J. Appl. Prob. **41,** 1230–1236 (2004)
Printed in Israel

THE KELLY SYSTEM MAXIMIZES MEDIAN FORTUNE

S. N. ETHIER,* *University of Utah*

Abstract

It is well known that the Kelly system of proportional betting, which maximizes the long-term geometric rate of growth of the gambler's fortune, minimizes the expected time required to reach a specified goal. Less well known is the fact that it maximizes the median of the gambler's fortune. This was pointed out by the author in a 1988 paper, but only under asymptotic assumptions that might cause one to question its applicability. Here we show that the result is true more generally, and argue that this is a desirable property of the Kelly system.

Keywords: Gambling; optimal proportional betting

2000 Mathematics Subject Classification: Primary 60G50
Secondary 60G40

1. Introduction

Consider a game of chance that is played repeatedly and is advantageous to the gambler. A proportional betting system is one in which the bettor wagers a fixed proportion $f \in [0, 1]$ of his current fortune at each trial. Let $X_1, X_2, \dots$ be independent and identically distributed (i.i.d.) $[-1, \infty)$-valued random variables with $0 < \mathrm{E}[X_1] < \infty$, X_l representing the proportional bettor's net gain (positive or negative) per unit bet at trial l. Then his fortune $F_n(f)$ after n trials is given by

$$F_n(f) = \prod_{l=1}^{n} (1 + fX_l), \tag{1}$$

assuming (without loss of generality) an initial fortune $F_0(f) := 1$. The strong law of large numbers implies that

$$\lim_{n\to\infty} n^{-1} \ln F_n(f) = \mathrm{E}[\ln(1 + fX_1)] \tag{2}$$

almost surely. The choice f^* of f that maximizes the right side of (2), which might be called the long-term geometric rate of growth of the proportional bettor's fortune, results in a betting system known as the Kelly (1956) system.

A well-known optimality property of the Kelly system, due to Breiman (1961), is that it minimizes the expected time required to reach a specified goal. Less well known is the fact that it maximizes the median of the proportional bettor's fortune. Actually, both of these results are true only in an asymptotic sense. The median result was pointed out by Ethier (1988), but the underlying asymptotic assumptions might cause one to question its applicability. Here we show that the conclusion is valid more generally than was previously realized. Maslov and Zhang (1998) argued that the Kelly system maximizes median fortune, but based their argument on

Received 15 March 2004; revision received 2 June 2004.
* Postal address: Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA.
Email address: ethier@math.utah.edu

the premise – typically false – that the mean and the median of a sum of i.i.d. random variables are equal.

We begin by reviewing the earlier median result.

2. Geometric Brownian motion approximation

We replace the i.i.d. sequence $X_1, X_2, \ldots$ by a one-parameter family of i.i.d. sequences $X_1(\varepsilon), X_2(\varepsilon), \ldots$ of nondegenerate $[-1, \infty)$-valued random variables, parametrized by the mean, that is, $\mathrm{E}[X_1(\varepsilon)] = \varepsilon$, where $0 \le \varepsilon < \varepsilon_0$. We assume that $X_1(\varepsilon) \overset{\mathrm{D}}{\to} X_1(0)$ in $\mathbb{R}$ as $\varepsilon \to 0$, where '$\overset{\mathrm{D}}{\to}$' denotes convergence in distribution, and we require two technical assumptions, namely that $\mathrm{E}[X_1(\varepsilon)/(1 + X_1(\varepsilon))] < 0$ (here and elsewhere, $-1/0 := -\infty$) for $0 < \varepsilon < \varepsilon_0$, and that $\{X_1^2(\varepsilon) \colon 0 < \varepsilon < \varepsilon_0\}$ is uniformly integrable. Let $\sigma^2(\varepsilon) := \mathrm{var}[X_1(\varepsilon)]$. Defining $f^*(\varepsilon)$ to be the unique $f \in (0, 1)$ that maximizes $\mathrm{E}[\ln(1 + f X_1(\varepsilon))]$, we observe that, given $\alpha > 0$ such that $\alpha f^*(\varepsilon) < 1$ for $0 < \varepsilon < \varepsilon_0$,

$$Y_\varepsilon(t) := \prod_{l=1}^{\lfloor \sigma^2(\varepsilon)t/\varepsilon^2 \rfloor} (1 + \alpha f^*(\varepsilon) X_l(\varepsilon))$$

represents the proportional bettor's fortune after $\lfloor \sigma^2(\varepsilon)t/\varepsilon^2 \rfloor$ trials (where $\lfloor \cdot \rfloor$ denotes the integer part of its argument), assuming a betting proportion of $\alpha f^*(\varepsilon)$, that is, α times the Kelly betting proportion.

Ethier (1988) showed that $Y_\varepsilon \overset{\mathrm{D}}{\to} Y$ in $D_{(0,\infty)}[0, \infty)$ as $\varepsilon \to 0$, where

$$Y(t) := \exp\{\alpha(1 - \alpha/2)t + \alpha W(t)\}, \qquad t \ge 0, \tag{3}$$

W being a standard Brownian motion and $D_{(0,\infty)}[0, \infty)$ the space of right-continuous sample paths with left limits and with values in $(0, \infty)$. A simple corollary of this is that, for fixed $t > 0$,

$$\lim_{\varepsilon \to 0} \mathrm{median}[Y_\varepsilon(t)] = \mathrm{median}[Y(t)] = \exp\{\alpha(1 - \alpha/2)t\}, \tag{4}$$

and this is maximized by $\alpha = 1$. In fact, the convergence in (4) holds uniformly for $\alpha \in [0, 2]$ and so, when $\varepsilon = \mathrm{E}[X_1(\varepsilon)]$ is small, $\mathrm{median}[Y_\varepsilon(t)]$ is maximized on $[0, 2]$ by α close to 1.

One might object that $\mathrm{E}[X_1(\varepsilon)]$ is fixed in practice and, while typically small, does not tend to zero in any sense. But then the same objection would apply to the widely accepted approximation based on the result that

$$f^*(\varepsilon) = \frac{\varepsilon}{\sigma^2(\varepsilon)}(1 + o(1)) \quad \text{as } \varepsilon \to 0.$$

Nevertheless, it is possible to show that median fortune is maximized when the wagers are fixed, and we take this up in what follows.

3. The case of two possible outcomes at each trial

We return to the situation in which there is a single i.i.d. sequence of $[-1, \infty)$-valued random variables $X_1, X_2, \ldots$ with $0 < \mathrm{E}[X_1] < \infty$. In this section we specialize to the case in which X_1 assumes only two values, namely $a > 0$ with probability p and -1 with probability $q := 1 - p$. In other words, the game pays odds of a to 1 on the occurrence of an event of probability p. The assumption of a positive mean tells us that $(a + 1)p - 1 > 0$.

Given $n \geq 1$, let $B_n := |\{1 \leq l \leq n\colon X_l = a\}|$ be the number of wins in the first n trials. Then B_n is binomial(n, p) and

$$F_n(f) = (1+af)^{B_n}(1-f)^{n-B_n} = \left(\frac{1+af}{1-f}\right)^{B_n}(1-f)^n.$$

Now the log of the median is the median of the log, so

$$\ln(\text{median}[F_n(f)]) = \text{median}[B_n]\ln\left(\frac{1+af}{1-f}\right) + n\ln(1-f).$$

Next we cite a theorem of Edelman (1979) and Hamza (1995), which states that the mean and median of a binomial differ by less than $\ln 2 = 0.693\cdots$. Thus, $\gamma_{n,p} := \text{median}[B_n] - np$ satisfies $|\gamma_{n,p}| < \ln 2$. Noting that

$$\mathrm{E}[\ln(F_n(f))] = np\ln(1+af) + nq\ln(1-f), \tag{5}$$

we find that

$$\begin{aligned}\ln(\text{median}[F_n(f)]) &= (np+\gamma_{n,p})\ln\left(\frac{1+af}{1-f}\right) + n\ln(1-f)\\ &= (np+\gamma_{n,p})\ln(1+af) + (nq-\gamma_{n,p})\ln(1-f)\\ &= \mathrm{E}[\ln(F_n^*(f))],\end{aligned}$$

where $F_n^*(f)$ is $F_n(f)$ with p replaced by $p_n := p + \gamma_{n,p}/n = \text{median}[B_n]/n$. Since (5) is uniquely maximized by the Kelly proportion

$$f^* = \frac{(a+1)p-1}{a} = \frac{\mathrm{E}[X_1]}{a},$$

it follows that median$[F_n(f)]$ is uniquely maximized at

$$\tilde{f}_n = \frac{(a+1)p_n - 1}{a} = f^* + \frac{a+1}{a}\frac{\gamma_{n,p}}{n}. \tag{6}$$

(Actually, median[binomial(n, p)], as a function of p, is a nondecreasing, integer-valued step function, the value of which is ambiguous at each of its n jumps. We adopt the convention that this function be right continuous, thereby eliminating the ambiguity.)

Thus, the median-maximizing proportion $\tilde{f}_n$ is $f^* + O(n^{-1})$. But notice that there are cases in which this approximation is exact: if the binomial mean np is an integer, then, since the binomial median is always an integer and they differ by less than 1, we have $\gamma_{n,p} = 0$ and therefore $\tilde{f}_n = f^*$.

For example, if $a = 1$ and $n = 100$, then for $p = 0.51$ the median-maximizing proportion $\tilde{f}_n$ is precisely the Kelly proportion $f^* = 0.02$, while for $p = 0.505$ we find that $\tilde{f}_n = 0.02$ and $f^* = 0.01$. This is a natural consequence of approximating a continuous function by a step function.

The latter example suggests that a more useful way to state (6) might be

$$\frac{\tilde{f}_n}{f^*} = 1 + \frac{a+1}{\mathrm{E}[X_1]}\frac{\gamma_{n,p}}{n},$$

which shows that, the smaller $\mathrm{E}[X_1]$ is, the larger n must be to ensure a relative error within specified bounds.

4. The case of more than two possible outcomes at each trial

As in Section 1, let $X_1, X_2, \dots$ be i.i.d. $[-1, \infty)$-valued random variables with $0 < \mathrm{E}[X_1] < \infty$, and define $F_n(f)$ for each $f \in [0, 1]$ and $n \geq 1$ by (1). Further, for $0 \leq f < 1$ let

$$\mu(f) := \mathrm{E}[\ln(1 + fX_1)], \qquad \sigma^2(f) := \mathrm{var}[\ln(1 + fX_1)].$$

Let $\{f_n\} \subset [0, 1)$ be a (nonrandom) sequence that converges to $f \in [0, 1)$. Then, by the Lindeberg–Feller theorem,

$$\frac{\ln(F_n(f_n)) - n\mu(f_n)}{\sqrt{n}} = \frac{1}{\sqrt{n}}\left(\sum_{l=1}^{n} \ln(1 + f_n X_l) - n\mu(f_n)\right) \xrightarrow{\mathrm{D}} \mathcal{N}(0, \sigma^2(f)),$$

where $\mathcal{N}(0, \sigma^2(f))$ is the normal distribution with mean zero and variance $\sigma^2(f)$. Here we are using the fact that $\sup_{n \geq 1} \mathrm{E}[|\ln(1 + f_n X_1) - \mu(f_n)|^3] < \infty$. It follows that

$$\begin{aligned}\frac{\ln(\mathrm{median}[F_n(f_n)]) - n\mu(f_n)}{\sqrt{n}} &= \mathrm{median}\left[\frac{\ln(F_n(f_n)) - n\mu(f_n)}{\sqrt{n}}\right] \\ &\to \mathrm{median}[\mathcal{N}(0, \sigma^2(f))] = 0.\end{aligned}$$

Since $\{f_n\}$ was arbitrary, we conclude that

$$\frac{\ln(\mathrm{median}[F_n(f)]) - n\mu(f)}{\sqrt{n}} \to 0,$$

uniformly in f in compact subsets of $[0, 1)$, or, equivalently, that

$$\mathrm{median}[F_n(f)] = \mathrm{e}^{n\mu(f) + o(\sqrt{n})}, \tag{7}$$

uniformly in f in compact subsets of $[0, 1)$.

Let us further assume that $\mu'(1-) = \mathrm{E}[X_1/(1 + X_1)] < 0$. This guarantees that there is a unique $f \in (0, 1)$ that maximizes $\mu(f)$; as before, we denote this Kelly proportion by f^*. Suppose that, for each $n \geq 1$, $\tilde{f}_n \in [0, 1)$ maximizes $\mathrm{median}[F_n(f)]$ as a function of f. Then

$$1 \geq \frac{\mathrm{median}[F_n(f^*)]}{\mathrm{median}[F_n(\tilde{f}_n)]} = \exp\{n[\mu(f^*) - \mu(\tilde{f}_n)] + o(\sqrt{n})\}.$$

Since $\mu(f^*) - \mu(\tilde{f}_n) \geq 0$ and $\mu''(f^*) < 0$, we have

$$o(n^{-1/2}) = \mu(f^*) - \mu(\tilde{f}_n) = -\tfrac{1}{2}(\tilde{f}_n - f^*)^2[\mu''(f^*) + o(1)]$$

and therefore

$$\tilde{f}_n = f^* + o(n^{-1/4}). \tag{8}$$

In view of (6), this may not be the best possible rate of convergence. To improve it we would need to improve the error term in (7).

Nevertheless, (7) is sufficient to tell us that, if $f_0 \in [0, 1)$ differs from f^* (with f^* as in the preceding paragraph), then

$$\frac{\mathrm{median}[F_n(f^*)]}{\mathrm{median}[F_n(f_0)]} = \exp\{n[\mu(f^*) - \mu(f_0)] + o(\sqrt{n})\} \to \infty$$

exponentially fast. This is one sense in which the Kelly system (asymptotically) maximizes median fortune.

We now turn to the question of whether the error term in (7) can be improved. A random variable Y_1 is said to be *lattice* if there exist $b \in \mathbb{R}$ and $h > 0$ such that $\mathrm{P}\{Y_1 \in b + h\mathbb{Z}\} = 1$, where $\mathbb{Z}$ is the set of integers; otherwise Y_1 is said to be *nonlattice*. If in the lattice case h is chosen maximally, it is called the *maximum span* of Y_1.

Typically, in gambling situations, X_1 is lattice, hence $1 + fX_1$ is lattice, and therefore, except in the case covered by Section 3, $\ln(1 + fX_1)$ is nonlattice for most f. For example, if X_1 has support $\{-1, 0, 1\}$, then $\ln(1 + fX_1)$ is lattice if and only if $\ln(1 + f)/\ln(1 - f)$ is rational, which is the case for at most countably many $f \in (0, 1)$. Furthermore, if it is lattice with $\ln(1 + f)/\ln(1 - f) = -r/s$, where r and s are positive, relatively prime integers, then its maximum span is given by $h = \ln(1 + f)/r < \ln 2$.

Hall (1980) proved that if $Y_1, Y_2, \ldots$ are i.i.d. nonlattice random variables with finite third absolute moment, mean μ, variance $\sigma^2 > 0$, and third central moment τ, then

$$\mathrm{median}[Y_1 + \cdots + Y_n] = n\mu - \frac{\tau}{6\sigma^2} + o(1).$$

Using an argument analogous to that of Hall, it is possible to obtain a related, albeit less precise, result in the lattice case. Let $Y_1, Y_2, \ldots$ be i.i.d. lattice random variables with maximum span h. Assume that Y_1 has finite third absolute moment, mean μ, variance $\sigma^2 > 0$, and third central moment τ. Then, by a theorem of Esseen and a local central limit theorem (see Section 43, Theorem 1, and Section 51, Theorem 2, respectively, of Gnedenko and Kolmogorov (1968)) we can write

$$\mathrm{median}[Y_1 + \cdots + Y_n] = n\mu - \frac{\tau}{6\sigma^2} + \theta_n h + o(1),$$

where $-\frac{1}{2} \le \theta_n \le \frac{3}{2}$.

Consequently, regardless of whether $\ln(1 + fX_1)$ is lattice or nonlattice,

$$\mathrm{median}[F_n(f)] = \mathrm{e}^{n\mu(f)+O(1)}. \tag{9}$$

The problem with (9) is that we have proved it only for fixed $f \in [0, 1)$. If we could prove that (9) holds uniformly in f in compact subsets of $[0, 1)$, thereby improving (7), we would then have $\tilde{f}_n = f^* + O(n^{-1/2})$, thereby improving (8).

5. Simultaneous wagers

Here we generalize the results of the previous section to the case in which there are several wagers to choose from at each trial, at least one of which has positive expectation. A typical example would be a sufficiently biased roulette wheel.

Given $d \ge 1$, let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ be i.i.d. $[-1, \infty)^d$-valued random variables with $\mathrm{E}[X_{1,i}] < \infty$ for $i = 1, \ldots, d$ and $\mathrm{E}[X_{1,i}] > 0$ for some $i \in \{1, \ldots, d\}$, where $\boldsymbol{X}_1 := (X_{1,1}, \ldots, X_{1,d})$. Let

$$\Delta := \{\boldsymbol{f} = (f_1, \ldots, f_d) \in [0, 1]^d : f_1 + \cdots + f_d \le 1\},$$

and define $F_n(\boldsymbol{f})$ for each $\boldsymbol{f} \in \Delta$ and $n \ge 1$ by

$$F_n(\boldsymbol{f}) := \prod_{l=1}^{n} (1 + \boldsymbol{f} \cdot \boldsymbol{X}_l).$$

We define $\Delta^\circ := \{\boldsymbol{f} \in \Delta \colon f_1 + \cdots + f_d < 1\}$, and for $\boldsymbol{f} \in \Delta^\circ$ we let

$$\mu(\boldsymbol{f}) := \mathrm{E}[\ln(1 + \boldsymbol{f} \cdot \boldsymbol{X}_1)].$$

The argument in Section 4 shows that

$$\operatorname{median}[F_n(\boldsymbol{f})] = \mathrm{e}^{n\mu(\boldsymbol{f})+o(\sqrt{n})}, \tag{10}$$

uniformly in $\boldsymbol{f}$ in compact subsets of Δ°.

Now let us further assume that the components of $\boldsymbol{X}_1$ have finite variance and are linearly independent (but not necessarily stochastically independent), that is, there does not exist $\boldsymbol{h} \in \mathbb{R}^d - \{\boldsymbol{0}\}$ (nonrandom) such that $\mathrm{P}\{\boldsymbol{h} \cdot \boldsymbol{X}_1 = 0\} = 1$. Then, for each $\boldsymbol{f} \in \Delta^\circ$, the matrix of second-order partial derivatives of $\mu(\boldsymbol{f})$, which we denote by $\nabla^2\mu(\boldsymbol{f})$, is negative definite. We can therefore adapt the argument of Section 4 to obtain that, if $\mu(\boldsymbol{f})$ has a unique maximum at $\boldsymbol{f}^* \in \Delta^\circ$, and if, for each $n \geq 1$, $\tilde{\boldsymbol{f}}_n \in \Delta^\circ$ maximizes $\operatorname{median}[F_n(\boldsymbol{f})]$ as a function of $\boldsymbol{f}$, then

$$\begin{aligned} o(n^{-1/2}) &= \mu(\boldsymbol{f}^*) - \mu(\tilde{\boldsymbol{f}}_n) \\ &= -(\tilde{\boldsymbol{f}}_n - \boldsymbol{f}^*) \cdot \nabla\mu(\boldsymbol{f}^*) - \tfrac{1}{2}(\tilde{\boldsymbol{f}}_n - \boldsymbol{f}^*)[\nabla^2\mu(\boldsymbol{f}^*) + \boldsymbol{o}(1)](\tilde{\boldsymbol{f}}_n - \boldsymbol{f}^*)^\top. \end{aligned} \tag{11}$$

Now, if $f_i^* = 0$ then $(\partial\mu/\partial f_i)(\boldsymbol{f}^*) \leq 0$, while if $f_i^* > 0$ then $(\partial\mu/\partial f_i)(\boldsymbol{f}^*) = 0$. This allows us to conclude that

$$\tilde{\boldsymbol{f}}_n = \boldsymbol{f}^* + \boldsymbol{o}(n^{-1/4}).$$

Alternatively, if $\mu(\boldsymbol{f})$ has a maximum at $\boldsymbol{f}^* \in \Delta^\circ$ (not necessarily unique), and if $\boldsymbol{f}_0 \in \Delta^\circ$ satisfies $\mu(\boldsymbol{f}_0) < \mu(\boldsymbol{f}^*)$, then (10) implies that

$$\frac{\operatorname{median}[F_n(\boldsymbol{f}^*)]}{\operatorname{median}[F_n(\boldsymbol{f}_0)]} \to \infty$$

exponentially fast. Thus, the results of Section 4 can be extended to this setting under suitable assumptions.

6. Remarks

Remark 1. Leib (2000) argued that the Kelly system is not special because it optimizes neither mean fortune nor the probability of a positive net gain. When it was suggested to him that he consider the median, he responded that "…maximizing median final bankroll is a concept with a face only a mathematician could love."

We disagree. The mean and the median are the two most widely used measures of central tendency, with the median often preferred for highly skewed populations such as annual incomes or housing prices. The proportional bettor's fortune is, by virtue of (3), approximately lognormal, which is itself a rather skewed distribution. For this reason, the median is not only an appropriate measure, it is arguably the most natural.

However, the median does not correspond to a utility function. To those, such as Maslov and Zhang (1998), who see utility functions as artificial, this would not be a problem. To economists, it might be seen as a drawback.

Remark 2. We have assumed throughout that $X_1 \geq -1$, that is, one cannot lose more than one bets. However, as Thorp (2000) pointed out, this is too restrictive in the financial markets.

It is also too restrictive in blackjack, due to the doubling and splitting options. Fortunately, the assumption can be relaxed to $X_1 \geq -K$, where $K > 0$.

To see this, it suffices to note that if $X_1 \geq -K$, then X_1/K satisfies the original assumption; moreover, denoting $F_n(f)$ of (1) by $F_n[X_1, \ldots, X_n](f)$, we have

$$F_n[X_1, \ldots, X_n](f/K) = F_n[X_1/K, \ldots, X_n/K](f).$$

For example, if f^* is the Kelly betting proportion for $X_1/K, \ldots, X_n/K$, then f^*/K is the Kelly betting proportion for $X_1, \ldots, X_n$. The same implication holds for median-maximizing betting proportions.

Remark 3. Kelly's system is widely used in practice – see Thorp (2000) for a survey of applications. Of course, in most cases the assumptions we have made are not literally satisfied. Specifically, the assumption of i.i.d. trials is usually not met, and the implicitly assumed infinite divisibility of capital is rarely if ever satisfied. Although we do not have a formal result to this effect, we believe that conclusions based on these assumptions are quite robust to minor departures from the assumptions.

References

Breiman, L. (1961). Optimal gambling systems for favorable games. In *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, Vol. 1, University of California Press, Berkeley, pp. 65–78.

Edelman, D. (1979). Supremum of mean–median differences for the binomial and Poisson distributions: ln 2. Tech. Rep., Department of Mathematical Statistics, Columbia University.

Ethier, S. N. (1988). The proportional bettor's fortune. In *Gambling Research: Proc. 7th Internat. Conf. Gambling Risk Taking*, Vol. 4, ed. W. R. Eadington, Bureau of Business and Economic Research, Reno, NV, pp. 375–383.

Gnedenko, B. V. and Kolmogorov, A. N. (1968). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Reading, MA.

Hall, P. (1980). On the limiting behaviour of the mode and median of a sum of independent random variables. *Ann. Prob.* **8,** 419–430.

Hamza, K. (1995). The smallest uniform upper bound on the distance between the mean and the median of the binomial and Poisson distributions. *Statist. Prob. Lett.* **23,** 21–25.

Kelly, J. L. Jr. (1956). A new interpretation of information rate. *Bell System Tech. J.* **35,** 917–926.

Leib, J. E. (2000). Limitations on Kelly, or the ubiquitous '$n \to \infty$'. In *Finding the Edge: Mathematical Analysis of Casino Games*, eds O. Vancura, J. A. Cornelius and W. R. Eadington, Institute for the Study of Gambling and Commercial Gaming, Reno, NV, pp. 233–258.

Maslov, S. and Zhang, Y.-C. (1998). Optimal investment strategy for risky assets. *Internat. J. Theoret. Appl. Finance* **1,** pp. 377–387.

Thorp, E. O. (2000). The Kelly criterion in blackjack, sports betting, and the stock market. In *Finding the Edge: Mathematical Analysis of Casino Games*, eds O. Vancura, J. A. Cornelius and W. R. Eadington, Institute for the Study of Gambling and Commercial Gaming, Reno, NV, pp. 163–213.