhancing yield, involves multiple simultaneous hands. Thus, consider playing H seats simultaneously at the same table, with a capital redistribution technique: initially, Player splits his total capital evenly among the seats, creating subpools, and at each seat bets a uniform fraction φ of the subpool capital. After this and every subsequent round, Player then repools his capital, again divides it evenly (or as evenly as feasible) among the seats, and repeats the betting process. In effect, the scheme uses the winning seats to replenish the losing ones. A scheme of this kind has been proposed for financial investing by Maslov and Zhang (1998).

Let the capital of the h th seat at the end of the n th round be C_n^h , so that the redistributed capital per seat is

$$\overline{C}_n = \frac{1}{H} \sum_{h=1}^H C_n^h. \quad (9.48)$$

Then, at that seat, the capital following the next round becomes

$$C_{n+1}^h = (1 + \varphi_H \Omega_{n+1}^h) \overline{C}_n. \quad (9.49)$$

The stochastic process Ω_n^h is uncorrelated from other rounds but, as seen in Sect. 7.1, has a nonvanishing covariance with the other hands on the same round. Thus, after N rounds the redistributed capital per seat has evolved to

$$\overline{C}_N / \overline{C}_0 = \prod_{n=1}^N (1 + \varphi_H \overline{\Omega}_n), \quad (9.50)$$

where $\overline{\Omega}_n \equiv H^{-1} \sum_{h=1}^H \Omega_n^h$. This is again a multiplicative process, like that for a single seat, and the mean of the redistributed capital is

$$\langle \overline{C}_N \rangle = \overline{C}_0 (1 + R\varphi_H)^N. \quad (9.51)$$

On the other hand, the median, \overline{C}_N^* , satisfies

$$\frac{1}{N} \ln \frac{\overline{C}_N^*}{\overline{C}_0} = \frac{1}{N} \sum_{n=1}^N \langle \ln (1 + \varphi_H \overline{\Omega}_n) \rangle. \quad (9.52)$$

Anticipating that the median rate's maximum occurs for $\varphi \ll 1$, $R \ll 1$, expand in this limit:

$$\begin{aligned} \frac{1}{N} \ln \frac{\overline{C}_N^*}{\overline{C}_0} &= \frac{1}{N} \sum_{n=1}^N \left(\varphi_H \langle \overline{\Omega}_n \rangle - \frac{1}{2} \varphi_H^2 \langle \overline{\Omega}_n^2 \rangle + \dots \right) \\ &= \varphi_H R - \frac{1}{2} \varphi_H^2 [\sigma^2 + (H-1)\Gamma]/H + \dots \end{aligned} \quad (9.53)$$

Thus, the maximum rate now occurs for the proportionality fraction $\varphi_H = H R / (\sigma^2 + (H-1)\Gamma) \approx H R / (1.26 + .47(H-1))$, where the effect of the higher order terms in (9.53) can be shown to be negligible. At this fraction,

$$\frac{1}{N} \ln \frac{\overline{C}_N^*}{\overline{C}_0} = \frac{H R^2 / 2}{\sigma^2 + (H-1)\Gamma}. \quad (9.54)$$

Remarkably, both the maximal rate of growth of the median capital and the betting fraction that achieves it have increased as a result of playing H hands simultaneously and redistributing capital among them, as compared with Kelly betting with the total capital at a single seat. As seen in (9.53), the single-hand variance is reduced by a factor of H , which in turn boosts the growth rate, although the effect is partially offset by the covariance. Thus, in the context of Kelly betting, playing multiple hands and systematically redistributing capital improve performance vs. that of just a single hand with the same total capital.

The log capital variance in this scheme, for small but otherwise arbitrary values of φ_H , is $\text{var} = N [\sigma^2 + (H-1)\Gamma] \varphi_H^2 / H$; it too would increase, proportionally, if the optimal bet fraction of the previous paragraph were employed. But since typical numerical values are $\Gamma \approx 0.47 < \sigma^2 \approx 1.26$, the coefficient of φ_H^2 in the variance expression decreases with increasing H . As a result, if Player maintains the same bet fraction as if playing only a single hand, namely $\varphi_1 = R / \sigma^2$, rather than the larger value from the optimality criterion, then playing several hands would both reduce the variance as well as raise the median's rate of increase! In this technique, the median capital and the variance become

$$\ln \left(\frac{\overline{C}_N^*}{\overline{C}_0} \right) = \frac{N R^2}{2\sigma^2} \left[1 + \frac{H-1}{H} \left(1 - \frac{\Gamma}{\sigma^2} \right) \right],$$

$$\text{var} = \frac{NR^2}{\sigma^2} \left[1 - \frac{H-1}{H} \left(1 - \frac{\Gamma}{\sigma^2} \right) \right].$$

Even with only two seats, the median rate is raised by more than 30% while the variance is reduced by the same percentage; with still more simultaneous hands the factors can exceed 50%.

9.3 Back-Counting and Table-Hopping

Section 4.3 has described the playing technique of table-hopping and its variant, back-counting. These maneuvers offer a significant improvement in yield, quantified in Table 4.5. Here the analysis underlying the table (following Werthamer (2006) and (2008)) is described for entry, exit, and departure, singly and in combination.

9.3.1 Entry

Begin with certain assumptions about Player and the table he is at:

1. The game is played with more than one deck, $D > 1$, so that the expected return R_0 for the first round after a shuffle is negative and has variance σ^2 .
2. The cards are reshuffled after a shoe penetration F .
3. Player is tracking the dealt cards using one of the usual balanced counting methods, such that the true count immediately after a shuffle is $\gamma = 0$ and the expected return becomes positive, $R > 0$, for true counts greater than a cross-over, $\gamma_0 > 0$.
4. Entry is allowed between any successive rounds, not just at shuffles.

The back-counter enters the game only when the true count first reaches the threshold $\gamma_E \geq \gamma_0$. Thus, we need the probability of this occurring, conditional on the true count being zero immediately following a shuffle. But this probability is mathematically congruent to the probability of ruin, i.e., of the player's capital C first reaching zero from its initial value of C_0 . Equation 9.6 has shown that the probability of ruin first occurring at round N , with a bet B per round, is

$$L_N = (C_0/NB) p_N(0), \quad (9.55)$$

where (9.18)

$$p_N(C) \equiv \frac{1}{\sqrt{2\pi N\sigma^2}} \exp \left[-\frac{(C - C_0 - NBR)^2}{2NB^2\sigma^2} \right] \quad (9.56)$$

is the distribution of capital after N rounds, in the absence of ruin. Furthermore, L_N can be shown to satisfy the drifting diffusion equation (like (9.34))

$$\left(\frac{\partial}{\partial N} - BR \frac{\partial}{\partial C_0} - \frac{1}{2} B^2 \sigma^2 \frac{\partial^2}{\partial C_0^2} \right) L_N = 0. \quad (9.57)$$

In parallel, the distribution of true counts was shown ((8.24) and (8.7), with $\Lambda = 1$) to be

$$\rho(\gamma, \tau) = \frac{1}{\sqrt{2\pi\tau}} \exp\left(-\frac{\gamma^2}{2\tau}\right), \quad (9.58)$$

contingent on the initial condition of zero true count at zero penetration. As a modification from the notation of Chap. 8, the parameter $\tau \equiv 52f/D(1-f) = (52\Delta)^2$ is used; the extra factor of 52 appears because the true count is defined as running count *per deck* undealt. We call τ “time” because the distribution can be verified to satisfy the diffusion equation

$$\left(\frac{\partial}{\partial \tau} - \frac{1}{2} \frac{\partial^2}{\partial \gamma^2}\right) \rho(\gamma, \tau) = 0, \quad (9.59)$$

where τ plays the mathematical role that time does in the physical diffusion process. (The fact that τ is a nonlinear function of depth, which *does* increase proportionally to real time, reflects the character of the underlying random card-dealing process: the true count, which is always zero just after a shuffle and then random walks, must return to zero if/when all cards in the shoe are dealt. Such a constrained process is sometimes called a “Brownian bridge.”)

Thus, there is a close analogy in their evolution between the distributions of capital and true count (compare (9.57) and (9.59)), which gives the probability of the true count first reaching γ_E at “time” τ_E as the analog of (9.55),

$$\rho_1(\gamma_E, \tau_E) = (\gamma_E/\tau_E) \rho(\gamma_E, \tau_E). \quad (9.60)$$

Furthermore, the distribution $\rho_E(\gamma, \tau)$ of true count γ at any “time” τ prior to entry must still satisfy (9.59) but now with the boundary condition $\rho_E(\gamma_E, \tau) = 0$. This boundary condition is the analog of that for “survival,” from Sect. 9.1; it is satisfied by including a “reflection” about γ_E :

$$\rho_E(\gamma, \tau) = (\rho(\gamma, \tau) - \rho(2\gamma_E - \gamma, \tau)) \Theta(\gamma_E - \gamma), \quad (9.61)$$

where Θ is again the unit step function. Note that any solution to the diffusion equation (and many other differential equations) is unique if it also satisfies the initial/boundary conditions.

This pair of expressions is further justified by showing that they satisfy conservation of probability. Thus, the total probability of *having* entered at any τ_E prior to τ is, from (9.60),

$$\tilde{P}_E(\tau) = \int_0^\tau d\tau_E \rho_1(\gamma_E, \tau_E) = 2 \operatorname{Erf}\left(\frac{\gamma_E}{\sqrt{\tau}}\right), \quad (9.62)$$

whereas the total probability of *not* having entered prior to τ is, from (9.61),

$$P_E(\tau) = \int_{-\infty}^{\infty} d\gamma \rho_E(\gamma, \tau) = 1 - 2 \operatorname{Erf} \left(\frac{\gamma_E}{\sqrt{\tau}} \right). \quad (9.63)$$

These indeed sum to unity, as required by conservation of probability.

In addition, the two expressions are also linked by the iterative “chain rule,”

$$\frac{1}{\tau_E} \int_0^{\tau_E} d\tau \int_{-\infty}^{\infty} d\gamma \rho_1(\gamma_E - \gamma, \tau_E - \tau) \rho_E(\gamma, \tau) = \rho_1(\gamma_E, \tau_E); \quad (9.64)$$

the integration steps needed for its proof also serve as a prototype for similar convolutions in succeeding sections, so are described in Appendix 2.

Next, given that the true count γ_E is first reached at τ_E , the subsequent evolution of the true count distribution is again a solution of (9.59) with those initial conditions, specifically $\rho(\gamma - \gamma_E, \tau - \tau_E)$, a generalization of (9.58). Then the distribution of true counts experienced by a back-counter who bets only *after* the true count has first reached γ_E is the convolution with the probability (9.60),

$$\int_0^{\tau} d\tau_E \rho(\gamma - \gamma_E, \tau - \tau_E) \rho_1(\gamma_E, \tau_E) = \rho(\Gamma_E, \tau), \quad (9.65)$$

$$\Gamma_E \equiv |\gamma - \gamma_E| + \gamma_E;$$

performing the integration uses the same methods as in Appendix 2. The back-counter, once he is entered, then bets an amount $B(\gamma)$ based on the true count, and plays until the next reshuffle at penetration F . With the true count distribution of (9.66), his yield is

$$Y_E = \int_{-\infty}^{\infty} d\gamma B(\gamma) R(\gamma) P(\gamma), \quad P(\gamma) \equiv \frac{1}{F} \int_0^F df \rho(\Gamma_E, \tau). \quad (9.66)$$

The entry threshold γ_E is chosen so as to maximize the yield. The distribution $P(\gamma)$ is graphed in Fig. 9.8.

9.3.2 Entry and Exit

Thus far we have assumed that the back-counter, once he has entered the game, remains in until the shoe is reshuffled. But some authorities suggest (e.g., Vancura and Fuchs 1998, p. 132) that he leave the table – we will call it “exit” – when the true count drops below the roughly +1 value at which the expected return crosses from positive to negative. We will analyze the more general situation of exit at a true count γ_X , and choose it jointly with γ_E to maximize the yield.

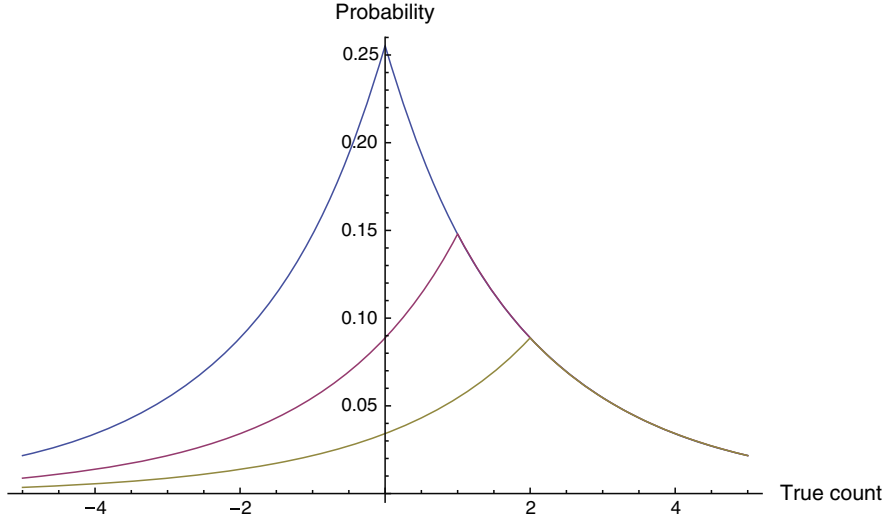


Fig. 9.8 Distributions of true counts, subsequent to entry at true counts 0, 1, 2

Without exit, the distribution of true counts following entry is given by (9.69), and satisfies the diffusion equation, (9.59). With exit, the modified distribution $\rho_{EX}(\gamma, \tau)$ for $\gamma \geq \gamma_X$ at time τ must still satisfy (9.59) but now with the boundary condition $\rho_{EX}(\gamma_X, \tau) = 0$. The solution, analogous to (9.61), is accomplished with a reflection about γ_X :

$$\rho_{EX}(\gamma, \tau) = [\rho(\Gamma_E, \tau) - \rho(|2\gamma_X - \gamma - \gamma_E| + \gamma_E, \tau)] \Theta(\gamma - \gamma_X). \quad (9.67)$$

Equations (9.67) and (9.66) are compared schematically in Fig. 9.9, for the same game parameters as in Fig. 9.8.

Furthermore, the probability of exiting at “time” τ_X , conditional on entry at τ_E , is the convolution of their respective probabilities,

$$\begin{aligned} P_X(\tau_X) &= \int_0^{\tau_X} d\tau_E \rho_1(\gamma_X - \gamma_E, \tau_X - \tau_E) \rho_1(\gamma_E, \tau_E) \\ &= \rho_1(2\gamma_E - \gamma_X, \tau_X). \end{aligned} \quad (9.68)$$

These two expressions, like (9.60) and (9.61) above, similarly satisfy conservation of probability but now conditional on entry:

$$\int_{\tau_E}^{\tau} d\tau_X P_X(\tau_X) + \int_{-\infty}^{\infty} d\gamma \rho_{EX}(\gamma, \tau) = 2\text{Erf}\left(\frac{\gamma_E}{\sqrt{\tau}}\right), \quad (9.69)$$

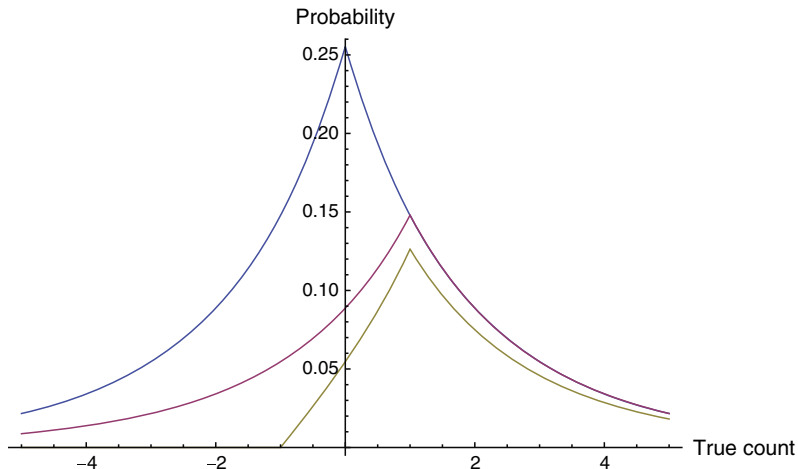


Fig. 9.9 Distributions of true counts: with neither entry nor exit, with entry at +1 but without exit, and with entry at +1 and exit at -1

where the right-hand side, by (9.62), is the probability of entry prior to τ .

The yield from this table is the analog of (9.66):

$$Y_{EX}^{(1)} = \frac{1}{F} \int_0^F df \int_{-\infty}^{\infty} d\gamma B(\gamma) R(\gamma) \rho_{EX}(\gamma, \tau). \quad (9.70)$$

But upon exit, a back-counter can “table-hop” to a second table with a freshly shuffled shoe and repeat the entry process. Hence, the most appropriate assessment of his yield is to include the cash flow from the second table during those rounds of the first that follow the exit and precede its reshuffle. Thus, the yield from the second table alone, weighting with the probability of exit from the first, is

$$Y_{EX}^{(2)} = \int_0^{\tau_F} d\tau_X P_X(\tau_X) \frac{1}{F} \int_0^{F-\tau_X} df \int_{-\infty}^{\infty} d\gamma B(\gamma) R(\gamma) \rho_{EX}(\gamma, \tau). \quad (9.71)$$

By combining the yields from the two tables, and carrying out the intermediate τ_X integration, the total yield becomes

$$\begin{aligned} Y_{EX} &= Y_{EX}^{(1)} + Y_{EX}^{(2)} \\ &= \frac{1}{F} \int_0^F df \left[1 + 2 \operatorname{Erf} \left(\frac{2\gamma_E - \gamma_X}{\sqrt{\tau}} \right) \right] \int_{-\infty}^{\infty} d\gamma B(\gamma) R(\gamma) \rho_{EX}(\gamma, \tau). \end{aligned} \quad (9.72)$$

where $\bar{\tau} \equiv 52(F - f)/D(1 - F + f)$. Although contributions from additional tables beyond the second should in principle be included, in practice these are negligible: the probability of exit from the second table is quite small and the number of rounds played at the third is typically too few to generate much additional value. The optimal entry and exit thresholds are determined, as in the entry-only case, by jointly maximizing the total yield.

9.3.3 Entry and Departure

Separately from a possible exit *following* entry, the table-hopper may independently choose to “depart” the table *prior* to entry. This might occur, for example, if much of the shoe has been dealt without reaching the entry threshold; or if the true count becomes decidedly negative and the probability of it swinging sufficiently positive to trigger entry is correspondingly low. Ideally, the departure decision should be based on a combination of these two circumstances. But Player decision-making predicated on both true count and depth parameters together seems difficult; rather, we consider departure based only on true count.

Exit, occurring after entry by definition, imposes the single boundary constraint $\gamma \geq \gamma_X$, satisfied as per (9.67) by a reflection about γ_X . Departure, occurring alternatively to entry, instead imposes the two simultaneous boundary constraints $\gamma_D \leq \gamma \leq \gamma_E$, which requires the more complex structure of an infinite array of reflections. A real-world example is that if one looks in a mirror one sees a single reflection; but if one instead looks in a mirror with a second one parallel to and facing it, one sees infinitely many, receding reflections.

Thus, the solution of the diffusion equation subject to the two boundary constraints is the generalization of (9.61),

$$\begin{aligned} \rho_{ED}(\gamma, \tau) &= \sum_{m=-\infty}^{\infty} [\rho(\gamma - a_m) - \rho(\gamma - 2\gamma_E - a_m)], \\ a_m &\equiv 2m(\gamma_E - \gamma_D). \end{aligned} \quad (9.73)$$

Then the probability of entry at τ_E is given by the generalization of the chain rule, (9.64),

$$\begin{aligned} \rho_{1D}(\gamma_E, \tau_E) &= \frac{1}{\tau_E} \int_0^{\tau_E} d\tau \int_{\gamma_D}^{\gamma_E} d\gamma \rho_1(\gamma_E - \gamma, \tau_E - \tau) \rho_{ED}(\gamma, \tau) \\ &= \frac{\gamma_E}{\tau_E} \rho(\gamma_E, \tau_E) - \frac{\gamma_E - \gamma_D}{\tau_E} \sum_{m=1}^{\infty} [\rho(a_m - \gamma_E, \tau_E) \\ &\quad - \rho(a_m + \gamma_E, \tau_E)]; \end{aligned} \quad (9.74)$$

the integrations use the same methods as for the chain rule, although the algebra is lengthier. Since the terms in the sum decrease exponentially with increasing m , a reasonable approximation is to keep only the larger of the two $m = 1$ terms; the result is

$$\rho_{1D}(\gamma_E, \tau_E) \cong \frac{\gamma_E}{\tau_E} \rho(\gamma_E, \tau_E) - \frac{\gamma_E - \gamma_D}{\tau_E} \rho(2\gamma_D - \gamma_E, \tau_E). \quad (9.75)$$

The corresponding result for the distribution of γ at τ subsequent to entry, generalizing (9.66), is

$$\begin{aligned} \rho_{ED}(\gamma, \tau) &= \int_0^\tau d\tau_E \rho(\gamma - \gamma_E, \tau - \tau_E) \rho_{1D}(\gamma_E, \tau_E) \\ &= \rho(\Gamma_E, \tau) - \sum_{m=1}^{\infty} \frac{\gamma_E - \gamma_D}{a_m - \gamma_E} [\rho(|\gamma - \gamma_E| + a_m - \gamma_E, \tau) \\ &\quad - \rho(|\gamma + \gamma_E| + a_m + \gamma_E, \tau)] \\ &\cong \rho(\Gamma_E, \tau) - \frac{\gamma_E - \gamma_D}{\gamma_E - 2\gamma_D} \rho(\Gamma_E - 2\gamma_D, \tau). \end{aligned} \quad (9.76)$$

Then the yield from the first table is

$$Y_{ED}^{(1)} = \frac{1}{F} \int_0^F df \int_{-\infty}^{\infty} d\gamma B(\gamma) R(\gamma) \rho_{ED}(\gamma, \tau). \quad (9.77)$$

Additionally, the probability of departure is the reverse of (9.74) and (9.75),

$$\begin{aligned} P_D(\tau_D) &= \frac{-\gamma_D}{\tau_D} \rho(\gamma_D, \tau_D) - \frac{\gamma_E - \gamma_D}{\tau_D} \sum_{m=1}^{\infty} [\rho(a_m + \gamma_D, \tau_D) \\ &\quad - \rho(a_m - \gamma_D, \tau_D)] P_D(\tau_D) \\ &\cong \frac{-\gamma_D}{\tau_D} \rho(\gamma_D, \tau_D) - \frac{\gamma_E - \gamma_D}{\tau_D} \rho(2\gamma_E - \gamma_D, \tau_D). \end{aligned} \quad (9.78)$$

Then the yield from a second table following departure becomes, like (9.71)

$$Y_{ED}^{(2)} = \int_0^{\tau_F} d\tau_D P_D(\tau_D) \frac{1}{F} \int_0^{F-f_D} df \int_{-\infty}^{\infty} d\gamma B(\gamma) R(\gamma) \rho_{ED}(\gamma, \tau) \quad (9.79)$$

and like (9.72) the total yield from both tables is

$$\begin{aligned}
Y_{ED} &= Y_{ED}^{(1)} + Y_{ED}^{(2)} \\
&\cong \frac{1}{F} \int_0^F df \left[1 + 2\text{Erf} \left(\frac{-\gamma_D}{\sqrt{\tau}} \right) - 2 \frac{\gamma_E - \gamma_D}{2\gamma_E - \gamma_D} \text{Erf} \left(\frac{2\gamma_E - \gamma_D}{\sqrt{\tau}} \right) \right] \\
&\quad \times \int_{-\infty}^{\infty} d\gamma B(\gamma) R(\gamma) \rho_{ED}(\gamma, \tau).
\end{aligned} \tag{9.80}$$

As with exit, the optimal entry and departure thresholds are determined by joint maximization of the yield.

9.3.4 Entry, Exit, and Departure

When the table-hopper employs both departure (prior to entry) and exit (subsequent to entry), the probability distribution of true counts at τ combines (9.67) and (9.76),

$$\begin{aligned}
\rho_{EDX}(\gamma, \tau) &= \int_0^{\tau} d\tau_E [\rho(\gamma - \gamma_E, \tau - \tau_E) \\
&\quad - \rho(2\gamma_X - \gamma - \gamma_E, \tau - \tau_E)] \rho_{1D}(\gamma_E, \tau_E) \\
&= [\rho_{ED}(\gamma, \tau) - \rho_{ED}(2\gamma_X - \gamma, \tau)] \Theta(\gamma - \gamma_X),
\end{aligned} \tag{9.81}$$

and the yield from the first table is

$$Y_{EDX}^{(1)} = \frac{1}{F} \int_0^F df \int_{-\infty}^{\infty} d\gamma B(\gamma) R(\gamma) \rho_{EDX}(\gamma, \tau). \tag{9.82}$$

The probability of departure remains independent of any possibility of later exit, so the departure contribution to the second table's yield is still given by (9.79); but the probability of exit is influenced by the possibility of previous departure, so that (9.69) generalizes to

$$\begin{aligned}
P_{XD}(\tau_X) &= \int_0^{\tau_X} d\tau_E \rho_1(\gamma_X - \gamma_E, \tau_X - \tau_E) \rho_{1D}(\gamma_E, \tau_E) \\
&\cong \rho_1(2\gamma_E - \gamma_X, \tau_X) \\
&\quad - \frac{\gamma_E - \gamma_D}{2\gamma_E - \gamma_D} \rho_1(2\gamma_E - \gamma_X - 2\gamma_D, \tau_X).
\end{aligned} \tag{9.83}$$

Thus, the total yield from the first two tables combined, generalizing (9.72) and (9.80), is

$$\begin{aligned}
Y_{EDX} = & \frac{1}{F} \int_0^F df \left\{ 1 + 2 \left[\text{Erf} \left(\frac{-\gamma_D}{\sqrt{\tau}} \right) + \text{Erf} \left(\frac{2\gamma_E - \gamma_X}{\sqrt{\tau}} \right) \right] \right. \\
& \left. - 2 \frac{\gamma_E - \gamma_D}{2\gamma_E - \gamma_D} \left[\text{Erf} \left(\frac{2\gamma_E - \gamma_D}{\sqrt{\tau}} \right) + \text{Erf} \left(\frac{2\gamma_E - \gamma_X - 2\gamma_D}{\sqrt{\tau}} \right) \right] \right\} \quad (9.84) \\
& \times \int_{-\infty}^{\infty} d\gamma B(\gamma) R(\gamma) \rho_{EDX}(\gamma, \tau).
\end{aligned}$$

As before, the three threshold true counts are determined jointly by maximizing the yield. Computations with (9.84), despite its apparent complexity, are feasible and examples of the resulting thresholds are listed in Table 4.5.

Appendix 1 The Risk of Ruin Formula

Concentrate here on the expression for risk of ruin, (9.12), which is slightly simpler than that for mean capital, (9.9), although similar methodologies apply to both. Begin by combining (9.12) with its sequential predecessors (9.6), (9.4), and (9.1). Then use the two Kronecker deltas to eliminate all but one of the three n_ω summations; the most convenient variable to remain is $\lambda \equiv n_+ + n_-$. Next, proceed to the asymptotic limit of large N as in Appendix 1 of Chap. 8: replace all factorials by their asymptotic limit (the Stirling approximation, $n! \approx \sqrt{2\pi n}(n/e)^n$); change the remaining two sums over discrete variables into integrations over continuous ones; and convert the integration over μ (seen in (9.12)) by the Method of Stationary Phase, with limits extended to $\pm\infty$, into an easily performed Gaussian. This leaves only the last integration, in the form

$$\begin{aligned}
W_N & \approx \int_{c_0}^N d\lambda \frac{c_0 \exp E(\lambda)}{\sqrt{2\pi\lambda(\lambda^2 - c_0^2)}}, \quad (9.85) \\
E(\lambda) & \equiv \sum_{\pm} \frac{\lambda \mp c_0}{2} \ln \frac{2\lambda\xi(\pm)}{\sigma^2(\lambda \mp c_0)}.
\end{aligned}$$

Next, apply the Stationary Phase method again: the maximum of $E(\lambda)$ with respect to λ is at $\lambda_0 \equiv c_0/|\hat{R}|$, where $\hat{R} \equiv (\xi(+)-\xi(-))/\sigma^2$; so that in the asymptotic limit it becomes $E(\lambda) \approx E_0 - \hat{R}^2(\lambda - \lambda_0)^2/2\lambda$, with $E_0 \equiv -c_0\Theta(\hat{R}) \ln \frac{1+\hat{R}}{1-\hat{R}}$. Thus, W_N asymptotically becomes

$$W_N \approx \int_0^{\hat{N}} d\lambda \frac{c_0}{\sqrt{2\pi\lambda^3}} \exp \left(E_0 - \frac{\hat{R}^2}{2\lambda} \left(\lambda - \frac{c_0}{|\hat{R}|} \right)^2 \right). \quad (9.86)$$

Although this integral is not itself a Gaussian, it nevertheless can be evaluated as a closed form: it can be looked up in tables or in the *Mathematica* software package, or converted into a sum of Gaussian integrals by changing the integration variable to $y = (c_0 - |\hat{R}|\lambda) / \sqrt{\lambda}$. The result is (9.12) in the text, QED.

Appendix 2 Chain Rule Convolution

Begin with the expression for the ratio,

$$J \equiv \frac{1}{\tau_E} \int_0^{\tau_E} d\tau \int_{-\infty}^{\infty} d\gamma \rho_1(\gamma_E - \gamma, \tau_E - \tau) \rho_E(\gamma, \tau) / \rho_1(\gamma_E, \tau_E). \quad (9.87)$$

After substitution from (9.60) and (9.58), the integrand can be rearranged into the form

$$J = \frac{1}{\gamma_E} \int_{-\infty}^{\gamma_E} d\gamma \int_0^{\tau_E} d\tau \sqrt{\frac{(\gamma_E - \gamma)^2 \tau_E}{2\pi(\tau_E - \tau)^3 \tau}} \left\{ \exp \left[-\frac{\left(\frac{\gamma}{\tau} - \frac{\gamma_E}{\tau_E}\right)^2}{2\left(\frac{1}{\tau} - \frac{1}{\tau_E}\right)} \right] - \exp \left[-\frac{\left(\frac{2\gamma_E - \gamma}{\tau} - \frac{\gamma_E}{\tau_E}\right)^2}{2\left(\frac{1}{\tau} - \frac{1}{\tau_E}\right)} \right] \right\}. \quad (9.88)$$

Next change integration variable: for the first term in the braces substitute $u = \gamma^2 (\tau^{-1} - \tau_E^{-1})$, while for the second term substitute $u' = (2\gamma_E - \gamma)^2 (\tau^{-1} - \tau_E^{-1})$. Then

$$J = \frac{1}{\gamma_E} \int_{-\infty}^{\gamma_E} d\gamma \left\{ \int_0^{\infty} du \frac{|\beta|}{\sqrt{2\pi u^3}} \exp \left(-\frac{(u + \beta)^2}{2u} \right) - \int_0^{\infty} du' \frac{|\beta'|}{\sqrt{2\pi u'^3}} \exp \left(-\frac{(u' + \beta')^2}{2u'} \right) \right\}, \quad (9.89)$$

where $\beta \equiv (\gamma - \gamma_E) \gamma / \tau_E$, $\beta' \equiv (\gamma_E - \gamma) (2\gamma_E - \gamma) / \tau_E$. But

$$\int_0^{\infty} du \frac{|\beta|}{\sqrt{2\pi u^3}} \exp \left(-\frac{(u + \beta)^2}{2u} \right) = \exp(-\beta - |\beta|) \quad (9.90)$$

as listed, e.g., in *Mathematica* or derivable via the further change of variable $v = (u + \beta) / \sqrt{u}$. Finally, shift the γ integration to $\tilde{\gamma} = -\gamma$ for the first term in the braces and to $\tilde{\gamma}' = \gamma_E - \gamma$ for the second, revealing a major cancellation between the two terms while the uncanceled part of the $\tilde{\gamma}$ integral becomes trivial. The result is just $J = 1$, QED.

Chapter 10

Play Strategies with Card Counting

10.1 Count-Dependent Playing Strategy

The analysis leading to Basic Strategy calculates the expected return, R , for a large number of possible playing strategies and identifies the one that maximizes it. The hand values s and s' , at or above which Player should stand, are identified as well as the initial values for which he should double down or split. Taken together, the values for standing, doubling, and splitting – for each Dealer upcard – constitute a set of parameters that specify each possible strategy. The family of all possible strategies numbers in the millions, although only one is maximal. Designate the family as π .

The task here is to develop a methodology for finding the optimal π dependent on the true count(s). The task is complicated by the fact that π is a disparate collection of individual numbers, each of which is an integer and none larger than 20: the usual techniques of the calculus are not suited to discrete variables of this kind. Also, a computer simulation approach would be very cumbersome, since for each true count many distinct strategy sets would need to be computed and compared to find the optimum; and the related task of finding the optimal counting vector(s) would then require repeating the entire computation for each of many candidate vectors. Although neither approach would be successful by itself, in combination they can give at least a qualitative insight.

Begin the maximization analysis as if π were a set of continuous, rather than discrete, elements labeled with the index ν . For this warm-up exercise, make π an explicit variable of R ; then the yield, generalizing (8.23), is

$$Y = \langle\langle B(\hat{\gamma}) \langle R_f(\mathbf{d}; \pi) \rangle_{\hat{\gamma}} \rangle\rangle. \quad (10.1)$$

Also designate π_0 as the optimal parameter set that maximizes Y at $\hat{\gamma} = 0$ (the analog of Optimal Basic Strategy). In addition, again recognize Δ^2 as small, as in Chap. 8, and truncate the Hermite expansion of (8.19) after $n = 2$. Then a series expansion in powers of $\delta\pi \equiv \pi - \pi_0$ gives

$$\begin{aligned}
\langle R_f(\mathbf{d}; \boldsymbol{\pi}) \rangle_{\hat{\gamma}} &\cong R^0 + \frac{1}{2} \delta \boldsymbol{\pi} \delta \boldsymbol{\pi} : \frac{\partial^2 R^0}{\partial \boldsymbol{\pi}_0 \partial \boldsymbol{\pi}_0} \\
&\quad - \sum_{\lambda} \hat{\gamma}_{\lambda} \hat{\boldsymbol{\alpha}}_{\lambda} \cdot \hat{\nabla} \left(R^0 + \delta \boldsymbol{\pi} \cdot \frac{\partial R^0}{\partial \boldsymbol{\pi}_0} \right) \\
&\quad + \frac{1}{2} \sum_{\lambda, \lambda'} (\hat{\gamma}_{\lambda} \hat{\gamma}_{\lambda'} - \Delta^2 \delta_{\lambda, \lambda'}) \hat{\boldsymbol{\alpha}}_{\lambda} \hat{\boldsymbol{\alpha}}_{\lambda'} : \hat{\nabla} \hat{\nabla} R^0
\end{aligned} \tag{10.2}$$

with $R^0 \equiv R_0(\mathbf{d}_0; \boldsymbol{\pi}_0)$. Maximizing with respect to $\delta \boldsymbol{\pi}$ leads to the optimal value

$$\begin{aligned}
\langle R_f(\mathbf{d}; \boldsymbol{\pi}_0) \rangle_{\hat{\gamma}} &= R^0 - \sum_{\lambda} \hat{\gamma}_{\lambda} \hat{\boldsymbol{\alpha}}_{\lambda} \cdot \hat{\nabla} R^0 \\
&\quad + \frac{1}{2} \sum_{\lambda, \lambda'} \hat{\boldsymbol{\alpha}}_{\lambda} \hat{\boldsymbol{\alpha}}_{\lambda'} : [(\hat{\gamma}_{\lambda} \hat{\gamma}_{\lambda'} - \Delta^2 \delta_{\lambda, \lambda'}) \hat{\nabla} \hat{\nabla} R^0 + \hat{\gamma}_{\lambda} \hat{\gamma}_{\lambda'} \mathbf{Q}],
\end{aligned} \tag{10.3}$$

where \mathbf{Q} involves derivatives of R with respect to $\boldsymbol{\pi}_0$,

$$\mathbf{Q} \equiv \sum_{v, v'} \frac{\partial \hat{\nabla} R^0}{\partial \boldsymbol{\pi}_{0, v}} \left(\frac{\partial^2 R^0}{\partial \boldsymbol{\pi}_0 \partial \boldsymbol{\pi}_0} \right)^{-1}_{v, v'} \frac{\partial \hat{\nabla} R^0}{\partial \boldsymbol{\pi}_{0, v'}}. \tag{10.4}$$

Furthermore, since $\hat{\mathbf{d}} \cdot \hat{\nabla} = 0$ as per (8.9), hence $\hat{\mathbf{d}} \cdot \mathbf{Q} = \mathbf{Q} \cdot \hat{\mathbf{d}} = 0$.

Next, some of this analysis is at least qualitatively correct in the actual situation that $\boldsymbol{\pi}$ is a set of discrete variables rather than continuous ones. The approach assumes that (10.3) is still appropriate, i.e., that $\langle R_f(\mathbf{d}; \boldsymbol{\pi}) \rangle_{\hat{\gamma}}$ is a quadratic form in $\sum_{\lambda} \hat{\gamma}_{\lambda} \hat{\boldsymbol{\alpha}}_{\lambda}$ with coefficients \mathbf{Q} at second order, even though (10.4) is disregarded and instead \mathbf{Q} is viewed as an array that must be determined computationally.

The assumption has already been validated numerically in Chap. 8. There, Fig. 8.1 displays results for the quantities

$$\begin{aligned}
R_{\text{opt}} &\equiv R_0(\langle \mathbf{d} \rangle_{\hat{\gamma}}; \boldsymbol{\pi}) - R^0, \\
R_{\text{bas}} &\equiv R_0(\langle \mathbf{d} \rangle_{\hat{\gamma}}; \boldsymbol{\pi}_0) - R^0,
\end{aligned} \tag{10.5}$$

as a function of a single count $\hat{\gamma}$. It uses (8.11) for $\langle \mathbf{d} \rangle_{\hat{\gamma}}$ with a counting vector \mathbf{v}_b as per (8.29); and for R_{opt} , $\boldsymbol{\pi}$ is continuously optimized with respect to $\hat{\gamma}$. The evidence of Fig. 8.1 is that both R_{opt} and R_{bas} can reasonably be approximated by cubic polynomials in true count, although Fig. 8.3 demonstrates the fine-grained kinkiness of R_{opt} .

Thus, returning to the computation of \mathbf{Q} , employ the following approach: while systematically varying $\hat{\boldsymbol{\alpha}}$ throughout its vector space, for each such $\hat{\boldsymbol{\alpha}}$ least-squares fit a cubic to $R_0(\langle \mathbf{d} \rangle_{\hat{\gamma}}; \boldsymbol{\pi}) - R_0(\langle \mathbf{d} \rangle_{\hat{\gamma}}; \boldsymbol{\pi}_0)$ as a function of $\hat{\gamma}$ and extract the coefficient of the quadratic term. This approach gives numerical values for all 55 independent elements of the \mathbf{Q} array.

10.1.1 Counting Vector Optimal for Play Variation Alone

Having established the validity of (10.3) with computed coefficients, the next step is to maximize Y with respect to the counting vectors. In this section, the task is simplified by assuming that no bet variation is carried out; the succeeding sections explore varying bet and play strategies simultaneously.

If the bet is fixed, then the yield, combining (10.1) and (10.3) together with (8.6) in the caret representation, reduces to

$$Y = B \left[R_0 + \frac{\psi(F)}{52D} \frac{1}{2} \sum_{\lambda=1}^{\Lambda} \hat{\alpha}_{\lambda} \hat{\alpha}_{\lambda} : \mathbf{Q} \right], \quad (10.6)$$

where $\psi(F)$, using (8.23), is

$$\psi(F) \equiv 52D \langle \langle \Delta^2 \rangle \rangle = \frac{1}{F} \int_0^F df \frac{f}{1-f} = -1 + \frac{1}{F} \ln \frac{1}{1-F}, \quad (10.7)$$

and is plotted in Fig. 10.1.

The techniques of Chap. 8, especially its Appendix 2, are applicable to finding the stationary points of Y with respect to $\hat{\alpha}_{\lambda}$. The solution is facilitated by introducing the eigenmodes of \mathbf{Q} ,

$$\mathbf{Q} = \sum_{\kappa=1}^9 Q_{\kappa} \mathbf{e}_{\kappa} \mathbf{e}_{\kappa}, \quad (10.8)$$

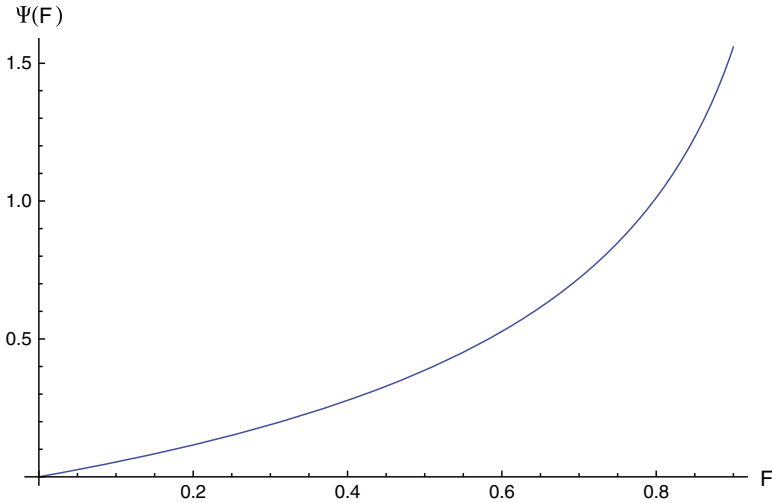


Fig. 10.1 The function $\Psi(F)$

with the index $\kappa = 1$ corresponding to the largest of the nonvanishing eigenvalues Q_κ and $\kappa = 9$ the smallest; the tenth eigenvector, $\hat{\mathbf{d}}$, has eigenvalue zero. It is easily shown that Y is stationary when each $\hat{\alpha}_\lambda$ aligns with an eigenvector \mathbf{e}_κ . Substitution into (10.6) shows that Y is maximized when the Λ largest of the eigenmodes are selected: the maximum Y is given by

$$Y = B \left[R_0 + \frac{\psi(F)}{52D} \frac{1}{2} \sum_{\kappa=1}^{\Lambda} Q_\kappa \right]. \quad (10.9)$$

An important consequence of (10.9) is that the number of useful counting vectors is no longer confined to just one. Although the first vector picks up the largest contribution to Y , additional vectors add further contributions. However, if the largest eigenvalue Q_1 is much larger than the sum of all the others, then that eigenvector provides the dominant effect and the others are insignificant; the key measure is the ratio $Q_1 / \text{tr} \mathbf{Q}$, where the trace operation is defined as $\text{tr} \mathbf{Q} \equiv \sum_{\kappa=1}^9 Q_\kappa$. Furthermore, a relevant measure of the worth of an approximate counting vector is its “play correlation,” $C_p \equiv \hat{\alpha}^* \cdot \mathbf{e}_1$, analogous to the betting correlation of (8.30).

Since the elements of \mathbf{Q} are available numerically, it is straightforward to compute their eigenmodes and so assess the incremental yield from varying the play strategy. First of all, the largest eigenvalue of \mathbf{Q} is $Q_1 = .833$, while the trace is $\text{tr} \mathbf{Q} = 1.256$. Thus, the largest eigenmode accounts for just about 2/3 of the trace: that single counting vector realizes much (but certainly not nearly all) of the total yield improvement available from play variation. Other numerical results, useful in other sections, are $\mathbf{v}_b \cdot \mathbf{Q} \cdot \mathbf{v}_b = .556$ and $|\mathbf{v}_b \cdot \mathbf{Q} \cdot (\mathbf{1} - \mathbf{v}_b \mathbf{v}_b)| = .241$.

Next, the eigenvector \mathbf{e}_1 and the optimal counting vector for play variation $\alpha^*(j) \equiv e_1(j) / \hat{d}(j)$ are listed in Table 10.1; the second row is also shown in Table 5.1.

As noted in Chap. 5, where it is labeled vp , this counting vector is similar to the optimal bet variation counting vector of Table 8.2, except for the striking reversal of sign for aces ($j = 1$) and the interchange of $j = 6$ with $j = 8$. Thus, rather than the two counting vectors being orthogonal, they have a strong correlation: $\mathbf{v}_b \cdot \mathbf{e}_1 = .833$.

Although α^* has irrational elements and so is not practical for actual counting, several approximations with integer or half-integer elements suggest themselves, analogous to those of Tables 3.1 and 3.2. The simplest, analogous to the Hi-Opt scheme but with the switch of $j = 3, 6$ with $j = 7, 8$, is also shown in Table 10.1, designated α_{HO}^* , and as the “simplified” row of Table 5.1; it has a play correlation of $C_p = .95$. The Hi-Opt vector itself has $C_p = .88$, larger than the $C_p = .80$ of

Table 10.1 Optimal counting vector for play variation

j	1	2	3	4	5	6	7	8	9	10
$e_1(j)$	+0.10	+0.09	+0.16	+0.30	+0.42	+0.06	+0.25	+0.18	−0.03	−0.77
$\alpha^*(j)$	+0.37	+0.33	+0.57	+1.10	+1.52	+0.22	+0.88	+0.66	−0.12	−1.38
$\alpha_{HO}^*(j)$	0	0	0	+1	+1	0	+1	+1	0	−1

the Hi-Lo vector and even the 0.83 of \mathbf{v}_b ; the higher play correlation of the Hi-Opt vector confirms the claim of its proponents that it does a better job than other simple approximations in taking advantage of play variation.

When counting with the vector α^* and adjusting play strategy, the yield improves according to (10.9); numerically,

$$Y/B = R_\infty + (.0067 + .0080 \psi(F)) / D. \quad (10.10)$$

For casinos that reshuffle frequently, e.g., after only half the pack is dealt, $\psi \approx .4$; at the other end of the range, few casinos will put off reshuffling until the penetration is as much as $F = 0.8$, where $\psi \approx 1$. More typical might be the intermediate value, $\psi \approx .7$. The result is that play variation improves the yield ratio only slightly, +0.002 at most, for $D = 4$ or more; it does not in itself shift the yield even close to positive. For $D = 1$, however, where the Optimal Basic Strategy return is already positive, play variation boosts the yield ratio by perhaps another +0.006.

10.1.2 Single Counting Vector Optimal for Bet and Play Together

It is disappointing, but not surprising, that varying the play strategy, based on the true count with a well-chosen counting vector, provides only a slight improvement in yield. This outcome is similar to the negligible improvement found in Sect. 2.1 from variations in optimal play with the number of decks or the composition of the hand. It seems that the yield is quite insensitive to changes in optimal play strategy arising from small changes in expected return. Nevertheless, pursuing the possibility to its logical end, consider how bet and play variations interact with each other in determining the best choice of counting vector(s) and the resulting yield.

The detailed analysis differs depending on whether only a single vector is considered, optimized to take best advantage of bet and play strategies together, or two vectors, one specific for betting and the other for playing. This section focuses on the former case, the next on the latter. Combine expressions (8.31) and (10.5) (generalized to a varying bet size) to obtain

$$Y = Y_0 + Y_1 \hat{\alpha} \cdot \mathbf{v}_b + \hat{\alpha} \hat{\alpha} : (Y_{21} \hat{\nabla} \hat{\nabla} R^0 + Y_{22} \mathbf{Q}) / 2, \quad (10.11)$$

where $Y_1 \equiv \langle \langle B(\hat{y}) \hat{y} \rangle \rangle \left| \hat{\nabla} R^0 \right|$, $Y_{21} \equiv \langle \langle B(\hat{y}) (\hat{y}^2 - \Delta^2) \rangle \rangle$, $Y_{22} \equiv \langle \langle B(\hat{y}) \hat{y}^2 \rangle \rangle$. Leaving aside the Y_{21} term, the maximization of Y with respect to $\hat{\alpha}$ is again facilitated by using the eigenmodes of \mathbf{Q} , leading to

$$\alpha_\kappa \equiv \hat{\alpha} \cdot \mathbf{e}_\kappa = Y_1 v_\kappa / (L - Y_{22} Q_\kappa). \quad (10.12)$$

The Lagrange parameter L is determined by the normalization of $\hat{\alpha}$, and so satisfies the implicit equation

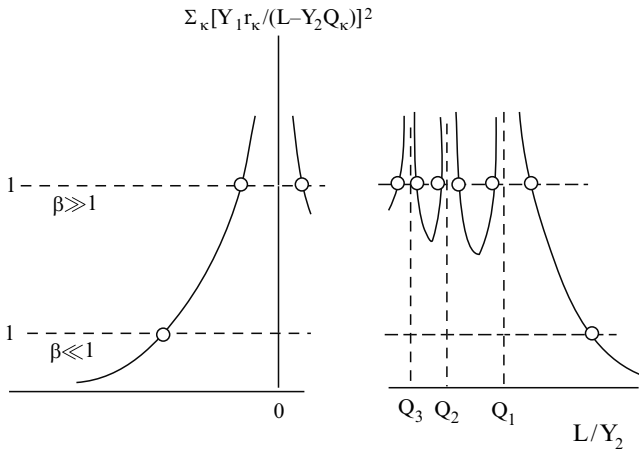


Fig. 10.2 Graphical solution to eigenvalue equation

$$1 = \sum_{\kappa} [Y_1 v_{\kappa} / (L - Y_{22} Q_{\kappa})]^2. \quad (10.13)$$

Equation (10.13) has at least two solutions, as can be seen by graphing its right hand side (rhs) as a function of L/Y_{22} (schematically, Fig. 10.2). Since the rhs tends to zero for L tending to $\pm\infty$, and tends to infinity (a pole) for L/Y_{22} tending to each Q_{κ} , there are always solutions for $L/Y_{22} > Q_1$ and for $L/Y_{22} < Q_{10} = 0$. The first of these maximizes Y and the second minimizes it. There may or may not also be solutions for L intermediate between two consecutive poles, depending on the depth of the minimum between those poles; the depth, in turn, is controlled by the ratio $\eta \equiv Y_{22}/Y_1$. If the ratio is large, solutions exist for L near every pole; in the limit, these tend to the solutions of the previous section. If the ratio is small, no intermediate solutions exist; in the limit, the two “outer” solutions tend to those of Chap. 8.

In principle, $\hat{\alpha}$ could be computed, for any given η , using the numerical eigenmodes of \mathbf{Q} and (10.11) and (10.12). But in practice η is small when bet variation is carried out, so insight is added by proceeding further analytically. A series expansion of these relations to first order in powers of η finds

$$\begin{aligned} \hat{\alpha} &\equiv \mathbf{v}_b (1 - |\alpha_{\perp}|^2)^{1/2} + \alpha_{\perp}, \\ \alpha_{\perp} &\cong \eta \mathbf{v}_b \cdot \mathbf{Q} \cdot (\mathbf{1} - \mathbf{v}_b \mathbf{v}_b). \end{aligned} \quad (10.14)$$

The presence of play variation shifts the optimal counting vector away from \mathbf{v}_b , by adding a small component perpendicular to it while shrinking the magnitude along it to preserve the normalization.

With the counting vector optimized, the resulting yield can be obtained by substituting back into (10.11); through second order in η and using the numerical results of the previous section,

$$\begin{aligned}
Y &\cong Y_0 + Y_1 + (Y_{22} \mathbf{v}_b \cdot \mathbf{Q} \cdot \mathbf{v}_b + Y_1 |\boldsymbol{\alpha}_\perp|^2) / 2 \\
&= Y_0 + Y_1 (1 + .28 \eta + .03 \eta^2).
\end{aligned} \tag{10.15}$$

The first three terms of this last expression are together just the yield if $\hat{\boldsymbol{\alpha}}$ were not shifted from \mathbf{v}_b as per (10.14): the shift increases Y only to second order in η – the last term of (10.15) – an increase that is essentially negligible. Thus, even if both play and bet size are varied, to a good approximation a single counting vector remains fixed at its optimum with respect to bet variation alone; and this conclusion still holds even with the more elaborate analysis that includes the Y_{21} term. Furthermore, the added yield from play variation even without the shift in counting vector (the third term in (10.15)) is almost exactly 2/3 of that available if the play variation were at its maximum, $C_p = 1$; maneuvers to increase yield still more have little left to gain.

10.1.3 Two Distinct Counting Vectors

As mentioned earlier, the strategy of varying both bet and play suggests the possibility of using two different counting vectors, one for choosing the bet size and the other for choosing the play. It seems intuitive that Y would be optimized by selecting \mathbf{v}_b as the counting vector for bet size and \mathbf{e}_1 for play; and that the resulting yield would be $Y = Y_0 + Y_1 + Y_{22} Q_1 / 2$, an improvement over that of (10.15) unless η is quite large (> 4.8). Although these simple suppositions are indeed correct, deriving them requires some care.

The mathematical challenge is that \mathbf{v}_b and \mathbf{e}_1 are not orthogonal: in fact, as seen earlier, they are substantially correlated. Thus, they will not each align with a distinct optimal orthonormal vector $\hat{\boldsymbol{\alpha}}_\lambda$. Structure this problem by designating $\hat{\boldsymbol{\alpha}}_2$ as the counting vector for play variation, and allowing the counting vector for bet variation to be some linear combination of the $\hat{\boldsymbol{\alpha}}_\lambda$. If so, then $\langle \langle B \hat{\gamma}_\lambda \rangle \rangle = Y_1 \theta_\lambda$ where, except for the normalization $\sum_\lambda \theta_\lambda^2 = 1$, the θ_λ remains to be determined. The yield in this situation is

$$Y = Y_0 + Y_1 \sum_\lambda \theta_\lambda \hat{\boldsymbol{\alpha}}_\lambda \cdot \mathbf{v}_b + Y_{22} \hat{\boldsymbol{\alpha}}_2 \hat{\boldsymbol{\alpha}}_2 : \mathbf{Q} / 2. \tag{10.16}$$

Thus, Y is stationary with respect to the $\hat{\boldsymbol{\alpha}}_\lambda$ provided

$$Y_1 \theta_\lambda \hat{\boldsymbol{\alpha}}_\lambda \cdot \mathbf{v}_b + Y_{22} \delta_{\lambda,2} \mathbf{Q} \cdot \hat{\boldsymbol{\alpha}}_2 = \sum_{\lambda,\lambda'} L_{\lambda,\lambda'} \hat{\boldsymbol{\alpha}}_{\lambda'}. \tag{10.17}$$

Setting $\hat{\boldsymbol{\alpha}}_2 = \mathbf{e}_1$, (10.17) separates into the two conditions

$$\begin{aligned}
Y_1 \theta_1 \mathbf{v}_b &= L_{11} \hat{\boldsymbol{\alpha}}_1 + L_{12} \mathbf{e}_1, \\
Y_2 \theta_2 \mathbf{v}_b + Y_2 Q_1 \mathbf{e}_1 &= L_{12} \hat{\boldsymbol{\alpha}}_1 + L_{22} \mathbf{e}_1.
\end{aligned} \tag{10.18}$$

Equations (10.18) is internally consistent only if $L_{11} \theta_2 = L_{12} \theta_1$ and $L_{11} (L_{22} - Y_2 Q_1) = (L_{12})^2$. The remaining elements of $L_{\lambda, \lambda'}$ are fixed from the constraints that $\hat{\alpha}_1$ is normalized and orthogonal to \mathbf{e}_1 ; thus $L_{12} = Y_1 \theta_1 \cos \varphi$ where $\cos \varphi \equiv \mathbf{v}_b \cdot \mathbf{e}_1$ and $L_{11} = Y_1 \theta_1 \sin \varphi$. Together, these determine the θ_λ : namely $\theta_1 = \sin \varphi$ and $\theta_2 = \cos \varphi$. These results, in turn, lead to

$$\hat{\alpha}_1 = \mathbf{v}_b \csc \varphi - \mathbf{e}_1 \cot \varphi. \quad (10.19)$$

Finally, using (10.18), the counting vector for bet variation $\sum_\lambda \theta_\lambda \hat{\alpha}_\lambda$ becomes just \mathbf{v}_b , as intuition suggested. Furthermore, the stationary value of Y , by substituting into (10.15), also matches expectation; substituting the numerical result for Q_1 lead to $Y = Y_0 + Y_1 (1 + .42 \eta)$.

10.1.4 Insurance with Variable Betting

Section 7.1 discussed insurance for the Basic Strategy Player, who does not count or vary his bets, and showed that the expected return from insurance is negative. However, the return becomes positive for large enough positive true counts. Analysis is needed to determine whether Player can gain from accepting insurance and under what circumstances. Based on (7.34) and (8.11), the expected return from insurance, conditional on a true count γ , is

$$\langle R_I \rangle_\gamma = \langle \langle d(1) (3d(10) - 1)/2 \rangle \rangle_\gamma. \quad (10.20)$$

The simplification adopted by most authorities is to replace this with

$$\begin{aligned} \langle R_I \rangle_\gamma^0 &\equiv \frac{1}{2} \langle \langle d(1) \rangle \rangle_\gamma (3 \langle \langle d(10) \rangle \rangle_\gamma - 1) \\ &= \frac{1}{338} \left(1 - \frac{\alpha(1)\gamma}{52} \right) \left(-1 - \frac{12\alpha(10)\gamma}{52} \right); \end{aligned} \quad (10.21)$$

this return is positive for $\gamma \geq \gamma_I^0 \equiv -52/12\alpha(10), \approx 4.053$ with the counting vector $\alpha(10) = -1.0692$ optimal for betting. Accepting insurance for true counts in this range, the added yield becomes

$$\begin{aligned} Y_I &= \langle \langle B(\gamma) \langle R_I \rangle_\gamma \ominus (\langle R_I \rangle_\gamma) \rangle \rangle \\ &\cong \langle \langle B(\gamma) \langle R_I \rangle_\gamma^0 \ominus (\gamma - \gamma_I^0) \rangle \rangle. \end{aligned} \quad (10.22)$$

Although the value of the simplified expression is small (depending specifically on the several parameters relevant to counting: the penetration, number of decks, and bet ramp), it is clearly positive and comparable to the yield gain from other

count-dependent play decisions. Authorities who discuss count-dependent play recommend including this insurance recipe.

However, (10.20) can be evaluated exactly, using (8.19), as

$$\langle R_I \rangle_\gamma = \frac{1}{338} \left[\left(1 - \frac{\alpha(1)\gamma}{52} \right) \left(-1 - \frac{12\alpha(10)\gamma}{52} \right) - 12\Delta^2\alpha(1)\alpha(10) \right]. \quad (10.23)$$

This more general expression can be computed straightforwardly and, with the optimal counting vector for betting, results in yield gain that is smaller than with (10.21) by roughly 20–25%, depending on the parameters mentioned above. The key cause is that the threshold for positive $\langle R_I \rangle_\gamma$, based on (10.23) and a counting vector for betting, increases monotonically from γ_I^0 with increasing depth. Thus, at depths approaching the penetration, the threshold becomes so high that it has little probability of being reached: the opportunities for insurance actually shrink as the pack is dealt, rather than expand. Conversely, retaining the γ_I^0 threshold at all depths imposes negative expectation bets at larger depths and so performs still worse by another 10% or so. A better choice of an approximate, depth-independent true count threshold is around 4.5 for four decks, 5.5 for one deck, which perform nearly as well as the exact depth-dependent threshold.

Griffin (1999, p. 71) and Wong (1994, p. 54), among others, note that a second counting vector, optimized and used exclusively for the insurance decision, might improve the yield vs. using a counting vector designed for betting. Working within the simplified (10.21), they recognize that the (normalized) second vector which produces the lowest true count threshold and most steeply increasing $\langle R_I \rangle_\gamma^0$, and hence the greatest yield, is $\alpha(10) = -3/2$, $\alpha(j \neq 10) = 2/3$; this is much like Thorp's Tens-count. This lowest threshold then is $\gamma_I^0 = 52/18 = 2.889$, well below that for the Hi-Lo vector.

Considering the exact insurance return, (10.23), and maximizing its yield, shows that the threshold for positive $\langle R_I \rangle_\gamma$ shifts downward with depth (rather than upward, as it does with the bet counting vector, because of the change in sign of $\alpha(1)$) from γ_I^0 , its value at zero. For Players who do not wish to accurately track the threshold down with depth, the best choice of a depth-independent threshold, with only slightly diminished performance, would correspond to an intermediate depth: a true count of about 2.7 for four decks and 2.1 for one deck. But even with such a second count, the insurance yield is still only improved by roughly 50% over that with just a bet vector, sufficiently modest that the additional counting effort seems unwarranted.

Strictly speaking, though, this Tens vector is the optimal counting vector only at or near zero depth, where the insurance option has very low likelihood. The optimal vector shifts with increasing depth, becoming noticeably different from the Tens (although its elements are no longer simple rationals, and so are impractical to use). Nevertheless, this refinement similarly makes only a slight improvement in yield. The bottom-line recommendation, reiterating Sect. 5.1, is to decline insurance unless the true count (with a bet counting vector such as Hi-Lo) exceeds roughly +5.

10.2 Counter Basic Strategy

Recall that Basic Strategy is the simple way to play a hand that maximizes the expected return from the reshuffle round. For Players who do not count or vary bets, Basic Strategy also maximizes yield. For card counters, though, the previous section considered the adjustment of play parameters depending on the true count, resulting in a complicated set of strategy indices and an only slightly improved yield. But if Player counts and also varies bets, a different approach to play variation suggests itself: a count-*independent* play strategy that maximizes the yield, averaged over all the rounds in a shoe. Such a strategy, termed Counter Basic Strategy (CBS) by Marcus (2007), has the important advantage of simplicity, like Basic Strategy itself, yet still captures much of the (modest) yield improvement possible from the strategy indices. The analysis given here follows Werthamer (2007).

In the notation of the previous section, consider those strategy parameters π_C that maximize the expected return from the reshuffle round at the true count γ_C . Then, if the π_C is regarded as independent of depth and used throughout the shoe, the yield will be

$$Y(\gamma_C) = \left\langle \left\langle B(\gamma) \langle R_f(\mathbf{d}; \pi_C) \rangle_\gamma \right\rangle \right\rangle, \quad (10.24)$$

as per (8.23). But since the bet sizes $B(\gamma)$ vary with true count and have a distribution that is peaked at some positive value, the yield should be a maximum when γ_C lies near that peak. Call that true count γ^* , the associated play parameters π^* , and the resulting maximal yield

$$Y^* = Y(\gamma^*) = \left\langle \left\langle B(\gamma) \langle R_f(\mathbf{d}; \pi^*) \rangle_\gamma \right\rangle \right\rangle. \quad (10.25)$$

Furthermore, this expression can be simplified using the Hermite expansion of (8.19):

$$\begin{aligned} Y^* &= \sum_{n=0}^{\infty} (1/n!) \langle \langle B(\gamma) \Delta^n H_n(\hat{\gamma}/\Delta) \rangle \rangle \left(-\hat{\alpha} \cdot \hat{\nabla} \right)^n R_0(\mathbf{d}_0; \pi^*) \\ &= \sum_{n=0}^{\infty} c_n \langle \langle B(\gamma) \Delta^n H_n(\hat{\gamma}/\Delta) \rangle \rangle. \end{aligned} \quad (10.26)$$

In this form, the coefficients c_n can be generated just from the reshuffle round, independently of the bet pattern.

The computations to implement this analytical program comprise several steps. In outline, the first step constructs a code for the expected return from the reshuffle round of a pack with D decks. The play parameters are adjusted to maximize the return at zero count, corresponding to $\pi(0)$, as in Chap. 8.

The second step generalizes to a nonvanishing count. The code is extended so that the probabilities of the cards drawn to the hand reflect the true count at its start, enabling computation of $R_0(\langle \mathbf{d} \rangle_\gamma; \pi)$ for any choice of parameters π . Then the

count-dependent play parameters $\pi(\gamma)$ are obtained from maximizing the expected return at that true count, giving $R_0(\langle \mathbf{d} \rangle_\gamma; \pi(\gamma))$. Implementing this computation leads to the curves of Fig. 8.1.

The third step least-squares fits the computed results for the expected returns of the previous step, $R_0(\langle \mathbf{d} \rangle_\gamma; \pi(\gamma))$, to a low-order polynomial in γ ; quite good accuracy (to about three significant figures throughout) is obtained with a cubic without DAS, a quartic with DAS. The four/five c_n coefficients resulting from the fit are then transcribed into a third/fourth order truncation of the corresponding Hermite polynomial expansion (8.19) and (10.26), giving an excellent numerical approximation to $\langle R_f(\mathbf{d}; \pi) \rangle_\gamma$. Because of the deviation of $R_0(\langle \mathbf{d} \rangle_\gamma; \pi(\gamma))$ from linearity in γ , the Hermite expansion introduces a dependence on depth. So far this is just the program of Chap. 8.

The fourth step now selects several positive trial values for true count γ_C . For each it generates $\pi(\gamma_C)$ from the results of the second step, i.e., by maximizing the expected return at that count. Then, for this fixed set of play parameters, $R_0(\langle \mathbf{d} \rangle_\gamma; \pi(\gamma_C))$ is computed as a function of γ . A truncated Hermite polynomial expression is generated just as in the previous paragraph, approximating the corresponding $\langle R_f(\mathbf{d}; \pi(\gamma_C)) \rangle_\gamma$; this quantity is maximal, of course, only at $\gamma = \gamma_C$.

The fifth step adopts one of the optimal bet ramps $B(\gamma)$ developed in Sect. 9.1; multiplies it with the expected return approximation from the fourth step; and from that uses (8.23) to compute the yield, Y_C , by integrating over γ , weighted by $p(\gamma)$, and averaging over the depth.

Lastly, the set of trial yield values Y_C are interpolated and the maximum is extracted; the corresponding π^* represents the desired Counter Basic Strategy, and Y^* the corresponding yield. Note, however, that CBS is dependent on the particular choice of bet strategy, as well as on the counting vector and the penetration. A representative case is exhibited in Table 5.4.

References

- Baldwin, Roger, Wilbert E. Cantey, Herbert Maisel, James P. McDermott, "The Optimum Strategy in Blackjack", *Journal of the American Statistical Association* 51, 429–439 (1956)
- Black, Fischer, Myron Scholes, "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy* 81, 637–654 (1973)
- Carlson, Bryce, "Blackjack for Blood", CompuStar Press (1992)
- Cox, J.C., S.A. Ross, M. Rubenstein, "Option Pricing: A Simplified Approach", *Journal of Financial Economics* 7, 229–263 (1979)
- Epstein, Richard A., "The Theory of Gambling and Statistical Logic", Academic, New York, revised edition (1995)
- Ethier, S.N., "The Kelly System Maximizes Median Fortune", *Journal of Applied Probability* 41, 230–6 (2004)
- Griffin, Peter A., "The Theory of Blackjack", Huntington Press, Las Vegas, NV, 6th edition (1999)
- Harris, B., "The Theory of Optimal Betting after Spreads; Janacek, K., "Theory of Optimal Betting"; Yamashita, W., "Optimal Betting Strategy": all at <http://www.bjmath.com/bjmath/Betsize> (1997)
- Hull, J.C., "Options, Futures and Other Derivatives", Prentice-Hall, Englewood Cliffs, NJ (2002)
- Ingersoll, J.E., Jr., "Theory of Financial Decision Making", Rowman & Littlefield, New York (1987)
- Ito, C., "On Stochastic Differential Equations", *Memoirs of the American Mathematical Society* 4, 1–51 (1951)
- Kelly, John L., Jr., "A New Interpretation of Information Rate", *Bell System Technical Journal* 35, 917–926 (1956)
- Manson, A.R., A.J. Barr, J.H. Goodnight, "Optimum Zero-Memory Strategy and Exact Probabilities for 4-Deck Blackjack", *The American Statistician* 29, 84–88 (1975)
- Marcus, H.I., "New Blackjack Strategy for Players Who Modify Their Bets Based on the Count", in Ethier, S.N. and W.R. Eadington (eds.), "Optimal Play: Mathematical Studies of Games and Gambling", Institute for the Study of Gambling and Commercial Gaming, University of Nevada, Reno (2007)
- Maslov, Sergei, Y.-C. Zhang, "Optimal Investment Strategy for Risky Assets", *International Journal of Theoretical and Applied Finance* 1, 377–387 (1998)
- Marsili, Matteo, Sergei Maslov, Y.-C. Zhang, *Physica A* 253, 403–418 (1998)
- Merton, Robert C., "Theory of Rational Option Pricing", *Bell Journal of Economics & Management Science* 4, 141 (1973)
- Mezrich, Ben, "Bringing Down the House", Free Press, New York (2002); "Busting Vegas", William Morrow, New York (2006)
- Morse, P.M., H. Feshbach, "Methods of Theoretical Physics", McGraw-Hill, New York (1953)
- Poundstone, William, "Fortune's Formula", Hill and Wang, New York, NY (2005)
- Revere, Lawrence, "Playing Blackjack as a Business", Carol Publishing, Secaucus, NJ, revised edition (1980)

- Samuelson, Paul A., "Proof that Properly Anticipated Prices Fluctuate Randomly", *Industrial Management Review* 6, 41 (1965)
- Scarne, John, "Scarne's Complete Guide to Gambling", Simon & Schuster, New York (1961)
- Schlesinger, Don, "Blackjack Attack: Playing the Pros' Way", RGE Publishing, North Las Vegas, NV, 3rd edition (2005)
- Scott, Jean, "The Frugal Gambler", Huntington Press, Las Vegas, NV, 2nd edition (2005)
- Shores, Thomas S., "Applied Linear Algebra and Matrix Analysis", Springer, Berlin (2007)
- Silberstang, Edwin, "Winning Blackjack for the Serious Player", Cardoza Publishing, Las Vegas, NV (1993)
- Sileo, Patrick, "The Evaluation of Blackjack Games Using a Combined Expectation and Risk Measure", in Eadington, W.R. and J.A. Cornelius (eds.), "Gambling and Commercial Gaming", University of Nevada, Reno (1992)
- Snyder, Arnold, "Blackbelt in Blackjack", Cardoza Publishing, Las Vegas, NV, 3rd edition (2005); "The Big Book of Blackjack", Cardoza Publishing, Las Vegas, NV (2006)
- Thorp, Edward O., "Beat the Dealer", Blaisdell, New York, revised edition (1966)
- Thorp, Edward O., "Does Basic Strategy Have the Same Expectation for Each Round?", in Vancura, O., J.A. Cornelius, and W.R. Eadington (eds.), "Finding the Edge: Mathematical Analysis of Casino Games", University of Nevada, Reno (2000)
- Vancura, Olaf, Ken Fuchs, "Knock-Out Blackjack", Huntington Press, Las Vegas, NV (1998)
- Werthamer, N. Richard, "Optimal Betting in Casino Blackjack", *International Gambling Studies* 5, 253 (2005); "Optimal Betting in Casino Blackjack II: Back-counting", *International Gambling Studies* 6, 111 (2006); "Optimal Betting in Casino Blackjack III: Table-hopping", *International Gambling Studies* 8, 63 (2008)
- Werthamer, N. Richard, "Basic Strategy for Card Counters: An Analytic Approach", in Ethier, S.N. and W.R. Eadington (eds.), "Optimal Play: Mathematical Studies of Games and Gambling", Institute for the Study of Gambling and Commercial Gaming, University of Nevada, Reno (2007)
- Wong, Stanford, "Professional Blackjack", Pi Yee Press, Las Vegas, NV, 5th edition (1994); "Blackjack Secrets", Pi Yee Press, Las Vegas, NV (1994)
- Zender, Bill, "Casino-ology: The Art of Managing Casino Games", Huntington Press, Las Vegas, NV (2008)

Index

- Ace adjustment, 32
- Back-counting, 29
- Balanced, 19
- Bankroll, 22
- Base bet, 21
- Basic Strategy, 9
- Betting correlation, BC, 18
- Blackjack, 6
- Break, 7
- Brownian bridge, 106
- Bust, 7
- Camouflage, 27
- Capital, 21
- Card counting, 15
- Catch, 22, 35
- Composition-dependent play, 14
- Count per deck, 17
- Counter Basic Strategy, CBS, 35
- Counting vector, 16
- Coverage, 25
- DAS, 8
- Departure, 29
- Depth, 67
- Double down, 7
- Draw, 7
- Effective true count, 35
- Effective yield, 22
- Eigenmode, 80
- Eigenvalue, 80
- Eigenvector, 80
- Element, 16
- Entry, 29
- Exit, 29
- Expected return, 10
- Generic Strategy, 9
- Hard, 7
- Hermite polynomials, 72
- Hi-Lo, 17
- Hi-Opt, 17
- Hit, 7
- HJY, 24
- Hole card, 6
- House, 5
- Illustrious 18, 35
- Initial Running Count, IRC, 78
- Insurance, 6
- Key count, 19
- Lifetimer, 23
- Marker, 5
- Method of Stationary Phase, 80
- Minmax, MM, 24
- Natural, 6
- Optimal Basic Strategy, 9
- Pack, 5
- Peek, 6
- Penetration, 8
- Pivot, 19
- Play parameters, 31
- Play-all, 29
- Push, 8
- Ramp, 22
- Reordering Theorem, 60
- Reshuffle, 8
- Reshuffle round, 10
- Resplit, 7
- Return on capital employed, ROCE, 43

Return on investment, ROI, 13
Round, 8
Ruin, 22
Ruin protection, RP, 26
Running count, 16

Shoe, 5
Side-count, 32
Soft, 7
Split, 7
Spread, 22
Staircase, 22
Stake, 22
Stand, 7
Stationary Phase, Method of, 80
Stick, 7
Strategy indices, 31

Surrender, 8
Survival, 22

Table-hopping, 29
Trip capital, 22
True count, 17

Unbalanced, 19
Upcard, 6

Weekender, 23
Wong Benchmark game, 20
Wonging, 29

Yield, 21
Yield ratio, 21
Yield reduction factor, YRF, 22

Index of Symbols

A	68	R_D, R_∞, R'	61	ρ	67
$A_{0,1,2}$	94	\overline{R}	84	σ	54
B	75	$R_{\text{opt}}, R_{\text{bas}}$	116	τ	106
B_\pm	93	R_I	62, 122	$\overline{\tau}$	110
c	47	s, s'	47	φ	100
c_0	81	t	53	Φ	80
\overline{c}	87	T	52, 53	ψ	117
c_n	74	u	49	Ψ	89
$C^{(4)}$	60	v	84, 85	Ω, Ω^*	57, 100
C_N	81	\mathbf{v}_b	77	\wp, \wp^*	55, 56
C_b	77	V	102	$\hat{\nabla}$	69
C_p	118	W_N	89		
d	47	Y	75		
$\hat{\mathbf{d}}$	69	α	68		
D	67	α^*	77		
$\hat{\mathbf{D}}$	70	β	93		
E	113	γ	68		
f	67	$\hat{\gamma}$	69		
F	75	γ_0	78		
G	50	γ_p, γ_k	79		
H	53, 103	Γ	54		
H_n	72	Γ_E	107		
J	114	$\delta(x)$	68		
K, K^*	48	$\delta(i, j)$	49, 84		
L_N	82	Δ	67		
\mathbf{m}	67	Δ_R	94		
M	59	ε	59		
N, N'	56	η	120		
p	68	Θ	82		
p_N, \hat{p}_N	82	Λ	68		
P, P'	47	ξ	53, 81		
q	89	Ξ	71		
Q	116	π	115		
R	49–52	Π, Π^*	49		