The Kelly Growth Criterion

Niels Wesselhöfft
Dr. Wolfgang K. Härdle

International Research Training Group 1792
Ladislaus von Bortkiewicz Chair of Statistics
Humboldt–Universität zu Berlin

http://irtg1792.hu-berlin.de
http://lvb.wiwi.hu-berlin.de

Portfolio choice

- Playing Blackjack

Figure 1: 'Ed' Thorp

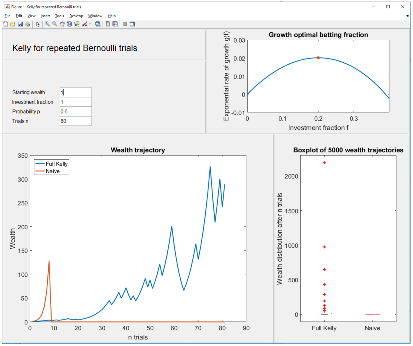

Figure 2: Matlab GUI

Portfolio choice

- Wealth for discrete returns $X_i \in \mathbb{R}^k$

$$W_n(f) = W_0 \prod_{i=1}^{n} \left(1 + \sum_{j=1}^{k} f_j X_{j,i} \right) \tag{1}$$

  - $W_0 \in \mathbb{R}^+$ starting wealth
  - $k \in \mathbb{N}^+$ assets with index j
  - $n \in \mathbb{N}^+$ periods with index i

- How to chose fraction vector $f \in \mathbb{R}^k$?

Managing Portfolio Risks

Two main strands

1. Mean-Variance approach: Markowitz (1952), Tobin (1958), Sharpe (1964) and Lintner (1965)
2. Kelly growth-optimum approach: Kelly (1956), Breiman (1961) and Thorp (1971)

Leo Breiman on BBI: 

Outline

1. Motivation ✓
2. Bernoulli - Kelly (1956)
3. Gaussian - Thorp (2006)
4. General i.i.d. - Breiman (1961)
5. Appendix

Arithmetic mean maximization

- Consider n favorable Bernoulli games with probability $\frac{1}{2} < p \leq 1$ $(q = 1 - p)$ and outcome $X = 1$ (-1)
- For $\mathrm{P}(X = 1) = p = 1$, investor bets everything

$$W_n = W_0 2^n \tag{2}$$

- Uncertainty - maximizing the expectation of wealth

$$\mathrm{E}(W_n) = W_0 + \sum_{i=1}^{n} (p - q)\,\mathrm{E}\left(fW_{n-1}\right), \tag{3}$$

- Leads to ruin asymptotically

$$\mathrm{P}\left(\{W_n \leq 0\}\right) = \mathrm{P}\left\{\lim_{n\to\infty} (1 - p^n)\right\} \to 1 \tag{4}$$

Minimizing risk of ruin

- Alternative: minimize the probability of ruin
- For $f = 0$

$$\mathrm{P}(\{W_n \leq 0\}) = 0 \tag{5}$$

- Minimum ruin strategy leads also to the minimization of the expected profits as no investment takes place

Geometric mean maximization

- Gambler bets a fraction of his wealth with m games won

$$W_n = W_0(1+f)^m(1-f)^{n-m} \tag{6}$$

- Exponential rate of growth per trial (log of the geometric mean)

$$G_n(f) = \log\left(\frac{W_n}{W_0}\right)^{\frac{1}{n}} = \log\left\{(1+f)^{\frac{m}{n}}(1-f)^{\frac{n-m}{n}}\right\} \tag{7}$$

$$= \left(\frac{m}{n}\right)\log(1+f) + \left(\frac{n-m}{n}\right)\log(1-f) \tag{8}$$

Geometric mean maximization

- By Borel's law of large numbers

$$\mathrm{E}\left\{G_n(f)\right\} = g(f) = p \cdot \log(1+f) + q \cdot \log(1-f) \tag{9}$$

- Maximizing $g(f)$ w.r.t. f:

$$g'(f) = \left(\frac{p}{1+f}\right) - \left(\frac{q}{1-f}\right) = \left\{\frac{p-q-f}{(1+f)(1-f)}\right\} = 0 \tag{10}$$

$$\star\ f = f^* = p - q, \quad p \geq q > 0 \tag{11}$$

- Second derivative according to f

$$g''(f) = -\left\{\frac{p}{(1+f)^2}\right\} - \left\{\frac{q}{(1-f)^2}\right\} < 0 \tag{12}$$

Closed form for Bernoulli trials

- Growth optimal fraction, under Bernoulli trials:

$$f^* = p - q \tag{13}$$

- Maximizes the expected value of the logarithm of capital at each trial

$$\begin{aligned} g(f^*) &= p \cdot \log(1 + p - q) + q \cdot \log(1 - p - q) && (14) \\ &= p \cdot \log(p) + q \cdot \log(q) + \log(2) > 0 && (15) \end{aligned}$$

A link to information theory

Bernoulli example, $p = 0.6$

- Exponential rate of asset growth for binary channel with p=0.6

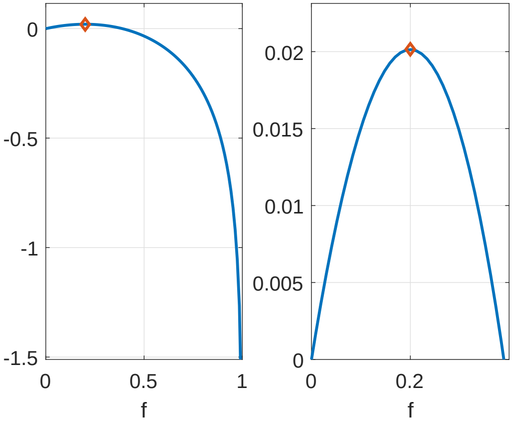

Figure 3: Bernoulli Exponential growth rate g(f)

Bernoulli

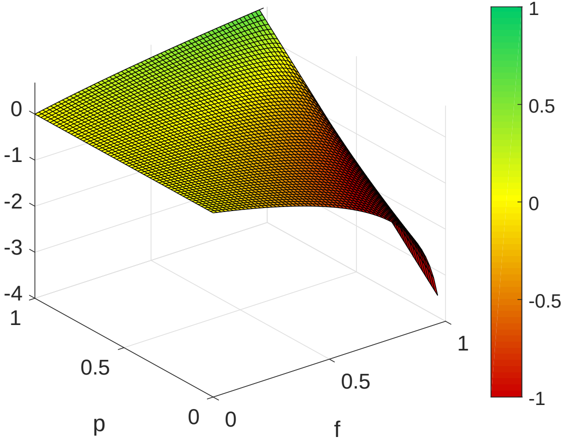

Figure 4: Bernoulli - Exponential growth rate g(f,p)

Gaussian (One-dimensional)

- $X \sim F$ with $\mathrm{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$
- Return of the risk free asset $r > 0$
- Wealth given investment fractions and restriction $\sum_{j=1}^{k} f_j = 1$

$$
\begin{aligned}
W(f) &= W_0 \left\{1 + (1 - f)r + fX\right\} && (16)\\
&= W_0 \left\{1 + r + f(X - r)\right\} && (17)
\end{aligned}
$$

Gaussian (One-dimensional)

- Maximize

$$g(f) = \mathsf{E}\{\log W_n(f)\} \quad = \mathsf{E}\{G(f)\} = \mathsf{E}\log\{W_n(f)/W_0\} \tag{18}$$

- Wealth after n periods

$$W_n(f) = W_0 \prod_{i=1}^{n} \{1 + r + f(X_i - r)\} \tag{19}$$

- Taylor expansion of

$$\mathsf{E}\left[\log\left\{\frac{W_n(f)}{W_0}\right\}\right] = \mathsf{E}\left[\sum_{i=1}^{n} \log\{1 + r + f(X_i - r)\}\right] \tag{20}$$

Gaussian (One-dimensional)

- Given $\log(1+x) = x - \frac{x^2}{2} + \cdots$

$$\log\{1+r+f(X-r)\} = r + f(X-r) - \frac{\{r+f(X-r)\}^2}{2} + \cdots \tag{21}$$

$$\approx r + f(X-r) - \frac{X^2 f^2}{2} \tag{22}$$

- Taking sum and expectation

$$\mathsf{E}\left[\sum_{i=1}^{n} \log\{1+r+f(X_i-r)\}\right] \approx r + f(\mu_n - r) - \frac{\sigma_n^2 f^2}{2} \tag{23}$$

- Myopia: taking $\sum_{i=1}^{n} X_i$ has no impact on the solution

Gaussian (One-dimensional)

- Result of the Taylor expansion

$$g(f) = r + f(\mu - r) - \sigma^2 f^2/2 + \mathcal{O}(n^{-1/2}). \tag{24}$$

- For $n \longrightarrow \infty$, $\mathcal{O}(n^{-1/2}) \longrightarrow 0$

$$g_\infty(f) = r + f(\mu - r) - \sigma^2 f^2/2. \tag{25}$$

- Differentiating $g(f)$ according to f

$$\frac{\partial g_\infty(f)}{\partial f} = \mu - r - \sigma^2 f = 0 \Rightarrow f^* = \frac{\mu - r}{\sigma^2} \tag{26}$$

- Betting the optimal fraction f^* leads to growth rate

$$g_\infty(f^*) = \frac{(\mu - r)^2}{2\sigma^2} + r. \tag{27}$$

- $g_\infty(f)$ is parabolic around f^* with range $0 \leq f^* \leq 2f^*$

Gaussian - $\mu = 0.03, \ \sigma = 0.15, \ r = 0.01$

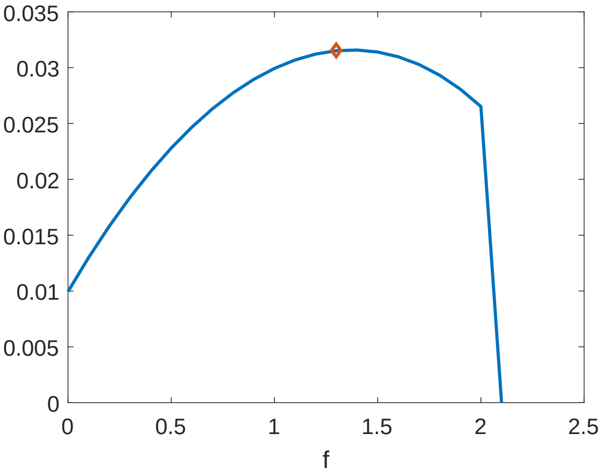

Figure 5: Gaussian approximation - Exponential growth rate g(f)

Gaussian (Multi-dimensional)

- $X \sim \mathrm{N}(\mu, \Sigma)$ and risk free rate $r > 0$

$$W_n(f) = W_0 \left\{ 1 + r + f^\top (X - r) \right\} \tag{28}$$

- Taking logarithm and expectations on both sides leads to $\mathrm{E}\left[\log\left\{W_n(f)/W_0\right\}\right]$, which is expanded in a Taylor series

$$g(f) = \mathrm{E}\left\{ \log(1+r) + \frac{1}{1+r}(\mu - 1r)^\top f - \frac{1}{2(1+r)^2} f^\top \Sigma f \right\} \tag{29}$$

- From quadratic optimization (Härdle and Simar, 2015)

$$f^* = \Sigma^{-1}(\mu - 1r) \tag{30}$$

$$g_\infty(f^*) = r + f^{*\top} \Sigma f^* / 2 \tag{31}$$

Gaussian - $\mu = [0.03\ 0.08],\ \sigma = [0.15\ 0.15],\ \rho = 0,\ r = 0.01$

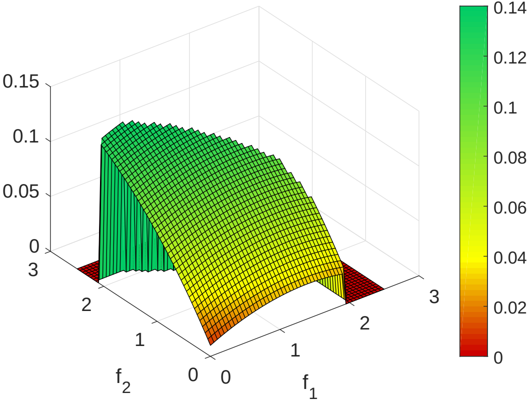

Figure 6: Gaussian approximation - Exponential growth rate g(f)

General i.i.d.

- Asymptotic dominance (in terms of wealth) of the Kelly strategy in a general i.i.d. setting in discrete time

Figure 7: Warren Buffett

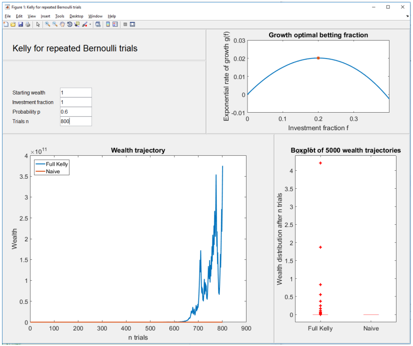

Figure 8: Matlab GUI

General i.i.d.

- Investment strategy $\Lambda = \begin{bmatrix} f_{i,j} & \cdots & f_{n,j} \\ \vdots & \ddots & \vdots \\ f_{i,k} & \cdots & f_{n,k} \end{bmatrix} = [f_i \cdots f_n]$
  - investment fractions f_i from time i to $n \in \mathbb{N}^+$
  - opportunities j to $k \in \mathbb{N}^+$
- Security price vector $p_i = \begin{bmatrix} p_{i,j} \\ \vdots \\ p_{i,k} \end{bmatrix}$
- Return per unit invested $x_i = \begin{bmatrix} \frac{p_{i,j}}{p_{i-1,j}} \\ \vdots \\ \frac{p_{i,k}}{p_{i-1,k}} \end{bmatrix}$.

Discrete i.i.d. setting

- Wealth of the investor in period n

$$W_n(f_n) = W_{n-1}(f_{n-1})\left\{f_n^\top x_n\right\} \tag{32}$$

- $W_n(f_n)$ increases exponentially
- Log-optimal fraction through growth rate maximization at each trial

$$f^* = \underset{f\in\mathbb{R}^k}{\operatorname{argmax}} \operatorname{E}\{\log(W_n)\} \tag{33}$$

Asymptotic outperformance

Theorem

- *Myopic log-optimal strategy* $\Lambda^* = [f^* \; \cdots \; f^*]$
- *Significantly different strategy* Λ

$$\mathrm{E}\{\log W_n(\Lambda^*)\} - \mathrm{E}\{\log W_n(\Lambda)\} \longrightarrow \infty, \tag{34}$$

- *Kelly investor dominates asymptotically*

$$\lim_{n\to\infty} \frac{W_n(\Lambda^*)}{W_n(\Lambda)} \xrightarrow{a.s.} \infty \tag{35}$$

Leo Breiman on BBI: 

Minimize time to reach goal g

Theorem

- *Let $N(g)$ be the smallest n, such that $W_i \geq g,\ \ g > 0$*
- *If equation (34) holds,*

$$\exists \alpha \geq 0 \perp\!\!\!\perp \Lambda,\ \ g \tag{36}$$

such that

$$\mathsf{E}\left\{N^*(g)\right\} - \mathsf{E}\left\{N(g)\right\} \leq \alpha, \tag{37}$$

- $\perp\!\!\!\perp$ *- independent of*
- Λ^* *asymptotically minimizes the time to reach goal g*

Time invariance

Theorem

- *Given a fixed set of opportunities the strategy is*
  - *fixed fraction*
  - *independent of the number of trials n*

$$\Lambda^* = [f_1^* \cdots f_n^*], \; f_1^* = \cdots = f_n^* \tag{38}$$

Bernoulli revisited

- For the repeated Bernoulli games of Kelly (1956)

Theorem

- *Two investors with equal initial endowment, investment fractions f_1 and f_2*
- *For exponential growth rates*

$$G_n(f_1) > G_n(f_2) \tag{39}$$

- *the Kelly bet dominates asymptotically*

$$\lim_{n \to \infty} \frac{W_n(f_1)}{W_n(f_2)} \xrightarrow{a.s.} \infty \tag{40}$$

Bernoulli revisited

Proof.

- Difference in exponential growth rates $G_n(f) = \log\left\{\frac{W_n(f)}{W_0}\right\}^{\frac{1}{n}}$

$$\log\left\{\frac{W_n(f_1)}{W_0}\right\}^{\frac{1}{n}} - \log\left\{\frac{W_n(f_2)}{W_0}\right\}^{\frac{1}{n}} = \log\left\{\frac{W_n(f_1)}{W_n(f_2)}\right\}^{\frac{1}{n}} \tag{41}$$

- by Borel strong law of large numbers

$$\mathrm{P}\left[\lim_{n\to\infty}\log\left\{\frac{W_n(f_1)}{W_n(f_2)}\right\}^{\frac{1}{n}}\right] > 0 \xrightarrow{a.s.} 1. \tag{42}$$

Bernoulli revisited

Proof.

- For $\omega \in \Omega$, there exists $N(\omega)$ such that for $n \geq N(\omega)$,

$$W_0 \exp\{nG(f_1)\} > W_0 \exp\{nG(f_2)\} \tag{43}$$

$$W_n(f_1) > W_n(f_2) \tag{44}$$

- Asymptotically

$$\lim_{n\to\infty} \frac{W_n(f_1)}{W_n(f_2)} \xrightarrow{a.s.} \infty \tag{45}$$

$\square$

Utility functions

- Three types of utility theories: Thorp (1971)
  - Descriptive utility - empirical data and mathematical fitting
  - Predictive utility - derives utility functions out of hypotheses
  - Normative utility - describe the behavior to achieve a certain goal
- The logarithmic utility function is used in a normative way

Conclusion

- Comparison of risk management theories
  - Markowitz-approach
    - arithmetic mean-variance efficient
    - maximizing single period returns
    - rests on two moments
  - Kelly-approach
    - geometric mean-variance efficient
    - maximize geometric rate of multi-period returns
    - utilizes the whole distribution

Information

▸ Closed form for Bernoulli trials

- Self-information (uncertainty) of outcome x

$$i(x) = -\log \mathrm{P}(x) = \log \frac{1}{\mathrm{P}(x)} \tag{46}$$

$$i(x) = 0, \ \text{for}\, \mathrm{P}(x) = 1 \tag{47}$$

$$i(x) > 1, \ \text{for}\, \mathrm{P}(x) < 1 \tag{48}$$

- Example: For a fair coin, the change of $\mathrm{P}\,(x = \{\text{tail}\}) = 0.5$

$$i(x) = -\log_2(1/0.5) = 1 \text{ bit}$$

Information

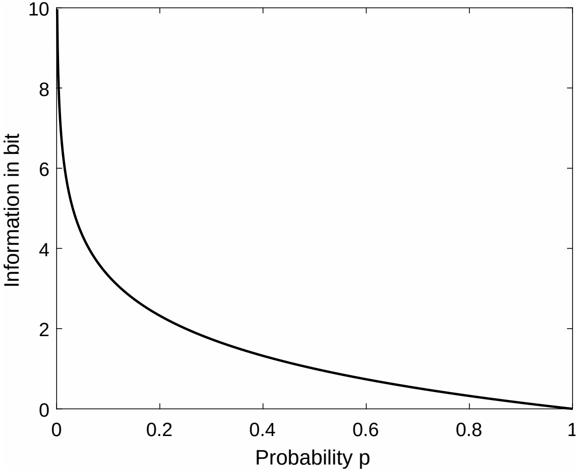

Figure 9: Self information of an outcome given probability p

Entropy

- Entropy as expectation of self-informations (average uncertainty), given outcomes $X = \{X_1, \ldots, X_n\}$

$$\begin{aligned} \mathrm{H}(X) &= \mathrm{E}\left[I(X)\right] = -\,\mathrm{E}\left\{\log \mathrm{P}(X)\right\} && (49)\\ &= -\sum_x \mathrm{P}(x) \log_2 \mathrm{P}(x) \geq 0 && (50) \end{aligned}$$

- For two outcomes and $p = q = 0.5$

$$\begin{aligned} H(X) &= -\left(p \log_2 p + q \log_2 q\right) \\ &= -\left(1/2 \log_2 1/2 + 1/2 \log_2 1/2\right) = 1 \text{ bit} \end{aligned}$$

Entropy

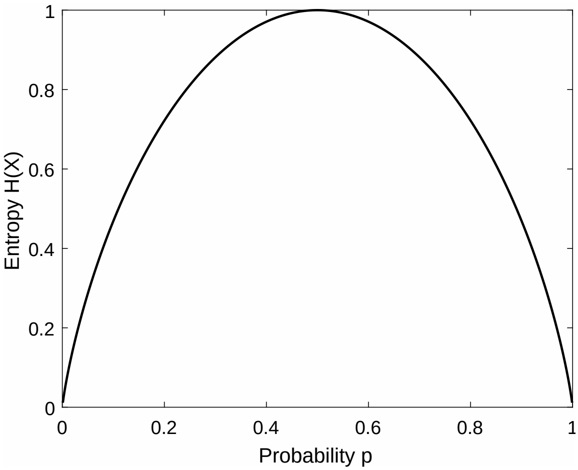

Figure 10: Entropy for two outcomes given probability p (1-p)

Entropy

- Joint entropy

$$H(X,Y) = -\operatorname{E}\{\log \operatorname{P}(X,Y)\} \quad (51)$$

$$= -\sum_{x,y} \operatorname{P}(x,y) \log \operatorname{P}(x,y) \quad (52)$$

- Conditional entropy

$$H(X \mid Y) = -\operatorname{E}\{\log \operatorname{P}(X \mid Y)\} \quad (53)$$

$$= -\sum_{x,y} \operatorname{P}(x \mid y) \log \operatorname{P}(x \mid y) \quad (54)$$

Noisy binary channel

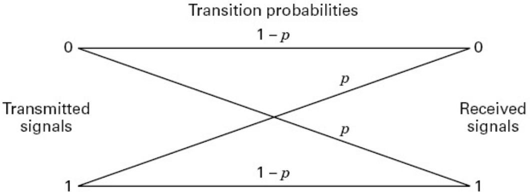

Figure 11: Noisy binary channel

Mutual information

- Mutual information

$$I(X;Y) = \mathrm{H}(X) - \mathrm{H}(X \mid Y) \tag{55}$$

$$= \mathrm{E}\left[\log \frac{\mathrm{P}(X \mid Y)}{\mathrm{P}(X)}\right] \tag{56}$$

- For the binary symmetric channel

$$I(X;Y) = \sum_x \sum_y \mathrm{P}(x,y) \log \frac{\mathrm{P}(x,y)}{\mathrm{P}(x)\,\mathrm{P}(y)} \tag{57}$$

$$= q \log(2q) + p \log(2p) \tag{58}$$

$$= p \log p + q \log q + \log(2) \tag{59}$$

Mutual information

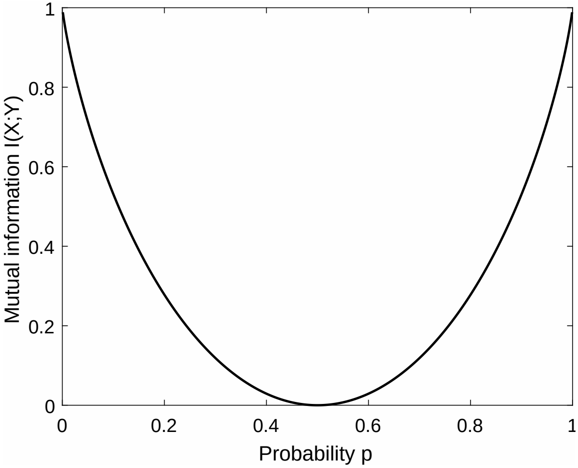

Figure 12: Mutual Information for a binary channel

Mutual information

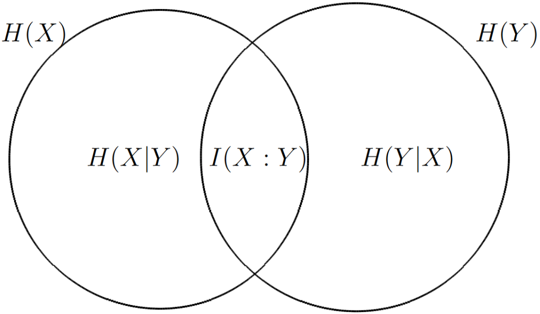

Figure 13: Relation of Entropy and Mutual Information

A link to information theory

- $I(X;Y)$ - mutual information
  - highest possible rate of information transmission in the presented channel
  - also called the channel's information carrying capacity or rate of transmission
- Equivalence to equation (14)

$$I(X;Y) = g(f^*) \tag{60}$$

Closed form for Bernoulli trials

A link to estimation theory

- Relative entropy or Kullback-Leibler divergence

$$D\left(\mathrm{P}(x)\,||\,\mathrm{Q}(x)\right) = -\,\mathrm{E}\left\{\log\frac{\mathrm{P}(x)}{\mathrm{Q}(x)}\right\} \tag{61}$$

$$= \sum_{x}\mathrm{P}(x)\log\frac{\mathrm{P}(x)}{\mathrm{Q}(x)} \geq 0 \tag{62}$$

- Relation to mutual information

$$I(X;Y) = D(\mathrm{P}(x,y)\,||\,\mathrm{P}(x)\,\mathrm{P}(y)) \tag{63}$$

The Kelly Growth Criterion

Niels Wesselhöfft
Dr. Wolfgang K. Härdle

International Research Training Group 1792
Ladislaus von Bortkiewicz Chair of Statistics
Humboldt–Universität zu Berlin

http://irtg1792.hu-berlin.de
http://lvb.wiwi.hu-berlin.de

For Further Reading

J. Kelly
A new interpretation of information rate
Bell System Technology Journal, 35, 1956

L. Breiman
Optimal gambling system for favorable games
Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, 1, 1961

E. O. Thorp
Portfolio choice and the Kelly criterion
Proceedings of the Business and Economics Section of the American Statistical Association, 1971

For Further Reading

R. Roll
Evidence on the growth optimum model
The Journal of Finance, 1973

L. C. MacLean, W. T. Ziemba and G. Blazenko
Growth versus Security in Dynamic Investment Analysis
Management Science, 38(11), 1992

E. O. Thorp
The Kelly criterion in Blackjack, Sports betting and the Stock Market
Handbook of Asset and Liability Management, 2006