Dani Chu, Yifan Wu and Tim B. Swartz*

Modified Kelly criteria

https://doi.org/10.1515/jqas-2017-0122

Abstract: This paper considers an extension of the Kelly criterion used in sports wagering. By recognizing that the probability p of placing a correct wager is unknown, modified Kelly criteria are obtained that take the uncertainty into account. Estimators are proposed that are developed from a decision theoretic framework. We observe that the resultant betting fractions can differ markedly based on the choice of loss function. In the cases that we study, the modified Kelly fractions are smaller than original Kelly.

Keywords: Bayes estimation; Kelly criterion; loss functions; minimax estimation.

1 Introduction

In its application to sports gambling, the Kelly criterion (Kelly 1956) provides a gambler with the optimal fraction of a bankroll that should be wagered on a given bet. In determining the Kelly criterion, a gambler needs to specify the probability p of placing a correct (i.e. winning) wager using a specified gambling *system*.

The Kelly criterion has received widespread attention, and some of the attention has been negative (Samuelson 1979). Experienced gamblers claim that the Kelly fraction is too high and often leads to financial loss (Murphy 2015). As a consequence, advice has been provided to instead use fractional Kelly approaches, such as "half-Kelly." With half-Kelly, a sports gambler wagers only half of the fraction of the bankroll that the Kelly criterion specifies.

The perplexing aspect of these negative experiences is that the Kelly criterion is based on mathematical proof. The Kelly criterion is optimal from several points of view; for example, it maximizes the exponential rate of growth and it provides the minimal expected time to reach an assigned balance (Breiman 1961). Therefore, how can it be that gamblers often experience losses when using the Kelly approach? The simple but often overlooked explanation is that the input p used in determining the Kelly fraction is an unknown quantity. Often, gamblers are overly optimistic concerning their gambling systems and the true p is less than the specified p.

In this paper, we take a statistical approach to the problem where we account for the uncertainty in p. The resulting fractions of the bankroll which we derive tend to be less than the Kelly criterion. Therefore, we provide a systematic approach for obtaining "modified Kelly criteria" rather than ad-hoc fractions such as half-Kelly or quarter-Kelly that have no theoretical underpinnings.

There are many papers that propose systems and insights with respect to gambling. For example, in greyhound racing, Schumaker (2013) used data mining techniques to identify profitable wagers. In basketball, Lopez and Matthews (2015) used logistic regression techniques to predict outcomes in the NCAA men's basketball tournament. In soccer, Feng, Polson, and Xu (2016) used the Skellam process to represent real-time betting odds for English Premier League soccer. However, there is only a scattered and limited scientific literature on the mathematical properties of sports gambling. Thorp (1969) provided probabilistic results concerning optimal systems for favourable games. Insley, Mok, and Swartz (2004) extended the Kelly criterion to situations where simultaneous wagers are placed. Under simultaneous wagering, the simple Kelly criterion is no longer applicable and computation is required to obtain optimal fractions. With respect to mathematical and probabilistic treatments related to finance and investing, MacLean, Thorp, and Ziemba (2011) provide a comprehensive edited volume with contributions that focus on various financial problems involving the Kelly criterion. For example, a number of papers in MacLean et al. (2011) use Kelly principles to assist in asset allocation.

In Section 2, we review the necessary terminology and foundations of sports gambling. We also review the derivation of the Kelly criterion. In Section 3, we develop modified Kelly criteria by gradually increasing our assumptions. This is accomplished in a decision theoretic framework with a loss function that is natural in the Kelly context. The first approach (requiring the fewest assumptions) is based on minimax estimation. We observe that the minimax approach is too conservative and does not provide useful betting fractions. The second approach is based on Bayes estimation which requires the introduction of a prior distribution on p. The approach is flexible since it accommodates different prior beliefs. Using a Beta

***Corresponding author: Tim B. Swartz,** Professor, Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby BC V5A1S6, Canada, e-mail: tim@stat.sfu.ca
Dani Chu and Yifan Wu: Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby BC V5A1S6, Canada

prior distribution, an analytical expression for the optimal betting fraction is obtained. In Section 4, we introduce alternative loss functions and investigate the corresponding Bayes estimators. In some cases, numerical methods are required for the resultant optimization problem. In Section 5, we investigate a number of examples where various estimators of the betting fraction f are compared. In the context of the Bayes estimators, we discuss the selection of prior distributions for p with particular emphasis given to a default prior which we hope is appealing to a wide audience. In the examples which we consider, the resulting fractions tend to be less than the Kelly criterion. A short discussion is provided in Section 6.

Before proceeding, there are two papers that deserve special mention. These papers also focus on the unknown aspect of p when considering the use of the Kelly criterion. In Baker and McHale (2013), a decision theory framework is developed where fractional Kelly systems are investigated. Although the language in Baker and McHale (2013) differs from ours (e.g. utility versus loss), our concern with the minimization of the loss function [see (6) ahead] is equivalent to their optimization problem. However, loss functions are complex in the sense that they depend on parameters, data and decisions. Consequently, treatment of the underlying estimation problem can differ, and at this stage, Baker and McHale (2013) take a distinctly alternative approach. They impose an assumed betting fraction $ks^*(q)$ where $s^*(p)$ is the Kelly fraction, q is an estimator of p and $0 < k < 1$ is a shrinkage factor. We make no such assumptions on the betting fraction. Baker and McHale (2013) then focus on their "utility" function (based on the imposed assumptions) and how the function is impacted by k, particular estimators q and their corresponding sampling distributions. Utility is analyzed for specified values of p. They also study alternative utility functions. A main takeaway from their paper is that shrunken Kelly (i.e. betting fractions that are reduced proportions of the Kelly fraction) seem preferable. In our development, the consideration of Bayes estimators leads to optimal betting fractions without imposing a specified form on the betting fraction.

Baker and McHale (2016) is more of a theoretical paper that builds on their previous work and argues that "frequentist risk minimization can be a useful technique." Baker and McHale (2016) provide conditions when bet fraction shrinkage used in their earlier paper may be preferred to minimizing expected posterior loss. We note that a major difference between the two papers and ours is that we make use of prior distributions whereas their approach is frequentist. Moreover, we provide explicit expressions and R code to evaluate optimal betting fractions.

2 Review of sports gambling

There are many types of wagers that can be placed on sporting events. The most common type of wager is known as a *point spread* wager. For example, consider game three of the 2017 Western Conference finals between the Golden State Warriors and the San Antonio Spurs of the National Basketball Association (NBA). One particular *line* took the form

$$\begin{aligned} \text{Warriors} \;\; -6.5 \;\; -110 \\ \text{Spurs} \;\; +6.5 \;\; -110. \end{aligned} \tag{1}$$

The line (1) is based on *American odds* and states that a wager of \$110 placed on the Warriors returns the original \$110 plus an additional \$100 if the Warriors win by more than 6.5 points. Alternatively, a wager of \$110 placed on the Spurs returns the original \$110 plus an additional \$100 if the Spurs win or if they lose by less than 6.5 points. One sees immediately from the point spread that the Warriors are the favorite (*chalk*) whereas the Spurs are the *dog*. If the point spread is an integer, then the betting outcome can be a *push* where the gambler neither wins nor loses, and the wager is returned.

Referring to the line (1), it is also apparent that if a *bookie* collects a bet of \$110 (Warriors) and a second bet of \$110 (Spurs), then the bookie ensures a profit (*vigorish*) of \$10 no matter which team wins. Therefore, an objective of the bookie is to set a competitive point spread that encourages a balance of the bets on both sides of the point spread. To move sentiment, bookies can modify the point spread and they can also offer different odds on the two teams (e.g. −115 and −105). When positive American odds are posted such as +110, this means that a \$100 winning wager returns the original \$100 plus an additional \$110.

European (decimal) odds provide an alternative representation of American odds. For example, European odds of $\theta = 1.909$ imply that a winning bet of \$110 returns \$110(1.909) = \$210. Therefore, European odds of $\theta = 1.909$ corresponds to American odds of −110 where the return on a winning \$110 wager is \$110 + \$100 = \$210.

Consider now a gambler who places wagers with European odds θ. Suppose further that the gambler is correct (i.e. makes winning wagers) with probability p. Then the gambling *system* is profitable if the expected return is positive. Using our notation and wagers of \$$x$, the gambling system is profitable if

$$(-x)(1-p) + (x\theta - x)p > 0$$

which implies

$$p > 1/\theta. \tag{2}$$

Therefore, under the common scenario of $\theta = 1.909$, a gambling system needs only be correct $p > 1/\theta \to 52.4\%$ of the time in order to be profitable.

Whereas the profitability condition (2) may seem remarkably low and plausible, gambling has been around for a very long time. Clearly, bookies cannot systematically be exploited by profitable gamblers and continue to exist. Therefore, the development of a gambling system that ensures (2) is difficult to attain. The focus of the remainder of the paper concerns an investigation of the optimal fraction of a gambler's bankroll that should be wagered. But we stress that this is only applicable under a gambling system that satisfies the profitability condition (2). When (2) does not hold, gambling is a losing proposition.

At this stage, it is useful to review both the statement and the proof leading to the Kelly criterion.

Proposition (Kelly 1956): Consider a gambling system that picks winners with probability p and places wagers at European odds θ. Then the optimal fraction of the bankroll for wagering (explained later) is known as the Kelly criterion and is given by

$$k(p) = \begin{cases} \frac{p\theta-1}{\theta-1} & p > 1/\theta \\ 0 & p \le 1/\theta \end{cases} . \tag{3}$$

Proof. Begin with an initial bankroll B_0 where we bet a fraction f of the bankroll. Let $w = 1(0)$ according to whether we win(lose) the wager. Then the subsequent bankroll is the random quantity

$$B_1(f) = (1 - f + \theta f)^w (1 - f)^{1-w} B_0 .$$

The Kelly approach attempts to maximize the expected log growth

$$\begin{aligned} \mathrm{E}\left[\log(B_1(f)/B_0)\right] &= \mathrm{E}\left[\log(1 - f + \theta f)^w (1 - f)^{1-w}\right] \\ &= \mathrm{E}\left[w \log(1 - f + \theta f) + (1 - w)\right. \\ &\qquad \left.\log(1 - f)\right] \\ &= p \log(1 - f + \theta f) + (1 - p) \\ &\qquad \log(1 - f) \end{aligned} \tag{4}$$

Differentiating (4) with respect to f and setting equal to zero leads to the Kelly criterion $f = k(p)$ in (3). $\square$

3 Development of modified Kelly criteria

We take the view that the determination of the optimal wagering fraction f is a statistical problem where the probability p of placing a winning wager is an unknown parameter. From the framework described in Section 2, we know that the Kelly criterion $k(p)$ is the optimal value of f. Hence, the problem is one of estimating the unknown $k(p)$ by an estimator $f = f(x)$.

The data x arises in the context of historical data. In practice, sports gamblers propose potential wagering systems and use historical data to estimate p which leads to $k(\hat{p})$. For example, a sports gambler may consider placing bets on road teams in the NBA when the home team is playing its first game back from a road trip. This constitutes a wagering system (i.e. a set of rules for wagering) where $\hat{p}$ is the proportion of winning hypothetical wagers from past seasons that satisfy the conditions of the system.

What is often overlooked in this standard approach is that the true parameter p is unknown. We therefore introduce a statistical model based on a proposed wagering system:

$$\begin{aligned} X &\equiv \text{number of winning historical matches} \\ &\sim \text{Binomial}(n, p). \end{aligned} \tag{5}$$

To assess the quality of an estimator, it is necessary to introduce a loss function. We denote $l(f, p)$ as the loss incurred by estimating the true Kelly criterion $k(p)$ with the fraction $f = f(x)$. As shown in Breiman (1961), the Kelly desiderata of maximizing expected log growth has various appealing properties. Therefore, it seems reasonable to consider a loss function which is the ratio of the optimal Kelly expected log growth to the expected log growth under an alternative fraction f. This natural loss function is given by

$$\begin{aligned} l_0(f, p) &= \mathrm{E}\left[\log\left(\frac{B_1(k(p))/B_0}{B_1(f)/B_0}\right)\right] \\ &= p \log\left(\frac{1 - k(p) + \theta k(p)}{1 - f + \theta f}\right) \\ &\quad + (1 - p) \log\left(\frac{1 - k(p)}{1 - f}\right). \end{aligned} \tag{6}$$

Our problem therefore reduces to minimizing the loss $l_0(f, p)$ given in (6). However, more structure is required in the minimization problem since f is a function of x and the parameter p is unknown. We therefore consider standard decision theoretic approaches with respect to the minimization [see chapter 12 of Wasserman (2004)].

3.1 Minimax estimators

Given the above framework, decision theoretic approaches typically begin with the introduction of the *risk function* of an estimator f given by

$$R_f(p) = \sum_{x=0}^{n} l_0(f, p)\, p(x \mid p) \tag{7}$$

where $p(x \mid p)$ is the binomial probability mass function that follows from (5). What is apparent is that $R_f(p)$ is an average over the sample space, and that the preference of an estimator f_1 over f_2 involves the comparison of risk functions. To simplify comparisons, minimax estimators are those which minimize the maximum risk

$$S(f) = \sup_p R_f(p);$$

the simplification arises because the comparison of $S(f_1)$ with $S(f_2)$ now involves the comparison of scalar quantities. Unfortunately, the minimax approach is not fruitful because it is overly conservative. When we look at the worst thing that can happen for a given f [i.e. $S(f)$], this quantity is minimized by $f(x) = 0$ for all x. To see this, note that $l_0(f, p)$ in (6) is non-negative for all f and p. Therefore, to make $R_f(p)$ in (7) as small as possible [i.e. $R_f(p) = 0$], this can be achieved by not betting (i.e. $f = 0$) such that no money is ever lost. In other words, we should never wager even when the system is profitable according to (2).

To see this mathematically, imagine that we do in fact have a profitable system where $p > 1/\theta$. Under this restriction,

$$\begin{aligned} S(f) &= \sup_p R_f(p) \\ &= R_f(p = 1/\theta) \\ &= \sum_{x=0}^{n} \left[\frac{1}{\theta} \log\left(\frac{1}{1 - f + \theta f} \right) + \frac{\theta - 1}{\theta} \right. \\ &\quad \left. \log\left(\frac{1}{1 - f} \right) \right] \tbinom{n}{x} p^x (1 - p)^{n-x}. \end{aligned} \tag{8}$$

Therefore, the minimization of (8) with respect to f requires the minimization of the term within the square parentheses for each value of x. Some straightforward calculus then yields $f = 0$.

3.2 Bayes estimators

To facilitate estimation, we introduce an additional ingredient to the framework; a prior density function $\pi(p)$ which describes our uncertainty in the parameter p.

With the addition of the prior distribution, it is customary to consider the *Bayes risk*

$$r_f = \int_0^1 R_f(p)\, \pi(p)\, dp \tag{9}$$

and *Bayes estimators* f which minimize the Bayes risk (9). This formulation is helpful since we are now comparing scalar quantities. For example, the estimator $f_1(x)$ is preferred to $f_2(x)$ if $r_{f_1(x)} < r_{f_2(x)}$. We use the well-known result that a Bayes estimator minimizes expected posterior loss. In other words, we attempt to find an estimator f which minimizes

$$G(f) = \int_0^1 l_0(f, p)\, \pi(p \mid x)\, dp. \tag{10}$$

The prior distribution that immediately comes to mind in this application is $p \sim \text{Beta}(a, b)$. The Beta is defined on the intended range $p \in (0, 1)$. Note also that for $a > 1$ and $b > 1$, the prior density is concave. Concavity is appealing to the sports gambler who has an apriori belief concerning the most likely value p_0 with decreasing probability as we move away from p_0. It may also be possible for a user to specify his or her subjective beliefs (a, b) through mean and variance considerations. In the case of $p \sim \text{Beta}(a, b)$, $\text{E}(p) = a/(a + b)$ and $\text{Var}(p) = ab/((a+b)^2(a+b+1))$. Furthermore, the $\text{Beta}(a, b)$ prior together with the historical observed data x in (5) gives the convenient posterior distribution

$$p \mid x \sim \text{Beta}(x + a, n - x + b) \tag{11}$$

with probability density function

$$\pi(p \mid x) = \frac{\Gamma(n + a + b)}{\Gamma(x + a)\Gamma(n - x + b)} p^{x+a-1} (1 - p)^{n-x+b-1}. \tag{12}$$

Therefore, based on (3), (6), (10) and (12), a Bayes estimator is the fraction of the bankroll $f = f(x)$ that minimizes

$$\begin{aligned} G(f) = \int_{1/\theta}^{1} &\left[p \log\left(\frac{1 - k(p) + \theta k(p)}{1 - f + \theta f} \right) \right. \\ &\left. + (1 - p) \log\left(\frac{1 - k(p)}{1 - f} \right) \right] \pi(p \mid x)\, dp \\ &+ \int_0^{1/\theta} \left[p \log\left(\frac{1}{1 - f + \theta f} \right) + (1 - p) \log\left(\frac{1}{1 - f} \right) \right] \\ &\pi(p \mid x)\, dp. \end{aligned} \tag{13}$$

Noting that f does not appear in the numerators of the logarithms in (13), minimizing (13) is therefore equivalent to maximizing

$$\begin{aligned} Q(f) &= \int_0^1 \left[p \log(1 - f + \theta f) + (1 - p) \log(1 - f) \right] \\ &\quad \pi(p \mid x)\, dp \\ &= \hat{p} \log(1 - f + \theta f) + (1 - \hat{p}) \log(1 - f) \end{aligned} \tag{14}$$

where $\hat{p} = (x + a)/(n + a + b)$ is the posterior mean of p corresponding to (11). Finally, the maximization of $Q(f)$ with respect to f in (14) is equivalent to the original Kelly formulation [i.e. the maximization of expected log growth (4)]. Therefore, based on the loss function $l_0(f, p)$ in (6) and the Beta(a, b) prior for p, we obtain the following attractive formula for the Bayes estimator of f,

$$f_0 = \begin{cases} \frac{\hat{p}\theta-1}{\theta-1} & \hat{p} > 1/\theta \\ 0 & \hat{p} \le 1/\theta \end{cases}. \tag{15}$$

We note the similarity between the Kelly fraction (3) and the Bayes estimator (15).

4 Alternative loss functions

Whereas we believe that the proposed loss function $l_0(f, p)$ in (6) is reasonable for the given application, there are many loss functions in the literature and various criteria underlying the choice of a loss function (Jafari Jozani and Tabrizi 2013). We now investigate Bayes estimators based on three alternative loss functions where we retain the same Beta(a, b) prior for the unknown parameter p. Whereas the proposed loss functions are individually appealing, we will observe that the choice of loss function can have a considerable impact on the resultant betting fraction.

Our first alternative loss function is absolute error loss which is a common loss function and is given by

$$l_1(f, p) \ = \ | \, f - k(p) \, |$$

where $k(p)$ is the Kelly fraction (3).

With the absolute error loss function, it is well known (Berger 1985) that the posterior median of $k(p)$ is the Bayes estimator. Some distribution theory gives the following distribution function for $k(p)$:

$$\begin{aligned} F(k) &= \text{Prob}(k(p) \le k) \\ &= \begin{cases} \int_0^{1/\theta} \pi(p \mid x)\, dp & k = 0 \\ \int_0^{\frac{k(\theta-1)+1}{\theta}} \pi(p \mid x)\, dp & 0 < k \le 1 \end{cases} \end{aligned}$$

from which $F(f_1) = 1/2$ provides the the Bayes estimator based on absolute error loss

$$\begin{aligned} f_1 \ &= \begin{cases} \frac{\tilde{p}\theta-1}{\theta-1} & \int_0^{1/\theta} \pi(p \mid x)\, dp < 1/2 \\ 0 & \int_0^{1/\theta} \pi(p \mid x)\, dp \ge 1/2 \end{cases} \\ &= \begin{cases} \frac{\tilde{p}\theta-1}{\theta-1} & \tilde{p} > 1/\theta \\ 0 & \tilde{p} \le 1/\theta \end{cases} \end{aligned} \tag{16}$$

where $\tilde{p}$ is the posterior median of p corresponding to (11). We again note the similarity of (16) with the Kelly fraction (3) and the Bayes estimator based on absolute error loss (15). In applications where the posterior distribution of p is nearly symmetric, there is little difference between the two Bayes estimators f_0 and f_1. Near symmetry occurs when the posterior Beta parameters $x + a$ and $n - x + b$ are large and comparable in magnitude. Given specified values of a, b, x, n and θ, we can easily obtain f_1 in (16) numerically.

Our second alternative loss function is squared error loss which is also a common loss function and is given by

$$l_2(f, p) \ = \ (f - k(p))^2$$

where $k(p)$ is the Kelly fraction (3).

With the squared error loss function, it is well known (Berger 1985) that the posterior mean of $k(p)$ is the Bayes estimator. Therefore, the Bayes estimator based on squared error loss is

$$\begin{aligned} f_2 \ &= \ \int_0^1 k(p)\, \pi(p \mid x)\, dp \\ &= \ \int_{1/\theta}^1 \left(\frac{p\theta - 1}{\theta - 1} \right) \pi(p \mid x)\, dp. \end{aligned} \tag{17}$$

Given specified values of a, b, x, n and θ, we note that f_2 in (17) can be easily obtained numerically.

Even though squared error loss is a common loss function, we see from (3), (15), (16) and (17) that f_2 (the Bayes estimator based on squared error loss) provides a fundamentally different betting fraction than $k(p)$, f_0 and f_1. With the other three fractions, there are always scenarios in which one will not bet (i.e. the betting fraction is zero). This is never the case with f_2.

Our third alternative loss function provides a compromise between absolute error loss and squared error loss (i.e. we consider an exponent $1 < k < 2$). The loss function is also motivated by common complaints involving the Kelly criterion. As mentioned, many gamblers claim that the Kelly fraction is too large. We therefore consider a loss function that introduces a penalty on overestimation and underestimation of the true Kelly fraction via the parameters $c_1 > 0$ and $c_2 > 0$. We define this general loss function

$$l_3(f, p) \ = \ (c_1 I_{f > k(p)} + c_2) \ | \, f - k(p) \, |^k \tag{18}$$

where I is the indicator function. With appropriate selections of c_1, c_2 and k, we observe that $l_1(f, p)$ and $l_2(f, p)$ are special cases of $l_3(f, p)$. For illustration of a different sort of loss function, we consider

$$c_1 = 1, c_2 = 1 \text{ and } k = 1.5 \tag{19}$$

such that k lies halfway between absolute error loss and squared error loss, and the penalty of overestimation is double the penalty of underestimation. The estimation penalties in (19) may be considered extreme. Therefore, we also consider the settings

$$c_1 = 1, c_2 = 2 \text{ and } k = 1.5 \tag{20}$$

such that the penalty of overestimation is 1.5 times the penalty of underestimation.

In the case of the loss function $l_3(f, p)$, the Bayes estimator is obtained by minimizing the expected posterior loss

$$G(f) = \int_0^1 l_3(f, p)\, \pi(p \mid x)\, dp \tag{21}$$

with respect to $f = f(x)$.

Obtaining an analytic expression for the minimum of $G(f)$ in (21) does not seem to be within our capabilities. Fortunately, simple quadrature rules such as Simpson's rule can approximate the integral (21). With quadrature rules, one does need to be careful of the discontinuity in (18) as a function of p. Also, the minimization problem is essentially a discrete optimization problem; in practice, we only need the optimal fraction to roughly three decimal points. Therefore, a brute force procedure can be used where f is incremented from 0.000 to 1.000 in steps of size 0.001, and $G(f)$ is calculated for each incremental value. We then obtain the Bayes estimator $f_{3a} = f_{3a}(x)$ which minimizes $G(f)$ for the observed x based on the loss function settings in (19). We also obtain the Bayes estimator $f_{3b} = f_{3b}(x)$ which minimizes $G(f)$ for the observed x based on the loss function settings in (20). R code which carries out the optimization is provided in the Appendix.

5 Examples

Example 5.1. Consider the case of European odds $\theta = 1.952$ which corresponds to American odds -105. We imagine a gambler who has proposed a wagering system where $x = 100$ correct wagers are observed at European odds $\theta = 1.952$ out of $n = 180$ historical wagers. The gambler rashly determines that the winning probability is $\hat{p} = x/n = 0.556$. This appears to be a profitable system since $\hat{p} > 1/\theta = 0.512$. Using $\hat{p}$, the standard application of the Kelly approach determines the Kelly fraction $k(\hat{p}) = 0.089$. Therefore, Kelly advises that the gambler should bet 9% of the bankroll on each wager, a substantial fraction!

In determining the modified Kelly criterion, the last step is the specification of the prior parameters a and b of the Beta distribution. In investing circles (of which gambling is a special case), there exists a theory of *efficient* markets. One version of the theory essentially states that market shares are valued at their proper valuation, and consequently, there is no systematic way to exploit the market. In our scenario, this suggests that for wagers of type (1) where both teams have the same odds, the probability p of placing a correct wager should be 0.5. Furthermore, internet searches related to sports gambling reveal that even the most positive *touts* do not claim to have gambling systems that pick winners at a rate greater than 60% (Moody 2015). Therefore, together with symmetry, these two insights suggest that most of the prior probability should be assigned to the interval (0.4, 0.6). Choosing $a = b = 50$ provides us with $\mathrm{E}(p) = 0.5$ and $\mathrm{SD}(p) \approx 0.05$. Therefore, the interval (0.4, 0.6) contains roughly 95% of the prior probability. According to the proposed loss functions, when we use this default prior, we obtain the modified Kelly criteria $f_0 = 0.048$, $f_1 = 0.048$, $f_2 = 0.056$, $f_{3a} = 0.034$ and $f_{3b} = 0.041$. These conservative fractions are smaller than the Kelly criterion $k(\hat{p}) = 0.089$. In particular, the estimate f_{3a} based on the extreme loss function is very small, and this highlights the importance and the effect of the choice of loss function. In this example, and in most of the examples that follow, we observe that the relationship amongst the Bayes estimators is $f_{3a} < f_{3b} < f_0 \approx f_1 < f_2$. In particular, f_0 (our preferred estimator) nearly corresponds to half-Kelly. Hence, we have provided a theoretical rationale for the use of half-Kelly.

Another idea in prior specification is to impose a tighter constraint $p \geq 0.5$ and then proceed with the same type of argumentation as above. Imposing $p \geq 0.5$ may be sensible from the point of view that coin flips (no knowledge) provided $p = 0.5$, and it may be argued that it is impossible to know less than a coin.

Example 5.2. In this example, we modify Example 5.1. We now set the European odds to $\theta = 1.909$ which corresponds to American odds -110; odds of -110 are more standard in the *sportsbook* industry and provide a greater vigorish. Consequently, the gambler will need to choose winners at a higher rate to ensure profitability. As before, we choose $a = b = 50$ and consider historical data with $x = 100$ and $n = 180$. We then obtain the Kelly criterion $k(\hat{p}) = 0.067$ and the modified Kelly criteria $f_0 = 0.025$,

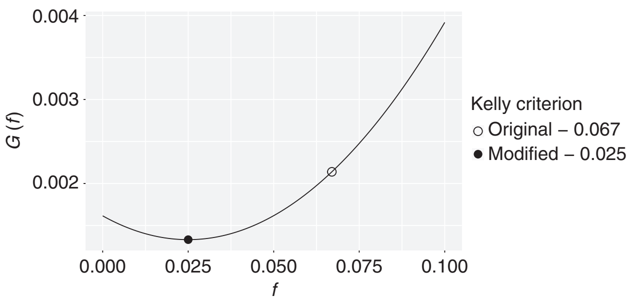

Figure 1: Plot of expected posterior loss $G(f)$ in (13) versus f for the conditions specified in Example 5.2 where original Kelly is compared versus modified Kelly f_0.

$f_1 = 0.025$, $f_2 = 0.039$, $f_{3a} = 0.018$ and $f_{3b} = 0.024$. These are all smaller fractions than the corresponding fractions in Example 5.1. Thus the increased vigorish causes a reduction in the fractions wagered. In the context of the loss function $l_0(f, p)$ in (6), Figure 1 provides a plot of $G(f)$ in (13) versus f in this example. We observe that that the function is convex where the expected posterior loss increases more rapidly for $f > f_0$ than for $f < f_0$.

Example 5.3. We make an adjustment to Example 5.2. As before, $\theta = 1.909$ and $a = b = 50$. However, this time, we have more historical data $x = 200$ and $n = 360$ but we retain the same proportion of successful wagers. The Kelly criterion remains at $k(\hat{p}) = 0.067$ but the modified Kelly criteria increase from $f_0 = 0.025$, $f_1 = 0.025$, $f_2 = 0.039$, $f_{3a} = 0.018$ and $f_{3b} = 0.024$ to $f_0 = 0.041$, $f_1 = 0.041$, $f_2 = 0.047$, $f_{3a} = 0.030$ and $f_{3b} = 0.036$, respectively. The increased fractions reflect our increased posterior support for larger values of p via the data.

As we collect more and more data (i.e. as $n \to \infty$), we gain confidence in the performance of a gambling system. Consequently, the posterior becomes concentrated at $\hat{p}$, and our modified Kelly criterion approaches the original Kelly criterion.

Example 5.4. We make another adjustment to Example 5.2 which concerns an examination of prior sensitivity. As before, $\theta = 1.909$, $x = 100$ and $n = 180$. However, this time, we are apriori more doubtful about our ability to have a winning system. We set $a = b = 100$ which leads to $\mathrm{SD}(p) \approx 0.035$. This implies that the interval $(0.43, 0.57)$ contains roughly 95% of the prior probability. In this case, the Kelly criterion remains at $k(\hat{p}) = 0.067$ but the modified Kelly criteria decrease from $f_0 = 0.025$, $f_1 = 0.025$, $f_2 = 0.039$, $f_{3a} = 0.018$ and $f_{3b} = 0.024$ to $f_0 = 0.005$, $f_1 = 0.005$, $f_2 = 0.024$, $f_{3a} = 0.008$ and $f_{3b} = 0.012$, respectively. The smaller fractions reflect our decreased prior belief in having a profitable system.

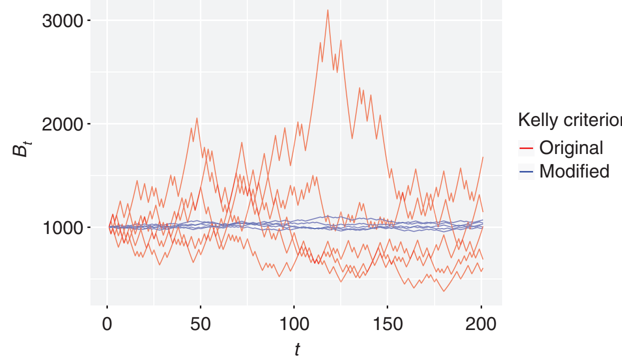

Figure 2: Plot of the bankroll B_t after t wagers using the original Kelly fraction and the modified Kelly fraction f_0 as described in Example 5.4 (five iterations of each scheme).

To get a sense of the performance of original Kelly versus modified Kelly, we carried out simulations based on hypothetical wagering. Following the settings in Example 5.4, the original Kelly fraction is $f_{orig} = 0.067$ and the modified Kelly fraction is $f_0 = 0.005$. Inverting (3), gives the corresponding probabilities of correct wagering, $p_{orig} = 0.556$ and $p_0 = 0.526$, respectively. For the simulation, we set the true $p = (p_{orig} + p_o)/2 = 0.541$. Under these conditions, and beginning with an initial bankroll of $B_0 = \$1000$, we generated 200 consecutive wagers following the two money management schemes (i.e. original Kelly and modified Kelly f_0). Five iterations of each scheme are presented in Figure 2 where we monitor the bankroll B_t after t wagers. We observe that original Kelly has more variability, with greater potential profits and greater potential losses. In the five simulation runs, three out of the five modified Kelly runs do better by the end of the wagering season than the corresponding original Kelly runs.

We then repeated the simulation procedure using 1000 runs. From these simulations, we provide density plots of the final bankroll B_{200} in Figure 3. We observe that modified Kelly f_0 is safer with 65% of the runs resulting in profits. In contrast, original Kelly is profitable only 53% of the time.

Finally, we considered simulation settings where the number of bets in a season was varied. We considered 1000 runs for independent seasons of length 100 to 5000 in increments of 100. Figure 4 shows the percentage of runs where the schemes were profitable, doubled and halved. Original Kelly is again compared against f_0. When the season length is large, the asymptotic properties begin to take over, and on average, modified Kelly f_0 outperforms original Kelly.

Example 5.5. This is a real data example based on an article (Professor MJ 2017) from the gambling website covers.com and his website. Professor MJ recommends a

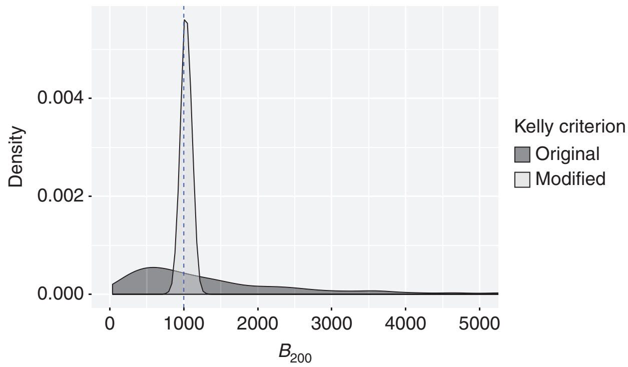

Figure 3: Plot of the estimated density function of the final bankroll B_{200} using the original Kelly fraction and the modified Kelly fraction f_0 as described in Example 5.4.

gambling system for the NBA playoffs that is applicable when a playoff team loses a game and the difference between the actual margin of victory and the point spread exceeds 12.5 points. In other words, the team has lost badly. Professor MJ's system is to bet on that team (using point spread) in the next game. This push-back phenomenom was also explored in the context of basketball and hockey by Swartz et al. (2011). Professor MJ's recommendation was based on studying all NBA playoff games from the 1991 playoff season up to game 2 of the 2017 Western Conference semi-finals. Under these conditions, there were $n = 484$ historical matches and $x = 271$ winning bets according to the binomial distribution in (5). The historical data provide an impressive winning proportion $\hat{p} = 0.560$. Professor MJ goes on to recommend betting on the San Antonio Spurs in game 2 of their Western Conference semi-final match against the Houston Rockets. The Rockets had crushed the Spurs 126-99 in game 1 where the Spurs were 6-point favorites. Game 2 satisfies the betting conditions as the difference between the actual result and the point spread was $(126 - 99) - (-6) = 33 > 12.5$ points.

Using European odds of $\theta = 1.909$ (the industry standard), the Kelly system advises a betting fraction of $k(\hat{p}) = 0.076$. However, if we use the same rationale for prior selection as in Example 5.1 ($a = b = 50$) and provide historical data inputs $x = 271$ and $n = 484$, we obtain modified Kelly fractions $f_0 = 0.054$, $f_1 = 0.054$, $f_2 = 0.056$, $f_{3a} = 0.042$ and $f_{3b} = 0.047$. We note that these are smaller fractions than original Kelly. In particular, f_0 is roughly 3/4 Kelly.

We now investigate the performance of Professor MJ's system on the remainder of the 2017 playoff matches. Admittedly, it is a small sample. Of the 32 subsequent playoff games beyond May 1/2017, there there were 14 playoff games that satisfied Professor MJ's betting criterion. In those games, wagering as stipulated by the system resulted in 4 wins, 8 losses and 2 pushes. Beginning with a \$1000 bankroll, the Kelly criterion would have produced a balance of \$694.10 whereas modified Kelly based on f_0 would have produced a balance of \$776.91. Therefore, the cautious modified Kelly approach was helpful in this example. The sequence of wins/losses/pushes was WLLWPLLWLLWLLP. Incidentally, in game 2 of the Western Conference semi-final, the Spurs were favorites by 5.5 points and they won the match 121-96, beating the spread.

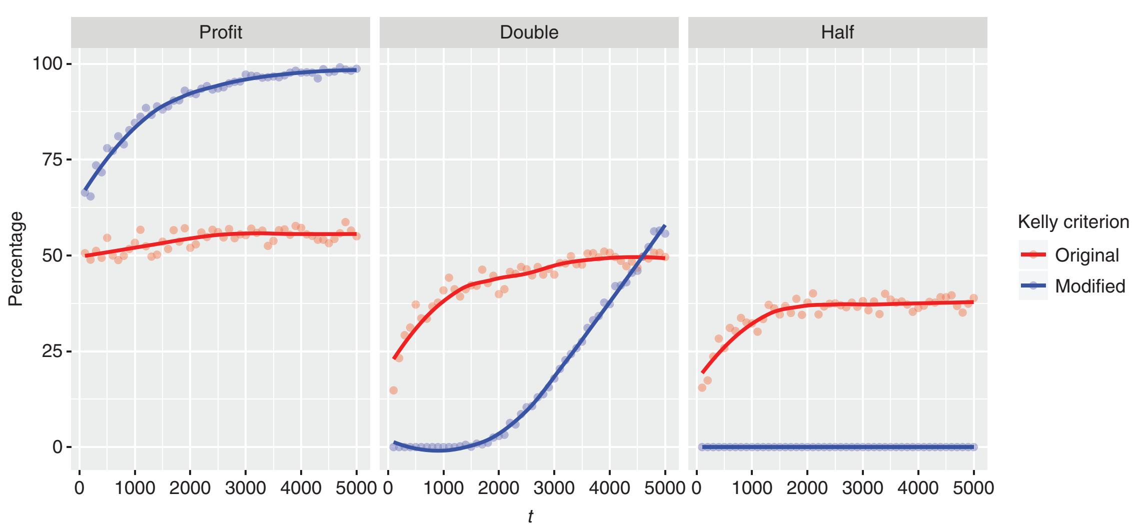

Figure 4: Corresponding to Example 5.4, percentages corresponding to the original Kelly and the modified Kelly schemes f_0 under betting seasons of length 100, . . . , 5000. We plot the percentages of profitable final bankrolls (i.e. exceeding \$1000), the percentages of final bankrolls that have at least doubled (i.e. exceeding \$2000) and the percentages of final bankrolls that have decreased by at least half (i.e. less than \$500). Points are smoothed using lowess curves.

6 Discussion

In terms of money management, the Kelly criterion has received considerable attention in sports gambling. However, for many gamblers the Kelly criterion is thought to be an excessive fraction which has led to recommendations such as half-Kelly. In this paper, modified Kelly criteria are suggested which require the specification of a Beta prior distribution on the gambler's probability p of selecting winning wagers. When the priors are conservative (i.e. place less probability on $p > 1/\theta$), then the modified Kelly criteria tend to be smaller than the original Kelly criterion. Moreover, in the examples we have considered, we have attempted to elicit reasonable prior opinion on p. And in these cases, the modified Kelly criteria are smaller than the original Kelly criterion.

We have seen that the choice of loss function also plays a role in determining an alternative to Kelly. Although the loss function $l_0(f, p)$ in (6) appears natural to us, individual gamblers may prefer to increase/decrease their betting fractions by considering the alternative loss functions presented in this paper.

Now there may be gamblers who do not follow a system such as described in Example 5.5. For example, they may have a sequence of potential bets where p is variable. In other words, they may have more confidence in some wagers than in others. It is also possible that these wagers correspond to variable odds θ. Note that the use of original Kelly is challenged in this setup because the estimation of p would be based on some smaller subset of games or perhaps on some regression analyses. How would modified Kelly work here? Well, the stumbling block would be the specification of x, n, a and b. Perhaps $x = n = 0$ is convenient; this would essentially lead to a criterion based on minimizing expected prior loss. Specifying a and b may be possible by using some of the same constraint argumentation used in Example 5.1.

In future work, it would be desirable to obtain modified Kelly fractions in the context of simultaneous wagers.

Acknowledgment: Chu and Swartz have been partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Undergraduate Student Research Awards (USRA) program and the Discovery Grants program, respectively. Wu has been partially supported by the Big Data Key Initiative at Simon Fraser University. The authors are most appreciative of the contribution of two anonymous reviewers who provided insightful comments. In particular, one of the reviewers recognized a simplification in equation (13) which resulted in the analytic expression (15). We also thank Mohammad Jafari Jozani who provided helpful discussions relating to loss functions.

Appendix

Below is an R function that takes as input the historical betting record (x, n), the European odds θ corresponding to the wager of interest, the Beta parameters (a, b) that describe the prior probability of a correct wager and the loss function parameters c_1, c_2 and k based on the general loss function $l_3(f, p)$ in (18). The output is the betting fraction f_3 which is a Bayes estimator.

```
library(cubature)

G <- function(f, x, n, theta, a, b, c1, c2, k){

 l_3_1 <- function(p){
   kelly <- ifelse(p > (1/theta), (p * theta - 1) / (theta - 1), 0)

   loss_func = ifelse(p < (f * (theta - 1) + 1)/theta,
                     yes = (c1 + c2) * (abs(f-kelly)^k),
                     no = 0)

   return(loss_func*dbeta(p, a + x, n - x + b))
 }

 l_3_2 <- function(p){

   kelly <- ifelse(p > (1/theta), (p * theta - 1) / (theta - 1), 0)
```

```
    loss_func = ifelse(p >= (f * (theta - 1) + 1)/theta,
                        yes = c2 * (abs(f-kelly)^k),
                        no = 0)

    return(loss_func*dbeta(p, a + x, n - x + b))
 }

# Integrate the loss function
part_1 <- adaptIntegrate(l_3_1, lowerLimit = 0, upperLimit = (f * (theta-1) + 1)/theta,
          tol = 1e-10)
part_2 <- adaptIntegrate(l_3_2, lowerLimit = (f * (theta-1) + 1)/theta , upperLimit = 1,
          tol = 1e-10)

 # Return the sum of the two integrals
 return(part_1$integral + part_2$integral )
}

f3 <- function(x, n, theta, a, b, c1, c2, k)
{
 i <- 0.001
 # f is a vector from 0 to 1
 f <- seq(0, 1, i)

 # G_ of_f contains the results of the function of G evalutated at all the f values
 G_of_f <- lapply(f, G, x = x, n = n, theta = theta, a = a, b = b, c1 = c1, c2 = c2, k = k)

 # Find the index of the minimum value and the corresponding f
 Modified_Kelly_Value <- which.min(G_of_f) * i - i

 return(Modified_Kelly_Value)
}
```

References

Baker, R. D. and I. G. McHale. 2013. “Optimal Betting Under Parameter Uncertainty: Improving the Kelly Criterion.” *Decision Analysis* 10:189–199.

Baker, R. D. and I.G. McHale. 2016. “Making Better Decisions: Can Minimizing Frequentist Risk Help?” *International Journal of Statistics and Probability* 5:80–90.

Berger, J. O. 1985. *Statistical Decision Theory and Bayesian Analysis, Second Edition*. New York: Springer-Verlag.

Breiman, L. 1961. “Optimal Gambling Systems for Favorable Games.” Volume 1: Contributions to the Theory of Statistics. in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman, 65–78. Berkeley, California: University of California Press.

Feng, G., N. Polson, and J. Xu. 2016. “The Market for English Premier League (EPL) Odds.” *Journal of Quantitative Analysis in Sports* 12:167–178.

Insley, R., L. Mok, and T. B. Swartz. 2004. “Issues Related to Sports Gambling.” *The Australian and New Zealand Journal of Statistics* 46:219–232.

Jafari Jozani, M. and N. J. Tabrizi. 2013. “Intrinsic Posterior Regret Gamma-Minimax Estimation for the Exponential Family of Distributions.” *Electronic Journal of Statistics* 7:1856–1874.

Kelly, J. L. 1956. “A New Interpretation of Information Rate.” *Bell System Technical Journal* 35:917–926.

Lopez, M. J. and G. J. Matthews. 2015. “Building an NCAA Men’s Basketball Predictive Model and Quantifying Its Success.” *Journal of Quantitative Analysis in Sports* 11:5–12.

MacLean, L. C., E. O. Thorp, and W. T. Ziemba, eds. 2011. *The Kelly Capital Growth Investment Criterion: Theory and Practice*, World Scientific Handbook in Financial Economics Series: Volume 3. Singapore: World Scientific Publishing.

Moody, A. 2015. The magic of 55% winners. *ThoughtCo.*, Accessed online June 2/2017 at https://www.thoughtco.com/the-magic-of-fiftyfive-percent-winners-3116838.

Murphy, A. 2015. “How to Use the Kelly Criterion in Online Sports Betting.” *MyBookie.ag*, Accessed online May 17/17 at https://mybookie.ag/sports-betting-guide/how-to-use-kelly-criterion/.

Professor MJ. 2017. “NBA bounce-back ATS betting trend points to stronger effort from the spurs in game 2.” Accessed online September 6/2017 at http://www.covers.com/Editorial/Article/ef2b2930-2f3a-11e7-bcfd-005056830dca/NBA-bounce-back-ATS-betting-trend-points-to-stronger-effort-from-the-Spurs-in-Game-2 and also at https://www.professormj.com/pages/nba-playoffs-the-bounce-back-effect.

Samuelson, P. A. 1979. “Why We Should Not Make Mean Log of Wealth Big Though Years to Act are Long.” *Journal of Banking and Finance* 3:305–307.

Schumaker, R. 2013. “Data Mining the Harness Track and Predicting Outcomes.” *Journal of International Technology and Information Management* 22:103–107.

Swartz, T. B., A. Tennakoon, F. Nathoo, M. Tsao, and P. S. Sarohia. 2011. “Ups and Downs: Team Performances in Best-of-Seven Playoff Series.” *Journal of Quantitative Analysis in Sports* 7(4):Article 2.

Thorp, E. O. 1969. “Optimal Gambling Systems for Favorable Games.” *Review of the International Statistical Institute* 37:273–293.

Wasserman, L. 2004. *All of Statistics: A Concise Course in Statistical Inference*. New York: Springer.