Kelly Criterion, Drawdowns, and Fat-Tails

This is a very, very rough draft with many incomplete thoughts, typos, and errors

Austin Pollok

15th January 2020

Contents

List of Code

List of Figures

Part I

History and Possible Uses

Chapter 1

History of Kelly Criterion

The Kelly Criterion was first formalized by John Kelly in 1956 (Kelly, 1956) and later popularized by Ed Thorp when he used it to determine bet sizes in his newly discovered card counting system (E. O. Thorp, 1966). Kelly thought of this as an application of Claude Shannon's theory of communication while working together at Bell Labs. There was no thought of utility maximization, it was developed solely from an information theoretic interpretation of gambling with an edge. Thorp later used it when gambling in the securities market operating as the first quantitative market neutral hedge fund. Note it can be quite useful to think of investing in markets as gambling in casinos (quote from (E. O. Thorp, 2018))...

(I believe) It was first called Kelly Criterion in paper by Thorp and Walden 1966, "A Favorable Side Bet in Nevada Baccarat." (E. Thorp & E. Walden, 1966)

The Kelly Criterion is nothing more than way to determine how to optimally bet when gambling in things with random returns so as to maximize our long term wealth. A so called "theory of re-investment," as opposed to a theory of investment (such as Markowitz theory), (Poundstone, 2006). This system only makes sense if we are placing a large number of bets and trying to compound our wealth. This opposes the Markowitz mean variance criterion which only applies to a single bet with no consideration of compounding. Intuitively, the Kelly Criterion tries to formalize the common gambler's saying of "playing with the house's money."...Note briefly compare the two strategies from (Brown, 2011)...

Summarize less technical writings on Kelly criterion from Nassim Taleb, Bill Gross, Mohnish Pabrai, and Ziemba's chapter on how Buffet and Simons bet like a Kelly gambler, etc....Also note the scandal with Paul Samuelson....

..."Kelly is not limited to two value payoffs, but applies generally to any gambling or investing situation **in which the probabilities are known or can be estimated** " . "...the investor or bettor generally avoids total loss; the bigger the edge, the larger the bet; the smaller the risk, the larger the bet." "The Kelly Criterion, not having been invented by the old-line academic economists, has generated considerable controversy." "Three caveats: the Kelly Criterion may lead to wide swings in the total wealth, so most users choose to bet some lesser fraction, typically one-half Kelly or less; for investors with short time horizons or who are averse to (volatility) risk, other approaches may be better; an exact application of Kelly requires exact probabilities of payoffs such as those in most casino games; to the extent these are uncertain, which is generally the case in the investment world, the Kelly bet should be based on a conservative estimate of the outcome." "The Kelly Criterion highlights the perils of overbetting even when you have the edge." (E. O. Thorp, 2018) .....See Aaron Brown's description of Kelly in his book and on post on Quora...Also see The Mathematics of Gambling book.

Q: How is this different from Modern Portfolio Theory? See Thorp's paper and see Ziemba's paper responding to Paul Samuelson.
Note it reinterprets risk as exposure or fraction of wealth in a gamble.

Note (risk, return) = $(f, g_r(f))$ implies that more "risk" as measured by exposure doesn't guarantee more "return" as measured by long-term growth rate. In fact, more risk can yield lower return or worse, ruin! This contradicts more risk, more return tradeoff.

What is the Sharpe ratio for Kelly type thinking? Sharpe ratio is meant to formalize the "risk/return" relationship using a ratio

$$\frac{\text{return}}{\text{risk}} = \frac{\text{long-term growth rate}}{\text{exposure}} = \frac{g_r(f)}{f}$$

Are there merits for considering

$$\frac{\mathbb{E}\big[\ln(1+fr)\big]}{f} \quad \text{vs} \quad \frac{\mathbb{E}\big[r\big]}{\sqrt{\mathrm{Var}(r)}} \quad ?$$

Also see Feynman's youtube video discussing same mathematical formulation with different stories can cause the thinking and questions to be entirely different.

- Note Feynman video on Knowing vs Understanding and how different ideas with equivalent formalizations can lead to very different questions
- Difference between adding more bets and reducing exposure
- Diversification practically means adding more bets that are uncorrelated to bets already taking. It reduces "noise/diffusion risk" or small jumps, but not large jump risk.
- Large jump risk can only be handled from reducing exposure or explicit hedges.
- Kelly more general than MPT, it amounts to reducing exposure, any way you do it. One particular way of reducing exposure is by adding simultaneous/contemporaneous bets which are truly independent or at least truly uncorrelated (especially in distressed times).
- MORE RISK $\implies$ LESS RETURN (or even RUIN)

Chapter 2

Possible Uses and Empirical Work

Need to add a motivation as to how and why particular types of traders/gamblers/investors would use this system. Document how it has been used in (E. O. Thorp, 1966) and in (E. Thorp & Kassouf, 1967)

Application to credit (loan) investing, loan portfolio construction.

Consider momentum and trend following strategies (or Carry and Trend in Lots of Places by Bhansali). See papers by Daniel and Moskowitz, Momentum Crashes (2015), and by Goufu Zhou, Taming Momentum Crashes: A Simple Stop Loss Strategy(2015). We should also investigate a survey of betting on unobserved events, or an empirical investigation of 'statistics of unobserved events' to help justify the fat tailed stable distribution assumptions.

Part II

Mathematical Formalization of Kelly Criterion Assuming Constant f Main tool: Law of Large Numbers

Chapter 3

Kelly Criterion in Discrete Time

We are essentially following (E. O. Thorp, 2006).
Consider the following scenario, we are repeatedly/sequentially placing bets on a sequence of random gambles/payoffs/returns (such as stocks, bonds, casino games, etc) and we are trying to decide how much, as fraction of current wealth, $f \in [0,1]$, we should bet to maximize our long term wealth. The constraint of being in $[0,1]$ corresponds to no short selling and no leverage/use of margin.
We further assume the (relative/net) returns,$\{r_k\}_{k\geq 1}$, of the gambles are independent and identically distributed, and

$$\forall k \geq 1, \qquad r_k := \frac{P_k - P_{k-1}}{P_{k-1}} = \frac{P_k}{P_{k-1}} - 1 \overset{d}{=} r \geq -1, \qquad \text{for some price process } P = \{P_k\}_{k\geq 1}, \quad P_k \geq 0$$

In a financial context it is clear what a price process is. In a gambling context we can think of the odds as the price, namely for $b:1$ odds, the initial "price" to play this game or take this bet is \$1, and the final "price" will be either $\$(b+1)$, or \$0.
Most importantly, we assume we have an edge,

$$\mathbf{E}[r] > 0$$

Note that we are apriori **assuming** that f is a fixed constant as opposed to a predictable process $\{f_{k-1}\}_{k\geq 1}$. This is intuitively justified by the fact that we are betting on independent and identically distributed gambles/returns. There are, however, rigorous justifications of this fact using dynamic programming arguments, see (Bellman & Kalaba, 1957), (Browne & Whitt, 1996), and paper by Kallsen.

(One possible question is what happens when we assume non-iid gambles, such as some sort of correlated process, or various distributions, such as fat tailed distributions?)

Let $\{W_n^f\}_{n\geq 1}$ be the wealth dynamics/process/trajectory achieved from betting the fraction of current wealth f in the sequence of gambles/returns $\{r_n\}_{n\geq 1}$, while compounding the returns (i.e. returns on returns much like interest on interest). In other words, the wealth process is modeling the process of compounding returns with a constant exposure(proportion) f to those returns.

Remark. *Note that we can interpret "discrete time" the same as discretely compounding over the unit interval* $[0,1]$*, for which we would use the notation* $W_1^{n,f}$*. This allows for an easy generalization to continuous time.*

The important part is that time is regarded as a point where we place a bet. So when we let $n \to \infty$, it can be interpreted as time (number of bets) increasing, or equivalently the frequency of bets per unit interval increasing to "continuous" betting.

3.1 Discrete Time Wealth Dynamics

$$
\begin{aligned}
W_0 &= w = \text{initial wealth/bankroll/portfolio size} \\
W_1^f &= W_0 + r_1(fW_0) = W_0(1 + fr_1) \\
W_2^f &= W_1^f + r_2(fW_1^f) = W_1^f(1 + fr_2) = W_0(1 + fr_1)(1 + fr_2) \\
&\vdots \\
W_n^f &= W_{n-1}^f + r_n(fW_{n-1}^f) = W_{n-1}^f(1 + fr_n) = W_0 \prod_{k=1}^{n}(1 + fr_k)
\end{aligned}
$$

See code 1 for a generic simulation of a wealth process assuming you can somehow simulate the returns.

3.2 Kelly Criterion

$$
\frac{W_n^f}{W_0} = e^{n \ln\left(\frac{W_n^f}{W_0}\right)^{\frac{1}{n}}} = e^{n \frac{1}{n} \sum_{k=1}^{n} \ln(1 + fr_k)} \xrightarrow[n\to\infty]{a.s.} e^{\infty \mathbf{E}[\ln(1 + fr)]}
$$

by the Law of Large Numbers.

Consider the growth rate at the n^{th} bet, which measures the exponential rate of increase per bet.

$$
G_n(f) := \frac{1}{n} \ln\left(\frac{W_n^f}{W_0}\right)
$$

Note it is the log geometric average of the first n returns.
Similarly, consider the long term exponential growth rate, which is the rate at which we expect, in the long run, to compound our wealth.

$$
g_r(f) := \mathbf{E}[\ln(1 + fr)]
$$

Since our goal is to maximize our long term wealth process, we see we can equivalently maximize our long term growth rate.
This leads to the so-called Kelly Criterion:

$$
\max_{f \in (0,1)} g_r(f) = \max_{f \in (0,1)} \mathbf{E}[\ln(1 + fr)] = \max_{f \in (0,1)} \int_{-1}^{\infty} \ln(1 + fx) p_r(x) dx
$$

In code 2 we simulate the wealth processes from betting on a coin toss and consider look at the actual growth rate and the long-term theoretical growth rate.

Remark. *This setup lends itself to see how this problem is thought about and how to formalize it for more complicated return sequences. Also we want to make this a vector of returns for more general settings.*

Remark. *The Kelly criterion has many other names depending on which point of view you are coming from. Most in the gambling or information theory community appear to call it the Kelly criterion, while most in the academic finance literature tend to call it optimal long term growth or log utility.*

(I need to investigate some of the other finance papers, such as those from Markowitz, Merton, Karatzas, Kardaras, etc)

3.3 Properties of Kelly Criterion in Discrete Time

Instead of limiting $f \in [0,1]$, it seems more general to assume $f \in \mathbb{R}$ such that $\mathbb{E}[|\ln(1+fr)|] < \infty$ of which $f \in [0,1]$ is a particular way to enforce this.

Consider the long-term growth rate, $g_r(f)$. Here, we investigate some basic properties of the structure of this function. For now, assume we can pass the limits and derivatives under the integral sign, and we assume whatever moment conditions we need for these functions to make sense. (Justify with a convergence theorem)

- $\frac{d}{df} g_r(f) = \mathbf{E}\left[\frac{r}{1+fr}\right] \implies \frac{d}{df} g_r(0) = \mathbf{E}[r] > 0,$ so $g_r(f)$ is increasing at $f = 0,$ and $g_r(0) = 0$
- $\frac{d}{df} g_r(1-) = \mathbf{E}\left[\frac{r}{1+r}\right] < 1$, if less than zero than we have existence of an f^* by Intermediate Value Theorem
- $\frac{d^2}{df^2} g_r(f) = -\mathbf{E}\left[\frac{r^2}{(1+fr)^2}\right] < 0 \iff g_r(f)$is convex $\iff \frac{d}{df} g_r(f)$ is decreasing, the question is does it ever decrease below zero?
- Claim $\forall c > 0,\ \mathbb{P}(r > c) > 0 \iff f^* > 0$ and $\forall \varepsilon > 0,\ \mathbb{P}(r < -1 + \varepsilon) > 0 \iff f^* < 1$
- Also, $\frac{d}{df} g_r(f)$ is decreasing $\implies$ if a solution exists, that is $\exists f^*$ such that $\frac{d}{df} g_r(f^*) = 0$, or more generally, $\exists f_c$ such that $\frac{d}{df} g_r(f_c) < 0$, then the solution is unique.

I need to summarize the properties of the Kelly as investigated by Breiman and later Cover. Also, here I need to put the reason that having an edge implies positive long term growth, see (E. O. Thorp, 2006) or Ethier's Doctrine of Chances page 359. Basically, $g_r''(f) < 0$ implying strict concavity and $g_r'(0) = \mathbf{E}[r] > 0$ with $g_r(0) = 0$ and $g_r(1-) = -\infty$ which implies there is an $f \in [0,1)$ such that $g_r(f) > 0$ by the intermediate value theorem. See Ethier's book page 361 for a CLT-type result for the geometric average growth rate, and try to get this using Discretization of Processes book for the continuous time results. See (Ethier, 2010).

3.4 Examples

3.4.1 Binary Bet (Binary Returns) $b:1$

Assume returns (payoffs) are binary, $b:1$. Note we think of the odds as the payoff/net-returns.
Let

$$r_k \stackrel{d}{=} r = \begin{cases} \text{b} & \text{with probability p} \\ -1 & \text{with probability 1-p} \end{cases} = b^{Y_k}(-1)^{1-Y_k}$$

Where $\{Y_k\}$ are iid Bernoulli(p) random variables.
We also assume we have an edge, that is,

$$\mathbf{E}[r] = bp + (-1)(1-p) = (b+1)p - 1 > 0 \iff p > \frac{1}{b+1}$$

Also note

$$\sigma_r = \sqrt{Var(r)} = (b+1)\sqrt{p(1-p)}$$

Then we are forced to ask, how should we compute the optimal betting fraction? We can try to explicitly evaluate the long term exponential growth rate and then somehow maximize it (take derivatives and solve explicitly or using Newton's method (etc), or gradient descent/other algorithms for finding the extrema). Another slightly less direct way is to use the Law of Large Numbers to try to approximate the integral and then search for the optimal fraction, i.e. Monte Carlo methods. An even less direct way would be to use backwards induction, that is, the dynamic programming principle. Lastly, we can try to use Taylor series expansions, under appropriate moment conditions, to approximate the logarithmic function. All of these methods have their own merit and some are better than others depending on assumptions about the control fraction and return distributions.
In summary, to solve we can try:

- Take derivative explicitly, solve for zeros
- Use Taylor series expansion, take derivative, solve for zeros
- Numerically solve via some gradient-descent type algorithm
- Simulate using (Monte-Carlo) Law of Large Numbers
- DPP

For this setup, consider

$$\begin{aligned} g_r(f) &= p\ln(1+fb) + (1-p)\ln(1-f) \\ \implies \frac{d}{df}g_r(f) &= \frac{bp}{1+fb} - \frac{1-p}{1-f} = 0 \\ \implies f^* &= \frac{bp-(1-p)}{b} = \frac{(b+1)p-1}{b} = \frac{\text{edge}}{\text{odds}} \end{aligned}$$

In code 3, we use Monte-Carlo simulation and a numerical solver in Python's package SciPy to search for f^*.

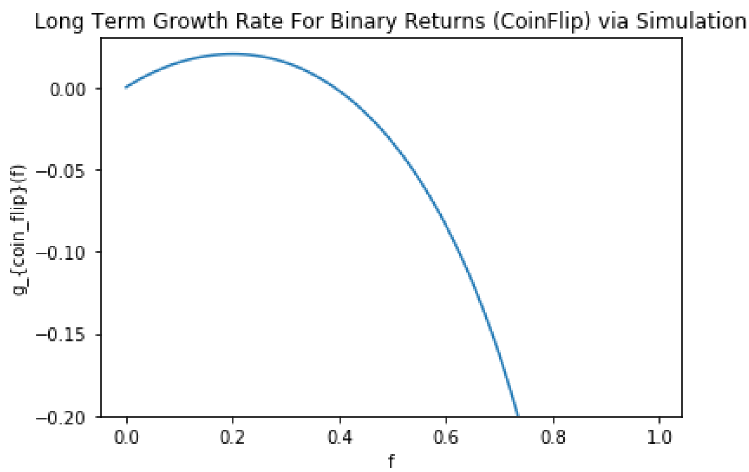


Figure 3.1: Long Term Growth Rate for Binary Bet with b = 1, and p = 0.6. This shows that more "risk" doesn't necessarily mean more "reward". In fact, too much "risk" means no "reward" and actually yields losses

(Note we need to work out the details here and explain the implications, also include a simulation of wealth processes and include a graph of growth rate vs. fraction to see how bet sizes affects growth rates. Also include various theorems about the optimal properties of the Kelly as explored in Breiman and Cover, etc. Also include the paper about essentially different strategies and how a non-Kelly fraction can dominate in long run (E. O. Thorp, 2012))

3.4.2 Multivariate Binary Returns

Applications to loan portfolio construction.

3.4.3 Uniform Returns

Assume (for whatever reason) that returns are distributed uniformly, $r \sim \text{Uniform}[-1, b]$.
We also assume we have an edge, that is,

$$\mathbf{E}[r] = \frac{b-1}{2} > 0 \iff b > 1$$

We are again forced to ask how we are to compute the optimal betting fraction? Yet this case isn't quite as straight forward as the binary betting scenario. Consider the long term growth rate,

$$\begin{aligned}
g_r(f) &= \mathbf{E}[\ln(1+fr)] \approx \frac{1}{n}\sum_{i=1}^{n} \ln(1+fr_i) \quad \text{which we can simulate and solve for } f^* \\
&= \int_{-1}^{b} \ln(1+fx) p_{\text{Uniform}}(x)dx = \frac{1}{b+1}\int_{-1}^{b} \ln(1+fx)dx \\
&= \frac{(1+fb)\ln(1+fb) - (1-f)\ln(1-f)}{f(b+1)} - 1 \\
\implies \frac{d}{df} g_r(f) &= \frac{fb - \ln(1+fb) + \ln(1-f)}{(b+1)f^2} = 0
\end{aligned}$$

This has to be solved numerically. Thus we can numerically solve for the zero of the derivative, or we can use simulation (LLN) methods to find the maximizer. In code 4, we carry this out.
In the case of $b = 2$, we see $f^* \approx .716...$

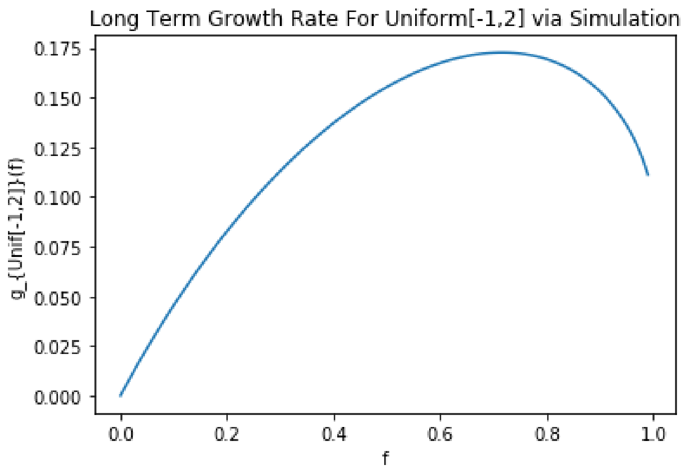


Figure 3.2: Long Term Growth Rate for Uniform Returns with b = 2

3.4.4 Cauchy$^2 - 1$ Returns

This is a special situation where you can compute the integral semi-explicitly with special functions and $f^* = \frac{1}{2}$. See graph, and table from book (ask Lototsky).

3.4.5 Log Normal (Gross) Returns

Let $P = \{P_k\}$ be the "price process" of some random outcome we are betting on. Consider the "net-returns," $r_k = \frac{P_k - P_{k-1}}{P_{k-1}}$.
Assume the log-(gross)returns are normally distributed. That is,

$$\ln\left(\frac{P_k}{P_{k-1}}\right) \sim \mathcal{N}(\mu, \sigma^2) \iff r_k \overset{d}{=} r \sim \mathrm{e}^{\mathcal{N}(\mu,\sigma^2)} - 1$$

Additionally, assume we have an edge,

$$\mathbf{E}[r] = \mathbf{E}[\mathrm{e}^{\mathcal{N}(\mu,\sigma^2)} - 1] =?\, \mathrm{e}^{\mu + \frac{\sigma^2}{2}} - 1 > 0 \iff \mu + \frac{\sigma^2}{2} > 0$$

Now consider

$$\begin{aligned} g_r(f) &= \mathbf{E}[\ln(1 + fr)] = \mathbf{E}\Bigg[\ln\Big(1 + f(\mathrm{e}^{\mathcal{N}(\mu,\sigma^2)} - 1)\Big)\Bigg] \approx \frac{1}{n}\sum_{i=1}^{n} \ln\Big(1 + f(\mathrm{e}^{N_i} - 1)\Big) \\ &= \int_{-\infty}^{\infty} \ln\Big(1 + f(\mathrm{e}^{x} - 1)\Big) p_{\mathcal{N}(\mu,\sigma^2)}(x)dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \ln\Big(1 + f(\mathrm{e}^{x} - 1)\Big)\, \mathrm{e}^{\frac{-1}{2\sigma^2}(x-\mu)^2}\, dx \end{aligned}$$

Before we try to solve explicitly, we can ask if a solution actually exists. Let $N \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\begin{aligned}
\frac{d}{df} g_r(1-) &= \mathbf{E}\left[\frac{r}{1+r}\right] = \mathbf{E}\left[\frac{\mathrm{e}^N - 1}{1 + \mathrm{e}^N - 1}\right] \\
&= 1 - \mathbf{E}[\mathrm{e}^{-N}] \\
&= 1 - \mathrm{m}_N(-1) = 1 - \mathrm{e}^{-\mu + \frac{1}{2}\sigma^2} < 0 \\
\iff &- \mu + \frac{1}{2}\sigma^2 > 0
\end{aligned}$$

Therefore, to have an edge, we must guarantee $\mu + \dfrac{\sigma^2}{2} > 0$, and to have an $f^* \in (0,1)$, we must guarantee $-\mu + \dfrac{\sigma^2}{2} > 0$.

Thus to guarantee a valid solution,

$$|\mu| < \frac{\sigma^2}{2}$$

To solve we can try:

- Take derivative explicitly, solve for zeros
- Use Taylor series expansion, take derivative, solve for zeros
- Numerically solve via some gradient-descent type algorithm
- Simulate using Law of Large Numbers
- DPP?

In progress...

3.4.6 Log Student-t (Gross) Returns

In progress...

3.4.7 Log Pareto (Gross) Returns

In progress...

3.4.8 Log Normal Inverse Gaussian (Gross) Returns

In progress...

3.4.9 Log α-Stable (Gross) Returns

Let $P = \{P_k\}$ be the "price process" of some random outcome we are betting on. Define the "net-returns" as $r_k = \frac{P_k - P_{k-1}}{P_{k-1}}$.
Assume the log-(gross)returns follow an $\alpha-$stable distribution. That is,

$$\ln\left(\frac{P_k}{P_{k-1}}\right) \sim \mathcal{S}(\alpha,\beta,\mu,\sigma) \iff r_k \overset{d}{=} r \sim \mathrm{e}^{\mathcal{S}(\alpha,\beta,\mu,\sigma)} - 1$$

Additionally, assume we have an edge,

$$\mathbf{E}[r] = \mathbf{E}[\mathrm{e}^{\mathcal{S}(\alpha,\beta,\mu,\sigma)} - 1] = ? > 0 \iff ? > 0$$

Now consider

$$\begin{aligned} g_r(f) &= \mathbf{E}[\ln(1+fr)] = \mathbf{E}\Big[\ln\big(1 + f(\mathrm{e}^{\mathcal{S}(\alpha,\beta,\mu,\sigma)} - 1)\big)\Big] \approx \frac{1}{n}\sum_{i=1}^{n} \ln\big(1 + f(\mathrm{e}^{S_i} - 1)\big) \\ &= \int_{-\infty}^{\infty} \ln\big(1 + f(\mathrm{e}^{x} - 1)\big) p_{\mathcal{S}(\alpha,\beta,\mu,\sigma)}(x)dx \\ &= ? \end{aligned}$$

This begs the question of what is the density for an α-Stable distribution? It turns out there is no closed form formula for the -stable distributions, except for $\alpha = 0.5, 1, 2$ which turns out to be the so-called Levy, Cauchy, and Gaussian distributions. They do however possess closed form formulas for their characteristic functions, as well as series representations. See (Samorodnitsky & Taqqu, 1994) for many more properties.

$$\begin{aligned} \varphi_{\mathcal{S}(\alpha,\beta,\mu,\sigma)}(t) &= \mathrm{e}^{i\mu t - |\sigma t|^{\alpha}\left(1 - i\beta\frac{t}{|t|}\Phi\right)} \\ \Phi &= \begin{cases} \tan\left(\frac{\alpha\pi}{2}\right) & ,\alpha \neq 1 \\ \frac{-1}{2\pi}\ln|\sigma t| & ,\alpha = 1 \end{cases} \end{aligned}$$

Recall if the characteristic function is integrable, then

$$p_{\mathcal{S}(\alpha,\beta,\mu,\sigma)}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-itx}\,\varphi_{\mathcal{S}(\alpha,\beta,\mu,\sigma)}(t)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{(\mu - x)it - |\sigma t|^{\alpha}\left(1 - i\beta\frac{t}{|t|}\Phi\right)}\,dt$$

Again, to solve we can try:

- Take derivative explicitly, solve for zeros (problem of integral taking too long to converge, possibly diverging, or being 'stiff')
- Use Taylor series expansion, take derivative, solve for zeros
- Numerically solve via some gradient-descent type algorithm
- Simulate using Law of Large Numbers (Use Chambers 1976, for simulating α-stable random variables)
- DPP?

Note the tail property of α-stable distributions, namely

$$p_{\mathcal{S}(\alpha,\beta,\mu,\sigma)}(x) = p_{\mathcal{S}}(x;\alpha,\beta,\mu,\sigma) \underset{|x|\to\infty}{\sim} \frac{C}{|x|^{1+\alpha}} \left(= \frac{C}{|x|^{d+\alpha}}, x \in \mathbb{R}^d \right)$$

Now consider the integrands behavior over different regions for large enough K,

$$\int_{-\infty}^{\infty} = \int_{-\infty}^{-K} + \int_{|x|\le K} + \int_{K}^{\infty}$$

What happens to the integrand at large values of x? For $f = 1$, we get $\ln(1 + f(\mathrm{e}^x - 1)) = x$. For $f \in (0,1)$,

$$\lim_{x\to\infty} \frac{\ln(1 + f(\mathrm{e}^x - 1))}{fx} \overset{L'H}{=} \lim_{x\to\infty} \frac{\mathrm{e}^x}{1 + f(\mathrm{e}^x - 1)} \overset{L'H}{=} \frac{1}{f}$$

$$\implies \ln(1 + f(\mathrm{e}^x - 1)) \underset{x\to\infty}{\sim} x$$

(plot this relationship)

In progress...

3.4.10 Log Tempered Stable (Gross) Returns

In progress...

3.4.11 Log (Gross) Factor Model Returns

Let $P = \{P_k\}$ be the "price process" of some random outcome we are betting on. Define the "net-returns" as $r_k = \frac{P_k - P_{k-1}}{P_{k-1}}$.
Naively, one can assume returns are completely random and none of the possible outcomes are "explainable" in which case we would model the log-gross returns as a random variable.

$$R_k = \ln\left(\frac{P_k}{P_{k-1}}\right) = \varepsilon_k$$

Which is effectively what was done in the previous sections for various distribution.
We can think of this random variable as encapsulating every possible reason about why and how the returns vary, and that we don't know any of them, or at least can't measure them.
But if we thought we did know, at least loosely, some of the reasons why the returns may vary, we could try to "pop" this reason out of ε and relate it to the returns in some functional form.
In other words, let's assume that returns aren't completely random, but somehow depend on a particular "factor" F (feature/co-variate), and further assume, for simplicity, returns depend on this factor linearly.

$$R_k = \ln\left(\frac{P_k}{P_{k-1}}\right) = \alpha_k + \beta_k F_k + \varepsilon_k$$

Repeating this process, we "pop" out various explainable/systematic reasons, called factors, from the return's noise, called "idiosyncratic" risks, and end up with the model below.
Assume the log-(gross)returns follow a factor model. The factors could be macroeconomic, fundamental, or statistical factors.

- Macroeconomic factors are observable economic and financial time series (observable factors)
- Fundamental factors are created from an observable asset characteristics, in other words, summary statistics of observable factors (judgement factors according to Tukey)
- Statistical factors are unobservable and extracted from asset returns by some statistical/data based procedure

That is,

$$R_k = \ln\left(\frac{P_k}{P_{k-1}}\right) = \alpha_k + \beta_{1,k}F_{1,k} + \cdots + \beta_{N,k}F_{N,k} + \varepsilon_k \iff r_k \overset{d}{=} r \sim \mathrm{e}^{\text{some distribution}} - 1$$

Where the factors follow particular distributions.

Chapter 4

Continuous Time Wealth Dynamics Under Levy Processes

4.1 Why Go To Continuous Time?

Explain uses of continuous time in finance. A common belief is continuous time is used to use the tools of stochastic calculus. For continuous time in gambling, see "Continuous-Time Casino Problems' by Pestien and Sudderth.Note we need to consider transaction costs and taxes!

4.2 From Discrete to Continuous Time

Assume we are gambling/investing/betting on a sequence of payoffs/returns at an arbitrary (deterministic) frequency/rate n per period.
Further assume the sequence of payoffs/returns that we're gambling/investing/betting in are independent and identically. In particular,

$$\forall n, \{r_{n,k}\}_{k=1}^{n} \quad \text{iid}, \qquad r_{n,k} \overset{d}{=} r_n \geq -1$$

Remark. *NOTE: We are first assuming $\{r_{n,k}\}$ i.i.d. (random walk), which should lead to Levy processes. Next, we should weaken i.i.d and assume $\{r_{n,k}\}$ independent, but not identically distributed, which should lead to additive or non-stationary independent increment processes. This would partially answer a question posed by Thorp. In this scenario, it would make sense for our f to be changing depending on the distribution we're currently betting on, so $\{f_{n,k}\}$. And finally, we would like to assume some dependence structure in the sequence $\{r_{n,k}\}$, for example $Cov(r_{n,k}, r_{n,j}) = \gamma_n(k,j)$, but the dependence doesn't need to be measured by the covariance(could be codifference or covariation for stable distributions). Finally, we would like to do these three scenarios but change the frequency n to be random N_n,i.e. a random time change. This is worked out in the i.i.d. case for the example of the binomial model in (Rachev & Rushendorf, 1994).*

Remark. *Here we only assume $\{r_{n,k}\} \in [-1, \infty)$, so we're making one gamble at a time as opposed to many simultaneous gambles at once, namely $\{\vec{r}_{n,k}\} = [r_{n,k}^1, \cdots, r_{n,k}^d]^\intercal \in [-1, \infty)^d$*

4.2.1 Single Period $[0,1]$

Let $\{W_1^{n,f}\}_{n\geq 1}$ be the wealth dynamics achieved from betting the fraction of current wealth f in the sequence of gambles/returns $\{r_{n,k}\}_{k=1}^n$, and compounding my bankroll/wealth over a single period. In other words, the wealth dynamics are modeling the process of compounding returns with a constant exposure of (proportion) f to those returns.

$$\begin{aligned}
W_0 &= w = \text{initial wealth/bankroll/portfolio size} \\
W_1^{1,f} &= W_0 + r_{n,1}(fW_0) = W_0(1 + fr_{n,1}) \\
W_1^{2,f} &= W_1^{1,f} + r_{n,2}(fW_1^{1,f}) = W_1^{1,f}(1 + fr_{n,2}) = W_0(1 + fr_{n,1})(1 + fr_{n,2}) \\
&\vdots \\
W_1^{n,f} &= W_1^{n-1,f} + r_{n,n}(fW_1^{n-1,f}) = W_1^{n-1,f}(1 + fr_{n,n}) = W_0 \prod_{k=1}^{n}(1 + fr_{n,k})
\end{aligned}$$

Now, we wish to find the dynamics of the limiting wealth variable as the frequency of bets/investments per period gets larger and larger.

$$W_1^{n,f} \xrightarrow[n\to\infty]{d} W_1^{\infty,f} =: W_1^f$$

$$\iff \ln\left(\frac{W_1^{n,f}}{W_0}\right) = \sum_{k=1}^{n} \ln(1 + fr_{n,k}) \xrightarrow[n\to\infty]{d} \ln\left(\frac{W_1^f}{W_0}\right)$$

We assume $f \in [0,1)$ as $f = 1$ is overbetting and will lead to ruin almost surely. Thus $fr_{n,k} > -1 \implies 1+fr_{n,k} > 0 \implies \ln(1 + fr_{n,k}) \in \mathbf{R}$.

Note that we need to prove convergence in distribution and justify that the limit, $\ln(\frac{W_1^f}{W_0})$, is infinitely divisible. Does this imply W_1^f is infinitely divisible? (Attempt to prove using both characteristic function methods and using results from (Jacod & Shiryaev, 2003)) Next, get convergence rates, something like Berry-Esseen, as there is a very practical interpretation to what games you can analyze with the continuous time results. Finally, we need many graphs and simulations showing tractability of these methods, including description and computation of the Levy characteristics of the limit from the approximating sequence.

4.2.2 Arbitrary Period $[0,t]$

Now we consider betting over a time horizon $[0,t]$ or, equivalently, betting over t-many periods, with n being the frequency/rate of bets being placed per period. Let $\{W_t^{n,f}\}_{n\geq 1}$ (note if t is free, we're dealing with a sequence of stochastic processes here as opposed to a sequence of random variables if t were fixed) be the wealth dynamics achieved from betting the fraction of current wealth f in the sequence of gambles/returns $\{r_{n,k}\}_{k=1}^n$, n-many times per period, and compounding my wealth over t-many periods. Let $W_0 = w$ be the initial wealth/bankroll/portfolio size. Then the wealth process given by compounding returns n-many times per period for $\lfloor t \rfloor$-many periods with a constant exposure of f to those returns is given by:

$$W_t^{n,f} = W_{t-1}^{n,f}\left(\prod_{k=1}^{n}(1 + fr_{n,k})\right), \text{ and } W_1^{n,f} = W_0 \prod_{k=1}^{n}(1 + fr_{1,k})$$

$$\implies W_t^{n,f} = W_0\left(\prod_{k=1}^{n}(1+fr_{n,k})\right)^{\lfloor t\rfloor} = W_0\prod_{k=1}^{\lfloor nt\rfloor}(1+fr_{n,k})$$

This time, instead of finding the limiting wealth variable, we wish to find the limiting wealth process as the frequency of my bets/trades per period gets larger and larger. That is,

$$W_t^{n,f} \xrightarrow[n\to\infty]{\mathcal{L}} W_t^{\infty,f} =: W_t^f \quad \text{in } \mathbf{D}[0,\infty)$$

$$\iff \ln\left(\frac{W_t^{n,f}}{W_0}\right) = \sum_{k=1}^{\lfloor nt\rfloor}\ln(1+fr_{n,k}) =: \sum_{k=1}^{\lfloor nt\rfloor}X_{n,k}^f \xrightarrow[n\to\infty]{\mathcal{L}} \ln\left(\frac{W_t^f}{W_0}\right) \quad \text{in } \mathbf{D}[0,\infty)$$

Again, note that we need to prove this convergence in the space of cadlag functions and also verify $\ln(\frac{W_t^{n,f}}{W_0})$ being a Levy process implies the actual wealth process is also Levy. One can hope to prove this convergence by the trick in section 2 of (Rachev & Rushendorf, 1994), or by using tightness, fdd convergence and characterization of limit by fdd. Or finally, what I am thinking is the most tractable as characteristics of Levy processes are supposedly tractable by FFT methods, prove tightness, convergence of triples, and characterization of limiting process by triples.
After establishing this, one would hope to get convergence rates that are uniform in t, and finally do many simulations.
Question: Should the sequence of returns depend on the time horizon? Maybe realistically, but for this model we assume each period, with number of gambles per period being fixed, we're betting on the same (statistically) sequence of gambles/investments.

Now that we get the continuous time wealth process, we should apply Ito's formula (for Levy processes) to get the dynamics to see what SDE the wealth process satisfies and one could use tools of stochastic calculus to help answer some of the above and below questions, as well as for simulation.

Remark. *Here we list some examples that quickly come to light which fall under our setting of* $\forall n$, $\{r_{n,k}\}$ *i.i.d., and* $r_{n,k} \geq -1$. *First we can consider a biased coin,* $r_{n,k} \overset{d}{=} r = \begin{cases} 1, & w/p\ p \\ -1, & w/p\ q \end{cases}$ *with* $p > q$. *Or more generally, a binary payoff, which could model sports betting, loans with simple terms (contractually simple loans), etc.,* $r_{n,k} \overset{d}{=} r = \begin{cases} b, & w/p\ p \\ -1, & w/p\ q \end{cases}$. *These were worked out in (E. O. Thorp, 2006), and also above.*

We could also consider betting on a financial product with some price or value process. Specifically, we would be placing bets on the random returns of this process. In this scenario, we have

$$r_{n,k} = \frac{P_{k/n} - P_{(k-1)/n}}{P_{(k-1)/n}} = e^{R_{k/n}-R_{(k-1)/n}} - 1 \overset{d}{=} e^{R_{1/n}} - 1 =: r_n \qquad \geq -1$$

Finally, we also have the binomial model as a primary example, and can be used in numerical approximation.

4.3 Obtain Limit Using Binomial Tree (Discrete Time, Discrete Space) Approximations to Continuous Time

Here, we initially follow (E. O. Thorp, 2006).
It suffices to use continuous time dynamics when we can think of our wealth at any point in time as the result of compounding many bets. Note, compounding bets could simply be thought of as readjusting portfolio holdings based off of a proportional strategy.
It is of interest to determine how quickly the discrete approximation converges to the continuous time wealth process to give us a precise gauge of about how many bets/readjustments are needed to use the continuous time dynamics(or at least the frequency of bets per unit time, whatever time may be: seconds, minutes, hours, days, weeks, months, years) (possibly see papers by Remi and Fanhui to get convergence rates of discretized SDEs).
Fix an arbitrary $t \in [0, \infty)$, and let n be the frequency/rate of bets placed in the window $[0, \lfloor nt \rfloor]$. In particular, n is the number of bets per unit time (i.e. in $[0, 1]$), and nt is the number of bets per time t.

Let $\mu \in \mathbf{R}$, $\sigma > 0$, $\{Y_k\}$ iid Bernoulli($\frac{1}{2}$),

$$r_{n,k} = \begin{cases} \frac{\mu}{n} + \frac{\sigma}{\sqrt{n}} & \text{with probability} \frac{1}{2} \\ \frac{\mu}{n} - \frac{\sigma}{\sqrt{n}} & \text{with probability} \frac{1}{2} \end{cases} = \left(\frac{\mu}{n} + \frac{\sigma}{\sqrt{n}} \right)^{Y_k} \left(\frac{\mu}{n} - \frac{\sigma}{\sqrt{n}} \right)^{1-Y_k}$$

Note $r_{n,k} = \frac{\mu}{n} + \zeta_{n,k}$ where $\zeta_{n,k} = (\frac{\sigma}{\sqrt{n}})^{Y_k}(-\frac{\sigma}{\sqrt{n}})^{1-Y_k}$

Assume we have an edge (note that this is the hard part to find in practice),

$$\mathbf{E}[r_{n,k}] = \frac{\mu}{n} > 0 \iff \mu > 0$$

Consider the approximation $\forall t$: (note we need to prove this convergence in law on the space $\mathbf{C}$ or $\mathbf{D}$ using methods, like characteristics, in (Jacod & Shiryaev, 2003))

$$W_t^{n,f} = W_0 \prod_{k=1}^{\lfloor nt \rfloor} (1 + f r_{n,k}) \xrightarrow[n \to \infty]{d} W_t^f$$

Equivalently, consider, $\forall t$, the log returns of wealth:

$$
\begin{aligned}
\ln\left(\frac{W_t^{n,f}}{W_0}\right) &= \sum_{k=1}^{\lfloor nt \rfloor} \ln(1+fr_{n,k}) \\
&= \sum_{k=1}^{\lfloor nt \rfloor} (fr_{n,k} - \frac{f^2}{2} r_{n,k}^2 + H.O.T) \\
&= (f\mu t) + f\sum_{k=1}^{\lfloor nt \rfloor} \zeta_{n,k} - \frac{f^2}{2}(\frac{\mu^2 t}{n} + 2\frac{\mu}{n}\sum_{k=1}^{\lfloor nt \rfloor} \zeta_{n,k} + \sigma^2 t) + H.O.T. \\
&\xrightarrow[n\to\infty]{d} f\mu t + f\sigma B_t - \frac{f^2}{2}\sigma^2 t = f(\mu t + \sigma B_t) - \frac{f^2}{2}\sigma^2 t =: \ln\left(\frac{W_t^f}{W_0}\right)
\end{aligned}
$$

$$
\iff W_t^{n,f} \xrightarrow[n\to\infty]{d} W_t^f = W_0\, \mathrm{e}^{f(\mu t + \sigma B_t) - \frac{f^2}{2}\sigma^2 t} = W_0 \mathcal{E}(f(\mu t + \sigma B_t))
$$

Where $\{B_t\}$ is Brownian motion, and $\mathcal{E}$ is the stochastic exponential operator, see (Jacod & Shiryaev, 2003) for relation between stochastic exponential and ordinary exponential.

$$
Y = \mathcal{E}(X) \iff dY_t = Y_{t-} dX_t
$$

To justify the above convergence in distribution, consider

$$
\varphi_{\sum_{k=1}^{\lfloor nt \rfloor} \zeta_{n,k}}(u) = (\varphi_{\zeta_{n,k}}(u))^{\lfloor nt \rfloor} = [\cos(\frac{u\sigma}{\sqrt{n}})^n]^t \xrightarrow[n\to\infty]{} \mathrm{e}^{-t(\frac{u^2\sigma^2}{2})} = \varphi_{\sigma B_t}(u)
$$

To reiterate the interpretation, note, $W_t^{n,f}$ is the wealth/bankroll/portfolio value corresponding to compounding at frequency/rate n per period, for t many periods, at (random) rate of return $\ln(\frac{W_t^{n,f}}{W_0})$. Then, as we send $n \to \infty$, i.e. the frequency of betting per period is getting larger and larger, we get W_t^f is the wealth/bankroll/portfolio value corresponding to continuously compounding over t many periods at random rate of return $\ln(\frac{W_t^f}{W_0})$.
Applying Ito's formula to the limiting wealth process gives us the linear (controlled) SDE:

$$
\begin{cases} dW_t^f = fW_t^f(\mu dt + \sigma dB_t) \\ W(0) = W_0 \end{cases}
$$

To extend this idea to more general distributions, not just lognormal, we follow (Rachev & Rushendorf, 1994) which characterizes all the infinitely divisible limits of the binomial model when you randomize some of the parameters. In particular, the frequency (number) of bets per period, and the up and down sizes of the returns.
Let $r_{n,k} = \begin{cases} u_n & \text{w.p. } p_n \\ d_n & \text{w.p. } q_n \end{cases} = (u_n)^{Y_k}(d_n)^{1-Y_k}$
Where $\{Y_k\}$ are iid Bernoulli(p_n) random variables.
Again, we consider $W_t^{n,f} = W_t^{n-1,f}(1 + fr_{n,\lfloor nt \rfloor}) = W_0 \prod_{k=1}^{\lfloor nt \rfloor}(1 + fr_{n,k})$

$$\implies \ln\left(\frac{W_t^{n,f}}{W_0}\right) = \sum_{k=1}^{\lfloor nt\rfloor} \ln(1+fr_{n,k}) =: \sum_{k=1}^{\lfloor nt\rfloor} X_{n,k}^f$$

Where

$$X_{n,k}^f := \ln(1+fr_{n,k}) = (Y_k)\ln(1+fu_n) + (1-Y_k)\ln(1+fd_n) \qquad 1 \le k \le \lfloor nt\rfloor$$

We need to make some assumptions on our up and down probabilities, namely, we need our summands to be uniformly asymptotically negligible (UAN), so we are actually looking in the class of infinitely divisible limits of the binomial model. In particular, $\forall \varepsilon > 0, \max_{1\le k\le\lfloor nt\rfloor} \mathbf{P}(|X_{n,k}^f| > \varepsilon) \underset{n\to\infty}{\longrightarrow} 0$. The condition of UAN puts conditions on the limit of the sequence of ups and downs, $\{u_n\}$ and $\{d_n\}$, by definition of $X_{n,k}^f$.

We need to prove

$$\ln\left(\frac{W_t^{n,f}}{W_0}\right) = \sum_{k=1}^{\lfloor nt\rfloor} X_{n,k}^f \xrightarrow[n\to\infty]{\mathcal{L}} \ln\left(\frac{W_t^f}{W_0}\right) \text{ on the space } \mathbf{D}([0,\infty))$$

NOTE: First, to prove the above convergence in distribution, for fixed $t \in [0,\infty)$, we need the tools of characteristic functions and infinitely divisible distributions. Second, to prove this convergence in law as random elements of function space, we can use two methods as described in the introduction of (Jacod & Shiryaev, 2003). The traditional Skorokhod way is done in section 2 in (Rachev & Rushendorf, 1994). We need to establish this convergence using the convergence of the characteristics as described in (Jacod & Shiryaev, 2003). Finally consider a random time change and check if it falls into the framework of the above sections.

(In process of adding computer simulations of these results)

4.4 Obtain Limit Using Continuous Time, Continuous Space (Stochastic Calculus) Approximations to Continuous Time

Here we make the (large) assumptions from the outset that we are continuously compounding and that we know the type of process we're betting on.

We can use stochastic exponential Levy models and relies more on the theory of stochastic calculus and methods of (Jacod & Shiryaev, 2003). Note the stochastic exponential can be non-positive, so one needs to ensure that this doesn't happen. Otherwise, we can use ordinary exponential Levy models which makes the stochastic calculus more difficult, though it makes more sense from a modeling and estimating from discrete returns point of view. Moreover, it turns out these are equivalent. See chapter 4 of (Barndorff-Nielsen & Shiryaev, 2010).

Assume a price process, for which we're betting/gambling on the incremental returns, follows an ordinary Levy exponential, that is, the log gross returns can be modeled by a Levy process R with characteristics (b, c, F),

$$P_t = P_0\, \mathrm{e}^{R_t}$$

and thus a Levy-Ito decomposition of

$$R_t = \underbrace{bt + \sqrt{c}W_t}_{\text{Brownian motion with drift}} + \underbrace{x\mathbb{1}_{\{|x|\le 1\}} * (\mu^R - \mathrm{mF})_t}_{\mathbb{L}^2\text{-martingale with countably many jumps}} + \underbrace{x\mathbb{1}_{\{|x|>1\}} * \mu_t^R}_{\text{compound Poisson process}}$$

Equivalently,

$$dP_t = P_{t-}(dR_t + \frac{c}{2}dt + \mathrm{e}^{\Delta R_t} - 1 - \Delta R_t)$$

with an initial price P_0. Note that $\mu^R(\omega; ds, dx) = \sum_{s>0} \delta_{(s,\Delta R_s(\omega))}(ds, dx)$ is the jump measure of the process R (Eberlein, 2009).

Consider compounding at frequency n over a period $[0, t]$. The relative returns (rate of return/net returns) are

$$r_{n,k} := \frac{P_{k/n} - P_{(k-1)/n}}{P_{(k-1)/n}} = \mathrm{e}^{R_{k/n} - R_{(k-1)/n}} - 1$$

Assume we have an edge,

$$\mathbf{E}[r_{n,k}] = \mathbf{E}\Big[\mathrm{e}^{R_{1/n}} - 1\Big] > 0$$

This leads to a wealth process

$$W_t^{n,f} = W_0 \prod_{k=1}^{\lfloor nt \rfloor}(1 + f r_{n,k}) = W_0 \prod_{k=1}^{\lfloor nt \rfloor}\left(1 + f\frac{P_{k/n} - P_{(k-1)/n}}{P_{(k-1)/n}}\right)$$

$$\iff \ln\left(\frac{W_t^{n,f}}{W_0}\right) = \sum_{k=1}^{\lfloor nt \rfloor} \ln\left(1 + f\frac{P_{k/n} - P_{(k-1)/n}}{P_{(k-1)/n}}\right) = \sum_{k=1}^{\lfloor nt \rfloor} \ln\left(1 + f\Big(\mathrm{e}^{R_{k/n} - R_{(k-1)/n}} - 1\Big)\right)$$

Remark. *Note here we embed this discrete problem into a continuous time framework to use the tools of stochastic calculus for Levy processes*

For $\frac{k-1}{n} \le s \le \frac{k}{n}$, set $r_s^{n,k} := \mathrm{e}^{R_s - R_{(k-1)/n}} - 1$, which is the relative return at time s on the k^{th} bet, while compounding at frequency n per period. Then by Ito's formula,

$$\begin{aligned}
\ln\left(1 + f\Big(\mathrm{e}^{R_s - R_{(k-1)/n}} - 1\Big)\right) &=: \ln(1 + f r_s^{n,k}) \\
&= \int_{\frac{k-1}{n}}^{s} \frac{f(r_{u-}^{n,k} + 1)}{1 + f r_{u-}^{n,k}} dR_u + \frac{c}{2} f(1-f) \int_{\frac{k-1}{n}}^{s} \frac{(r_{u-}^{n,k} + 1)}{(1 + f r_{u-}^{n,k})^2} du \\
&+ \int_{\frac{k-1}{n}}^{s} \int_{\mathbf{R}} \left[\ln(1 + f(\mathrm{e}^{R_{u-} + x - R_{(k-1)/n}} - 1)) - \ln(1 + f r_{u-}^{n,k}) - f x \frac{r_{u-}^{n,k} + 1}{1 + f r_{u-}^{n,k}}\right] \mu^R(du, dx)
\end{aligned}$$

Now, let $s = \frac{k}{n}$, then

$$\begin{aligned}
\ln\left(\frac{W_t^{n,f}}{W_0}\right) &= \sum_{k=1}^{\lfloor nt \rfloor} \ln(1 + f\Big(\mathrm{e}^{R_{k/n} - R_{(k-1)/n}} - 1\Big)) \\
&= \sum_{k=1}^{\lfloor nt \rfloor}\left(\int_{\frac{k-1}{n}}^{\frac{k}{n}} h_1(\frac{k-1}{n}, R_{s-}) dR_s + \int_{\frac{k-1}{n}}^{\frac{k}{n}} h_2(\frac{k-1}{n}, R_{s-}) ds + \int_{\frac{k-1}{n}}^{\frac{k}{n}} \int_{\mathbf{R}} h_3(\frac{k-1}{n}, x, R_{s-}) \mu^R(ds, dx)\right) \\
\implies \ln\left(\frac{W_t^{n,f}}{W_0}\right) &= \int_0^t H_1^n(s, R_{s-}) dR_s + \int_0^t H_2^n(s, R_{s-}) ds + \int_0^t \int_{\mathbf{R}} H_3^n(s, x, R_{s-}) \mu^R(ds, dx)
\end{aligned}$$

Where

$$H_i^n(s, y) := \sum_{k=1}^{\lfloor nt \rfloor} h_i(\frac{k-1}{n}, y) \mathbb{1}_{\left[\frac{k-1}{n}, \frac{k}{n}\right)}(s), \quad \text{for } i = 1, 2$$

and

$$H_3^n(s,x,y) = \sum_{k=1}^{\lfloor nt \rfloor} h_3(\frac{k-1}{n},x,y)\mathbb{1}_{\left[\frac{k-1}{n},\frac{k}{n}\right)}(s)$$

Now we need to prove convergence of three stochastic integrals by showing $\forall i \in \{1,2,3\}$, $H_i^n \underset{n\to\infty}{\longrightarrow} H_i$ implies the corresponding stochastic integrals converge in some sense, most likely in probability. Remi suggests trying Lenglart's inequality to get convergence in probability or convergence in expectation with appropriate moment assumptions. Also, see chapter 4 of Applebaums book.

Lemma 1. *Lenglart's Inequality:*
Let X be a cadlag, adapted process and A an increasing process such that $\mathbf{E}[X_t] \le \mathbf{E}[A_t], \quad \forall t$.
Then $\forall t, \epsilon, \eta > 0$, we have

- *A predictable* $\implies \mathbf{P}(\sup_{s\le t}|X_s| \ge \epsilon) \le \frac{\eta}{\epsilon} + \mathbf{P}(A_t \ge \eta)$

- *A adapted* $\implies \mathbf{P}(\sup_{s\le t}|X_s| \ge \epsilon) \le \frac{1}{\epsilon}\left[\eta + \mathbf{E}[\sup_{s\le t}\Delta A_s]\right] + \mathbf{P}(A_t \ge \eta)$

I claim that (need to prove)

$$\ln\left(\frac{W_t^{n,f}}{W_0}\right) \underset{n\to\infty}{\xrightarrow{uc\mathbf{P}}} fR_t + \frac{1}{2}f(1-f)[R^c,R^c]_t + \int_0^t \int_{\mathbf{R}} \left[\ln(1+f(\mathrm{e}^x - 1)) - fx\right]\mu^R(ds,dx) =: \ln\left(\frac{W_t^f}{W_0}\right)$$

Thus we get the continuous time wealth dynamics, when we're continuously betting/trading (re-adjusting portfolio) on the increments of the price, for a price process following a log-Levy process.

$$\implies W_t^f := W_0\,\mathrm{e}^{fR_t + \frac{1}{2}f(1-f)[R^c,R^c]_t + \int_0^t \int_{\mathbf{R}} \left[\ln(1+f(\mathrm{e}^x - 1)) - fx\right]\mu^R(ds,dx)}$$

One question needing to be answered is what are the characteristics of the continuous time wealth process? If not, we should be able to track the characteristic starting from the price process and see how a C^2 mapping of them alters the characteristics to get the wealth process. See appendix in Kallsen and Kallenberg paper.
Note, for $Q = Q(\omega,s,x)$, we have equivalent notation for the stochastic integral of Q with respect to some random measure (in particular the jump measure of a process $\mu^X(\omega;ds,dx) := \sum_{s>0}\delta_{(s,\Delta X_s(\omega))}(ds,dx)$). $\int_0^t \int_{\mathbf{R}} Q(\omega,s,x)\mu^X(\omega;ds,dx) = (Q * \mu^X)_t(\omega) = \sum_{s\le t} Q(\omega,s,\Delta X_s(\omega))$. See (Jacod & Shiryaev, 2003)

Now that we have conjectured the continuous time wealth dynamics, corresponding to continuous compounding, we apply Ito's formula to see what type of SDE it solves. Recall, for a Levy process X, we have

$$d(\mathrm{e}^{X_t}) = \mathrm{e}^{X_{t-}}(dX_t + \frac{1}{2}d[X^c,X^c]_t + \mathrm{e}^{\Delta X_t} - 1 - \Delta X_t)$$

$$\implies dW_t^f = fW_{t-}^f\big(dR_t + \frac{1}{2}d[R^c,R^c]_t + \mathrm{e}^{\Delta R_t} - 1 - \Delta R_t\big)$$

Remark. *This agrees with a quick heuristic sanity check. Recall*

$$W_t^f = W_{t-\delta}^f + fW_{t-\delta}^f\frac{P_t - P_{t-\delta}}{P_{t-\delta}} \underset{\delta\to 0}{\approx} dW_t^f = fW_{t-}^f\frac{dP_t}{P_{t-}}$$

Assume a price process, which we're betting/gambling on the returns or investing in, follows a stochastic exponential, $P_t = P_0\mathcal{E}(\tilde{R})_t$, for $\tilde{R}$ a Levy process $\iff dP_t = P_{t-}d\tilde{R}_t$, with initial price P_0. Note the relation between ordinary exponential and stochastic exponential, $\mathrm{e}^{R_t} = \mathcal{E}(\tilde{R})_t$, where $\tilde{R}_t = R_t + \frac{1}{2}[R^c, R^c]_t + \sum_{s\le t} \mathrm{e}^{\Delta R_s} - 1 - \Delta R_s$, for more details, see either (Jacod & Shiryaev, 2003), or (Cont & Tankov, 2004). By these results, the two methods should be equivalent with the theorem relating ordinary versus stochastic exponential serving as the bridge. Then consider the compounding at frequency n, and set $r_{n,k} := \dfrac{P_{k/n} - P_{(k-1)/n}}{P_{(k-1)/n}}$,

$$W_t^{n,f} = W_0 \prod_{k=1}^{\lfloor nt \rfloor}(1 + f r_{n,k}) = W_0 \prod_{k=1}^{\lfloor nt \rfloor}\left(1 + f\frac{P_{k/n} - P_{(k-1)/n}}{P_{(k-1)/n}}\right)$$

$$\iff \ln\Big(\frac{W_t^{n,f}}{W_0}\Big) = \sum_{k=1}^{\lfloor nt \rfloor}\ln(1 + f\frac{P_{k/n} - P_{(k-1)/n}}{P_{(k-1)/n}})$$

Then, $W_t^{n,f} \underset{n\to\infty}{\longrightarrow} W_0\mathcal{E}(f\tilde{R})_t =: W_t^f$ in the sense of uniform convergence in probability.
This implies $dW_t^f = fW_{t-}^f d\tilde{R}_t$
See (Kardaras & Platen, 2013).
I also claim that $\mathcal{E}(f\tilde{R})_t = \mathrm{e}^{fR_t + \frac{1}{2}f(1-f)[R^c,R^c]_t + \int_0^t\int_{\mathbf{R}}\left[\ln(1+f(\mathrm{e}^x-1)) - fx\right]\mu^R(ds,dx)}$.
Thus these two modeling approaches are equivalent, yet when working with data, one would want to use the ordinary exponential to get the distribution of the log returns as it is interpretable from data.

Goal: Verify this setup and the equivalency of these two methods with a few simple processes with increasing complexity/behavior. First take a (simple) Poisson process as my return process, $R_t = N_t = \sum_{s\le t}\Delta N_s \implies \tilde{R}_t = \sum_{s\le t}\mathrm{e}^{\Delta N_s} - 1$. Next, let Z_t be a process with finite variation and consider a finite variation process with drift $R_t = bt + Z_t$. In particular if $\int_0^t\int_{\mathbf{R}}|x|\mu^R(ds,dx) < \infty$, then as an example, take $Z_t = \int_0^t\int_{\mathbf{R}} x\mu^R(ds,dx) = \sum_{s\le t}\Delta R_s$, so $R_t = bt + \sum_{s\le t}\Delta R_s \implies \tilde{R}_t = bt + \sum_{s\le t}(\mathrm{e}^{\Delta R_s} - 1)$. Finally, we consider a continuous, infinite variation process with drift. Let $R_t = (b - \frac{c}{2})t + \sqrt{c}W_t \implies \tilde{R}_t = bt + \sqrt{c}W_t$. Also, need to determine the decomposition of an alpha stable process, namely its jump measure.

Chapter 5

Kelly Criterion in Continuous Time Under Levy Processes

(Summarize Kelly's conclusion and chapter of book about using it in investing.)

5.1 Kelly Criterion Setup

Now that we have our continuous time wealth dynamics, we would like to see how the wealth dynamics evolve in the long run while continuously having exposure f to the returns.

$$\frac{W_t^f}{W_0} = \mathrm{e}^{t\ln\left(\frac{W_t^f}{W_0}\right)^{\frac{1}{t}}} := \mathrm{e}^{tG_t(f)} = \mathrm{e}^{t\frac{1}{t}\ln\left(\frac{W_t^f}{W_0}\right)} \xrightarrow[t\to\infty]{a.s.} \exp\left(\infty \lim_{t\to\infty}\frac{1}{t}\ln\left(\frac{W_t^f}{W_0}\right)\right) =: \mathrm{e}^{\infty g(f)}$$

We see if we wish to choose a constant exposure f^* which maximizes my long term wealth, we need choose f^* to maximize my long term growth rate, or equivalently maximize the long term log-geometric average. More formally,

$$f^* = \arg\sup_{f\in[0,1)} \lim_{t\to\infty} \frac{W_t^f}{W_0} = \arg\sup_{f\in[0,1)} \lim_{t\to\infty} \frac{1}{t}\ln\left(\frac{W_t^f}{W_0}\right)$$

Set

$$g(f) := \lim_{t\to\infty}\frac{1}{t}\ln\left(\frac{W_t^f}{W_0}\right)$$

Where $g(f)$ is the long term growth rate if we can continuously compound (at least approximately, i.e. $n\to\infty$ is justified) over an infinite (or at least long enough) horizon. Note in a finance context, $n\to\infty$ means continuous trading or continuously updating/re-balancing a portfolio's position.

The continuous time Kelly Criterion is, therefore,

$$sup_{f\in[0,1)}g(f) = \sup_{f\in[0,1)}\lim_{t\to\infty}\frac{1}{t}\ln\left(\frac{W_t^f}{W_0}\right) = \sup_{f\in[0,1)}\mathbf{E}\left[\ln\left(\frac{W_1^f}{W_0}\right)\right]$$

Where the last equality is true for Levy processes with a first moment (essentially SLLN's).

Remark. *Note that we need to establish convergence rates as $t \to \infty$ to tell us how long we have to keep placing our bets for this approximation to be relevant. Another question to be answered is whether or not this strategy is myopic, in other words, is maximizing my growth rate at any horizon/time period the same as maximizing my long term growth rate? For example, is maximizing my wealth over one year the same as maximizing my wealth in the long run? If not, how are they related? More precisely, how is* $\sup_f \mathbf{E}\left[\frac{1}{t}\ln\left(\frac{W_t^f}{W_0}\right)\right]$ *related to* $\sup_f g(f)$*? What's difference between the finite horizon problem and the infinite horizon problem?*

Working in continuous time means we are continuously placing bets, or continuously trading and adjusting our portfolio's exposure. Obviously this is a mathematical idealization. Thus to justify our use of this idealization, we need to verify:

$$\lim_{t\to\infty}\lim_{n\to\infty}\frac{1}{t}\ln\left(\frac{W_t^{n,f}}{W_0}\right) = \lim_{n\to\infty}\lim_{t\to\infty}\frac{1}{t}\ln\left(\frac{W_t^{n,f}}{W_0}\right)$$

The left hand side corresponds to continuous trading and then letting the horizon go to infinity, a mathematical idealization. While the right hand side corresponds to discretely trading, which is what actually happens, and letting the horizon get larger and larger, then letting the frequency of compounding go to infinity. Note that we should verify that $f_n^* \to f^*$, and that $f_t^* \to f^*$.
Set,

$$\begin{aligned}
G_t^n(f) &:= \frac{1}{t}\ln\left(\frac{W_t^{n,f}}{W_0}\right) = \ln\left(\frac{W_t^{n,f}}{W_0}\right)^{\frac{1}{t}} \\
G_t(f) &:= \lim_{n\to\infty} G_t^n \\
G^n(f) &:= \lim_{t\to\infty} G_t^n
\end{aligned}$$

We note that $G_t^n(f)$ is the log-geometric average over horizon $[0,t]$ corresponding to discretely compounding n-many times per period $[0,1]$ with constant exposure of f to the bets/returns. In other words, it is the growth rate at time t corresponding to discretely many compounds. Then $G^n(f)$ is the long term growth corresponding to compounding at rate n per period with exposure of f to the gambles. And $G_t(f)$ is the growth rate at time t corresponding to continuous compounding with exposure of f to the continuous returns (i.e. continuous trading).

Remark. *Note that the conditions $n \to \infty$ and $t \to \infty$ means we are placing lots and lots of bets for a long long time. (Is this reasonable, and if so, what types of gambles? Stock market, debt/bond/mortgage markets, casino games, currencies, making loans?)*

Remark. *Add pictures of the discrete growth rate converging to the continuous growth rate, and then the continuous growth rage converging to the long term growth rate! And compare to the discrete time growth rate functions.*

5.2 Solve Using Binomial Tree Approximations

Follow idea in(E. O. Thorp, 2006). In progress...

5.3 Solve Using Direct Evaluation of Long Run Limit

Recall a price process $P_t = P_0 \,\mathrm{e}^{R_t}$, where R is a Levy process with characteristics (b, c, F) and thus decomposition

$$R_t = bt + \sqrt{c} W_t + \int_0^t \int_{\mathbf{R}} x \mathbb{1}_{|x| \le 1} (\mu^R - mF)(ds, dx) + \int_0^t \int_{\mathbf{R}} x \mathbb{1}_{|x| > 1} \mu^R(ds, dx)$$

Remark. *For this section, as of now, we are assuming all the moments and smoothness we need from the processes involved for the operations we use to make sense. For example, differentiating under integral, finiteness of integrals, or LLN.*

Now, consider the average geometric rate of growth over the period $[0, t]$

$$G_t(f) := \frac{1}{t} \ln\left(\frac{W_t^f}{W_0}\right) = \frac{1}{t}\left(fR_t + \frac{f(1-f)}{2}[R^c, R^c]_t + \int_0^t \int_{\mathbf{R}} \big[\ln(1 + f(\mathrm{e}^x - 1)) - fx\big] \mu^R(ds, dx)\right)$$

Then by a Law of Large Numbers result (need finite mean) for Levy processes, we get convergence in some sense (a.s. or in $\mathbf{P}$)

$$G_t(f) \underset{t\to\infty}{\longrightarrow} f\mathbf{E}\big[R_1\big] + \frac{1}{2} f(1-f)\mathbf{E}\Big[[R^c, R^c]_1\Big] + \mathbf{E}\Big[\int_0^1 \int_{\mathbf{R}} \big[\ln(1 + f(\mathrm{e}^x - 1)) - fx\big] \mu^R(ds, dx)\Big] =: g_R(f)$$

We can express the long term growth rate, defined above, in terms of the characteristics of R, so that

$$g_R(f) = g_R(f; b, c, F) = f\left(b + \int_{\mathbf{R}} x \mathbb{1}_{|x|>1} F(dx)\right) + \frac{c}{2} f(1-f) + \int_{\mathbf{R}} \big[\ln(1 + f(\mathrm{e}^x - 1)) - fx\big] F(dx)$$

The Kelly criterion seeks to maximize this long term growth rate function. Here we assume we can differentiate under the integral and that the integral is finite, which really is a a condition of the small jumps according to F as we've already assumed things about the large jumps when we assumed finite mean.

$$\max_f g_R(f) \implies \frac{d}{df} g_R(f) = \left(b + \int_{\mathbf{R}} x \mathbb{1}_{|x|>1} F(dx)\right) + \frac{c}{2} - fc + \int_{\mathbf{R}} \left[\frac{\mathrm{e}^x - 1}{1 + f(\mathrm{e}^x - 1)} - x\right] F(dx) = 0$$

$$\iff \frac{d}{df} g_R(f) = \left(b + \frac{c}{2}\right) - fc + \int_{\mathbf{R}} \left[\frac{\mathrm{e}^x - 1}{1 + f(\mathrm{e}^x - 1)} - x \mathbb{1}_{|x| \le 1}\right] F(dx) = 0$$

We need to somehow (using a numerical solver) solve this above equation for $f^* = f^*(b, c, F)$ and see how it depends on the the characteristics (b, c, F). We also need to plot $g_R(f; b, c, F)$. In progress...

5.4 Properties of Kelly Criterion in Continuous Time

In progress...

Notes to myself:
Need to investigate more of the math finance literature (as they frequently work in continuous time) such as papers by Markowitz, Merton, Karatzas, Kallsen, Goll, and Kardaras).
If you choose to use this decision criterion, what are some properties of your wealth you can expect to happen? See Ethier's book page 361, (Ethier, 2010), for a CLT-type result for the geometric average growth rate, and try to get this using Discretization of Processes book for the continuous time results.

Part III

Mathematical Formulation of Kelly Criterion Assuming Predictable f Main tool: Dynamic Programming and Optimal Control

Chapter 6

Kelly Criterion in Discrete Time

I need to further investigate ideas shown in (Algoet & Cover, 1988) and generalized in (Algoet, 1994), as well as in (Bellman & Kalaba, 1957).

6.1 Primary Problem

The Kelly Criterion hopes to maximize the "long run growth rate of wealth," which is very much open to interpretation of how we formalize "long run," and "growth rate."

The problem set-up is to consider a wealth process/trajectory $W_n^f := W_0 \prod_{k=1}^{n}(1 + f_{k-1}r_k)$, with a predictable fraction $0 \leq f_{k-1} := f(r_1, \cdots, r_{k-1}) \leq 1, \quad \forall k \geq 1$,
or more generally, $0 \leq f_{k-1} := f_{k-1}(r_1, \cdots, r_{k-1}) \leq 1, \quad \forall k \geq 1$, that is, the function not only depends on your horizon but also on the time/bet number. Also, we define the exponential growth rate (log-geometric average) at bet n as

$$G_n\big(f_0, \cdots, f_{n-1}\big) := \frac{1}{n} \ln\left(\frac{W_n^f}{W_0}\right)$$

Typically the problem is to maximize the **true** long-term growth rate for every possible path (wealth trajectory):

$$\sup_{f_0, f_1, f_2, \cdots} \liminf_{n\to\infty} \frac{1}{n} \ln\left(\frac{W_n^f}{W_0}\right) = \sup_{f_0, f_1, f_2, \cdots} \liminf_{n\to\infty} G_n\big(f_0, \cdots, f_{n-1}\big)$$

(Note there may be a measurability problem with the fractions here?)
If for whatever reason this problem isn't solvable, or is too difficult at first, then one typically chooses to maximize the **expected** long-term growth rate:

$$\sup_{f_0, f_1, f_2, \cdots} \liminf_{n\to\infty} \mathbf{E}\left[\frac{1}{n} \ln\left(\frac{W_n^f}{W_0}\right)\right] = \sup_{f_0, f_1, f_2, \cdots} \liminf_{n\to\infty} \mathbf{E}\Big[G_n\big(f_0, \cdots, f_{n-1}\big)\Big]$$

It may also be the case that this problem isn't solvable, or is too difficult at first, then one typically chooses to first approximate this problem on a large finite horizon, and then subsequently take the limit of the approximate optimal controls and optimization problem, hoping it will equal the optimal control and problem for the infinite horizon case.

6.2 Approximate Problem

Here, we wish to solve the problem of

$$\sup_{f_0^n, f_1^n, \cdots, f_{n-1}^n} \mathbf{E}\left[\frac{1}{n}\ln\left(\frac{W_n^f}{W_0}\right)\right]$$

Equivalently,

$$\sup_{f_0^n, f_1^n, \cdots, f_{n-1}^n} \mathbf{E}\left[\ln(W_n^f)\right] =: V^n(0, W_0) = V_0^n(W_0)$$

With $0 \le f_{k-1}^n := f^n(r_1, \cdots, r_{k-1}) \le 1, \quad \forall 1 \le k \le n,$
or more generally, $0 \le f_{k-1}^n := f_{k-1}^n(r_1, \cdots, r_{k-1}) \le 1, \quad \forall 1 \le k \le n$, that is, the function not only depends on your horizon but also on the time/bet number,
and $W_n^f := W_0 \prod_{k=1}^{n}(1 + f_{k-1}^n r_k)$.

We hope to find that maximizing the true long run growth rate is equivalent (at least under certain models), to maximizing the terminal expected log wealth. If and when we establish this equivalency of the different goals, we can stop trying to use long run limit theorems (such as Law of Large Numbers) and use the tools of dynamic programming and stochastic control. See (Browne & Whitt, 1996).

6.3 Assume $n = 2$, even odds, favorable coin flip

Let $r_k = \begin{cases} 1 & \text{with probability p} \\ -1 & \text{with probability q} \end{cases} = (1)^{Y_k}(-1)^{1-Y_k}$
Where $\{Y_k\}$ are iid Bernoulli(p) random variables.
We also assume we have an edge, i.e. $\mathbf{E}[r_k] > 0$.

(Need to fix and expand on the graphics)

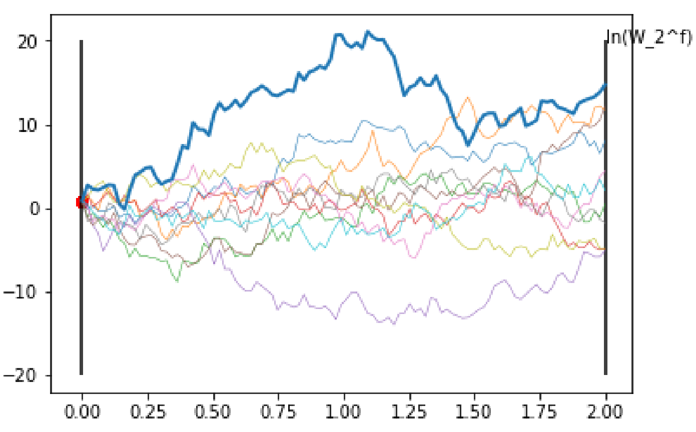


Figure 6.1: Traditional Optimization without DPP

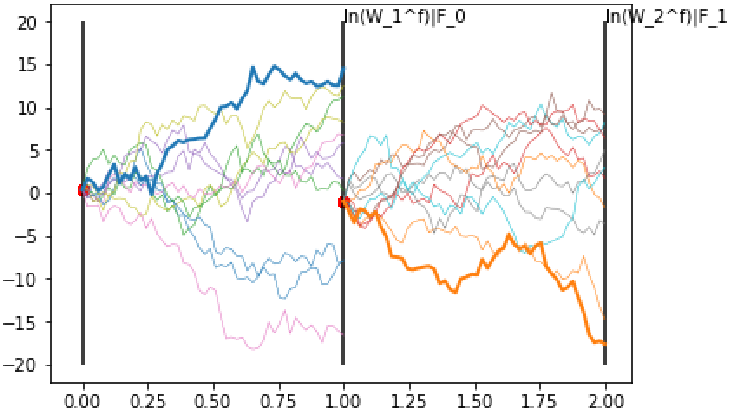


Figure 6.2: Principle of Optimality / DPP

Goal:

$$V^2(0, W_0) := \sup_{f_0^2, f_1^2} \mathbf{E}_{0,W_0}\big[\ln(W_2^f)\big]$$

Also, from here on we drop the superscript of 2.
Consider breaking up the problem into smaller pieces and working backwards, holding the prior pieces fixed! This is essentially the method of dynamic programming.

First we determine what we wish to optimize in the second period, at $t = 1$, given that we're holding the first period arbitrarily fixed:

$$\begin{aligned}
\mathbf{E}_{1,W_1^f}\big[\ln(W_2^f)\big] &= \mathbf{E}_{1,W_1^f}\big[\ln(W_1^f) + \ln(1 + f_1 r_2)\big] \\
&= \ln(W_1^f) + \mathbf{E}_{1,W_1^f}\big[\ln(1 + fr)\big]\Big|_{f=f_1} \\
&=: J(1, W_1^f, f_1)
\end{aligned}$$

Next, we maximize over all paths in the second period, temporarily holding the first period's outcome arbitrarily fixed.

$$\begin{aligned}
\sup_{f_1} \mathbf{E}_{1,W_1^f}\big[\ln(W_2^f)\big] = \cdots &= \sup_{f_1}\Big\{\ln(W_1^f) + \mathbf{E}_{1,W_1^f}\big[\ln(\frac{W_2^f}{W_1^f})\big]\Big\} \\
&= \sup_{f_1}\Big\{\ln(W_1^f) + \mathbf{E}_{1,W_1^f}\big[\ln(1 + fr)\big]\Big|_{f=f_1}\Big\} \\
&=: \sup_{f_1} J(1, W_1^f, f_1) \\
&=: V(1, W_1^f)
\end{aligned}$$

Where we have defined the value after the first period as

$$
\begin{aligned}
V(1, W_1^f) &:= \sup_{f_1} \left\{ \ln(W_1^f) + \mathbf{E}_{1,W_1^f}\left[\ln\left(\frac{W_2^f}{W_1^f} \right) \right] \right\} \\
&= \ln(W_1^f) + \sup_{f_1} \int_{-1}^{\infty} \ln(1 + f_1 x) \mathbf{P}(r \in dx) \\
&= \ln(W_1^f) + p \ln(1 + f_1^*) + q \ln(1 - f_1^*)
\end{aligned}
$$

Where $f_1^* = p - q$.
Next, we back up one step to $t = 0$, take an average, and then maximize over all choices at time zero.

$$
\begin{aligned}
\sup_{f_0} \mathbf{E}_{0,W_0}\left[V(1, W_1^f) \right] &= \sup_{f_0} \mathbf{E}_{0,W_0}\left[\ln(W_1^f) + \sup_{f_1} \mathbf{E}_{1,W_1^f}\left[\ln\left(\frac{W_2^f}{W_1^f} \right) \right] \right] \\
&= \ln(W_0) + \sup_{f_0} \mathbf{E}\big[\ln(1 + fr) \big] \Big|_{f = f_0} + \mathbf{E}\big[\ln(1 + f_1^* r) \big] \\
&= \ln(W_0) + 2\mathbf{E}\big[\ln(1 + f^* r) \big]
\end{aligned}
$$

Where $f^* := f_1^* = f_0^* = p - q$.
The question one must answer is, does the above equation equal the initial optimization problem, in other words does Bellman's 'Principle of Optimality' hold? Mathematically speaking, can we move the sup under the expectation?

$$
\sup_{f_0, f_1} \mathbf{E}[\ln(W_2^f)] := V(0, W_0) \underset{?}{=} \sup_{f_0} \mathbf{E}_{0,W_0}[V(1, W_1^f)]
$$

To prove LHS $\leq$ RHS:

$$
\begin{aligned}
\mathbf{E}[\ln(W_2^f)] &= \mathbf{E}\big[\mathbf{E}_{1,W_1^f}[\ln(W_1^f) + \ln(1 + f_1 r_2)] \big] \\
&= \mathbf{E}\big[\ln(W_1^f) + \mathbf{E}_{1,W_1^f}[\ln(1 + f_1 r_2)] \big] \\
&\leq \mathbf{E}\big[\ln(W_1^f) + \sup_{f_1} \mathbf{E}_{1,W_1^f}[\ln(1 + f_1 r_2)] \big] \\
&= \mathbf{E}\big[V(1, W_1^f) \big] \qquad \text{(depends on} f_0) \\
&\leq \sup_{f_0} \mathbf{E}\big[V(1, W_1^f) \big], \quad \forall (f_0, f_1) \in (0,1) \times (0,1) \\
\text{Take sup of LHS} \implies \sup_{f_0, f_1} \mathbf{E}[\ln(W_2^f)] &=: V(0, W_0) \leq \sup_{f_0} \mathbf{E}_{0,W_0}\big[V(1, W_1^f) \big]
\end{aligned}
$$

To prove RHS $\leq$ LHS:
By definition of $V(1, W_1^f)$, we have $\forall \varepsilon, \quad \exists f_1^\varepsilon \quad$ s.t. $V(1, W_1^f) \leq \ln(W_1^f) + \mathbf{E}_{1,W_1^f}[\ln(1 + f_1^\varepsilon r_2)] + \varepsilon$

$$
\begin{aligned}
\implies \mathbf{E}[V(1, W_1^f)] &\le \mathbf{E}\Big[\ln(W_1^f) + \mathbf{E}_{1,W_1^f}[\ln(1+f_1^\varepsilon)] + \varepsilon\Big] \\
&= \mathbf{E}[\ln(W_1^f(1+f_1^\varepsilon r_2))] + \varepsilon \\
&= \mathbf{E}[\ln(W_2^{f^\varepsilon})] + \varepsilon \\
&\le \sup_{f_0, f_1} \mathbf{E}[\ln(W_2^f)] + \varepsilon \\
&= V(0, W_0) + \varepsilon \qquad \forall f_0
\end{aligned}
$$

Take sup over f_0

$$
\implies \sup_{f_0} \mathbf{E}[V(1, W_1^f)] \le V(0, W_0) + \varepsilon \qquad \forall \varepsilon
$$

Send $\varepsilon \to 0$

$$
\implies \sup_{f_0} \mathbf{E}[V(1, W_1^f)] \le V(0, W_0)
$$

Thus LHS = RHS, and Bellman's 'Principle of Optimality' holds for this particular problem.

6.4 Arbitrary n, Arbitrary Distribution

Recall our goal of finding the 'fraction/proportion/exposure' process $\{f_0^n, f_1^n, \cdots, f_{n-1}^n\}$ that maximizes the 'long run growth rate of wealth,' which we need to show is equivalent the expected long run growth rate of wealth, which we also need to show is approximated by maximizing $\mathbf{E}[\ln(W_n^f)]$. Note for now we don't assume any special structure of $\{r_k\}$ besides $r_k \ge -1$. Later, we'll assume i.i.d.

$$V^n(0,W_0) := \sup_{f_0^n, f_1^n, \cdots, f_{n-1}^n} \mathbf{E}[\ln(W_n^f)] = \sup_{f_0^n, f_1^n, \cdots, f_{n-1}^n} \mathbf{E}\Big[\mathbf{E}_{n-1,W_{n-1}^f}\big[\ln(W_{n-1}^f) + \ln(1 + f_{n-1}r_n)\big]\Big]$$

$$= \sup_{f_0^n, f_1^n, \cdots, f_{n-1}^n} \Big\{\mathbf{E}\Big[\ln(W_{n-1}^f) + \mathbf{E}_{n-1,W_{n-1}^f}\big[\ln(1 + fr_n)\big]\Big|_{f=f_{n-1}^n}\Big]\Big\}$$

$$= \sup_{f_0^n, f_1^n, \cdots, f_{n-1}^n} \Big\{\mathbf{E}\Big[\ln(W_{n-1}^f)\Big] + \mathbf{E}\Big[\mathbf{E}_{n-1,W_{n-1}^f}\big[\ln(1 + fr_n)\big]\Big|_{f=f_{n-1}^n}\Big]\Big\} \qquad \text{then, by the same logic...}$$

$$= \sup_{f_0^n, f_1^n, \cdots, f_{n-1}^n} \Big\{\mathbf{E}\Big[\mathbf{E}_{n-2,W_{n-2}^f}\big[\ln(W_{n-2}^f) + \ln(1 + fr_{n-1})\big]\Big|_{f=f_{n-2}^n}\Big] + \mathbf{E}\Big[\mathbf{E}_{n-1,W_{n-1}^f}\big[\ln(1 + fr_n)\big]\Big|_{f=f_{n-1}^n}\Big]\Big\}$$

$$= \sup_{f_0^n, f_1^n, \cdots, f_{n-1}^n} \Big\{\mathbf{E}\Big[\ln(W_{n-2}^f) + \mathbf{E}_{n-2,W_{n-2}^f}\big[\ln(1 + fr_{n-1})\big]\Big|_{f=f_{n-2}^n}\Big] + \mathbf{E}\Big[\mathbf{E}_{n-1,W_{n-1}^f}\big[\ln(1 + fr_n)\big]\Big|_{f=f_{n-1}^n}\Big]\Big\}$$

$$= \sup_{f_0^n, f_1^n, \cdots, f_{n-1}^n} \Big\{\mathbf{E}\Big[\ln(W_{n-2}^f)\Big] + \mathbf{E}\Big[\mathbf{E}_{n-2,W_{n-2}^f}\big[\ln(1 + fr_{n-1})\big]\Big|_{f=f_{n-2}^n}\Big] + \mathbf{E}\Big[\mathbf{E}_{n-1,W_{n-1}^f}\big[\ln(1 + fr_n)\big]\Big|_{f=f_{n-1}^n}\Big]\Big\}$$

$\vdots$ continue with this logic $\vdots$

$$\begin{aligned} = \sup_{f_0^n, f_1^n, \cdots, f_{n-1}^n} \Big\{ & \ln(W_0) + \mathbf{E}\Big[\mathbf{E}_{0,W_0}\big[\ln(1 + fr_1)\big]\Big|_{f=f_0^n}\Big] + \mathbf{E}\Big[\mathbf{E}_{1,W_1^f}\big[\ln(1 + fr_2)\big]\Big|_{f=f_1^n}\Big] + \cdots \\ & + \mathbf{E}\Big[\mathbf{E}_{n-2,W_{n-2}^f}\big[\ln(1 + fr_{n-1})\big]\Big|_{f=f_{n-2}^n}\Big] + \mathbf{E}\Big[\mathbf{E}_{n-1,W_{n-1}^f}\big[\ln(1 + fr_n)\big]\Big|_{f=f_{n-1}^n}\Big]\Big\} \end{aligned}$$

$$\begin{aligned} = \sup_{f_0^n, f_1^n, \cdots, f_{n-1}^n} \Big\{\mathbf{E}\Big[& \ln(W_0) + \mathbf{E}_{0,W_0}\big[\ln(1 + fr_1)\big]\Big|_{f=f_0^n} + \mathbf{E}_{1,W_1^f}\big[\ln(1 + fr_2)\big]\Big|_{f=f_1^n} + \cdots \\ & + \mathbf{E}_{n-2,W_{n-2}^f}\big[\ln(1 + fr_{n-1})\big]\Big|_{f=f_{n-2}^n} + \mathbf{E}_{n-1,W_{n-1}^f}\big[\ln(1 + fr_n)\big]\Big|_{f=f_{n-1}^n}\Big]\Big\} \end{aligned}$$

$$\text{heuristically} \approx \sup_{f_0^n}\sup_{f_1^n}\cdots\sup_{f_{n-1}^n}\Big\{\mathbf{E}\Big[\cdots\text{same as above}\cdots\Big]\Big\}$$

DPP makes this next heuristic a true equality, which can be proved similarly as n = 2.

$$\approx \sup_{f_0^n}\Big\{\mathbf{E}\Big[\ln(W_0)+\mathbf{E}_{0,W_0}\big[\ln(1+fr_1)\big]\big|_{f=f_0^n}+\sup_{f_1^n}\mathbf{E}_{1,W_1^f}\big[\ln(1+fr_2)\big]\big|_{f=f_1^n}\Big]+\cdots$$

$$+\mathbf{E}\Big[\sup_{f_{n-1}^n}\mathbf{E}_{n-1,W_{n-1}^f}\big[\ln(1+fr_n)\big]\big|_{f=f_{n-1}^n}\Big]\Big\}$$

If identically distributed $\quad r_k \overset{d}{=} r,$ then $\cdots$

$$=\sup_{f_0^n}\Big\{\mathbf{E}_{0,W_0}\Big[\ln(W_0)+\ln(1+fr)\Big]\Big|_{f=f_0^n}+\mathbf{E}\Big[\sup_{f_1^n}\mathbf{E}_{1,W_1^f}\big[\ln(1+fr)\big]\Big|_{f=f_1^n}\Big]+\cdots$$

$$+\mathbf{E}\Big[\sup_{f_{n-1}^n}\mathbf{E}_{n-1,W_{n-1}^f}\big[\ln(1+fr)\big]\big|_{f=f_{n-1}^n}\Big]\Big\}$$

If independent, then $\cdots$

$$\begin{aligned}
&=\sup_{f_0^n}\Big\{\ln(W_0)+\mathbf{E}\Big[\ln(1+fr)\Big]\Big|_{f=f_0^n}+\mathbf{E}\Big[\sup_{f_1^n}\mathbf{E}\big[\ln(1+fr)\big]\Big|_{f=f_1^n}\Big]+\cdots+\mathbf{E}\Big[\sup_{f_{n-1}^n}\mathbf{E}\big[\ln(1+fr)\big]\Big|_{f=f_{n-1}^n}\Big]\Big\}\\
&=\mathbf{E}\Big[\ln(W_0)+\ln(1+f_0^{n*}r)+\mathbf{E}[\ln(1+f_1^{n*}r)]+\cdots+\mathbf{E}[\ln(1+f_{n-1}^{n*}r)]\Big]\\
&=\ln(W_0)+\sum_{k=1}^{n}\mathbf{E}\big[\ln(1+f_{k-1}^{n*}r)\big]\\
&=\ln(W_0)+n\mathbf{E}\big[\ln(1+f^{n*}r)\big]
\end{aligned}$$

Note (see William's book page 93)

$$\mathbf{E}\big[\ln(1+f_t r_{t+1})\big|r_1,\cdots,r_t,W_t^f=x\big]:=\mathbf{E}_{t,x}\big[\ln(1+f_t r_{t+1})\big]=\mathbf{E}\big[\ln(1+fr_t)\big]\Big|_{f=f_t}=\mathbf{E}\big[\ln(1+fr)\big]\Big|_{f=f_t}$$

Where the last two equalities are by independence and by identical distribution of $\{r_k\}$.
Set $g(f):=\mathbf{E}\big[\ln(1+fr)\big]$.

We define

$$J^n(k,x,(f_k^n,\cdots,f_{n-1}^n)):=\mathbf{E}_{k,x}\Big[\ln(x)+\ln\Big(\frac{W_n^f}{W_k^f}\Big)\Big]=\mathbf{E}_{k,x}\Big[\ln(x)+\ln\Big(\prod_{j=k+1}^{n}(1+f_{j-1}^n r_j)\Big)\Big]$$

$$V^n(k,x):=\sup_{f_k^n,f_{k+1}^n,\cdots,f_{n-1}^n}\mathbf{E}_{k,x}\Big[\ln(x)+\ln\Big(\frac{W_n^f}{W_k^f}\Big)\Big]:=\sup_{f_k^n,f_1^n,\cdots,f_{n-1}^n}J^n(k,x,(f_k^n,\cdots,f_{n-1}^n))\quad\forall k\geq 1$$

And recall,

$$V^n(0,x):=\sup_{f_0^n,f_1^n,\cdots,f_{n-1}^n}\mathbf{E}_{0,x}\Big[\ln(x)+\ln\Big(\frac{W_n^f}{W_0}\Big)\Big]=\sup_{f_0^n,f_1^n,\cdots,f_{n-1}^n}\mathbf{E}\big[\ln(W_n^f)\big]$$

We need to prove Bellman's 'Principle of Optimality/DPP' for this particular problem. That is, $\forall k \geq 0$,

$$V^n(k,x) = \sup_{f_k^n} \mathbf{E}_{k,x}\Big[V^n\big(k+1, x\big(1 + f_k^n r_{k+1}\big)\big)\Big]$$

6.5 Convergence of Controls and Value Function

In progress...

6.6 Specific Distribution

In progress...

6.7 Convergence to Continuous Time

In progress... Change $k = t, k+1 = t+\delta$ and take the limit as $\delta \to 0$.

Chapter 7

Continuous Time Wealth Dynamics under Levy Processes

Need to replace the computations in chapter with a predictable "fraction/proportion" process and I think the computations should follow through similarly and the main change is anytime you see an f times any process, e.g. fR_t, it would be replaced by an integral, e.g. $\int_0^t f_{s-} dR_s$

Derive an HJB equation, see Jupyter Notebook.

Chapter 8

Kelly Criterion in Continuous Time Under Levy Processes

Do we need to work in the finite horizon case first and use stochastic control techniques, or can we use LLN's and maximize the limit??
We would be working with stochastic integrals with respect to Lévy processes, I think there should be a long run limit but have yet to work out the details.

Part IV

Mathematical Formulation of Kelly Criterion With Drawdown Constraint Main tool: Dynamic Programming and Optimal Control

Chapter 9

Motivation and History

In progress...

Here, we need to discuss how major drawdowns can affect your wealth aand long term growth/compounding rate. We need to formally flush out the idea that a single ten percent drop is way worse than ten, one percent drops spread out over some time because of the non-linearity of the wealth process. Cite the example of tax hedging in Fortune's Formula book. This idea leads us to wanting to add a drawdown constraint to the Kelly criterion. Show empirical graphs, etc.

There are lots of writings about fluctuation theory and gambling, see (Feller, 1968), (Feller, 1971), (Wilson, 1970), maybe Ziemba's Beat the Racetrack. Basically says your wealth process can experience surprisingly large amounts of fluctuations away from the long term expected value. So one has to be able to weather these possible bad streaks, and take advantage of the good streaks.

Also see Meb Faber's article "A Quantitative Approach to Tactical Asset Allocation" for a discussion on drawdowns and CTA, managed futures and the drawdown constraints.

For example, a 75% drawdown must be followed by a 300% return to break even, or compound at 10% for fifteen years which shows how drawdowns can be disastrous.

For example, let $r_{\text{dradown}} \in (-1, 0)$, most likely something less than -0.30, that is, more than a 30% return

$$(1 + r_{\text{dradown}})(1 + r) = 1$$

Solving this for r will tell you the single period return you need to mitigate this loss of r_{dradown}.
Similarly, it would take T-many years compounding at a rate of r to break even from a drop of r_{dradown}

$$(1 + r_{\text{dradown}})(1 + r)^T = 1$$

Chapter 10

Kelly Criterion in Discrete Time with Drawdown Constraint

10.1 Drawdown Constraint

Remark. *We should use conditional probabilities for our constraint. Also note how drawdowns are like negative rates of return.*

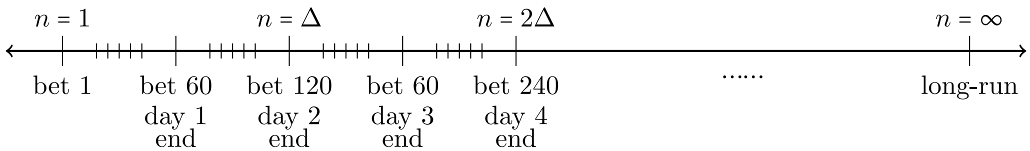
In this section we are trying to capture (formalize) the notion of 'short-term/time' risk while using a long-run growth rate optimization Kelly-type strategy. Thus we are trying to formalize the notion of "fractional Kelly" as described in (E. O. Thorp, 2006). We would hope to 'minimize,' in a sense, the short-term drawdown risk while still maximizing the 'long-term growth rate.'

As a simple motivating narrative, say we've identified a biased coin flipping game where we have an edge. Say we can place bets every minute for one hour a day, so 60 bets a day, and we are trying to limit our two day drawdowns for whatever reason, maybe our investor is demanding a report from us every two days, or maybe we are impatient and can't handle things going too bad for more than two days.

(Note this example ignores liquidity/fat-tail issues, alpha/edge decay, etc.)
(Also note this example should get us to continuous time as normalized sums of coin tosses $\approx$ Brownian Motion, and a random sum of coin tosses, thought of as coin tosses in trading time, $\approx$ alpha-Stable Process (see Rachev '94).)

In this hypothetical, $\delta = 120$ bets, so δ quantifies how often you check your 'drawdown status.'

Note I do not check my 'drawdown status' after every bet ($\forall n$), but I check it after every two days of betting, that is, every bet that is a multiple of Δ ($\forall n = \ell\Delta$, $\forall \ell \in \mathbb{N}$).



This leads to considering different types of drawdown conditions.

Also, let us remark on the question of if we should condition or not on the probabilities of drawdowns, i.e should we use $\mathbf{P}_0$ or $\mathbf{P}_{\ell\Delta}$, $\forall \ell \le L-1$? Note if we do not condition by using $\mathbf{P}_0 \quad \forall \ell \le L-1$, then for any period ℓ, $\exists\ (2^{\Delta})^{\ell+1}$ possible paths that we must average over to compute the drawdown probability at every period our drawdown status is checked. On the other hand, if we do condition by using $\mathbf{P}_{\ell\Delta}$, $\forall \ell \le L-1$, then for any period ℓ, $\exists \quad 2^{\Delta}$ possible paths that we must average over to compute the drawdown probability at every period our drawdown status is checked. Also note, our 'investor' treats $W^f_{\ell\Delta}$ as a constant when computing the drawdown performance, which is in fact conditioning. That is the 'investor' only cares about not losing too much over any period $[\ell\Delta+1, (\ell+1)\Delta]$, for all periods $\ell \le L-1$

Before we formally define the drawdown constraint, let us recall our approximate problem of (large) finite horizon maximization of expected log-wealth subject to a drawdown constraint over all periods that fit into the finite horizon. For simplicity, assume the (large) finite horizon is some multiple of how often the drawdown status is checked, i.e. the horizon is $[0, L\Delta]$ where L is the number of periods you gamble. More precisely,

$$\sup_{\mathcal{A}_L} \mathbf{E}\big[\ln(W^f_{L\Delta})\big]$$

Where the set of admissible controls is all predictable fractions that satisfy the drawdown constraints for all periods of length Δ in the horizon of length $L\Delta$:

$$\mathcal{A}_L := \left\{ \{f^L_{k-1}\}_{k=1}^{L\Delta} \middle| \mathbf{P}_{\ell\Delta}\big(\mathrm{d}_{(\ell+1)\Delta}(f) \le \beta\big) \ge 1-\varepsilon, \quad \forall 0 \le \ell \le L-1, \quad f^L_{k-1} = f^L_{k-1}(r_1, \cdots, r_{k-1}) \in (0,1) \right\}$$

Note we hope to solve the limiting problem by dividing by $L\Delta$ and send $L \to \infty$, hoping the optimal problem and controls converge.
Upon further inspection, we see after appropriate conditioning

$$\begin{aligned}
\mathbf{E}\left[\frac{1}{L\Delta}\ln\left(\frac{W^f_{L\Delta}}{W_0}\right)\right] &= \mathbf{E}\left[\frac{1}{L\Delta}\ln\left(\frac{W^f_{\Delta}}{W_0}\frac{W^f_{2\Delta}}{W^f_{\Delta}}\cdots\frac{W^f_{L\Delta}}{W^f_{(L-1)\Delta}}\right)\right] = \frac{1}{L}\sum_{\ell=0}^{L-1}\mathbf{E}\left[\frac{1}{\Delta}\ln\left(\frac{W^f_{(\ell+1)\Delta}}{W^f_{\ell\Delta}}\right)\right] \\
&= \frac{1}{L}\sum_{\ell=0}^{L-1}\frac{1}{\Delta}\sum_{k=\ell\Delta+1}^{(\ell+1)\Delta}\mathbf{E}\Big[g\big(f^L_{k-1}(r_1,\cdots,r_{k-1})\big)\Big], \quad \text{where} \quad g(f) := \mathbf{E}[\ln(1+fr)]
\end{aligned}$$

Note that $\varepsilon = 1$ corresponds to the typical gambler that doesn't worry about and types of drops in their wealth. And $\varepsilon = 0$ corresponds to the gambler who "buys insurance" to guarantee that no drop in wealth will ever exceed a certain threshold β. This buying of insurance could be through a reduction of their exposure f, or through the systematic purchase of some sort of instruments such as put options, swaps, utilization of stop loss orders, or blackjack insurance. See work on CPPI, constant proportion portfolio insurance (see Theory of constant proportion portfolio insurance by Fisher Black and Perold), which can be though of as reducing risky asset exposure, as opposed to tail risk hedging which can be though of as explicitly buying insurance-like products to be able to take on more risky asset exposure (see Protecting Portfolio Value: Constant Proportion Portfolio Insurance Versus Tail Risk Hedging).

Thus maximizing the finite horizon expected growth rate is equivalent to maximizing the above.
This leads us to the question of **What Definition of Drawdown Should We Be Using?**

10.1.1 Comparing Starting to Ending Value in a Given Period Δ

Remark. *Note that starting of new period is really ending of of previous period, so this should really be called Comparing Ending to Ending Value. For now we drop the superscript off the controls.*

If the person evaluating our performance, could be our emotions (i.e. risk tolerance) or our investor, is considering how much we have at the end of the period of length Δ compared to how much we had at the beginning of the period Δ, they may impose a condition saying they don't want to see, too often, our bankroll (wealth or portfolio-value) drop more than $100\times\beta\%$ of what we started with at the beginning of the period.

Formally, this means: $\forall\Delta\in\mathbb{N}$(in our exapmle = 120) (quantifies how often we check our drawdown status), $\forall\beta\in(0,1)$ (quantifies how large of a drawdown we are allowed), $\forall\varepsilon\in[0,1]$ (quantifies "too often" in the above sentence, i.e. how often we can tolerate violating the β constraint)

$$\begin{aligned}
&\mathbf{P}_n\left(\left(\frac{W_n^f - W_{n+\Delta}^f}{W_n^f}\right)^+ \le \beta\right) \ge 1-\varepsilon \qquad \forall n = \ell\Delta, \quad \forall \ell \le L-1 \\
\iff &\mathbf{P}_{\ell\Delta}\left(\left(\frac{W_{\ell\Delta}^f - W_{(\ell+1)\Delta}^f}{W_{\ell\Delta}^f}\right)^+ \le \beta\right) \ge 1-\varepsilon \qquad \forall \ell \le L-1 \\
\iff &\mathbf{P}_{\ell\Delta}\left(\prod_{k=\ell\Delta+1}^{(\ell+1)\Delta}(1+f_{k-1}r_k) \ge 1-\beta\right) \ge 1-\varepsilon \qquad \forall \ell \le L-1 \\
\iff &\mathbf{P}_{\ell\Delta}\left(D_{\ell\Delta}\right) \ge 1-\varepsilon \qquad \forall \ell \le L-1 \qquad \text{where} \qquad D_{\ell\Delta} := \left\{\prod_{k=\ell\Delta+1}^{(\ell+1)\Delta}(1+f_{k-1}r_k) \ge 1-\beta\right\}
\end{aligned}$$

Remark. *Note in some literature this is called "Shortfall" instead of drawdown. See book by Martin Leibowitz on Shortfall.*

10.1.2 Comparing Starting Value to Minimum in a Given Period δ

Literal Ruin
For any given period you might want to track how much you've dropped from your starting wealth. That is, you are concerned with how much of your starting wealth you might lose. Formally, this means

$$\begin{aligned}
&\mathbb{P}_{\ell\Delta}\left(\frac{W_{\ell\Delta}^f - \inf\limits_{\ell\Delta\le k\le(\ell+1)\Delta} W_k^f}{W_{\ell\Delta}^f} \le \beta\right) \ge 1-\varepsilon \\
\iff &\mathbf{P}_{\ell\Delta}\left(1-\beta \le \inf_{\ell\Delta\le k\le(\ell+1)\Delta}\prod_{i=\ell\Delta+1}^{k}(1+f_{i-1}^L r_i)\right) = \mathbf{P}_{\ell\Delta}\left(\bigcap_{k=\ell\Delta+1}^{(\ell+1)\Delta}\left\{\prod_{i=\ell\Delta+1}^{k}(1+f_{i-1}^L r_i) \ge 1-\beta\right\}\right)
\end{aligned}$$

10.1.3 Maximum Drawdowns in a Given Period Δ

Note this corresponds to the drawdowns reported in managed futures/trend following, CTA (Commodity Trading Advisor) reports.

$$\mathbf{P}_{\ell\Delta}\Bigg(\underbrace{\sup_{\ell\Delta \le j \le (\ell+1)\Delta} \underbrace{\sup_{\ell\Delta \le i \le j} \left[\frac{W_i^f - W_j^f}{W_i^f} \right]^+}_{\text{drawdown at time j}}}_{\text{maximum drawdown over horizon}} \le \beta \Bigg) = \mathbf{P}_{\ell\Delta}\Bigg(\sup_{\ell\Delta \le i < j \le (\ell+1)\Delta} \left[\frac{W_i^f - W_j^f}{W_i^f} \right]^+ \le \beta \Bigg) \ge 1 - \varepsilon, \quad \ell \le L - 1$$

Effective Ruin

Similarly, if the person evaluating our performance is considering how much we have at the end of the period of length Δ compared to the most we had all period Δ, they may impose a condition saying they don't want to see, too often, our bankroll (wealth or portfolio-value) drop more than $100\times\beta\%$ of our largest bankroll over that period.

Formally, this means: $\forall\Delta \in \mathbb{N}$(in our exapmle = 120) (quantifies how often we check our drawdown status), $\forall\beta \in (0,1)$ (quantifies how large of a drawdown we are allowed), $\forall\varepsilon \in [0,1]$ (quantifies "too often" in the above sentence, i.e. how often we can tolerate violating the β constraint)

Let $\beta \in (0,1)$ be the threshold for the relative (percentage) drawdown from the top. Thus $1-\beta$ is the percentage of remaining wealth left after a drawdown, i.e. percentage from the bottom. Now, consider the space of all possible trajectories (of wealth or of returns?) with the relative drawdown (from the top), over any given horizon, bounded by the threshold β:

$$\begin{aligned}
\left\{ \frac{\sup_{n \le j \le n+\Delta} W_j^f - W_{n+\Delta}^f}{\sup_{n \le j \le n+\Delta} W_j^f} \le \beta \right\} &= \left\{ \frac{W_{n+\Delta}^f}{\sup_{n \le j \le n+\Delta} W_j^f} \ge 1 - \beta \right\} = \left\{ \inf_{n \le j \le n+\Delta} \frac{W_{n+\Delta}^f}{W_j^f} \ge 1 - \beta \right\} \\
&= \left\{ \inf_{n \le j \le n+\Delta} \prod_{k=j+1}^{n+\Delta} \left(1 + f_{k-1} r_k\right) \ge 1 - \beta \right\} \\
&= \left\{ \frac{\cancel{W_n^f} \prod_{k=n+1}^{n+\Delta} \left(1 + f_{k-1} r_k\right)}{\cancel{W_n^f} \sup\{1, \left(1 + f_n r_{n+1}\right), \cdots, \prod_{k=n+1}^{n+\Delta} \left(1 + f_{k-1} r_k\right)\}} \ge 1 - \beta \right\} \\
&= \left\{ \prod_{k=n+1}^{n+\Delta} \left(1 + f_{k-1} r_k\right) \ge 1 - \beta \right\} \cap \left\{ \prod_{k=n+2}^{n+\Delta} \left(1 + f_{k-1} r_k\right) \ge 1 - \beta \right\} \cap \cdots \cap \left\{ 1 \ge 1 - \beta \right\} \\
&= \bigcap_{j=1}^{\Delta} \left\{ \prod_{k=n+j}^{n+\Delta} \left(1 + f_{k-1} r_k\right) \ge 1 - \beta \right\}
\end{aligned}$$

We would like to have a low probability of seeing our wealth trajectory drop too much over any given period of predefined length Δ. Thus, we would like to constrain our set of fractions to those that give wealth trajectories a high probability of drawdowns staying below the threshold β, while still maximizing the expected logarithm of wealth.

Constraint: $\forall \beta, \varepsilon \in (0,1), \forall \Delta \in \mathbb{N}$

$$\mathbf{P}_n\left(\frac{\sup\limits_{n\le j\le n+\Delta} W_j^f - W_{n+\Delta}^f}{\sup\limits_{n\le j\le n+\Delta} W_j^f} \le \beta\right) = \mathbf{P}_n\left(\frac{W_{n+\Delta}^f}{\sup\limits_{n\le j\le n+\Delta} W_j^f} \ge 1-\beta\right)$$
$$= \mathbf{P}_n\left(\bigcap_{j=1}^{\Delta}\left\{\prod_{k=n+j}^{n+\Delta}\left(1+f_{k-1}r_k\right) \ge 1-\beta\right\}\right) \ge 1-\varepsilon, \qquad \forall n = \ell\Delta, \quad \ell \le L-1$$

10.2 Problem Formulation

WLOG drop the superscript L and set $W_0 = 1$.
Recall our goal

$$\begin{aligned}\text{maximize: } & \mathbf{E}\left[\ln(W_{L\Delta}^f)\right] \\ \text{subject to: } & \mathbf{P}_{\ell\Delta}\left(\left(\frac{W_{\ell\Delta}^f - W_{(\ell+1)\Delta}^f}{W_{\ell\Delta}^f}\right)^+ < \beta\right) > 1-\varepsilon, \qquad \forall 0 \le \ell \le L-1\end{aligned}$$

From this point, we can reformulate the constraint as a terminal function in the spirit of Lagrange multipliers. Then establish a first order condition, which is highly distributional dependent, and a DPP recursion relation. We then will try to numerically solve with deep reinforcement learning/neuro-dynamic programming.

10.3 What Controls (Fractions) Should We Be Using?

Remark. *This section is mainly about considering different types of controls for any type of drawdown definition. In the next section we consider different types of drawdowns for appropriate controls*

In general, our filtration is generated from the sequence of returns. That is $\mathbb{F} = \mathbb{F}^r = \{\mathcal{F}_n^r\} = \{\sigma(r_1, \cdots, r_n)\}_{n\ge1}$. Also, compare with $\mathbb{F}^{W^f}$. That is, "strong" versus "weak" formulation.

Therefore, in general, $0 \le f_{k-1}^L := f_{k-1}^L(r_1, \cdots, r_{k-1}) \le 1, \quad \forall 1 \le k \le L\delta$

Now why would we want our fraction to depend on the returns or on our wealth? (I need to summarize on the exposition in (E. O. Thorp, 2012)) Note we aren't truly interested in the returns (noise) that we're betting on for their own sake, but as a proxy for the wealth trajectory which we are trying to control.

This rationale leads us to believe our fraction should explicitly depend on our wealth trajectory somehow, and not explicitly on the return trajectory, only implicitly through the wealth trajectory.

10.3.1 Constant Control $\mathcal{A}_L^{\mathbf{const}}$

Assume the fraction (control) is a constant,

$$\forall 1 \le k \le L\delta, \quad f_{k-1}^L(r_1, \cdots, r_{k-1}) = f^L$$

$$\mathcal{A}_L^{\text{const}} := \left\{ f^L \in (0,1) \middle| \mathbf{P}_{\ell\delta}\big(\mathrm{d}_{(\ell+1)\delta}(f) \le \beta\big) \ge 1-\varepsilon, \quad \forall 0 \le \ell \le L-1 \right\}$$

This is the traditional Kelly/Thorp assumption and makes both intuitive and rigorous (can be proved using DPP) sense in the case of i.i.d. returns with no drawdown constraints.

In this case, the **expected** growth rate over a large, but finite, horizon is

$$\begin{aligned}\mathbf{E}\left[\frac{1}{L\delta}\ln\left(\frac{W_{L\delta}^f}{W_0}\right)\right] &= \frac{1}{L\delta}\sum_{k=1}^{L\delta}\mathbf{E}\big[\ln(1+f^L r_k)\big] \\ &= \mathbf{E}\big[\ln(1+f^L r)\big] \quad \text{by i.i.d.}\end{aligned}$$

10.3.2 Constant Control per Period Depending on Initial Period Wealth (Turns out to be same as constant control)

Consider the case where we can only adjust our betting fraction (exposure percentage) before the start of any given betting period, using only knowledge of our "initial wealth" for that period of length δ. Thus, we're using a constant fraction in a given period, though it can change at the start of every period.
Formally,

$$\forall 0 \le \ell \le L-1, \quad \forall (\ell\delta+1) \le k \le (\ell+1)\delta, \quad f_{k-1}^L(r_1,\cdots,r_{k-1}) = f_\ell^L(W_{\ell\delta}^f)$$

Equivalently,

$$f_{k-1}^L(r_1,\cdots,r_{k-1}) = \sum_{\ell=0}^{L-1} f^L(W_{\ell\delta}^f)\mathbb{1}_{[\ell\delta+1,(\ell+1)\delta]}(k)$$

$$\begin{aligned}\mathcal{A}_L^{init} := \Bigg\{ \{f_{k-1}^L\}_{k=1}^{L\delta} \Bigg| \mathbf{P}_{\ell\delta}\big(\mathrm{d}_{(\ell+1)\delta}(f) \le \beta\big) \ge 1-\varepsilon, \quad \forall 0 \le \ell \le L-1, \\ f_{k-1}^L(r_1,\cdots,r_{k-1}) = \sum_{\ell=0}^{L-1} f^L(W_{\ell\delta}^f)\mathbb{1}_{[\ell\delta+1,(\ell+1)\delta]}(k) \in (0,1) \Bigg\}\end{aligned}$$

In this case, the **expected** growth rate over a large, but finite, horizon is

$$\begin{aligned}\mathbf{E}\left[\frac{1}{L\delta}\ln\left(\frac{W_{L\delta}^f}{W_0}\right)\right] &= \frac{1}{L\delta}\sum_{k=1}^{L\delta}\sum_{\ell=0}^{L-1}\mathbf{E}\big[\ln(1+f^L(W_{\ell\delta}^f)r_k)\big]\mathbb{1}_{[\ell\delta+1,(\ell+1)\delta]}(k) \\ &= \cdots\end{aligned}$$

10.3.3 δ-Control Depending on Wealth

(I think this makes the most sense) Assume the fraction only depends on my wealth process/path/trajectory occurring in a given δ-period,

$$\forall 0 \le \ell \le L-1, \quad \forall (\ell\delta+1) \le k \le (\ell+1)\delta, \quad f_{k-1}^L(r_1,\cdots,r_{k-1}) = f_{k-1}^L(W_{\ell\delta}^f,\cdots,W_{k-1}^f)$$

Equivalently,

$$f_{k-1}^L(r_1,\cdots,r_{k-1}) = \sum_{\ell=0}^{L-1} f_{k-1}^L(W_{\ell\delta}^f,\cdots,W_{k-1}^f)\mathbb{1}_{[\ell\delta+1,(\ell+1)\delta]}(k)$$

$$\mathcal{A}_L^{\delta,W^f} := \left\{ \{f_{k-1}^L\}_{k=1}^{L\delta} \middle| \mathbf{P}_{\ell\delta}\big(\mathrm{d}_{(\ell+1)\delta}(f) \le \beta\big) \ge 1-\varepsilon, \quad \forall 0 \le \ell \le L-1, \right.$$

$$\left. f_{k-1}^L(r_1, \cdots, r_{k-1}) = \sum_{\ell=0}^{L-1} f_{k-1}^L(W_{\ell\delta}^f, \cdots, W_{k-1}^f)\mathbb{1}_{[\ell\delta+1,(\ell+1)\delta]}(k) \in (0,1) \right\}$$

In this case, the **expected** growth rate over a large, but finite, horizon is

$$\mathbf{E}\left[\frac{1}{L\delta}\ln\left(\frac{W_{L\delta}^f}{W_0}\right)\right] = \frac{1}{L\delta}\sum_{k=1}^{L\delta}\sum_{\ell=0}^{L-1}\mathbf{E}\big[\ln(1+f_{k-1}^L(W_{\ell\delta}^f,\cdots,W_{k-1}^f)r_k)\big]\mathbb{1}_{[\ell\delta+1,(\ell+1)\delta]}(k)$$

$$=$$

10.3.4 Factor Based Control Depending on Some External Factors

See notes from Chris Jones about portfolio weights depending on some variable that is monotonically related to edge $\mathbf{E}[r]$. For simplicity, one can assume $f = f(\mathbf{E}[r])$ (is this already the Kelly assumption where f depends on the return sequence we're betting on, or some linear (economic) factor model $f = f(\alpha + \sum_{k=1}^N \beta_k \mathbf{E}[F_k])$. More precisely, this is really assuming the returns follow some factor model.

10.3.5 Relations Among Controls

Consider at the filtration level.
Conjecture:

$$\mathbf{F}^r = \mathbf{F}^{W^f}$$

So it doesn't matter which filtration we use.

10.4 Cases

10.4.1 Start-to-End, Constant Control

GOAL:

$$\sup_{\mathcal{A}_L^{\text{const}}} \mathbf{E}[\ln(W_{L\delta}^f)]$$

$$\mathcal{A}_L^{\text{const}} := \left\{ f^L \in (0,1) \middle| \mathbf{P}_{\ell\delta}\big(D_{\ell\delta}\big) \ge 1-\varepsilon, \quad \forall 0 \le \ell \le L-1 \right\}$$

Where $D_{\ell\delta} := \left\{ \prod_{k=\ell\delta+1}^{(\ell+1)\delta}(1+fr_k) \ge 1-\beta \right\}$, and $\{D_{\ell\delta}\}_{\ell=0}^{L-1}$ is a sequence of independent events, since the returns are i.i.d.
Note this is the traditional Kelly/Thorp control space.

Next, the constraint becomes

$$
\begin{aligned}
\mathbf{P}_{\ell\delta}\bigg(D_{\ell\delta}\bigg) &= \mathbf{P}\bigg(D_{\ell\delta}\bigg) \qquad \text{by independence} \\
&= \mathbf{P}\bigg(\prod_{k=1}^{\delta}(1+fr_k) \geq 1-\beta\bigg) \qquad \text{since}(r_1,\cdots,r_\delta) \overset{d}{=} (r_{\ell\delta+1},\cdots,r_{(\ell+1)\delta}) \\
&= \mathbf{P}\bigg(\frac{W_\delta^f}{W_0} \geq 1-\beta\bigg) \\
&= \mathbf{P}\bigg(\ln\frac{W_\delta^f}{W_0} \geq \ln(1-\beta)\bigg) \\
&= \mathbf{P}\bigg(\sum_{k=1}^{\delta}(\ln(1+fr_k) \geq \ln(1-\beta)\bigg) \geq 1-\varepsilon \qquad \forall \ell \leq L-1
\end{aligned}
$$

Thus, in the constant fraction case, the constraint boils down to a constraint on a random walk.

10.4.2 Start-to-End, δ-Control

Assume $\mathbf{E}[r] > 0$, $r \geq -1$, $F(x) := \mathbf{P}(r \leq x)$.
WLOG drop the superscript L and set $W_0 = 1$.
Recall our goal

$$
\begin{aligned}
\text{maximize: } & \mathbf{E}\big[\ln(W_{L\delta}^f)\big] \\
\text{subject to: } & \mathbf{P}_{\ell\delta}\Bigg(\bigg(\frac{W_{\ell\delta}^f - W_{(\ell+1)\delta}^f}{W_{\ell\delta}^f}\bigg)^+ < \beta\Bigg) > 1-\varepsilon, \qquad \forall 0 \leq \ell \leq L-1
\end{aligned}
$$

Investigating the constraint further, we see

$$
\begin{aligned}
& \mathbf{P}_{\ell\delta}\Bigg(\bigg(\frac{W_{\ell\delta}^f - W_{(\ell+1)\delta}^f}{W_{\ell\delta}^f}\bigg)^+ < \beta\Bigg) > 1-\varepsilon, \qquad \forall 0 \leq \ell \leq L-1 \\
\iff & \mathbf{P}\bigg(\frac{W_{(\ell+1)\delta}^f}{W_{\ell\delta}^f} < 1-\beta\bigg) \leq \varepsilon, \qquad \forall 0 \leq \ell \leq L-1 \\
\iff & \mathbf{E}\left[F\left(\frac{\frac{1-\beta}{\left(\frac{W_{(\ell+1)\delta-1}^f}{W_{\ell\delta}^f}\right)} - 1}{f_{(\ell+1)\delta-1}}\right)\right] \leq \varepsilon, \qquad \forall 0 \leq \ell \leq L-1
\end{aligned}
$$

Due to the iid returns, we can further assume without loss of generality that $\ell = 0$ and the inequality constraint is actually an equality constraint as the optimal control, if it exists must necessarily be on the boundary (need to prove).

Thus we can simplify our goal to

$$\text{maximize: } \mathbf{E}\big[\ln(W^f_{L\delta})\big]$$
$$\text{subject to: } \mathbf{P}_{\ell\delta}\left(\left(\frac{W^f_{\ell\delta}-W^f_{(\ell+1)\delta}}{W^f_{\ell\delta}}\right)^+ < \beta\right) > 1-\varepsilon, \qquad \forall 0\le \ell\le L-1$$

Arbitrary Distribution, $\delta = 2$

Let $\ell = 0$, that is, we're only considering the first period, and drop the superscript from the fractions. Then our controls are $f_0(W^f_0)$, and $f_1(W^f_0, W^f_1)$, and the constraint is

$$\mathbf{P}_0\Big((1+f_0(W^f_0)r_1)(1+f_1(W^f_0,W^f_1)r_2) \le 1-\beta\Big) \le \varepsilon$$

Set $\tilde{f}_1(x) \coloneqq f_1(w_0, w_0 x)$, $F(x) = \mathbf{P}(r\le x)$, $\xi_1 = \xi_1(r_1) \coloneqq \tilde{f}_1(1+f_0 r_1) \in \mathcal{F}^r_1$, and condition on the first return, then the constraint becomes

$$\mathbf{E}\left[\mathbf{E}_1\Big[\mathbb{1}_{\left\{r_2 \le \frac{\frac{1-\beta}{1+f_0 r_1}-1}{\tilde{f}_1(1+f_0 r_1)}\right\}}\Big]\right] = \mathbf{E}\left[\mathbf{P}\left(r \le \frac{\frac{1-\beta}{1+f_0 r_1}-1}{\tilde{f}_1(1+f_0 r_1)}\right)\right] = \mathbf{E}\left[F\left(\frac{\frac{1-\beta}{1+f_0 r_1}-1}{\xi_1}\right)\right] \le \varepsilon.$$

The goal is

$$\sup_{f_0,\tilde{f}_1} \mathbf{E}\left[\ln\left(\frac{W^f_2}{W_0}\right)\right] = \sup_{f_0\in(0,1)} \sup_{\tilde{f}_1\in\mathcal{A}_{f_0}} \mathbf{E}\Big[\ln(1+f_0 r_1) + \mathbf{E}_1\big[\ln(1+\tilde{f}_1(1+f_0 r_1)r_2\big]\Big] \quad \text{can't move sup under expectation}$$
$$= \sup_{f_0\in(0,1)} \left\{ g(f_0) + \sup_{\tilde{f}_1\in\mathcal{A}_{f_0}} \mathbf{E}\big[g(\tilde{f}_1(1+f_0 r_1))\big]\right\} \quad \text{no DPP due to constraint}$$

First consider

$$\sup_{\tilde{f}_1\in\mathcal{A}_{f_0}} \mathbf{E}\Big[g\big(\tilde{f}_1\big(1+f_0 r_1\big)\big)\Big] = \sup_{\xi_1\in(0,1)} \mathbf{E}\Big[g\big(\xi_1\big)\Big], \quad \xi_1 \coloneqq \tilde{f}_1(1+f_0 r_1) \in \mathcal{F}^r_1$$
$$\text{subject to: } \mathbf{E}\left[F\left(\frac{\frac{1-\beta}{1+f_0 r_1}-1}{\xi_1}\right)\right] \le \varepsilon \iff \mathbf{E}\left[F\left(\frac{\frac{1-\beta}{1+f_0 r_1}-1}{\xi_1}\right)\right] - \varepsilon \le 0$$

We recast this problem to an unconstrained optimization problem where the constraint is added to the objective function with an appropriate parameter, thus using the tools of Lagrange multipliers.

$$\sup_{\xi_1\in(0,1)} \mathbf{E}\left[g\big(\xi_1\big) + \lambda\left(F\left(\frac{\frac{1-\beta}{1+f_0 r_1}-1}{\xi_1}\right) - \varepsilon\right)\right]$$

Remark. *We are also investigating using tools in stochastic programming with a so-called chance constraint.*

Arbitrary Distribution, Arbitrary δ

Fix a given betting period $\ell \in [0, L-1]$.

Chapter 11

Convergence to Continuous Time

In progress...

Chapter 12

Kelly Criterion in Continuous Time with Drawdown Constraint

12.1 Formalism/Formulation

Let Ω be the set of all paths, $\mathbf{D}([0,\infty);\mathbf{R}^d)$, representing the space of all possible trajectories of the instantaneous relative returns (rate of return, net return) $r_t = \dfrac{dP_t}{P_{t-}}$ = Levy Process, and $\mathbf{P}$ the observed, objective, true distribution/measure of the log-returns. In other words, $\mathbf{P}$ is the law of the Levy process. (Or would we define the sample space as the space of all possible price trajectories?). On which we can define $W : [0,\infty) \times \mathbf{D}([0,\infty;\mathbf{R}^d) \times \mathcal{P} \longrightarrow [0,\infty)$ as the wealth/bankroll/portfolio value trajectories,

$$W_t^f(r) = W_t^f(\omega) = W_0\,\mathrm{e}^{r_t^f},$$

where

$$r_t^f := \int_0^t f_{s-}dr_s + \frac{1}{2}\int_0^t f_{s-}(1-f_{s-})d[r^c,r^c]_s + \int_0^t\int_{\mathbf{R}}\big[\ln(1+f_{s-}(\mathrm{e}^x-1)) - f_{s-}x\big]\mu^r(ds,dx).$$

Note we need to prove this, see other paper. Also, $\mathcal{P}$ is the space of all predictable fractions/controls.
We can also define the relative drawdown $d_t^\delta(f)$

Note the difference between a relative (to either largest point or to starting point) and an absolute drawdown. My guess is I would consider relative drawdown if I was working with the wealth process, but I would consider absolute drawdown if I was working with the log-wealth process, namely the log-return process. We would like the probability of a "drawdown" being smaller than a prescribed threshold, over a particular window of time, to be close to one. One way of formalizing this is in a pathwise sense: Fix $\delta \geq 0, t > 0, \beta \in (0,1), \varepsilon \in (0,1)$,

$$\mathbf{P}(\mathrm{d}_t^\delta(f) \leq \beta, \forall t > 0) > 1-\varepsilon$$

$$\Longleftrightarrow \mathbf{P}\left(\frac{\sup\limits_{t-\delta\leq s\leq t} W_s^f - W_t^f}{\sup\limits_{t-\delta\leq s\leq t} W_s^f} \leq \beta, \forall t > 0\right) > 1-\varepsilon$$

$$\iff \mathbf{P}\left(\frac{W_t^f}{\sup\limits_{t-\delta\le s\le t} W_s^f} \ge 1-\beta, \forall t>0\right) > 1-\varepsilon$$

$$\iff \mathbf{P}\left(\inf_{t>0}\frac{W_t^f}{\sup\limits_{t-\delta\le s\le t} W_s^f} \ge 1-\beta\right) > 1-\varepsilon$$

$$\iff \mathbf{P}\left(\inf_{t>0}\frac{W_t^f}{\sup\limits_{t-\delta\le s\le t} W_s^f} \le 1-\beta\right) \le \varepsilon$$

Note, it appears (need to prove) that there is no uniform ε bound, meaning we conjecture (which should follow from iid increments) that the following

$$\mathbf{P}\left(\sup_{0\le t\le T}\frac{\sup\limits_{0\le s\le t} W_s^f - W_t^f}{\sup\limits_{0\le s\le t} W_s^f} \le \beta\right) > 1-\varepsilon, \quad \forall T\in(0,\infty), \forall \varepsilon\in(0,1)$$

is impossible to satisfy since it is (?) equivalent to

$$\lim_{T\to\infty}\mathbf{P}\left(\sup_{0\le t\le T}\frac{\sup\limits_{0\le s\le t} W_s^f - W_t^f}{\sup\limits_{0\le s\le t} W_s^f} \le \beta\right) = 0 > 1-\varepsilon$$

Consider the space

$$(\beta, f, \mathbf{P}\left(\inf_{t>0}\frac{W_t^f}{\sup\limits_{t-\delta\le s\le t} W_s^f} \le 1-\beta\right))$$

and for a given ε, β we will get a set of admissible f (i.e. a constrained region) to which we maximize $g_R(f)$ over the admissible set.

Remark. *I need to justify these two formulations with a practical application story. Also, it seems plausible that given these constraints, f will no longer be a fixed fraction. We should, from the outset, consider a predictable process $f = \{f_t\}$ with $f_t \in (0,1), \forall t$. To do this we need to consider the discrete time Kelly problem for a predictable fraction. See Lieb's counter-example in (E. O. Thorp, 2012). Also consider Bonferroni's inequality to interpret these as confidence interval curves and see the derivation at the back of (E. O. Thorp, 2006).*

Another way of formulating this is a uniform pointwise approach: Fix $\delta \ge 0, t>0, \beta\in(0,1), \varepsilon\in(0,1)$,

$$\mathbf{P}(\mathrm{d}_t^\delta(f) \le \beta) > 1-\varepsilon, , \forall t>0$$

$$\iff \mathbf{P}\left(\frac{\sup\limits_{t-\delta\le s\le t} W_s^f - W_t^f}{\sup\limits_{t-\delta\le s\le t} W_s^f} \le \beta\right) > 1-\varepsilon, \forall t>0$$

$$\iff \mathbf{P}\left(\frac{W_t^f}{\sup\limits_{t-\delta\le s\le t} W_s^f} \ge 1-\beta\right) > 1-\varepsilon, \forall t>0$$

$$\iff \mathbf{P}\left(\frac{W_t^f}{\sup\limits_{t-\delta\le s\le t} W_s^f} \le 1-\beta\right) \le \varepsilon, \forall t > 0$$

12.2 First Attempt of a Solution using Dynamic Programming

In progress...

12.3 Second Attempt of a Solution using Fluctuation and Ruin Theory

See (Feller, 1968), (Feller, 1971), (Asmussen & Albrecher, 2010), (Kyprianou, 2014). Also see Feller's 1948 paper called Fluctuation Theory of Recurrent Events (especially sections 5, 8, and 9).

Chapter 13

Tail Risk Hedging

Another way to limit drawdowns besides reducing exposures is to systematically "buy insurance" whether it be some sort of financial contract such as out of the money options or swaps, or utilizing some sort of stop loss strategy. This is a way of limiting drawdowns by adding "cheap" convexity to your portfolio as opposed to as opposed to simply reducing your exposure (fractional Kelly). Thus your wealth process would take something of the form, at least in discrete time,

$$W_1^{n,f} = \prod_{k=1}^{\lfloor 1t \rfloor} (1 + f^0 r^0 + f^1 r^1_{n,k} + f^2 r^2_{n,k}), \quad f^0 + f^1 + f^2 = 1$$

(This is probably not exactly the right formulation, need to think more)
Where r^1 is the rate of return on our "risky" gamble where we have an edge, and r^2 is our rate of return of our insurance product, note this is a function of our risky gamble, $r^2 = F(r^1)$, which has positive return when the risky gamble returns are below a certain threshold.
Also note r^0 is the return in a riskless bond or cash.

$$R^{\vec{f},portfolio} = f^0 R^{riskfree} + f^1 R^{aggressive} + f^2 R^{tailhedge}$$

WRITE IN GROSS RETURNS
Investigate tail-hedging strategies in (Bhansali, 2014).

Part V

Future Work

Chapter 14

Other Questions Posed by Thorp Worth Investigating

Some additional research questions posed by Thorp in (E. O. Thorp, 2012) are in regards to an investment scenario faced by fund manager Mohnish Pabrai (see The Dhandho Investor: The Low-Risk Value Method to High Return). Thorp points out why, in this particular situation, he wouldn't necessarily use the Kelly Criterion without particular adaptations. His discussion of the possible problems with the Kelly motivate many interesting and practical questions that should be resolved as they all make the case for reducing the optimal fraction. They are the following:

14.1 Opportunity Costs

This is a problem of a gambler placing simultaneous bets at a given instance in time, such as holding a portfolio with assets that have some dependence structure between them. In particular, when making our decision, to avoid overbetting, we need to know all the bets that we currently have and are currently considering adding, as well as their joint properties. We could attempt to solve this issue by finding, either explicitly or numerically, the optimal $\vec{f}^*$ when we have a (log-gross) return vector $\vec{R}_t = (R_t^1, \cdots, R_t^d)^\top \in \mathbf{R}^d$ that has a particular dependence structure, for example, $\gamma_t(i,j) = Cov(R_t^i, R_t^j)$. I think the mathematics of heavy tailed copulas could help answer this question. See (Cont & Tankov, 2004), chapter 5. Also, for intuition on how this can be used and why it would be important, see section 10.2 of (Ethier, 2010). At the end of this section, Ethier has an example where he has two betting scenarios, one which is more diversified than the other, with an edge, but you would choose the one with lower edge per trial since the lower edge per trial strategy leads to more diversification and hence a higher long term growth rate with lower volatility.
The opportunity cost refers to over-diversifying in the presence of large edges as this can lead to missed opportunities for compounding. Considering available alternative investments at any given time will help avoid this problem, see chapter 5 of (Brown, 2011).

14.2 Risk (volatility and ruin) Tolerance

As discussed in (E. O. Thorp, 2006), and in many articles in the main reference book, betting your full optimal fraction f^* can lead to large swings in your wealth in the short run, i.e. "there are drawdowns which are too large for the comfort of many investors" in the short run. So a heuristic that Thorp implemented was to use a percentage of the optimal fraction, called "fractional Kelly," which tamed the wild drawdowns while also reducing the long term growth rate.

One way this can be formalized is by adding a constraint to the long term grown rate optimization problem. Note that a large drawdown can lead to effective ruin if someone decides to pull all the money on the investment strategy. Thus drawdown risk is really just effective ruin risk. We are also interested in the most important type of risk which is risk of literal ruin for a given initial wealth and betting fraction, $\mathbf{P}(\inf_{t\geq 0} W_t^f \leq 0 | W_0 = w) = \mathbf{P}_{0,w}(\inf_{t\geq 0} W_t^f \leq 0) =: \Psi(w, f)$.

It appears that one way these problems can be handled is with the tools of fluctuation theory for Levy processes, see chapter 11 especially in (Asmussen & Albrecher, 2010). In addition to answering questions about ruin and drawdown probabilities, it appears this theory can answer questions (involving asymptotic and non-asymptotic superior properties) like the chance your wealth will be larger than any goal on or before a particular time, $\mathbf{P}(\max_{s\leq t} W_s^f > B | W_0 = w)$. Moreover, chapter 8 (Level Dependent Risk Processes) might be able to answer questions about transaction costs and taxes. We need to include these into our wealth model as they really affect compounding, again see $(E.O.Thorp, 2006)$. Another way to handle these problems is with stochastic control methods.

Remark. *We claim that one of the worst things that can happen when trying to compound gains is having "large" draw-downs. As a simple example, it is worse to have a ten percent drop in my wealth today than to have 10 one percent drops in my wealth spread out over time, say daily or monthly, due to the non-linearity of the exponential. Thus the moral is that steep (implied by the slope) drawdowns destroy compounding!*

14.3 Model Uncertainty

Another reason for using a "fractional Kelly" strategy (besides minimizing short term drawdown risk) is when we are unsure about our return/payoff distribution. If we get the distribution wrong then our optimal fraction could still lead to over-betting.

See (Browne & Whitt, 1996), and Optimal betting under parameter uncertainty: Improving the Kelly Criterion by Baker and McHale, and papers by S. Peng on model uncertainty for a candidate theoretical framework.

In particular, this section is essential for connecting Kelly criterion with forecasting models or factor models for returns of various strategies

$$r \approx \hat{r}(\vec{X}; \vec{\theta})$$

14.4 Black Swans

This is one piece we are trying to solve above. It is solving the Kelly Criterion under more general, including fat-tailed, Levy processes.

14.5 The "Long Run"

Here one would be concerned with convergence rates of the Strong Law of Large Numbers and its variants. In other words, how long do I need to keep compounding my wealth (or about how many bets do I need to place/frequency of bets) to achieve my long term growth rate. Note that not every gambler can make enough bets for the long run to kick in, at least with a high probability. These asymptotic properties will not hold true if the gambler/investor doesn't have enough opportunities to make it into the long run. As of now, I am not too sure how to go about finding these convergence rates, yet it would be a very important question to have answered.

14.6 Predictable f, using stochastic control methods

It is of interest to investigate processes that have independent increments and non-stationary increments, see (E. O. Thorp, 2006). Note these could help an investor formalize economic cycles or environments and change their incremental distributions in a deterministic way. It also makes sense in the case of card counting, where your payoff distribution is changing in a random way and is being measured by the count. (See Bill Gross's quote about knowing where you are in the cycle is like knowing the count). Additionally, in this framework we wouldn't have a reason for our optimal fraction to be a constant as every increment has a different environment, so we would be solving for an optimal fraction process $\{f_t^*\}$ and we leave the setting of constant proportion/myopic strategies. And in this framework, it may be possible to find a strategy that dominates the Kelly strategy and is "essentially different," see "Leib's Paradox" in (E. O. Thorp, 2012). This might be mathematically formalized using additive processes.

Remark. *Of course the practical question of estimating the characteristics of a Levy process from data and numerically computing all of these different quantities is also a question needing to be answered.*

Appendix A

Levy Processes, Convergence of Discrete Models, and Stochastic Calculus for Levy Processes

Definition 1. *Levy Process:*

Theorem 1. *Levy-Ito Decomposition:*

Theorem 2. *Itô's Formula for Lévy Processes from Random Walks:*

A.1 Introduction to Lévy Processes

(We follow Matthias Winkel's notes) Consider a random walk, $S_n = \sum_{k=1}^{n} \xi_k$, with $\{\xi_k\}$ being a sequence of independent and identically distributed random variables. Note the increments, $S_{k+h} - S_k$, $\forall h, k \geq 1$, are independent and stationary.
That is, $S_{k+h} - S_k \perp\!\!\!\perp (S_k, S_{k-1}, \cdots, S_1)$, and $S_{k+h} - S_k \sim S_h$. Also, note $S_0 = 0$.

For the random walk, there exists limit theorems such as Law of Large numbers, $\frac{S_n}{n} \xrightarrow[n\to\infty]{a.s.} \mathbf{E}[Y_1]$, whenever $\mathbf{E}[|\xi_1|] < \infty$, and the Central Limit Theorem.

Theorem 3 (Gauss)**.** *Scratch*

$$\int_{-\infty}^{\infty} \mathrm{e}^{-x^2}\,\mathrm{d}x = \sqrt{\pi}. \tag{A.1}$$

Theorem 4 (CLT)**.** $Var(\xi_1) < \infty \implies \frac{S_n - \mathbb{E}\left[S_n\right]}{\sqrt{n}} \xrightarrow[n\to\infty]{d} N(0, Var(\xi_1))$

Proof. Use (Jacod & Shiryaev, 2003, Remark VII.5.3). □

But what if we took a more and more remote look at our random walk, that is, zoomed out further and further. More specifically, imagine these discrete observations $\{\xi_k\}_{k \leq n}$ happening during a continuous time interval, for simplicity say $[0,1]$. Thus we have n−many observations per time period. "Zooming out" is equivalent to speeding up the frequency of our observations per time period. A theorem by Donsker describes what happens to our random walk as the frequency per time period gets larger and larger.

Set the centered and scaled pre-limiting process to be

$$Y_t^n := \frac{S_{\lfloor nt \rfloor} - \mathbb{E}\big[S_{\lfloor nt \rfloor}\big]}{\sqrt{n}}$$

and the limiting process to be

$$Y_t := \sigma B_t \overset{d}{=} N(0, \sigma^2 t), \quad \sigma^2 := \mathrm{Var}(\xi_1)$$

where B is a Brownian motion process.

Theorem 5. *Donsker's Invariance Principle/Functional CLT:*
$\sigma^2 := Var(\xi_1) < \infty \implies Y^n \xrightarrow[n\to\infty]{\mathcal{L}} Y$ (*convergence in law as stochastic processes*)

Proof. Prove using convergence of characteristic triple. See (Jacod & Shiryaev, 2003, Corollary VII.3.11) □

Note we need to define Brownian motion and convergence in law of stochastic processes. A simple binomial model is a particular example. Also, add simulations of the convergence of discrete to continuous time. It can be shown that X has continuous sample paths, and has independent and stationary increments, thus showing a continuous time version of a random walk.

What happens if the observations have fat enough tails that the variance is infinite, $\mathrm{Var}(\xi_1) = \infty$?

Theorem 6. *Generalized CLT:*

Proof. Prove using convergence of characteristic triple □

Theorem 7. *Functional Generalized CLT:*

Proof. □

Most generally, consider limit theorems of traingular arrays and their functional version, which shows how to get Lévy processes from random walks.

Appendix B

Dynamic Programming

Discuss history of dynamic programming and it's formalization.

Appendix C

Code

```
import numpy as np
w = 1 #initial wealth
n = 1000 #number of times we gamble, frequency of betting
f = 0.2 #fixed exposure fraction

W_f = np.arange(0,n+1,dtype=np.float) #initialize wealth path
W_f[0] = w #initial wealth
#below we  need to define a function called relative_returns
#that simulates the returns of the distribution we're gambling on
r = relative_returns #return simulation for the fixed path

for i in range(1,len(W_f)):
    W_f[i] = W_f[i-1]*(1+f*r[i]) #recursive wealth process for given f
```

Listing 1: Simulation of Discrete Time Wealth Processes

```python
#Simulation of wealth process for a b:1 payoff sequence of bets with constant proportional (f) betting
import numpy as np
import matplotlib.pyplot as plt

paths = 20 #number of different omegas
w = 1 #initial wealth
n = 200 #number of times we gamble, frequency of betting
b = 1 #this is b in the b:1 odds
p = 0.6 #this is p
f = np.arange(0, 1, 0.1)
optimal_f = .2 #this is f, exposure of wealth/bankroll/reserves to gamble payoffs

#simulating returns from a binary bet
def binary_returns(n, p, up, down):
    r = np.random.uniform(size=n+1)
    r1=np.where(r<p,up,down)
    return r1

for k in range(1, len(f)): #fix f
    W_f = np.arange(0,n+1,dtype=np.float) #initial wealth process for given f
    W_f[0] = w #initial wealth
    for j in range(1, paths): #fix path/fix omega
        r = binary_returns(n,p,b,-1) #return generation for the fixed path

        for i in range(1,len(W_f)):
            W_f[i] = W_f[i-1]*(1+(f[k]*r[i])) #recursive wealth process for given f

        plt.subplot(2,1,1)
        plt.plot(W_f/w)
        plt.title('Wealth/PortfolioValue on Ordinary and Log Scale, for f = ' + str(f[k]))
        plt.xlabel('n')
        plt.ylabel('W_n^f/W_0')

        plt.subplot(2,1,2)
        plt.plot(np.log(W_f/w))
        plt.xlabel('n')
        plt.ylabel('ln(W_n^f/W_0)')

        G_n_f = (1/n)*np.log(W_f[n]/w) #observed geometric average at end of betting
    plt.show()
    print('observed geometric average is ', G_n_f, ' for f = ', f[k])
    print('theoretical long run growth rate is ', p*np.log(1+b*f[k]) + (1. - p)*np.log(1-f[k]), 'for f = ' + str(f[k]))
plt.show()
```

Listing 2: Simulation of Wealth Process for Coin Toss, Observed Growth Rate, Theoretical Growth Rate

```python
import numpy as np
import matplotlib.pyplot as plt
import scipy
from scipy.optimize import minimize_scalar
#Binary Bet/Payoff (Coin Toss) b:1
b = 1
p = 0.6
f_binary = np.arange(0, 1, .01)
g_f_binary = np.arange(len(f_binary), dtype='float')
N = 1000000
def coinflip_returns(n, p, up, down):
    r = np.random.uniform(size=n+1)
    r1=np.where(r<p,up,down)
    return r1

print('This is verifying LLN holds for binary return')
for k in range(10):
    r = coinflip_returns(N, p, b, -1)
    print(np.mean(r))

for i in range(len(f_binary)):
    g_f_binary[i] = np.mean(np.log(1+f_binary[i]*r))

print('Edge = ', np.mean(r))
plt.figure()
plt.hist(r)
plt.title('Relative Returns Histogram')

print('Binary bet, f* = ', f_binary[np.argmax(g_f_binary)])
plt.figure()
plt.plot(f_binary, g_f_binary)
plt.ylim(-0.2,.03)
plt.xlabel('f')
plt.ylabel('g_{coin_flip}(f)')
plt.title('Long Term Growth Rate For Binary Returns (CoinFlip) via Simulation')

def minus_g_binary(f):
    return (-1)*(p*np.log(1+b*f)+(1-p)*np.log(1-f))

print(minimize_scalar(minus_g_binary, bounds=(0,1), method='bounded'))
plt.figure()
plt.plot(np.linspace(0,1,len(f_binary)), -minus_g_binary(np.linspace(0,1,len(f_binary)))) #'.'
plt.title('Long Term Growth Rate For Binary Returns (CoinFlip) via Numerical Solver' )
plt.ylim(-0.2,0.03)
plt.xlabel('f')
plt.ylabel('g_{coin_flip}(f)')
```

Listing 3: Kelly Criterion Binary Returns

```python
import numpy as np
import matplotlib.pyplot as plt
import scipy
import scipy.integrate as integrate
import scipy.optimize as op
from scipy.optimize import minimize_scalar

#uniform[-1,2] distrubution
f_unif = np.arange(0, 1, .01)
g_f_unif = np.arange(len(f_unif), dtype='float')
N = 1000000
#this is verifying LLN holds
print('This is verifying LLN holds for uniform return')
for k in range(10):
    r = np.random.uniform(-1,2, size=N)
    print(np.mean(r))

for i in range(len(f_unif)):
    g_f_unif[i] = np.mean(np.log(1+f_unif[i]*r))

print('Edge = ', np.mean(r))
plt.figure()
plt.hist(r, bins = 500, density=True)
plt.title('Relative Returns Histogram')
print('Uniform returns, f* = ', f_unif[np.argmax(g_f_unif)], 'and g_r(f*) = ', g_f_unif[np.argmax(g_f_unif)])
plt.figure()
plt.plot(f_unif, g_f_unif)
plt.xlabel('f')
plt.ylabel('g_{Unif[-1,2]}(f)')
plt.title('Long Term Growth Rate For Uniform[-1,2] via Simulation')

def minus_g_uniform(f):
    return (-1)*scipy.integrate.quad(lambda x: np.log(1+f*x), -1,2)[0]/3.

print(minimize_scalar(minus_g_uniform, bounds=(0,1), method='bounded'))

plt.figure()
plt.plot(f_unif, [-minus_g_uniform(f) for f in f_unif])
plt.xlabel('f')
plt.ylabel('g_{Unif[-1,2]}(f)')
plt.title('Long Term Growth Rate For Uniform[-1,2] via Numerical Solver')
```

Listing 4: Kelly Criterion Uniform Returns

```
In progress...
```

Listing 5: Simulation of α-Stable Process

```
template
```

Listing 6: template

Bibliography

Algoet, P. H. (1994). The strong law of large numbers for sequential decisions under uncertainty. *IEEE Trans. Inform. Theory*, *40*(3), 609–633. doi:10.1109/18.335876

Algoet, P. H. & Cover, T. M. (1988). Asymptotic optimality and asymptotic equipartition properties of log-optimum investment. *Ann. Probab. 16*(2), 876–898. Retrieved from http://links.jstor.org/sici?sici=0091-1798(198804)16:2%3C876:AOAAEP%3E2.0.CO;2-D&origin=MSN

Asmussen, S. & Albrecher, H. (2010). *Ruin probabilities* (Second). Advanced Series on Statistical Science & Applied Probability. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ. doi:10.1142/9789814282536

Barndorff-Nielsen, O. E. & Shiryaev, A. (2010). *Change of time and change of measure*. Advanced Series on Statistical Science & Applied Probability. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ. doi:10.1142/7928

Bellman, R. & Kalaba, R. (1957). Dynamic programming and statistical communication theory. *Proc. Nat. Acad. Sci. U.S.A. 43*, 749–751. doi:10.1073/pnas.43.8.749

Bhansali, V. (2014). *Tail-risk hedging*. McGraw-Hill Education.

Brown, A. (2011). *Red-blooded risk: The secret history of wall street*. Wiley.

Browne, S. & Whitt, W. (1996). Portfolio choice and the Bayesian Kelly criterion. *Adv. in Appl. Probab. 28*(4), 1145–1176. doi:10.2307/1428168

Cont, R. & Tankov, P. (2004). *Financial modelling with jump processes*. Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL.

Eberlein, E. (2009). Jump type levy processes. In *Handbook of financial time series* (pp. 439–456). Springer.

Ethier, S. N. (2010). *The doctrine of chances*. Probability and its Applications (New York). Probabilistic aspects of gambling. Springer-Verlag, Berlin. doi:10.1007/978-3-540-78783-9

Feller, W. (1968). *An introduction to probability theory and its applications. vol. i.* Third edition. John Wiley & Sons, Inc., New York-London-Sydney.

Feller, W. (1971). *An introduction to probability theory and its applications. vol. ii.* Second edition. John Wiley & Sons, Inc., New York-London-Sydney.

Jacod, J. & Shiryaev, A. N. (2003). *Limit theorems for stochastic processes* (Second). Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin. doi:10.1007/978-3-662-05265-5

Kardaras, C. & Platen, E. (2013). Multiplicative approximation of wealth processes involving no-short-sales strategies via simple trading. *Math. Finance*, *23*(3), 579–590. doi:10.1111/j.1467-9965.2011.00511.x

Kelly, J. (1956). A new interpretation of information rate. In *The kelly capital growth investment criterion* (Chap. 2, pp. 25–34). World Scientific.

Kyprianou, A. E. (2014). *Fluctuations of Lévy processes with applications* (Second). Universitext. Introductory lectures. Springer, Heidelberg. doi:10.1007/978-3-642-37632-0

Poundstone, W. (2006). *Fortune's formula: The untold story of the scientific betting system that beat the casinos and wall street.* Hill and Wang.

Rachev, S. T. & Rushendorf, L. (1994). Models for option prices. *Teor. Veroyatnost. i Primenen. 39*(1), 150–190. doi:10.1137/1139005

Samorodnitsky, G. & Taqqu, M. S. (1994). *Stable non-Gaussian random processes.* Stochastic Modeling. Stochastic models with infinite variance. Chapman & Hall, New York.

Thorp, E. O. (1966). *Beat the dealer: A winning strategy for the game of twenty-one.* Vintage.

Thorp, E. O. (2006). The kelly criterion in blackjack, sports betting, and the stock market. In *The kelly capital growth investment criterion* (Chap. 54, pp. 791–834). World Scientific.

Thorp, E. O. (2012). Understanding the kelly criterion. In *The kelly capital growth investment criterion* (Chap. 36, pp. 511–525). World Scientific.

Thorp, E. O. (2018). *A man for all markets.* From Las Vegas to Wall Street, how I beat the dealer and the market, With a foreword by Nassim Nicholas Taleb, Paperback edition of the 2017 original. Random House, New York.

Thorp, E. & E. Walden, W. (1966, June). A favorable side bet in nevada baccarat. *Journal of The American Statistical Association - J AMER STATIST ASSN*, *61*, 313–328. doi:10.1080/01621459.1966.10480867

Thorp, E. & Kassouf, S. (1967). *Beat the market: a scientific stock market system.*

Wilson, A. N. (1970). *The casino gambler's guide.* Harper Collins.