Derivation and application of the generalized Kelly criterion

Jan Oberwahrenbrock

February 2025

1 Introduction

Imagine you get the opportunity to play a manipulated coin toss game. You know that $P(\text{Head}) = 0.6$ so $P(\text{Tail}) = 1 - P(\text{Head}) = 0.4$. You pay an arbitrary amount as buy-in and predict whether Head or Tail will be the result. If your prediction is correct, you get your buy-in doubled back; if it's wrong, you lose your buy-in.

You have $100 in your wallet and wonder if you should play and, if so, what your strategy should be. It is intuitive that you should play because the odds are in your favor if you bet on Head. However, finding the right strategy seems harder. Let's keep in mind that each coin flip is independent of the previous ones, so our betting amount shouldn't depend on any previous results. Thus, we successfully prevent ourselves from falling for the gambler's fallacy.

The only variable remaining that determines the buy-in amount is our capital. If we bet nothing, we miss out on the probability in our favor. If we bet all our capital in every round, we have a 40% probability to go completely broke in every round. To go all-in every round only works out if we win every round, which probability goes extremely fast to 0 as we play more rounds (see Figure 1).

Note:
If your capital is K_n going all-in maximizes the expected value of the new capital K_{n+1} that you have after playing one round:

$$E[K_{n+1}] = (K_n + \text{buy-in}) \cdot 0.6 + (K_n - \text{buy-in}) \cdot 0.4$$

with buy-in $\in [0, K_n]$. Although maximizing the expected value of new capital for each round might sound good at first, it also maximizes our probability of bankruptcy at a really impressive velocity as we play more and more games. So it would be disastrous to maximize for the expected value of the new capital for each round.

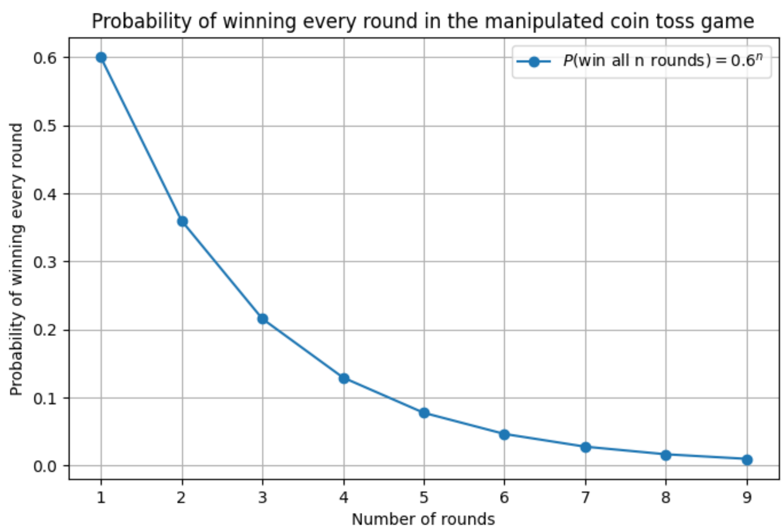


Figure 1: Probability of Winning Every Round in a Coin Toss Game

So we know that for each round we should bet more than nothing and less than everything. So there has to be a fraction f^* of our capital K with $0 < f^* < 1$ that is optimal to bet each round.

Let's find that optimal fraction f^* of our capital K:

2 Kelly criterion for manipulated coin toss game

You play for n rounds and you get Head h times and Tail t times with $n = h+t$. Your capital after n rounds is:

$$K_n = K_0 \cdot (1+f)^h \cdot (1-f)^t \tag{1}$$

For example, let's say you start with $K_0 = 100$ dollars, and you randomly choose $f = 0.1$. If you play $n = 5$ rounds and get the following sequence of outcomes: H, T, T, H, H, then your capital progression is:

$$K_1 = 100 \cdot (1.1) = 110.00$$

$$K_2 = 110.00 \cdot (0.9) = 99.00$$

$$K_3 = 99.00 \cdot (0.9) = 89.10$$

$$K_4 = 89.10 \cdot (1.1) = 98.01$$

$$K_5 = 98.01 \cdot (1.1) = 107.81$$

So after 5 rounds, your capital has grown to \$107.81, which is exactly what the closed-form formula (see Eq. (1)) gives:

$$K_5 = 100 \cdot (1.1)^3 \cdot (0.9)^2 = 100 \cdot 1.331 \cdot 0.81 = 107.81$$

Note:
As we can see, K_4 is lower than our initial capital, even though we doubled our buy-in twice and lost it twice. The problem is that, for example, a 10% gain does not offset a 10% loss, since after the loss, the nominal stake is lower. Likewise, a 10% loss after a 10% gain is worse because in this case, a higher nominal stake is lost (see Figure 2). This exact problem arises when considering the expected value of the capital. While a percentage loss of $x\%$ is mathematically offset by a percentage gain of $x\%$ in the expected value calculation, this approach fails to account for the compounding nature of gains and losses. Therefore, multiplication must be used instead of addition, making the expected value an unsuitable metric in this context.

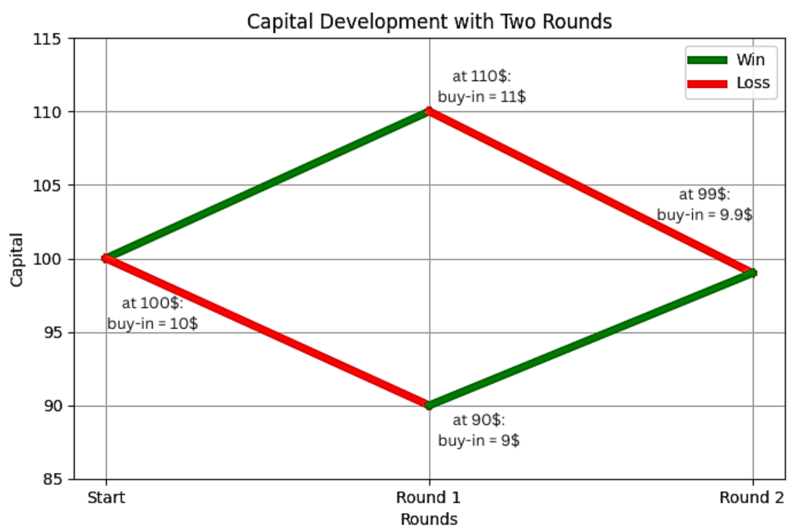


Figure 2: Impact of Consecutive 10% Gain and 10% Loss

Here we randomly selected the values for h, t and f. Now let's keep f as a variable and think about the properties of h and t:

As n grows, the fractions $\frac{h}{n}$ converges to $P(\text{Head})$ and $\frac{t}{n}$ converges to $P(\text{Tail})$, so we approximate:

$$h \approx n \cdot P(\text{Head}), \quad t \approx n \cdot P(\text{Tail})$$

Thus, we expect:

$$\begin{aligned}K_n &= K_0 \cdot (1+f)^h \cdot (1-f)^t \\ &\approx K_0 \cdot (1+f)^{nP(\text{Head})} \cdot (1-f)^{nP(\text{Tail})} \\ &= K_0 \cdot \Big((1+f)^{P(\text{Head})}\Big)^n \cdot \Big((1-f)^{P(\text{Tail})}\Big)^n \\ &= K_0 \cdot \Big((1+f)^{P(\text{Head})} \cdot (1-f)^{P(\text{Tail})}\Big)^n\end{aligned}$$

which leads to exponential growth with a growth factor of:

$$(1+f)^{P(\text{Head})} \cdot (1-f)^{P(\text{Tail})}$$

Thus, the optimal fraction f^* is the value that maximizes the growth factor. To find the optimal fraction f^*, we determine the maximum by taking the derivative and setting it to zero. Alternatively, the maximum can be found numerically or graphically.

$$\begin{aligned}0 &= \frac{d}{df^*}(1+f^*)^{P(\text{Head})} \cdot (1-f^*)^{P(\text{Tail})} \\ &= \frac{d}{df^*} e^{\ln((1+f^*)^{P(\text{Head})} \cdot (1-f^*)^{P(\text{Tail})})} \\ &= \frac{d}{df^*} e^{\ln((1+f^*)^{P(\text{Head})}) + \ln((1-f^*)^{P(\text{Tail})})} \\ &= \frac{d}{df^*} e^{P(\text{Head})\ln(1+f^*) + P(\text{Tail})\ln(1-f^*)} \\ &= \left(\frac{P(\text{Head})}{1+f^*} - \frac{P(\text{Tail})}{1-f^*}\right) \cdot e^{P(\text{Head})\ln(1+f^*) + P(\text{Tail})\ln(1-f^*)} \\ \Leftrightarrow \quad 0 &= \frac{P(\text{Head})}{1+f^*} - \frac{P(\text{Tail})}{1-f^*}\end{aligned}$$

The above transformations are based solely on logarithm laws, differentiation rules, and the fact that e^x is always greater than zero.

Now, the next step is to solve the equation for f^*.

$$
\begin{aligned}
0 &= \frac{P(\text{Head})}{1+f^*} - \frac{P(\text{Tail})}{1-f^*} \\
\Leftrightarrow \quad & \frac{P(\text{Head})}{1+f^*} = \frac{P(\text{Tail})}{1-f^*} \\
\Leftrightarrow \quad & P(\text{Head})(1-f^*) = P(\text{Tail})(1+f^*) \\
\Leftrightarrow \quad & P(\text{Head}) - P(\text{Head})f^* = P(\text{Tail}) + P(\text{Tail})f^* \\
\Leftrightarrow \quad & P(\text{Head}) - P(\text{Tail}) = P(\text{Head})f^* + P(\text{Tail})f^* \\
\Leftrightarrow \quad & P(\text{Head}) - P(\text{Tail}) = f^*(P(\text{Head}) + P(\text{Tail})) \\
\Leftrightarrow \quad & f^* = \frac{P(\text{Head}) - P(\text{Tail})}{P(\text{Head}) + P(\text{Tail})}
\end{aligned}
$$

Now, we can substitute the values for $P(\text{Head})$ and $P(\text{Tail})$ to determine f^*.

$$
f^* = \frac{P(\text{Head}) - P(\text{Tail})}{P(\text{Head}) + P(\text{Tail})} = \frac{0.6-0.4}{0.6+0.4} = \frac{0.2}{1.0} = 0.2
$$

We therefore know that the growth factor is maximized when we bet 20% of our capital in each round.

If we stake 20%, we expect a growth factor per round of:

$$
(1+f^*)^{P(\text{Head})} \cdot (1-f^*)^{P(\text{Tail})} = 1.2^{P(\text{Head})} \cdot 0.8^{P(\text{Tail})} = 1.2^{0.6} \cdot 0.8^{0.4} \approx 1.0203
$$

So we expect a capital gain of about 2% per game in the long run.

The graphical determination of both f^* and the corresponding growth factor, using f^* as the bet size, confirms our calculations (see Figure 3 and 4).

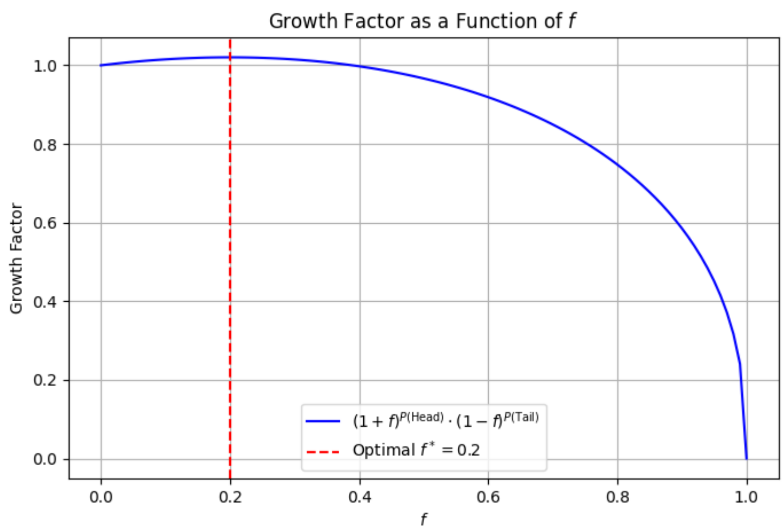


Figure 3: growth factors for different f values

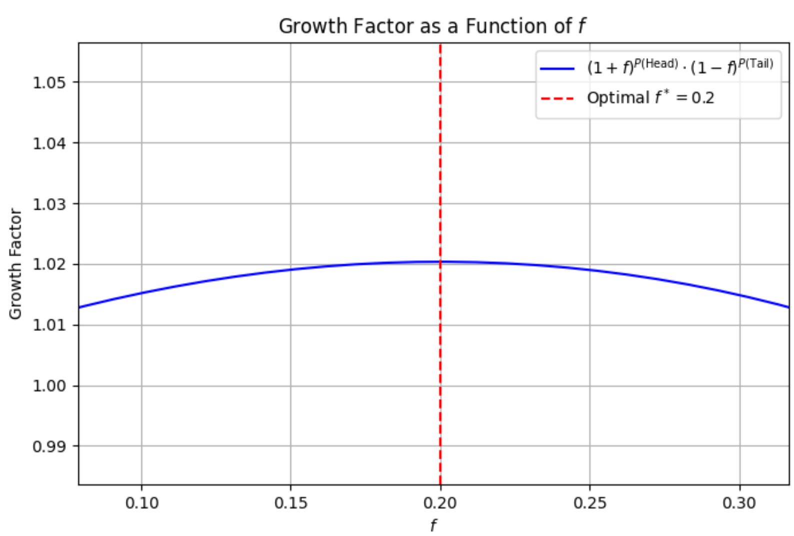


Figure 4: growth factors for different f values close to f^*

3 Limitation of This Approach in Reality

Keep in mind that we approximated $h \approx n \cdot P(\text{Head})$ and $t \approx n \cdot P(\text{Tail})$, leading to:

$$K_n = K_0 \cdot (1+f)^h \cdot (1-f)^t \approx K_0 \cdot (1+f)^{nP(\text{Head})} \cdot (1-f)^{nP(\text{Tail})}$$

This approximation holds well when the number of rounds is sufficiently large for the law of large numbers to take effect. However, if the number of rounds is too small, randomness significantly influences the returns, which is undesirable.

While f^* is the optimal fraction in the long run, for a relatively small number of rounds, it may be beneficial to use a fraction $f' < f^*$, accepting a lower expected growth rate in exchange for reduced variance in final returns. This strategy is known as the fractional Kelly approach.

To examine this for the manipulated coin toss game, we simulate the outcomes for two betting fractions: $f^* = 0.2$ (full Kelly) and $f' = \frac{f^*}{4} = 0.05$ (fractional Kelly) and 100$ as initial capital.

The simulation consists of 1000 independent runs for each of the following cases: $n = 100$, $n = 500$, and $n = 1000$. The resulting final capital distributions are visualized using boxplots.

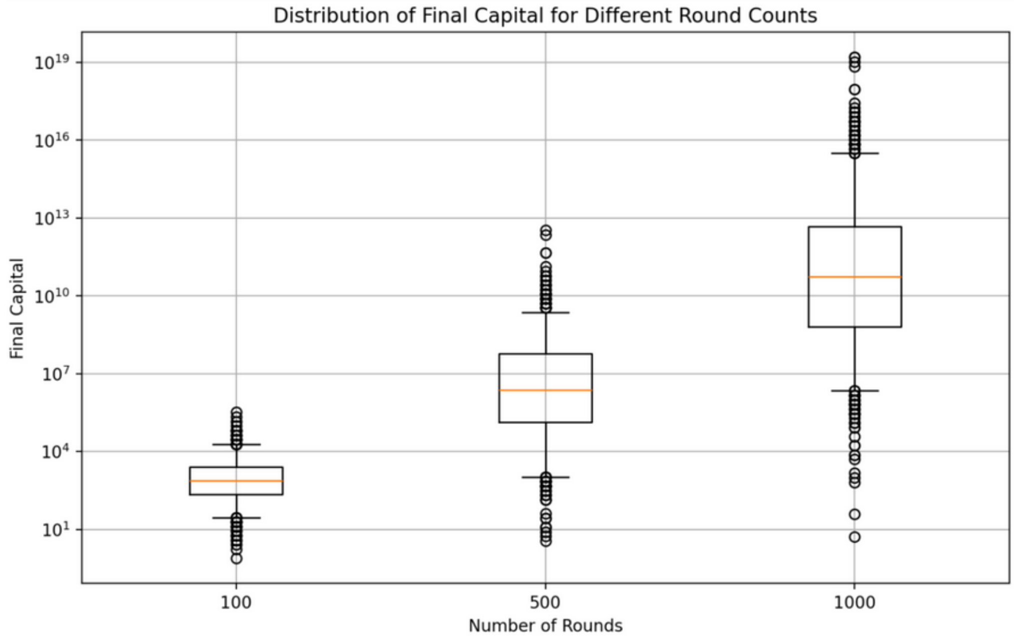
- The yellow line represents the median.
- The box covers 50% of the final capital values.
- The whiskers extend to all data points except for the bottom 5% and top 5%, which are marked as outliers (circles).

The table below the boxplot presents the key statistical values derived from the boxplots. The underlying growth factors were computed using the formula:

$$\text{growth factor} = \sqrt[n]{\frac{\text{final capital}}{\text{initial capital}}}$$

The calculated growth factors are displayed in the second table.

Manipulated coin toss game simulation with f = 0.2



Key Statistics of the Boxplot Distributions:

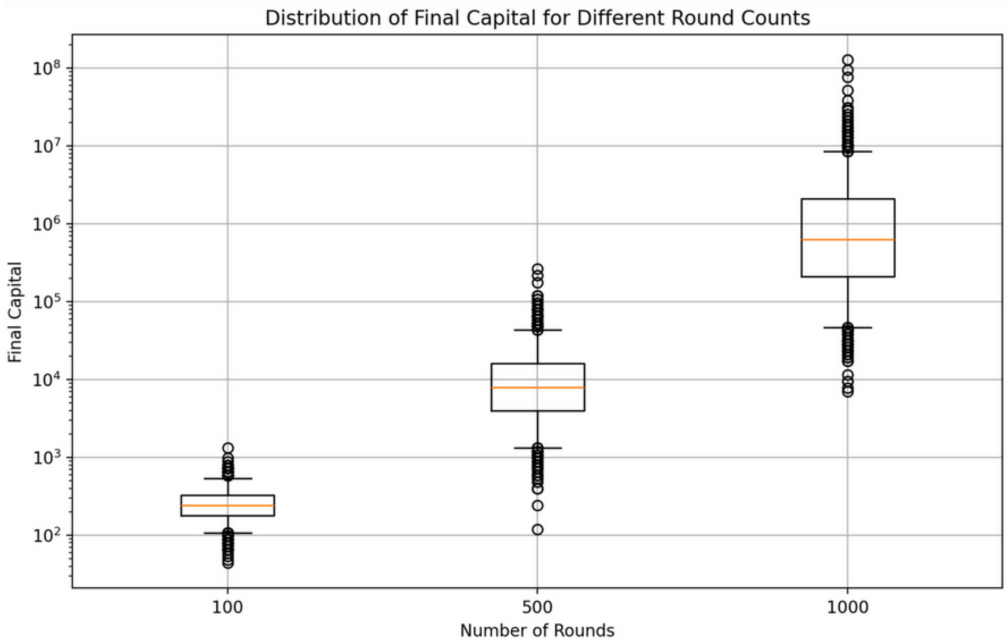
	Whisker Low (5%)	Q1 (25%)	Median (50%)	Q3 (75%)	Whisker High (95%)
100	2.92e+01	2.22e+02	7.49e+02	2.53e+03	1.92e+04
500	1.06e+03	1.38e+05	2.36e+06	6.04e+07	2.38e+09
1,000	2.20e+06	6.42e+08	5.56e+10	4.81e+12	3.16e+15

Computed Growth Factors for Boxplot Statistics:

	Whisker Low Growth (%)	Q1 Growth (%)	Median Growth (%)	Q3 Growth (%)	Whisker High Growth (%)
100	-1.222634 %	0.800342 %	2.033960 %	3.282676 %	5.397921 %
500	0.473903 %	1.456404 %	2.033960 %	2.698051 %	3.455431 %
1,000	1.004904 %	1.579890 %	2.033960 %	2.490060 %	3.157119 %

Figure 5: Simulation results for the manipulated coin toss game with $f = 0.2$ (Full Kelly)

Manipulated coin toss game simulation with f = 0.05



Key Statistics of the Boxplot Distributions:

	Whisker Low (5%)	Q1 (25%)	Median (50%)	Q3 (75%)	Whisker High (95%)
100	1.08e+02	1.78e+02	2.40e+02	3.24e+02	5.37e+02
500	1.32e+03	3.96e+03	7.97e+03	1.61e+04	4.37e+04
1,000	4.71e+04	2.11e+05	6.35e+05	2.11e+06	8.57e+06

Computed Growth Factors for Boxplot Statistics:

	Whisker Low Growth (%)	Q1 Growth (%)	Median Growth (%)	Q3 Growth (%)	Whisker High Growth (%)
100	0.075039 %	0.577086 %	0.879523 %	1.182870 %	1.695813 %
500	0.516708 %	0.738273 %	0.879523 %	1.020972 %	1.223385 %
1,000	0.617359 %	0.768524 %	0.879523 %	1.000753 %	1.142371 %

Figure 6: Simulation results for the manipulated coin toss game with $f = 0.05$ (Fractional Kelly)

A general observation for both f^* and f' is that as the number of rounds increases, the final capital values converge toward an underlying growth factor that corresponds to the expected theoretical value. This demonstrates the law of large numbers in action.

For 100 rounds, we observe that there are more final capital values below the initial capital when using f^* compared to f', as the variance is higher for f^*. Therefore, if an investor plans to play only 100 rounds, choosing f' might be preferable to reduce the probability of a net loss, as losing money is generally perceived as worse than winning slightly less (see Log Utility Theory).

When comparing the results for 1000 rounds, however, the argument for choosing f' becomes less relevant. Both strategies yield significantly higher final capital than the initial capital, except for a few outliers in f^*. The highest final capital values for f' are comparable to the lowest values for f^*, meaning that over longer time horizons, f^* clearly leads to better outcomes.

The only argument in favor of f' would be if an investor aims to maximize the probability of reaching a specific capital threshold, such as exceeding 500$. In the simulation, all values for f' remain above 500$, while f^* has a few outliers below this threshold.

For a relatively small number of rounds, a fractional Kelly approach reduces variance and increases the likelihood of staying above a certain capital threshold. However, for sufficiently large numbers of rounds, using f^* is optimal.

Thus, in all cases, f^* serves as an upper bound, as betting more than f^* reduces the expected growth rate and increases variance, which is highly undesirable for the rational investor.

4 Generalization of the Kelly Criterion

Now, let's generalize this approach so that it is applicable in real-life scenarios where situations are significantly more complex than in the manipulated coin toss game.

4.1 Definition of the Outcome Space X

In the manipulated coin toss game, we only had Head and Tail as possible outcomes. Now, we want to allow:

- Discrete sets of any finite size
- Continuous intervals as possible outcomes

Thus, X can either be a finite set or a continuous interval and contains all possible outcomes of the game.

- If X is a finite set, we define it as:

$$X = \{x_1, x_2, \ldots, x_{|X|}\}$$

where $|X|$ represents the number of elements in X.

- If X is a continuous interval, we define it as:

$$X = [a, b]$$

where a and b are the lower and upper bounds of the interval.

4.2 Definition of the Probability Function p

To describe the likelihood of different outcomes, we define a probability function p.

- If X is a finite set, then p is a probability mass function:

  $$p : X \to [0, 1]$$

  that assigns a probability $p(x_i)$ to each outcome x_i. Since probabilities must sum to one, we require:

  $$\sum_{i=1}^{|X|} p(x_i) = 1.$$

- If X is a continuous interval, then p is a probability density function (PDF):

  $$p : X \to \mathbb{R}_{\geq 0}$$

  such that the total probability integrates to one:

  $$\int_a^b p(x)\, dx = 1.$$

4.3 Definition of the Return Function r

Next, we define the return function $r : X \to \mathbb{R}$, which determines the return on the buy-in for every outcome $x \in X$:

- $r(x) = -1 \quad \to$ Buy-in is lost.
- $r(x) = 0 \quad \to$ You get exactly your buy-in back (break-even).
- $r(x) = 1 \quad \to$ You make your buy-in as a profit.

4.4 Generalized Growth Factor

With these definitions, the generalized growth factor is given by:

$$\prod_{x\in X}(1+f\cdot r(x))^{p(x)}$$

We apply a few transformations to make the calculation easier:

$$\prod_{x\in X}(1+f\cdot r(x))^{p(x)}$$

$$= e^{\ln\left(\prod_{x\in X}(1+f\cdot r(x))^{p(x)}\right)}$$

$$= e^{\sum_{x\in X}\ln(1+f\cdot r(x)^{p(x)})}$$

$$= e^{\sum_{x\in X}p(x)\cdot\ln(1+f\cdot r(x))}$$

Thus, the generalized growth factor simplifies to:

$$e^{\sum_{x\in X}p(x)\ln(1+f\cdot r(x))}$$

4.5 Finding the Optimal Fraction f^*

f^* is the value that maximizes the growth factor. That means:

$$f^* = \arg\max_f \prod_{x\in X}(1+f\cdot r(x))^{p(x)} = \arg\max_f e^{\sum_{x\in X}p(x)\ln(1+f\cdot r(x))}$$

Since the exponential function e^x is strictly increasing, this maximization problem reduces to maximizing the sum inside the exponent:

$$f^* = \arg\max_f \sum_{x\in X}p(x)\ln(1+f\cdot r(x))$$

- If X is a finite set $X=\{x_1,x_2,\ldots,x_{|X|}\}$, then:

$$f^* = \arg\max_f \sum_{i=1}^{|X|}p(x_i)\ln(1+f\cdot r(x_i))$$

- If X is a continuous interval $[a, b]$, then:

$$f^* = \arg\max_f \int_a^b p(x) \ln(1 + f \cdot r(x)) \, dx.$$

These maximizations can no longer be solved analytically and must be performed numerically.

4.6 Calculating the Growth Factor

Thus, the optimal growth factor in both cases is:

$$e^{\max\left(\sum_{x \in X} p(x) \ln(1+f \cdot r(x))\right)} = e^{\sum_{x \in X} p(x) \ln(1+f^* \cdot r(x))}$$

5 Application Example: Options Market

In this section, we illustrate how an investor who has developed a probability model for stock prices at a specific expiration date can use the generalized Kelly criterion to evaluate Bull Put Spreads.

The investor's probability model assigns probabilities to different stock prices at expiration, forming a continuous probability distribution over an interval X (see Figure 7). This allows for an assessment of potential option positions expiring on the same date.

A Bull Put Spread consists of a long put (purchased) at a lower strike price and a short put (sold) at a higher strike price. The return function corresponding to a specific selection of put options can be derived based on their strike prices and market prices (see Figure 8).

Given the probability model (Figure 7) and the return function (Figure 8), we define X as the interval of all possible stock prices at expiration. The generalized Kelly criterion can then be used to:

- Determine whether a Bull Put Spread with a positive growth factor exists.
- Identify the combination of put options that maximizes the growth factor.
- Calculate the optimal fraction of capital to allocate to such a trade.

This is done by iterating over all possible Bull Put Spreads and applying the generalized Kelly criterion to each of them.

This example does not aim to construct a functional trading system but rather demonstrates how an investor could systematically apply the generalized Kelly criterion in the context of evaluating options strategies.

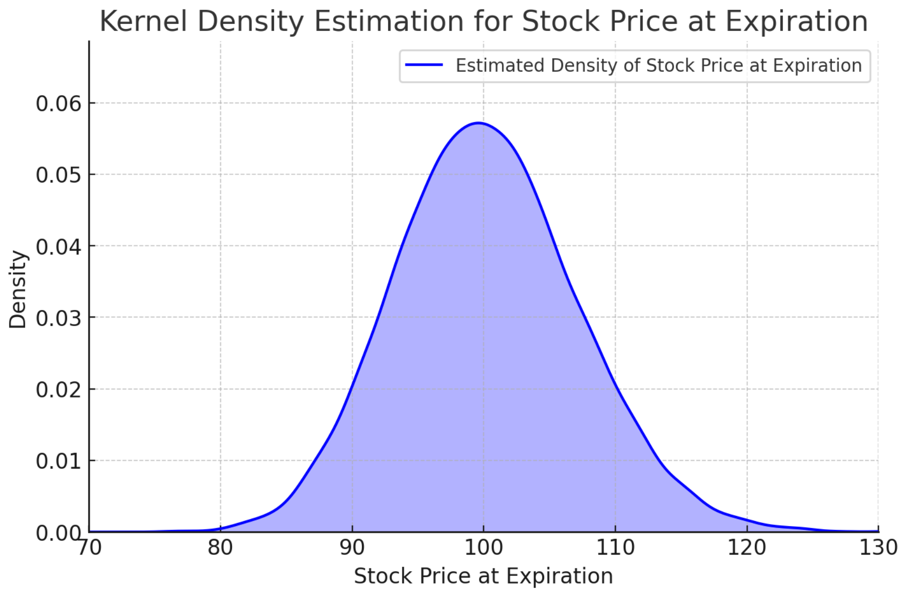


Figure 7: Estimated probability distribution of stock price at expiration

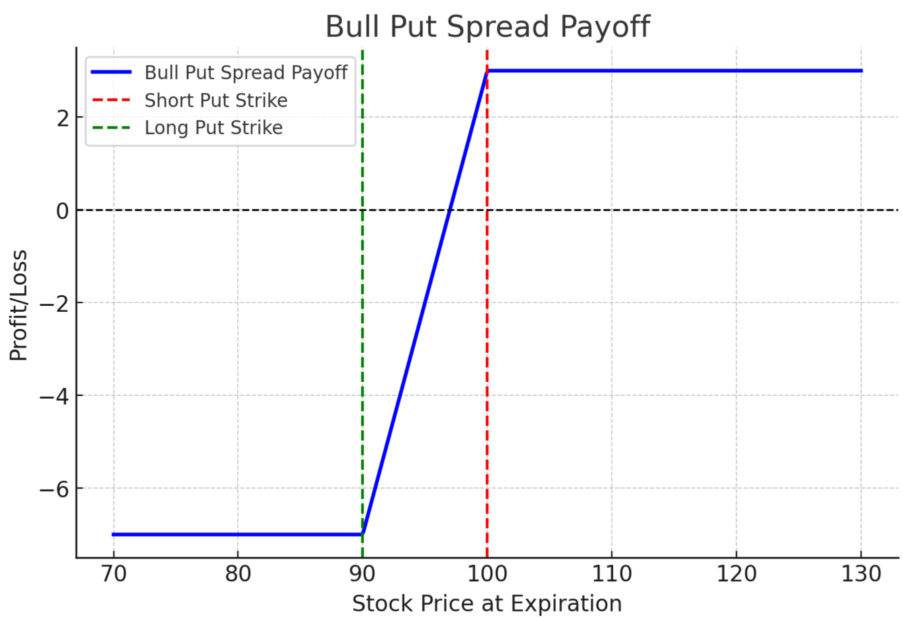


Figure 8: Return function of a Bull Put Spread