Portfolios that Contain Risky Assets 13: Indepedent, Identically Distributed Models for Portfolios

C. David Levermore

University of Maryland, College Park, MD

Math 420: *Mathematical Modeling*
April 19, 2021 version

Portfolios that Contain Risky Assets
Part II: Stochastic Models

11. Independent, Identically-Distributed Models for Assets
12. Assessment of Independent, Identically-Distributed Models
13. Independent, Identically-Distributed Models for Portfolios
14. Kelly Objectives for Markowitz Portfolios
15. Cautious Objectives for Markowitz Portfolios
16. Optimization of Mean-Variance Objectives
17. Fortune's Formulas

Independent, Identically-Distributed Models for Portfolios

1. Independent, Identically-Distributed Models for Markets
2. Independent, Identically-Distributed Models for Portfolios
3. Growth Rates
4. Law of Large Numbers
5. Normal Growth Rate Model
6. Portfolio Selection
7. Kelly Criterion for a Simple Game
8. Kelly Criterion in Practice

Independent, Identically-Distributed Models for Markets

We now consider a market with N risky assets. Let $\{s_i(d)\}_{d=0}^{D}$ be the share price history of asset i. The associated return and growth rate histories are $\{r_i(d)\}_{d=1}^{D}$ and $\{x_i(d)\}_{d=1}^{D}$ where

$$r_i(d) = \frac{s_i(d)}{s_i(d-1)} - 1\,, \qquad x_i(d) = \log\left(\frac{s_i(d)}{s_i(d-1)}\right) .$$

Because each $s_i(d)$ is positive, each $r_i(d)$ is in $(-1,\infty)$, and each $x_i(d)$ is in $(-\infty,\infty)$. Let $\mathbf{r}(d)$ and $\mathbf{x}(d)$ be the N-vectors

$$\mathbf{r}(d) = \begin{pmatrix} r_1(d) \\ \vdots \\ r_N(d) \end{pmatrix}, \qquad \mathbf{x}(d) = \begin{pmatrix} x_1(d) \\ \vdots \\ x_N(d) \end{pmatrix} .$$

The market return and growth rate histories can then be expressed simply as $\{\mathbf{r}(d)\}_{d=1}^{D}$ and $\{\mathbf{x}(d)\}_{d=1}^{D}$ respectively.

Independent, Identically-Distributed Models for Markets

An IID model for this market draws D random vectors $\{\mathbf{R}_d\}_{d=1}^D$ from a fixed probablity density $q(\mathbf{R})$ over $(-1, \infty)^N$. Such a model is reasonable when the points $\{(d, \mathbf{r}(d))\}_{d=1}^D$ are distributed uniformly in d. This is hard to visualize when N is not small.

You might think a necessary condition for the entire market to have an IID model is that each asset has an IID model. *This can be visualized for each asset by plotting the points $\{(d, r_i(d))\}_{d=1}^D$ in the dr-plane and seeing if they appear to be distributed uniformly in d.*

Similar visual tests based on pairs of assets can be carried out by plotting the points $\{(d, r_i(d), r_j(d))\}_{d=1}^D$ in $\mathbb{R}^3$ with an interactive 3D graphics package.

Independent, Identically-Distributed Models for Markets

Visual tests like those described above often show that funds behave more like IID models than individual stocks or bonds. This means that portfolio balancing strategies based on IID models might work better for portfolios composed largely of funds. This is one reason why some investors prefer investing in funds over investing in individual stocks and bonds.

A better lesson to be drawn from the observation in the last paragraph is that every sufficiently diverse portfolio of assets in a market will behave more like an IID model than many of the individual assets in that market. In other words, IID models for a market can be used to develop portfolio balancing strategies when the portfolios considered are sufficiently diverse, even when the behavior of individual assets in that market may not be well described by the model. This is another reason to prefer holding diverse, broad-based portfolios.

Independent, Identically-Distributed Models for Markets

More importantly, this suggests that it is better to apply visual tests like those described above to representative portfolios rather than to individual assets in the market.

Remark. Such visual tests can only warn you when IID models might not be appropriate for describing the data. There are also statistical tests that can play this role. *There is no visual or statistical test that can insure the validity of using an IID model for a market. However, due to their simplicity, IID models are often used unless there is a good reason not to use them.*

Independent, Identically-Distributed Models for Markets

After we have decided to use an IID model for the market, we must gather statistical information about the return probability density $q(\mathbf{R})$. The mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Xi}$ of $\mathbf{R}$ are given by

$$\boldsymbol{\mu} = \int \mathbf{R}\, q(\mathbf{R})\,\mathrm{d}\mathbf{R}\,, \qquad \boldsymbol{\Xi} = \int (\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})^{\mathrm{T}} q(\mathbf{R})\,\mathrm{d}\mathbf{R}\,.$$

Given any sample $\{\mathbf{R}_d\}_{d=1}^{D}$ drawn from $q(\mathbf{R})$, these have the unbiased estimators

$$\hat{\boldsymbol{\mu}} = \sum_{d=1}^{D} w_d\,\mathbf{R}_d\,, \qquad \hat{\boldsymbol{\Xi}} = \sum_{d=1}^{D} \frac{w_d}{1-\bar{w}}\,(\mathbf{R}_d-\hat{\boldsymbol{\mu}})\,(\mathbf{R}_d-\hat{\boldsymbol{\mu}})^{\mathrm{T}}\,.$$

If we assume that such a sample is given by the return history $\{\mathbf{r}(d)\}_{d=1}^{D}$ then these estimators are given in terms of the vector $\mathbf{m}$ and matrix $\mathbf{V}$ by

$$\hat{\boldsymbol{\mu}} = \mathbf{m}\,, \qquad \hat{\boldsymbol{\Xi}} = \frac{1}{1-\bar{w}}\,\mathbf{V}\,.$$

Independent, Identically-Distributed Models for Portfolios

Recall that the value of a portfolio that holds a risk-free balance $b_{\mathrm{rf}}(d)$ with return μ_{rf} and $n_i(d)$ shares of asset i during trading day d is

$$\pi(d) = b_{\mathrm{rf}}(d)\,(1 + \mu_{\mathrm{rf}}) + \sum_{i=1}^{N} n_i(d) s_i(d)\,.$$

We will assume that $\pi(d) > 0$ for every d. Then the return $r(d)$ and growth rate $x(d)$ for this portfolio on trading day d are given by

$$r(d) = \frac{\pi(d)}{\pi(d-1)} - 1\,, \qquad x(d) = \log\left(\frac{\pi(d)}{\pi(d-1)}\right)\,.$$

Recall that the return $r(d)$ for the Markowitz portfolio with allocation $\mathbf{f}$ can be expressed in terms of the vector $\mathbf{r}(d)$ as

$$r(d) = (1 - \mathbf{1}^{\mathrm{T}}\mathbf{f})\mu_{\mathrm{rf}} + \mathbf{f}^{\mathrm{T}}\mathbf{r}(d)\,.$$

Independent, Identically-Distributed Models for Portfolios

This implies that if the underlying market has an IID model with return probability density $q(\mathbf{R})$ then the Markowitz portfolio with allocation $\mathbf{f}$ has the IID model with return probability density $q_{\mathbf{f}}(R)$ given by

$$q_{\mathbf{f}}(R) = \int \delta\Big(R - (1 - \mathbf{1}^{\mathrm{T}}\mathbf{f})\mu_{\mathrm{rf}} - \mathbf{R}^{\mathrm{T}}\mathbf{f}\Big)\; q(\mathbf{R})\,\mathrm{d}\mathbf{R}\,.$$

Here $\delta(\,\cdot\,)$ denotes the *Dirac delta distribution*, which can be defined by the property that for every sufficiently nice function $\psi(R)$

$$\int \psi(R)\,\delta\Big(R - (1 - \mathbf{1}^{\mathrm{T}}\mathbf{f})\mu_{\mathrm{rf}} - \mathbf{R}^{\mathrm{T}}\mathbf{f}\Big)\,\mathrm{d}R = \psi\Big((1 - \mathbf{1}^{\mathrm{T}}\mathbf{f})\mu_{\mathrm{rf}} + \mathbf{R}^{\mathrm{T}}\mathbf{f}\Big)\;.$$

Independent, Identically-Distributed Models for Portfolios

Hence, by combining the foregoing formula for $q_{\mathbf{f}}(R)$ with the defining property of the Dirac delta distribution, we see that for every sufficiently nice function $\psi(R)$ we have the formula

$$\begin{aligned}
\mathrm{Ex}(\psi(R)) &= \int \psi(R)\, q_{\mathbf{f}}(R)\, \mathrm{d}R \\
&= \int \psi(R) \left[\int \delta\Big(R - (1 - \mathbf{1}^{\mathrm{T}}\mathbf{f})\mu_{\mathrm{rf}} - \mathbf{R}^{\mathrm{T}}\mathbf{f}\Big)\, q(\mathbf{R})\, \mathrm{d}\mathbf{R} \right] \mathrm{d}R \\
&= \int \left[\int \psi(R)\, \delta\Big(R - (1 - \mathbf{1}^{\mathrm{T}}\mathbf{f})\mu_{\mathrm{rf}} - \mathbf{R}^{\mathrm{T}}\mathbf{f}\Big)\, \mathrm{d}R \right] q(\mathbf{R})\, \mathrm{d}\mathbf{R} \\
&= \int \psi\Big((1 - \mathbf{1}^{\mathrm{T}}\mathbf{f})\mu_{\mathrm{rf}} + \mathbf{R}^{\mathrm{T}}\mathbf{f}\Big)\, q(\mathbf{R})\, \mathrm{d}\mathbf{R}\,.
\end{aligned}$$

This formula can also be viewed as defining $q_{\mathbf{f}}(r)$.

Independent, Identically-Distributed Models for Portfolios

In particular, we can compute the mean μ of $q_{\mathbf{f}}(R)$ as

$$\begin{aligned}\mu = \mathrm{Ex}(R) &= \int \left((1-\mathbf{1}^{\mathrm{T}}\mathbf{f})\mu_{\mathrm{rf}} + \mathbf{R}^{\mathrm{T}}\mathbf{f}\right) q(\mathbf{R})\,\mathrm{d}\mathbf{R} \\ &= (1-\mathbf{1}^{\mathrm{T}}\mathbf{f})\mu_{\mathrm{rf}} \int q(\mathbf{R})\,\mathrm{d}\mathbf{R} + \left(\int \mathbf{R}\, q(\mathbf{R})\,\mathrm{d}\mathbf{R}\right)^{\mathrm{T}} \mathbf{f} \\ &= (1-\mathbf{1}^{\mathrm{T}}\mathbf{f})\mu_{\mathrm{rf}} + \boldsymbol{\mu}^{\mathrm{T}}\mathbf{f}\,,\end{aligned}$$

where in the last step we have used the facts that

$$\int q(\mathbf{R})\,\mathrm{d}\mathbf{R} = 1\,, \qquad \int \mathbf{R}\, q(\mathbf{R})\,\mathrm{d}\mathbf{R} = \boldsymbol{\mu}\,.$$

Independent, Identically-Distributed Models for Portfolios

This formula for μ can then be used to compute the variance ξ of $q_{\mathbf{f}}(R)$ as

$$\begin{aligned}\xi = \mathrm{Ex}((R-\mu)^2) &= \int \left((1-\mathbf{1}^{\mathrm{T}}\mathbf{f})\mu_{\mathrm{rf}} + \mathbf{R}^{\mathrm{T}}\mathbf{f} - \mu\right)^2 q(\mathbf{R})\,\mathrm{d}\mathbf{R} \\ &= \int \left(\mathbf{R}^{\mathrm{T}}\mathbf{f} - \boldsymbol{\mu}^{\mathrm{T}}\mathbf{f}\right)^2 q(\mathbf{R})\,\mathrm{d}\mathbf{R} \\ &= \int \mathbf{f}^{\mathrm{T}}(\mathbf{R}-\boldsymbol{\mu})\,(\mathbf{R}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{f}\,q(\mathbf{R})\,\mathrm{d}\mathbf{R} \\ &= \mathbf{f}^{\mathrm{T}}\left(\int (\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})^{\mathrm{T}} q(\mathbf{R})\,\mathrm{d}\mathbf{R}\right)\mathbf{f} = \mathbf{f}^{\mathrm{T}}\boldsymbol{\Xi}\,\mathbf{f}\,,\end{aligned}$$

where in the last step we have used the fact that

$$\int (\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})^{\mathrm{T}} q(\mathbf{R})\,\mathrm{d}\mathbf{R} = \boldsymbol{\Xi}\,.$$

Independent, Identically-Distributed Models for Portfolios

If we assume that the return history $\{\mathbf{r}(d)\}_{d=1}^{D}$ is an IID sample drawn from a probability density $q(\mathbf{R})$ then unbiased estimators of the associated mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Xi}$ are given in terms of $\mathbf{m}$ and $\mathbf{V}$ by

$$\hat{\boldsymbol{\mu}} = \mathbf{m}\,, \qquad \hat{\boldsymbol{\Xi}} = \frac{1}{1-\bar{w}}\,\mathbf{V}\,.$$

Moreover, the Markowitz portfolio with allocation $\mathbf{f}$ has the return history $\{r(d)\}_{d=1}^{D}$ where

$$r(d) = (1 - \mathbf{1}^{\mathrm{T}}\mathbf{f})\mu_{\mathrm{rf}} + \mathbf{f}^{\mathrm{T}}\mathbf{r}(d)\,.$$

This return history is an IID sample drawn from the probability density $q_{\mathbf{f}}(R)$ and the formulas on the last two pages show that the mean μ and variance ξ of $q_{\mathbf{f}}(R)$ have the unbiased estimators

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

Growth Rate Probability Densities

Now suppose that the returns of an IID model for a portfolio are drawn from a return probability density $q(R)$. Given D samples $\{R_d\}_{d=1}^{D}$ that are drawn from $q(R)$, the associated simulated portfolio values $\{\Pi_d\}_{d=1}^{D}$ satisfy

$$\Pi_d = (1 + R_d)\,\Pi_{d-1}\,, \quad \text{for } d = 1, \cdots, D\,. \tag{3.1}$$

If the inital portfolio value $\pi(0)$ is known then we set $\Pi_0 = \pi(0)$ and use induction to show that

$$\Pi_d = \prod_{d'=1}^{d} (1 + R_{d'})\,\pi(0)\,. \tag{3.2}$$

Growth Rate Probability Densities

The *growth rate* X_d is related to the return R_d by

$$e^{X_d} = 1 + R_d \,. \tag{3.3}$$

In other words, X_d is the growth rate that yeilds a return R_d on trading day d. The formula for Π_d then takes the form

$$\Pi_d = \exp\left(\sum_{d'=1}^{d} X_{d'} \right) \pi(0) \,. \tag{3.4}$$

Growth Rate Probability Densities

If the samples $\{R_d\}_{d=1}^{D}$ are drawn from a density $q(R)$ over $(-1,\infty)$ then the $\{X_d\}_{d=1}^{D}$ are drawn from a density $p(X)$ over $(-\infty,\infty)$ where

$$p(X)\,\mathrm{d}X = q(R)\,\mathrm{d}R\,,$$

with X and R related by

$$X = \log(1+R)\,, \qquad R = e^X - 1\,.$$

More explicitly, the densities $p(X)$ and $q(R)$ are related by

$$p(X) = q\left(e^X - 1\right) e^X\,, \qquad q(R) = \frac{p(\log(1+R))}{1+R}\,.$$

Growth Rate Probability Densities

Because our models will involve means and variances, we will require that

$$\int_{-\infty}^{\infty} X^2 p(X)\,\mathrm{d}X = \int_{-1}^{\infty} \log(1+R)^2\, q(R)\,\mathrm{d}R < \infty\,,$$
$$\int_{-\infty}^{\infty} \left(e^X - 1\right)^2 p(X)\,\mathrm{d}X = \int_{-1}^{\infty} R^2 q(R)\,\mathrm{d}R < \infty\,.$$

Then the mean γ and variance θ of X are

$$\gamma = \mathrm{Ex}(X) = \int_{-\infty}^{\infty} X\, p(X)\,\mathrm{d}X\,,$$
$$\theta = \mathrm{Vr}(X) = \mathrm{Ex}\Big((X-\gamma)^2\Big) = \int_{-\infty}^{\infty} (X-\gamma)^2\, p(X)\,\mathrm{d}X\,.$$

Growth Rate Probability Densities

The big advantage of working with $p(X)$ rather than $q(R)$ is the fact that

$$\log\left(\frac{\Pi_d}{\pi(0)}\right) = \sum_{d'=1}^{d} X_{d'} \,.$$

In other words, $\log(\Pi_d/\pi(0))$ *is a sum of an IID process.* It is easy to compute the mean and variance of this quantity in terms of those of X.

For the mean of $\log(\Pi_d/\pi(0))$ we find that

$$\mathrm{Ex}\left(\log\left(\frac{\Pi_d}{\pi(0)}\right)\right) = \sum_{d'=1}^{d} \mathrm{Ex}(X_{d'}) = d\,\gamma \,,$$

Growth Rate Probability Densities

For the variance of $\log(\Pi_d/\pi(0))$ we find that

$$\begin{aligned}
\mathrm{Vr}\left(\log\left(\frac{\Pi_d}{\pi(0)}\right)\right) &= \mathrm{Ex}\left(\left(\sum_{d'=1}^{d} X_{d'} - d\,\gamma\right)^2\right) \\
&= \mathrm{Ex}\left(\left(\sum_{d'=1}^{d} (X_{d'} - \gamma)\right)^2\right) \\
&= \mathrm{Ex}\left(\sum_{d'=1}^{d}\sum_{d''=1}^{d} (X_{d'} - \gamma)\,(X_{d''} - \gamma)\right) \\
&= \sum_{d'=1}^{d} \mathrm{Ex}\left((X_{d'} - \gamma)^2\right) = d\,\theta\,.
\end{aligned}$$

Growth Rate Probability Densities

Remark. The off-diagonal terms in the foregoing double sum vanish because

$$\mathrm{Ex}\Big((X_{d'} - \gamma)\,(X_{d''} - \gamma)\Big) = 0 \qquad \text{when } d'' \neq d' .$$

Hence, the growth rate expected value and variance of the IID model portfolio at day d is

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

Law of Large Numbers

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

$$Y_d = \frac{1}{d} \sum_{d'=1}^{d} X_{d'} \qquad \text{for every } d = 1,\, 2,\, \cdots \quad .$$

It is easy to check that

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

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

$$\lim_{d\to\infty} \mathrm{Pr}\{|Y_d - \gamma| \geq \delta\sqrt{\theta}\} = 0\,. \tag{4.5}$$

Law of Large Numbers

The convergence rate of this limit can be estimated by the *Chebyshev inequality*, which yields the δ-dependent upper bound

$$\Pr\{|Y_d - \gamma| \geq \delta\sqrt{\theta}\} \leq \frac{\mathrm{Vr}(Y_d)}{\delta^2\,\theta} = \frac{1}{\delta^2}\,\frac{1}{d}\,. \tag{4.6}$$

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

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

Law of Large Numbers

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

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

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

Remark. *The IID model suggests that the growth rate mean γ is a good proxy for the reward of a portfolio and that $\sqrt{\theta}$ is a good proxy for its risk. However, these are not the proxies chosen by MPT.*

Law of Large Numbers

The proxies γ and $\sqrt{\theta}$ can be approximated by $\hat{\gamma}$ and $\sqrt{\hat{\theta}}$ where $\hat{\gamma}$ and $\hat{\theta}$ are the unbiased estimators of γ and θ given by

$$\hat{\gamma} = \sum_{d=1}^{D} w_d \, X_d \,, \qquad \hat{\theta} = \sum_{d=1}^{D} \frac{w_d}{1 - \bar{w}} \left(X_d - \hat{\gamma} \right)^2 .$$

Normal Growth Rate Model

We can illustrate what is going on with the simple IID model where $p(X)$ is the *normal* or *Gaussian* density with mean γ and variance θ, which is given by

$$p(X) = \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(X-\gamma)^2}{2\theta}\right) .$$

Let $\{X_d\}_{d=1}^{\infty}$ be a sequence of IID random variables drawn from $p(X)$ and let $\{Y_d\}_{d=1}^{\infty}$ be the sequence of random variables defined by

$$Y_d = \frac{1}{d} \sum_{d'=1}^{d} X_{d'} \qquad \text{for every } d = 1, 2, \cdots .$$

Normal Growth Rate Model

We can easily check that

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

We can also check that

$$\mathrm{Ex}(Y_d | Y_{d-1}) = \frac{d-1}{d}\, Y_{d-1} + \frac{1}{d}\,\gamma\,.$$

So the variables Y_d are neither independent nor identically distributed.

It can be shown (the details are not given here) that Y_d is drawn from the normal density with mean γ and variance θ/d, which is given by

$$p_d(Y) = \sqrt{\frac{d}{2\pi\theta}}\, \exp\left(-\frac{(Y-\gamma)^2 d}{2\theta}\right)\,.$$

Normal Growth Rate Model

Because $\Pi_d/\pi(0) = e^{d\,Y_d}$, the mean return at day d is

$$\begin{aligned}\mathrm{Ex}\Big(e^{d\,Y_d}\Big) &= \sqrt{\frac{d}{2\pi\theta}}\int \exp\left(-\frac{(Y-\gamma)^2 d}{2\theta} + d\,Y\right)\,\mathrm{d}Y \\ &= \sqrt{\frac{d}{2\pi\theta}}\int \exp\left(-\frac{(Y-\gamma-\theta)^2 d}{2\theta} + d(\gamma+\tfrac{1}{2}\theta)\right)\,\mathrm{d}Y \\ &= \exp\Big(d(\gamma+\tfrac{1}{2}\theta)\Big)\,.\end{aligned}$$

Because $p_d(Y)$ becomes sharply peaked around $Y = \gamma$ as d increases, most investors will see the lower growth rate γ rather than $\gamma + \frac{1}{2}\theta$.

By setting $d = 1$ in the above formula, we see that the return mean is

$$\mu = \mathrm{Ex}(R) = \mathrm{Ex}\Big(e^X - 1\Big) = \exp(\gamma + \tfrac{1}{2}\theta) - 1\,.$$

Hence, $\mu \approx \gamma + \frac{1}{2}\theta$ when $|\gamma + \frac{1}{2}\theta| << 1$ and $\mu > \gamma + \frac{1}{2}\theta$ when $\mu \neq 0$.

Normal Growth Rate Model

Therefore most investors will see a return that is below the return mean μ — far below in volatile markets. This is because e^X amplifies the tail of the normal density. For a more realistic IID model with a density $p(X)$ that decays more slowly than a normal density as $X \to \infty$, this difference can be more striking. Said another way, most investors will not see the same return as Warren Buffett, but his return will boost the mean.

The normal growth rate model confirms that γ is a better proxy for how well a risky asset might perform than μ because $p_d(Y)$ becomes more peaked around $Y = \gamma$ as d increases. The Law of Large Numbers extends this result to IID models that are more realistic.

Portfolio Selection

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

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

Portfolio Selection

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

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

Kelly Criterion for a Simple Game

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

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

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

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

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

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

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

Kelly Criterion for a Simple Game

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

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

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

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

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

Kelly Criterion for a Simple Game

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

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

The law of large numbers implies that

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

Therefore for large n we see that

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

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

$$\gamma(f) = p \log(1 + fr) + (1 - p) \log(1 - f) \,. \tag{7.7}$$

This is now an exercise from first semester calculus.

Kelly Criterion for a Simple Game

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

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

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

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

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

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

Kelly Criterion for a Simple Game

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

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

Upon solving this equation for f_* we find that

$$f_* = \frac{p(1+r) - 1}{r} \,. \tag{7.8}$$

Kelly Criterion for a Simple Game

Remark. We see from (7.8) that if $p(1+r)-1>0$ then

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

Therefore the Kelly crtierion yields the optimal betting strategy

$$f_* = \begin{cases} 0 & \text{if } p(1+r)-1 \leq 0\,, \\ \dfrac{p(1+r)-1}{r} & \text{if } p(1+r)-1 > 0\,. \end{cases} \tag{7.9}$$

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

$$\gamma(f_*) = p\log\big(p(1+r)\big) + (1-p)\log\left((1-p)\,\frac{1+r}{r}\right). \tag{7.10}$$

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

Kelly Criterion for a Simple Game

Remark. Some bettors call r the *odds* because the return r on a winning wager is usually chosen so that the ratio $r : 1$ reflects a probability of winning. The expected return on each amount wagered is $pr - (1 - p)$. This is the probability of winning, p, times the return of a win, r, plus the probability of losing, $1 - p$, times the return of a loss, -1. Some bettors call this quantity the *edge* when it is postive. Notice that $pr - (1 - p) = p(1 + r) - 1$ is the numerator of f_* given by (7.8), while r is the denominator of f_* given by (7.8). Then strategy (7.9) can be expressed in this language as follows.

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

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

Kelly Criterion in Practice

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

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

Kelly Criterion in Practice

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

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

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

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

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

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

Kelly Criterion in Practice

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

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

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

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

Kelly Criterion in Practice

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

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

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

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

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

Kelly Criterion in Practice

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

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