Portfolios that Contain Risky Assets 13: Law of Large Numbers (Kelly) Objectives

C. David Levermore

University of Maryland, College Park, MD

Math 420: *Mathematical Modeling*
April 22, 2018 version

Portfolios that Contain Risky Assets
Part II: Stochastic Models

Law of Large Numbers (Kelly) Objectives

1. Introduction
2. Law of Large Numbers
3. Kelly Criterion for a Simple Game
4. Kelly Criterion in Practice

Introduction

An IID model for the Markowitz portfolio with allocation **f** satisfies

$$\mathrm{Ex}\left(\log\left(\frac{\pi(d)}{\pi(0)}\right)\right) = d\,\gamma\,, \qquad \mathrm{Var}\left(\log\left(\frac{\pi(d)}{\pi(0)}\right)\right) = d\,\theta\,,$$

where γ and θ are the growth rate mean and variance. These are estimated from a share price history, for example by

$$\hat{\gamma} = \hat{\mu} - \tfrac{1}{2}\big(\hat{\mu}^2 + \hat{\xi}\big)\,, \quad \hat{\theta} = \hat{\xi}\,,$$

where $\hat{\mu}$ and $\hat{\xi}$ are the return mean and variance estimators

$$\hat{\mu} = \mu_{\mathrm{rf}}\left(1 - \mathbf{1}^{\mathrm{T}}\mathbf{f}\right) + \mathbf{m}^{\mathrm{T}}\mathbf{f}\,, \qquad \hat{\xi} = \frac{\mathbf{f}^{\mathrm{T}}\mathbf{V}\mathbf{f}}{1 - \bar{w}}\,.$$

Introduction

Our general approach to portfolio management will be to select an allocation $\mathbf{f}$ that maximizes some objective function. The IID model suggests that we might want to pick $\mathbf{f}$ to maximize γ. Below we will show that an important tool from probability, the *Law of Large Numbers*, confirms that this strategy is ideal. However, a difficulty with using this strategy is that we do not know γ. Rather, we will develop strategies that maximize one of a family objective functions that are built from $\hat{\gamma}$ and $\hat{\theta}$.

In 1956 John Kelly, a colleague of Claude Shannon at Bell Labs, used the Law of Large Numbers to devise optimal betting strategies for a class of games of chance. A strategy that tries to maximize γ became known as the *Kelly criterion, Kelly strategy*, or *Kelly bet*. In practice they employed modifications of the Kelly criterion.

Introduction

Such strategies were subsequently adopted by Claude Shannon, Ed Thorp, and others to win at blackjack, roullette, and other casino games. These exploits are documented in Ed Thorpe's 1962 book *Beat the Dealer*. Because many casinos were controlled by organized crime at that time, using these strategies could adversely affect the user's health.

Claude Shannon, Ed Thorp, and others soon realized that it was better for both their health and their wealth to apply the Kelly criterion to winning on Wall Street. Ed Thorpe laid out a strategy to do this in his 1967 book *Beat the Market*. He went on to run the first quantiative hedge fund, Princeton Newport Partners, which introduced statistical arbitrage strategies to Wall Street. This history is told in Scott Peterson's 2010 book *The Quants* and in Ed Thorp's 2017 book *A Man for All Markets.*

Law of Large Numbers

Let $\{X(d)\}_{d=1}^{\infty}$ be any sequence of IID random variables drawn from a probability density $p(X)$ with mean γ and variance $\theta > 0$. Let $\{Y(d)\}_{d=1}^{\infty}$ be the sequence of random variables defined by

$$Y(d) = \frac{1}{d} \sum_{d'=1}^{d} X(d') \qquad \text{for every } d = 1, \cdots, \infty .$$

It is easy to check that

$$\mathrm{Ex}(Y(d)) = \gamma , \qquad \mathrm{Var}(Y(d)) = \frac{\theta}{d} .$$

Given any $\delta > 0$ the *Law of Large Numbers* states that

$$\lim_{d \to \infty} \mathrm{Pr}\{|Y(d) - \gamma| \geq \delta\} = 0 . \qquad (2.1)$$

Law of Large Numbers

This limit is not uniform in δ. Its convergence rate can be estimated by the *Chebyshev inequality*, which yields the (not uniform in δ) upper bound

$$\Pr\{|Y(d)-\gamma| \geq \delta\} \leq \frac{\mathrm{Var}(Y(d))}{\delta^2} = \frac{1}{\delta^2}\,\frac{\theta}{d}\,. \tag{2.2}$$

Remark. The Chebyshev inequality is easy to derive. Suppose that $p_d(Y)$ is the (unknown) probability density for $Y(d)$. Then

$$\begin{aligned}\Pr\{|Y(d)-\gamma| \geq \delta\} &= \int_{\{Y\,:\,|Y-\gamma|\geq\delta\}} p_d(Y)\,\mathrm{d}Y \\ &\leq \int_{\{Y\,:\,|Y-\gamma|\geq\delta\}} \frac{|Y-\gamma|^2}{\delta^2}\, p_d(Y)\,\mathrm{d}Y \\ &\leq \frac{1}{\delta^2}\int |Y-\gamma|^2\, p_d(Y)\,\mathrm{d}Y = \frac{\mathrm{Var}(Y(d))}{\delta^2} = \frac{1}{\delta^2}\,\frac{\theta}{d}\,.\end{aligned}$$

Law of Large Numbers

Remark. The unknown probability density $p_d(Y)$ can be expressed in terms of the unknown probability density $p(X)$ as

$$p_d(Y) = \int \cdots \int \delta\left(Y - \frac{1}{d} \sum_{d'=1}^{d} X_{d'} \right) p(X_1) \cdots p(X_d) \, \mathrm{d}X_1 \cdots \mathrm{d}X_d \,,$$

where $\delta(\,\cdot\,)$ is the Dirac delta distribution introduced earlier.

Law of Large Numbers

If $\{X(d)\}_{d=1}^{D}$ is any sequence of IID random variables drawn from an unknown probability density $p(X)$ with unknown mean γ and variance θ then γ and θ have the unbiased estimators $\widehat{\gamma}$ and $\widehat{\theta}$ given by

$$\widehat{\gamma} = \frac{1}{D}\sum_{d=1}^{D} X(d)\,, \qquad \widehat{\theta} = \frac{1}{D-1}\sum_{d=1}^{D}\left(X(d)-\widehat{\gamma}\right)^2 .$$

The law of large numbers (2.1) states that the estimator $\widehat{\gamma}$ will converge to γ as $D \to \infty$. However, in practice D will be finite. The Chebyshev bound (2.2) can be used to assess the quality of the estimator $\widehat{\gamma}$ for finite D.

When $\gamma > 0$ the relative error of the estimate $\widehat{\gamma}$ is

$$\frac{|\widehat{\gamma}-\gamma|}{\gamma} .$$

We would like to know how big D should be to insure that this relative error is less than some $\eta \in (0,1)$ with a certain confidence.

Law of Large Numbers

By setting $\delta = \eta\gamma$ in the Chebyshev bound (2.2) we obtain

$$\Pr\left\{ \frac{|\widehat{\gamma} - \gamma|}{\gamma} \geq \eta \right\} \leq \frac{1}{\eta^2} \frac{\theta}{\gamma^2} \frac{1}{D} .$$

We then replace θ and γ on the right-hand side by $\widehat{\theta}$ and $\widehat{\gamma}$ and pick D large enough to achieve the desired confidence.

For example, if we want to know γ to within 20% with a confidence of 90% then we set $\eta = \frac{1}{5}$ and pick D so large that

$$25 \frac{\widehat{\theta}}{\widehat{\gamma}^2} \frac{1}{D} \leq \frac{1}{10} .$$

Because there are about 250 trading days in a year, this shows that we must average $X(d)$ over a period of $\widehat{\theta}/\widehat{\gamma}^2$ years before we can know γ that well with this much confidence. In practice $\widehat{\theta}/\widehat{\gamma}^2$ is not small.

Kelly Criterion for a Simple Game

Before showing how the Kelly criterion is applied to balancing portfolios with risky assets, we will show how it is applied to a simple betting game.

Consider a game in which each time that we place a bet:

(i) the probability of winning is $p \in (0, 1)$,

(ii) the probability of losing is $q = 1 - p$,

(iii) when we win there is a positive return r on our bet.

We start with a bankroll of cash and the game ends when the bankroll is gone. Suppose that you know p and r. We would like answers to the following questions.

1. When should we play?,
2. When we do play, what fraction of our bankroll should we bet?,

Kelly Criterion for a Simple Game

The game is clearly an IID process. Because each time we play we are faced with the same questions and will have no addtional helpful information, the answers will be the same each time. Therefore we only consider strategies in which we bet a fixed fraction f of our bankroll. If $f = 0$ then we are not betting. If $f = 1$ then we are betting out entire bankroll. (This is clearly a foolish strategy in the long run because we will go broke the first time we lose.) Then

$$\begin{aligned} &\text{when we win our bankroll increases by a factor of } 1 + fr\,, \\ &\text{when we lose our bankroll decreases by a factor of } 1 - f\,. \end{aligned}$$

Therefore if we bet n times and win m times (hence, lose $n - m$ times) then our bankroll changes by a factor of

$$(1 + fr)^m (1 - f)^{n-m}\,.$$

The Kelly criterion is to pick $f \in [0, 1)$ to maximize this factor for large n.

Kelly Criterion for a Simple Game

This is equivalent to maximizing the log of this factor, which is

$$m \log(1 + fr) + (n - m) \log(1 - f)\,.$$

The law of large numbers implies that

$$\lim_{n \to \infty} \frac{m}{n} = p\,.$$

Therefore for large n we see that

$$\begin{aligned} & m \log(1 + fr) + (n - m) \log(1 - f) \\ & \qquad \sim \big(p \log(1 + fr) + (1 - p) \log(1 - f)\big) n\,. \end{aligned}$$

Hence, the Kelly criterion says that we want to pick $f \in [0, 1)$ to maximize the growth rate

$$\gamma(f) = p \log(1 + fr) + (1 - p) \log(1 - f)\,. \tag{3.3}$$

This is now an exercise from first semester calculus.

Kelly Criterion for a Simple Game

Notice that $\gamma(0) = 0$ and that

$$\lim_{f \nearrow 1} \gamma(f) = -\infty \,.$$

Also notice that for every $f \in [0, 1)$ we have

$$\gamma'(f) = \frac{pr}{1 + fr} - \frac{1 - p}{1 - f} \,,$$
$$\gamma''(f) = -\frac{pr^2}{(1 + fr)^2} - \frac{1 - p}{(1 - f)^2} \,.$$

Because $\gamma''(f) < 0$ over $[0, 1)$, we see that $\gamma(f)$ is strictly concave over $[0, 1)$ and that $\gamma'(f)$ is strictly decreasing over $[0, 1)$.

If $\gamma'(0) = pr - (1 - p) = p(1 + r) - 1 \leq 0$ then $\gamma(f)$ is strictly deceasing over $[0, 1)$ because $\gamma'(f)$ is strictly decreasing over $[0, 1)$. In that case the maximizer for $\gamma(f)$ over $[0, 1)$ is $f = 0$ and the maximun is $\gamma(0) = 0$.

Kelly Criterion for a Simple Game

If $\gamma'(0) = pr - (1 - p) = p(1 + r) - 1 > 0$ then $\gamma(f)$ has a unique maximizer at $f = f_* \in (0, 1)$ that satifies

$$0 = \gamma'(f_*) = \frac{pr}{1 + f_* r} - \frac{1 - p}{1 - f_*}$$
$$= \frac{pr(1 - f_*) - (1 - p)(1 + f_* r)}{(1 + f_* r)(1 - f_*)}$$
$$= \frac{p(1 + r) - f_* r}{(1 + f_* r)(1 - f_*)} .$$

Upon solving this equation for f_* we find that

$$f_* = \frac{p(1 + r) - 1}{r} . \tag{3.4}$$

Kelly Criterion for a Simple Game

Remark. We see from (3.4) that if $p(1+r)-1>0$ then

$$0 < f_* = \frac{p(1+r)-1}{r} = p - \frac{1-p}{r} < p < 1\,.$$

Therefore the Kelly crtierion yields the optimal betting strategy

$$f_* = \begin{cases} 0 & \text{if } p(1+r)-1 \leq 0\,, \\ \dfrac{p(1+r)-1}{r} & \text{if } p(1+r)-1 > 0\,. \end{cases} \tag{3.5}$$

The maximum growth rate (details not shown) when $p(1+r)-1>0$ is

$$\gamma(f_*) = p\log\big(p(1+r)\big) + (1-p)\log\left((1-p)\,\frac{1+r}{r}\right). \tag{3.6}$$

Remark. In practice this strategy is is far from ideal for reasons that we will discuss in the next section.

Kelly Criterion for a Simple Game

Remark. Some bettors call r the *odds* because the return r on a winning wager is usually chosen so that the ratio $r:1$ reflects a probability of winning. The expected return on each amount wagered is $pr-(1-p)$. This is the probability of winning, p, times the return of a win, r, plus the probability of losing, $1-p$, times the return of a loss, -1. Some bettors call this quantity the *edge* when it is postive. Notice that $pr-(1-p)=p(1+r)-1$ is the numerator of f_* given by (3.4), while r is the denominator of f_* given by (3.4). Then strategy (3.5) can be expressed in this language as follows.

1. Do not bet unless we have an edge.
2. If we have an edge then bet $f_* = \dfrac{\text{edge}}{\text{odds}}$ of our bankroll.

This view of the Kelly criterion is popular, but is not very helpful when trying to apply it to more complicated games.

Kelly Criterion in Practice

In most betting games played at casinos the players do not have an edge unless they can use information that is not used by the house when computing the odds. For example, card counting strategies can allow a blackjack player to compute a more accurate probability of winning than the one used by the house when it computed the odds.

Kelly bettors will not make a serious wager until they are very sure that they have an edge, and then they will use the Kelly criterion to size their bet. Because their algorithm yields an approximation of their edge, they are not sure of their true Kelly optimal bet. Because there is a big downside to betting more that the true Kelly optimal bet, their bet is typically a fraction of the Kelly optimal bet.

Kelly Criterion in Practice

We will illustrate these ideas with a modification of the simple game from the last section. Specifically, suppose that the game is the same except for the fact that we are not told p. Rather, we are told that $r = .125$ and that the player won 225 times the last 250 times the game was played.

Based on the information that the player won 225 times the last 250 times the game was played, we guess that $p = .9$. If we use this value of p then we see that

$$p(1 + r) - 1 = .9(1 + .125) - 1 = \tfrac{9}{10} \cdot \tfrac{9}{8} - 1 = \tfrac{1}{80} \,.$$

Based on this calculation, we have an edge, so we will play and the optimal bet is

$$f_* = \frac{p(1 + r) - 1}{r} = \frac{\frac{1}{80}}{\frac{1}{8}} = \frac{1}{10} \,.$$

Therefore the Kelly strategy is to bet $\frac{1}{10}$ of our bankroll each time.

Kelly Criterion in Practice

However, suppose that the previous players had just gotten lucky and that in fact $p = .875$. If we use this value of p then we see that

$$p(1 + r) - 1 = .875(1 + .125) - 1 = \frac{7}{8} \cdot \frac{9}{8} - 1 = -\frac{1}{64} .$$

Therefore we do not have an edge and we should not play!

The difference between .9 and .875 is not large in the sense that it is not an unreasonable error based on only 250 observations. If we bet $\frac{1}{10}$ of our bankroll each time then our bankroll will be significantly diminished before we have played the game enough to realize that there is no edge!

Kelly Criterion in Practice

Now suppose that in fact $p = .895$. If we use this value of p then we see that

$$p(1 + r) - 1 = .895(1 + .125) - 1 = .006875\,.$$

So in fact, we have an edge. However, the optimal bet is

$$f_* = \frac{p(1+r) - 1}{r} = \frac{.006875}{.125} = .055\,.$$

If we bet $\frac{1}{10}$ of our bankroll each time then our bankroll will be significantly diminished before we have played the game enough to realize that p is lower than .9.

Kelly Criterion in Practice

In this game both the edge and the odds are small. Small uncertainties in our estimation of p can lead to large uncertainties in our estimation of f_*. If we overestimate f_* enough then we are almost certain to lose. Betting more than the true f_* is called *overbetting*. If we underestimate f_* then we will certainly win, just a less than the optimal amount.

Because of this asymmetry, it is wise to bet a fraction of the optimal Kelly bet when we are uncertain of our edge. The greater the uncertainty, the smaller the fraction that should be used. Fractions ranging from $\frac{1}{3}$ to $\frac{1}{10}$ are common, depending on the uncertainty. These are called *fractional Kelly strategies*.