Risk-Constrained Kelly Gambling

Enzo Busseti   Ernest K. Ryu   Stephen Boyd

March 19, 2016

Abstract

We consider the classic Kelly gambling problem with general distribution of outcomes, and an additional risk constraint that limits the probability of a drawdown of wealth to a given undesirable level. We develop a bound on the drawdown probability; using this bound instead of the original risk constraint yields a convex optimization problem that guarantees the drawdown risk constraint holds. Numerical experiments show that our bound on drawdown probability is reasonably close to the actual drawdown risk, as computed by Monte Carlo simulation. Our method is parametrized by a single parameter that has a natural interpretation as a risk-aversion parameter, allowing us to systematically trade off asymptotic growth rate and drawdown risk. Simulations show that this method yields bets that out perform fractional-Kelly bets for the same drawdown risk level or growth rate. Finally, we show that a natural quadratic approximation of our convex problem is closely connected to the classical mean-variance Markowitz portfolio selection problem.

1 Introduction

In 1956 John Kelly proposed a systematic way to allocate a total wealth across a number of bets so as to maximize the long term growth rate when the gamble is repeated [Kel56, MTZ11]. Similar results were later derived in the finance literature, under the name of *growth-optimum portfolio*; see, *e.g.*, [Mer90, Ch. 6]). It is well known that with Kelly optimal bets there is a risk of the wealth dropping substantially from its original value before increasing, *i.e.*, a *drawdown.* Several ad hoc methods can be used to limit this drawdown risk, at the cost of decreased growth rate. The best known method is fractional Kelly betting, in which only a fraction of the Kelly optimal bets are made [DL12]. The same method has been proposed in the finance literature [Bro00]. Another ad hoc method is Markowitz's mean-variance portfolio optimization [Mar52], which trades off two objectives that are related to, but not the same as, long term growth rate and drawdown risk.

In this paper we directly address drawdown risk and show how to find bets that trade off drawdown risk and growth rates. We introduce the risk-constrained Kelly gambling problem, in which the long term wealth growth rate is maximized with an additional constraint that limits the probability of a drawdown to a specified level. This idealized problem captures

what we want, but seems very difficult to solve. We then introduce a convex optimization problem that is a *restriction* of this problem; that is, its feasible set is smaller than that of the risk-constrained problem. This problem is tractable, using modern convex optimization methods.

Our method can be used in two ways. First, it can be used to find a conservative set of bets that are guaranteed to satisfy a given drawdown risk constraint. Alternatively, its single parameter can be interpreted as a risk-aversion parameter that controls the trade-off between growth rate and drawdown risk, analogous to Markowitz mean-variance portfolio optimization [Mar52], which trades off mean return and (variance) risk. Indeed, we show that a natural quadratic approximation of our convex problem can be closely connected to Markowitz mean-variance portfolio optimization.

In §2 we review the Kelly gambling problem, and describe methods for computing the optimal bets using convex optimization. In simple cases, such as when there are two possible outcomes, the Kelly optimal bets are well known. In others, for example when the returns come from infinite distributions, the methods do not seem to be well known. In §3, we define the drawdown risk, and in §4, we derive a bound on the drawdown risk. In §5, we use this bound to form the risk-constrained Kelly gambling problem, which is a tractable convex optimization problem. In §6, we derive a quadratic approximation of the risk-constrained Kelly gambling problem, and relate it to classical Markowitz portfolio optimization. Finally, in §7 we give some numerical examples to illustrate the methods.

2   Kelly gambling

In Kelly gambling, we place a fixed fraction of our total wealth (assumed positive) on n bets. We denote the fractions as $b \in \mathbf{R}^n$, so $b \geq 0$ and $\mathbf{1}^T b = 1$, where $\mathbf{1}$ is the vector with all components 1. The n bets have a random nonnegative payoff or return, denoted $r \in \mathbf{R}^n_+$, so the wealth after the bet changes by the (random) factor $r^T b$. We will assume that all bets do not have infinite return in expectation, *i.e.*, that $\mathbf{E} r_i < \infty$ for $i = 1, \ldots, n$. We will also assume that the bet n has a certain return of one, *i.e.*, $r_n = 1$ almost surely. This means that b_n represents the fraction of our wealth that we do not wager, or hold as cash. The bet vector $b = e_n$ corresponds to not betting at all. We refer to the bets $1, \ldots, n-1$ as the risky bets.

We mention some special cases of this general Kelly gambling setup.

- *Two outcomes.* We have $n = 2$, and r takes on only two values: $(P, 1)$, with probability π, and $(0, 1)$, with probability $1 - \pi$. The first outcome corresponds to winning the bet (π is the probability of winning) and $P > 1$ is the payoff.

- *Mutually exclusive outcomes.* There are $n-1$ mutually exclusive outcomes, with return vectors: $r = (P_k e_k, 1)$ with probability $\pi_k > 0$, for $k = 1, \ldots, n-1$, where $P_k > 1$ is the payoff for outcome k, and e_k is the unit vector with kth entry one and all others 0. Here we bet on which of outcomes $1, \ldots, n-1$ will be the winner (*e.g.*, the winner of a horse race).

- *General finite outcomes.* The return vector takes on K values $r_1, \ldots, r_K$, with probabilities $\pi_1, \ldots, \pi_K$. This case allows for more complex bets, for example in horse racing show, place, exacta, perfecta, and so on.

- *General returns.* The return r comes from an arbitrary infinite distribution (with $r_n = 1$ almost surely). If the returns are log-normal, the gamble is a simple model of investing (long only) in $n - 1$ assets with log-normal returns; the nth asset is risk free (cash). More generally, we can have $n - 1$ arbitrary derivatives (*e.g.*, options) with payoffs that depend on an underlying random variable.

2.1 Wealth growth

The gamble is repeated at times (epochs) $t = 1, 2, \ldots$, with IID (independent and identically distributed) returns. Starting with initial wealth $w_1 = 1$, the wealth at time t is given by

$$w_t = (r_1^T b) \cdots (r_{t-1}^T b),$$

where r_t denotes the realized return at time t (and not the tth entry of the vector). The wealth sequence $\{w_t\}$ is a stochastic process that depends on the choice of the bet vector b, as well as the distribution of return vector r. Our goal is to choose b so that, roughly speaking, the wealth becomes large.

Note that $w_t \geq 0$, since $r \geq 0$ and $b \geq 0$. The event $w_t = 0$ is called *ruin*, and can only happen if $r^T b = 0$ has positive probability. The methods for choosing b that we discuss below all preclude ruin, so we will assume it does not occur, *i.e.*, $w_t > 0$ for all t, almost surely. Note that if $b_n > 0$, ruin cannot occur since $r^T b \geq b_n$ almost surely.

With $v_t = \log w_t$ denoting the logarithm of the wealth, we have

$$v_t = \log(r_1^T b) + \cdots + \log(r_{t-1}^T b).$$

Thus $\{v_t\}$ is a random walk, with increment distribution given by the distribution of $\log(r^T b)$. The drift of the random walk is $\mathbf{E} \log(r^T b)$; we have $\mathbf{E} v_t = (t - 1) \mathbf{E} \log(r^T b)$, and $\mathbf{var}\, v_t = (t - 1) \mathbf{var} \log(r^T b)$. The quantity $\mathbf{E} \log(r^T b)$ can be interpreted as the average growth rate of the wealth; it is the drift in the random walk $\{v_t\}$. The (expected) growth rate $\mathbf{E} \log(r^T b)$ is a function of the bet vector b.

2.2 Kelly gambling

In Kelly gambling, we choose b to maximize $\mathbf{E} \log(r^T b)$, the growth rate of wealth. This leads to the optimization problem

$$\begin{array}{ll} \text{maximize} & \mathbf{E} \log(r^T b) \\ \text{subject to} & \mathbf{1}^T b = 1, \quad b \geq 0, \end{array} \tag{1}$$

with variable b. We call a solution $b^\star$ of this problem a set of *Kelly optimal* bets. The Kelly gambling problem is always feasible, since $b = e_n$ (which corresponds to not placing

any bets) is feasible. This choice achieves objective value zero, so the optimal value of the Kely gambling problem is always nonnegative. The Kelly gambling problem (1) is a convex optimization problem, since the objective is concave and the constraints are convex.

Kelly optimal bets maximize growth rate of the wealth. If $\tilde{b}$ is a bet vector that is not Kelly optimal, with associated wealth sequence $\tilde{w}_t$, and $b^\star$ is Kelly optimal, with associated wealth sequence $w_t^\star$, then $w_t^\star/\tilde{w}_t \to \infty$ with probability one as $t \to \infty$. (This follows immediately since the random walk $\log w_t^\star - \log \tilde{w}_t$ has positive drift [Fel71, §XII.2]. Also see [CT12, §16] for a general discussion of Kelly gambling.)

We note that the bet vector $b = e_n$ is Kelly optimal if and only if $\mathbf{E}r_i \leq 1$ for $i = 1, \ldots, n-1$. Thus we should not bet at all if all the bets are losers in expectation; conversely, if just one bet is a winner in expectation, the optimal bet is not the trivial one e_n, and the optimal growth rate is positive. We show this in the appendix.

2.3   Computing Kelly optimal bets

Here we describe methods to compute Kelly optimal bets, *i.e.*, to solve the Kelly optimization problem (1). It can be solved analytically or semi-analytically for simple cases; the general finite outcomes case can be handled by standard convex optimization tools, and the general case can be handled via stochastic optimization.

Two outcomes.   For a simple bet with two outcomes, with win probability π and payoff P, we obtain the optimal bet with simple minimization of a univariate function. We have

$$b^\star = \left(\frac{\pi P - 1}{P - 1}, \frac{P - \pi P}{P - 1} \right),$$

provided $\pi P > 1$; if $\pi P \leq 1$, the optimal bet is $b^\star = (0, 1)$. Thus we should bet a fraction $(\pi P - 1)/(P - 1)$ of our wealth each time, if this quantity is positive.

General finite outcomes.   When the return distribution is finite the Kelly gambling problem reduces to

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^{K} \pi_i \log(r_i^T b) \\ \text{subject to} & \mathbf{1}^T b = 1, \quad b \geq 0, \end{array} \tag{2}$$

which is readily solved using convex optimization methods [BV04]. Convex optimization software systems like CVX [GB14], CVXPY [DB16] and Convex.jl [UMZ$^+$14], based on DCP (Disciplined Convex Programming) [GBY06], or others like YALMIP [Lö4], can handle such problems directly. In our numerical simulation we use CVXPY with the open source solver ECOS [DCB13], recently extended to handle exponential cone constraints [Ser15].

General returns.   We can solve the Kelly gambling problem (1) even in the most general case, when r takes on an infinite number of values, provided we can generate samples from the distribution of r. In this case we can use a projected stochastic gradient method with averaging [RM51, NY83, Pol87, KY03, Bub15].

As a technical assumption we assume here that the Kelly optimal bet $b^\star$ satisfies $(b^\star)_n > 0$, *i.e.*, the optimal bet involves holding some cash. We assume we know $\varepsilon > 0$ that satisfies $\varepsilon < (b^\star)_n$ (which implies that $r^T b > \varepsilon$ a.s.) and define $\Delta_\varepsilon = \{b \mid \mathbf{1}^T b = 1, \ b \geq 0, \ b_n \geq \varepsilon\}$. Then the gradient of the objective is given by

$$\nabla_b \mathbf{E} \log(r^T b) = \mathbf{E} \nabla_b \log(r^T b) = \mathbf{E} \frac{1}{r^T b} r$$

for any $b \in \Delta_\varepsilon$. So if $r^{(k)}$ is an IID sample from the distribution,

$$\frac{1}{r^{(k)T} b} r^{(k)} \tag{3}$$

is an unbiased estimate of the gradient, *i.e.*, a stochastic gradient, of the objective at b. (Another unbiased estimate of the gradient can be obtained by averaging the expression (3) over multiple return samples.)

The (projected) stochastic gradient method with averaging computes the iterates

$$\bar{b}^{(k+1)} = \Pi \left(\bar{b}^{(k)} + \frac{t_k}{(r^{(k)T} \bar{b}^{(k)})} r^{(k)} \right), \quad k = 1, 2, \dots,$$

where the starting point $\bar{b}^{(1)}$ is any vector in Δ_ε, $r^{(k)}$ are IID samples from the distribution of r, and Π is (Euclidean) projection onto Δ_ε (which is readily computed; see Lemma 3). The step sizes $t_k > 0$ must satisfy

$$t_k \to 0, \qquad \sum_{k=1}^{\infty} t_k = \infty.$$

(For example, $t_k = C/\sqrt{k}$ with any $C > 0$ satisfies this condition.) Then the (weighted) running average

$$b^{(k)} = \frac{\sum_{i=1}^{k} t_i \bar{b}^{(i)}}{\sum_{i=1}^{k} t_i}$$

converges to Kelly optimal. The stochastic gradient method can be slow to converge, but it always works; that is $\mathbf{E} \log(r^T b^{(k)})$ converges to the optimal growth rate.

In practice, one does not know a priori how small ε should be. One way around this is to choose a small ε, and then check that $(b^{(k)})_n > \varepsilon$ holds for large k, in which case we know our guess of ε was valid. A more important practical variation on the algorithm is *batching*, where we replace the unbiased estimate of the gradient with the average over some number of samples. This does not affect the theoretical convergence of the algorithm, but can improve convergence in practice.

5

3 Drawdown

We define the *minimum wealth* as the infimum of the wealth trajectory over time,

$$W^{\min} = \inf_{t=1,2,\ldots} w_t.$$

This is a random variable, with distribution that depends on b. With $b = e_n$, we have $w_t = 1$ for all t, so $W^{\min} = 1$. With b for which $\mathbf{E}\log(r^T b) > 0$ (which we assume), $W^{\min}$ takes values in $(0, 1]$. Small $W^{\min}$ corresponds to a case where the initial wealth drops to a small value before eventually increasing.

The *drawdown* is defined as $1 - W^{\min}$. A drawdown of 0.3 means that our wealth dropped 30% from its initial value (one), before increasing (which it eventually must do, since $v_t \to \infty$ with probability one). Several other definitions of drawdown are used in the literature. A large drawdown means that $W^{\min}$ is small, *i.e.*, our wealth drops to a small value before growing.

The *drawdown risk* is defined as $\mathbf{Prob}(W^{\min} < \alpha)$, where $\alpha \in (0, 1)$ is a given target (undesired) minimum wealth. This risk depends on the bet vector b in a very complicated way. There is in general no formula for the risk in terms of b, but we can always (approximately) compute the drawdown risk for a given b using Monte Carlo simulation. As an example, a drawdown risk of 0.1 for $\alpha = 0.7$ means the probability of experiencing more than 30% drawdown is only 10%. The smaller the drawdown risk (with any target), the better.

3.1 Fractional Kelly gambling

It is well known that Kelly optimal bets can lead to substantial drawdown risk. One ad hoc method for handling this is to compute a Kelly optimal bet $b^\star$, and then use the so-called fractional Kelly [DL12] bet given by

$$b = fb^\star + (1 - f)e_n, \tag{4}$$

where $f \in (0, 1)$ is the fraction. The fractional Kelly bet scales down the (risky) bets by f. Fractional Kelly bets have smaller drawdowns than Kelly bets, at the cost of reduced growth rate. We will see that trading off growth rate and drawdown risk can be more directly (and better) handled.

3.2 Kelly gambling with drawdown risk

We can add a drawdown risk constraint to the Kelly gambling problem (1), to obtain the problem

$$
\begin{aligned}
\text{maximize} \quad & \mathbf{E}\log(r^T b) \\
\text{subject to} \quad & \mathbf{1}^T b = 1, \quad b \geq 0, \\
& \mathbf{Prob}(W^{\min} < \alpha) < \beta,
\end{aligned}
\tag{5}
$$

with variable b, where $\alpha, \beta \in (0, 1)$ are given parameters. The last constraint limits the probability of a drop in wealth to value α to be no more than β. For example, we might

take $\alpha = 0.7$ and $\beta = 0.1$, meaning that we require the probability of a drawdown of more than 30% to be less than 10%. (This imposes a constraint on the bet vector b.)

Unfortunately the problem (5) is, as far as we know, a difficult optimization problem in general. In the next section we will develop a bound on the drawdown risk that results in a tractable convex constraint on b. We will see in numerical simulations that the bound is generally quite good.

4 Drawdown risk bound

In this section we derive a condition that bounds the drawdown risk. Consider any $\lambda > 0$ and bet b. For any $\alpha \in (0,1)$ and $\beta \in (0,1)$ that satisfies $\lambda = \log \beta / \log \alpha$ we have

$$\mathbf{E}(r^T b)^{-\lambda} \leq 1 \implies \mathbf{Prob}(W^{\min} < \alpha) < \beta. \tag{6}$$

In other words, if our bet satisfies $\mathbf{E}(r^T b)^{-\lambda} \leq 1$, then its drawdown risk $\mathbf{Prob}(W^{\min} < \alpha)$ less than β.

To see this, consider the stopping time

$$\tau = \inf\{t \geq 1 \,|\, w_t < \alpha\},$$

and note $\tau < \infty$ if and only if $W^{\min} < \alpha$. From Lemma 5 of the appendix, we get

$$1 \geq \mathbf{E}\left[\exp(-\lambda \log w_\tau - \tau \log \mathbf{E}(r^T b)^{-\lambda}) \,|\, \tau < \infty\right] \mathbf{Prob}(W^{\min} < \alpha).$$

Since $-\tau \log \mathbf{E}(r^T b)^{-\lambda} \geq 0$ when $\tau < \infty$, we have

$$1 \geq \mathbf{E}\left[\exp(-\lambda \log w_\tau) \,|\, \tau < \infty\right] \mathbf{Prob}(W^{\min} < \alpha).$$

Since $w_\tau < \alpha$ when $\tau < \infty$, we have

$$1 > \exp(-\lambda \log \alpha)\mathbf{Prob}(W^{\min} < \alpha).$$

So we have

$$\mathbf{Prob}(W^{\min} < \alpha) < \alpha^\lambda = \beta.$$

5 Risk-constrained Kelly gambling

Replacing the drawdown risk constraint in the problem (5) with the lefthand side of (6), with $\lambda = \log \beta / \log \alpha$, yields the risk-constrained Kelly gambling problem (RCK)

$$
\begin{array}{ll}
\text{maximize} & \mathbf{E}\log(r^T b), \\
\text{subject to} & \mathbf{1}^T b = 1, \quad b \geq 0, \\
& \mathbf{E}(r^T b)^{-\lambda} \leq 1,
\end{array}
\tag{7}
$$

with variable b. We refer to a solution of this problem as an RCK bet. The RCK problem is a *restriction* of problem (5), since it has a smaller feasible set: any b that is feasible for RCK must satisfy the drawdown risk constraint $\mathbf{Prob}(W^{\min} < \alpha) < \beta$. In the limit when either $\beta \to 1$ or $\alpha \to 0$ we get $\lambda \to 0$. For $\lambda = 0$, the second constraint is always satisfied, and the RCK problem (7) reduces to the (unconstrained) Kelly gambling problem (1).

Let us now show that the RCK problem (7) is convex. The objective is concave and the constraints $\mathbf{1}^T b = 1$, $b \geq 0$ are convex. To see that the function $\mathbf{E}(r^T b)^{-\lambda}$ is convex in b, we note that for $r^T b > 0$, $(r^T b)^{-\lambda}$ is a convex function of b; so the expectation $\mathbf{E}(r^T b)^{-\lambda}$ is a convex function of b (see [BV04, §3.2]). We mention that the last constraint can also be written as $\log \mathbf{E}(r^T b)^{-\lambda} \leq 0$, where the lefthand side is a convex function of b.

The RCK problem (7) is always feasible, since $b = e_n$ is feasible. Just as in the Kelly gambling problem, the bet vector $b = e_n$ is optimal for RCK (7) if and only if $\mathbf{E}r_i \leq 1$ for $i = 1, \ldots, n-1$. In other words we should not bet at all if all the bets are losers in expectation; conversely, if just one bet is a winner in expectation, the solution of the RCK problem will have positive growth rate (and of course respect the drawdown risk constraint). We show this in the appendix.

5.1   Risk aversion parameter

The RCK problem (7) depends on the parameters α and β only through $\lambda = \log \beta / \log \alpha$. This means that, for fixed λ, our one constraint $\mathbf{E}(r^T b)^{-\lambda} \leq 1$ actually gives us a family of drawdown constraints that must be satisfied:

$$\mathbf{Prob}(W^{\min} < \alpha) < \alpha^\lambda \tag{8}$$

holds for all $\alpha \in (0,1)$. For example $\alpha = 0.7$ and $\beta = 0.1$ gives $\lambda = 6.46$; thus, our constraint also implies that the probability of a drop in wealth to $\alpha = 0.5$ (*i.e.*, we lose half our initial wealth) is no more than $(0.5)^{6.46} = 0.011$. Another interpretation of (8) is that our risk constraint actually bounds the entire CDF (cumulative distribution function) of $W^{\min}$: it stays below the function $\alpha \mapsto \alpha^\lambda$.

The RCK problem (7) can be used in two (related) ways. First, we can start from the original drawdown specifications, given by α and β, and then solve the problem using $\lambda = \log \beta / \log \alpha$. In this case we are guaranteed that the resulting bet satisfies our drawdown constraint. An alternate use model is to consider λ as a *risk-aversion parameter*; we vary it to trade off growth rate and drawdown risk. This is very similar to traditional Markowitz portofolio optimization [Mar52] [BV04, §4.4.1], where a risk aversion parameter is used to trade off risk (measured by portfolio return variance) and return (expected portfolio return). We will see another close connection between our method and Markowitz mean-variance portfolio optimization in §6.

5.2   Light and heavy risk aversion regimes

In this section we give provide an interpretation of the drawdown risk constraint

$$\mathbf{E}(r^T b)^{-\lambda} \leq 1 \tag{9}$$

in the light and heavy risk aversion regimes, which correspond to small and large values of λ, respectively.

Heavy risk aversion. In the heavy risk aversion regime, *i.e.*, in the limit $\lambda \to \infty$, the constraint (9) reduces to $r^T b \geq 1$ almost surely. In other words, problem (7) only considers risk-free bets in this regime.

Light risk aversion. Next consider the light risk aversion regime, *i.e.*, in the limit $\lambda \to 0$. Note that constraint (9) is equivalent to

$$\frac{1}{\lambda} \log \mathbf{E} \exp(-\lambda \log(r^T b)) \leq 0.$$

As $\lambda \to 0$ we have

$$
\begin{aligned}
\frac{1}{\lambda} \log \mathbf{E} \exp(-\lambda \log(r^T b)) &= \frac{1}{\lambda} \log \mathbf{E} \left[1 - \lambda \log(r^T b) + \frac{\lambda^2}{2} (\log(r^T b))^2 + O(\lambda^3) \right] \\
&= -\mathbf{E} \log(r^T b) + \frac{\lambda}{2} \mathbf{E} (\log(r^T b))^2 - \frac{\lambda}{2} (\mathbf{E} \log(r^T b))^2 + O(\lambda^2) \\
&= -\mathbf{E} \log(r^T b) + \frac{\lambda}{2} \mathbf{var} \log(r^T b) + O(\lambda^2).
\end{aligned}
$$

(In the context of stochastic control, $(1/\lambda) \log \mathbf{E}_X \exp(-\lambda X)$ is known as the exponential disutility or loss and λ as the risk-sensitivity parameter. This asymptotic expansion is well-known see *e.g.* [Whi81, Whi90].) So constraint (9) reduces to

$$\frac{\lambda}{2} \mathbf{var} \log(r^T b) \leq \mathbf{E} \log(r^T b)$$

in the limit $\lambda \to 0$. Thus the (restricted) drawdown risk contraint (9) limits the ratio of variance to mean growth in this regime.

5.3 Computing RCK bets

Two outcomes. For the two outcome case we can easily solve the problem (7), almost analytically. The problem is

$$
\begin{aligned}
\text{maximize} \quad & \pi \log(b_1 P + (1 - b_1)) + (1 - \pi)(1 - b_1), \\
\text{subject to} \quad & 0 \leq b_1 \leq 1, \\
& \pi(b_1 P + (1 - b_1))^{-\log \beta / \log \alpha} + (1 - \pi)(1 - b_1)^{-\log \beta / \log \alpha} \leq 1.
\end{aligned}
\tag{10}
$$

If the solution of the unconstrained problem,

$$\left(\frac{\pi P - 1}{P - 1}, \frac{P - \pi P}{P - 1} \right),$$

9

satisfies the risk constraint, then it is the solution. Otherwise we reduce b_1 to find the value for which

$$\pi(b_1 P + (1 - b_1))^{-\lambda} + (1 - \pi)(1 - b_1)^{-\log \lambda} = 1.$$

(This can be done by bisection since the lefthand side is monotone in b_1.) In this case the RCK bet is a fractional Kelly bet (4), for some $f < 1$.

Finite outcomes case. For the finite outcomes case we can restate the RCK problem (7) in a convenient and tractable form. We first take the log of the last constraint and get

$$\log \sum_{i=1}^{K} \pi_i (r_i^T b)^{-\lambda} \le 0,$$

we then write it as

$$\log \left(\sum_{i=1}^{K} \exp(\log \pi_i - \lambda \log(r_i^T b)) \right) \le 0.$$

To see that this constraint is convex we note that the log-sum-exp function is convex and increasing, and its arguments are all convex functions of b (since $\log(r_i^T b)$ is concave), so the lefthand side function is convex in b [BV04, §3.2]. Moreover, convex optimization software systems like CVX, CVXPY, and Convex.jl based on DCP (disciplined convex programming) can handle such compositions directly. Thus we have the problem

$$
\begin{array}{ll}
\text{maximize} & \sum_{i=1}^{K} \pi_i \log(r_i^T b), \\
\text{subject to} & \mathbf{1}^T b = 1, \quad b \ge 0, \\
& \log \left(\sum_{i=1}^{K} \exp(\log \pi_i - \lambda \log(r_i^T b)) \right) \le 0.
\end{array}
\tag{11}
$$

In this form the problem is readily solved; its CVXPY specification is given in appendix B.

General returns. As with the Kelly gambling problem, we can solve the RCK problem (7) using a stochastic optimization method. We use a primal-dual stochastic gradient method (from [NY78, NJLS09]) applied to the Lagrangian

$$L(b, \kappa) = -\mathbf{E} \log(r^T b) + \kappa(\mathbf{E}(r^T b)^{-\lambda} - 1),
\tag{12}$$

with $b \in \Delta_\varepsilon$ and $\kappa \ge 0$. (As in the unconstrained Kelly optimization case, we make the technical assumption that $b_n^\star > \varepsilon > 0$.) In the appendix we show that the RCK problem (7) has an optimal dual variable $\kappa^\star$ for the constraint $\mathbf{E}(r^T b)^{-\lambda} \le 1$, which implies that solving problem (7) is equivalent to finding a saddle point of (12). We also assume we know an upper bound M on the value of the optimal dual variable $\kappa^\star$.

Our method computes the iterates

$$\bar{b}^{(k+1)} = \Pi \left(\bar{b}^{(k)} + t_k \frac{(r^{(k)T} \bar{b}^{(k)})^\lambda + \lambda \bar{\kappa}^{(k)}}{(r^{(k)T} \bar{b}^{(k)})^{\lambda+1}} r^{(k)} \right),$$

$$\bar{\kappa}^{(k+1)} = \left(\bar{\kappa}^{(k)} + t_k \frac{1 - (r^{(k)T} \bar{b}^{(k)})^\lambda}{(r^{(k)T} \bar{b}^{(k)})^\lambda} \right)_{[0,M]},$$

where the starting points $\bar{b}^{(1)}$ and $\bar{\kappa}^{(1)}$ are respectively in Δ_ε and $[0, M]$, $r^{(k)}$ are IID samples from the distribution of r, Π is Euclidean projection onto Δ_ε, and $(a)_{[0,M]}$ is projection onto $[0, M]$, *i.e.*,

$$(a)_{[0,M]} = \max\{0, \min\{M, a\}\}.$$

The step sizes $t_k > 0$ must satisfy

$$t_k \to 0, \qquad \sum_{k=1}^{\infty} t_k = \infty.$$

We use the (weighted) running averages

$$b^{(k)} = \frac{\sum_{i=1}^{k} t_i \bar{b}^{(i)}}{\sum_{i=1}^{k} t_i}, \qquad \kappa^{(k)} = \frac{\sum_{i=1}^{k} t_i \bar{\kappa}^{(i)}}{\sum_{i=1}^{k} t_i}$$

as our estimates of the optimal bet and $\kappa^\star$, respectively.

Again, this method can be slow to converge, but it always works; that is $\mathbf{E}\log(r^T b^{(k)})$ converges to the optimal value and $\max\{\mathbf{E}(r^T b^{(k)}))^{-\lambda} - 1, 0\} \to 0$. As in the unconstrained Kelly case, we do not know a priori how small ε should be, or how large M should be. We can choose a small ε and large M and later verify that $(\bar{b}^{(k)})_n > \varepsilon$, and $\bar{\kappa}^{(k)} < M$; if this holds, our guesses of ε and M were valid. Also as in the unconstrained case, batching can be used to improve the practical convergence. In this case, we replace our unbiased estimates of the gradients of the two expectations with an average over some number of them.

Finally, we mention that the optimality conditions can be independently checked. As we show in Lemma 4 of the appendix, a pair $(b^\star, \kappa^\star)$ is a solution of the RCK problem if and only if it satisfies the following optimality conditions:

$$
\begin{aligned}
\mathbf{1}^T b^\star &= 1, \quad b^\star \geq 0, \quad \mathbf{E}(r^T b^\star)^{-\lambda} \leq 1 \\
\kappa^\star &\geq 0, \quad \kappa^\star(\mathbf{E}(r^T b^\star)^{-\lambda} - 1) = 0 \\
\mathbf{E}\frac{r_i}{r^T b^\star} + \kappa^\star \lambda \mathbf{E}\frac{r_i}{(r^T b^\star)^{\lambda+1}} &\begin{cases} \leq 1 + \kappa^\star \lambda & b_i = 0 \\ = 1 + \kappa^\star \lambda & b_i > 0. \end{cases}
\end{aligned}
\tag{13}
$$

These conditions can be checked for a computed approximate solution of RCK, using Monte Carlo simulation to evaluate the expectations. (The method above guarantees that $\mathbf{1}^T b = 1$, $b \geq 0$, and $\kappa \geq 0$, so we only need to check the other three conditions.)

6 Quadratic approximation

In this section we form a quadratic approximation of the RCK problem (7), which we call the *quadratic RCK problem* (QRCK), and derive a close connection to Markowitz portfolio optimization. We use the notation $\rho = r - \mathbf{1}$ for the (random) excess return, so (with

11

$\mathbf{1}^T b = 1$) we have $r^T b - 1 = \rho^T b$. Assuming $r^T b \approx 1$, or equivalently $\rho^T b \approx 0$, we have the (Taylor) approximations

$$\log(r^T b) = \rho^T b - \frac{1}{2}(\rho^T b)^2 + O((\rho^T b)^3),$$

$$(r^T b)^{-\lambda} = 1 - \lambda \rho^T b + \frac{\lambda(\lambda+1)}{2}(\rho^T b)^2 + O((\rho^T b)^3).$$

Substituting these into the RCK problem (7) we obtain the QRCK problem

$$
\begin{array}{ll}
\text{maximize} & \mathbf{E}\rho^T b - \frac{1}{2}\mathbf{E}(\rho^T b)^2 \\
\text{subject to} & \mathbf{1}^T b = 1, \quad b \geq 0 \\
& -\lambda\mathbf{E}\rho^T b + \frac{\lambda(\lambda+1)}{2}\mathbf{E}(\rho^T b)^2 \leq 0.
\end{array}
\tag{14}
$$

This approximation of the RCK problem is a convex quadratic program (QP) which is readily solved. We expect the solution to be a good approximation of the RCK solution when the basic assumption $r^T b \approx 1$ holds.

This approximation can be useful when finding a solution to the RCK problem (7). We first estimate the first and second moments of ρ via Monte Carlo, and then solve the QRCK problem (14) (with the estimated moments) to get a solution b^{qp} and a Langrange multipler κ^{qp}. We take these as good approximations for the solution of (7), and use them as the starting points for the primal-dual stochastic gradient method, i.e., we set $b^{(1)} = b^{\text{qp}}$ and $\kappa^{(1)} = \kappa^{\text{qp}}$. This gives no theoretical advantage, since the method converges no matter what the initial points are; but it can speed up the convergence in practice.

We now connect the QRCK problem (14) to classical Markowitz portfolio selection. We start by defining $\mu = \mathbf{E}\rho$, the mean excess return, and

$$S = \mathbf{E}\rho\rho^T = \Sigma + (\mathbf{E}\rho)(\mathbf{E}\rho)^T,$$

the (raw) second moment of ρ (with Σ the covariance of the return). We say that an allocation vector b is a Markowitz portfolio if it solves

$$
\begin{array}{ll}
\text{maximize} & \mu^T b - \frac{\gamma}{2}b^T\Sigma b \\
\text{subject to} & \mathbf{1}^T b = 1, \quad b \geq 0,
\end{array}
\tag{15}
$$

for some value of the (risk-aversion) parameter $\gamma \geq 0$. A solution to problem (14) is a Markowitz portfolio, provided there are no arbitrages. By no arbitrage we mean that $\mu^T b > 0$, $\mathbf{1}^T b = 1$, and $b \geq 0$, implies $b^T\Sigma b > 0$.

Let us show this. Let b^{qp} be the solution to the QRCK problem (14). By (strong) Lagrange duality [Ber09], b^{qp} is a solution of

$$
\begin{array}{ll}
\text{maximize} & \mu^T b - \frac{1}{2}b^T S b + \nu(\mu^T b - \frac{\lambda+1}{2}b^T S b) \\
\text{subject to} & \mathbf{1}^T b = 1, \quad b \geq 0
\end{array}
$$

for some $\nu \geq 0$, which we get by dualizing only the constraint $-\lambda\mathbf{E}\rho^T b + \frac{\lambda(\lambda+1)}{2}\mathbf{E}(\rho^T b)^2 \leq 0$. We divide the objective by $1 + \nu$ and substitute $S = \mu\mu^T + \Sigma$ to get that b^{qp} is a solution of

$$
\begin{array}{ll}
\text{maximize} & \mu^T b - \frac{\eta}{2}(\mu^T b)^2 - \frac{\eta}{2}b^T\Sigma b \\
\text{subject to} & \mathbf{1}^T b = 1, \quad b \geq 0,
\end{array}
\tag{16}
$$

for some $\eta > 0$. In turn, b^{qp} is a solution of

$$\begin{array}{ll} \text{maximize} & (1 - \eta\mu^T b^{\text{qp}})\mu^T b - \frac{\eta}{2}b^T\Sigma b \\ \text{subject to} & \mathbf{1}^T b = 1, \quad b \geq 0, \end{array} \tag{17}$$

since the objectives of problem (16) and (17) have the same gradient at b^{qp}. If $\mu^T b^{\text{qp}} < 1/\eta$ then problem (17) is equivalent to problem (15) with $\gamma = \eta/(1 - \eta\mu^T b^{\text{qp}})$.

Assume for contradiction that $\mu^T b^{\text{qp}} \geq 1/\eta$, which implies $(b^{\text{qp}})^T\Sigma b^{\text{qp}} > 0$ by the no artibtrage assumption. Then for problem (17) the bet e_n achieves objective value 0, which is better than that of b^{qp}. As this contradicts the optimality of b^{qp}, we have $\mu^T b^{\text{qp}} < 1/\eta$. So we conclude a solution to problem (14) is a solution to problem (15), *i.e.*, b^{qp} is a Markowitz portfolio.

7 Numerical simulations

In this section we report results for two specific problem instances, one with finite outcomes and one with infinite outcomes, but our numerical explorations show that these results are typical.

7.1 Finite outcomes

We consider a finite outcomes case with $n = 20$ (so there are 19 risky bets) and $K = 100$ possible outcomes. The problem data is generated as follows. The probabilities π_i, $i = 1, \ldots, K$ are drawn uniformly on $[0,1]$ and then normalized so that $\sum_i \pi_i = 1$. The returns $r_{ij} \in \mathbf{R}_{++}$ for $i = 1, \ldots, K$ and $j = 1, \ldots, n-1$ are drawn from a uniform distribution in $[0.7, 1.3]$. Then, 30 randomly selected returns r_{ij} are set equal to 0.2 and other 30 equal to 2. (The returns r_{in} for $i = 1, \ldots, K$ are instead all set to 1.) The probability that a return vector r contains at least one "extreme" return (*i.e.*, equal to 0.2 or 2) is $1 - (1 - 60/(19 \cdot 100))^{n-1} \approx 0.45$.

7.1.1 Comparison of Kelly and RCK bets

We compare the Kelly optimal bet with the RCK bets for $\alpha = 0.7$ and $\beta = 0.1$ ($\lambda = 6.456$). We then obtain the RCK bets for $\lambda = 5.500$, a value chosen so that we achieve risk close to the specified value $\beta = 0.1$ (as discussed in §5.1). For each bet vector we carry out 10000 Monte Carlo simulations of w_t for $t = 1, \ldots, 100$. This allows us to estimate (well) the associated risk probabilities. Table 1 shows the results. The second column gives the growth rate, the third column gives the bound on drawdown risk, and the last column gives the drawdown risk computed by Monte Carlo simulation.

The Kelly optimal bet experiences a drawdown exceeding our threshold $\alpha = 0.7$ around 40% of the time. For all the RCK bets the drawdown risk (computed by Monte Carlo) is less than our bound, but not dramatically so. (We have observed this over a wide range of problems.) The RCK bet with $\lambda = 6.456$ is *guaranteed* to have a drawdown probability not

Bet	$\mathbf{E}\log(r^T b)$	$e^{\lambda\log\alpha}$	$\mathbf{Prob}(W^{\min} < \alpha)$
Kelly	0.062	-	0.397
RCK, $\lambda = 6.456$	0.043	0.100	0.073
RCK, $\lambda = 5.500$	0.047	0.141	0.099

Table 1: Comparison of Kelly and RCK bets. Expected growth rate and drawdown risk are computed with Monte Carlo simulations.

exceeding 10%; Monte Carlo simulation shows that it is (approximately) 7.3%. For the third bet vector in our comparison, we decreased the risk aversion parameter until we obtained a bet with (Monte Carlo computed) risk near the limit 10%.

The optimal value of the (hard) Kelly gambling problem with drawdown risk (9) must be less than 0.062 (the unconstrained optimal growth rate) and greater than 0.043 (since our second bet vector is guaranteed to satisfy the risk constraint). Since our third bet vector has drawdown risk less than 10%, we can further refine this result to state that the optimal value of the (hard) Kelly gambling problem with drawdown risk (9) is between 0.062 and 0.047.

Figure 1 shows ten trajectories of w_t in our Monte Carlo simulations for the Kelly optimal bet (left) and the RCK bet obtained with $\lambda = 5.5$ (right). Out of these ten simulations, four of the Kelly trajectories dip below the threshold $\alpha = 0.7$, and one of the other trajectories does, which is consistent with the probabilities reported above.

Figure 2 shows the sample CDF of $W^{\min}$ over the 10000 simulated trajectories of w_t, for the Kelly optimal bets and the RCK bets with $\lambda = 6.46$. The upper bound α^λ is also shown. We see that the risk bound is not bad, typically around 30% or so higher than the actual risk. We have observed this to be the case across many problem instances.

7.1.2   Comparison of RCK and QRCK

Table 2 shows the Monte Carlo values of growth rate and drawdown risk for the QRCK bets with $\lambda = 6.456$ (to compare with the RCK solution in table 1). The QRCK bets come with no guarantee on the drawdown risk, but with $\lambda = 6.456$ the drawdown probability (evaluated by Monte Carlo) is less than $\beta = 0.1$. The value $\lambda = 2.800$ is selected so that the risk is approximately 0.10; we see that its growth rate is smaller than the growth rate of the RCK bet with the same drawdown risk.

Figure 3 shows the values of each b_i with $i = 1, \ldots, 20$, for the Kelly, RCK, and QRCK bets. We can see that the Kelly bet concentrates on outcome 4; the RCK and QRCK bets still make a large bet on outcome 4, but also spread their bets across other outcomes as well.

7.1.3   Risk-growth trade-off of RCK, QRCK, and fractional Kelly bets

In figure 4 we compare the trade-off between drawdown risk and expected growth rate of the RCK and QRCK problems for multiple choices of λ and the fractional Kelly bets (4) for
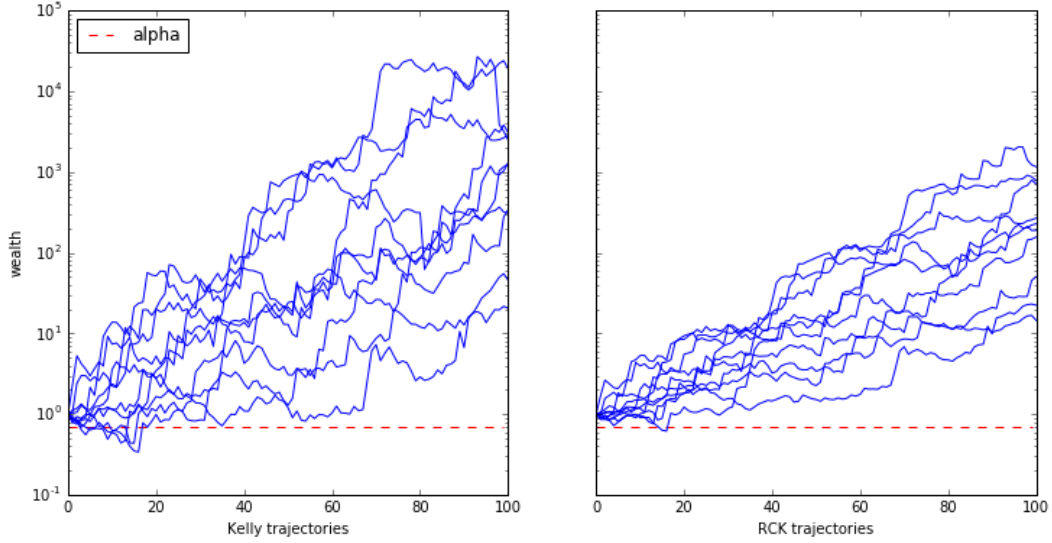
Figure 1: Wealth trajectories for the Kelly optimal bet (left) and the restricted risk-constrained bet with $\lambda = 5.5$. The dashed line shows the wealth threshold $\alpha = 0.7$.

Bet	$\mathbf{E}\log(r^T b)$	$e^{\lambda\log\alpha}$	$\mathbf{Prob}(W^{\min} < \alpha)$
QRCK, $\lambda = 0.000$	0.054	1.000	0.218
QRCK, $\lambda = 6.456$	0.027	0.100	0.025
QRCK, $\lambda = 2.800$	0.044	0.368	0.100

Table 2: Statistics for QRCK bets. Expected growth rate and drawdown risk are computed with Monte Carlo simulations.
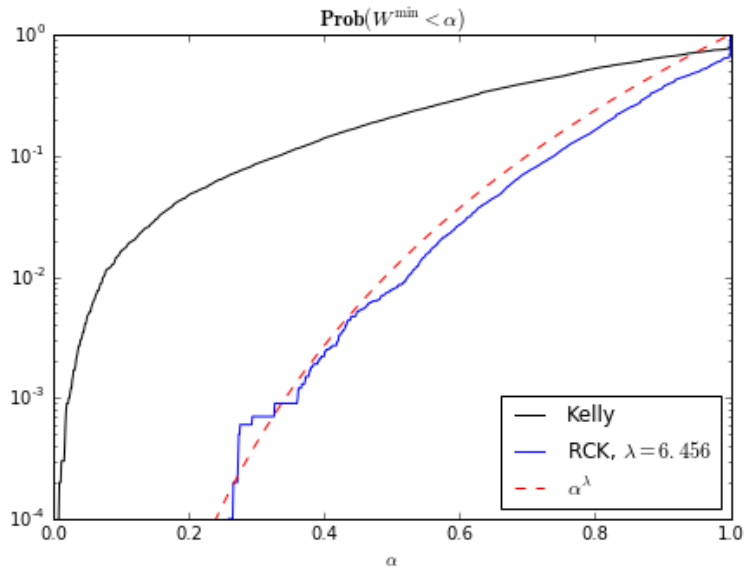
Figure 2: Sample CDF of $W^{\min}$ (*i.e.,* $\mathbf{Prob}(W^{\min} < \alpha)$) for the Kelly optimal bets and RCK bets with $\lambda = 6.46$. The upper bound α^{λ} is also shown.
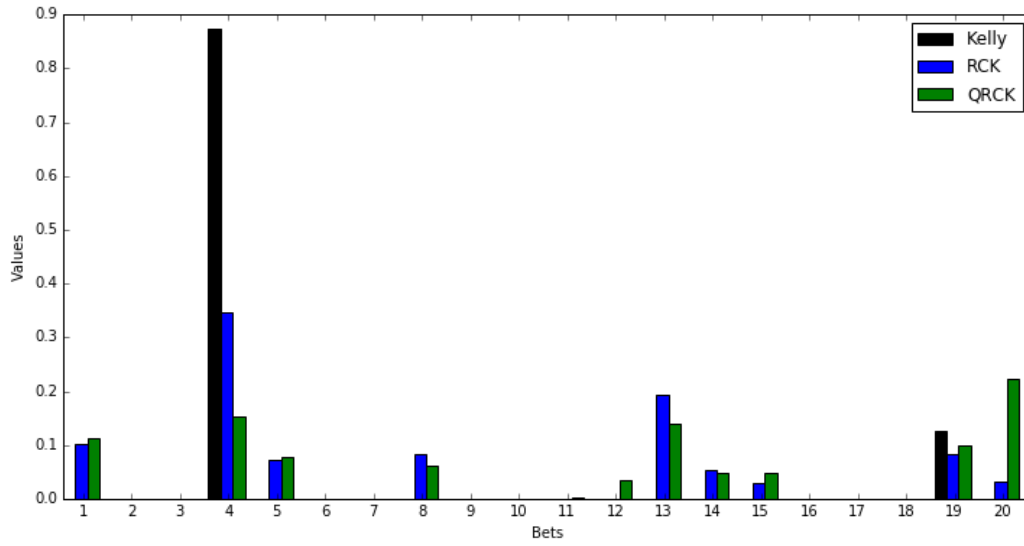

Figure 3: Comparison of the values b_i, $i = 1, \ldots, 20$, for different bet vectors. The RCK and QRCK bets are both obtained with $\lambda = 6.456$.
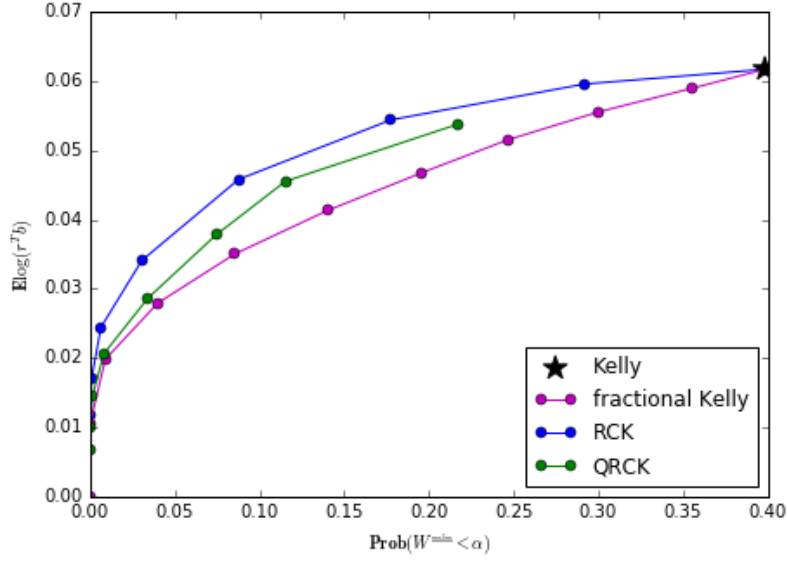
16

Figure 4: Trade-off of drawdown risk and expected growth rate for the RCK, QRCK, and fractional Kelly bets. The Kelly bet is also shown.

multiple choices of f. The plots are akin to the risk-return tradeoff curves that are typical in Markowitz portfolio optimization. We see that RCK yields superior bets than QRCK and fractional Kelly (in some cases substantially better). For example, the Kelly fractional bet that achieves our risk bound 0.1 has a growth rate around 0.035, compared with RCK, which has a growth rate 0.047.

7.2 General returns

We show here an instance of the problem with an infinite distribution of returns, defined as a mixture of lognormals

$$r \sim \begin{cases} \log \mathcal{N}(\nu_1, \Sigma_1) & \text{w.p. } 0.5 \\ \log \mathcal{N}(\nu_2, \Sigma_2) & \text{w.p. } 0.5, \end{cases}$$

with $n = 20$, $\nu_1, \nu_2 \in \mathbf{R}^n$, and $\Sigma_1, \Sigma_2 \in \mathbf{S}_+^n$. We have $\nu_{1n} = \nu_{2n} = 0$, and the matrices Σ_1, Σ_2 are such that r_n has value 1 with probability 1.

We generate a sample of 10^6 observations from this returns distribution and use it to solve the Kelly, RCK, and QRCK problems. In the algorithms for solving Kelly and RCK we use step sizes $t_k = C/\sqrt{k}$ for some $C > 0$ and batching over 100 samples per iteration (so that we run 10^4 iterations). We initialize these algorithms with the QRCK solutions to speed up convergence. To solve the QRCK problem we compute the first and second moments of $\rho = r - 1$ using the same sample of 10^6 observations. We generate a separate

17

Bet	$\mathbf{E}\log(r^T b)$	$e^{\lambda\log\alpha}$	$\mathbf{Prob}(W^{\min} < \alpha)$
Kelly	0.077	-	0.569
RCK, $\lambda = 6.456$	0.039	0.100	0.080
RCK, $\lambda = 5.700$	0.043	0.131	0.101

Table 3: Comparison of Kelly and RCK for the infinite outcome case. Expected growth rate and drawdown risk are computed with Monte Carlo simulations.

sample of returns for Monte Carlo simulations, consisting of 10^4 simulated trajectories of w_t for $t = 1, \ldots, 100$.

7.2.1   Comparison of RCK and Kelly bets

In our first test we compare the Kelly bet with RCK bets for two values of the parameter λ. The first RCK bet has $\lambda = 6.456$, which guarantees that the drawdown risk at $\alpha = 0.7$ is smaller or equal than $\beta = 0.1$. Table 3 shows that this is indeed the case, the Monte Carlo simulated risk is 0.08, not very far from the theoretical bound. The second value of λ is instead chosen so that the Monte Carlo risk is approximately equal to 0.1.

Figure 5 shows 10 simulated trajectories w_t for the Kelly bet and for the RCK bet with $\lambda = 5.700$. In this case 6 of the Kelly trajectories fall below $\alpha = 0.7$ and one of the RCK trajectories does, consistently with the values obtained above. Figure 6 shows the sample CDF of $W^{\min}$ for the Kelly bet and the RCK bet with $\lambda = 6.456$, and the theoretical bound given by α^{λ}.

7.2.2   Risk-growth trade-off of RCK, QRCK, and fractional Kelly bets

We compare the Monte Carlo simulated drawdown risk ($\alpha = 0.7$) and expected growth rate of the Kelly, RCK, QRCK, and fractional Kelly bets. We select multiple values of λ (for RCK and QRCK) and f (for fractional Kelly) and plot them together in figure 7. We observe, as we did in the finite outcome case, that RCK yields superior bets than QRCK. This is particularly significant since QRCK is closely connected to the classic Markowitz portfolio optimization model, and this example resembles a financial portfolio selection problem (a bet with infinite return distributions). The fractional Kelly bets, in this case, show instead the (essentially) same performance as RCK.
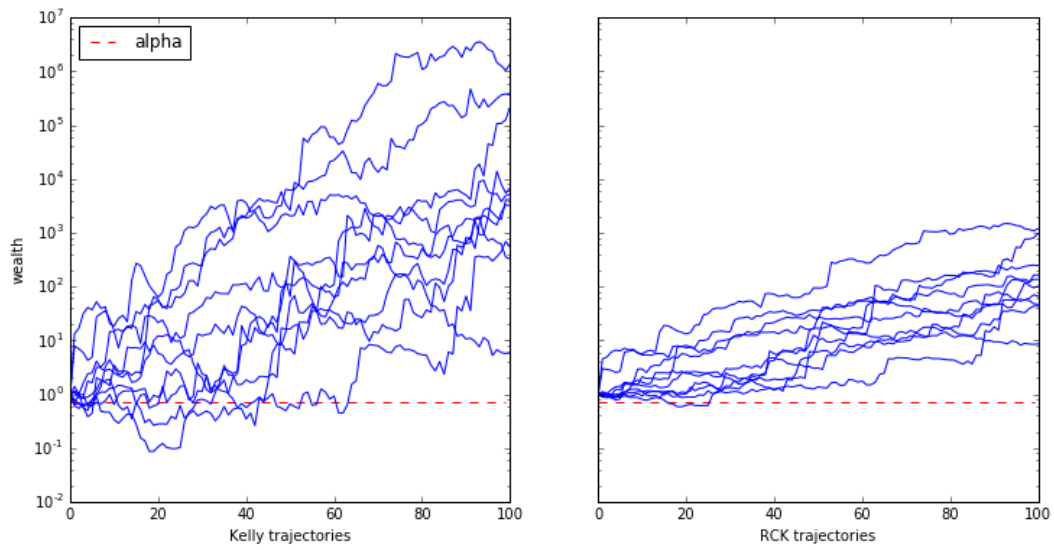
Figure 5: Wealth trajectories for the Kelly optimal bet (left) and the RCK bet with $\lambda = 5.700$, for the infinite returns distribution. The wealth threshold $\alpha = 0.7$ is also shown.
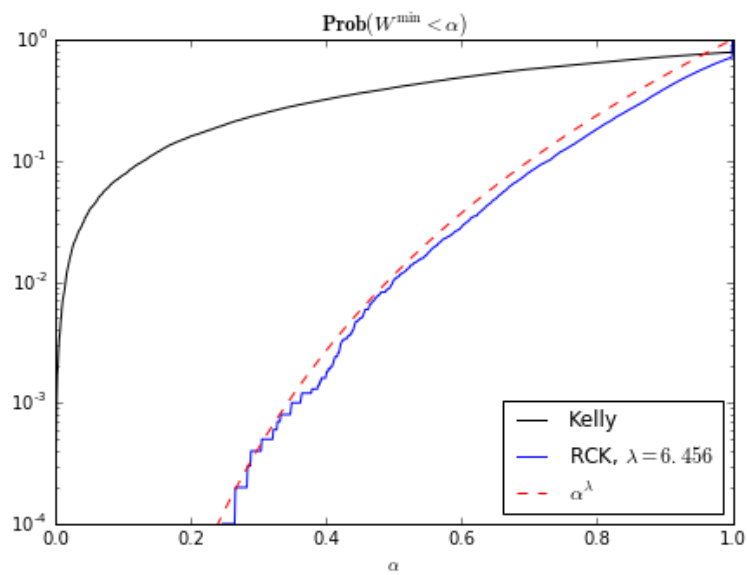
Figure 6: Sample CDF of $W^{\min}$ for the Kelly optimal bet and RCK bet with $\lambda = 6.46$, for the infinite returns distribution. The bound α^λ is also shown.
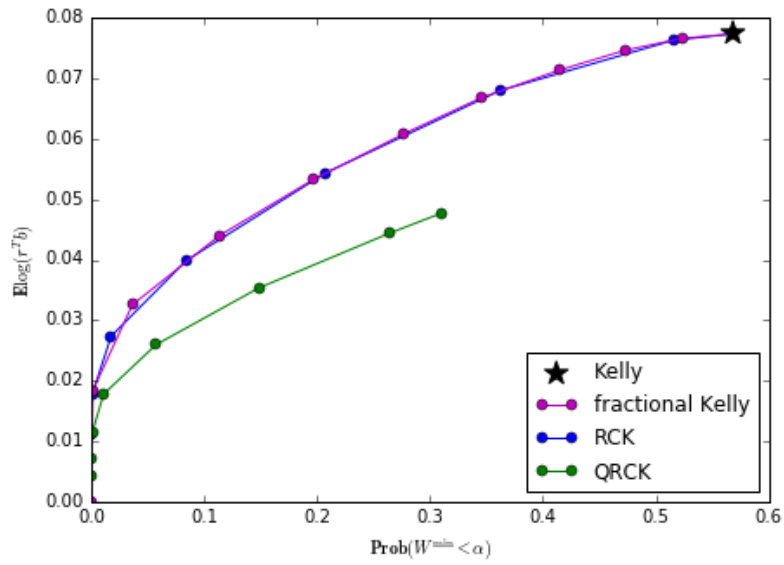
Figure 7: Trade-off of drawdown risk and expected growth rate for the RCK, QRCK, and fractional Kelly bets, with the infinite return distribution. The Kelly bet is also shown.

A Miscellaneous lemmas

Here we collect the technical details and derivations of several results from the text.

Lemma 1. *The bet e_n is optimal for problem* (1) *if and only if* $\mathbf{E}r_i \leq 1$ *for* $i = 1, \ldots, n-1$. *Likewise, the bet e_n is optimal for problem* (7) *if and only if* $\mathbf{E}r_i \leq 1$ *for* $i = 1, \ldots, n-1$.

Proof. Assume $\mathbf{E}r_i \leq 1$ for $i = 1, \ldots, n-1$. Then by Jensen's inequality,

$$\mathbf{E}\log(r^T b) \leq \log \mathbf{E}r^T b \leq 0$$

for any $b \in \{b \mid \mathbf{1}^T b = 1, \ b \geq 0\}$. So e_n, which achieves objective value 0, is optimal for both problem (1) and problem (7).

Next assume $\mathbf{E}r_i > 1$ for some i. Consider the bet $b = fe_i + (1-f)e_n$ and

$$\phi(f) = \mathbf{E}\log(r^T b)$$

for $f \in [0, 1)$. By [Ber73, Proposition 2.1] we have the right-sided derivative

$$\partial_+ \phi(f) = \mathbf{E}\frac{r(e_i - e_n)}{r^T b},$$

and we have

$$\partial_+ \phi(0) = \mathbf{E}r(e_i - e_n) = \mathbf{E}r_i - 1 > 0.$$

Since $\phi(0) = 0$, we have $\phi(f) > 0$ for a small enough f. So $b = fe_i + (1-f)e_n$ has better objective value than e_n for small enough f, i.e., e_n is not optimal for problem (1).

Again, assume $\mathbf{E}r_i > 1$ for some i. Write $b^\star$ for the Kelly optimal bet. (We have already established that $b^\star \neq e_n$ and that $\mathbf{E}\log(r^T b^\star) > 0$.) Consider the bet $\tilde{b} = fb^\star + (1-f)e_n$ and

$$\psi(f) = \mathbf{E}(r^T \tilde{b})^{-\lambda}$$

for $f \in [0, 1)$. By a similar argument as before, we have

$$\partial_+ \psi(f) = -\lambda \mathbf{E}(r^T \tilde{b})^{-\lambda-1} r^T (b^\star - e_n),$$

and

$$\partial_+ \psi(0) = -\lambda \left(\mathbf{E}(r^T b^\star) - 1 \right).$$

By Jensen's inequality we have

$$0 \leq \mathbf{E}\log(r^T b^\star) \leq \log \mathbf{E}r^T b^\star.$$

So $\partial_+ \psi(0) < 0$. Since $\psi(0) = 1$, we have $\psi(f) < 1$ for small enough f. So $\tilde{b} = fb^\star + (1-f)e_n$ is feasible for small enough f. Furthermore, $\tilde{b}$ has strictly better objective value than e_n for $f \in (0, 1)$ since the objective is concave. So e_n is not optimal. $\square$

Lemma 2. *Problem* (7) *always has an optimal dual variable.*

Proof. Assume e_n is optimal. By Lemma 1, e_n is optimal even without the constraint $\mathbf{E}(r^T b)^{-\lambda} \le 1$. So $\kappa^\star = 0$ is an optimal dual variable.

Now assume e_n is not optimal. By the reasoning with the function ψ of Lemma 1's proof, there is a point strictly feasible with respect to the constraint $\mathbf{E}(r^T b)^{-\lambda} \le 1$. So by [Roc74, Theorem 17], we conclude a dual solution exists. $\qquad\square$

Lemma 3. *Let $\varepsilon \in [0,1]$, and write $\Pi(\cdot)$ for the projection onto Δ_ε. Define the function $(\cdot)_{+,\varepsilon} : \mathbf{R}^n \to \mathbf{R}^n$ as $((x)_{+,\varepsilon})_i = \max\{x_i, 0\}$ for $i = 1, \ldots, n-1$ and $((x)_{+,\varepsilon})_n = \max\{x_n, \varepsilon\}$. Then*

$$\Pi(z) = (z - \nu\mathbf{1})_{+,\varepsilon},$$

where ν is a solution of the equation

$$\mathbf{1}^T (z - \nu\mathbf{1})_{+,\varepsilon} = 1.$$

The left-hand side is a nonincreasing function of ν, so ν can be obtained efficiently via bisection on with the starting interval $[\max_i z_i - 1, \max_i z_i]$.

Proof. By definition of the projection, $x = \Pi(z)$ is the solution of

$$\begin{aligned}
\text{minimize} \quad & \tfrac{1}{2}\|x - z\|_2^2 \\
\text{subject to} \quad & \mathbf{1}^T x = 1 \\
& x \ge 0, \quad x_n \ge \varepsilon.
\end{aligned}$$

This problem is equivalent to

$$\begin{aligned}
\text{minimize} \quad & \tfrac{1}{2}\|x - z\|_2^2 + \nu(\mathbf{1}^T x - 1) \\
\text{subject to} \quad & x \ge 0, \quad x_n \ge \varepsilon
\end{aligned}$$

for an optimal dual variable $\nu \in \mathbf{R}$. This follows from dualizing with respect to the constraint $\mathbf{1}^T x = 1$ (and not the others), applying strong Lagrange duality, and using fact that the objective is strictly convex [Ber09]. This problem is in turn equivalent to

$$\begin{aligned}
\text{minimize} \quad & \tfrac{1}{2}\|x - (z - \nu\mathbf{1})\|_2^2 \\
\text{subject to} \quad & x \ge 0, \quad x_n \ge \varepsilon,
\end{aligned}$$

which has the analytic solution

$$x^\star = (z - \nu\mathbf{1})_{+,\varepsilon}.$$

An optimal dual variable ν must satisfy

$$\mathbf{1}^T (z - \nu\mathbf{1})_{+,\varepsilon} = 1,$$

by the KKT conditions. Write $h(\nu) = (z - \nu\mathbf{1})_{+,\varepsilon}$. Then $h(\nu)$ a continuous nonincreasing function with

$$h(\max_i z_i - 1) \ge 1, \quad h(\max_i z_i) = \varepsilon.$$

So a solution of $h(\nu) = 1$ is in the interval $[\max_i z_i - 1, \max_i z_i]$. $\qquad\square$

Lemma 4. *A pair* $(b^\star, \kappa^\star)$ *is a solution of the RCK problem if and only if it satisfies conditions* (13).

Proof. Assume $(b^\star, \kappa^\star)$ is a solution of the RCK problem. This immediately gives us $\mathbf{1}^T b^\star = 1$, $b^\star \geq 0$, $\mathbf{E}(r^T b^\star)^{-\lambda} \leq 1$, and $\kappa^\star \geq 0$. Moreover $b^\star$ is a solution to

$$
\begin{array}{ll}
\text{minimize} & -\mathbf{E}\log(r^T b) + \kappa^\star \mathbf{E}(r^T b)^{-\lambda} \\
\text{subject to} & \mathbf{1}^T b = 1, \quad b \geq 0
\end{array}
\tag{18}
$$

and $\kappa^\star(\mathbf{E}(r^T b^\star)^{-\lambda} - 1) = 0$, by Lagrange duality.

Consider $b^\varepsilon = (1 - \varepsilon)b^\star + \varepsilon b$, where $b \in \mathbf{R}^n$ satisfies $\mathbf{1}^T b = 1$ and $b \geq 0$. Define $\varphi(\varepsilon) = -\mathbf{E}\log(r^T b^\varepsilon) + \kappa^\star \mathbf{E}(r^T b^\varepsilon)^{-\lambda}$. Since b^ε is feasible for problem (18) for $\varepsilon \in [0, 1)$ and since $b^\star$ is optimal, we have $\varphi(0) \leq \varphi(\varepsilon)$ for $\varepsilon \in [0, 1)$. We later show $\varphi(\varepsilon) < \infty$ for $\varepsilon \in [0, 1)$. So $\partial_+ \varphi(0) \geq 0$, where ∂_+ denotes the right-sided derivative.

By [Ber73, Proposition 2.1] we have

$$
\partial_+ \varphi(\varepsilon) = -\mathbf{E}\frac{r^T(b - b^\star)}{r^T b^\varepsilon} - \kappa^\star \lambda \mathbf{E}\frac{r^T(b - b^\star)}{(r^T b^\varepsilon)^{\lambda+1}}.
$$

So we have

$$
0 \leq \partial_+ \varphi(0) = 1 + \kappa^\star \lambda \mathbf{E}(r^T b^\star)^{-\lambda} - \mathbf{E}\frac{r^T b}{r^T b^\star} - \kappa^\star \lambda \mathbf{E}\frac{r^T b}{(r^T b^\star)^{\lambda+1}}.
$$

Using $\kappa^\star(\mathbf{E}(r^T b^\star)^{-\lambda} - 1) = 0$, and reorganizing we get

$$
\left(\mathbf{E}\frac{r}{r^T b^\star} + \kappa^\star \lambda \mathbf{E}\frac{r}{(r^T b^\star)^{\lambda+1}} \right)^T b \leq 1 + \kappa^\star \lambda,
\tag{19}
$$

for any $b \in \mathbf{R}^n$ that satisfies $\mathbf{1}^T b = 1$ and $b \geq 0$. With $b = e_i$ for $i = 1, \ldots, n$, we get

$$
\mathbf{E}\frac{r_i}{r^T b^\star} + \kappa^\star \lambda \mathbf{E}\frac{r_i}{(r^T b^\star)^{\lambda+1}} \leq 1 + \kappa^\star \lambda,
\tag{20}
$$

for $i = 1, \ldots, n$.

Now assume $b_i^\star > 0$ for some i and let $b = e_i$. Then b^ε is feasible for problem (18) for small negative ε. We later show $\varphi(\varepsilon) < \infty$ for small negative ε. This means $\partial_+ \varphi(0) = 0$ in this case, and we conclude

$$
\mathbf{E}\frac{r_i}{r^T b^\star} + \kappa^\star \lambda \mathbf{E}\frac{r_i}{(r^T b^\star)^{\lambda+1}} \begin{cases} \leq 1 + \kappa^\star \lambda & b_i = 0 \\ = 1 + \kappa^\star \lambda & b_i > 0. \end{cases}
$$

It remains to show that $\varphi(\varepsilon) < \infty$ for appropriate values of ε. First note that $b^\star$, a solution, has finite objective value, *i.e.*, $-\mathbf{E}\log(r^T b^\star) < \infty$ and $\mathbf{E}(r^T b^\star)^{-\lambda} < \infty$. So for $\varepsilon \in [0, 1)$, we have

$$
\begin{aligned}
\varphi(\varepsilon) &= -\mathbf{E}\log((1 - \varepsilon)r^T b^\star + \varepsilon r^T b) + \kappa^\star \mathbf{E}((1 - \varepsilon)r^T b^\star + \varepsilon r^T b)^{-\lambda} \\
&\leq -\mathbf{E}\log((1 - \varepsilon)r^T b^\star) + \kappa^\star \mathbf{E}((1 - \varepsilon)r^T b^\star)^{-\lambda} \\
&= -\log(1 - \varepsilon) - \mathbf{E}\log(r^T b^\star) + \kappa^\star(1 - \varepsilon)^{-\lambda}\mathbf{E}(r^T b^\star)^{-\lambda} < \infty.
\end{aligned}
$$

Next assume $b_i > 0$ for some $i \in \{1, \ldots, n\}$. Then for $\varepsilon \in (-b_i, 0)$ we have

$$
\begin{aligned}
\varphi(\varepsilon) &\leq -\mathbf{E}\log(r^T b^\star + \varepsilon r_i) + \kappa^\star \mathbf{E}(r^T b^\star + \varepsilon r_i)^{-\lambda} \\
&\leq -\mathbf{E}\log(((b_i + \varepsilon)/b_i) r^T b^\star) + \kappa^\star \mathbf{E}(((b_i + \varepsilon)/b_i) r^T b^\star)^{-\lambda} \\
&= -\log((b_i + \varepsilon)/b_i) - \mathbf{E}\log(r^T b^\star) + \kappa^\star((b_i + \varepsilon)/b_i)^{-\lambda} \mathbf{E}(r^T b^\star)^{-\lambda} < \infty.
\end{aligned}
$$

Finally, assume conditions (13) for $(b^\star, \kappa^\star)$, and let us go through the argument in reverse order to show the converse. Conditions (13) implies condition (20), which in turn implies condition (19) as $\mathbf{1}^T b = 1$. Note $b^\star$ has finite objective value becuase $\mathbf{E}(r^T b^\star)^{-\lambda} < \infty$ by assumption and $-\mathbf{E}\log(r^T b^\star) \leq (1/\lambda)\log \mathbf{E}(r^T b^\star)^{-\lambda} < \infty$ by Jensen's inequality. So $\varphi(0) \leq \varphi(\varepsilon)$ for $\varepsilon \in [0, 1)$ by the same argument as before, *i.e.*, $b^\star$ is optimal for problem (18). This fact, together with conditions (13), give us the KKT conditions of the RCK problem, and we conclude $(b^\star, \kappa^\star)$ is a solution of the RCK problem by Lagrange duality [Ber09]. $\qquad\square$

Lemma 5. *Consider an IID sequence $X_1, X_2, \ldots$ from probability measure μ, its random walk $S_n = X_1 + X_2 + \cdots + X_n$, a stopping time τ, and*

$$
\psi(\lambda) = \log \mathbf{E}\exp(-\lambda X) = \log \int \exp(-\lambda x)\, d\mu(x).
$$

Then we have

$$
\mathbf{E}\left[\exp(-\lambda S_\tau - \tau\psi(\lambda)) \mid \tau < \infty\right]\mathbf{Prob}(\tau < \infty) \leq 1.
$$

This lemma is a modification of an identity from [Wal44], [Fel71, §XVIII.2], and [Gal13, §9.4].

Proof. Consider the tilted probability measure

$$
d\mu_\lambda(x) = \exp(-\lambda x - \psi(\lambda))d\mu(x)
$$

and write $\mathbf{Prob}_{\mu_\lambda}$ for the probability under the tilted measure μ_λ. Then we have

$$
\begin{aligned}
\mathbf{Prob}_{\mu_\lambda}(\tau = n) &= \int I_{\{\tau = n\}} d\mu_\lambda(x_1, x_2, \ldots, x_n) \\
&= \int I_{\{\tau = n\}} \exp(-\lambda s_n - n\psi(\lambda)) d\mu(x_1, x_2, \ldots, x_n) \\
&= \mathbf{E}\left[\exp(-\lambda S_\tau - \tau\psi(\lambda)) I_{\{\tau = n\}}\right]
\end{aligned}
$$

By summing through $n = 1, 2, \ldots$ we get

$$
\begin{aligned}
\mathbf{Prob}_{\mu_\lambda}(\tau < \infty) &= \mathbf{E}\left[\exp(-\lambda S_\tau - \tau\psi(\lambda)) I_{\{\tau < \infty\}}\right] \\
&= \mathbf{E}\left[\exp(-\lambda S_\tau - \tau\psi(\lambda)) \mid \tau < \infty\right]\mathbf{Prob}(\tau < \infty).
\end{aligned}
$$

Since $\mathbf{Prob}_{\mu_\lambda}(\tau < \infty) \leq 1$, we have the desired result. $\qquad\square$

B DCP specification

The finite outcome RCK problem (11) can be formulated and solved in CVXPY as

```
b = Variable(n)
lambda_risk = Parameter(sign = 'positive')
growth = pi.T*log(r.T*b)
risk_constraint = (log_sum_exp (log(pi) - lambda_risk * log(r.T*b)) <= 0)
constraints = [ sum_entries(b) == 1, b >= 0, risk_constraint ]
risk_constr_kelly = Problem(Maximize(growth),constraints)
risk_constr_kelly.solve()
```

Here `r` is the matrix whose columns are the return vectors, and `pi` is the vector of probabilities. The second to last line forms the problem (object), and in the last line the problem is solved. The optimal bet is written into `b.value`. We note that if we set `lambda_risk` equal to 0 this problem formulation is equivalent (computationally) to the Kelly problem.

Acknowledgments

References

[Ber73] D. P. Bertsekas. Stochastic optimization problems with nondifferentiable cost functionals. *Journal of Optimization Theory and Applications*, 12(2):218–231, 1973.

[Ber09] D. P. Bertsekas. *Convex Optimization Theory*. Athena Scientific, 2009.

[Bro00] S. Browne. Risk-constrained dynamic active portfolio management. *Management Science*, 46(9):1188–1199, 2000.

[Bub15] S. Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends in Machine Learning*, 8(3–4):231–357, 2015.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[CT12] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012.

[DB16] S. Diamond and S. Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 2016.

[DCB13] A. Domahidi, E. Chu, and S. Boyd. ECOS: An SOCP solver for embedded systems. In *European Control Conference (ECC)*, pages 3071–3076, 2013.

[DL12]   M. Davis and S. Lleo. Fractional Kelly strategies in continuous time: Recent developments. In L. C. MacLean and W. T. Ziemba, editors, *Handbook of the Fundamentals of Financial Decision Making*, pages 753–788. World Scientific Publishing, 2012.

[Fel71]   W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, 2nd edition, 1971.

[Gal13]   R. G. Gallager. *Stochastic Processes: Theory for Applications*. Cambridge University Press, 2013.

[GB14]   M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

[GBY06]   M. Grant, S. Boyd, and Y. Ye. Disciplined convex programming. In L. Liberti and N. Maculan, editors, *Global Optimization: From Theory to Implementation*, Nonconvex Optimization and its Applications, pages 155–210. Springer, 2006.

[Kel56]   J. Kelly, Jr. A new interpretation of information rate. *IRE Transactions on Information Theory*, 2(3):185–189, 1956.

[KY03]   H. J. Kushner and G. G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, 2003.

[Lö04]   J. Löfberg. YALMIP: A toolbox for modeling and optimization in MATLAB. In *Proceedings of the IEEE International Symposium on Computer Aided Control Systems Design*, pages 284–289, 2004.

[Mar52]   H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[Mer90]   R. C. Merton. *Continuous-Time Finance*. Wiley-Blackwell, 1990.

[MTZ11]   L. C. MacLean, E. O. Thorp, and W. T. Ziemba. *The Kelly Capital Growth Investment Criterion: Theory and Practice*, volume 3. World Scientific Publishing, 2011.

[NJLS09]   A. S. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.

[NY78]   A. S. Nemirovski and D. B. Yudin. On Cesari's convergence of the steepest descent method for approximating saddle points of convex-concave functions. *Doklady Akademii Nauk SSSR*, 239:1056–1059, 1978.

[NY83]   A. S. Nemirovski and D. B. Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley, 1983.

[Pol87]   B. T. Polyak. *Introduction to Optimization.* Optimization Software, New York, 1987.

[RM51]   H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, 1951.

[Roc74]   R. T. Rockafellar. *Conjugate Duality and Optimization.* Society for Industrial and Applied Mathematics, Philadelphia, 1974.

[Ser15]   S. A. Serrano. *Algorithms for Unsymmetric Cone Optimization and an Implementation for Problems with the Exponential Cone.* PhD thesis, Stanford University, 2015.

[UMZ$^+$14] M. Udell, K. Mohan, D. Zeng, J. Hong, S. Diamond, and S. Boyd. Convex optimization in Julia. *SC14 Workshop on High Performance Technical Computing in Dynamic Languages*, 2014.

[Wal44]   A. Wald. On cumulative sums of random variables. *The Annals of Mathematical Statistics*, 15(3):283–296, 1944.

[Whi81]   P. Whittle. Risk-sensitive linear/quadratic/Gaussian control. *Advances in Applied Probability*, 13(4):764–777, 1981.

[Whi90]   P. Whittle. *Risk-sensitive Optimal Control.* Wiley, 1990.