Musings on the Kelly Criterion

Arham Habib

March 2025

1 Motivation

Bet sizing is the bridge between having an edge and realizing its value. Even with perfect forecasts, the question of how much to bet — and on what — can dramatically alter long-term outcomes. This write-up explores two key frameworks for sizing bets: the Kelly criterion, which maximizes long-term capital growth, and mean-variance optimization, which trades off return against risk. Though these frameworks arise from different philosophies — one rooted in utility theory, the other in portfolio theory — they ultimately lead to similar structures. Understanding where they overlap, and where they diverge, can provide a clearer lens for thinking about capital allocation in uncertain environments.

2 Binary Bets

2.1 Problem Statement

Let's say that I'm betting on sports. I train a "perfect" model – that is, if the model says a team will win 65% of the time, and we replay that game 10,000 times, they will win around 6,500 of those games. I say "around" because the model is the perfect forecast of the **probability**, not of the outcome. In the context of making predictions, we'd call this "perfect calibration". Over time, anything we say will happen 65% of the time will happen exactly 65% of the time.

Even with this perfect model, there are some open questions to how we want to execute the trades. Let's say we started at $1000, and we could borrow more from the bank any time we wanted. If we were betting against an imperfect sportsbook, we would just bet the $1000 every single time. If we lost, we'd borrow more. With the knowledge that our model is perfect, we'd certainly come out on top in the long run.

Still, we might have to tolerate a large drawdown in interim. Imagine someone is giving you 1:1 odds for a coin that flips heads 55% of the time. Over time, betting on heads is certainly the winning strategy. Still, it is not inconceivable that we'd be negative in the first 10, 100, or 500 bets. That edge might be statistically sound, but when playing a 55%/45% game, there is plenty of unavoidable randomness.

Now, imagine that we only get one grant of $1000. If we lose it all, we can no longer bet and take advantage of our perfect model. How would this change our betting?

The Kelly criterion provides a solution to a fundamental problem: determining the optimal bet size to maximize long-term capital growth. It makes some assumptions, namely, that the investor wants to maximize the logarithmic growth rate of capital. Implicitly, this assumes something called a logarithmic utility function – I won't cover that in detail here, but this has some preferrable qualities because it is concave. In plain English, a concave utility function means that our gambler doesn't bet his entire bankroll every time on 1% edge.

Consider a bet with known odds. Suppose a sportsbook offers a bet at +200 odds, meaning that a $100 wager will return $300 if successful (the original $100 stake plus $200 in winnings). These odds correspond to an implied probability of:

$$\frac{100}{300} = \frac{1}{3}$$

If we believe the true probability of winning is 40%, and we start with a bankroll of $1000, the question is: What fraction of our bankroll should we bet to maximize long-term growth?

2.2 Formalizing the Problem

Define the following variables:

$$\begin{aligned} p &= \text{probability of winning the bet} \\ o &= \text{odds multiple (e.g., 2 for +200)} \\ x &= \text{fraction of bankroll bet} \\ A_t &= \text{bankroll at time } t \end{aligned}$$

Our bankroll after n bets depends on the number of wins w and losses $n-w$, following:

$$A_{t+n} = A_t(1+ox)^w(1-x)^{n-w}$$

We win a bet p% of the time, and lose it (1-p)% of the time. So, if we take the **expected value** of our bankroll after n bets, we get:

$$
\begin{aligned}
E\left[\frac{A_{t+n}}{A_t}\right] &= E\left[(1+ox)^w(1-x)^{n-w}\right] \\
E\left[\frac{A_{t+n}}{A_t}\right] &= (1+ox)^{np}(1-x)^{n(1-p)} \\
E\left[\log\left(\frac{A_{t+n}}{A_t}\right)\right] &= np\log(1+ox) + n(1-p)\log(1-x)
\end{aligned}
$$

2.3 Maximization

We want to maximize the rate at which our capital grows. To maximize the expected logarithmic growth rate, we differentiate with respect to x. The question we're answering is how the expected growth rate changes as we bet more or less of our bankroll. If we bet none of it, it will never grow. If we bet it all and lose, we're also stuck at 0. The maximization of growth will be somewhere in between:

$$\frac{d}{dx}\left[np\log(1+ox) + n(1-p)\log(1-x)\right] = 0$$

Taking the derivatives,

$$np\frac{o}{1+ox} - n(1-p)\frac{1}{1-x} = 0$$

Solving for x,

$$
\begin{aligned}
p\frac{o}{1+ox} &= (1-p)\frac{1}{1-x} \\
po(1-x) &= (1-p)(1+ox) \\
po - pox &= 1 - p + ox - pox \\
po - 1 - p &= -ox + pox \\
po - 1 - p &= x(po - o) \\
x &= \frac{po-(1-p)}{o} \\
x &= \boxed{p - \frac{1-p}{o}}
\end{aligned}
$$

This is the optimal Kelly bet fraction for a binary bet. So, for the example above, if the odds are 2:1 and we think the real probability is 40%, we'd bet 10% of our bankroll:

$$.4 - .6(1/2) = .1$$

In sports betting, odds are often expressed in percentage form as the **book-implied probability**, which we denote as:

$$b = \frac{o}{1+o}$$

where o represents the decimal odds multiple. Our goal is to derive the Kelly criterion bet fraction using b instead of o.

First, solving for o in terms of b:

$$\begin{aligned}
b &= \frac{o}{1+o} \\
b(1+o) &= o \\
b + bo &= o \\
b &= o - bo \\
b &= o(1-b) \\
o &= \frac{1-b}{b}
\end{aligned}$$

The standard Kelly criterion formula for a binary bet is:

$$x = p - \frac{1-p}{o}$$

Substituting $o = \frac{1-b}{b}$:

$$\begin{aligned}
x &= p - \frac{1-p}{\frac{1-b}{b}} \\
x &= p - \left((1-p) \times \frac{b}{1-b}\right) \\
x &= p - \frac{b(1-p)}{1-b} \\
x &= p - \frac{b-pb}{1-b}
\end{aligned}$$

Expanding the numerator:

$$\begin{aligned}
x &= p - \frac{b-pb}{1-b} \\
x &= \frac{p(1-b) - (b-pb)}{1-b} \\
x &= \frac{p - pb - b + pb}{1-b} \\
x &= \boxed{\frac{p-b}{1-b}}
\end{aligned}$$

Again, for the example above, the book odds would be 33% and we might think our true odds are 40%. This formulation would suggest 10% of our bankroll, identical to the answer above:

$$\frac{.4 - 1/3}{2/3} = .1 = 10\%$$

Notice that the bet size scales linearly with the percent difference in our "ground truth" odds p and b. Further, notice that the bet size is also increasing as b increases – if the bet was already likely to hit as implied by the book, and is even more likely to hit with our "ground truth" understanding, we should size up. In other words, bet more when you are 95% sure on 90% lines than when you are 10% sure on 5% lines. Both of these should feel intuitive. We bet more with more edge, and we allocate more of the portfolio when the bet isn't a long shot.

Below is a table of how much to bet on a single, binary outcome based on the "perfect model" versus the book odds. I don't include any rows where the model is equal to or lower than the book odds – Kelly would tell us to not bet in those situations:

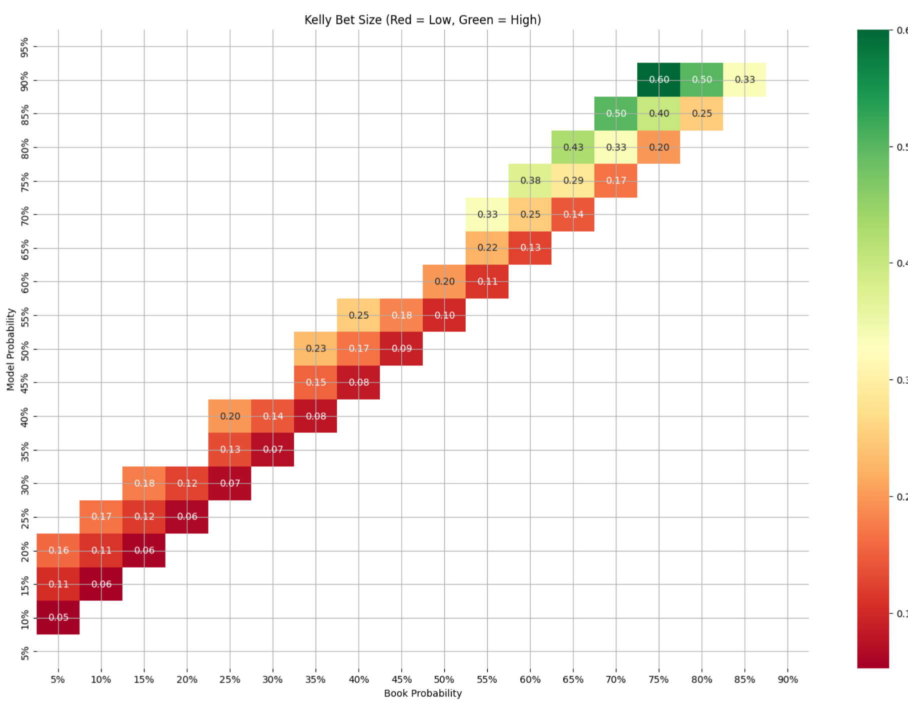

Figure 1: Kelly Criterion Binary Sizing

3 Single Normally Distributed Bet

3.1 Problem Statement

Let's instead say that we've graduated out of our degenerate gambler era in college and started professionally gambling instead at a hedge fund by investing in stocks. Our stocks' return profile is very different from the binary bets. They could go up or down by an arbitrary amount, so the outcome of our bet is not 0 or 1. Still, larger moves are less likely.

In a world where we have perfect estimates on the **distribution** of returns, how might we best grow our portfolio?

3.2 Setup

Suppose a bet's return R is modeled as a normal distribution (to be clear, this is not how stocks behave – it is just a toy example):

$$R \sim \mathcal{N}(\mu, \sigma^2)$$

where:

- $\mu = E[R]$ is the expected return of the bet,
- $\sigma^2 = \mathrm{Var}(R)$ is the variance of the return.

Let x be the fraction of our bankroll we bet. Then, our **log wealth growth** over time is:

$$G = \log(1 + xR).$$

3.3 Expected Growth and Taylor Expansion

To determine the optimal bet fraction x^*, we maximize:

$$E[G] = E[\log(1 + xR)].$$

Since the logarithm function is nonlinear, we approximate it using a second-order Taylor expansion around $x = 0$ (no need to dwell on this – it trades off some precision for making the math more convenient):

$$\log(1 + xR) \approx xR - \frac{1}{2}x^2R^2.$$

Taking the expectation,

$$E[\log(1 + xR)] \approx E[xR - \frac{1}{2}x^2R^2].$$

Using the linearity of expectation,

$$E[\log(1 + xR)] = xE[R] - \frac{1}{2}x^2 E[R^2].$$

Expanding $E[R^2]$ using the variance identity:

$$E[R^2] = \mu^2 + \sigma^2.$$

Thus,

$$E[\log(1 + xR)] = x\mu - \frac{1}{2}x^2(\mu^2 + \sigma^2).$$

3.4 Maximization

Again, we are trying to answer which percentage of portfolio invested will maximize our portfolio's logarithmic growth. To maximize this expression with respect to x, we take the derivative:

$$\frac{d}{dx}\left(x\mu - \frac{1}{2}x^2(\mu^2 + \sigma^2)\right) = \mu - x(\mu^2 + \sigma^2).$$

Setting it to zero,

$$\mu - x(\mu^2 + \sigma^2) = 0.$$

Solving for x,

$$\boxed{x^* = \frac{\mu}{\mu^2 + \sigma^2}.}$$

3.5 Relation to the Sharpe Ratio

Notice that the number above is extremely similar to a Sharpe ratio, except with an additional $\mu^2\sigma$ term in the denominator. This should line up with the intuition that a higher Sharpe trade deserves a higher allocation.

Below is a plot of trades by monthly Sharpe ratio (annualized Sharpe ratios of the same amount will often imply that you should put your entire portfolio into something), because Kelly is supposed to represent a bet at a single point in time, not a bet in a portfolio.

Specifically, the shape of the curve is **almost** linear:

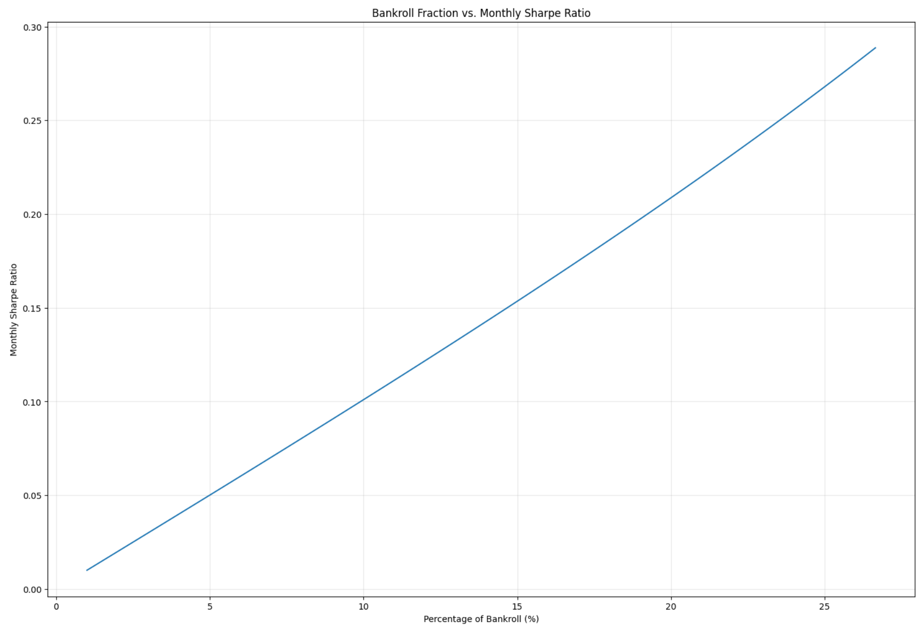

Figure 2: Kelly Criterion Sizing for Normally Distributed Bet

4 Kelly Size Across Many Normally Distributed Bet Opportunities

4.1 Problem Statement

Let's say that now, we have access to multiple great analysts' recommendations each month. So, instead of betting on a single stock, we can bet on mulitple. How does that change the problem?

4.2 Setup

Suppose we have N independent bets available in the same time period, each with normally distributed returns:

Let

$$\mathbf{R} = (R_1, R_2, \ldots, R_N)^\top$$

denote the vector of returns on N assets in one period, each modeled as

$$R_i \sim \mathcal{N}(\mu_i, \sigma_i^2),$$

with mean vector

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N)^\top$$

and covariance matrix

$$\Sigma \quad \text{where} \quad \Sigma_{ij} \;=\; \mathrm{Cov}(R_i,\, R_j).$$

We allocate a fraction x_i of our bankroll to asset i, collected in the vector

$$\mathbf{x} \;=\; (x_1,\, x_2,\, \ldots,\, x_N)^{\top}.$$

Our *wealth multiplier* is approximately

$$1 + \mathbf{x}^{\top}\mathbf{R},$$

so the *log wealth growth* after one period is

$$G(\mathbf{x}) \;=\; \mathbb{E}\big[\log\big(1 + \mathbf{x}^{\top}\mathbf{R}\big)\big].$$

We seek the $\mathbf{x}$ that maximizes $G(\mathbf{x})$.

4.3 Second-Order Taylor Expansion

For small $\mathbf{x}^{\top}\mathbf{R}$, we use the second-order expansion:

$$\log\big(1 + \mathbf{x}^{\top}\mathbf{R}\big) \;\approx\; \mathbf{x}^{\top}\mathbf{R} \;-\; \frac{1}{2}\,\big(\mathbf{x}^{\top}\mathbf{R}\big)^2.$$

Taking expectation,

$$G(\mathbf{x}) \;=\; \mathbb{E}\big[\log\big(1 + \mathbf{x}^{\top}\mathbf{R}\big)\big] \;\approx\; \mathbb{E}\big[\mathbf{x}^{\top}\mathbf{R}\big] \;-\; \frac{1}{2}\,\mathbb{E}\big[(\mathbf{x}^{\top}\mathbf{R})^2\big].$$

Since $\mathbf{x}^{\top}\mathbf{R}$ is a linear combination of the R_i,

$$\mathbb{E}\big[\mathbf{x}^{\top}\mathbf{R}\big] \;=\; \mathbf{x}^{\top}\boldsymbol{\mu}.$$

Second Moment. The term $\mathbb{E}[(\mathbf{x}^{\top}\mathbf{R})^2]$ expands to

$$\mathbb{E}\Big[\sum_{i,j} x_i R_i\, x_j R_j\Big] \;=\; \sum_{i,j} x_i x_j\, \mathbb{E}[R_i R_j].$$

But

$$\mathbb{E}[R_i R_j] \;=\; \mathrm{Cov}(R_i, R_j) + \mathbb{E}[R_i]\,\mathbb{E}[R_j] \;=\; \Sigma_{ij} + \mu_i\,\mu_j.$$

Hence

$$\mathbb{E}\big[(\mathbf{x}^{\top}\mathbf{R})^2\big] \;=\; \mathbf{x}^{\top}\Sigma\,\mathbf{x} \;+\; (\mathbf{x}^{\top}\boldsymbol{\mu})^2 \;=\; \mathbf{x}^{\top}\Sigma\,\mathbf{x} \;+\; \mathbf{x}^{\top}(\boldsymbol{\mu}\,\boldsymbol{\mu}^{\top})\,\mathbf{x}.$$

Therefore, the approximate objective is

$$G(\mathbf{x}) \;\approx\; \mathbf{x}^{\top}\boldsymbol{\mu} \;-\; \frac{1}{2}\Big[\mathbf{x}^{\top}\Sigma\,\mathbf{x} \;+\; \mathbf{x}^{\top}(\boldsymbol{\mu}\,\boldsymbol{\mu}^{\top})\,\mathbf{x}\Big] \;=\; \mathbf{x}^{\top}\boldsymbol{\mu} \;-\; \frac{1}{2}\,\mathbf{x}^{\top}\big[\Sigma + \boldsymbol{\mu}\,\boldsymbol{\mu}^{\top}\big]\mathbf{x}.$$

4.4 Maximization

To find the maximizing $\mathbf{x}$, we take the gradient w.r.t. $\mathbf{x}$ and set it to zero:

$$\nabla_{\mathbf{x}} G(\mathbf{x}) \;=\; \boldsymbol{\mu} \;-\; \left[\Sigma + \boldsymbol{\mu}\,\boldsymbol{\mu}^\top\right]\mathbf{x} \;=\; 0.$$

Thus the optimum must satisfy

$$\left[\Sigma + \boldsymbol{\mu}\,\boldsymbol{\mu}^\top\right]\mathbf{x}^* \;=\; \boldsymbol{\mu}.$$

Hence

$$\boxed{\mathbf{x}^* \;=\; \left[\Sigma + \boldsymbol{\mu}\,\boldsymbol{\mu}^\top\right]^{-1}\boldsymbol{\mu}.}$$

4.5 Special Case: Independent Assets

If the assets are independent, Σ is diagonal with entries σ_i^2, and $\boldsymbol{\mu}\,\boldsymbol{\mu}^\top$ is also diagonal if we only care about each asset's own mean.[1] In that case,

$$\Sigma + \boldsymbol{\mu}\,\boldsymbol{\mu}^\top$$

is diagonal with element $\sigma_i^2 + \mu_i^2$ in the i-th row and column, so its inverse is also diagonal. Therefore,

$$\boxed{x_i^* \;=\; \frac{\mu_i}{\sigma_i^2 + \mu_i^2} \quad \text{for each } i,}$$

which matches the single-asset formula $x^* = \frac{\mu}{\mu^2+\sigma^2}$ applied independently to each asset.

4.6 Comparison to the "Textbook" MVO Formula

In classical "mean–variance" or simplified log-utility treatments, one often sees the objective $\mathbf{x}^\top\boldsymbol{\mu} - \frac{1}{2}\,\mathbf{x}^\top\Sigma\,\mathbf{x}$ *without* the $\boldsymbol{\mu}\,\boldsymbol{\mu}^\top$ term. That version yields

$$\mathbf{x}^* \;=\; \Sigma^{-1}\,\boldsymbol{\mu}.$$

Mathematically, this omits the μ^2 terms in the second moment expansion. If $\|\boldsymbol{\mu}\|$ is relatively small, ignoring μ^2 may be acceptable. However, to be consistent with the single-asset derivation (where $\mu^2+\sigma^2$ appears in the denominator), one should include the $\boldsymbol{\mu}\,\boldsymbol{\mu}^\top$ matrix to capture the full second moment of returns.

Practically, most bets one will take have a much greater volatility than expected return:

$$\sigma_i >> \mu_i$$

This is certainly consistent with most bets at a hedge fund, or props in a sportsbook. When the standard deviation term of our bet dominates, the inclusion or exclusion of the mean vector will not practically change target sizing.

[1] Technically, $\boldsymbol{\mu}\,\boldsymbol{\mu}^\top$ is always rank-1, but for pure 'coordinate-wise' independence you'd treat cross-terms as negligible if $\mu_i\mu_j$ is small for $i \neq j$.

5 Comparison to Portfolio Construction

For readers familiar with portfolio theory, these results should look similar. That is no coincidence.

The Kelly criterion for multiple bets with normally distributed outcomes closely resembles **mean-variance portfolio optimization**, a fundamental concept in modern portfolio theory (MPT). Both approaches determine optimal allocations based on expected returns and risk, but they differ in their underlying objectives.

5.1 Mean-Variance Portfolio Optimization

In Markowitz portfolio theory, an investor chooses portfolio weights w_i to maximize the **risk-adjusted return**, typically formulated as:

$$\max_{\boldsymbol{w}} \quad \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\lambda}{2} \boldsymbol{w}^\top \Sigma \boldsymbol{w}$$

where:

- $\boldsymbol{w}$ is the vector of asset weights,
- $\boldsymbol{\mu}$ is the vector of expected returns,
- Σ is the covariance matrix of returns,
- λ is a **risk aversion** parameter.

Solving the first-order condition,

$$\boldsymbol{\mu} - \lambda \Sigma \boldsymbol{w} = 0,$$

yields the optimal portfolio weights:

$$\boldsymbol{w}^* = \frac{1}{\lambda} \Sigma^{-1} \boldsymbol{\mu}.$$

5.2 Kelly Criterion vs. Portfolio Optimization

The **Kelly criterion for multiple bets** follows a similar structure:

$$\boldsymbol{x}^* = \Sigma^{-1} \boldsymbol{\mu}.$$

Both use expected returns and the inverse of the covariance matrix to determine allocations. Both penalize the correlation between bets / assets by adjusting for risk exposure. Finally, if bets (or assets) are independent allocations reduce to $x_i^* = \frac{\mu_i}{\sigma_i^2}$, similar to how uncorrelated assets receive independent allocations in portfolio theory.

However, the Kelly criterion maximizes long-term logarithmic growth while portfolio optimization balances return against risk through a risk aversion parameter λ. In other words, Kelly will tell you how much risk to put on to maximize an objective. MPT asks you how much risk you're comfortable with, and optimizes within those constraints. These are both useful tools, but in different scenarios.

Below is a chart of the Kelly-optimal bet size for a portfolio of three assets with equal mean return and standard deviation return.

Specifically, these are three assets that each independently have a Sharpe ratio of around.5 annualized – that is, if taking this bet over a year, the **mean** return is about half of what we expect the **standard deviation** of returns to be. That's a reasonable, but not out of the park, signal. Each asset has the exact same mean and standard deviation of return. The graph visualizes the recommended bet sizes on each asset.

What is the recommended bet size for different **correlations** among these signals? A positively correlated set of bets is where the three bets perform well in similar worlds. For example, a portfolio with three semiconductor stocks might be positively correlated. A portfolio with negative correlation might be one where the outcomes are very unrelated – for example, a bet on who will win the NBA championship (three bets are necessarily mutually exclusive). Kelly will tell us to bet more when the assets are **negatively correlated**, because we get the benefit of a high mean outcome and compensatory risk profiles.

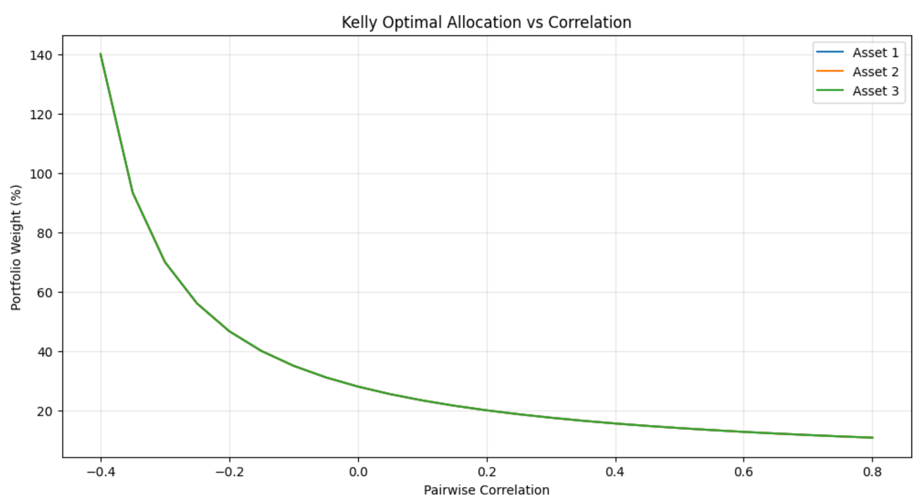

Figure 3: Kelly Criterion Bet Size versus Asset Pairwise Correlation

The correlation = 0 case is a solid baseline – Kelly tells us to bet roughly

27% of our portfolio on each signal to be optimal. When the bets are correlated, we see less diversification benefit – Kelly tells us to bet roughly 10% on each asset instead. However, as the correlation becomes more and more negative, Kelly tells us to bet around 140% of our portfolio **on each asset** – or over 400% of our portfolio across the 3 assets. That's the power of diversification: the math implies that negatively correlated, positive expectancy bets are good enough to be worth mortgaging your house.

Of course, correlations cannot be directly observed. The instability of a correlation matrix is a notorious impediment to mean-variance optimization. Correlations of a larger magnitude are, more often than not, spurious. Practically, one should practically mortgage their house for that noisy -.4 correlation. Perhaps such correlations exist in the sports world (e.g. player parlays), but they're rare in the trading world.

However, that (simple) example should illustrate how correlations affect bet sizes. As our bets become less correlated, they become more diversifying as a portfolio – therefore, we ought to bet more.

6 Conclusion

All the math aside, these frameworks for bet sizing come to three intuitive conclusions.

1. Bet more when you have more edge
2. Bet more when underlying return streams are less correlated
3. Bet to avoid ruin (no bankroll left)