

R 语言和金融量化投资中凯利公式的应用

MatrixSpk

目录

1 凯利公式简介	1
1.1 凯利公式的诞生	2
1.2 传奇人物爱德华·索普	2
1.3 凯利公式的基本原理	3
1.4 用 R 语言模拟凯利公式	4
2 凯利公式的扩展及其在股票投资中的利用	7
2.1 凯利公式的扩展	7
2.2 凯利公式在股票投资中的应用	8
3 多资产组合下的凯利公式	15
3.1 多资产组合下的凯利公式简介	15
3.2 多资产组合下凯利公式的 R 语言实现	15
4 分数凯利策略推导	17
5 结语	18

1 凯利公式简介

凯利公式是由信息学家凯利最早提出的一个用于解决信号传输过程中噪声干扰问题的数学公式，但由于该公式良好的通用性，其在游戏下注及证券投资中都发挥了巨大的作用，下面我们对凯利公式进行简单的介绍。

1.1 凯利公式的诞生

1956 年，在美国贝尔实验室工作的物理学家约翰·拉里·凯利为了解决长途电话信号传输中的噪声干扰问题¹，提出了凯利公式。后来，凯利发现这个公式也同样适用于赌博和投资领域。该公式可以帮助投资者在长期重复博弈中通过优化每次投注的比例来最大化长期资本增长率，同时也有助于避免破产风险。

凯利在凯利公式的诞生中居功甚伟，但凯利公式在投资领域的大规模应用却归功于另一位金融大鳄，他就是在 1987 年史称“黑色星期一”大股灾中逆势获利的投资家爱德华·索普。

1.2 传奇人物爱德华·索普

爱德华·索普是举世闻名的金融大鳄，也是一位数学家。他同时也是投资书籍《战胜市场》(Beat the Market) 和《战胜庄家》(Beat the Dealer) 的作者。此人是坚定的量化投资派。他坚信金融市场上的行为可以通过数学模型和概率统计来描述和预测。他主张人们在金融市场上要基于数据进行投资决策，而不是凭借主观感觉。

在索普眼中，投资无非是一场“概率游戏”，人们可以通过对金融市场建立模型来捕捉金融市场的规律，并寻找在大数据意义上有显著优势的投资策略。

索普在投资中还格外注重资金的重要性，这一点和巴菲特不谋而合。他认为获取长期稳定投资收益的核心在控制单笔投资的风险，而不是追求单笔投资的最高汇报。因而，找到一种方式来确定单笔投资的投资比例就十分重要。

很久以来，索普并未思索出来确定投资比例的最佳数学方法，直到有一天他的至交好友信息学家香农告诉了他关于凯利和凯利公式的故事。他们两人一起利用凯利公式击败了拉斯维加斯赌场的庄家。好莱坞电影《决胜 21 点》讲述的正是他们的故事。

后来索普又将凯利公式加以扩展并运用到股票投资中，并在《战胜市场》一书中大肆推广。他倡导投资者们基于凯利公式来确定最优投资比例以平衡

¹约翰·拉里·凯利，《噪声中信息率的新解释》

投资风险和收益，避免人们在投资过程中过于保守或者过度杠杆。

1.3 凯利公式的基本原理

假如我们正在玩一个纸牌游戏，在知悉游戏规则的情况下，我们清楚该游戏的胜率为 p ，对应的赔率为 b 。那么，我们该如何投注才能够使得我们在整个游戏中获得的资本增值最大呢？显然，这是一个数学问题。

我们可以假设初始资金为 C_0 游戏总共进行 N 轮，每次投注比例为 f ，获胜时资金将变为投资前的 $(1 + bf)$ 倍，失败时资金将变成 $1 - f$ 倍，进一步假设获胜次数为 n_1 ，那么失败次数为 n_2 ，且 $n = n_1 + n_2$ ，那么，最终资金 C_n 可表示为：

$$C_n = C_0(1 + bf)^{n_1} \cdot (1 - f)^{n_2}$$

对 C_n 取对数得：

$$\ln C_n = \ln C_0 + n_1 \ln(1 + bf) + n_2 \ln(1 - f)$$

根据大数定律，当 n 足够大时，有：

$$n_1 = pn$$

$$n_2 = qn$$

$$q = 1 - p$$

于是：

$$\ln C_n = \ln C_0 + npln(1 + bf) + nqln(1 - f)$$

对 $\ln C_n$ 求导数并令导数为零。容易知道答案是：

$$f^* = \frac{p \cdot b - q}{b}$$

上式即为凯利公式。

1.4 用 R 语言模拟凯利公式

下面用最简单好用的 R 语言来模拟一下凯利公式的效果：

```
set.seed(97) # 锁定随机种子，保证结果可重复
n <- 1000
p <- 0.52
b <- 1.05
initial_capital <- 10000

# 计算凯利公式的投资比例
q <- 1-p
f_kelly <- (b*p-q)/b
f_half_kelly <- f_kelly/2
fixed_fraction <- 0.1

# 模拟函数
simulate_kelly <-
  function(fraction,initial_capital,n){
    capital <- rep(initial_capital,n+1)
    for (i in 1:n){
      bet <- capital[i]*fraction
      outcome <- rbinom(1,1,p)
      if(outcome == 1){
        capital[i+1] <- capital[i]+bet*b
      }
      else {
        capital[i+1] <- capital[i]-bet
      }
    }
  }
```

```

    }
    return(capital)
  }

capital_kelly <-simulate_kelly(f_kelly,initial_capital,n)
capital_half_kelly <-simulate_kelly(f_half_kelly,initial_capital,n)
capital_fixed <- simulate_kelly(fixed_fraction,initial_capital,n)

# 创建数据框
df <- data.frame(
  step = 0:n,
  kelly = capital_kelly,
  half_kelly = capital_half_kelly,
  fixed = capital_fixed
)

# 转换为长格式
df_long <- reshape2::melt(df, id.vars = "step")

```

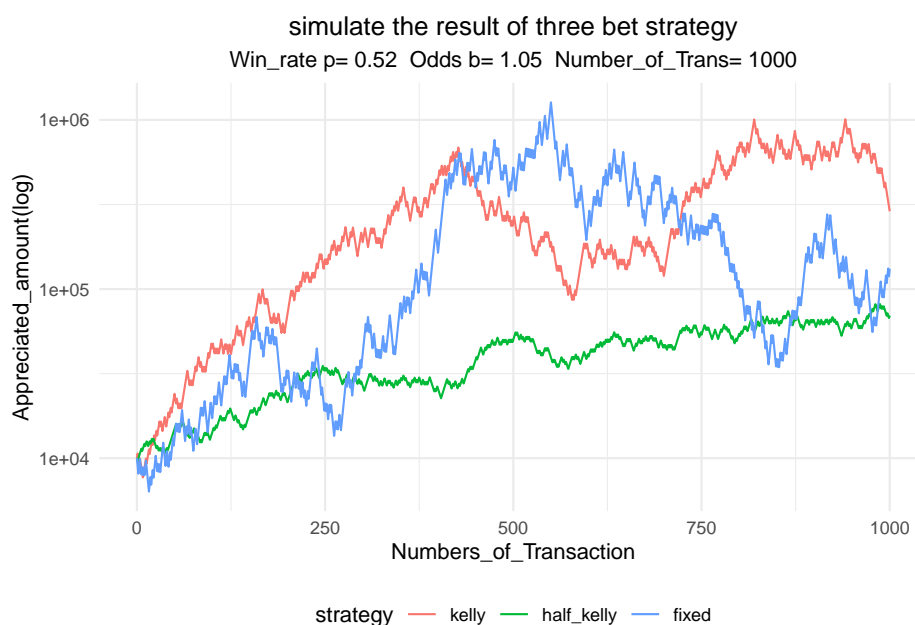
为了直观感受凯利公式的效果，可以绘制一幅简单的线图。

```

library(ggplot2)
ggplot(df_long, aes(x = step, y = value, color = variable)) +
  geom_line(linewidth = 0.5) +
  scale_y_log10() +
  labs(
    title = "simulate the result of three bet strategy",
    subtitle = paste("Win_rate p=", p, " Odds b=", b, " Number_of_Trans=", n),
    x = "Numbers_of_Transaction",
    y = "Appreciated_amount(log)",
    color = "strategy"
  ) +
  theme_minimal() +
  theme(legend.position = "bottom",

```

```
plot.title = element_text(hjust=0.5),
plot.subtitle = element_text(hjust=0.5)
)
```



查看一下三种下注策略各自对应的赌博结果：

```
# 输出最终结果
cat(" 凯利公式比例：", round(f_kelly, 4), "\n",
    " 半凯利比例：", round(f_half_kelly, 4), "\n",
    " 固定比例:", fixed_fraction, "\n",
    " 最终账户余额：\n",
    " 全凯利策略：", format(round(capital_kelly[n+1], 2), big.mark = ","), "\n",
    " 半凯利策略：", format(round(capital_half_kelly[n+1], 2), big.mark = ","), "\n",
    " 固定策略：", format(round(capital_fixed[n+1], 2), big.mark = ","), "\n"
)
```

```
## 凯利公式比例： 0.0629
```

```
## 半凯利比例： 0.0314
```

```
## 固定比例： 0.1
```

```
## 最终账户余额：
```

全凯利策略：288,432.5

半凯利策略：69,690.41

固定策略：131,676.3

由上图可以直观的看出，在利用凯利公式下注的情况下，长期资本增值率完胜其他两种方式。

2 凯利公式的扩展及其在股票投资中的利用

原始版本的凯利公式适用于赌博投注，因为它假设每次投注失败后下注者将完全失去下注的筹码。然而，人们在进行股票投资时，并非完全损失投资仓位，而只是损失了一部分比例。因此，如果想在股票投资中应用凯利公式，必须对凯利公式加以扩展。

2.1 凯利公式的扩展

假设我们已经掌握了一个投资策略，每次应用该投资策略时盈利的概率为 p ，失败的概率为 q ；盈利时的净盈利率为 r_w ，失败时的净损失率为 r_l ；则盈利时资本金增加为 $(1 + r_w)$ 倍，失败时本金变为 $(1 - r_l)$ 倍。

如果我们的初始资本为 C_0 ，进行 n 次投资，投资成功次数为 n_1 ，失败次数为 n_2 ，那么，最总资金 C_n 可表示为：

$$C_n = C_0(1 + r_w f)^{n_1}(1 - r_l f)^{n_2}$$

对上式取对数，并利用大数定律可知：

$$\ln C_n = \ln C_0 + n p \ln(1 + r_w f) - n q \ln(1 - r_l f)$$

对 $\ln C_n$ 关于 f 求导，并令导函数为零，可得：

$$f^* = \frac{p r_w - q r_l}{r_w r_l}$$

2.2 凯利公式在股票投资中的应用

熟悉蜡烛图形态分析的投资者应当知道，“早晨之星”是很常见的一个用来预示后续股市上涨的形态。我们可以利用 R 语言中的 `candlesticks` 包来甄别股票市场上满足“早晨之星”形态的股票，并将其纳入我们的投资组合。

以下是具体步骤和示例代码：

步骤 1：安装并加载包

```
install.packages("candlesticks") # 用于识别蜡烛图形态
install.packages("quantmod")    # 用于获取股票数据
library(candlesticks)
library(quantmod)
```

步骤 2：获取股票数据

```
# 以苹果公司 (AAPL) 为例
getSymbols("AAPL", from = "2023-01-01")
head(AAPL)
```

步骤 3：识别射击之星形态

`candlestick` 包中提供了 `CSPStar()` 函数来用以识别“早晨之星”和“黄昏之星”。举例如下：

```
MorningStar <- CSPStar(AAPL)[,"MorningStar"]
head(MorningStar)
```

	MorningStar
2020-01-02	NA
2020-01-03	NA
2020-01-06	FALSE
2020-01-07	FALSE
2020-01-08	FALSE
2020-01-09	FALSE


```
count <- sum(MorningStar[!is.na(MorningStar)])
cat("MorningStar 出现的次数是 :", count, "\n")
Date_MorningStar <- index(AAPL)[which(MorningStar[!is.na(MorningStar)]==TRUE)]
cat("MorningStar 出现的时间是 :", as.character(Date_MorningStar), "\n")
```

接下来我们来看一个常见的单一指标策略：假设我们在早晨之星形态出现当天以收盘价买入某支股票，盈利超过 15% 时止盈，亏损超过 10% 时止损，持有时间超过 60 个交易日止损。用 R 语言实现上述投资策略：

```
# 创建交易信号数据框
trades <- data.frame(
  entry_date = as.Date(character()),
  entry_price = numeric(),
  exit_date = as.Date(character()),
  exit_price = numeric(),
  type = character()
)

# 遍历每个信号
for(i in which(MorningStar==1)) {
  # 买入条件：当前为 Morning Star
  entry_date <- index(AAPL)[i]
  entry_price <- as.numeric(AAPL$Close[i])

  # 寻找卖出条件（从次日开始）
  for(j in (i+1):nrow(AAPL)) {
    current_return <- as.numeric(AAPL$Close[j])/entry_price - 1

    # 条件 1：收益率超过 15%
    if(current_return >= 0.15) {
      trades <- rbind(trades, data.frame(
        entry_date = entry_date,
        entry_price = entry_price,
        exit_date = index(AAPL)[j],
```

```
        exit_price = as.numeric(AAPL$Close[j]),
        type = " 止盈 "
    ))
    break
}

# 条件 2: 收益率低于-10%
if(current_return <= -0.10) {
    trades <- rbind(trades, data.frame(
        entry_date = entry_date,
        entry_price = entry_price,
        exit_date = index(AAPL)[j],
        exit_price = as.numeric(AAPL$Close[j]),
        type = " 止损 "
    ))
    break
}

# 条件 3: 持有 60 交易日
if(j - i == 60) {
    trades <- rbind(trades, data.frame(
        entry_date = entry_date,
        entry_price = entry_price,
        exit_date = index(AAPL)[j],
        exit_price = as.numeric(AAPL$Close[j]),
        type = " 时间止损 "
    ))
    break
}
}
}
```

4. 结果分析

```
# 计算总收益率
trades$returns <- (trades$exit_price / trades$entry_price) - 1
total_return <- sum(trades$returns)
cat(" 总收益率:", round(total_return*100, 2), "%\n")
cat(' 胜率:', round(sum(trades$returns>0)/length(trades$returns)*100,2), "%\n")
cat(' 单笔盈利:', round(total_return/length(trades$returns),2), "%\n")
# 查看交易记录
print(trades)
```

总收益率: 65.31 %

胜率: 64.29 %

单笔盈利: 0.05 %

	entry_date	entry_price	exit_date	exit_price	type	returns
1	2020-03-24	61.72	2020-04-14	71.7625	止盈	0.16271056
2	2020-11-11	119.49	2021-01-22	139.0700	止盈	0.16386317
3	2021-03-22	123.39	2021-06-16	130.1500	时间止损	0.05478559
4	2021-10-14	143.76	2021-12-07	171.1800	止盈	0.19073455
5	2021-12-15	179.30	2022-01-25	159.7800	止损	-0.10886784
6	2022-03-09	162.95	2022-05-11	146.5000	止损	-0.10095120
7	2022-04-04	178.44	2022-04-26	156.8000	止损	-0.12127325
8	2022-04-28	163.64	2022-05-11	146.5000	止损	-0.10474211
9	2022-09-19	154.48	2022-09-30	138.2000	止损	-0.10538581
10	2022-12-21	135.45	2023-03-16	155.8500	止盈	0.15060915
11	2023-04-27	168.41	2023-06-30	193.9700	止盈	0.15177244
12	2023-05-05	173.57	2023-08-02	192.5800	时间止损	0.10952350
13	2023-08-14	179.46	2023-11-07	181.8200	时间止损	0.01315057
14	2024-05-02	173.03	2024-06-11	207.1500	止盈	0.19719121

上述策略的胜率 p 等于 0.6531, r_w 等于 0.1327, r_l 等于 0.1082。根据扩展后的凯利公式可计算出最佳投资比例为：

$$f = 0.09$$

基于上述策略参数, 用 R 模拟长期投资结果：

```
set.seed(97) # 锁定随机种子，保证结果可重复
n <- 1000
p <- 0.6531 # 早晨之星的历史胜率
r_w <- 0.1327 # 早晨之星的历史盈利率均值
r_l <- 0.1082 # 早晨之星的历史损失率均值
initial_capital <- 10000

# 计算凯利公式的投资比例
q <- 1-p
f_kelly <- (p*r_w-q*r_l)/r_w*r_l
f_half_kelly <- f_kelly/2

# 模拟函数
simulate_kelly <- function(fraction,initial_capital,n,p){
  capital <- rep(initial_capital,n+1)
  for (i in 1:n){
    bet <- capital[i]*fraction
    outcome <- rbinom(1,1,p)
    if(outcome == 1){
      capital[i+1] <- capital[i]+bet*r_w
    }
    else {
      capital[i+1] <- capital[i]-bet*r_l
    }
  }
  return(capital)
}

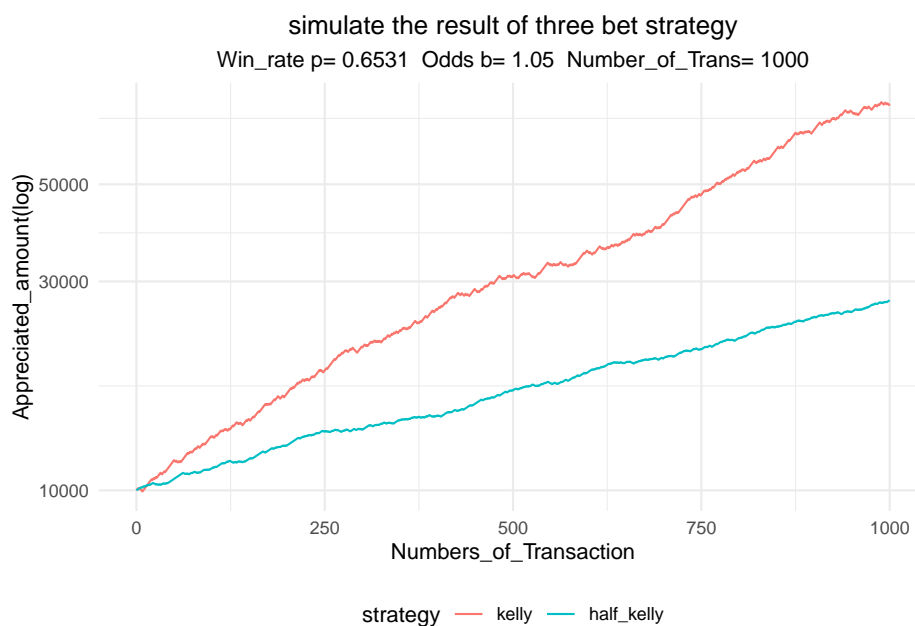
capital_kelly <-simulate_kelly(f_kelly,initial_capital,n,p)
capital_half_kelly <-simulate_kelly(f_half_kelly,initial_capital,n,p)

# 创建数据框
df <- data.frame(
```

```
step = 0:n,  
kelly = capital_kelly,  
half_kelly = capital_half_kelly  
)  
  
# 转换为长格式  
df_long <- reshape2::melt(df, id.vars = "step")
```

为了直观感受凯利公式的效果，可以绘制一幅简单的线图。

```
library(ggplot2)  
ggplot(df_long, aes(x = step, y = value, color = variable)) +  
  geom_line(linewidth = 0.5) +  
  scale_y_log10() +  
  labs(  
    title = "simulate the result of three bet strategy",  
    subtitle = paste("Win_rate p=", p, " Odds b=", b, " Number_of_Trans=", n),  
    x = "Numbers_of_Transaction",  
    y = "Appreciated_amount(log)",  
    color = "strategy"  
  ) +  
  theme_minimal() +  
  theme(legend.position = "bottom",  
        plot.title = element_text(hjust=0.5),  
        plot.subtitle = element_text(hjust=0.5)  
  )
```



查看一下上述策略对应的投资结果：

```
# 输出最终结果
cat(" 凯利公式比例：", round(f_kelly, 4), "\n",
    " 半凯利比例：", round(f_half_kelly, 4), "\n",
    " 最终账户余额：\n",
    " 全凯利策略：", format(round(capital_kelly[n+1], 2), big.mark = ","), "\n",
    " 半凯利策略：", format(round(capital_half_kelly[n+1], 2), big.mark = ","), "\n"
)
```

```
## 凯利公式比例： 0.0401
## 半凯利比例： 0.02
## 最终账户余额：
## 全凯利策略： 76,292.89
## 半凯利策略： 27,178.01
```

3 多资产组合下的凯利公式

在实际投资中，投资者通常不会投资单一股票，而是持有一个投资组合，即多资产组合。那么，如何将凯利公式应用到多资产组合情况下的投资头寸控制呢？事实上，凯利公式很容易从单一资产（例如，股票）扩展到多资产组合情况的。

3.1 多资产组合下的凯利公式简介

假设我们计划投资 m 种资产（例如， m 支股票），其对应的收益率向量为 $R = (R_1, R_2, \dots, R_m)^T$ ，其中 R_i 是第 i 种资产的收益率；投资比例向量为 $f = (f_1, f_2, \dots, f_m)^T$ ，其中 $\sum_{i=1}^m f_i = 1$ ， f_i 是投资于第 i 种资产的比例。

那么，资产组合的收益率为 $R_p = f^T R = \sum_{i=1}^m f_i R_i$ 。

为找到最优投资组合，可以引入对数效用函数 $U(f) = E[\ln(1 + f^T R)]$ ，目标是找到使 $U(f)$ 最大化的 f 。

对 $U(f)$ 关于 f 求梯度（多元函数求导），并令梯度为零向量。

假设收益率的均值向量为 $\mu = E[r]$ ，协方差矩阵为 $\Sigma = Cov(R, R)$ 。那么，可以解出最优投资比例向量 f^* 满足的方程，进而求解得到各资产的最优投资比例。

由于这个计算很复杂，所以在真正求解的时候，往往用利用数值求解方法。当然了，对我们而言，直接使用 R 语言求解即可。

3.2 多资产组合下凯利公式的 R 语言实现

假设我们已有各资产的历史收益率数据，并且存储在一个数据框中，每列代表一种资产的收益率。这里用随机数据模拟：

```
# 设定随机数种子，确保结果可复现
set.seed(27)
# 模拟 5 种资产，100 个时间周期的收益率数据
```

```

num_assets <- 5
num_periods <- 100
returns <- matrix(rnorm(num_assets * num_periods, mean = 0.05, sd = 0.2),
                  nrow = num_periods, ncol = num_assets)
colnames(returns) <- paste0("Asset", 1:num_assets)
returns_df <- as.data.frame(returns)

```

计算资产收益率均值和协方差矩阵

```

# 计算各资产的平均收益率
mean_returns <- colMeans(returns_df)
# 计算资产收益率的协方差矩阵
cov_matrix <- cov(returns_df)

```

定义广义凯利模型函数

```

generalized_kelly_model <- function(mean_returns, cov_matrix) {
  # 求解广义凯利模型的最优投资比例
  # 这里使用数值优化方法，因为解析解求解复杂
  library(quadprog)
  n <- length(mean_returns)
  Dmat <- 2 * cov_matrix
  dvec <- -2 * mean_returns
  Amat <- cbind(rep(1, n), diag(n))
  bvec <- c(1, rep(0, n))
  sol <- solve.QP(Dmat, dvec, Amat, bvec)
  return(sol$solution)
}

```

使用上述函数求解：

```

# 计算最优投资比例
optimal_weights <- generalized_kelly_model(mean_returns, cov_matrix)
# 展示结果
optimal_weights

```



```
## [1] 1.631395e-01 3.669821e-20 4.517315e-01 3.851290e-01 0.000000e+00
```

上述代码中：

- 首先，模拟生成资产收益率数据，实际应用中需替换为真实历史数据。
- 然后，计算资产收益率均值和协方差矩阵，这是模型输入关键参数。
- 定义的 `generalized_kelly_model` 函数，借助 `quadprog` 包（用于求解二次规划问题），根据广义凯利模型原理计算最优投资比例。该函数中，构建二次规划问题的目标函数和约束条件，目标是最大化资产组合的预期对数收益，约束包括投资比例之和为 1 以及各比例非负。
- 最后调用函数得到最优投资比例。

需注意，实际应用时：

- 要确保历史数据的准确性和代表性，数据质量影响模型效果。
- 模型假设市场是平稳的，但实际市场存在不确定性，可考虑结合其他风险评估方法调整投资组合。

4 分数凯利策略推导

由于凯利公式计算出投资比例在实际应用中，可能因为胜率、赔率等估计不准确导致过度投资，从而给投资者带来破产风险。故人们又总结出分数凯利策略。分数凯利策略并非严格数学推导，而是基于风险控制考虑对凯利公式的调整。

分数凯利公式建议投资者在每次投资时，只投资凯利公式建议投资额的一定比例 $k(0 \leq k \leq 1)$ ，即实际投注比例 $f_s = kf^*$ 。

例如，若凯利公式算出应投注资金的 40% ($f^* = 0.4$)，采用分数凯利策略，如 $k = 0.5$ ，则实际只投注 20% ($f_s = 0.2$)。这样在降低投注比例同时，减小风险和波动性，平衡收益与稳定性。

5 结语

本文是 R 语言与金融量化投资的系列文章之一，希望对读者有用。本文更新于 2025 年 4 月 9 日。