Chapter 9

THE KELLY CRITERION IN BLACKJACK SPORTS BETTING, AND THE STOCK MARKET*

EDWARD O. THORP

Edward O. Thorp and Associates, Newport Beach, CA 92660, USA

Contents

* Paper presented at: The 10th International Conference on Gambling and Risk Taking, Montreal, June 1997, published in: Finding the Edge: Mathematical Analysis of Casino Games, edited by O. Vancura, J.A. Cornelius, W.R. Eadington, 2000. Corrections added April 20, 2005.

Handbook of Asset and Liability Management, Volume 1
Edited by S.A. Zenios and W.T. Ziemba

DOI: 10.1016/S1872-0978(06)01009-X

Abstract

The central problem for gamblers is to find positive expectation bets. But the gambler also needs to know how to manage his money, i.e., how much to bet. In the stock market (more inclusively, the securities markets) the problem is similar but more complex. The gambler, who is now an "investor", looks for "excess risk adjusted return". In both these settings, we explore the use of the Kelly criterion, which is to maximize the expected value of the logarithm of wealth ("maximize expected logarithmic utility"). The criterion is known to economists and financial theorists by names such as the "geometric mean maximizing portfolio strategy", maximizing logarithmic utility, the growth-optimal strategy, the capital growth criterion, etc. The author initiated the practical application of the Kelly criterion by using it for card counting in blackjack. We will present some useful formulas and methods to answer various natural questions about it that arise in blackjack and other gambling games. Then we illustrate its recent use in a successful casino sports betting system. Finally, we discuss its application to the securities markets where it has helped the author to make a thirty year total of 80 billion dollars worth of "bets".

Keywords

Kelly criterion, betting, long run investing, portfolio allocation, logarithmic utility, capital growth

JEL classification: C61, D81, G1

1. Introduction

The fundamental problem in gambling is to find positive expectation betting opportunities. The analogous problem in investing is to find investments with excess risk-adjusted expected rates of return. Once these favorable opportunities have been identified, the gambler or investor must decide how much of his capital to bet. This is the problem which we consider here. It has been of interest at least since the eighteenth century discussion of the St. Petersburg Paradox (Feller, 1966) by Daniel Bernoulli.

One approach is to choose a goal, such as to minimize the probability of total loss within a specified number of trials, N. Another example would be to maximize the probability of reaching a fixed goal on or before N trials (Browne, 1996).

A different approach, much studied by economists and others, is to value money using a utility function. These are typically defined for all non-negative real numbers, have extended real number values, and are non-decreasing (more money is at least as good as less money). Some examples are $U(x) = x^a$, $0 \leqslant a < \infty$, and $U(x) = \log x$, where log means $\log_e$, and $\log 0 = -\infty$. Once a utility function is specified, the object is to maximize the expected value of the utility of wealth.

Daniel Bernoulli used the utility function $\log x$ to "solve" the St. Petersburg Paradox. (But his solution does not eliminate the paradox because every utility function which is unbounded above, including log, has a modified version of the St. Petersburg Paradox.) The utility function $\log x$ was revisited by Kelly (1956) where he showed that it had some remarkable properties. These were elaborated and generalized in an important paper by Breiman (1961). Markowitz (1959) illustrates the application to securities. For a discussion of the Kelly criterion (the "geometric mean criterion") from a finance point of view, see McEnally (1986). He also includes additional history and references.

I was introduced to the Kelly paper by Claude Shannon at M.I.T. in 1960, shortly after I had created the mathematical theory of card counting at casino blackjack. Kelly's criterion was a bet on each trial so as to maximize $E \log X$, the expected value of the logarithm of the (random variable) capital X. I used it in actual play and introduced it to the gambling community in the first edition of Beat the Dealer (Thorp, 1962). If all blackjack bets paid even money, had positive expectation and were independent, the resulting Kelly betting recipe when playing one hand at a time would be extremely simple: bet a fraction of your current capital equal to your expectation. This is modified somewhat in practice (generally down) to allow for having to make some negative expectation "waiting bets", for the higher variance due to the occurrence of payoffs greater than one to one, and when more than one hand is played at a time.

Here are the properties that made the Kelly criterion so appealing. For ease of understanding, we illustrate using the simplest case, coin tossing, but the concepts and conclusions generalize greatly.

2. Coin tossing

Imagine that we are faced with an infinitely wealthy opponent who will wager even money bets made on repeated independent trials of a biased coin. Further, suppose that on each trial our win probability is $p > 1/2$ and the probability of losing is $q = 1 - p$. Our initial capital is X_0. Suppose we choose the goal of maximizing the expected value $E(X_n)$ after n trials. How much should we bet, B_k, on the kth trial? Letting $T_k = 1$ if the kth trial is a win and $T_k = -1$ if it is a loss, then $X_k = X_{k-1} + T_k B_k$ for $k = 1, 2, 3, \ldots,$ and $X_n = X_0 + \sum_{k=1}^{n} T_k B_k$. Then

$$E(X_n) = X_0 + \sum_{k=1}^{n} E(B_k T_k) = X_0 + \sum_{k=1}^{n} (p - q) E(B_k).$$

Since the game has a positive expectation, i.e., $p - q > 0$ in this even payoff situation, then in order to maximize $E(X_n)$ we would want to maximize $E(B_k)$ at each trial. Thus, to maximize expected gain we should bet *all of our resources* at each trial. Thus $B_1 = X_0$ and if we win the first bet, $B_2 = 2X_0$, etc. However, the probability of ruin is given by $1 - p^n$ and with $p < 1$, $\lim_{n\to\infty}[1 - p^n] = 1$ so ruin is almost sure. Thus the "bold" criterion of betting to maximize expected gain is usually undesirable.

Likewise, if we play to minimize the probability of eventual ruin (i.e., "ruin" occurs if $X_k = 0$ on the kth outcome) the well-known gambler's ruin formula in Feller (1966) shows that we minimize ruin by making a *minimum* bet on each trial, but this unfortunately also minimizes the expected gain. Thus "timid" betting is also unattractive.

This suggests an intermediate strategy which is somewhere between maximizing $E(X_n)$ (and assuring ruin) and minimizing the probability of ruin (and minimizing $E(X_n)$). An asymptotically optimal strategy was first proposed by Kelly (1956).

In the coin-tossing game just described, since the probabilities and payoffs for each bet are the same, it seems plausible that an "optimal" strategy will involve always wagering the same fraction f of your bankroll. To make this possible we shall assume from here on that capital is infinitely divisible. This assumption usually does not matter much in the interesting practical applications.

If we bet according to $B_i = f X_{i-1}$, where $0 \leqslant f \leqslant 1$, this is sometimes called "fixed fraction" betting. Where S and F are the number of successes and failures, respectively, in n trials, then our capital after n trials is $X_n = X_0(1+f)^S(1-f)^F$, where $S + F = n$. With f in the interval $0 < f < 1$, $\Pr(X_n = 0) = 0$. Thus "ruin" in the technical sense of the gambler's ruin problem cannot occur. "Ruin" shall henceforth be reinterpreted to mean that for arbitrarily small positive ε, $\lim_{n\to\infty}[\Pr(X_n \leqslant \varepsilon)] = 1$. Even in this sense, as we shall see, ruin *can* occur under certain circumstances.

We note that since

$$\mathrm{e}^{n \log\left[\frac{X_n}{X_0}\right]^{1/n}} = \frac{X_n}{X_0},$$

the quantity

$$G_n(f) = \log\left[\frac{X_n}{X_0}\right]^{1/n} = \frac{S}{n}\log(1+f) + \frac{F}{n}\log(1-f)$$

measures the exponential rate of increase per trial. Kelly chose to maximize the expected value of the growth rate coefficient, $g(f)$, where

$$g(f) = E\left\{\log\left[\frac{X_n}{X_0}\right]^{1/n}\right\} = E\left\{\frac{S}{n}\log(1+f) + \frac{F}{n}\log(1-f)\right\}$$
$$= p\log(1+f) + q\log(1-f).$$

Note that $g(f) = (1/n)E(\log X_n) - (1/n)\log X_0$ so for n fixed, maximizing $g(f)$ is the same as maximizing $E\log X_n$. We usually will talk about maximizing $g(f)$ in the discussion below. Note that

$$g'(f) = \frac{p}{1+f} - \frac{q}{1-f} = \frac{p-q-f}{(1+f)(1-f)} = 0$$

when $f = f^* = p - q$.

Now

$$g''(f) = -p/(1+f)^2 - q/(1-f)^2 < 0$$

so that $g'(f)$ is monotone strictly decreasing on [0, 1). Also $g'(0) = p - q > 0$ and $\lim_{f\to 1^-} g'(f) = -\infty$. Therefore by the continuity of $g'(f)$, $g(f)$ has a unique maximum at $f = f^*$, where $g(f^*) = p\log p + q\log q + \log 2 > 0$. Moreover, $g(0) = 0$ and $\lim_{f\to q^-} g(f) = -\infty$ so there is a unique number $f_c > 0$, where $0 < f^* < f_c < 1$, such that $g(f_c) = 0$. The nature of the function $g(f)$ is now apparent and a graph of $g(f)$ versus f appears as shown in Figure 1.

The following theorem recounts the important advantages of maximizing $g(f)$. The details are omitted here but proofs of (i)–(iii), and (vi) for the simple binomial case can be found in Thorp (1969); more general proofs of these and of (iv) and (v) are in Breiman (1961).

Theorem 1. (i) *If* $g(f) > 0$, *then* $\lim_{n\to\infty} X_n = \infty$ *almost surely, i.e., for each* M, $\Pr[\liminf_{n\to\infty} X_n > M] = 1$;

(ii) *If* $g(f) < 0$, *then* $\lim_{n\to\infty} X_n = 0$ *almost surely*; *i.e., for each* $\varepsilon > 0$, $\Pr[\limsup_{n\to\infty} X_n < \varepsilon] = 1$;

(iii) *If* $g(f) = 0$, *then* $\limsup_{n\to\infty} X_n = \infty$ *a.s. and* $\liminf_{n\to\infty} X_n = 0$ *a.s.*

(iv) *Given a strategy* Φ^* *which maximizes* $E\log X_n$ *and any other "essentially different" strategy* Φ (*not necessarily a fixed fractional betting strategy*), *then* $\lim_{n\to\infty} X_n(\Phi^*)/X_n(\Phi) = \infty$ *a.s.*

(v) *The expected time for the current capital* X_n *to reach any fixed preassigned goal* C *is, asymptotically, least with a strategy which maximizes* $E\log X_n$.

(vi) *Suppose the return on one unit bet on the* ith *trial is the binomial random variable* U_i; *further, suppose that the probability of success is* p_i, *where* $1/2 < p_i < 1$.

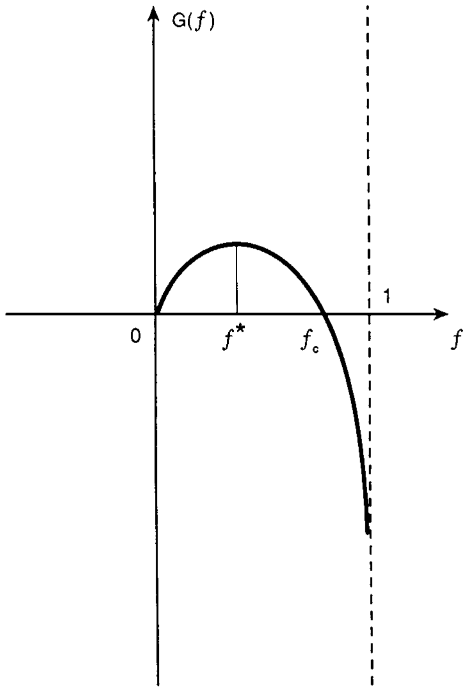

Fig. 1.

Then $E \log X_n$ *is maximized by choosing on each trial the fraction* $f_i^* = p_i - q_i$ *which maximizes* $E \log(1 + f_i U_i)$.

Part (i) shows that, except for a finite number of terms, the player's fortune X_n will exceed any fixed bound M when f is chosen in the interval $(0, f_c)$. But, if $f > f_c$, part (ii) shows that ruin is almost sure. Part (iii) demonstrates that if $f = f_c$, X_n will (almost surely) oscillate randomly between 0 and $+\infty$. Thus, one author's statement that $X_n \to X_0$ as $n \to \infty$, when $f = f_e$, is clearly contradicted. Parts (iv) and (v) show that the Kelly strategy of maximizing $E \log X_n$ is asymptotically optimal by two important criteria. An "essentially different" strategy is one such that the difference $E \ln X_n^* - E \ln X_n$ between the Kelly strategy and the other strategy grows faster than the standard deviation of $\ln X_n^* - \ln X_n$, ensuring $P(\ln X_n^* - \ln X_n > 0) \to 1$. Part (vi) establishes the validity of utilizing the Kelly method of choosing f_i^* on each trial (even if the probabilities change from one trial to the next) in order to maximize $E \log X_n$.

Example 2.1. Player A plays against an infinitely wealthy adversary. Player A wins even money on successive independent flips of a biased coin with a win probability of $p = .53$ (no ties). Player A has an initial capital of X_0 and *capital is infinitely divisible*.

Applying Theorem 1(vi), $f^* = p - q = .53 - .47 = .06$. Thus 6% of current capital should be wagered on each play in order to cause X_n to grow at the fastest rate possible consistent with zero probability of ever going broke. If Player A continually bets a fraction smaller than 6%, X_n will also grow to infinity but the rate will be slower.

If Player A repeatedly bets a fraction larger than 6%, up to the value f_c, the same thing applies. Solving the equation $g(f) = .53\log(1+f) + .47\log(1-f) = 0$ numerically on a computer yields $f_c = .11973^-$. So, if the fraction wagered is more than about 12%, then even though Player A may temporarily experience the pleasure of a faster win rate, eventual downward fluctuations will inexorably drive the values of X_n toward zero. Calculation yields a growth coefficient of $g(f^*) = f(.06) = .001801$ so that after n successive bets the log of Player A's average bankroll will tend to $.001801n$ times as much money as he started with. Setting $.001801n = \log 2$ gives an expected time of about $n = 385$ to double the bankroll.

The Kelly criterion can easily be extended to uneven payoff games. Suppose Player A wins b units for every unit wager. Further, suppose that on each trial the win probability is $p > 0$ and $pb - q > 0$ so the game is advantageous to Player A. Methods similar to those already described can be used to maximize

$$g(f) = E\log(X_n/X_0) = p\log(1+bf) + q\log(1-f).$$

Arguments using calculus yield $f^* = (bp-q)/b$, the optimal fraction of current capital which should be wagered on each play in order to maximize the growth coefficient $g(f)$.

This formula for f^* appeared in Thorp (1984) and was the subject of an April 1997 discussion on the Internet at Stanford Wong's website, http://bj21.com (miscellaneous free pages section). One claim was that one can only lose the amount bet so there was no reason to consider the (simple) generalization of this formula to the situation where a unit wager wins b with probability $p > 0$ and loses a with probability q. Then if the expectation $m \equiv bp - aq > 0$, $f^* > 0$ and $f^* = m/ab$. The generalization does stand up to the objection. One can buy on credit in the financial markets and lose much more than the amount bet. Consider buying commodity futures or selling short a security (where the loss is potentially unlimited). See, e.g., Thorp and Kassouf (1967) for an account of the E.L. Bruce short squeeze.

For purists who insist that these payoffs are not binary, consider selling short a binary digital option. These options are described in Browne (1996).

A criticism sometimes applied to the Kelly strategy is that capital is not, in fact, infinitely divisible. In the real world, bets are multiples of a minimum unit, such as \$1 or \$.01 (penny "slots"). In the securities markets, with computerized records, the minimum unit can be as small as desired. With a minimum allowed bet, "ruin" in the standard sense is always possible. It is not difficult to show, however (see Thorp and Walden, 1966) that if the minimum bet allowed is small relative to the gambler's initial capital, then the probability of ruin in the standard sense is "negligible" and also that the theory herein described is a useful approximation. This section follows Rotando and Thorp (1992).

3. Optimal growth: Kelly criterion formulas for practitioners

Since the Kelly criterion asymptotically maximizes the expected growth rate of wealth, it is often called the optimal growth strategy. It is interesting to compare it with the other fixed fraction strategies. I will present some results that I have found useful in practice. My object is to do so in a way that is simple and easily understood. These results have come mostly from sitting and thinking about "interesting questions". I have not made a thorough literature search but I know that some of these results have been previously published and in greater mathematical generality. See, e.g., Browne (1996, 1997) and the references therein.

3.1. The probability of reaching a fixed goal on or before n trials

We first assume coin tossing. We begin by noting a related result for standard Brownian motion. Howard Tucker showed me this in 1974 and it is probably the most useful single fact I know for dealing with diverse problems in gambling and in the theory of financial derivatives.

For standard Brownian motion $X(t)$, we have

$$P\bigl(\sup\bigl[X(t) - (at + b)\bigr] \geqslant 0,\ 0 \leqslant t \leqslant T\bigr) = N(-\alpha - \beta) + \mathrm{e}^{-2ab} N(\alpha - \beta) \tag{3.1}$$

where $\alpha = a\sqrt{T}$ and $\beta = b/\sqrt{T}$. See Figure 2. See Appendix B for Tucker's derivation of (3.1).

In our application $a < 0, b > 0$ so we expect $\lim_{T\to\infty} P(X(t) \geqslant at+b,\ 0 \leqslant t \leqslant T) = 1$.

Let f be the fraction bet. Assume independent identically distributed (i.d.d.) trials Y_i, $i = 1, \ldots, n$, with $P(Y_i = 1) = p > 1/2$, $P(Y_i = -1) = q < 1/2$; also assume $p < 1$ to avoid the trivial case $p = 1$.

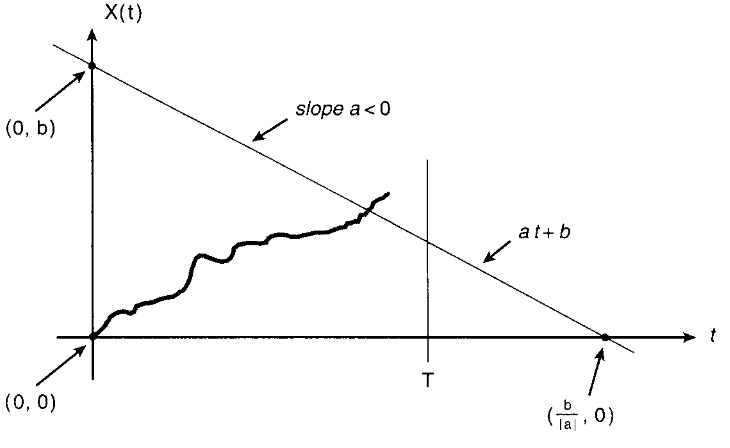

Fig. 2.

Bet a fixed fraction f, $0 < f < 1$, at each trial. Let V_k be the value of the gambler or investor's bankroll after k trials, with initial value V_0. Choose initial stake $V_0 = 1$ (without loss of generality); number of trials n; goal $C > 1$.

What is the probability that $V_k \geqslant C$ for some k, $1 \leqslant k \leqslant n$? This is the same as the probability that $\log V_k \geqslant \log C$ for some k, $1 \leqslant k \leqslant n$. Letting $\ln = \log_e$ we have:

$$V_k = \prod_{i-1}^{k}(1 + Y_i f) \quad \text{and}$$

$$\ln V_k = \sum_{i=1}^{k} \ln(1 + Y_i f),$$

$$E \ln V_k = \sum_{i=1}^{k} E \ln(1 + Y_i f),$$

$$\mathrm{Var}(\ln V_k) = \sum_{i=1}^{k} \mathrm{Var}\big(\ln(1 + Y_i f)\big),$$

$$E \ln(1 + Y_i f) = p \ln(1 + f) + q \ln(1 - f) \equiv m \equiv g(f),$$

$$\begin{aligned}
\mathrm{Var}\big[\ln(1 + Y_i f)\big] &= p\big[\ln(1 + f)\big]^2 + q\big[\ln(1 - f)\big]^2 - m^2 \\
&= \big(p - p^2\big)\big[\ln(1 + f)\big]^2 + \big(q - q^2\big)\big[\ln(1 - f)\big]^2 \\
&\quad - 2pq \ln(1 + f) \ln(1 - f) \\
&= pq\big\{\big[\ln(1 + f)\big]^2 - 2 \ln(1 + f) \ln(1 - f) + \big[\ln(1 - f)\big]^2\big\} \\
&= pq\big\{\ln\big[(1 + f)/(1 - f)\big]\big\}^2 \equiv s^2.
\end{aligned}$$

Drift in n trials: mn.
Variance in n trials: $s^2 n$.

$$\ln V_k \geqslant \ln C,\ \ 1 \leqslant k \leqslant n, \quad \text{iff}$$

$$\sum_{i=1}^{k} \ln(1 + Y_i f) \geqslant \ln C,\ \ 1 \leqslant k \leqslant n, \quad \text{iff}$$

$$S_k \equiv \sum_{i=1}^{k} \big[\ln(1 + Y_i f) - m\big] \geqslant \ln C - mk, \quad 1 \leqslant k \leqslant n,$$

$$E(S_k) = 0, \quad \mathrm{Var}(S_k) = s^2 k.$$

We want $\mathrm{Prob}(S_k \geqslant \ln C - mk,\ 1 \leqslant k \leqslant n)$.

Now we use our Brownian motion formula to approximate S_n by $\mathrm{Prob}(X(t) \geqslant \ln C - mt/s^2,\ 1 \leqslant t \leqslant s^2 n)$ where each term of S_n is approximated by an $X(t)$, drift 0 and variance s^2 ($0 \leqslant t \leqslant s^2,\ s^2 \leqslant t \leqslant 2s^2,\ \ldots,\ (n-1)s^2 \leqslant t \leqslant ns^2$). Note: the approximation is only "good" for "large" n.

Then in the original formula (3.1):

$$T = s^2 n,$$
$$b = \ln C,$$
$$a = -m/s^2,$$
$$\alpha = a\sqrt{T} = -m\sqrt{n}/s,$$
$$\beta = b/\sqrt{T} = \ln C/s\sqrt{n}.$$

Example 3.1.

$$C = 2,$$
$$n = 10^4,$$
$$p = .51,$$
$$q = .49,$$
$$f = .0117,$$
$$m = .000165561,$$
$$s^2 = .000136848.$$

Then

$$P(\cdot) = .9142.$$

Example 3.2. Repeat with

$$f = .02,$$

then

$$m = .000200013, \quad s^2 = .000399947 \quad \text{and} \quad P(\cdot) = .9214.$$

3.2. The probability of ever being reduced to a fraction x of this initial bankroll

This is a question that is of great concern to gamblers and investors. It is readily answered, approximately, by our previous methods.

Using the notation of the previous section, we want $P(V_k \leqslant x$ for some $k,\ 1 \leqslant k \leqslant \infty)$. Similar methods yield the (much simpler) continuous approximation formula:

$$\text{Prob}(\cdot) = \mathrm{e}^{2ab} \quad \text{where } a = -m/s^2 \text{ and } b = -\ln x$$

which can be rewritten as

$$\text{Prob}(\cdot) = x^{\wedge}(2m/s^2) \quad \text{where } \wedge \text{ means exponentiation.} \tag{3.2}$$

Example 3.3.

$$p = .51, \quad f = f^* = .02,$$
$$2m/s = 1.0002,$$
$$\text{Prob}(\cdot) \doteq x.$$

We will see in Section 7 that for the limiting continuous approximation and the Kelly optimal fraction f^*, $P(V_k(f^*) \leqslant x \text{ for some } k \geqslant 1) = x$.

My experience has been that most cautious gamblers or investors who use Kelly find the frequency of substantial bankroll reduction to be uncomfortably large. We can see why now. To reduce this, they tend to prefer somewhat less than the full betting fraction f^*. This also offers a margin of safety in case the betting situations are less favorable than believed. The penalty in reduced growth rate is not severe for moderate underbetting. We discuss this further in Section 7.

3.3. The probability of being at or above a specified value at the end of a specified number of trials

Hecht (1995) suggested setting this probability as the goal and used a computerized search method to determine optimal (by this criterion) fixed fractions for $p - q = .02$ and various c, n and specified success probabilities.

This is a much easier problem than the similar sounding in Section 3.1. We have for the probability that $X(T)$ at the end exceeds the goal:

$$\begin{aligned} P\big(X(T) \geqslant aT + b\big) &= \frac{1}{\sqrt{2\pi T}} \int_{aT+b}^{\infty} \exp\big\{-x^2/2T\big\}\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi T}} \int_{aT^{1/2}+bT^{-1/2}}^{\infty} \exp\big\{-u^2/2\big\}\,\mathrm{d}u \end{aligned}$$

where $u = x/\sqrt{T}$ so $x = aT + b$ gives $u\sqrt{T} = aT + b$ and $U = aT^{1/2} + bT^{-1/2}$. The integral equals

$$\begin{aligned} 1 - N\big(aT^{1/2} + bT^{-1/2}\big) &= N\big(-\big(aT^{1/2} + bT^{-1/2}\big)\big) \\ &= 1 - N(\alpha + \beta) = N(-\alpha - \beta). \end{aligned} \tag{3.3}$$

For example (3.1) $f = .0117$ and $P = .7947$. For example (3.2) $P = .7433$. Example (3.1) is for the Hecht optimal fraction and example (3.2) is for the Kelly optimal fraction. Note the difference in P values.

Our numerical results are consistent with Hecht's simulations in the instances we have checked.

Browne (1996) has given an elegant continuous approximation solution to the problem: What is the strategy which maximizes the probability of reaching a fixed goal C on or before a specified time n and what is the corresponding probability of success? Note that the optimal strategy will in general involve betting varying fractions, depending on the time remaining and the distance to the goal.

As an extreme example, just to make the point, suppose $n = 1$ and $C = 2$. If $X_0 < 1$ then no strategy works and the probability of success is 0. But if $1 \leqslant X_0 < 2$ one should bet at least $2 - X_0$, thus any fraction $f \geqslant (2 - X_0)/X_0$, for a success probability of p. Another extreme example: $n = 10$, $C = 2^{10} = 1024$, $X_0 = 1$. Then the only strategy which can succeed is to bet $f = 1$ on every trial. The probability of success is p^{10} for this strategy and 0 for all others (if $p < 1$), including Kelly.

3.4. *Continuous approximation of expected time to reach a goal*

According to Theorem 1(v), the optimal growth strategy asymptotically minimizes the expected time to reach a goal. Here is what this means. Suppose for goal C that $m(C)$ is the greatest lower bound over all strategies for the expected time to reach C. Suppose $t^*(C)$ is the expected time using the Kelly strategy. Then $\lim_{C\to\infty}(t^*(c)/m(c)) = 1$.

The continuous approximation to the expected number of trials to reach the goal $C > X_0 = 1$ is

$$n(C, f) = (\ln C)/g(f)$$

where f is any fixed fraction strategy. Appendix C has the derivation. Now $g(f)$ has a unique maximum at $g(f^*)$ so $n(C, f)$ has a unique minimum at $f = f^*$. Moreover, we can see how much longer it takes, on average, to reach C if one deviates from f^*.

3.5. *Comparing fixed fraction strategies: the probability that one strategy leads another after n trials*

Theorem 1(iv) says that wealth using the Kelly strategy will tend, in the long run, to an infinitely large multiple of wealth using any "essentially different" strategy. It can be shown that any fixed $f \neq f^*$ is an "essentially different" strategy. This leads to the question of how fast the Kelly strategy gets ahead of another fixed fraction strategy, and more generally, how fast one fixed fraction strategy gets ahead of (or behind) another.

If W_n is the number of wins in n trials and $n - W_n$ is the number of losses,

$$G(f) = (W_n/n)\ln(l + f) + (1 - W_n/n)\ln(1 - f)$$

is the actual (random variable) growth coefficient.

As we saw, its expectation is

$$g(f) = E\big(G(f)\big) = p\log(1 + f) + q\log(1 - f) \tag{3.4}$$

and the variance of $G(f)$ is

$$\operatorname{Var} G(f) = \big((pq)/n\big)\big\{\ln\big((1 + f)/(1 - f)\big)\big\}^2 \tag{3.5}$$

and it follows that $G(f)$, which has the form $G(f) = a(\sum T_k)/n + b$, is approximately normally distributed with mean $g(f)$ and variance $\operatorname{Var} G(f)$. This enables us to give the distribution of X_n and once again answer the question of Section 3.3. We illustrate this with an example.

Example 3.4.

$$p = .51, \quad q = .49, \quad f^* = .02, \quad N = 10{,}000 \quad \text{and}$$
$$s = \text{standard deviation of } G(f)$$

g/s	f	g	s	$\Pr(G(f) \leqslant 0)$
1.5	.01	.000150004	.0001	.067
1.0	.02	.000200013	.0002	.159
.5	.03	.000149977	.0003	.309

Continuing, we find the distribution of $G(f_2) - G(f_1)$. We consider two cases.

Case 1. The *same* game

Here we assume both players are betting on the same trials, e.g., betting on the same coin tosses, or on the same series of hands at blackjack, or on the same games with the same odds at the same sports book. In the stock market, both players could invest in the same "security" at the same time, e.g., a no-load S&P 500 index mutual fund.

We find

$$E\big(G(f_2) - G(f_1)\big) = p \log\big((1+f_2)/(1+f_1)\big) + q \log\big((1-f_2)/(1-f_1)\big)$$

and

$$\operatorname{Var}\big(G(f_2) - G(f_1)\big) = (pq/n)\left\{\log\left[\left(\frac{1+f_2}{1-f_2}\right)\left(\frac{1-f_1}{1+f_1}\right)\right]\right\}^2$$

where $G(f_2) - G(f_1)$ is approximately normally distributed with this mean and variance.

Case 2. Identically distributed independent games

This corresponds to betting on two different series of tosses with the same coin. $E(G(f_2) - G(f_1))$ is as before. But now $\operatorname{Var}(G(f_2) - G(f_1)) = \operatorname{Var}(G(f_2)) + \operatorname{Var}(G(f_1))$ because $G(f_2)$ and $G(f_1)$ are now independent. Thus

$$\operatorname{Var}\big(G(f_2) - G(f_1)\big) = (pq/n)\left\{\left[\log\left(\frac{1+f_2}{1-f_2}\right)\right]^2 + \left[\log\left(\frac{1+f_1}{1-f_1}\right)^2\right]\right\}.$$

Let

$$a = \log\left(\frac{1+f_1}{1-f_1}\right), \qquad b = \log\left(\frac{1+f_2}{1-f_2}\right).$$

Then in Case 1, $V_1 = (pq/n)(a-b)^2$ and in Case 2, $V_2 = (pq/n)(a^2+b^2)$ and since $a, b > 0$, $V_1 < V_2$ as expected. We can now compare the Kelly strategy with other fixed fractions to determine the probability that Kelly leads after n trials. Note that this probability is always greater than $1/2$ (within the accuracy limits of the continuous approximation, which is the approximation of the binomial distribution by the normal, with its well known and thoroughly studied properties) because $g(f^*) - g(f) > 0$ where $f^* = p - q$ and $f \neq f^*$ is some alternative. This can fail to be true for small n, where the approximation is poor. As an extreme example to make the point, if $n = 1$, any $f > f^*$ beats Kelly with probability $p > 1/2$. If instead $n = 2$, $f > f^*$ wins with probability p^2 and $p^2 > 1/2$ if $p > 1/\sqrt{2} \doteq .7071$. Also, $f < f^*$ wins with probability $1 - p^2$ and $1 - p^2 > 1/2$ if $p^2 < 1/2$, i.e., $p < 1/\sqrt{2} = .7071$. So when $n = 2$, Kelly always loses more than half the time to some other f unless $p = 1/\sqrt{2}$.

We now have the formulas we need to explore many practical applications of the Kelly criterion.

4. The long run: when will the Kelly strategy "dominate"?

The late John Leib wrote several articles for Blackjack Forum which were critical of the Kelly criterion. He was much bemused by "the long run". What is it and when, if ever, does it happen?

We begin with an example.

Example 4.1.

$$p = .51, \quad n = 10{,}000,$$

V_i and s_i, $i = 1, 2$, are the variance and standard deviation, respectively, for Section 3.5 Cases 1 and 2, and $R = V_2/V_1 = (a^2 + b^2)/(a - b)^2$ so $s_2 = s_1\sqrt{R}$. Table 1 summarizes some results. We can also approximate $\sqrt{R}$ with a power series estimate using only the first term of a and of b: $a \doteq 2f_1$, $b \doteq 2f_2$ so $\sqrt{R} \doteq \sqrt{f_1^2 + f_2^2}/|f_1 - f_2|$. The approximate results, which agree extremely well, are 2.236, 3.606 and 1.581, respectively.

The first two rows show how nearly symmetric the behavior is on each side of the optimal $f^* = .02$. The column $(g_2 - g_1)/s_1$ shows us that $f^* = .02$ only has a .5 standard deviation advantage over its neighbors $f = .01$ and $f = .03$ after $n = 10{,}000$

Table 1
Comparing strategies

f_1	f_2	$g_2 - g_1$	s_1	$(g_2 - g_1)/s_1$	$\sqrt{R}$
.01	.02	.00005001	.00010000	.50	2.236
.03	.02	.00005004	.00010004	.50	3.604
.03	.01	.00000003	.00020005	.00013346	1.581

Table 2
The long run: $(g_2 - g_1)/s$ after n trials

f_1	f_2	$n = 10^4$	$n = 4 \times 10^4$	$n = 16 \times 10^4$	$n = 10^6$
.01	.02	.5	1.0	2.0	5.0
.03	.02	.5	1.0	2.0	5.0
.03	.01	.000133	.000267	.000534	.001335

trials. Since this advantage is proportional to $\sqrt{n}$, the column $(g_2 - g_1)/s_1$ from Table 1 gives the results of Table 2.

The factor $\sqrt{R}$ in Table 1 shows how much more slowly f_2 dominates f_1 in Case 2 versus Case 1. The ratio $(g_2 - g_1)/s_2$ is $\sqrt{R}$ times as large so the same level of dominance takes R times as long. When the real world comparisons of strategies for practical reasons often use Case 2 comparisons rather than the more appropriate Case 1 comparisons, the dominance of f^* is further obscured. An example is players with different betting fractions at blackjack. Case 1 corresponds to both betting on the same sequence of hands. Case 2 corresponds to them playing at different tables (not the same table, because Case 2 assumes independence). (Because of the positive correlation between payoffs on hands played at the same table, this is intermediate between Cases 1 and 2.)

It is important to understand that "the long run", i.e., the time it takes for f^* to dominate a specified neighbor by a specified probability, can vary without limit. Each application requires a separate analysis. In cases such as Example 4.1, where dominance is "slow", one might argue that using f^* is not important. As an argument against this, consider two coin-tossing games. In game 1 your edge is 1.0%. In game 2 your edge is 1.1%. With one unit bets, after n trials the difference in expected gain is $E_2 - E_1 = .001n$ with standard deviation s of about $\sqrt{2n}$ hence $(E_2 - E_1)/s \doteq .001\sqrt{n}/\sqrt{2}$ which is 1 when $n = 2 \times 10^6$. So it takes two million trials to have an 84% chance of the game 2 results being better than the game 1 results. Does that mean it's unimportant to select the higher expectation game?

5. Blackjack

For a general discussion of blackjack, see Thorp (1962, 1966), Wong (1994) and Griffin (1979). The Kelly criterion was introduced for blackjack by Thorp (1962). The analysis is more complicated than that of coin tossing because the payoffs are not simply one to one. In particular the variance is generally more than 1 and the Kelly fraction tends to be less than for coin tossing with the same expectation. Moreover, the distribution of various payoffs depends on the player advantage. For instance the frequency of pair splitting, doubling down, and blackjacks all vary as the advantage changes. By binning the probability of payoff types according to ex ante expectation, and solving the Kelly equations on a computer, a strategy can be found which is as close to optimal as desired.

There are some conceptual subtleties which are noteworthy. To illustrate them we'll simplify to the coin toss model.

At each trial, we have with probability .5 a "favorable situation" with gain or loss X per unit bet such that $P(X = 1) = .51$, $P(X = -1) = .49$ and with probability .5 an unfavorable situation with gain or loss Y per unit bet such that $P(Y = 1) = .49$ and $P(Y = -1) = .51$. We know before we bet whether X or Y applies.

Suppose the player must make small "waiting" bets on the unfavorable situations in order to be able to exploit the favorable situations. On these he will place "large" bets. We consider two cases.

Case 1. Bet f_0 on unfavorable situations and find the optimal f^* for favorable situations. We have

$$g(f) = .5\bigl(.51\log(1+f) + .49\log(1-f)\bigr) + .5\bigl(.49\log(1+f_0) + .51\log(1-f_0)\bigr). \tag{5.1}$$

Since the second expression in (5.1) is constant, f maximizes $g(f)$ if it maximizes the first expression, so $f^* = p - q = .02$, as usual. It is easy to verify that when there is a spectrum of favorable situations the same recipe, $f_i^* = p_i - q_i$ for the ith situation, holds. Again, in actual blackjack f_i^* would be adjusted down somewhat for the greater variance. With an additional constraint such as $f_i \leqslant kf_0$, where k is typically some integral multiple of f_0, representing the betting spread adopted by a prudent player, then the solution is just $f_i \leqslant \min(f_i^*, kf_0)$.

Curiously, a seemingly similar formulation of the betting problem leads to rather different results.

Case 2. Bet f in favorable situations and af in unfavorable situations, $0 \leqslant a \leqslant 1$.

Now the bet sizes in the two situations are linked and both the analysis and results are more complex. We have a Kelly growth rate of

$$g(f) = .5\bigl(.51\log(1+f) + .49\log(1-f)\bigr) + .5\bigl(.49\log(1+af) + .51\log(1-af)\bigr). \tag{5.2}$$

If we choose $a = 0$ (no bet in unfavorable situations) then the maximum value for $g(f)$ is at $f^* = .02$, the usual Kelly fraction.

If we make "waiting bets", corresponding to some value of $a > 0$, this will shift the value of f^* down, perhaps even to 0. The expected gain divided by the expected bet is $.02(1-a)/(1+a)$, $a \geqslant 0$. If $a = 0$ we get .02, as expected. If $a = 1$, we get 0, as expected: this is a fair game and the Kelly fraction is $f^* = 0$. As a increases from 0 to 1 the (optimal) Kelly fraction f^* decreases from .02 to 0. Thus the Kelly fraction for favorable situations is less *in this case* when bets on unfavorable situations reduce the overall advantage of the game.

Arnold Snyder called to my attention the fact that Winston Yamashita had (also) made this point (March 18, 1997) on the "free" pages, miscellaneous section, of Stanford Wong's web site.

Table 3
f^* versus a

a	f^*	a	f^*	a	f^*
0	.0200	1/3	.0120	.7	.0040
.1	.0178	.4	.0103	.8	.0024
.2	.0154	.5	.0080	.9	.0011
.3	.0128	.6	.0059	1.0	.0000

For this example, we find the new f^* for a given value of a, $0 < a < 1$, by solving $g'(f) = 0$. A value of $a = 1/3$, for instance, corresponds to a bet of $1/3$ unit on Y and 1 unit on X, a betting range of 3 to 1. The overall expectation is .01. Calculation shows $f^* = .012001$. Table 3 shows how f^* varies with a.

To understand why Cases 1 and 2 have different f^*, look first at Equation (5.1). The part of $g(f)$ corresponding to the unfavorable situations is fixed when f_0 is fixed. Only the part of $g(f)$ corresponding to the favorable situations is affected by varying f. Thus we maximize $g(f)$ by maximizing it over just the favorable situations. Whatever the result, it is then reduced by a fixed quantity, the part of g containing f_0. On the other hand, in Equation (5.2) both parts of $g(f)$ are affected when f varies, because the fraction af used for unfavorable situations bears the constant ratio a to the fraction f used in favorable situations. Now the first term, for the favorable situations, has a maximum at $f = .02$, and is approximately "flat" nearby. But the second term, for the unfavorable situations, is negative and decreasing moderately rapidly at $f = .02$. Therefore, it we reduce f somewhat, this term increases somewhat, while the first term decreases only very slightly. There is a net gain so we find $f^* < .02$. The greater a is, the more important is the effect of this term so the more we have to reduce f to get f^*, as Table 3 clearly shows. When there is a spectrum of favorable situations the solution is more complex and can be found through standard multivariable optimization techniques.

The more complex Case 2 corresponds to what the serious blackjack player is likely to need to do in practice. He will have to limit his current maximum bet to some multiple of his current minimum bet. As his bankroll increases or decreases, the corresponding bet sizes will increase or decrease proportionately.

6. Sports betting

In 1993 an outstanding young computer science Ph.D. told me about a successful sports betting system that he had developed. Upon review I was convinced. I made suggestions for minor simplifications and improvements. Then we agreed on a field test. We found a person who was extremely likely to always be regarded by the other sports bettors as a novice. I put up a test bankroll of $50,000 and we used the Kelly system to estimate our bet size.

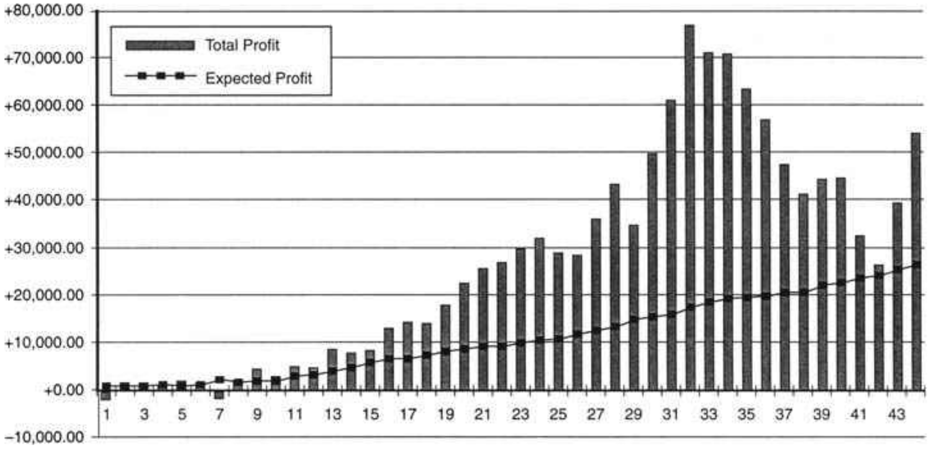

Fig. 3. Betting log Type 2 sports.

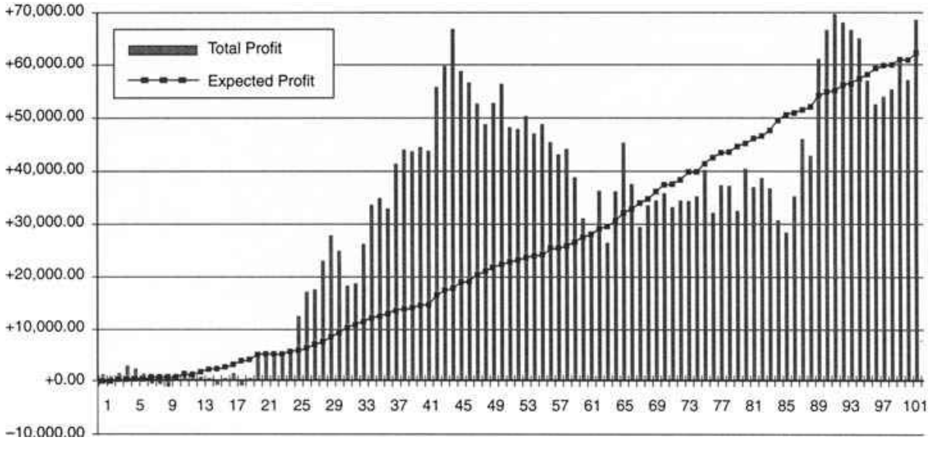

Fig. 4. Betting log Type 1 sports.

We bet on 101 days in the first four and a half months of 1994. The system works for various sports. The results appear in Figures 3 and 4. After 101 days of bets, our \$50,000 bankroll had a profit of \$123,000, about \$68,000 from Type 1 sports and about \$55,000 from Type 2 sports. The expected returns are shown as about \$62,000 for Type 1 and about \$27,000 for Type 2. One might assign the additional \$34,000 actually won to luck. But this is likely to be at most partly true because our expectation estimates from the model were deliberately chosen to be conservative. The reason is that using too large an f^* and overbetting is much more severely penalized than using too small an f^* and underbetting.

Though \$123,000 is a modest sum for some, and insignificant by Wall Street standards, the system performed as predicted and passed its test. We were never more than a few thousand behind. The farthest we had to invade our bankroll to place bets was about \$10,000.

Our typical expectation was about 6% so our total bets ("action") were about \$2,000,000 or about \$20,000 per day. We typically placed from five to fifteen bets a day and bets ranged from a few hundred dollars to several thousand each, increasing as our bankroll grew.

Though we had a net win, the net results by casino varied by chance from a substantial loss to a large win. Particularly hard hit was the "sawdust joint" Little Caesar's. It "died" towards the end of our test and I suspect that sports book losses to us may have expedited its departure.

One feature of sports betting which is of interest to Kelly users is the prospect of betting on several games at once. This also arises in blackjack when (a) a player bets on multiple hands or (b) two or more players share a common bankroll. The standard techniques readily solve such problems. We illustrate with:

Example 6.1. Suppose we bet simultaneously on two independent favorable coins with betting fractions f_1 and f_2 and with success probabilities p_1 and p_2, respectively. Then the expected growth rate is given by

$$\begin{aligned} g(f_1, f_2) = {} & p_1 p_2 \ln(1 + f_1 + f_2) + p_1 q_2 \ln(1 + f_1 - f_2) \\ & + q_1 p_2 \ln(1 - f_1 + f_2) + q_1 q_2 \ln(1 - f_1 - f_2). \end{aligned}$$

To find the optimal f_1^* and f_2^* we solve the simultaneous equations $\partial g/\partial f_1 = 0$ and $\partial g/\partial f_2 = 0$. The result is

$$\begin{aligned} f_1 + f_2 &= \frac{p_1 p_2 - q_1 q_2}{p_1 p_2 + q_1 q_2} \equiv c, \\ f_1 - f_2 &= \frac{p_1 q_2 - q_1 p_2}{p_1 q_2 + q_1 p_2} \equiv d, \\ f_1^* &= (c + d)/2, \qquad f_2^* = (c - d)/2. \end{aligned} \tag{6.1}$$

These equations pass the symmetry check: interchanging 1 and 2 throughout maps the equation set into itself.

An alternate form is instructive. Let $m_i = p_i - q_i$, $i = 1, 2$ so $p_i = (1 + m_i)/2$ and $q_i = (1 - m_i)/2$. Substituting in (6.1) and simplifying leads to:

$$\begin{aligned} c &= \frac{m_1 + m_2}{1 + m_1 m_2}, \qquad d = \frac{m_1 - m_2}{1 - m_1 m_2}, \\ f_1^* &= \frac{m_1(1 - m_2^2)}{1 - m_1^2 m_2^2}, \qquad f_2^* = \frac{m_2(1 - m_1^2)}{1 - m_1^2 m_2^2} \end{aligned} \tag{6.2}$$

which shows clearly the factors by which the f_i^* are each reduced from m_i^*. Since the m_i are typically a few percent, the reduction factors are typically very close to 1.

In the special case $p_1 = p_2 = p$, $d = 0$ and $f^* = f_1^* = f_2^* = c/2 = (p-q)/(2(p^2+q^2))$. Letting $m = p-q$ this may be written $f^* = m/(1+m^2)$ as the optimal fraction to bet on each coin simultaneously, compared to $f^* = m$ to bet on each coin sequentially.

Our simultaneous sports bets were generally on different games and typically not numerous so they were approximately independent and the appropriate fractions were only moderately less than the corresponding single bet fractions. Question: Is this always true for independent simultaneous bets? Simultaneous bets on blackjack hands at different tables are independent but at the same table they have a pairwise correlation that has been estimated at .5 (Griffin, 1979, p. 142). This should substantially reduce the Kelly fraction per hand. The blackjack literature discusses approximations to these problems. On the other hand, correlations between the returns on securities can range from nearly -1 to nearly 1. An extreme correlation often can be exploited to great advantage through the techniques of "hedging". The risk averse investor may be able to acquire combinations of securities where the expectations add and the risks tend to cancel. The optimal betting fraction may be very large.

The next example is a simple illustration of the important effect of covariance on the optimal betting fraction.

Example 6.2. We have two favorable coins as in the previous example but now their outcomes need not be independent. For simplicity assume the special case where the two bets have the same payoff distributions, but with a joint distribution as in Table 4.

Now $c+m+b = (1+m)/2$ so $b = (1-m)/2 - c$ and therefore $0 \leqslant c \leqslant (1-m)/2$.

Calculation shows $\mathrm{Var}(X_i) = 1 - m^2$, $\mathrm{Cor}(X_1, X_2) = 4c - (1-m)^2$ and $\mathrm{Cor}(X_1, X_2) = [4c - (1-m)^2]/(1-m^2)$. The symmetry of the distribution shows that $g(f_1, f_2)$ will have its maximum at $f_1 = f_2 = f$ so we simply need to maximize $g(f) = (c+m)\ln(1+2f) + c\ln(1-2f)$. The result is $f^* = m/(2(2c+m))$. We see that for m fixed, as c decreases from $(1-m)/2$ and $\mathrm{cor}(X_1, X_2) = 1$, to 0 and $\mathrm{cor}(X_1, X_2) = -(1-m)/(1+m)$, f^* for each bet increases from $m/2$ to $1/2$, as in Table 5.

Table 4
Joint distribution of two "identical" favorable coins with correlated outcomes

X_1 :	X_2 : 1	-1
1	$c+m$	b
-1	b	c

Table 5
f^* increases as $\mathrm{Cor}(X_1, X_2)$ decreases

$\mathrm{Cor}(X_1, X_2)$	c	f^*
1	$(1-m)/2$	$m/2$
0	$(1-m^2)/4$	$m/(1+m^2)$
$-(1-m)/(1+m)$	0	$1/2$

It is important to note that for an exact solution or an arbitrarily accurate numerical approximation to the simultaneous bet problem, covariance or correlation information is not enough. We need to use the entire joint distribution to construct the g function.

We stopped sports betting after our successful test for reasons including:

(1) It required a person on site in Nevada.

(2) Large amounts of cash and winning tickets had to be transported between casinos. We believed this was very risky. To the sorrow of others, subsequent events confirmed this.

(3) It was not economically competitive with our other operations.

If it becomes possible to place bets telephonically from out of state and to transfer the corresponding funds electronically, we may be back.

7. Wall street: the biggest game

To illustrate both the Kelly criterion and the size of the securities markets, we return to the study of the effects of correlation as in Example 6.2. Consider the more symmetric and esthetically pleasing pair of bets U_1 and U_2, with joint distribution given in Table 6.

Clearly $0 \leqslant a \leqslant 1/2$ and $\mathrm{Cor}(U_1, U_2) = \mathrm{Cor}(U_1, U_2) = 4a - 1$ increases from -1 to 1 as a increases from 0 to 1/2. Finding a general solution for (f_1^*, f_2^*) appears algebraically complicated (but specific solutions are easy to find numerically), which is why we chose Example 6.2 instead. Even with reduction to the special case $m_1 = m_2 = m$ and the use of symmetry to reduce the problem to finding $f^* = f_1^* = f_2^*$, a general solution is still much less simple. But consider the instance when $a = 0$ so $\mathrm{Cor}(U_1, U_2) = -1$. Then $g(f) = \ln(1 + 2mf)$ which increases without limit as f increases. This pair of bets is a "sure thing" and one should bet as much as possible.

This is a simplified version of the classic arbitrage of securities markets: find a pair of securities which are identical or "equivalent" and trade at disparate prices. Buy the relatively underpriced security and sell short the relatively overpriced security, achieving a correlation of -1 and "locking in" a riskless profit. An example occurred in 1983. My investment partnership bought \$ 330 million worth of "old" AT&T and sold short \$332.5 million worth of when-issued "new" AT&T plus the new "seven sisters" regional telephone companies. Much of this was done in a single trade as part of what was then the largest dollar value block trade ever done on the New York Stock Exchange (December 1, 1983).

In applying the Kelly criterion to the securities markets, we meet new analytic problems. A bet on a security typically has many outcomes rather than just a few, as in

Table 6
Joint distribution of U_1 and U_2

U_1 :	$U_2 : m_2 + 1$	$m_2 - 1$
$m_1 + 1$	a	$1/2 - a$
$m_1 - 1$	$1/2 - a$	a

most gambling situations. This leads to the use of continuous instead of discrete probability distributions. We are led to find f to maximize $g(f) = E\ln(1 + fX) = \int \ln(1 + fx)\,\mathrm{d}P(x)$ where $P(x)$ is a probability measure describing the outcomes. Frequently the problem is to find an optimum portfolio from among n securities, where n may be a "large" number. In this case x and f are n-dimension vectors and fx is their scalar product. We also have constraints. We always need $1 + fx > 0$ so $\ln(\cdot)$ is defined, and $\sum f_i = 1$ (or some $c > 0$) to normalize to a unit (or to a $c > 0$) investment. The maximization problem is generally solvable because $g(f)$ is concave. There may be other constraints as well for some or all i such as $f_i \geqslant 0$ (no short selling), or $f_i \leqslant M_i$ or $f_i \geqslant m_i$ (limits amount invested in ith security), or $\sum |f_i| \leqslant M$ (limits total leverage to meet margin regulations or capital requirements). Note that in some instances there is not enough of a good bet or investment to allow betting the full f^*, so one is forced to underbet, reducing somewhat both the overall growth rate and the risk. This is more a problem in the gaming world than in the much larger securities markets. More on these problems and techniques may be found in the literature.

7.1. Continuous approximation

There is one technique which leads rapidly to interesting results. Let X be a random variable with $P(X = m + s) = P(X = m - s) = .5$. Then $E(X) = m$, $\mathrm{Var}(X) = s^2$. With initial capital V_0, betting fraction f, and return per unit of X, the result is

$$V(f) = V_0\bigl(1 + (1 - f)r + fX\bigr) = V_0\bigl(1 + r + f(X - r)\bigr),$$

where r is the rate of return on the remaining capital, invested in, e.g., Treasury bills. Then

$$\begin{aligned} g(f) &= E\bigl(G(f)\bigr) = E\bigl(\ln\bigl(V(f)/V_0\bigr)\bigr) = E\ln\bigl(1 + r + f(X - r)\bigr) \\ &= .5\ln\bigl(1 + r + f(m - r + s)\bigr) + .5\ln\bigl(1 + r + f(m - r - s)\bigr). \end{aligned}$$

Now subdivide the time interval into n equal independent steps, keeping the same drift and the same total variance. Thus m, s^2 and r are replaced by m/n, s^2/n and r/n, respectively. We have n independent X_i, $i = 1, \ldots, n$, with

$$P\bigl(X_i = m/n + sn^{-1/2}\bigr) = P\bigl(X_i = m/n - sn^{-1/2}\bigr) = .5.$$

Then

$$V_n(f)/V_0 = \prod_{i=1}^{n}\bigl(1 + (1 - f)r + fX_i\bigr).$$

Taking $E(\log(\cdot))$ of both sides gives $g(f)$. Expanding the result in a power series leads to

$$g(f) = r + f(m - r) - s^2 f^2/2 + \mathrm{O}\bigl(n^{-1/2}\bigr) \tag{7.1}$$

where $\mathrm{O}(n^{-1/2})$ has the property $n^{1/2}\mathrm{O}(n^{-1/2})$ is bounded as $n \to \infty$. Letting $n \to \infty$ in (7.1) we have

$$g_\infty(f) \equiv r + f(m - r) - s^2 f^2/2. \tag{7.2}$$

The limit $V \equiv V_\infty(f)$ of $V_n(f)$ as $n \to \infty$ corresponds to a log normal diffusion process, which is a well-known model for securities prices. The "security" here has instantaneous drift rate m, variance rate s^2, and the riskless investment of "cash" earns at an instantaneous rate r. Then $g_\infty(f)$ in (7.2) is the (instantaneous) growth rate of capital with investment or betting fraction f. There is nothing special about our choice of the random variable X. Any bounded random variable with mean $E(X) = m$ and variance $\mathrm{Var}(X) = s^2$ will lead to the same result. Note that f no longer needs to be less than or equal to 1. The usual problems, with $\log(\cdot)$ being undefined for negative arguments, have disappeared. Also, $f < 0$ causes no problems. This simply corresponds to selling the security short. If $m < r$ this could be advantageous. Note further that the investor who follows the policy f must now adjust his investment "instantaneously". In practice this means adjusting in tiny increments whenever there is a small change in V. This idealization appears in option theory. It is well known and does not prevent the practical application of the theory (Black and Scholes, 1973). Our previous growth functions for finite sized betting steps were approximately parabolic in a neighborhood of f^* and often in a range up to $0 \leqslant f \leqslant 2f^*$, where also often $2f^* \doteq f_c$. Now with the limiting case (7.2), $g_\infty(f)$ is exactly parabolic and very easy to study.

Lognormality of $V(f)/V_0$ means $\log(V(f)/V_0)$ is $N(M, S^2)$ distributed, with mean $M = g_\infty(f)t$ and variance $S^2 = \mathrm{Var}(G_\infty(f))t$ for any time t. From this we can determine, for instance, the expected capital growth and the time t_k required for $V(f)$ to be at least k standard deviations above V_0. First, we can show by our previous methods that $\mathrm{Var}(G_\infty(f)) = s^2 f^2$, hence $\mathrm{Sdev}(G_\infty(f)) = sf$. Solving $t_k g_\infty = k t_k^{1/2}\,\mathrm{Sdev}(G_\infty(f))$ gives $t_k g_\infty^2$ hence the expected capital growth $t_k g_\infty$, from which we find t_k. The results are summarized in Equations (7.3).

$$
\begin{aligned}
&f^* = (m-r)/s^2, \qquad g_\infty(f) = r + f(m-r) - s^2 f^2/2,\\
&g_\infty(f^*) = (m-r)^2/2s^2 + r,\\
&\mathrm{Var}\big(G_\infty(f)\big) = s^2 f^2, \qquad \mathrm{Sdev}\big(G_\infty(f)\big) = sf,\\
&t_k g_\infty(f) = k^2 s^2 f^2 / g_\infty,\\
&t_k = k^2 s^2 f^2 / g_\infty^2.
\end{aligned} \tag{7.3}
$$

Examination of the expressions for $t_k g_\infty(f)$ and t_k show that each one increases as f increases, for $0 \leqslant f < f_+$ where f_+ is the positive root of $s^2 f^2/2 - (m-r)f - r = 0$ and $f_+ > 2f^*$.

Comment: The capital asset pricing model (CAPM) says that the market portfolio lies on the Markowitz efficient frontier E in the (s, m) plane at a (generally) unique point $P = (s_0, m_0)$ such that the line determined by P and $(s = 0, m = r)$ is tangent to E (at P). The slope of this line is the Sharpe ratio $S = (m_0 - r_0)/s_0$ and from (7.3) $g_\infty(f^*) = S^2/2 + r$ so the maximum growth rate $g_\infty(f^*)$ depends, for fixed r, only on the Sharpe ratio. (See Quaife (1995).) Again from (7.3), $f^* = 1$ when $m = r + s^2$ in which case the Kelly investor will select the market portfolio without borrowing or lending. If $m > r + s^2$ the Kelly investor will use leverage and if $m < r + s^2$ he will invest partly in T-bills and partly in the market portfolio. Thus the Kelly investor will dynamically reallocate as f^* changes over time because of fluctuations in the forecast m, r and s^2, as well as in the prices of the portfolio securities.

From (7.3), $g_\infty(1) = m - s^2/2$ so the portfolios in the (s, m) plane satisfying $m - s^2/2 = C$, where C is a constant, all have the same growth rate. In the continuous approximation, the Kelly investor appears to have the utility function $U(s, m) = m - s^2/2$. Thus, for any (closed, bounded) set of portfolios, the best portfolios are exactly those in the subset that maximizes the one parameter family $m - s^2/2 = C$. See Kritzman (1998), for an elementary introduction to related ideas.

Example 7.1. The long run revisited. For this example let $r = 0$. Then the basic equations (7.3) simplify to

$$\begin{aligned}
&r = 0: \quad f^* = m/s^2, \quad g_\infty(f) = mf - s^2 f^2/2,\\
&g_\infty(f^*) = m^2/2s^2,\\
&\mathrm{Var}\big(G_\infty(f)\big) = s^2 f^2, \qquad \mathrm{Sdev}\big(G_\infty(f)\big) = sf.
\end{aligned} \tag{7.4}$$

How long will it take for $V(f^*) \geqslant V_0$ with a specified probability? How about $V(f^*/2)$? To find the time t needed for $V(f) \geqslant V_0$ at the k standard deviations level of significance ($k = 1$, $P = 84\%$; $k = 2$, $P = 98\%$, etc.) we solve for $t \equiv t_k$:

$$t g_\infty(f) = k t^{1/2}\,\mathrm{Sdev}\big(G_\infty(f)\big). \tag{7.5}$$

We get more insight by normalizing all f with f^*. Setting $f = cf^*$ throughout, we find when $r = 0$

$$\begin{aligned}
&r = 0: \quad f^* = m/s^2, \quad f = cm/s^2,\\
&g_\infty(cf^*) = m^2\big(c - c^2/2\big)/s^2,\\
&\mathrm{Sdev}\big(G_\infty(cf^*)\big) = cm/s,\\
&t g_\infty(cf^*) = k^2 c/(1 - c/2),\\
&t(k, cf^*) = k^2 s^2/\big(m^2(1 - c/2)^2\big).
\end{aligned} \tag{7.6}$$

Equations (7.6) contain a remarkable result: $V(f) \geqslant V_0$ at the k standard deviation level of significance occurs when expected capital growth $t g_\infty = k^2 c/(1 - c/2)$ and this result is *independent of m and s*. For $f = f^*$ ($c = 1$ in (7.6)), this happens for $k = 1$ at $t g_\infty = 2$ corresponding to $V = V_0 e^2$ and at $k = 2$ for $t g_\infty = 8$ corresponding to $V = V_0 e^8$. Now $e^8 \doteq 2981$ and at a 10% annual (instantaneous) growth rate, it takes 80 years to have a probability of 98% for $V \geqslant V_0$. At a 20% annual instantaneous rate it takes 40 years. However, for $f = f^*/2$, the number for $k = 1$ and 2 are $t g_\infty = 2/3$ and $8/3$, respectively, just $1/3$ as large. So the waiting times for $\mathrm{Prob}(V \geqslant V_0)$ to exceed 84% and 98% become 6.7 years and 26.7 years, respectively, and the expected growth rate is reduced to $3/4$ of that for f^*.

Comment: Fractional Kelly versus Kelly when $r = 0$

From Equations (7.6) we see that $g_\infty(cf^*)/g_\infty(f^*) = c(2-c)$, $0 \leqslant c < \infty$, showing how the growth rate relative to the maximum varies with c. The relative risk $\text{Sdev}(G_\infty(cf^*))/\text{Sdev}(G_\infty(f^*)) = c$ and the relative time to achieve the same expected total growth is $1/c(2-c)$, $0 < c < 2$. Thus the relative "spread" for the same expected total growth is $1/(2-c)$, $0 < c < 2$. Thus, even by choosing c very small, the spread around a given expected growth cannot be reduced by $1/2$. The corresponding results are not quite as simple when $r > 0$.

7.2. *The (almost) real world*

Assume that prices change "continuously" (no "jumps"), that portfolios may be revised "continuously", and that there are no transactions costs (market impact, commissions, "overhead"), or taxes (Federal, State, city, exchange, etc.). Then our previous model applies.

Example 7.2. The S&P 500 Index. Using historical data we make the rough estimates $m = .11$, $s = .15$, $r = .06$. The equations we need for $r \neq 0$ are the generalizations of (7.6) to $r \neq 0$ and $f = cf^*$, which follow from (7.3):

$$
\begin{aligned}
&cf^* = c(m-r)/s^2,\\
&g_\infty(cf^*) = \big((m-r)^2\big(c - c^2/2\big)\big)/s^2 + r,\\
&\text{Sdev}\big(G_\infty(cf^*)\big) = c(m-r)/s,\\
&tg_\infty(cf^*) = k^2c^2/\big(c - c^2/2 + rs^2/(m-r)^2\big),\\
&t(k, cf^*) = k^2c^2\big((m-r)^2/s^2\big)/\big(\big((m-r)^2/s^2\big)\big(c-c^2/2\big)+r\big)^2.
\end{aligned}
\tag{7.7}
$$

If we define $\widetilde{m} = m - r$, $\widetilde{G}_\infty = G_\infty - r$, $\tilde{g}_\infty = g_\infty - r$, then substitution into Equations (7.7) give Equations (7.6), showing the relation between the two sets. It also shows that examples and conclusions about $P(V_n > V_0)$ in the $r = 0$ case are equivalent to those about $P(\ln(V(t)/V_0) > rt)$ in the $r \neq 0$ case. Thus we can compare various strategies versus an investment compounding at a constant riskless rate r such as zero coupon U.S. Treasury bonds.

From Equations (7.7) and $c = 1$, we find

$$
f^* = 2.2\bar{2}, \qquad g_\infty(f^*) = .11\bar{5}, \qquad \text{Sdev}\big(G_\infty(f^*)\big) = .3\bar{3},
$$

$$
tg_\infty(f^*) = .96k^2, \qquad t = 8.32k^2 \text{ years.}
$$

Thus, with $f^* = 2.2\bar{2}$, after 8.32 years the probability is 84% that $V_n > V_0$ and the expected value of $\log(V_n/V_0) = .96$ so the median value of V_n/V_0 will be about $e^{.96} = 2.61$.

With the usual unlevered $f = 1$, and $c = .45$, we find using (7.3)

$$
g_\infty(1) = m - s^2/2 = .09875, \qquad \text{Sdev}\big(G_\infty(1)\big) = .15,
$$

$$
tg_\infty(1) = .23k^2, \qquad t(k, .45f^*) = 2.31k^2 \text{ years.}
$$

Writing $tg_\infty = h(c)$ in (7.7) as

$$h(c) = k^2/\bigl(1/c + rs^2/\bigl((m-r)^2c^2\bigr) - 1/2\bigr)$$

we see that the measure of riskiness, $h(c)$, increases as c increases, at least up to the point $c = 2$, corresponding to $2f^*$ (and actually beyond, up to $1 + \sqrt{1 + \frac{rs^2}{(m-r)^2}}$).

Writing $t(k, cf^*) = t(c)$ as

$$t(c) = k^2\bigl((m-r)^2/s^2\bigr)/\bigl((m-r)^2/s^2\bigr)(1-c/2) + r/c^2$$

shows that $t(c)$ also increases as c increases, at least up to the point $c = 2$. Thus for smaller (more conservative) $f = cf^*$, $c \leqslant 2$, specified levels of $P(V_n > V_0)$ are reached earlier. For $c < 1$, this comes with a reduction in growth rate, which reduction is relatively small for f near f^*.

Note: During the period 1975–1997 the short term T-bill total return for the year, a proxy for r if the investor lends (i.e., $f < 1$), varied from a low of 2.90% (1993) to a high of 14.71% (1981). For details, see Ibbotson Associates, 1998 (or any later) Yearbook.

A large well connected investor might be able to borrow at broker's call plus about 1%, which might be approximated by T-bills plus 1%. This might be a reasonable estimate for the investor who borrows ($f > 1$). For others the rates are likely to be higher. For instance the prime rate from 1975–1997 varied from a low of 6% (1993) to a high of 19% (1981), according to Associates First Capital Corporation (1998).

As r fluctuates, we expect m to tend to fluctuate inversely (high interest rates tend to depress stock prices for well known reasons). Accordingly, f^* and g_∞ will also fluctuate so the long term S&P index fund investor needs a procedure for periodically re-estimating and revising f^* and his desired level of leverage or cash.

To illustrate the impact of $r_b > r$, where r_b is the investor's borrowing rate, suppose r_b in example (7.2) is $r + 2\%$ or .08, a choice based on the above cited historical values for r, which is intermediate between "good" $r_b \doteq r + 1\%$, and "poor" $r_b \doteq$ the prime rate $\doteq r + 3\%$. We replace r by r_b in Equations (7.7) and, if $f^* > 1$, $f^* = 1.3\bar{3}$, $g_\infty(f^*) = .100$, $\mathrm{Sdev}(G_\infty(f^*)) = .20$, $tg_\infty(f^*) = .4k^2$, $t = 4k^2$ years. Note how greatly f^* is reduced.

Comment: Taxes

Suppose for simplicity that all gains are subject to a constant continuous tax rate T and that all losses are subject to a constant continuous tax refund at the same rate T. Think of the taxing entities, collectively, as a partner that shares a fraction T of all gains and losses. Then Equations (7.7) become:

$$cf^* = c(m-r)/s^2(1-T),$$

$$g_\infty(cf^*) = \bigl((m-r)^2\bigl(c - c^2/2\bigr)\bigr)/s^2 + r(1-T),$$

$$\mathrm{Sdev}\bigl(G_\infty(cf^*)\bigr) = c(m-r)/s,$$

$$tg_\infty(cf^*) = k^2c^2/\big(c - c^2/2 + r(1-T)s^2/(m-r^2)\big),$$
$$t(k, cf^*) = k^2c^2\big((m-r)^2/s^2\big)/\Big(\big((m-r)^2/s^2\big)\big(c - c^2/2\big) + r(1-T)\Big)^2. \tag{7.7T}$$

It is interesting to see that cf^* increases by the factor $1/(1-T)$. For a high income California resident, the combined state and federal marginal tax rate is 45% so this factor is $1/.55 = 1.82$. The amplification of cf^* leads to the same growth rate as before except for a reduction by rT. The Sdev is unchanged and $t(k, cf^*)$ is increased slightly. However, as a practical matter, the much higher leverage needed with a high tax rate is typically not allowed under the margin regulation or is not advisable because the inability to continuously adjust in the real world creates dangers that increase rapidly with the degree of leverage.

7.3. The case for "fractional Kelly"

Figure 5 shows three g curves for the true $m : m_t = .5m_e$, $1.0m_e$ and $1.5m_e$, where m_e is the estimated value of m. The vertical lines and the slanting arrows illustrate the reduction in g for the three choices of: $f = .5f_e^*$, f_e^* and $1.5f_e^*$. For example with $f = .5f_e^*$ or "half Kelly", we have no loss and achieve the maximum $g = .25$, in case $m_t = .5m_e$. But if $m_t = m_e$ then $g = .75$, a loss of .25 and if $m_t = 1.5m_e$ then $g = 1.25$, a loss of 1.0, where all g are in units of $m_e^2/2s^2$. This is indicated both by LOSS_1 and LOSS_2 on the vertical line above $f/f_e^* = .5$, and by the two corresponding arrows which lead upward, and in this case to the right, from this line. A disaster occurs when $m_t = .5m_e$ but we choose $f = 1.5f_e^*$. This combines overbetting f_e^* by 50% with the overestimate of $m_e = 2m_t$. Then $g = -.75$ and we will be ruined. It is still bad to choose $f = f_e^*$ when $m_t = .5m_e$ for then $g = 0$ and we suffer increasingly wild oscillations, both up and down, around our initial capital. During a large downward oscillation experience shows that bettors will generally either quit or be eliminated by a minimum bet size requirement.

Some lessons here are:

(1) To the extent m_e is an uncertain estimate of m_t, it is wise to assume $m_t < m_e$ and to choose $f < f_e^*$ by enough to prevent $g \leqslant 0$.

Estimates of m_e in the stock market have many uncertainties and, in cases of forecast excess return, are more likely to be too high than too low. The tendency is to regress towards the mean. Securities prices follow a "non-stationary process" where m and s vary somewhat unpredictably over time. The economic situation can change for companies, industries, or the economy as a whole. Systems that worked may be partly or entirely based on data mining so m_t may be substantially less than m_e. Changes in the "rules" such as commissions, tax laws, margin regulations, insider trading laws, etc., can also affect m_t. Systems that do work attract capital, which tends to push exceptional m_t down towards average values. The drift down means $m_e > m_t$ is likely.

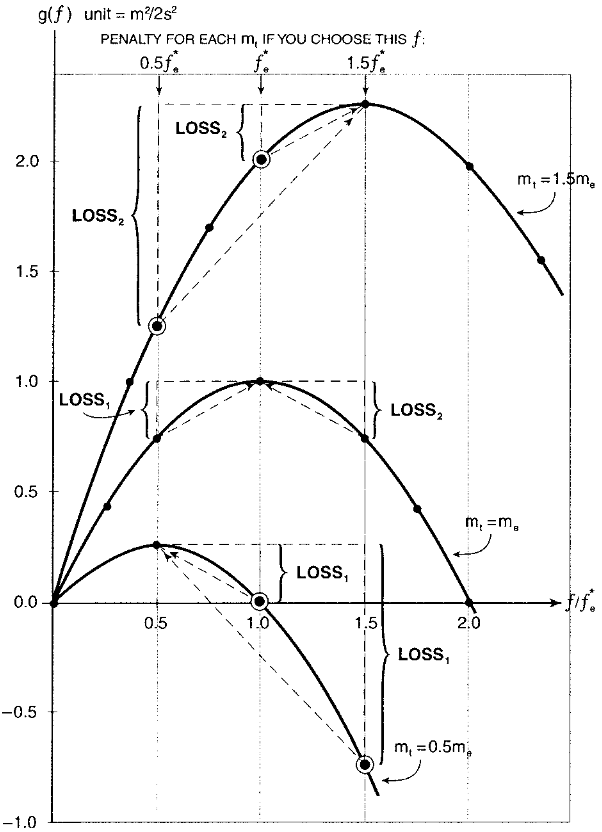

Fig. 5. Penalties for choosing $f = f_e \neq f^* = f_t$.

Sports betting has much the same caveats as the securities markets, with its own differences in detail. Rules changes, for instance, might include: adding expansion teams; the three point rule in basketball; playing overtime sessions to break a tie; changing types of bats, balls, gloves, racquets or surfaces.

Blackjack differs from the securities and sports betting markets in that the probabilities of outcomes can in principle generally be either calculated or simulated to any desired degree of accuracy. But even here m_t is likely to be at least somewhat less than m_e. Consider player fatigue and errors, calculational errors and mistakes in applying either blackjack theory or Kelly theory (e.g., calculating f^* correctly, for which some of the issues have been discussed above), effects of a fixed shuffle point, non-random shuffling, preferential shuffling, cheating, etc.

(2) Subject to (1), choosing f in the range $.5f_e^* \leqslant f < f_e^*$ offers protection against $g \leqslant 0$ with a reduction of g that is likely to be no more than 25%.

Example 7.3. The great compounder. In 1964 a young hedge fund manager acquired a substantial interest in a small New England textile company called Berkshire Hathaway. The stock traded then at 20. In 1998 it traded at 70,000, a multiple of 3500, and an annualized compound growth rate of about 27%, or an instantaneous rate of 24%. The once young hedge fund manager Warren Buffett is now acknowledged as the greatest investor of our time, and the world's second richest man. You may read about Buffett in Buffett and Clark (1997), Hagstrom (1994, 2004), Kilpatrick (1994) and Lowenstein (1995). If, as I was, you were fortunate enough to meet Buffett and identify the Berkshire opportunity, what strategy does our method suggest? Assume (the somewhat smaller drift rate) $m = .20$, $s = .15$, $r = .06$. Note: Plausible arguments for a smaller future drift rate include regression towards the mean, the increasing size of Berkshire, and risk from the aging of management. A counter-argument is that Berkshire's compounding rate has been as high in its later years as in its earlier years. However, the S&P 500 Index has performed much better in recent years so the spread between the growth rates of the Index and of Berkshire has been somewhat less. So, if we expect the Index growth rate to revert towards the historical mean, then we expect Berkshire to do so even more. From Equations (7.3) or (7.7),

$$f^* = 6.2\bar{2}, \qquad g_\infty(f^*) = .49\bar{5}, \qquad \mathrm{Sdev}\big(G_\infty(f^*)\big) = .9\bar{3},$$
$$t g_\infty(f^*) = 1.76k^2, \qquad t = 3.54k^2 \text{ years.}$$

Compare this to the unlevered portfolio, where $f = 1$ and $c = 1/6.2\bar{2} \doteq .1607$. We find:

$$f = 1, \qquad g_\infty(f) = .189, \qquad \mathrm{Sdev}\big(G_\infty(f)\big) = .15,$$
$$t_k g_\infty(f) = .119k^2, \qquad t_k = .63k^2 \text{ years.}$$

Leverage to the level $6.2\bar{2}$ would be inadvisable here in the real world because securities prices may change suddenly and discontinuously. In the crash of October, 1987, the S&P 500 index dropped 23% in a single day. If this happened at leverage of 2.0, the new leverage would suddenly be $77/27 = 2.85$ before readjustment by selling part of the portfolio. In the case of Berkshire, which is a large well-diversified portfolio, suppose we chose the conservative $f = 2.0$. Note that this is the maximum initial leverage allowed "customers" under current regulations. Then $g_\infty(2) = .295$. The values in 30 years for median V_∞/V_0 are approximately: $f = 1$, $V_\infty/V_0 = 288$; $f = 2$, $V_\infty/V_0 = 6,974$; $f = 6.2\bar{2}$, $V_\infty/V_0 = 2.86 \times 10^6$. So the differences in results of leveraging are enormous in a generation. (Note: Art Quaife reports $s = .24$ for 1980–1997. The reader is invited to explore the example with this change.)

The results of Section 3 apply directly to this continuous approximation model of a (possibly) leveraged securities portfolio. The reason is that both involve the same "dynamics", namely $\log G_n(f)$ is approximated as (scaled) Brownian motion with drift. So we can answer the same questions here for our portfolio that were answered in Section 3 for casino betting. For instance (3.2) becomes

$$\mathrm{Prob}\big(V(t)/V_0 \leqslant x \text{ for some } t\big) = x^\wedge\big(2g_\infty/\mathrm{Var}(G_\infty)\big) \tag{7.8}$$

where $\wedge$ means exponentiation and $0 < x < 1$. Using (7.4), for $r = 0$ and $f = f^*$, $2g_\infty/\operatorname{Var}(G_\infty) = 1$ so this simplifies to

$$\operatorname{Prob}(\cdot) = x. \tag{7.9}$$

Compare with Example 3.3 For $0 < r < m$ and $f = f^*$ the exponent of x in (7.9) becomes $1 + 2rs^2/(m - r)^2$ and has a positive first derivative so, as r increases, $P(\cdot)$ decreases since $0 < x < 1$, tending to 0 as r tends to m, which is what we expect.

7.4. A remarkable formula

In earlier versions of this chapter the exponent in Equations (3.2), (7.8) and (7.9) were off by a factor of 2, which I had inadvertently dropped during my derivation. Subsequently Don Schlesinger posted (without details) two more general continuous approximation formulas for the $r = 0$ case on the Internet at www.bjmath.com dated June 19, 1997.

If V_0 is the initial investment and $y > 1 > x > 0$ then for f^* the probability that $V(t)$ reaches yV_0 before xV_0 is

$$\operatorname{Prob}\bigl(V(t,\ f^*) \text{ reaches } yV_0 \text{ before } xV_0\bigr) = (1 - x)/\bigl(1 - (x/y)\bigr) \tag{7.10}$$

and more generally, for $f = cf^*$, $0 < c < 2$,

$$\begin{aligned}&\operatorname{Prob}\bigl(V(t,\ cf^*) \text{ reaches } yV_0 \text{ before } xV_0\bigr)\\ &\quad = \bigl[1 - x^\wedge(2/c - 1)\bigr]/\bigl[1 - (x/y)^\wedge(2/c - 1)\bigr]\end{aligned} \tag{7.11}$$

where $\wedge$ means exponentiation.

Clearly (7.10) follows from (7.11) by choosing $c = 1$. The $r = 0$ case of our Equation (7.8) follows from (7.11) and the $r = 0$ case of our Equation (7.9) follows from (7.10). We can derive a generalization of (7.11) by using the classical gambler's ruin formula (Cox and Miller, 1965, p. 31, Equation (2.0)) and passing to the limit as step size tends to zero (Cox and Miller, 1965, pp. 205–206), where we think of $\log(V(t,\ f)/V_0)$ as following a diffusion process with mean g_∞ and variance $v(G_\infty)$, initial value 0, and absorbing barriers at $\log y$ and $\log x$. The result is

$$\operatorname{Prob}\bigl(V(t, cf^*) \text{ reaches } yV_0 \text{ before } xV_0\bigr) = [1 - x^\wedge a]/\bigl[1 - (x/y)^\wedge a\bigr] \tag{7.12}$$

where $a = 2g_\infty/V(G_\infty) = 2M/V$ where M and V are the drift and variance, respectively, of the diffusion process per unit time. Alternatively, (7.12) is a simple restatement of the known solution for the Wiener process with two absorbing barriers (Cox and Miller, 1965, Example 5.5).

As Schlesinger notes, choosing $x = 1/2$ and $y = 2$ in (7.10) gives $\operatorname{Prob}(V(t,\ f^*)$ doubles before halving$) = 2/3$. Now consider a gambler or investor who focuses only on values $V_n = 2^n V_0$, $n = 0, \pm 1, \pm 2, \ldots$, multiples of his initial capital. In log space, $\log(V_n/V_0) = n \log 2$ so we have a random walk on the integer multiples of $\log 2$, where the probability of an increase is $p = 2/3$ and of a decrease, $q = 1/3$. This gives us a convenient compact visualization of the Kelly strategy's level of risk.

If instead we choose $c = 1/2$ ("half Kelly"), Equation (7.11) gives Prob($V(t, f^*/2)$ doubles before halving) $= 8/9$ yet the growth rate $g_\infty(f^*/2) = .75 g_\infty(f^*)$ so "half Kelly" has 3/4 the growth rate but much less chance of a big loss.

A second useful visualization of comparative risk comes from Equation (7.8) which gives

$$\mathrm{Prob}\big(V(t, cf^*)/V_0 \leqslant x \text{ for some } t\big) = x^\wedge(2/c - 1). \tag{7.13}$$

For $c = 1$ we had Prob$(\cdot) = x$ and for $c = 1/2$ we get Prob$(\cdot) = x^3$. Thus "half Kelly" has a much lessened likelihood of severe capital loss. The chance of ever losing half the starting capital is 1/2 for $f = f^*$ but only 1/8 for $f = f^*/2$. My gambling and investment experience, as well as reports from numerous blackjack players and teams, suggests that most people strongly prefer the increased safety and psychological comfort of "half Kelly" (or some nearby value), in exchange for giving up 1/4 of their growth rate.

8. A case study

In the summer of 1997 the XYZ Corporation (pseudonym) received a substantial amount of cash. This prompted a review of its portfolio, which is shown in Table 7 in the column 8/17/97. The portfolio was 54% in Biotime, ticker BTIM, a NASDAQ biotechnology company. This was due to existing and historical relationships between people in XYZ Corp. and in BTIM. XYZ's officers and directors were very knowledge-

Table 7
Statistics for logs of monthly wealth relatives, 3/31/92 through 6/30/97

		Berkshire	BioTime	SP500	T-bills
Monthly	Mean	.0264	.0186	.0146	.0035
	Standard deviation	.0582	.2237	.0268	.0008
Annual	Mean	.3167	.2227	.1753	.0426
	Standard deviation	.2016	.7748	.0929	.0028
Monthly	Covariance	.0034	−.0021	.0005	1.2E−06
			.0500	−.0001	3.2E−05
				.0007	5.7E−06
					6.7E−07
Monthly	Correlation	1.0000	−.1581	.2954	.0257
			1.0000	−.0237	.1773
				1.0000	.2610
					1.0000

able about BTIM and felt they were especially qualified to evaluate it as an investment. They wished to retain a substantial position in BTIM.

The portfolio held Berkshire Hathaway, ticker BRK, having first purchased it in 1991.

8.1. The constraints

Dr. Quaife determined the Kelly optimal portfolio for XYZ Corp. subject to certain constraints. The list of allowable securities was limited to BTIM, BRK, the Vanguard 500 (S&P 500) Index Fund, and T-bills. Being short T-bills was used as a proxy for margin debt. The XYZ broker actually charges about 2% more, which has been ignored in this analysis. The simple CAPM (capital asset pricing model) suggests that the investor only need consider the market portfolio (for which the S&P 500 is being substituted here, with well known caveats) and borrowing or lending. Both Quaife and the author were convinced that BRK was and is a superior alternative and their knowledge about and long experience with BRK supported this.

XYZ Corp. was subject to margin requirements of 50% initially and 30% maintenance, meaning for a portfolio of securities purchased that initial margin debt (money lent by the broker) was limited to 50% of the value of the securities, and that whenever the value of the account net of margin debt was less than 30% of the value of the securities owned, then securities would have to be sold until the 30% figure was restored.

In addition XYZ Corp. wished to continue with a “significant” part of its portfolio in BTIM.

8.2. The analysis and results

Using monthly data from 3/31/92 through 6/30/97, a total of 63 months, Quaife finds the means, covariances, etc. given in Table 7.

Note from Table 7 that BRK has a higher mean and a lower standard deviation than BTIM, hence we expect it to be favored by the analysis. But note also the negative correlation with BTIM, which suggests that adding some BTIM to BRK may prove advantageous.

Using the statistics from Table 7, Quaife finds the following optimal portfolios, under various assumptions about borrowing.

As expected, BRK is important and favored over BTIM but some BTIM added to the BRK is better than none.

If unrestricted borrowing were allowed it would be foolish to choose the corresponding portfolio in Table 8. The various underlying assumptions are only approximations with varying degrees of validity: Stock prices do not change continuously; portfolios can’t be adjusted continuously; transactions are not costless; the borrowing rate is greater than the T-bill rate; the after tax return, if different, needs to be used; the process which generates securities returns is not stationary and our point estimates of the statistics in Table 7 are uncertain. We have also noted earlier that because “over-

Table 8
Optimal portfolio allocations with various assumptions about borrowing

Security fraction			
Security	No borrowing	50% margin	Unrestricted borrowing
Berkshire	.63	1.50	6.26
BioTime	.37	.50	1.18
S&P 500	.00	.00	12.61
T-bills	.00	−1.00	−19.04
Portfolio growth rate			
Mean	.36	.62	2.10
Standard deviation	.29	.45	2.03

betting" is much more harmful than underbetting, "fractional Kelly" is prudent to the extent the results of the Kelly calculations reflect uncertainties.

In fact, the data used comes from part of the period 1982–1997, the greatest bull market in history. We would expect returns in the future to regress towards the mean so the means in Table 7 are likely to be overestimates of the near future. The data set is necessarily short, which introduces more uncertainty, because it is limited by the amount of BTIM data. As a sensitivity test, Quaife used conservative (mean, std. dev.) values for the price relatives (not their logs) for BRK of (1.15, .20), BTIM of (1.15, 1.0) and the S&P 500 from 1926–1995 from Ibbotson (1998) of (1.125, .204) and the correlations from Table 7. The result was fractions of 1.65, .17, .18 and −1.00 respectively for BRK, BTIM, S&P 500 and T-bills. The mean growth rate was .19 and its standard deviation was .30.

8.3. The recommendation and the result

The 50% margin portfolio reallocations of Table 8 were recommended to XYZ Corp.'s board on 8/17/97 and could have been implemented at once. The board elected to do nothing. On 10/9/97 (in hindsight, a good sale at a good price) it sold some BTIM and left the proceeds in cash (not good). Finally on 2/9/98 after a discussion with both Quaife and the author, it purchased 10 BRK (thereby gaining almost $140,000 by 3/31/98, as it happened). The actual policy, led to an increase of 73.5%. What would have happened with the recommended policy with no rebalance and with one rebalance on 10/6/97? The gains would have been 117.6% and 199.4%, respectively. The gains over the suboptimal board policy were an additional $475,935 and $1,359,826, respectively.

The optimal policy displays three important features in this example: the use of leverage, the initial allocation of the portfolio, and possible rebalancing (reallocation) of the portfolio over time. Each of these was potentially important in determining the final result. The potential impact of continuously rebalancing to maintain maximum margin is illustrated in Thorp and Kassouf (1967), Appendix A, The Avalanche Effort.

The large loss from the suboptimal policy was much more than what would have been expected because BRK and BTIM appreciated remarkably. In .62 years, BRK was up 60.4% and BTIM was up 62.9%. This tells us that—atypically—in the absence of rebalancing, the relative initial proportions of BRK and BTIM did not matter much over the actual time period. However, rebalancing to adjust the relative proportions of BRK and BTIM was important, as the actual policy's sale of some BTIM on 10/9/97 illustrated. Also, rebalancing was important for adjusting the margin level as prices, in this instance, rose rapidly.

Table 8 illustrates what we might have normally expected to gain by using 50% margin, rather than no margin. We expect the difference in the medians of the portfolio distributions to be \$1,080,736[exp(.62 × .62) − exp(.36 × .62)] = \$236,316 or 21.9% which is still large.

8.4. *The theory for a portfolio of securities*

Consider first the unconstrained case with a riskless security (T-bills) with portfolio fraction f_0 and n securities with portfolio fractions $f_1, \ldots, f_n$. Suppose the rate of return on the riskless security is r and, to simplify the discussion, that this is also the rate for borrowing, lending, and the rate paid on short sale proceeds. Let $C = (s_{ij})$ be the matrix such that s_{ij}, $i, j = 1, \ldots, n$, is the covariance of the ith and jth securities and $M = (m_1, m_2, \ldots, m_n)^T$ be the row vector such that m_i, $i = 1, \ldots, n$, is the drift rate of the ith security. Then the portfolio satisfies

$$\begin{aligned} & f_0 + \cdots + f_n = 1, \\ & m = f_0 r + f_1 m_1 + \cdots + f_n m_n = r + f_1(m_1 - r) + \cdots + f_n(m_n - r) \\ & \quad = r + F^T(M - R), \\ & s^2 = F^T C F \end{aligned} \tag{8.1}$$

where $F^T = (f_1, \ldots, f_n)$ and ^{T} means "transpose", and R is the column vector $(r, r, \ldots, r)^T$ of length n.

Then our previous formulas and results for one security plus a riskless security apply to $g_\infty(f_1, \ldots, f_n) = m - s^2/2$. This is a standard quadratic maximization problem. Using (8.1) and solving the simultaneous equations $\partial g_\infty / \partial f_i = 0$, $i = 1, \ldots, n$, we get

$$\begin{aligned} & F^* = C^{-1}[M - R], \\ & g_\infty(f_1^*, \ldots, f_n^*) = r + (F^*)^T C F^*/2 \end{aligned} \tag{8.2}$$

where for a unique solution we require C^{-1} to exist, i.e., $\det C \neq 0$. When all the securities are uncorrelated, C is diagonal and we have $f_i^* = (m_i - r)/s_{ii}$ or $f_i^* = (m_i - r)/s_i^2$, which agrees with Equation (7.3) when $n = 1$.

Note: BRK issued a new class of common, ticker symbol BRK.B, with the old common changing its symbol to BRK.A. One share of BRK.A can be converted to 30 shares of BRK.B at any time, but not the reverse. BRK.B has lesser voting rights and no right to assign a portion of the annual quota of charitable contributions. Both we and the market consider these differences insignificant and the A has consistently traded at about 30 times the price of the B.

If the price ratio were always exactly 30 to 1 and both these securities were included in an analysis, they would each have the same covariances with other securities, so $\det C = 0$ and C^{-1} does not exist.

If there is an initial margin constraint of q, $0 \leqslant q \leqslant 1$, then we have the additional restriction

$$|f_1| + \cdots + |f_n| \leqslant 1/q. \tag{8.3}$$

The n-dimensional subset in (8.3) is closed and bounded.

If the rate for borrowing to finance the portfolio is $r_b = r + e_b$, $e_b \geqslant 0$, and the rate paid on the short sale proceeds is $r_s = r - e_s$, $e_s \geqslant 0$, then the m in Equation (8.1) is altered. Let $x^+ = \max(x, 0)$ and $x^- = \max(0, -x)$ so $x = x^+ - x^-$ for all x. Define $f^+ = f_1^+ + \cdots + f_n^+$, the fraction of the portfolio held long. Let $f^- = f_1^- + \cdots + f_n^-$, the fraction of the portfolio held short.

Case 1. $f^+ \leqslant 1$

$$m = r + f_1(m_1 - r) + \cdots + f_n(m_n - r) - e_s f^-. \tag{8.4.1}$$

Case 2. $f^+ > 1$

$$m = r + f_1(m_1 - r) + \cdots + f_n(m_n - r) - e_b(f^+ - 1) - e_s f^-. \tag{8.4.2}$$

9. My experience with the Kelly approach

How does the Kelly-optimal approach do in practice in the securities markets? In a little-known paper (Thorp, 1971) I discussed the use of the Kelly criterion for portfolio management. Page 220 mentions that "On November 3, 1969, a private institutional investor decided to ... use the Kelly criterion to allocate its assets". This was actually a private limited partnership, specializing in convertible hedging, which I managed. A notable competitor at the time (see Institutional Investor (1998)) was future Nobel prize winner Harry Markowitz. After 20 months, our record as cited was a gain of 39.9% versus a gain for the Dow Jones Industrial Average of +4.2%. Markowitz dropped out after a couple of years, but we liked our results and persisted. What would the future bring?

Up to May 1998, twenty eight and a half years since the investment program began. The partnership and its continuations have compounded at approximately 20% annually with a standard deviation of about 6% and approximately zero correlation with the market ("market neutral"). Ten thousand dollars would, tax exempt, now be worth 18 million dollars. To help persuade you that this may not be luck, I estimate that during this period I have made about $80 billion worth of purchases and sales ("action", in casino language) for my investors. This breaks down into something like one and a quarter million individual "bets" averaging about $65,000 each, with on average hundreds of "positions" in place at any one time. Over all, it would seem to be a moderately "long run" with a high probability that the excess performance is more than chance.

10. Conclusion

Those individuals or institutions who are long term compounders should consider the possibility of using the Kelly criterion to asymptotically maximize the expected compound growth rate of their wealth. Investors with less tolerance for intermediate term risk may prefer to use a lesser function. Long term compounders ought to avoid using a greater fraction ("overbetting"). Therefore, to the extent that future probabilities are uncertain, long term compounders should further limit their investment fraction enough to prevent a significant risk of overbetting.

Acknowledgements

I thank Dr. Jerry Baesel, Professor Sid Browne, Professor Peter Griffin, Dr. Art Quaife, and Don Schlesinger for comments and corrections and to Richard Reid for posting this chapter on his website. I am also indebted to Dr. Art Quaife for allowing me to use his analysis in the case study.

This chapter has been revised and expanded since its presentation at the 10th International Conference on Gambling and Risk Taking.

Appendix A. Integrals for deriving moments of E_∞

$$
\begin{aligned}
I_0\left(a^2, b^2\right) &= \int_0^\infty \exp\left[-\left(a^2x^2 + b^2/x^2\right)\right] \mathrm{d}x, \\
I_n\left(a^2, b^2\right) &= \int_0^\infty x^n \exp\left[-\left(a^2x^2 + b^2/x^2\right)\right] \mathrm{d}x.
\end{aligned}
$$

Given I_0 find I_2

$$
\begin{aligned}
I_0\left(a^2, b^2\right) &= \int_0^\infty \exp\left[-\left(a^2x^2 + b^2/x^2\right)\right] \mathrm{d}x \\
&= -\int_\infty^0 \exp\left[-\left(a^2/u^2 + b^2u^2\right)\right]\left(-\mathrm{d}u/u^2\right)
\end{aligned}
$$

where $x = 1/u$ and $\mathrm{d}x = -\mathrm{d}u/u^2$ so

$$I_0(a^2, b^2) = \int_0^\infty x^{-2} \exp\left[-\left(b^2x^2 + a^2/x^2\right)\right] = I_{-2}\left(b^2, a^2\right),$$

hence

$$I_{-2}\left(a^2, b^2\right) = I_0\left(b^2, a^2\right) = \frac{\sqrt{\pi}}{2\mid b\mid}\mathrm{e}^{-2|ab|},$$

$$I_0 = \int_0^\infty \exp\left[-\left(a^2x^2 + b^2/x^2\right)\right]\mathrm{d}x = U \cdot V|_0^\infty - \int_0^\infty V\,\mathrm{d}U \tag{A.1}$$

where $U = \exp[\,\cdot\,]$, $\mathrm{d}V = \mathrm{d}x$, $\mathrm{d}U = (\exp[\,\cdot\,])(-2a^2x + 2b^2x^{-3})$ and $V = x$ so

$$\begin{aligned} I_0 &= \exp\left[-\left(a^2x^2 + b^2/x^2\right)\right] \cdot x|_0^\infty \\ &\quad - \int_0^\infty \left(-2a^2x^2 + 2b^2/x^2\right)\exp\left[-\left(a^2x^2 + b^2/x^2\right)\right]\mathrm{d}x \\ &= 2a^2 I_2\left(a^2, b^2\right) - 2b^2 I_{-2}\left(a^2, b^2\right). \end{aligned}$$

Hence:

$$I_0\left(a^2, b^2\right) = 2a^2 I_2\left(a^2, b^2\right) - 2b^2 I_{-2}\left(a^2, b^2\right)$$

and $I_{-2}(a^2, b^2) = I_0(b^2, a^2)$ so substituting and solving for I_2 gives

$$I_2\left(a^2, b^2\right) = \frac{1}{2a^2}\left\{I_0\left(a^2, b^2\right) + 2b^2 I_0\left(b^2, a^2\right)\right\}.$$

Comments.

(1) We can solve for all even n by using I_0, I_{-2} and I_2, and integration by parts.

(2) We can use the indefinite integral J_0 corresponding to I_0, and the previous methods, to solve for J_{-2}, J_2, and then for all even n. Since

$$I_0\left(a^2, b^2\right) = \frac{\sqrt{\pi}}{2|a|}\mathrm{e}^{-2|ab|}$$

then

$$I_{-2}\left(a^2, b^2\right) = \frac{\sqrt{\pi}}{2|b|}\mathrm{e}^{-2|ab|} \quad \text{and}$$

$$I_2\left(a^2, b^2\right) = \frac{1}{2a^2}\left\{\frac{\sqrt{\pi}}{2|a|} + 2b^2\frac{\sqrt{\pi}}{2|b|}\right\}\mathrm{e}^{-2|ab|} = \frac{\sqrt{\pi}}{4a^2}\mathrm{e}^{-2|ab|}\left\{1/|a| + 2|b|\right\}.$$

Appendix B. Derivation of formula (3.1)

This is based on a note from Howard Tucker. Any errors are mine.

From the paper by Paranjape and Park, if $x(t)$ is standard Brownian motion, if $a \neq 0$, $b > 0$,

$$P\big(X(t) \leqslant at + b,\ 0 \leqslant t \leqslant T \mid X(T) = s\big)$$
$$= \begin{cases} 1 - \exp\left\{-\dfrac{2b}{T}(aT + B - s)\right\} & \text{if } s \leqslant aT + b, \\ 0 & \text{if } s > aT + B. \end{cases}$$

Write this as:

$$P\big(X(t) \leqslant at + b,\ 0 \leqslant t \leqslant T \mid X(T)\big)$$
$$\overset{\text{a.s.}}{=} 1 - \exp\left\{-2b\big(aT + b - X(T)\big)\frac{1}{T}\right\} \quad \text{if } X(T) \leqslant aT + b.$$

Taking expectations of both sides of the above, we get

$$P\big(X(t) \leqslant at + b,\ 0 \leqslant t \leqslant T\big)$$
$$= \int_{-\infty}^{aT+b} \big(1 - e^{-2b(aT+b-s)1/T}\big) \frac{1}{\sqrt{2\pi T}} e^{-s^2/2T}\, ds$$
$$= \frac{1}{\sqrt{2\pi T}} \int_{-\infty}^{aT+b} e^{-s^2/2T}\, ds - \frac{e^{-2ab}}{\sqrt{2\pi T}} \int_{-\infty}^{aT+b} e^{-(s-2b)^2/2T}\, ds.$$

Hence

$$P(X \text{ going above line } at + b \text{ during } [0, T]) = 1 - \text{previous probability}$$
$$= \frac{1}{\sqrt{2\pi T}} \int_{aT+b}^{\infty} e^{-s^2/2T}\, ds + e^{-2ab} \cdot \frac{1}{\sqrt{2\pi T}} \int_{-\infty}^{aT-b} e^{-u^2/2T}\, du,$$
$$\text{where } u = s - 2b. \tag{B.1}$$

Now, when $a = 0$, $b > 0$,

$$P\Big[\sup_{0 \leqslant t \leqslant T} X(t) \geqslant b\Big] = \sqrt{\frac{2}{\pi T}} \int_b^{\infty} e^{-v^2/2T}\, dv,$$

which agrees with a known formula (see, e.g., p. 261 of Tucker (1967)). In the case $a > 0$, when $T \to \infty$, since $\sqrt{T}/T \to 0$ and $\sqrt{T} = s.d.$ of $X(T)$, the first integral $\to 0$, the second integral $\to 1$, and $P(X$ ever rises above line $at + b) = e^{-2ab}$. Similarly, in the case $a < 0$, $P($ever rises above line $at + b) = 1$.

The theorem it comes from is due to Sten Malmquist, On certain confidence contours for distribution functions, Ann. Math. Stat. 25 (1954), pp. 523–533. This theorem is stated in S.R. Paranjape and C. Park, Distribution of the supremum of the two-parameter Yeh–Wiener process on the boundary, J. Appl. Prob. 10 (1973).

Letting $\alpha = a\sqrt{T}$, $\beta = b/\sqrt{T}$, formula (B.1) becomes

$$P(\cdot) = N(-\alpha - \beta) + e^{-2\alpha\beta} N(\alpha - \beta) \text{ where } \alpha, \beta > 0 \quad \text{or}$$

$$P\big(X(t) \leqslant at + b,\ 0 \leqslant t \leqslant T\big) = 1 - P(\cdot) = N(\alpha + \beta) - \mathrm{e}^{-2\alpha\beta} N(\alpha - \beta)$$

for the probability the line is never surpassed. This follows from:

$$\frac{1}{\sqrt{2\pi T}} \int_{aT+b}^{\infty} \mathrm{e}^{-s^2/2T}\, \mathrm{d}s = \frac{1}{\sqrt{2\pi}} \int_{a\sqrt{T}+b/\sqrt{T}}^{\infty} \mathrm{e}^{-x^2/2}\, \mathrm{d}x = N(-\alpha - \beta) \quad \text{and}$$

$$\frac{1}{\sqrt{2\pi T}} \int_{-\infty}^{aT-b} \mathrm{e}^{-u^2/2T}\, \mathrm{d}u = N(\alpha - \beta)$$

where $s = aT + b$, $x = s/\sqrt{T} = a\sqrt{T} + b/\sqrt{T}$, $\alpha = a\sqrt{T}$ and $\beta = b/\sqrt{T}$.

The formula becomes:

$$\begin{aligned} &P\big(\sup\big[X(t) - (at + b)\big] \geqslant 0\colon\ 0 \leqslant t \leqslant T\big) \\ &\quad = N(-\alpha - \beta) + \mathrm{e}^{-2ab} N(\alpha - \beta) \\ &\quad = N(-\alpha - \beta) + \mathrm{e}^{-2\alpha\beta} N(\alpha - \beta), \quad \alpha, \beta > 0. \end{aligned}$$

Observe that

$$\begin{aligned} P(\cdot) &< N(-\alpha - \beta) + N(\alpha - \beta) = \big\{1 - N(\alpha + \beta)\big\} + N(\alpha - \beta) \\ &= \int_{-\infty}^{\alpha-\beta} \alpha(x)\, \mathrm{d}x + \int_{\alpha+\beta}^{\infty} \alpha(x)\, \mathrm{d}x < 1 \end{aligned}$$

as it should be.

Appendix C. Expected time to reach goal

Reference: Handbook of Mathematical Functions, Abramowitz and Stegun, Editors, N.B.S. Applied Math. Series 55, June 1964.

P. 304, 7.4.33 gives with $\operatorname{erf} z \equiv \frac{2}{\sqrt{\pi}} \int_0^z \mathrm{e}^{-t^2}\, \mathrm{d}t$ the integral:

$$\begin{aligned} &\int \exp\big\{-\big(a^2x^2 + b^2/x^2\big)\big\}\, \mathrm{d}x \\ &\quad = \frac{\sqrt{\pi}}{4a}\big[\mathrm{e}^{2ab} \operatorname{erf}(ax + b/x) + \mathrm{e}^{-2ab} \operatorname{erf}(ax - b/x)\big] + C, \quad a \neq 0. \end{aligned} \tag{C.1}$$

Now the left side is >0 so for real a, we require $a > 0$ otherwise the right side is <0, a contradiction.

We also note that p. 302, 7.4.3. gives

$$\int_0^{\infty} \exp\big\{-\big(at^2 + b/t^2\big)\big\}\, \mathrm{d}t = \frac{1}{2}\sqrt{\frac{\pi}{a}}\, \mathrm{e}^{-2\sqrt{ab}} \tag{C.2}$$

with $\Re a > 0$, $\Re b > 0$.

To check (C.2) v. (C.1), suppose in (C.1) $a > 0$, $b > 0$ and find $\lim_{x\to 0}$ and $\lim_{x\to\infty}$ of $\operatorname{erf}(ax + b/x)$ and $\operatorname{erf}(ax - b/x)$,

$$\lim_{x\downarrow 0+}(ax + b/x) = +\infty, \qquad \lim_{x\downarrow 0+}(ax - b/x) = -\infty,$$
$$\lim_{x\to\infty}(ax + b/x) = +\infty, \qquad \lim_{x\to\infty}(ax - b/x) = +\infty.$$

Equation (C.1) becomes

$$\frac{\sqrt{\pi}}{4a}\mathrm{e}^{-2ab}\big[\operatorname{erf}(\infty) - \operatorname{erf}(-\infty)\big] = \frac{\sqrt{\pi}}{4a}\mathrm{e}^{-2ab}2\operatorname{erf}(\infty) = \frac{\sqrt{\pi}}{2a}\mathrm{e}^{-2ab}$$

since we know $\operatorname{erf}(\infty) = 1$.

In (C.2) replace a by a^2, b by b^2 to get

$$I_0(a^2, b^2) \equiv \int_0^\infty \exp\{-(a^2t^2 + b^2/t^2)\}\,\mathrm{d}t = \frac{1}{2}\frac{\sqrt{\pi}}{|a|}\mathrm{e}^{-2|ab|}$$

which is the same.

Note: if we choose the lower limit of integration to be 0 in (C.1), then we can find C:

$$\begin{aligned}0 &= \int_0^{0^+} \exp\{-(a^2x^2 + b^2/x^2)\}\,\mathrm{d}x = \frac{\sqrt{\pi}}{4a}\big[\mathrm{e}^{2ab}\operatorname{erf}(\infty) + \mathrm{e}^{-2ab}\operatorname{erf}(-\infty)\big] + C\\ &= \frac{\sqrt{\pi}}{4a}\big[\mathrm{e}^{2ab} - \mathrm{e}^{-2ab}\big] + C.\end{aligned}$$

Whence

$$\begin{aligned}F(x) &\equiv \int_0^x \exp\{-(a^2x^2 + b^2/x^2)\}\,\mathrm{d}x\\ &= \frac{\sqrt{\pi}}{4a}\big\{\mathrm{e}^{2ab}\big[\operatorname{erf}(ax + b/x) - 1\big] + \mathrm{e}^{-2ab}\big[\operatorname{erf}(ax - b/x) + 1\big]\big\}.\end{aligned} \tag{C.3}$$

To see how (C.3) might have been discovered, differentiate:

$$\begin{aligned}F'(x) &= \exp\{-(a^2x^2 + b^2/x^2)\}\\ &= \frac{\sqrt{\pi}}{4a}\big\{\mathrm{e}^{2ab}(a - b/x^2)\operatorname{erf}'(ax + b/x)\\ &\quad + \mathrm{e}^{-2ab}(a + b/x^2)\operatorname{erf}'(ax - b/x)\big\}.\end{aligned}$$

Now $\operatorname{erf}'(z) = \frac{2}{\sqrt{\pi}}\exp(-z^2)$ so

$$\begin{aligned}\operatorname{erf}'(ax + b/x) &= \frac{2}{\sqrt{\pi}}\exp\big[-(ax + b/x)^2\big] = \frac{2}{\sqrt{\pi}}\exp\{-(a^2x^2 + b^2/x^2 + 2ab)\}\\ &= \frac{2}{\sqrt{\pi}}\mathrm{e}^{-2ab}\exp\{-(a^2x^2 + b^2/x^2)\}\end{aligned}$$

and, setting $b \leftarrow -b$,

$$\mathrm{erf}'(ax - b/x) = \frac{2}{\sqrt{\pi}} \mathrm{e}^{2ab} \exp\{-(a^2x^2 + b^2/x^2)\}$$

whence

$$\begin{aligned} F'(x) &= \frac{\sqrt{\pi}}{4a} \left\{ \frac{2}{\sqrt{\pi}} (a - b/x^2) + \frac{2}{\sqrt{\pi}} (a + b/x^2) \right\} \exp\{-(a^2x^2 + b^2/x^2)\} \\ &= \frac{1}{2a} \{2a\} \exp\{-(a^2x^2 + b^2/x^2)\} = \exp\{-(a^2x^2 + b^2/x^2)\}. \end{aligned}$$

Case of interest: $a < 0$, $b > 0$.
Expect:

$$\begin{aligned} b > 0,\ a \leqslant 0 \quad &\Rightarrow \quad F(T) \uparrow 1 \text{ as } T \to \infty, \\ b > 0,\ a > 0 \quad &\Rightarrow \quad F(T) \uparrow c < 1 \text{ as } T \to \infty. \end{aligned}$$

If $b > 0$, $a = 0$:

$$F(T) = N(-\beta) + N(-\beta) = 2N(-b/\sqrt{T}) \uparrow 2N(0) = 1 \quad \text{as } T \uparrow \infty.$$

Also, as expected $F(T) \uparrow 1$ as $b \downarrow 0$.
If $b > 0$, $a < 0$: See below.
If $b > 0$, $a > 0$:

$$\begin{aligned} F(T) &= N(-a\sqrt{T} - b/\sqrt{T}) + \mathrm{e}^{-2ab} N(a\sqrt{T} - b/\sqrt{T}) \\ &\to N(-\infty) + \mathrm{e}^{-2ab} N(\infty) \\ &= \mathrm{e}^{-2ab} < 1 \quad \text{as } T \uparrow \infty. \end{aligned}$$

This is correct.

If $b = 0$: $F(T) = N(-a\sqrt{T}) + N(a\sqrt{T}) = 1$. This is correct.

Let $F(T) = P(X(t) \geqslant at + b$ for some $t, 0 \leqslant t \leqslant T)$ which equals $N(-\alpha - \beta) + \mathrm{e}^{-2ab} N(\alpha - \beta)$ where $\alpha = a\sqrt{T}$ and $\beta = b/\sqrt{T}$ so $ab = \alpha\beta$; we assume $b > 0$ and $a < 0$ in which case $0 \leqslant F(T) \leqslant 1$ and $\lim_{T\to\infty} F(T) = 1$, $\lim_{T\to 0} F(T) = 0$; F is a probability distribution function:

$$\begin{aligned} &\lim_{T\to 0} F(T) = N(-\infty) + \mathrm{e}^{-2ab} N(-\infty) = 0, \\ &\lim_{T\to\infty} F(T) = N(+\infty) + \mathrm{e}^{-2ab} N(-\infty) = 1. \end{aligned}$$

The density function is

$$f(T) = F'(T) = \frac{\partial}{\partial T}(-\alpha - \beta) N'(-\alpha - \beta) + \mathrm{e}^{-2ab} \frac{\partial}{\partial T}(\alpha - \beta) N'(\alpha - \beta)$$

where

$$\frac{\partial \alpha}{\partial T} = \frac{1}{2} a T^{-1/2}, \qquad \frac{\partial \beta}{\partial T} = -\frac{1}{2} b T^{-3/2},$$

$$N'(-\alpha-\beta)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-(\alpha+\beta)^2/2}=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(a^2T+b^2/T+2ab)}{2}\right\},$$

$$N'(\alpha-\beta)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-(\alpha-\beta)^2/2}=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(a^2T+b^2/T-2ab)}{2}\right\},$$

$$\begin{aligned}Tf(T)&=T\left(-\frac{1}{2}aT^{-1/2}+\frac{1}{2}bT^{-3/2}\right)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-ab}\exp\left\{-\frac{(a^2T+b^2/T)}{2}\right\}\\&\quad+T\mathrm{e}^{-2ab}\left(\frac{1}{2}aT^{-1/2}+\frac{1}{2}bT^{-3/2}\right)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{ab}\exp\left\{-\frac{(a^2T+b^2/T)}{2}\right\}\\&=\frac{\mathrm{e}^{-ab}}{2\sqrt{2\pi}}\left[(-aT^{+1/2}+bT^{-1/2})\exp\left\{-\frac{(a^2T+b^2/T)}{2}\right\}\right.\\&\quad\left.+\left(aT^{+1/2}+bT^{-1/2}\right)\exp\left\{-\frac{(a^2T+b^2/T)}{2}\right\}\right]\\&=\frac{b\mathrm{e}^{-ab}}{\sqrt{2\pi}}T^{1/2}\exp\left\{\frac{-(a^2T+b^2/T)}{2}\right\}.\end{aligned}$$

The expected time to the goal is

$$E_\infty=\int_0^\infty Tf(T)\,\mathrm{d}T=\frac{b\mathrm{e}^{-ab}}{\sqrt{2\pi}}\int_0^\infty T^{-1/2}\exp\left\{-\frac{(a^2T+b^2/T)}{2}\right\}\mathrm{d}T,$$

$$\left.\begin{array}{l}T^{1/2}=x\\T=x^2\\\mathrm{d}T=2x\,\mathrm{d}x\end{array}\right\}=\frac{2b\mathrm{e}^{-ab}}{\sqrt{2\pi}}\int_0^\infty\exp\left\{-\left[\left(\frac{a}{\sqrt{2}}\right)^2x^2+\left(\frac{b}{\sqrt{2}}\right)^2x^{-2}\right]\right\}\mathrm{d}x$$

$$=\frac{2b\mathrm{e}^{-ab}}{\sqrt{2\pi}}I_0\left(\left(\frac{a}{\sqrt{2}}\right)^2,\left(\frac{b}{\sqrt{2}}\right)^2\right).$$

Now

$$I_0\left(a^2,b^2\right)=\frac{\sqrt{\pi}}{2|a|}\mathrm{e}^{-2|ab|}\quad\text{so}$$

$$I_0\left(\left(\frac{a}{\sqrt{2}}\right)^2,\left(\frac{b}{\sqrt{2}}\right)^2\right)=\frac{\sqrt{\pi}}{\sqrt{2}|a|}\mathrm{e}^{-|ab|}\quad\text{whence}$$

$$E_\infty=\frac{2b\mathrm{e}^{-ab}}{\sqrt{2\pi}}\frac{\sqrt{\pi}}{\sqrt{2}|a|}\mathrm{e}^{-|ab|}=\frac{b}{|a|},\quad a<0,\ b>0.$$

Note:

$$f(T)\equiv F'(T)=\frac{b\mathrm{e}^{-ab}}{\sqrt{2\pi}}T^{-3/2}\exp\left\{\frac{-(a^2T+b^2/T)}{2}\right\}>0$$

for all a, e.g., $a<0$, so $F(T)$ is monotone increasing. Hence, since $\lim_{T\to\infty}F(T)=1$ for $a<0$ and $\lim_{T\to\infty}F(T)<1$ for $a>0$, $0\leqslant F(T)\leqslant 1$ for all T so we have more confidence in using the formula for $a < 0$ too.

$$\begin{array}{lll} \text{Check:} & E_\infty(a,b)\downarrow 0 \text{ as } \downarrow -\infty & \text{yes,} \\ & E_\infty(a,b)\uparrow \text{ as } b\uparrow & \text{yes,} \\ & E_\infty(a,b)\uparrow \text{ as} |a|\downarrow & \text{yes,} \\ \text{note} & \lim_{a\downarrow 0^+} E_\infty(a,b) = +\infty & \text{as suspected.} \end{array}$$

This leads us to believe that in a fair coin toss (fair means no drift) and a gambler with finite capital, the expected time to ruin is infinite.

This is correct. Feller gives $D = z(a-z)$ as the duration of the game, where z is initial capital, ruin is at 0, and a is the goal. Then $\lim_{a\to\infty} D(a) = +\infty$.

Note: $E_\infty = b/|a|$ means the expected time is the same as the point where $aT + b$ crosses $X(t) = 0$. See Figure 2.

$$E_\infty = b/|a|, \quad a = -m/s^2, \quad b = \ln\lambda,$$

$$\lambda = C/X_0 = \text{normalized goal},$$

$$m = p\ln(1+f) + q\ln(1-f) \equiv g(f),$$

$$s^2 = pq\{\ln[(1+f)/(1-f)]\}^2,$$

$$\text{Kelly fraction } f^* = p - q, \quad g(f^*) = p\ln 2p + q\ln 2q,$$

$$\text{For } m > 0, \quad E_\infty = (\ln\lambda)s^2/g(f).$$

Now this is the expected time in variance units. However s^2 variance units $= 1$ trial so

$$n(\lambda, f) \equiv \frac{E_\infty}{s^2} = \frac{\ln\lambda}{g(f)} = \frac{\ln\lambda}{m}$$

is the expected number of trials.

$$\begin{array}{lll} \text{Check:} & n(\lambda,f)\uparrow & \text{as } \lambda\uparrow, \\ & n(\lambda,f)\to\infty & \text{as } \lambda\to\infty, \\ & n(\lambda,f)\uparrow & \text{as } m\downarrow 0, \\ & n(\lambda,f)\to\infty & \text{as } m\to 0. \end{array}$$

Now $g(f)$ has unique maximum at $g(f^*)$ where $f^* = p - q$, the "Kelly fraction", therefore $n(\lambda, f)$ has a unique minimum for $f = f^*$. Hence f^* reaches a fixed goal in least expected time in this, the continuous case, so we must be asymptotically close to least expected time in the discrete case, which this approximates increasing by well in the sense of the CLT (Central Limit Theorem) and its special case, the normal approximation to the binomial distribution. The difference here is the trials are asymmetric. The positive and negative step sizes are unequal.

References

Associates First Capital Corporation, 1998. 1997 Annual Report. Associates First Capital Corporation, Dallas, TX.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–659.

Breiman, L., 1961. Optimal gambling systems for favorable games. In: Fourth Berkeley Symposium on Probability and Statistics, vol. I, pp. 65–78.

Browne, S., 1996. Reaching Goals by a Deadline: Continuous-Time Active Portfolio Management. Columbia University, New York.

Browne, S., 1997. The return on investment from proportional investment strategies. Advances in Applied Probability 30 (1), 216–238.

Buffett, M., Clark, D., 1997. Buffettology. Rawson Associates, Simon and Schuster, New York.

Cox, D.R., Miller, H.D., 1965. The Theory of Stochastic Processes. Wiley, New York.

Feller, W., 1966. An Introduction to Probability Theory and Its Applications, vol. I, Revised. Wiley, New York.

Griffin, P.A., 1979. The Theory of Blackjack. Huntington Press, Las Vegas. Revised 1995.

Hagstrom, R.G. Jr., 1994. The Warren Buffett Way. Wiley, New York.

Hagstrom, R.G. Jr., 2004. The Warren Buffett Way, second ed. Wiley, New York.

Hecht, R., 1995. Private correspondence.

Ibbotson Associates, 1998. Yearbook: Stocks, Bonds, Bills and Inflation (or any later edition). Ibbotson Associates, Chicago.

Institutional Investor, 1998. Ivory Tower Investing, pp. 43–55 (see p. 44), March.

Kelly, J.L., 1956. A new interpretation of information rate. Bell System Technical Journal 35, 917–926.

Kilpatrick, A., 1994. Of Permanent Value, the Story of Warren Buffet. Distributed by Southern Publishers Group, Birmingham, AL.

Kritzman, M., 1998. Risk and utility: basics. In: Bernstein, Damodaran (Eds.), Investment Management. Wiley, New York. Chapter 2.

Lowenstein, R., 1995. The Making of an American Capitalist. Random House, New York.

Markowitz, H., 1959. Portfolio Selection. Cowles Monograph, vol. 16. Wiley, New York.

McEnally, R.W., 1986. Latané's bequest: the best of portfolio strategies. Journal of Portfolio Management 12 (2), 21–30, Winter.

Quaife, A., 1995. Using the Sharpe ratio to evaluate investments. The Trans Times 4 (1), February. Trans Time Inc., Oakland, CA.

Rotando, L.M., Thorp, E.O., 1992. The Kelly criterion and the stock market. American Mathematical Monthly 99, 922–931, December.

Thorp, E.O., 1962. Beat the Dealer. Random House, New York.

Thorp, E.O., 1966. Beat the Dealer, second ed. Vintage, New York.

Thorp, E.O., 1969. Optimal gambling systems for favorable games. Review of the International Statistical Institute 37, 273–293.

Thorp, E.O., 1971. Portfolio choice and the Kelly criterion. In: Proceedings of the 1971 Business and Economics Section of the American Statistical Association, pp. 215–224.

Thorp, E.O., 1984. The Mathematics of Gambling. Lyle Stuart, Secaucus, NJ.

Thorp, E.O., Kassouf, S.T., 1967. Beat the Market. Random House, New York.

Thorp, E.O., Walden, W., 1966. A winning bet in Nevada baccarat, part I. Journal of the American Statistical Association 61, 313–328.

Tucker, H., 1967. A Graduate Course in Probability. Academic Press, San Diego, CA.

Wong, S., 1994. Professional Blackjack. Pi Yee Press, La Jolla, CA.