

Optimal betting under parameter uncertainty: improving the Kelly criterion

Rose D. Baker^{*} and Ian G. McHale[†]

Centre for Sports Business, Salford Business School, University of
Salford, M5 4WT, UK.

March 21, 2013

Abstract

The Kelly betting criterion ignores uncertainty in the probability of winning the bet, and uses an estimated probability. In general, such replacement of population parameters by sample estimates gives poorer out-of-sample than in-sample performance. We show that to improve out-of-sample performance the size of the bet should be shrunk in the presence of this parameter uncertainty, and compare some estimates of the shrinkage factor. From a simulation study and from an analysis of some tennis betting data we show that the shrunk Kelly approaches developed here offer an improvement over the ‘raw’ Kelly criterion. One approximate estimate of the shrinkage factor gives a ‘back of envelope’ correction to the Kelly criterion that could easily be used by bettors. We also study bet shrinkage and swelling for general risk-averse utility functions, and discuss the general implications of such results for decision theory.

Keywords

Kelly criterion, parameter risk, expected utility, bootstrap, tennis betting, shrinkage

1 Introduction

Statisticians are well aware that models fitted to data perform less well out of sample. For example, decisions such as how to invest made on the basis of limited experience will be suboptimal; this is ‘parameter risk’. Decision theory teaches us to act so as to maximize our expected utility (e.g. Berger, 1985), and whether we are adopting

^{*}email address: r.d.baker@salford.ac.uk

[†]email address: i.mchale@salford.ac.uk

a Bayesian or a frequentist approach, our optimal decision can only be based on our limited experience and on any knowledge we bring to the situation. Is the resulting decision based on maximizing expected utility the best that can be made by an agent, or can the maximized utility be modified to improve its out-of-sample performance, so giving on average a better decision?

The perhaps surprising answer is that maximized utilities can sometimes indeed be modified to increase their expected out-of-sample value, as exemplified by Kan and Zhou (2007). Assuming a multivariate normal distribution of returns, they showed that an optimal portfolio choice made au Markowitz could be modified for higher out-of-sample expected utility, and that their improved investment choice dominated the Bayesian choice, i.e. it gave a higher expected utility for all choices of parameter value.

We wish here to improve betting choices, where most simply the choice is what proportion s of a bankroll to bet on an event, given the quoted fractional odds b , and the probability p of winning. The practical value of this analysis goes beyond sports betting, into predictive markets and financial markets such as financial spread betting.

Our approach in this paper is frequentist but we start with the Bayesian formulation, to illustrate the problem. In general, if the utility of a win is u_w and the utility of a loss is u_l , given a probability p of winning, the expected utility is $pu_w + (1 - p)u_l$. When p is not known, the Bayesian approach is to supply a ‘prior’ distribution that embodies our belief about p . The expected utility then becomes $E(p)u_w + (1 - E(p))u_l$, with the expectation taken over the prior distribution. Because the utility is linear in p , the expected utility is a function only of $E(p)$, and so conventional utility maximization ignores the uncertainty in p .

In the frequentist approach, the estimated probability of a win is \hat{p} , and the bettor’s expected utility is $\hat{p}u_w + (1 - \hat{p})u_l$, where \hat{p} is the bettor’s estimate of p , based maybe on some formal experiment, such as tossing a biased coin a number of times, or maybe just a guess, informed or not. Clearly, the only effect of adopting a Bayesian rather than a frequentist approach to this problem is that \hat{p} is replaced by $E(p)$, which may not then be the observed proportion of wins but may be shifted towards the centre of the prior distribution for p , often $1/2$. But whether we are a Bayesian or a frequentist, we always effectively replace the unknown population parameter p by an estimate which we shall call \hat{p} . This is the ‘plug in’ approach. Often, the Bayesian use of a prior distribution outperforms the plug-in method, but here the two approaches are identical.

The uncertainty about p causes a problem when utility is maximized with respect to a decision variable, such as the proportion s of one’s ready money to bet on the event. For any value of the decision variable, the computed expected utility $u(s)$ may well be an unbiased estimator of the true expected utility. However, the computed maximized

expected utility $u(s^*(\hat{p}))$, where $s^*(\hat{p})$ is the bettor's optimum value of s , will not be an unbiased estimator of the out-of-sample maximized expected utility, and will in general over-estimate it, as will be seen later. This is the problem with which we are concerned: how can we improve the out-of-sample optimum value of the decision variable $s^*(\hat{p})$, given that our knowledge of p is uncertain?

Our approach can be thought of as bootstrapping the optimized expected utility. The bootstrapped optimized utility can then be further optimized by scaling the size of the investment s . Hence the logic is very similar to that of bootstrap adaptive sampling, where for example one might estimate a trimmed mean, choosing the trimming proportion to minimize the variance of the mean in the bootstrapped sample (Léger and Romano, 1990).

The famous Kelly betting criterion (Kelly, 1956), which maximizes a logarithmic utility function, is widely used to choose s . Hence we focus on the use of the Kelly criterion, although our results here have general implications for bettors and indeed for all decision makers who maximize utility functions under parameter uncertainty. The more general case is discussed in section 4.

We shall show that for the logarithmic utility function used in deriving the Kelly criterion, the required scaling of the bet is always shrinkage, but that for other utility functions, in rare cases with very favourable bets, the optimum bet size can swell. The optimum amount of bet shrinkage depends on our uncertainty about the probability of winning. When there is some knowledge of this, e.g. when data are available on the historical performance of a model for this probability, the error on the estimated probability can be estimated, allowing bets to be rescaled.

We now give a brief account of previous work on the Kelly criterion. After its introduction by Kelly (1956), the formula for the optimum proportion s^* has since been extensively studied (e.g. Thorp, 2006) and has been dubbed 'fortune's formula' by Poundstone (Poundstone, 2005). This is because, on always betting a constant proportion of one's wealth, the Kelly choice maximizes long-run wealth from a series of similar gambles when p is known. This follows because the expected utility using a logarithmic utility function is proportional to the logarithm of long-run expected wealth.

The Kelly criterion has a wide range of applications. For example, Johnson et al. (2006) use it as an investment strategy in their assessment of betting market efficiency. Griffin (1984) studied different measures of win rate when an investor adopts the Kelly strategy. There are extensive applications to card games (Thorp, 2006), and Kelly betting has been widely applied outside sport and gambling. There are applications to finance and the stock market (Rotando and Thorp, 1992, Medo et al. 2008), and to law (Barnett, 2010). It would seem applicable to insurance also; for example, in house contents

insurance one can pay more annually to insure contents to a higher value. Paying more insurance is analogous to raising the stake in betting. It has also been applied to situations where the outcomes are more complex, e.g. spread betting (Haigh, 2000, and Chapman, 2007). Other variations on the simple scheme are multiple outcomes (Barnett, 2010) and simultaneous betting (Whitrow, 2007).

There is very little in the literature on Kelly betting that addresses the problem of uncertain probabilities, but it is generally recommended that one should scale down one's bet in response to uncertainty in the probability of winning. Bettors often use a 'half-Kelly' strategy in which the Kelly-optimal bet is calculated and then halved. The rationale for partial Kelly strategies is also partly the fact that in the real world there is a minimum bet size. It is therefore necessary to be more cautious than the Kelly criterion implies, because one's bankroll is not infinitely divisible, and so ruin is possible.

Half-Kelly and more generally fractional Kelly strategies are used in betting on blackjack games (Thorp, 2006), and MacLean et al. (1992) study their use in investment analysis. A great deal of work has been done on Kelly and partial Kelly strategies in investment by MacLean, Ziemba and coworkers, e.g. MacLean, Ziemba and Blazenko (1992), MacLean and Ziemba (1999), MacLean, Sanagre, Zhao and Ziemba (2004), MacLean, Ziemba and Li (2005), MacLean, Thorp and Ziemba (2010) and MacLean, Thorp, Zhao and Ziemba (2011). A comprehensive collection of papers including a review of work on Kelly-based investment strategies has now been published (MacLean, Thorp and Ziemba, 2012).

Kadane (2011) shows that half-Kelly strategies do not correspond exactly to the optimization of any utility function, but that constant relative risk aversion utility, with the constant relative risk parameter equal to the reciprocal of the partial Kelly parameter is a good approximation to the utility function, particularly when the edge is small. That is, with utility function $U(f) = \frac{1-f^{1-\theta}}{\theta-1}$, where f is fraction of wealth invested, $\theta \simeq 1/\lambda$, where λ is the Kelly fraction.

The only paper that directly addresses the issue of uncertainty in winning probability is Medo et al. (2008), who discuss the problem of limited information in Kelly games. For example, they consider a sequence of wagers where the probability of winning is known to be randomly (with probability 1/2) $p+\Delta$ or $p-\Delta$, yet the odds offered are held constant. They consider an 'insider' who knows the probability of winning, and an 'outsider' who does not. The somewhat trivial finding is that the insider will accrue superior returns to the outsider. In general, given different types of information, more or less 'sharp' forecasts can be made (Johnstone et al., 2011). One can easily imagine other cases where 'insider' information would be available. For example, given a computer equipped with a webcam, one could conceivably make an improved prediction of how a tossed coin would

land, based on its motion before ‘heads’ or ‘tails’ were called. However, the ‘outsider’ only knows the probability that a tossed coin will come up heads.

In these examples even the ‘outsider’ knows the exact probability of winning a game. Having only a limited type of information of course means that a bettor will gain less from a bet, but there is no uncertainty about the probability of winning and thus the use of the Kelly criterion is still the best strategy, given a logarithmic utility function. Medo’s example shows that simply having imperfect information does not mean that we need to shrink the Kelly bet, and leads us to clarify what we mean by an ‘error in a probability’. The error in probability that we consider is the error in the estimate \hat{p} of the probability p that would be computed if we had perfect information *of the type allowed us*, i.e. not ‘insider’ knowledge.

McCardle and Winkler (1992) address the problem of optimal strategy under repeated gambles, where there is a similar setup to Medo et al. (2008), i.e. tosses of ‘good’ and ‘bad’ coins, although here the same coin would be used for all gambles. If few tosses are made, it is wise to be cautious. This paper is interesting but did not consider the logarithmic utility used in Kelly betting.

The next section introduces the logic of bet shrinkage, and section 3 presents some complications, and solutions, arising in real world applications. In section 4 we generalize the concept of bet shrinkage for the case of general risk-averse utility functions. In sections 5 and 6 we illustrate the method’s performance with a Monte-Carlo simulation, and with data from tennis betting, where the probabilities of a player winning are estimated from a simple model. Closing remarks are given in section 7.

2 Bet Shrinkage

In general, let the probability of an event occurring be p , and suppose that the sampling distribution of an estimate q (written as Q when considered as a random variable) of p has pdf $f(q)$, with mean p and variance σ^2 . In practice, Q may be any available estimate of p , even a wild guess.

If p were known, using a logarithmic utility function we would maximize the expected utility

$$E(u(s)) = p \ln(1 + bs) + (1 - p) \ln(1 - s)$$

where b is the fractional odds offered by the bookmaker, so that for a successful bet the bettor receives b units of the stake as profit. Calculus shows the optimum proportion of the bankroll to invest to be $s^*(p) = \frac{(b+1)p-1}{b}$, which is the Kelly betting formula.

If p has been estimated as Q we obtain the optimum bet as $s^*(Q)$, and hence the

expected maximized expected utility is

$$E(u^*) = \int_0^1 f(q) \{p \ln(1 + bs^*(q)) + (1 - p) \ln(1 - s^*(q))\} dq. \quad (1)$$

The utility $E(u^*)$ will be somewhat lower than the naïve estimate of the maximized utility, as the integrand would be maximized by setting $s^*(q) = s^*(p)$, when $E(u^*)$ would reduce to the naïve estimate of expected utility,

$$p \ln(1 + bs^*(p)) + (1 - p) \ln(1 - s^*(p)).$$

This is the situation that this paper is concerned with: when p is not known exactly, the expected gain is smaller than the maximization of the expected utility would suggest.

We now seek to increase the expected utility by scaling the bet from the Kelly value of $s^*(Q)$ to $ks^*(Q)$, when the expected utility becomes

$$E(u^*) = \int_0^1 f(q) \{p \ln(1 + bks^*(q)) + (1 - p) \ln(1 - ks^*(q))\} dq. \quad (2)$$

With a slight abuse of notation we continue to write $E(u^*)$ to denote the expected maximized utility even when $s^*(Q)$ has been scaled by k .

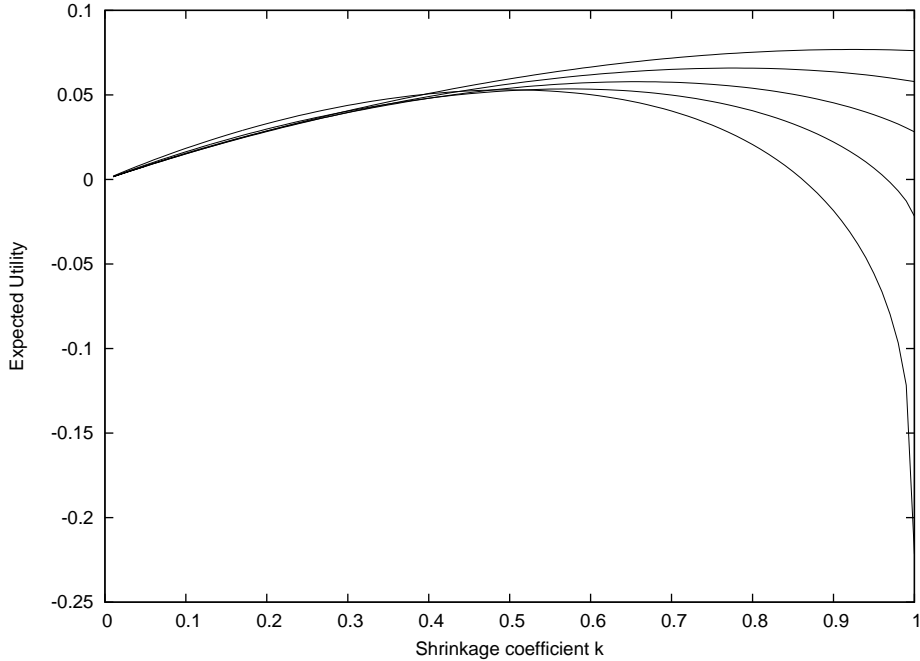


Figure 1: Expected utility of the optimum Kelly bet shrunk by the shrinkage coefficient k , when $b = 1$, $p = 0.7$. The curves have $\sigma = 0.05, 0.1, 0.15, 0.2, 0.25$ (lowest at top).

We shall be able to increase $E(u^*)$ by optimizing with respect to k . Figure 1 shows

the utility of a bet for different values of σ , where f is taken as a beta distribution. It shows how optimal shrinkage increases as σ increases. It can be seen that bet shrinkage can convert a negative utility (a disutility) of betting into a positive utility.

Theorem 1 *Expected maximized utility $E(u^*)$ can be increased by shrinking the bet size.*

Proof. To prove that utility can be increased by shrinking the bet size, we show that $dE(u^*)/dk|_{k=1} < 0$ for non-degenerate $f(q)$. We have that

$$dE(u^*)/dk|_{k=1} = \int_0^1 f(q)s^*(q)\left\{\frac{pb}{1+bs^*(q)} - \frac{1-p}{1-s^*(q)}\right\}dq,$$

which reduces to

$$dE(u^*)/dk|_{k=1} = 1 - (1/(b+1))E\{p/Q + (1-p)b/(1-Q)\}.$$

Since $1/q$ and $1/(1-q)$ are convex functions, Jensen's inequality, here that $E(1/Q) > 1/E(Q)$ and $E(1/(1-Q)) > 1/E(1-Q)$ applies. One could also derive these inequalities directly from the Cauchy-Schwarz inequality. Then the fact that $E(Q) = p$ gives the required result. When there is no uncertainty in the probability of winning, the inequality becomes an identity. ■

It is straightforward to show that the curvature $d^2E(u^*)/dk^2$ is negative and that $dE(u^*)/dk|_{k=0} > 0$, so that there is a unique maximum utility for $0 < k < 1$.

The next step is to find the optimum value of the scaling parameter k . This parameter satisfies the equation

$$dE(u^*)/dk = \int_0^1 f(q)s^*(q)\left\{\frac{pb}{1+bks^*(q)} - \frac{1-p}{1-ks^*(q)}\right\}dq = 0. \quad (3)$$

and the optimum value k^* may be found using the Newton-Raphson iteration

$$k_{n+1} = k_n - \frac{dE(u^*)/dk}{d^2E(u^*)/dk^2}.$$

To carry out this procedure in practice one requires a specific form for the pdf $f(q)$. Assuming that only the variance of the sampling distribution can be specified, we can only choose a 2-parameter distribution, and the beta distribution is the obvious choice. Herein we refer to this method as the beta method. In Bayesian inference, the beta distribution is the posterior distribution for the probability p following Bernoulli trials if a conjugate prior is used. However, here in our frequentist approach it is used instead to approximate the sampling distribution of \hat{p} . The beta pdf is $f(q) \propto q^{\alpha-1}(1-q)^{\beta-1}$, where $\alpha = p\{p(1-p)/\sigma^2 - 1\}$, $\beta = (1-p)\{p(1-p)/\sigma^2 - 1\}$. The maximum value of

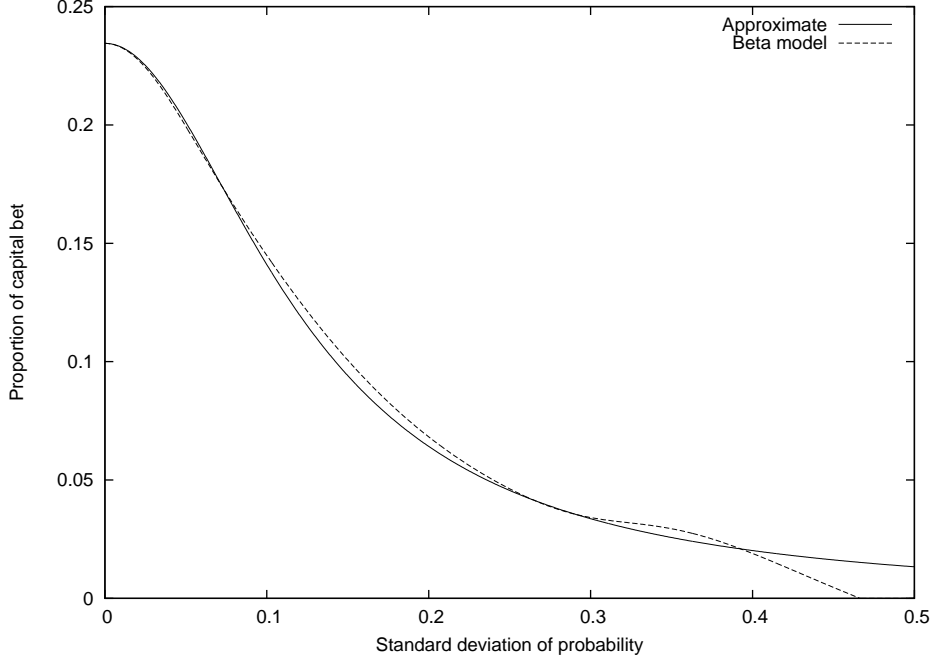


Figure 2: Bet size s^* as a function of standard deviation of win probability estimate for tennis betting, where expected win probability is $p = 0.599$ and the odds are $b_1 = 1.1$, $b_2 = 0.72$. The curves are for the approximate method and for the beta-distribution method.

σ that can be attained is $1/2$ (Johnson, Kotz and Balakrishnan, 1995) and this is also the value resulting from a probability p that is 0 or 1 with equal probability. A uniform distribution of probability has $\alpha = \beta = 1$ and $\sigma = (1/12)^{-1/2} \simeq 0.289$.

First and second-order approximations to k^* can be derived, and the first-order approximation (small σ) is a simple formula that gives a reasonable approximation to k .

Expanding u^* about $s^*(p)$, we have

$$E(u^*) = E(u(s^*(p))) + (1/2)\partial^2 E(u(x))/\partial x^2|_{\{x = s^*(p)\}} \times \int_0^1 (ks^*(q) - s^*(p))^2 f(q) dq + \dots \quad (4)$$

where the omitted term $\partial E_p(u(x))/\partial x|_{\{x = s^*(p)\}} = 0$. Differentiating (4) with respect to k and equating to zero, we obtain the optimum value k^* as

$$k^* = \frac{((b+1)p-1)^2}{((b+1)p-1)^2 + (b+1)^2\sigma^2} = \frac{s^*(p)^2}{s^*(p)^2 + ((b+1)/b)^2\sigma^2}. \quad (5)$$

From this equation, the half-Kelly method ($k = \frac{1}{2}$) is optimal when approximately $\sigma = p - 1/(1+b)$. Figure 2 shows how bet size decreases with error on probability using this approximate method and the ‘beta method’.

In reality, we do not know p and σ^2 , and we have only estimates \hat{p} and $\hat{\sigma}^2$ of these two parameters. The optimum shrinkage parameter k^* is of course also just an estimate of the

optimum amount of shrinkage. But given that some shrinkage is needed, the examples will show that using this method is still better than using the ‘raw’ Kelly criterion.

3 Some complications

We have so far ignored the fact that negative bets are not always possible in the real world. If $Q < 1/(1 + b)$ in (1), then the Kelly solution is $s^*(Q) < 0$. It is possible to lay bets (make such negative bets, which are analogous to short-selling) via betting exchanges. This is the case that has been implicitly considered so far. However, at the racetrack one would simply not bet and the expected utility would be zero. For such a bettor, the integral over $f(q)$ for example in (1) would have a lower limit of $1/(1 + b)$.

The inability to place negative bets does not affect the approximate formula (5) for k^* that applies when σ is small, but it does invalidate the very general proof that $k^* < 1$ for any unbiased sampling pdf $f(q)$. The optimum value of k derived from solving (3) using a beta distribution for f is then increased somewhat from what it would be if bets could be laid. However, we have never found an example where bet swelling rather than bet shrinkage occurs.

If negative bets are impossible, at least two different cases can be distinguished. In one, if the odds are unfavourable so that $Q < 1/(1 + b)$ the bettor simply does not bet. If so, the lower limit of the integral in (3) changes to $1/(1 + b)$. In betting on which of two players will win a match, however, where the bookies offer odds of b_1, b_2 on the two players, if $Q < 1/(1 + b_2)$ one would bet on player 2. As the predicted probability of player 1 winning increases, one switches from betting on player 2 to not betting, then to betting on player 1. The integrand of (3) changes accordingly. If the odds for bets on both players to win were such that both bets would be worthwhile, one would make the choice that maximized utility. One might wish to bet on both players given favourable odds on each, but we ignore this remote possibility; bookmakers are hardly likely to offer such odds! The decision tree is then:

```

if ( $s_1(p) < 0$  &  $s_2(1 - p) < 0$ ) then no bet
else if ( $s_1(p) > 0$  &  $s_2(1 - p) < 0$ )
or  $s_1(p) > 0$  &  $p > \ln\{b_1(1 + b_1 2)/(1 + b_1)\}/\ln(b_1 b_2)$  then
bet on player 1
else
bet on player 2

```

and the integrand of (3) is modified to reflect this decision tree. The point is that the estimated expected utility of making a scaled Kelly bet has to correspond to the bettor’s

actual practice.

4 Bet rescaling for general risk-averse utility functions

The Kelly method uses a logarithmic utility function. We now generalize to bet rescaling using general risk-averse utility functions u .

In general,

$$E(u) = pu(1 + bs^*) + (1 - p)u(1 - s^*),$$

from which at $s^*(p)$ the first derivative $E(u') = 0$. For a risk-averse utility function with $u'' < 0$, the second derivative $E(u'') < 0$, so $s^*(p)$ maximizes the utility rather than minimizing it. To study bet scaling for general risk averse utility functions we again consider $dE(u^*)/dk|_{k=1}$; bet shrinkage occurs if this differential is negative. Differentiating

$$E(u^*) = E\{pu(1 + bs^*(Q)) + (1 - p)u(1 - ks^*(Q))\}$$

we obtain

$$dE(u^*)/dk|_{k=1} = E\{s^*(Q)[pbu'(1 + bs^*(Q)) - (1 - p)u'(1 - s^*(Q))]\}.$$

We note that

$$dE(u^*)/dk|_{k=0} = (bp - (1 - p))u'(1)E\{s^*(Q)\},$$

which is positive if the bet would give a positive expected gain. Thus $k^* > 0$.

We shall henceforth denote the n th derivative of s^* with respect to its argument as s_n^* . Using the identity

$$bqu'(1 + bs^*(q)) = (1 - q)u'(1 - s^*(q)) \tag{6}$$

we obtain

$$dE(u^*)/dk|_{k=1} = -E\left\{\frac{(Q - p)s^*(Q)u'(1 - s^*(Q))}{Q}\right\}. \tag{7}$$

This equation can be recast as

$$dE(u^*)/dk|_{k=1} = -E\left\{\frac{(Q - p)s^*(Q)bu'(1 + bs^*(Q))}{1 - Q}\right\}$$

or as

$$dE(u^*)/dk|_{k=1} = -E[(Q-p)s^*(Q)\{bu'(1+bs^*(Q)) + u'(1-s^*(Q))\}]. \quad (8)$$

To show shrinkage, writing in general

$$dE(u^*)/dk|_{k=1} = -E\{(Q-p)h(Q)\}$$

a sufficient condition for shrinkage is that $h'(Q) > 0 \forall Q$. Then

$$E\{(Q-p)h(Q)\} = E\{(Q-p)(h(Q) - h(p))\} > 0$$

as $h(Q) > h(p)$ if $Q > p$, else $h(Q) < h(p)$.

We can simplify further by specialising to the small σ case, when for example

$$dE(u^*)/dk|_{k=1} = -\sigma^2 s_1^* \{bu'(1+bs^*) + b^2 s^* u''(1+bs^*) + u'(1-s^*) - s^* u''(1-s^*)\} \quad (9)$$

where s^* and s_1^* are evaluated at p . However, this simplification is just for ease of presentation, and the results only assume $h(q)' > 0$, which is true also when σ is not small.

Theorem 2 *The condition $s_1 > s/p$ is sufficient for shrinkage.*

Proof. From (7) for small σ we have that

$$dE(u^*)/dk|_{k=1} = -\frac{\sigma^2}{p} \{s_1^* u'(1-s^*) - s^* s_1^* u''(1-s^*) - s^* u'(1-s^*)/p\}.$$

From this equation, as $u'' < 0$, a sufficient condition for shrinkage is that $s_1^* > s^*/p$. ■

This condition is of limited use, as it is not expressed directly in terms of the utility function. We can see that it is satisfied for the logarithmic utility function.

Theorem 3 *The condition*

$$R_r(1+bs^*) < \frac{b^2}{b^2-1}/(1-p). \quad (10)$$

is sufficient for shrinkage, where R_r is the Arrow-Pratt measure of relative risk aversion.

Proof. By considering the shrinkage condition $-(b^2-1)s^* u''(1+bs^*) < bu'(1+bs^*)/(1-p)$ we see that shrinkage must occur if the Arrow-Pratt measure of relative risk aversion, $R_r(x) = -xu''(x)/u'(x)$ satisfies (10). ■

Shrinkage will thus always occur either if the relative risk aversion coefficient of the utility function satisfies (10). This condition covers the logarithmic utility function and isoelastic utility function, $u(x) = (x^\alpha - 1)/\alpha$, where $0 < \alpha < 1$, for which $R_r \leq 1$.

From differentiating (6), $s_1^* > 0$. The only negative term in the sum in (9) is $b^2 s^* u''(1 + bs^*)$. Shrinkage does not occur for the exponential utility function $u(x) = -\exp(-\lambda x)$, where λ is the risk aversion parameter. For this utility, from the identity (6) we have that

$$s^*(p) = \ln \frac{bp}{1-p} / (b+1)\lambda. \quad (11)$$

The proportion of wealth bet can exceed unity, as the utility function is defined for negative wealth, and the bettor can borrow. This has sadly happened many times, e.g. with the Nigerian 419 scam, when duped individuals have borrowed heavily from friends and family to take advantage of a seemingly wonderful opportunity.

Equation (11) for $s^*(p)$ gives

$$dE(u^*)/dk|_{k=1} \propto -[1 + \ln \frac{bp}{1-p} \{p - b/(b+1)\}].$$

Clearly, if the odds are very good so that $b \gg 1$, then $b/(b+1) > p$ and $dE(u^*)/dk|_{k=1} > 0$ and the optimum bet will swell instead of shrinking. Intuitively, this swelling is because this utility function does not penalize a loss as severely as does the logarithmic utility function, so when the odds are very favourable, it is preferable to swell one's bet rather than shrink it. This situation will be very rare in practice, as excellent odds of winning are not often on offer.

Knowing that risk-averse utility functions can sometimes produce swelling rather than shrinking of bets, we now proceed to find another sufficient condition for shrinkage.

Theorem 4 *The conditions*

$$p \leq \frac{2b^2 + b - 1}{b^3 + b^2 - b - 1}$$

and $u''' > 0$ are sufficient for shrinkage.

Proof. Taylor's theorem gives $g(a) = g(x) - g'(\xi)(x - a)$, where $a < \xi < x$, so that if $g'(\xi)$ decreases with ξ , $g(a) > g(x) - g'(x)(x - a)$. In our case we have $u'(1 - s^*) > u'(1 + bs^*) - u''(1 + bs^*)(b + 1)s^*$. This inequality follows if u'' decreases with s . We have that $u''(\xi) > -u'(a)/(x - a)$, so u'' increases towards zero, but the increase need not be monotonic. Hence we must also assume smoothness of u , expressed by $u''' > 0$. Then $u''(x) = u''(a) + u'''(\xi)(x - a) > u''(a)$. ■

All utility functions known to us have this property. Further, $u'(1 + bs^*) = \frac{1-p}{bp} u'(1 -$

s^*), so

$$u'(1 - s^*) + bu'(1 + bs^*) = u'(1 - s^*)/p > \frac{-(b+1)s^*u''(1 + bs^*)}{1 - (1-p)/bp}. \quad (12)$$

Using this inequality, we have that for small σ^2

$$E(u^*)/dk|_{k=1} < -\sigma^2 s_1^* \{-s^*u''(1 - s^*) + (b^2 - \frac{b+1}{p - (1-p)/b})s^*u''(1 + bs^*)\}.$$

Since $u''(1 + bs^*) > u''(1 - s^*)$,

$$E(u^*)/dk|_{k=1} < -\sigma^2 s_1^* (b^2 - 1 - \frac{b+1}{p - (1-p)/b})s^*u''(1 + bs^*). \quad (13)$$

From (13) a sufficient condition for shrinkage is that

$$b^2 - \frac{b+1}{p - (1-p)/b} \leq 1,$$

or

$$p \leq \frac{2b^2 + b - 1}{b^3 + b^2 - b - 1}.$$

For $b = 2$ this equation yields $p \leq 1$, so bet shrinkage rather than swelling must always occur for $b \leq 2$ for any smooth risk-averse utility function.

Finally, the approximation to the optimum scaling factor k^* derived from (4) can be derived readily for any utility function, not just the logarithmic.

Constructing the bootstrapped estimator $E(u^*(ks(Q)))$, expanding by Taylor series about $s^*(p)$ and maximizing with respect to k , we obtain

$$k^* = \frac{s^*(p)E(s^*(Q))}{E(s^*(Q)^2)} \quad (14)$$

which is valid for small $ks^*(Q) - s^*(p)$, i.e. for small variance σ^2 . The form (14) does not depend explicitly on the form of u , but of course does so implicitly, because $s^*(Q)$ is a function of u . Expanding $s^*(Q)$ in a Taylor series about p we obtain

$$k^* = \frac{s^*(p)\{s^*(p) + (1/2)s_2^*(p)\sigma^2\}}{s^*(p)^2 + (s_1^*(p))^2 + s^*(p)s_2^*(p)\sigma^2}. \quad (15)$$

Thus, in this section, we have obtained the general approximate formula for bet rescaling under risk-averse utility functions (15) and have asked whether bet shrinkage rather than swelling under probability uncertainty is bound to occur under any risk-averse utility function. Surprisingly, we found that bet swelling can occur for some utility functions (of which the exponential is an example) when the odds are very good.

5 Examples

5.1 A Simulated gambling study

We measured the performance of the bet shrinkage methods introduced here by examining the mean utility of bets, compared with the ‘raw’ Kelly method. The first example is a simulated gambling study.

Suppose that a bettor obtains even odds on an event, such as whether a six can be thrown in 5 tosses of a die. The probability is $1 - (5/6)^5 \simeq 0.57$. This bet is a more favourable version of a famous bet made by the Chavalier de Méré, that he could throw an ‘ace’ in 4 tosses of a die, for which the probability is $1 - (5/6)^4 \simeq 0.5177$ (e.g., David, 1987).

We imagine that the bettor possesses ‘bounded rationality’, so that he does not know the true probability of winning. However, he estimates the probability of winning by retiring to a quiet corner and performing sets of tosses. To obtain the results in table 1, we generated 10000 simulations. Each simulation contained 100 trials, where half the time the number of sets of tosses performed was drawn uniformly between 8 and 20. The bettor then estimates the true probability, and also calculates a standard error for his estimate. For the other half of the simulations, the bettor was assumed to know the exact probability of winning. This simulation was intended to represent a situation where the probability of winning may sometimes be known very accurately, but may sometimes be known with various degrees of error.

Method	Mean final bankroll	Median	S.d.	Mean utility
Kelly (and shrinkage)	44.5	7.48	189.5	1.98
Half-Kelly	6.88	4.41	8.05	1.47
Kelly	57.1	2.1	669.0	0.682
Half-Kelly	7.4	3.5	14.3	1.238
Approx shrinkage	26.3	2.7	189.9	0.936
Beta shrinkage	27.2	3.0	183.1	1.061
Binomial shrinkage	26.5	3.1	171.8	1.090

Table 1: Final bankroll (mean, median and standard deviation) and logarithmic utility from 100 sequential bets, starting with a unit (favourable) bet, in simulated gambling using various methods. In the first two rows the probability of winning is known exactly.

Looking at the expected utilities, in simulations where the probability of success was known exactly, table 1 shows firstly that the ‘raw’ Kelly method outperforms the half-Kelly method as it must (in the first two rows of results). However, the half-Kelly method outperforms the Kelly method when the true probability of success is not known exactly.

Although the average final bankroll was greater using the Kelly criterion, the mean utility was less than unity, so it would have been better not to bet at all.

We tested a small-error approximate shrinkage formula (5) and the ‘beta’ method, based on solving (3) exactly for the optimum shrinkage k^* , but unavoidably then assuming a functional form for the distribution of estimated probability, here the beta distribution. Equation (3) was modified for ‘no bet’ if $\hat{p} < 1/2$, and the (correct) binomial sampling distribution was also used.

Our ‘beta’ method outperforms the Kelly method when the probability of success is not known exactly, and gives a final mean utility greater than unity, so it is better to bet using this system than not to bet. The approximate method does not perform as well, but is still better than the ‘raw’ Kelly method. The binomial sampling distribution performs best of all. The point is, of course, that the proposed method is adaptive and dominates the Kelly method, in that it never performs worse in terms of mean utility, whereas the half-Kelly method is not adaptive and does not dominate the Kelly method.

Table 2 shows results for the same situation, except that the bet is now that a six can be thrown in 3 tosses of a die, for which the probability is only $1 - (5/6)^3 \simeq 0.42$. Now the Kelly bet is zero, but with an imperfect knowledge of the probability of winning one will bet. The losses are reduced using shrinkage methods, and again the half-Kelly method performs well. This tallies with the results of Grant and Johnstone (2010) who found that a 40% Kelly bet based on expert opinion gave the best returns for betting on Australian Football League matches.

Method	Mean final bankroll	Median	S.d.	Mean utility
Kelly (and shrinkage)	1	1	0	0
Half-Kelly	1	1	0	0
Kelly	0.709	0.527	0.657	-0.698
Half-Kelly	0.841	0.785	0.345	-0.253
Approx shrinkage	0.855	0.786	0.450	-0.296
Beta shrinkage	0.835	0.758	0.451	-0.322
Binomial shrinkage	0.838	0.763	0.443	-0.313

Table 2: Final bankroll (mean, median and standard deviation) and logarithmic utility from 100 sequential bets, starting with a unit (unfavourable) bet, in simulated gambling using various methods. In the first two rows the probability of winning is known exactly.

For both favourable and unfavourable bets, shrinkage increases the expected utility over the ‘raw’ Kelly criterion. For favourable bets, the unshrunk Kelly criterion gives a high mean bankroll, but this is deceptive; many bettors will lose, and only a few will make large winnings. Shrinkage however increases the median final bankroll.

5.2 Tennis betting

We obtained data on the results of matches from the top tier of men’s professional tennis, the ATP Tour, for 2005-2011 (15,794 matches) from www.tennis-data.co.uk. The data included the participants’ names and ATP world rankings points at the time of the match, the match result and up to six bookmaker’s odds for each game. Following Boulier and Stekler (1999) and Clarke and Dyte (2000), we adopt a simple logistic regression model for estimating the probability of victory for the higher ranked (better) player using the natural logarithm of the ratio two players’ world rankings points as the only covariate. We use data from 2005-2008 (the in-sample) to make predictions for matches from 2009 (the out-sample). In order to reduce any bias in the resulting predictions, for each predicted match in the out-sample we randomly selected 1000 matches from the in-sample to estimate the probability of the higher ranked player winning the match.

In applying our methodology, the first problem is to estimate the value of the error on the predicted probability. This is a complex issue. To cut the Gordian knot, we simply calculated the log-odds ratios for bookmaker and model predictions, and partitioned the error equally between the bookmaker and the model, estimating the error variance of the predicted log-odds as half the variance of the difference between these two sets of log-odds. The delta-method was then used to find the variance σ^2 of \hat{p} . This can only be a crude estimate of σ^2 , partly because bookmaker probabilities may not be conditioned on the same information set as the model, so some of the disagreement may arise from the bookmaker’s predictions being ‘sharper’. Needless to say, simply using the statistical error of the model probability computed from the Hessian would greatly underestimate σ^2 , because that computation assumes that the model is correct.

The results of applying various modifications of the Kelly method to these data are given in table 3. As can be seen, the shrinkage methods outperform the ‘raw’ Kelly method, especially in terms of mean utility. The ‘beta’ method and the approximate method perform similarly well, and again the half-Kelly method does well. Whereas the approximate method can only shrink the bet, the ‘beta’ method sometimes shrinks it to zero. In fact, the optimum value of the shrinkage coefficient k was negative, so the optimum feasible value was zero.

One might be disappointed that these methods do not enable one to ‘beat the bookie’. Despite the mean bankroll for the approximated shrinkage method exceeding unity, the excess is not statistically significant. This lack of significance is probably because our ranking model is not as good as the bookmaker’s ‘model’. It gave a slightly larger Brier score (Brier, 1950) of 0.2271, compared to 0.1886 for the bookmaker’s probabilities. However, the aim of this paper is merely to demonstrate the advantage of bet shrinkage when probabilities are in error, not to beat the bookmaker.

Method	Mean	Mean utility	% on player 1	% on player 2	% no bets
Kelly	0.984	-0.0482	30.85	49.65	19.49
Half-Kelly	0.992	-0.0161	30.85	49.65	19.49
Approx shrinkage	1.001	-0.0024	30.85	49.65	19.49
‘Beta’ shrinkage	1.000	-0.0027	9.98	35.85	54.17

Table 3: Final bankroll and mean utility from tennis betting using various methods described in the text.

6 Conclusions

In this paper we have highlighted an issue in decision theory that when maximizing a utility function, optimal investment size should change once one acknowledges that the probabilities involved are not known with certainty. The basis of our method is the fact that optimized utility is an upwards biased estimator of expected maximum utility, and so optimized utility can be further optimized by bet rescaling, which is usually shrinkage.

Demonstrating this concept using the Kelly criterion betting formula, we show, in theory and in practice, that shrinking the investment size provides an improved expected utility in the presence of such uncertainty. We generalize this idea of shrinkage to that of all risk-averse utility functions, and find that, somewhat surprisingly, for utility functions with a large relative risk aversion coefficient, utility could sometimes be increased by swelling bet size rather than shrinking it. It may seem paradoxical that the response to increased uncertainty could be a bolder action, but the very favourable bets that should be swelled under some utility functions are frankly unlikely to be on offer in the real world.

There is a general implication for decision theory: regarding the practical problem of maximizing utilities, one should be able to get higher expected utilities using our method than using naive maximization, and this methodology can help in other cases even when the utility is not exactly linear in a parameter. This has already been demonstrated for Markowitz optimization by Kan and Zhou (2007). They were able to find an investment strategy that dominated the Bayesian strategy. Our case is in some ways simpler than theirs, because there is a scalar decision variable rather than a vector of monies to be invested in different shares. However, for Kelly betting the shrinkage factor is the solution of a nonlinear equation involving an integral, making the Kelly case more complex, and it has not been possible to demonstrate dominance of a shrunken bet over the Bayesian solution.

Our method of shrinkage relies on there being an estimate of the uncertainty in probability, σ , and of course this is itself hard to quantify. It can be estimated from fitting

models to data, but if the model is wrong, the statistical error on \hat{p} will be a spuriously low estimate of the true error. Unfortunately, it is when σ is largest that it is likely to be hardest to estimate. One can at least carry out a sensitivity analysis by using (3) to see how the optimum size of bet s^* changes as the error σ increases.

Shrinkage also works in more complex instances of betting, for example the multiple outcomes case, such as eachway betting in horse races, and the simultaneous betting case. This latter is a multivariate problem, because the sizes of each of a number of bets may be shrunk by differing amounts. There is much further work that could be done in this area, both in deriving shrinkage formulae for more general decision problems, and also in evaluating their performance.

Finally, Poundstone (2005) dubbed the Kelly criterion ‘fortune’s formula’. Can it be made even more fortunate? Our view is that the equation

$$s^* = \frac{((b+1)p-1)^3}{b\{((b+1)p-1)^2 + (b+1)^2\sigma^2\}}$$

derived from (5) would be a strong contender.

Acknowledgements

We thank the editor, associate editor and anonymous referees for their comments and suggestions which have improved the paper.

References

- Barnett T 2010 Applying the Kelly criterion to lawsuits. *Law, Probability and Risk* 9(2):139-147.
- Berger JO 1985 *Statistical Decision Theory and Bayesian Analysis* 2nd ed. Springer, New York.
- Boulier BL and Stekler HO 1999 Are sports seedings good predictors? An evaluation. *International Journal of Forecasting*, 15(1):83-91.
- Brier GW 1950 Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78(1) 1-3.
- Chapman S J 2007 The Kelly criterion for spread bets. *IMA Journal of Applied Mathematics* 72(1):43-51.

- Clarke SR and Dyte D 2000 Using official ratings to simulate major tennis tournaments. *International Transactions in Operational Research* 7(6):585-594.
- David F N 1987 *Games, gods and gambling*. Griffin, London.
- Grant A and Johnstone D 2010 Finding profitable forecast combinations using probability scoring rules. *International Journal of Forecasting* 26(3):498-510.
- Griffin P A 1984 Different measures of win rate for optimal proportional betting. *Management Science*, 30(12):1540-1547.
- Haigh J 2000 The Kelly criterion and bet comparisons in spread betting. *Journal of the Royal Statistical Society series D* 49(4):531-540.
- Johnson J E V, Jones O and Tang L 2006 Exploring decision makers' use of price information in a speculative market. *Management Science* 52(6):897-908.
- Johnson N L, Kotz S and Balakrishnan N 1994 *Continuous Univariate Distributions* vol. 2 Wiley, New York.
- Johnstone DJ, Jose VRR and Winkler R L 2011 Tailored scoring rules for probabilities. *Decision Analysis* 8(4):256-268.
- Kadane JB 2011 Partial-Kelly strategies and expected utility: small-edge asymptotics. *Decision Analysis* 8(4):4-9.
- Kan R and Zhou G 2007 Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis* 42(3):621-656.
- Kelly JL Jr 1956 A new interpretation of information rate. *Information Theory, IRE Transactions on* 2(3):185-189.
- Léger C and Romano JP 1990 Bootstrap adaptive estimation: The trimmed-mean example. *The Canadian Journal of Statistics* 18(4):297-314.
- MacLean LC, Sanegre R, Zhao Y and Ziemba WT 2004 Capital growth with security. *Journal of Economic Dynamics & Control* 28(5):937-954.
- MacLean LC, Thorp, EO, Zhao Y and Ziemba WT 2011 How Does the Fortune's Formula Kelly Capital Growth Model Perform? *Journal of Portfolio Management* 37(4):96-111.

- Maclean LC, Thorp EO, Ziemba WT 2010 Long-term capital growth: the good and bad properties of the Kelly and fractional Kelly capital growth criteria. *Quantitative Finance* 10(7): 681-687.
- MacLean LC, Ziemba WT 1999 Growth versus security tradeoffs in dynamic investment analysis. *Annals of Operations Research* 85(1):193-225.
- MacLean LC, Thorp EO, Ziemba WT 2012 Eds. *The Kelly Capital Growth Criterion: Theory and Practice*. World Scientific, Singapore.
- Maclean LC, Ziemba WT and Blazenko G 1992 Growth versus security in dynamic investment analysis. *Management Science* 38(11):1562-1585.
- MacLean LC, Ziemba WT, Li YM 2005 Time to wealth goals in capital accumulation *Quantitative Finance* 5(4):343-355.
- McCardle KF and Winkler RL 1992 Repeated gambles, learning, and risk aversion. *Management Science* 38(6):807-818.
- Medo M, Pis'mak YM and Zhang Y 2008 Diversification and limited information in the Kelly game. *Physica A* 387(24):6151-6158.
- Poundstone W 2005 *Fortune's Formula: The Untold Story of the Scientific Betting System that Beat the Casinos and Wall Street*. Hill and Wang, New York.
- Rotando LM and Thorp EO 1992 The Kelly criterion and the Stock Market. *American Mathematical Monthly* 99(10):922-931.
- Thorp EO 2006 The Kelly criterion in blackjack, sports betting, and the stock market, ch. 9 in *Handbook of Asset and Liability Management: Theory and Methodology*. Zenios S. A., Ziemba W. A. Eds.
- Whitrow C 2007 Algorithms for optimal allocation of bets on many simultaneous events. *Appl. Statist.* 56(5):607-623.