
Portfolio Construction and Analytics 读书笔记

目录	4
Contents	1
Contents	1
1 资产管理的介绍	1
2 随机变量、概率分布和重要的统计概念	1
2.1 随机变量	1
2.2 伯努利试验和概率分布函数	1
2.3 n重伯努利试验	1
2.4 正态分布和概率分布函数	1
2.5 累积分布函数	1
2.6 描述分布	1
2.6.1 集中趋势的度量	1
2.6.2 风险的度量	2
2.6.3 偏度	3
2.6.4 峰度	3
2.7 协方差和相关系数	4
2.8 随机变量和的性质	4
2.9 联合概率分布和条件概率	4
2.10 Copulas	5
2.11 概率分布和取样	6
2.11.1 中心极限定理	7
2.11.2 置信区间	7
2.11.3 Bootstrapping	7
2.11.4 假设检验	7
3 常见的分布函数介绍	8
3.1 分布函数的样例	8
3.1.1 记号说明	8

3.1.2	离散型和连续型均匀分布	8
3.1.3	t分布	8
3.1.4	对数正态分布	9
3.1.5	泊松分布	10
3.1.6	指数分布	10
3.1.7	卡方分布	11
3.1.8	伽玛分布	11
3.1.9	贝塔分布	12
3.2	金融回报率的分布模型	12
3.2.1	椭圆分布族	12
3.2.2	稳定Paretian分布族	12
3.2.3	广义 λ 分布族	13
3.3	金融回报率的尾部风险模型	13
3.3.1	广义极值分布	13
3.3.2	广义帕累托分布	14
3.3.3	极值模型	14
4	统计学模型	15
4.1	经典收益模型	15
4.2	回归分析	15
4.2.1	一个简单的例子	15
4.2.2	回归分析在投资中的应用	16
4.3	因子分析	16
4.4	主成分分析	16
4.5	自回归条件方差模型	17
5	模型模拟	18
5.1	蒙特卡罗模拟	18
5.1.1	选择分布函数	18
5.1.2	理解蒙特卡罗模拟的输出	18
5.2	为什么采用蒙特卡罗模拟	19
5.2.1	多个输入变量和混合分布	19
5.2.2	合并相关	19

5.2.3	模型评估	19
5.2.4	模拟多少次?	20
5.2.5	随机数的生成	20
6	模型优化	21
6.1	优化公式	21
6.1.1	最大化和最小化	21
6.1.2	局部最优和全局最优	21
6.1.3	多目标优化	21
6.2	重要的优化问题	23
6.2.1	凸优化	23
6.2.2	线性规划	23
6.2.3	二次规划	23
6.2.4	二阶锥规划	24
6.2.5	整数规划	24
6.3	一个简单的优化例子:资产分配	24
6.4	优化算法	26
6.5	优化软件	27
6.6	一个求解的例子	27
6.6.1	Excel求解	27
6.6.2	求解结果	27
7	非确定优化	28
7.1	动态规划	28
7.2	随机规划	28
7.2.1	多级模型	28
7.2.2	风险中性随机模型	30
7.2.3	机会约束模型	31
7.3	鲁棒优化	32
8	资产多样化	33
8.1	资产分散的例子	33
8.2	经典均值方差优化模型	35

8.3	有效边界理论	36
8.4	经典均值方差优化模型	36
8.4.1	预期回报公式	36
8.4.2	风险厌恶公式	38
8.5	资本市场线	38
8.6	期望效用函数理论	39
8.6.1	二次效用函数	39
8.6.2	线性效用函数	39
8.6.3	指数效用函数	40
8.6.4	幂效用函数	40
8.6.5	logistic效用函数	40
8.6.6	再定义多样化	40
9	因子模型	41
9.1	金融经济学文献中的因子模型	41
9.2	因子模型的均值方差优化	42
9.3	实践中的因子选择	43
9.4	阿尔法构建中的因子模型	43
9.5	因子模型用于风险评估	45
9.5.1	宏观经济因子模型	45
9.5.2	基本面因子模型	45
9.5.3	统计因子模型	46
9.5.4	混合因子模型	46
9.5.5	选择“适合的”因子模型	47
9.6	数据管理与质量问题	47
9.6.1	数据调整	47
9.6.2	幸存者偏差	48
9.6.3	前视偏差	48
9.6.4	数据迁就	48
9.7	风险分解、风险归因和业绩归因	49
9.8	因子投资	49

1 资产管理的介绍

2 随机变量、概率分布和重要的统计概念

2.1 随机变量

随机变量：定义在样本空间 ω 上的实值函数。

2.2 伯努利试验和概率分布函数

设伯努利试验一次成功的概率为 p 那么一重伯努利试验有如下的概率分布函数：

$$\Pr(\tilde{X} = x) = \begin{cases} 1 - p & x=0 \\ p & x=1 \end{cases}$$

2.3 n重伯努利试验

设伯努利试验一次成功的概率为 p 那么 n 重伯努利试验成功 x 次的概率为：

$$\Pr(\tilde{X} = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, x = 0, \dots, n$$

2.4 正态分布和概率分布函数

正态分布：

$$f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

概率分布函数(PDF): 表示随机变量在样本空间上的概率分布：

$$\Pr[a \leq X \leq b] = \int_a^b f_X(x) dx$$

2.5 累积分布函数

累积分布函数(CDF):

$$F(x) = \Pr[X \leq b] = \int_{-\infty}^b f_X(x) dx$$

2.6 描述分布

2.6.1 集中趋势的度量

2.6.1.1 均值：

$$\mu = E[X] = \int_{-\infty}^{\infty} xP(x) dx,$$

2.6.1.2 方差:

$$\text{Var}(X) = E[(X - E[X])^2]$$

2.6.1.3 k阶中心矩:

$$\mu_k = E[(X - \mu)^k] = \int_{-\infty}^{\infty} (x - \mu)^k P(x) dx,$$

4.矩量母函数

$$M_t(X) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} P(x) dx,$$

2.6.2风险的度量

2.6.2.1 方差和标准差

在度量投资组合的风险时，首要考虑的就是投资组合的方差和标准差：

1.方差:

$$\begin{aligned}\text{Var}(X) &= E[(X - E[X])^2] \\ &= E[X^2 - 2X E[X] + E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 \\ &= E[X^2] - E[X]^2\end{aligned}$$

2.标准差: $\sigma_X = \sqrt{\text{Var}(X)}$

2.6.2.2 变异系数

当需要比较两组数据离散程度大小的时候，如果两组数据的测量尺度相差太大，或者数据量纲的不同，直接使用标准差来进行比较不合适，此时就应当消除测量尺度和量纲的影响，而变异系数可以做到这一点，它是原始数据标准差与原始数据平均数的比。**CV**没有量纲，这样就可以进行客观比较了。事实上，可以认为变异系数和极差、标准差和方差一样，都是反映数据离散程度的绝对值。其数据大小不仅受变量值离散程度的影响，而且还受变量值平均水平大小的影响。

变异系数的定义：

$$c_v = \frac{\sigma}{\mu}$$

例如：投资组合A的变异系数为0.8,投资组合B的变异系数为0.5，那么我们可以认为投资组合A的风险比较大。

2.6.2.3 范围

随机变量的范围：即随机变量的取值范围。例如正态分布的取值范围是负无穷到正无穷。

2.6.2.4 百分位数

随机变量 X 或它的概率分布的分位数 Z_α ，是指满足条件 $\Pr(X \leq Z_\alpha) = \alpha$ 的实数

2.6.2.5 风险价值

风险价值 $\text{VaR}(\text{Value at Risk})$ ：在市场正常波动下，某一金融资产或证券组合的最大可能损失。更为确切的是指，在一定概率水平（置信度）下，某一金融资产或证券组合价值在未来特定时期内的最大可能损失。

给定置信度 α ：

$$\text{VaR}_\alpha(X) = \inf \{x \in \mathbb{R} : F_X(x) > \alpha\} = F_Y^{-1}(1 - \alpha).$$

2.6.2.6 条件风险价值

在投资组合超过某个给定 VaR 值的条件下，该投资组合的平均损失值。

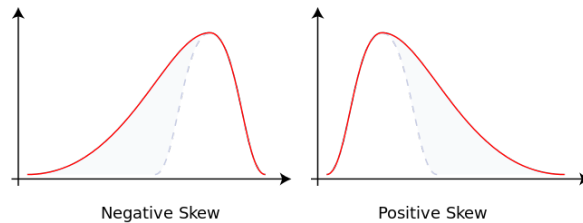
$$\text{CVaR}_\alpha(X) = E[-X \mid X \leq -\text{VaR}_\alpha(X)]$$

2.6.3 偏度

偏度（skewness），是统计数据分布偏斜方向和程度的度量，是统计数据分布非对称程度的数字特征。我们可以从图片中（来自wiki百科）直观地看出正偏度和负偏度：

1. 负偏度。密度函数左边的尾巴更加厚实，随机变量主要的取值分布在右边，通常我们也把他称为”右倾斜”。

2. 正偏度。密度函数右边的尾巴更加厚实，随机变量主要的取值分布在左边，通常我们也把他称为”左倾斜”。



计算公式：

$$\gamma_1 = E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right] = \frac{\mu_3}{\sigma^3} = \frac{E[(X - \mu)^3]}{(E[(X - \mu)^2])^{3/2}} = \frac{\kappa_3}{\kappa_2^{3/2}}$$

2.6.4 峰度

峰度是描述总体中所有取值分布形态陡缓程度的统计量。这个统计量需要与正态分布相比较，峰度为3表示该总体数据分布与正态分布的陡缓程度相同；峰度大于3表示该总体数据分布与正态分布相比较为陡峭，为尖顶峰；峰度小于3表示该总体数据分

布与正态分布相比较为平坦，为平顶峰。峰度的绝对值数值越大表示其分布形态的陡缓程度与正态分布的差异程度越大。

计算公式：

$$\text{Kurt}[X] = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] = \frac{\mu_4}{\sigma^4} = \frac{E[(X - \mu)^4]}{(E[(X - \mu)^2])^2}$$

2.7 协方差和相关系数

协方差:用于衡量两个变量的总体误差。而方差是协方差的一种特殊情况，即当两个变量是相同的情况。期望值分别为 $E[X]$ 与 $E[Y]$ 的两个实随机变量 X 与 Y 之间的协方差 $\text{Cov}(X, Y)$ 定义为：

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

相关系数：研究变量之间线性相关程度的量。随机变量 X 和 Y 的相关系数定义为：

$$\rho_{X,Y} = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

2.8 随机变量和的性质

1.随机变量和的期望。

$$E[aX + bY] = a \cdot E[X] + b \cdot E[Y]$$

2.随机变量和的方差。

$$\text{Var}[aX + bY] = a^2 \cdot \text{Var}[X] + b^2 \cdot \text{Var}[Y] + 2 \cdot a \cdot b \cdot \text{Cov}(X, Y)$$

2.随机变量和的分布。对于独立的随机变量 X 和 Y , 随机变量 $Z = X + Y$ 的密度函数就是 X 的密度函数和 Y 的密度函数的卷积。

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(z - x)f_X(x) dx \quad f_Z(z) = \int_{-\infty}^{\infty} f_Y(z - x)f_X(x) dx$$

2.9 联合概率分布和条件概率

对于离散型随机变量 X, Y ，当 $X=x$ 时， Y 的条件分布为：

$$P_Y(y | X = x) = P(Y = y | X = x) = \frac{P(X = x \cap Y = y)}{P(X = x)}$$

同样对于连续型随机变量，当 $X=x$ 时， Y 的条件分布为：

$$f_Y(y | X = x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

一般的我们有如下结论：

1. 重期望公式： $E(E(X | \mathcal{H})) = E(X)$
2. 条件方差公式： $\text{Var}(X) = E(\text{Var}(X | \mathcal{H})) + \text{Var}(E(X | \mathcal{H}))$
3. 条件协方差公式： $\text{cov}(X, Y) = E(\text{cov}(X, Y | Z)) + \text{cov}(E(X | Z), E(Y | Z))$

边缘分布和密度函数

对于离散型随机变量：

$$\Pr(X = x) = \sum_y \Pr(X = x, Y = y) = \sum_y \Pr(X = x | Y = y) \Pr(Y = y)$$

对于连续型随机变量：

$$p_X(x) = \int_y p_{X,Y}(x, y) dy = \int_y p_{X|Y}(x | y) p_Y(y) dy$$

2.10 Copulas

copula函数描述的是变量间的相关性，实际上是一类将联合分布函数与它们各自的边缘分布函数连接在一起的函数，因此也有人将它称为连接函数。相关理论的提出可以追溯到1959年，Sklar通过定理形式将多元分布与Copula函数联系起来。

Copula函数的定义： $C : [0, 1]^d \rightarrow [0, 1]$ 称为是一个d维copula函数如果：

- $C(u_1, \dots, u_{i-1}, 0, u_{i+1}, \dots, u_d) = 0$, 如果某个分量为零，函数为零，
- $C(1, \dots, 1, u, 1, \dots, 1) = u$, 如果函数的d-1个分量为1，那么函数是把u映成u，
- C 对于它的每一个分量是非减的。

Sklar定理

Sklar 定理（二元形式）：若 $H(x, y)$ 是一个具有连续边缘分布的 $F(x)$ 与 $G(y)$ 的二元联合分布函数，那么存在唯一的copula函数 C ，使得 $H(x, y) = C(F(x), G(y))$ 。反之，如果 C 是一个copula函数，而 F 和 G 是两个任意的概率分布函数，那么由上式定义的 H 函数一定是一个联合分布函数，且对应的边缘分布刚好就是 F 和 G 。

Copulas函数族

1. 高斯Copula函数

高斯Copula函数是定义在单位立方体 $[0, 1]^d$ 上面的函数。它是通过定义在 \mathbb{R}^d 上的多元正态函数构造的。

给定一个相关系数矩阵 $R \in [-1, 1]^{d \times d}$, 参数为 R 的高斯Copula函数可以写成:

$$C_R^{\text{Gauss}}(u) = \Phi_R(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)),$$

其中 Φ^{-1} 标准正态分布函数的逆而 Φ_R 是均值为零、协方差为 R 的多元正态分布。

2.阿基米德Copula函数

函数 C 被称为阿基米德copula 如果它满足如下条件:

$$C(u_1, \dots, u_d; \theta) = \psi^{[-1]}(\psi(u_1; \theta) + \dots + \psi(u_d; \theta); \theta)$$

其中 $\psi: [0, 1] \times \Theta \rightarrow [0, \infty)$ 是连续、严格递减的凸函数, 并且满足: $\psi(1; \theta) = 0$ 。 θ 从属于某个参数空间 Θ 。

常用的阿基米德copula函数:

名称	$C_\theta(u, v)$	参数 θ
Ali-Mikhail-Haq	$\frac{uv}{1 - \theta(1-u)(1-v)}$	$\theta \in [-1, 1)$
Clayton	$[\max\{u^{-\theta} + v^{-\theta} - 1; 0\}]^{-1/\theta}$	$\theta \in [-1, \infty) \setminus \{0\}$
Frank	$-\frac{1}{\theta} \log \left[1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1} \right]$	$\theta \in \mathbb{R} \setminus \{0\}$
Gumbel	$\exp \left[- \left((-\log(u))^\theta + (-\log(v))^\theta \right)^{1/\theta} \right]$	$\theta \in [1, \infty)$
Independence	uv	
Joe	$1 - [(1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta]^{1/\theta}$	$\theta \in [1, \infty)$

2.11 概率分布和取样

在通常情况下, 我们无法知道总体的分布情况。所以我们通常通过样本来估计总体。例如, 我们独立地观测到 n 个样本数据: X_1, \dots, X_n , 通常采用如下的公式估计总体的均值、方差等参数:

样本均值:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

样本方差:

$$s_2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

样本标准差:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

样本协方差:

$$sCov(\bar{X}, \bar{Y}) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

样本相关系数:

$$r(\bar{X}, \bar{Y}) = \frac{sCov(\bar{X}, \bar{Y})}{s_X s_Y}$$

2.11.1 中心极限定理

独立同分布的中心极限定理:

设随机变量 X_1, X_2, \dots, X_n 独立同分布, 并且具有有限的数学期望和方差: $E(X_i) = \mu$, $Var[X_i] = \sigma^2 < \infty$,

则对任意 z :

$$\lim_{n \rightarrow \infty} \Pr [\sqrt{n}(S_n - \mu) \leq z] = \Phi\left(\frac{z}{\sigma}\right),$$

其中

$$S_n := \frac{X_1 + \dots + X_n}{n}$$

为样本的均值。

2.11.2 置信区间

置信区间是一种常用的区间估计方法, 所谓置信区间就是分别以统计量的置信上限和置信下限为上下界构成的区间。对于一组给定的样本数据, 其平均值为 μ , 标准偏差为 σ , 则其整体数据的平均值的 $100(1-\alpha)\%$ 置信区间为 $(\mu - Z_{\frac{\alpha}{2}}\sigma, \mu + Z_{\frac{\alpha}{2}}\sigma)$, 其中 α 为非置信水平在正态分布内的覆盖面积 $Z_{\frac{\alpha}{2}}$ 即为对应的标准分数。

2.11.3 Bootstrapping

Bootstrapping 算法, 指的就是利用有限的样本资料经由多次重复抽样, 重新建立起足以代表母体样本分布的新样本。我们会在后续的章节中结合蒙特卡洛模拟给出详细的介绍。

2.11.4 假设检验

假设检验(Hypothesis Testing)是数理统计学中根据一定假设条件由样本推断总体的一种方法。具体作法是: 根据问题的需要对所研究的总体作某种假设, 记作 H_0 ; 选取合适的统计量, 这个统计量的选取要使得在假设 H_0 成立时, 其分布为已知; 由实测的样本, 计算出统计量的值, 并根据预先给定的显著性水平进行检验, 作出拒绝或接受假设 H_0 的判断。常用的假设检验方法有 u 检验法、 t 检验法、 χ^2 检验法(卡方检验)、 F 检验法, 秩和检验等。我们会在后续的章节中给出详细的介绍。

3 常见的分布函数介绍

3.1 分布函数的样例

3.1.1 记号说明

Gama 函数的定义如下：

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx$$

Beta函数的定义如下：

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt$$

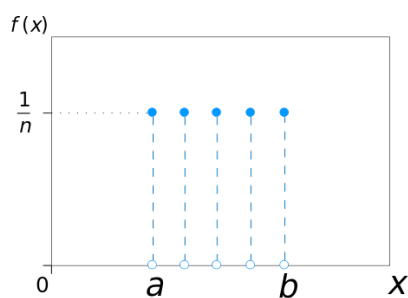
3.1.2 离散型和连续型均匀分布

离散型均匀分布：

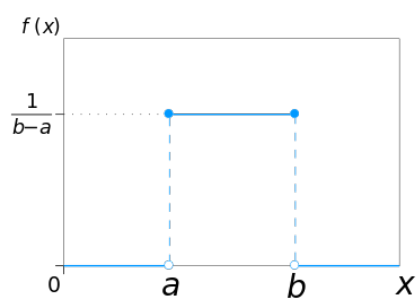
$$\Pr(X = x) = \frac{1}{N}$$

连续型均匀分布：

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$



3-1 离散型均匀分布



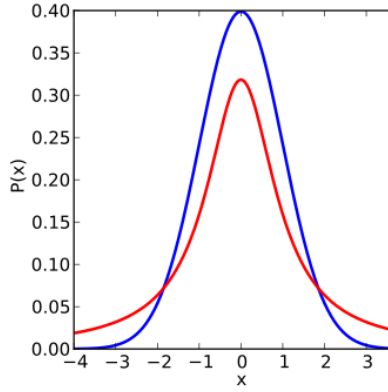
3-2 连续型均匀分布

3.1.3 t分布

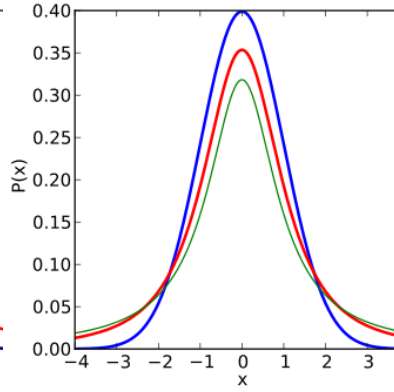
密度函数：

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

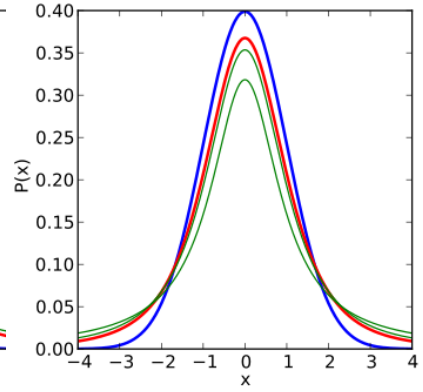
自由度分别为1, 2, 3的t分布的密度函数。其中蓝色的线条表示正态分布的密度函数。绿色的线条表示上一幅图中t分布的密度函数。



3-3 自由度为1



3-4 自由度为2



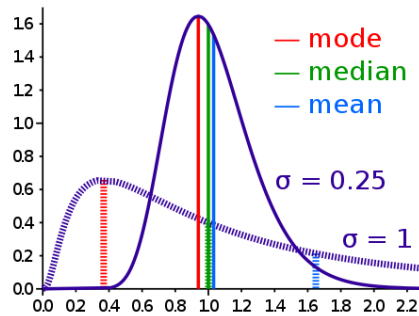
3-5 自由度为3

3.1.4 对数正态分布

对数正态分布（logarithmic normal distribution）是指一个随机变量的对数服从正态分布，则该随机变量服从对数正态分布。对数正态分布从短期来看，与正态分布非常接近。但长期来看，对数正态分布向上分布的数值更多一些。

密度函数：

$$\begin{aligned}
 f_X(x) &= \frac{d}{dx} \Pr(X \leq x) = \frac{d}{dx} \Pr(\ln X \leq \ln x) \\
 &= \frac{d}{dx} \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \\
 &= \varphi\left(\frac{\ln x - \mu}{\sigma}\right) \frac{d}{dx} \left(\frac{\ln x - \mu}{\sigma}\right) \\
 &= \varphi\left(\frac{\ln x - \mu}{\sigma}\right) \frac{1}{\sigma x} \\
 &= \frac{1}{x} \cdot \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right).
 \end{aligned}$$

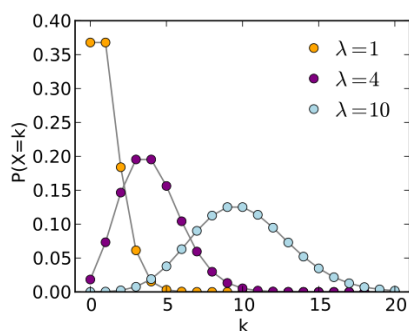


3-6 两种不同偏度的对数正态分布

3.1.5 泊松分布

密度函数：

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$



3-7 不同 λ 值的泊松分布

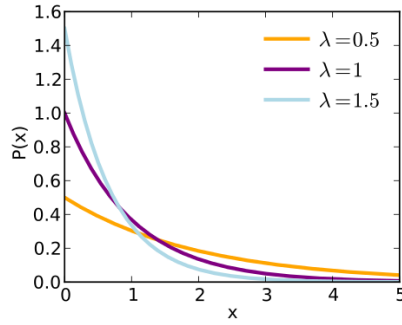
值得一提的是，随着 λ 的递增，泊松分布会不断逼近均值为 λ ，方差为 $\sqrt{\lambda}$ 的正态分布，这一结论在金融模型中有重要的应用。

3.1.6 指数分布

在概率理论和统计学中，指数分布（也称为负指数分布）是描述泊松过程中的事件之间的时间的概率分布，即事件以恒定平均速率连续且独立地发生的过程。这是伽马分布的一个特殊情况。

密度函数：

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$



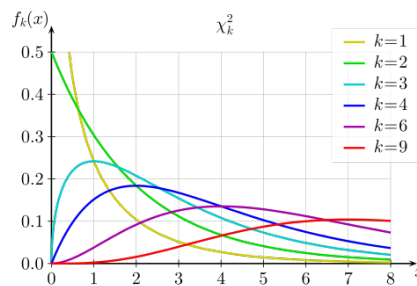
3-8 不同参数的指数分布

3.1.7 卡方分布

若 n 个相互独立的随机变量 $\xi_1, \xi_2, \dots, \xi_n$ ，均服从标准正态分布（也称独立同分布于标准正态分布），则这 n 个服从标准正态分布的随机变量的平方和构成一新的随机变量，其分布规律称为卡方分布（chi-square distribution）。

密度函数：

$$f(x; k) = \begin{cases} \frac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})}, & x > 0; \\ 0, & \text{otherwise.} \end{cases}$$



3-9 不同参数的卡方分布

3.1.8 伽玛分布

伽玛分布（Gamma Distribution）是统计学的一种连续概率函数，是概率统计中一种非常重要的分布。”指数分布”和” χ^2 分布”都是伽马分布的特例。

Gamma分布中的参数 α 称为形状参数（shape parameter）， β 称为尺度参数（scale parameter）。

密度函数：

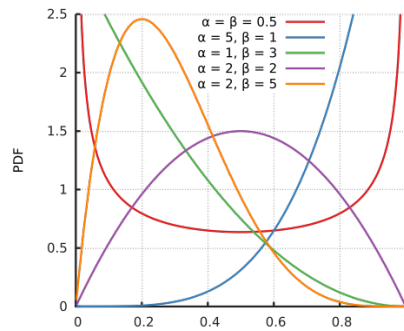
$$f(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \quad \text{for } x > 0 \text{ and } \alpha, \beta > 0,$$

3.1.9 贝塔分布

在概率论中，贝塔分布，也称 B 分布，是指一组定义在(0,1) 区间的连续概率分布。

密度函数：

$$\begin{aligned} f(x; \alpha, \beta) &= \text{constant} \cdot x^{\alpha-1}(1-x)^{\beta-1} \\ &= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1} du} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \\ &= \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} \end{aligned}$$



3-10 不同参数的贝塔分布

3.2 金融回报率的分布模型

3.2.1 椭圆分布族

椭圆分布的密度函数具有如下形式：

$$f(x) = \frac{c}{\sqrt{|\Sigma|}} \cdot g((x - \mu)' \Sigma^{-1} (x - \mu))$$

例如，我们所熟知的多元正态分布就属于椭圆分布族：

$$f_{\mathbf{X}}(x_1, \dots, x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}}$$

3.2.2 稳定Paretian分布族

稳定Paretian分布族主要包括如下三个分布：正态分布，柯西分布，列维分布。例

如我们所熟知的柯西分布，其密度函数为：

$$f(x) = \frac{1}{\pi\gamma \left[1 + \left(\frac{x-\mu}{\gamma}\right)^2\right]} = \frac{1}{\pi\gamma} \left[\frac{\gamma^2}{(x-\mu)^2 + \gamma^2} \right],$$

以及列维分布，其密度函数为：

$$f(x; \mu, \gamma) = \sqrt{\frac{\gamma}{2\pi}} \frac{e^{-\frac{\gamma}{2(x-\mu)}}}{(x-\mu)^{3/2}}$$

我们把 μ 称为位置参数，而把 γ 称为尺度参数。

3.2.3 广义 λ 分布族

Tukey λ 分布：

$$F^{-1}(p) = Q(p; \lambda) = \begin{cases} \frac{1}{\lambda} [p^\lambda - (1-p)^\lambda], & \text{if } \lambda \neq 0 \\ \log(\frac{p}{1-p}), & \text{if } \lambda = 0, \end{cases}$$

Tukey λ 分布可以近似一些常见的分布：

$\lambda = -1$: 接近柯西分布 $C(0, \Pi)$

$\lambda = 0$: logistic分布

$\lambda = 0.14$: 接近正态分布 $N(0, 2.142)$

$\lambda = 1$: 均匀分布 $U(-1, 1)$

利用Tukey分布我们可以定义广义 λ 分布(GLD)：

$$F^{-1}(p) = Q(p; \lambda) = \lambda_1 + \frac{1}{\lambda_2} [p_3^\lambda - (1-p)_4^\lambda]$$

上述分布的VaR和CVaR都是容易计算的。

3.3 金融回报率的尾部风险模型

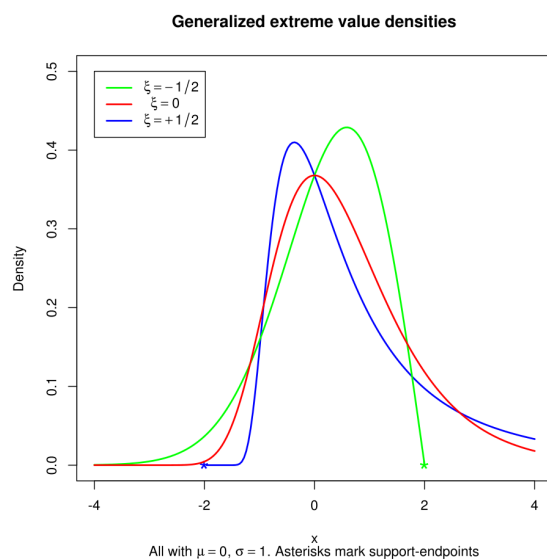
3.3.1 广义极值分布

采用标准化的方法：

$$s = (x - \mu)/\sigma$$

可以得到：标准广义极值分布的密度函数：

$$f(s; \xi) = \begin{cases} (1 + \xi s)^{(-1/\xi)-1} \exp(-(1 + \xi s)^{-1/\xi}) & \xi \neq 0 \\ \exp(-s) \exp(-\exp(-s)) & \xi = 0 \end{cases}$$



3-11 不同参数下的极值分布

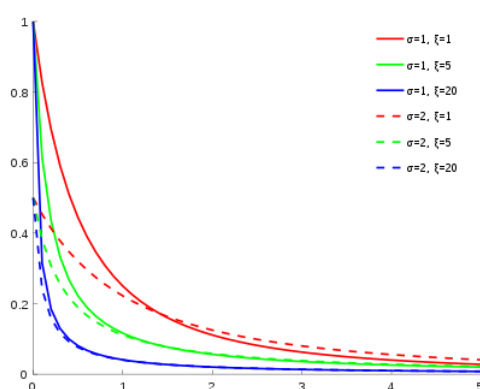
3.3.2 广义帕累托分布

同样的采用标准化的方法：

$$s = (x - \mu) / \sigma$$

可以得到：标准广义帕累托分布的密度函数：

$$f_{(\xi, \mu, \sigma)}(x) = \frac{1}{\sigma} \left(1 + \frac{\xi(x - \mu)}{\sigma} \right)^{(-\frac{1}{\xi} - 1)},$$



3-12 不同参数下的帕累托分布

3.3.3 极值模型

为了拟合GDP模型，我们通常采用极大似然法估计参数（MLE）一旦GDP的参数

估计完成以后，我们就可以计算模型的VaR和CVaR了：

$$(100 - \epsilon)\%VaR = u + \frac{\theta}{\xi} \left(\left(\frac{n}{N_u} \left(1 - \frac{\epsilon}{100} \right) \right)^{-\xi} - 1 \right)$$

其中 u 是门限， n 是观测数量， N_u 是观测值大于门限值的数量。

4 统计学模型

4.1 经典收益模型

最为经典的线性回归模型可以表示为：

$$r_i = \alpha_i + f_1\beta_{i1} + \cdots + f_p\beta_{ip} + \varepsilon_i \quad i = 1, \dots, n,$$

r_i :第 i 份资产收益率

f_k : 影响因子

β :回归系数（灵敏度）

α :常数

ε_i 随机扰动

我们之所以给出上述的模型，是假设收益率可以通过一些可观测的因子线性表出。上述的模型的衍生是多种多样的。例如：上述的模型是静态的，我们可以在不同的时间下观测数据，从而建立模型：

$$r_{i,t+1} = \alpha_i + f_{1t}\beta_{i1} + \cdots + f_{pt}\beta_{ip} + \varepsilon_{it} \quad i = 1, \dots, n,$$

4.2 回归分析

4.2.1 一个简单的例子

假设我们想研究保洁公司股票的收益率和标普500指数的关系，我们可以建立如下的回归模型：

$$r_{P\&G} = \alpha + \beta r_{S\&P500} + \varepsilon$$

其中：

$r_{P\&G}$: 保洁公司股票的收益率。

$r_{P\&G500}$:标普500指数收益率。

我们选取了过去78个月份的数据进行回归分析，得到如下的回归方程：

$$r_{P\&G} = 0.0021 + 0.4617r_{S\&P500}$$

经过P值检验， β 系数显著不为0，经过F检验回归方程显著。为了确保回归模型的有效性，我们还需要对残差进行如下的检验：

- 1.对 ε 进行正态性检验
- 2.对 ε 进行方差齐性检验
- 3.对 ε 进行自相关检验

4.2.2 回归分析在投资中的应用

回归分析在投资中的应用主要在于以下四个方面：

- 1.建立投资策略
- 2.选择投资策略
- 3.选择标的资产
- 4.评估策略表现

4.3 因子分析

一个因子分析模型和回归分析模型十分相似，例如资产回报率模型可以写成：

$$r = \alpha + B \cdot f + \varepsilon$$

其中 α 是N维列向量，表示资产的平均回报率。 f 是K维因子向量， B 是 $N \times K$ 维因子载荷矩阵。

因子分析存在的问题：尽管我们可以计算出具体的因子，但是有时候我们难以对因子给出合理的解释。

4.4 主成分分析

主成分分析（Principal Component Analysis, PCA），是一种统计方法。通过正交变换将一组可能存在相关性的变量转换为一组线性不相关的变量，转换后的这组变量叫主成分。

设是数据的协方差矩阵，PCA的主要步骤：

求解矩阵 W^T 满足如下的条件。

$$\text{Max } W^T \Sigma W$$

$$\text{s.t. } W^T W = 1$$

作为一个例子，我们考虑10支股票（AXP, T, BA, CAT, CVX, CSCO, KO, DD, XOM, GE）构成的投资组合在过去78个月内的收益率。我们计算得到各个主成分如下：

EXHIBIT 4.4 Principal components.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
AXP	-0.3125	-0.3965	-0.0053	-0.1065	0.2329	-0.3591	0.7197	-0.0855	-0.0244	0.1563
T	-0.262	0.2583	-0.5274	-0.5126	0.5049	0.2171	-0.0773	0.0167	0.0072	-0.1233
BA	-0.319	-0.1623	0.3003	0.4218	0.51	0.3489	-0.2128	-0.2478	-0.3287	0.0829
CAT	-0.3592	-0.0749	0.0022	-0.3208	-0.3152	-0.3924	-0.4024	-0.092	-0.5604	0.1562
CVX	-0.316	0.4582	0.2953	-0.0732	0.0018	-0.0931	-0.0679	-0.1833	0.4842	0.5622
CSCO	-0.3206	-0.1258	0.1474	-0.2682	-0.4821	0.6971	0.2597	0.0245	-0.0325	-0.0057
KO	-0.2625	0.2633	-0.5851	0.5335	-0.3049	-0.0072	0.1629	-0.3324	-0.0654	-0.0178
DD	-0.3654	-0.3085	0.081	-0.0091	-0.0691	-0.162	-0.2745	-0.2502	0.5104	-0.5788
XOM	-0.2559	0.5627	0.3717	0.0788	0.0302	-0.1638	0.256	0.2926	-0.2266	-0.4931
GE	-0.3634	-0.1923	-0.1883	0.2768	-0.0081	-0.0166	-0.1704	0.7934	0.1595	0.1854

4-13 主成分

EXHIBIT 4.5 Standard deviations of the 10 principal components.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Standard deviation	2.3003	1.1724	0.9020	0.7931	0.7036	0.6943	0.5649	0.4933	0.4500	0.3865
Proportion of variance	0.5291	0.1374	0.0814	0.0629	0.0495	0.0482	0.0319	0.0243	0.0203	0.0149
Cumulative proportion	0.5291	0.6666	0.7480	0.8109	0.8604	0.9086	0.9405	0.9648	0.9851	1.0000

4-14 各个主成分占比

从中我们可以看到，前三个主成分解释了总体74 %的方差。通常我们选取占比较高的前几个成分作为我们的主成分，一旦我们确定了主成分我们就可以采用如下的公式计算得分了：

$$x_k = \sum_{i=1}^N \beta_{ik} r_i$$

4.5 自回归条件方差模型

传统的回归模型假设残差的方差是相同的，然而这样的假设存在如下的问题：

1.金融资产的回报率的振幅随时间变化，并且在某一时期，巨大的资产振幅往往预示着未来一段时间的振幅也会比较大。这也是我们所说的波动集群效应。

2.高频金融数据往往具有厚尾性，这与传统的数据服从正态分布的假设相矛盾。

基于上述的问题，我们引入了ARCH和GARCH模型。

ARCH(q):

以 ε_t 表示收益或者收益残差，假设 $\varepsilon_t = \sigma_t z_t$ ，此处 $z_t \sim i.i.d N(0, 1)$ （即独立同分布，均符合期望为0，方差为1的正态分布）。在此条件下 σ_t^2 可以写成：

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2,$$

其中 $\alpha_i \geq 0, i > 0$.

GARCH(p,q):

$$y_t = x_t' b + \epsilon_t$$

$$\epsilon_t | \psi_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2 = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2$$

对于ARCH和GARCH模型，我们容易计算他们的VaR和CVaR:

$$VaR_{1-\varepsilon} = V_t(-u_r + q_{1-\varepsilon} \sigma_r)$$

其中 $q_{1-\varepsilon}$ 为标准正态分布 $100(1 - \varepsilon)$ 分位数。同时:

$$CVaR_{1-\varepsilon} = V_t(-u_r + \frac{\phi(q_{1-\varepsilon})}{\varepsilon} \sigma_r)$$

5 模型模拟

5.1 蒙特卡罗模拟

假设你现在有1000美元。你计划用这笔资金投资标普500指数。 C_0, C_1 分别表示期初和期末（一年以后）的资金。 $r_{0,1}$ 表示期间资金的回报率。则有:

$$C_1 = (1 + r_{0,1})C_0$$

在 $[t, t + 1]$ 区间内资金的回报率可以表示成:

$$r_t = \frac{P_{t+1} - P_t + D_t}{P_t}$$

P_t :标普500指数在 t 时刻的价格

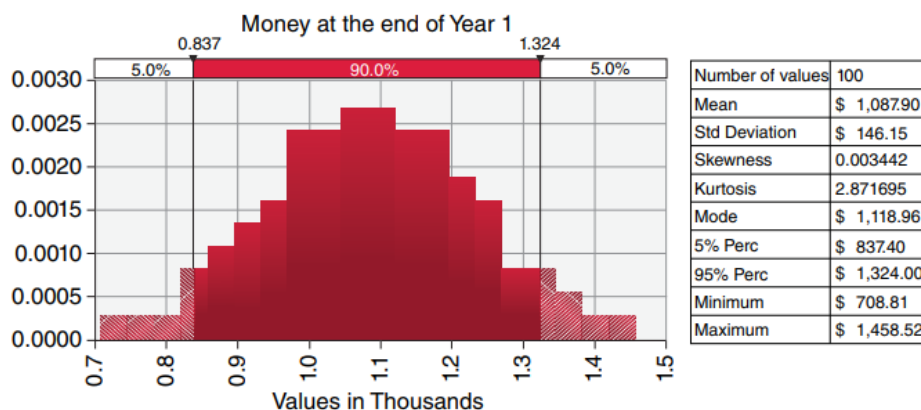
D_t :在区间 $[t, t + 1]$ 内的标普500指数的股息

5.1.1 选择分布函数

为了采用蒙特卡罗模拟我们的模型，我们首先要确定模拟数据的分布函数。这里有两种方法确定数据的分布函数。第一种是采用历史数据，第二种是随机产生一种给定分布的数据序列。例如，我们可以生成一组均值为 μ ，方差为 σ 的正态分布的数据。

5.1.2 理解蒙特卡罗模拟的输出

作为蒙特卡罗模拟的一个例子，我们模拟了标普指数一年以后的收益情况。我们生成了100个服从正态分布的随机数（均值为8.79%,方差为14.65%）图表中的直方图展示了未来一年的收益率分布:



5-15 未来一年标普500收益率分布

根据统计学的知识我们可以知道，我们有95%的置信度认为一年后的收益率分布的区间：

$$\left(\bar{x} - z^* \frac{\sigma}{\sqrt{n}}, \bar{x} + z^* \frac{\sigma}{\sqrt{n}} \right) = (1058.9, 1116.9)$$

5.2 为什么采用蒙特卡罗模拟

上面的例子给出了蒙特卡罗模拟的基本方法，我们采用上述方法，给出了一年以后收益率的分布情况。值得注意的是，如果标普500收益率不满足正态分布，那么我们的模拟就不准确了。下面的例子是一个更加复杂的模拟。

5.2.1 多个输入变量和混合分布

假设你为未来养老而计划做一笔投资，投资的本金为1000美元，投资的周期为30年。假设标普500指数服从均值为 μ ，方差为 σ 的正态分布。记期初的资金为 C_0 ，期末的资金为 C_{30} 。容易知道，期末的收益可以写成：

$$r_{0,t} = (1 + r_{0,1})(1 + r_{1,2}) \dots (1 + r_{t-1,t}) - 1$$

期末的收益取决于这30个正态分布。

5.2.2 合并相关

如果我们投资的是两类标的：国债和股票。这二者的收益率存在着负相关。那么我们又该如何模拟呢？不妨假设股票和国债收益率的相关系数为-0.2。国债收益率服从均值为4%，方差为7%的正态分布。我们采用蒙特卡罗进行了500次模拟，下图展示了30年后收益率的分布情况：

5.2.3 模型评估

如何评估模型的好坏呢？我们考虑下面两种不同的策略：

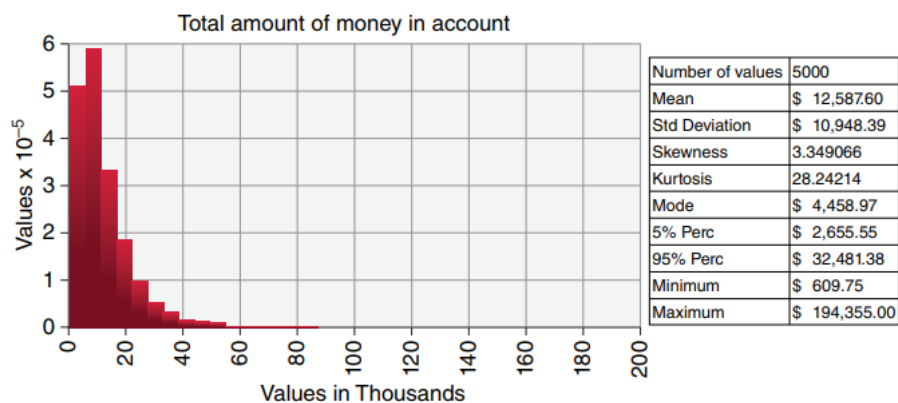
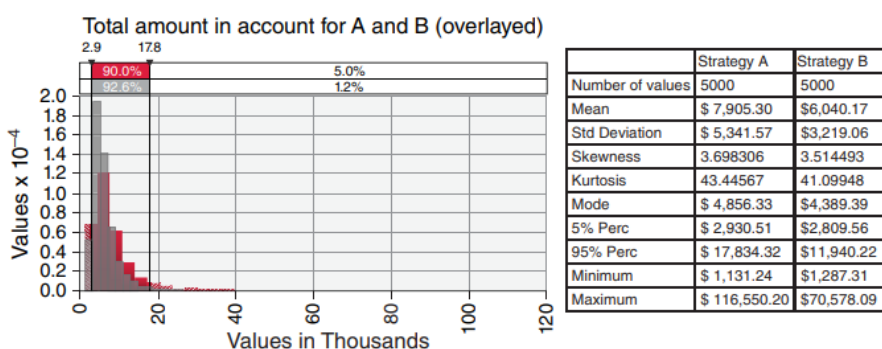


Exhibit 5.3 Output distribution for amount of capital after 30 years.

A: 50%资金投资股票，50%资金投资债券

B: 30%资金投资股票，70%资金投资债券

我们对投资组合A和投资组合B都进行了400次模拟。从模拟的结果我们可以看出，30年后，投资组合A的平均收益大于投资组合B。但是投资组合A的变异系数也大于投资组合B。这说明投资组合A的风险比B大。课本上的图标罗列了二者的均值、方差、峰度、偏度。从中我们可以直观地看到两个投资组合的差异。



5.2.4模拟多少次？

模拟的数据至少要大于等于30才具有统计学意义。在上述的例子中我们模拟了400个样本用于分析。一般来说提高样本的数量有助于我们提高模拟的精度。

5.2.5随机数的生成

Excel、R、MATLAB 都可以很方便地生成随机数，可参考各软件的文档、也可百度。

6 模型优化

本章我们主要讨论的是优化问题——在一系列限制条件下如何使得我们的模型是最优的。

6.1 优化公式

一个优化问题的数学表达式主要由以下三个部分组成：

1. 一系列的决策变量（一般由一个 $N \times 1$ 维的向量组成）
2. 一个目标函数
3. 一系列的约束条件 $(g_i(x), h_j(x))$ 满足： $g_i(x) \leq 0, h_j(x) = 0$

一般来说我们的目标函数总是可以写成：

Maximize: 资产预期回报率

6.1.1 最大化和最小化

通常来说，最优化问题可以写成如下的表达式：

$$\begin{aligned} & \text{minimize } f(x) \\ & \text{subject to } g_i(x) \leq 0 \quad i \in \{1, \dots, I\} \\ & \quad \quad \quad h_j(x) = 0 \quad j \in \{1, \dots, J\}. \end{aligned}$$

同时最大化问题和最小化问题也是可以相互转化的：

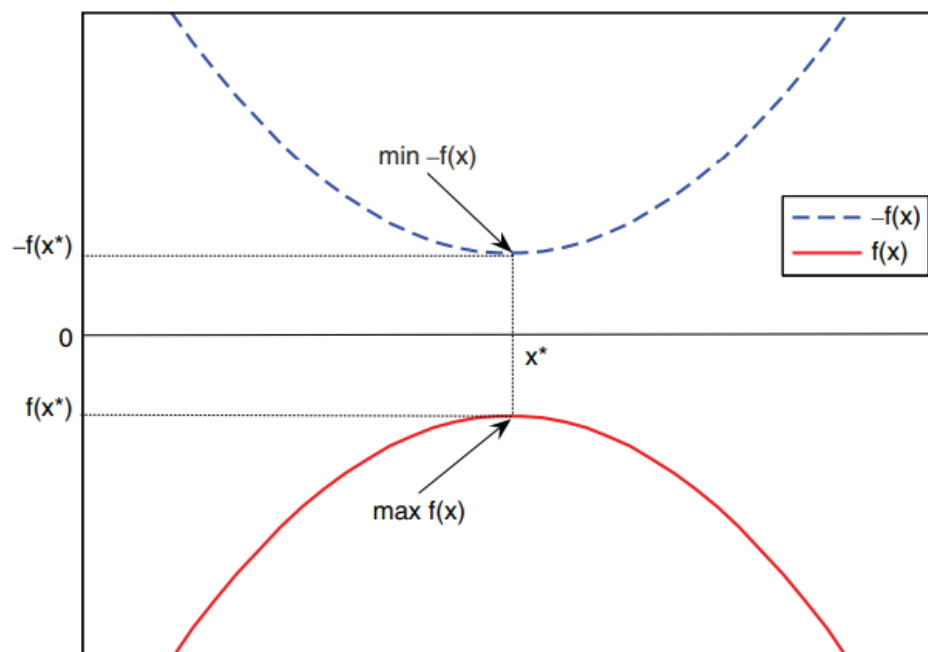
$$\max_x f(x) = -\min_x -f(x)$$

6.1.2 局部最优和全局最优

在求解最优解的时候，部分算法（例如梯度下降法）容易陷入局部最优解而无法得到全局最优解。

6.1.3 多目标优化

多目标规划是数学规划的一个分支。研究多于一个的目标函数在给定区域上的最优化。又称多目标最优化。通常记为MOP(multi-objective programming)。求解多目标线性规划的基本思想是将多目标转化为单目标，常见的方法有理想点法、线性加权法、最大最小法、目标规划法、模糊数学解法等。



6-16 最大化 $f(x)$ 相当于最小化 $-f(x)$

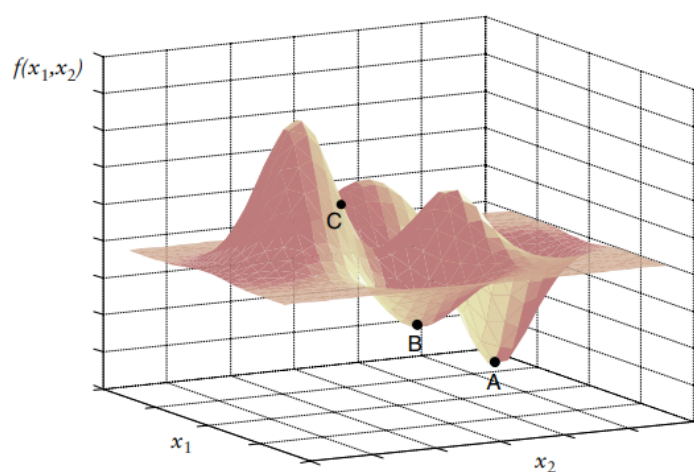


Exhibit 6.2 Global (point A) versus local (point B) minimum for a function of two variables x_1 and x_2 .

6.2 重要的优化问题

6.2.1 凸优化

所谓的凸优化问题是指我们的目标函数和约束条件都是凸函数：

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0 \quad i = 1, \dots, I \\ & Ax \leq b \end{aligned}$$

其中 $f(x)$ 和 $g_i(x)$ 都是凸函数。

6.2.2 线性规划

线性规划（Linear programming, 简称LP）是运筹学中研究较早、发展较快、应用广泛、方法较成熟的一个重要分支，它是辅助人们进行科学管理的一种数学方法。研究线性约束条件下线性目标函数的极值问题的数学理论和方法。线性规划问题的标准形式：

$$\begin{aligned} \min_x \quad & \mathbf{c}^T \mathbf{x} \\ \text{subject to} \quad & A\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

6.2.3 二次规划

二次规划的一般形式可以表示为：

$$\begin{aligned} \min_x \quad & \frac{1}{2} \mathbf{x}^T Q \mathbf{x} + \mathbf{c}^T \mathbf{x} \\ \text{subject to} \quad & A\mathbf{x} \leq \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

其中：

x 是一个 N 维决策变量

Q 是一个 $N \times N$ 维矩阵

c 是一个 N 维向量

A 是一个 $J \times N$ 维矩阵

b 是一个 J 维矩阵

6.2.4 二阶锥规划

一个二阶锥规划问题(SOCP)是指具有如下形式的图规划问题:

$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & \|A_i x + d_i\|_2 \leq c_i^T x + e_i, \quad i = 1, \dots, I \end{aligned}$$

其中:

c 是一个 N 维变量

A 是一个 $J \times N$ 的矩阵

b 是一个 J 维向量

C_i 是一个 $I_i \times N$ 维矩阵

d_i 是一个 I_i 维矩阵

e_i 是标量

6.2.5 整数规划

整数规划是指规划中的变量（全部或部分）限制为整数，若在线性模型中，变量限制为整数，则称为整数线性规划。

整数规划又分为:

- 1、纯整数规划：所有决策变量均要求为整数的整数规划
- 2、混合整数规划：部分决策变量均要求为整数的整数规划
- 3、纯0—1整数规划：所有决策变量均要求为0—1的整数规划
- 4、混合0—1规划：部分决策变量均要求为0—1的整数规划

6.3 一个简单的优化例子:资产分配

现在有一个资产管理者计划投资1千万美元于下面四种基金:

EXHIBIT 6.4 Data for the portfolio manager's problem.

Fund Type	Growth	Index	Bond	Money Market
Fund #	1	2	3	4
Expected return	20.69%	5.87%	10.52%	2.43%
Risk level	4	2	2	1
Max investment	40%	40%	40%	40%

他对自己的投资分配有如下的限定：投资于任何一种基金的比例不超过40%、投资于基金1和基金3的金额之和不超过总投资的60%、平均的投资风险水平不能超过2。

我们假定 $x = (x_1, x_2, x_3, x_4)$ 表示投资于四种基金的金额。那么我们的目标函数可以写成：

$$f(x) = \mu^T x = (20.69\%)x_1 + (5.87\%)x_2 + (10.52\%)x_3 + (2.43\%)x_4$$

我们的约束条件可以写成：

1. 总的投资金额应该等于1千万美元：

$$x_1 + x_2 + x_3 + x_4 = 10,000,000$$

2. 投资于基金1和基金3的金额之和不超过总投资的60%：

$$x_1 + x_3 \leq 6,000,000$$

3. 平均的投资风险水平不能超过2：

$$\frac{4x_1 + 2x_2 + 2x_3 + x_4}{x_1 + x_2 + x_3 + x_4} \leq 2$$

由于 $x_1 + x_2 + x_3 + x_4 = 10,000,000$ ，所以我们化简得到：

$$4x_1 + 2x_2 + 2x_3 + x_4 \leq 20,000,000$$

4. 投资于各个基金的比例不超过40%：

$$x_1 \leq 4,000,000, x_2 \leq 4,000,000, x_3 \leq 4,000,000, x_4 \leq 4,000,000$$

5. 当然，投资金额是非负的：

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0$$

最后我们的问题可以写成如下的形式：

$$\max_{x_1, x_2, x_3, x_4} \begin{bmatrix} 0.2069 & 0.0587 & 0.1052 & 0.0243 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = 10,000,000$$

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 4 & 2 & 2 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \leq \begin{bmatrix} 6,000,000 \\ 20,000,000 \\ 4,000,000 \\ 4,000,000 \\ 4,000,000 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

6.4 优化算法

求解线性规划问题的基本方法是单纯形法。单纯形算法利用多面体的顶点构造一个可能的解，然后沿着多面体的边走到目标函数值更高的另一个顶点，直至到达最优解为止。虽然这个算法在实际上很有效率，在小心处理可能出现的“循环”的情况下，可以保证找到最优解，但它的最坏情况可以很坏：可以构筑一个线性规划问题，单纯形算法需要问题大小的指数倍的运行时间才能将之解出。事实上，有一段时期内人们曾不能确定线性规划问题是NP完全问题还是可以在多项式时间里解出的问题。

第一个在最坏情况具有多项式时间复杂度的线性规划算法在1979年由前苏联数学家Leonid Khachiyan提出。这个算法建基于非线性规划中Naum Shor发明的椭球法（ellipsoid method），该法又是Arkadi Nemirovski（2003年冯诺伊曼运筹学理论奖得主）和D. Yudin的凸集最优化椭球法的一般化。

理论上，“椭球法”在最恶劣的情况下所需要的计算量要比“单形法”增长的缓慢，有希望用之解决超大型线性规划问题。但在实际应用上，Khachiyan的算法令人失望：一般来说，单纯形算法比它更有效率。它的重要性在于鼓励了对内点算法的研究。内点算法是针对单形法的“边界趋近”观念而改采“内部逼近”的路线，相对于只沿着可行域的边沿进行移动的单纯形算法，内点算法能够在可行域内移动。

1984年，贝尔实验室印度裔数学家卡马卡（Narendra Karmarkar）提出了投影尺度法（又名Karmarkar's algorithm）。这是第一个在理论上和实际上都表现良好的算法：它的最坏情况仅为多项式时间，且在实际问题中它比单纯形算法有显著的效率提升。自此之后，很多内点算法被提出来并进行分析。一个常见的内点算法为Mehrotra predictor-corrector method。尽管在理论上对它所知甚少，在实际应用中它却表现出色。

6.5 优化软件

matlab、python、c++、java 理论上都可以用于求解。

6.6 一个求解的例子

6.6.1 Excel求解

这里我采用的是matlab求解。代码如下：

```

1  f=[-0.2069  -0.0587  -0.1052  -0.0243];
2  Aineq=[1 0 1 0;4 2 2 1;1 0 0 0 ;0 1 0 0 ;0 0 1 0 ;0 0 0
      1];
3  bineq=[6 20 4 4 4 4];
4  Aeq=[1,1,1,1];
5  beq=10;
6  lb=[0 0 0 0 ];
7  %% Start with the default options
8  options = optimoptions('linprog');
9  %% Modify options setting
10 options = optimoptions(options,'Display','off');
11 options = optimoptions(options,'Algorithm','interior-
      point');
12 [x,fval,exitflag,output,lambda] = ...
13 linprog(f,Aineq,bineq,Aeq,beq,lb,[],[],options);

```

6.6.2 求解结果

运行上述代码得到求解的结果：

$$x = \begin{bmatrix} 2,000,000 \\ 0 \\ 4,000,000 \\ 4,000,000 \end{bmatrix}$$

$$f_{max} = 931,800$$

所以我们投资于四种基金的金额分别为：2,000,000美元，0美元，4,000,000美元，4,000,000美元。一年后我们的预期收益为931,800美元。

7 非确定优化

7.1 动态规划

所谓的动态规划，是指在 t 时刻的动态系统的状态 x_t 依赖于过去的状态：

$$x_{t+1} = g_t(x_t, \mu_t, \xi_t)$$

其中， μ_t 是一个控制变量， ξ_t 而是一个随机变量。

在金融领域，我们通常认为我们所研究的动态系统是有界的。我们记 $f_t(x_t, \mu_t, \xi_t)$ 是一个动态系统在 t 时刻的代价（收益）。那么从0时刻到 T 时刻，总的代价函数可以写成：

$$\sum_{t=0}^T f_t(x_t, \mu_t, \xi_t)$$

如果我们想要最大化总的收益，相当于我们要选取 μ_t 使得：

$$V_t(x_t) = \max_{\mu_t} \{c(\mu_t) + d \cdot V_{t+1}(\mu_t)\}$$

（注：此处作者没有指明 $c(\mu_t)$ 表示什么意思，有点儿看不懂）

7.2 随机规划

7.2.1 多级模型

课本的写法过于简略并且又没有指明每个变量的意义。这里结合wiki百科给出Multistage Models 更为详细的介绍。

假设在 $t = 0$ 时刻我们有资金 w_0 。我们计划投资于 n 件标的资产。令 $x_0 = (x_{10}, \dots, x_{n0})$ 为初始的资金分配， $\sum_{i=1}^n x_{i0} = W_0$, $x_{i0} \geq 0$ 。假设我们允许在 $t = 1, \dots, T-1$ 时刻调整我们的资产分配。

令 $\xi_t = (\xi_{1t}, \dots, \xi_{nt})$ 表示 t 时刻的资产回报率， $t = 1, \dots, T$ 。显然 ξ_1, \dots, ξ_T 构成了一组随机变量。并且 t 时刻的资产分配 x_t 依赖于过去的信息。或者说， x_t 是 $\xi_{[t]}$ 的函数：

$$x_t = x_t(\xi_{[t]})$$

其中 $\xi_{[t]} = (\xi_1, \dots, \xi_t)$ 是一组随机变量。并且对 $t = 0, \dots, T-1$ 我们有：

$$\sum_{i=1}^n x_{it}(\xi_{[t]}) = W_0,$$

在 $t = 1, \dots, T$ 时刻资产的价值 W_t 为：

$$W_t = \sum_{i=1}^n \xi_{it} x_{i,t-1}(\xi_{[t-1]}),$$

假设我们的目标是最大化我们的总资产：

$$\max E[U(W_T)].$$

显然，在 $t = T - 1$ 时刻，收益率 $\xi_{[T-1]} = (\xi_1, \dots, \xi_{T-1})$ 是已知的。并且资产分配的权重 x_{T-2} 也是已知的。那么，我们研究的问题就可以转化为：

$$\begin{aligned} \max_{x_{T-1}} \quad & E[U(W_T)|\xi_{[T-1]}] \\ \text{subject to} \quad & W_T = \sum_{i=1}^n \xi_{iT} x_{i,T-1} \\ & \sum_{i=1}^n x_{i,T-1} = W_{T-1} \\ & x_{T-1} \geq 0 \end{aligned}$$

我们所优化的目标函数的最优值取决于 W_{T-1} 和 $\xi_{[T-1]}$ 。我们把它记为 $Q_{T-1}(W_{T-1}, \xi_{[T-1]})$ 。同样的，在时刻 $t = T - 2, \dots, 1$ ，我们的优化问题可以写成：

$$\begin{aligned} \max_{x_t} \quad & E[Q_{t+1}(W_{t+1}, \xi_{[t+1]})|\xi_{[t]}] \\ \text{subject to} \quad & W_{t+1} = \sum_{i=1}^n \xi_{i,t+1} x_{i,t} \\ & \sum_{i=1}^n x_{i,t} = W_t \\ & x_t \geq 0 \end{aligned}$$

而在 $t = 0$ 时刻，我们的优化问题成为：

$$\begin{aligned} \max_{x_0} \quad & E[Q_1(W_1, \xi_{[1]})] \\ \text{subject to} \quad & W_1 = \sum_{i=1}^n \xi_{i,1} x_{i0} \\ & \sum_{i=1}^n x_{i0} = W_0 \\ & x_0 \geq 0 \end{aligned}$$

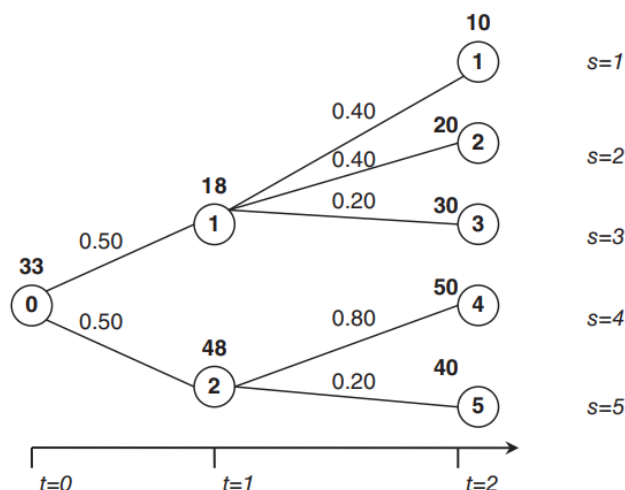
下图展示了一个简单的场景树例子：这里有三个阶段，两个待优化阶段。每一个节点表示资产的价格。例如，在 $t = 0$ 时刻，资产的价格为5是一个确定值。在 $t = 2$ 时刻，资产的价格有5种可能的结果。

给定上述场景树的情况下，我们可以按如下的步骤建立随机优化模型：

(i) 定义变量 $x_t^{(s)}$ ： t 时刻出现 s 场景时资产的价格。在本例中 x_0^0 表示第0阶段资产的价格， x_1^1, x_1^2 表示第二阶段资产的价格， x_2^1, \dots, x_2^5 表示第三阶段资产的价格。

(ii) 写出目标函数：

$$\max \sum_{i=0}^2 \sum_s \pi_i^{(s)} c_i^{(s)} x_i^{(s)}$$



其中 $\pi_i^{(s)}$ 表示 i 时刻出现 s 场景的概率。

(iii) 写出每个阶段的约束条件：

$$f_t(x_t^{(s)}) = 0 \quad t = 0, 1, 2$$

注记：课本中以线性函数的形式给出了每一个阶段的约束条件，但是肯本没有说明 A , B 表示的是什么意思。

那么，我们又是如何生成场景树的呢？通常来说又下面的四种方法：

1. 从历史数据中自足抽样。
2. 使用带参数的概率模型。
3. 生成服从历史数据分布的数据。
4. 采用自回归模型。

7.2.2 风险中性随机模型

假设现在有 N 件标的资产，回报率为 $\tilde{r}_1, \dots, \tilde{r}_N$ 。令 w_1, \dots, w_N 表示我们在 N 件资产中的投资比例。我们希望最大化我们投资组合的预期回报并且极小化投资组合的风险（方差）。假定未来资产回报率的情况有 S 种可能。比如，在情况 S 下，资产的回报率为：

$$\sum_{i=1}^N r_i^s w_i$$

我们用 π_1, \dots, π_N 表示 S 种情形可能出现的概率：

$$\sum_{s=1}^S \pi_s = 1$$

令 μ_1, \dots, μ_N 表示 N 种资产的预期回报率，那么总的期望汇报率为：

$$\sum_{i=1}^N \mu_i w_i$$

我们投资的总方差为：

$$\sum_{s=1}^S \pi_s \left(\sum_{i=1}^N r_i^s w_i - \sum_{i=1}^N \mu_i w_i \right)^2$$

于是，我们可以定义我们的目标函数为：

$$\max_w \sum_{i=1}^N \mu_i w_i - k \left[\sum_{s=1}^S \pi_s \left(\sum_{i=1}^N r_i^s w_i - \sum_{i=1}^N \mu_i w_i \right)^2 \right]$$

在这里 k 为惩罚系数。通常 k 由投资者自己给定。一般来说，投资者的风险厌恶程度越小， k 的取值越小。

我们不妨讨论一个具体的例子：

下表展示了四种基金未来收益率可能出现的两种情况。我们假定情景1出现的概率为0.7，情景2出现的概率为0.3。

EXHIBIT 7.2 Scenarios for the returns of the four funds.

Fund #	1	2	3	4
Scenario 1	50.39%	15.69%	23.29%	2.50%
Scenario 2	-9.02%	-3.96%	-2.25%	2.36%
Mean return	20.69%	5.87%	10.52%	2.43%

那么我们的目标函数就可以写成：

$$\begin{aligned} & \max_w 20.69w_1 + 5.87w_2 + 10.52w_3 + 2.43w_4 \\ & -0.3k(50.39w_1 + 15.69w_2 + 23.29w_3 + 2.5w_4 - (20.69w_1 + 5.87w_2 + 10.52w_3 + 2.43w_4))^2 \\ & -0.7k(-9.02w_1 - 3.96w_2 - 2.25w_3 + 2.36w_4 - (20.69w_1 + 5.87w_2 + 10.52w_3 + 2.43w_4))^2 \end{aligned}$$

7.2.3 机会约束模型

如果原始模型的约束条件为：

$$a^T x \leq b$$

那么机会约束模型的约束条件就可以写成：

$$\Pr(a^T x) \leq \varepsilon$$

显然，一个概率约束模型不是一种标准的优化公式，我们需要将上述的条件进一步写成可以计算的问题。我们假定我们的资产有 S 种不同汇报的情景。我们可以把约束条件写成：

$$\begin{aligned} a^s x &\leq b + M y_s \quad s = 1, \dots, S \\ \sum_{s=1}^S y_s &\leq [\varepsilon S] \\ y_s &\in \{0, 1\} \quad s = 1, \dots, S \end{aligned}$$

其中 M 是由投资者给出的常数（通常很大）。

当 $y_s = 0$ 时，我们需要满足如下的约束条件：

$$a^s x \leq b + M y_s \quad s = 1, \dots, S$$

当 $y_s = 1$ 时，由于 $M y_s$ 很大，我们的问题成为一个无约束问题。

而附加条件：

$$\sum_{s=1}^S y_s \leq [\varepsilon S]$$

约定了1最多出现的次数。

例如，我们预期的资产汇报有115种可能，并且我们的置信度为 $\varepsilon = 5\%$ ，那么我们允许出现最多 $[0.05 * 115] = 5$ 个1。也就是说，我们在95%的情况下是不违背约束条件的。

如果 x 服从正态分布，那么我们的约束条件可以写成：

$$a^s x + q_{1-\varepsilon} \sqrt{x^T \Sigma x} \leq b$$

这是一类二阶锥规划问题，那么我们该如何求解呢？鲁棒优化给出了很好的答案。

7.3 鲁棒优化

在随机规划中我们常常难以确定数据所服从的概率分布。鲁棒优化引入一个不确定集来帮助我们解决上述问题。鲁棒优化不同于线性规划，鲁棒优化的系数是一个不确定的集合，其目的是在最糟糕的条件下为目标函数找到一个最优解。

例如，考虑线性约束问题：

$$\hat{a}^T x \leq b$$

其中随机变量 $\hat{a} = (\hat{a}_1, \dots, \hat{a}_N)$ ，其95%的置信区间为： $(a_i - \delta_i, a_i + \delta_i), i = 1, \dots, N$ 。

记：

$$U_\delta(\hat{a}) = \{a \mid |a - \hat{a}_i| \leq \delta_i, i = 1, \dots, N\}$$

鲁棒优化的目标函数可以写成：

$$\max_{a \in U_\delta(\hat{a})} a^T x \leq b$$

可以将上述式子写成一个易于求解的公式：

$$a^T x + \delta^T |x| \leq b$$

在求解实际问题中，我们常常不需要将所有的系数都取到最糟糕的情况，由于我们容易获得系数的标准差和期望。我们常常采用经典的椭圆不确定集合：

$$U_\delta(\hat{a}) = \{a | (a - \hat{a})^T \Sigma^{-1} (a - \hat{a}) \leq \delta^2\}$$

等价地，我们可以写成：

$$U_\delta(\hat{a}) = \{a | \|\Sigma^{-\frac{1}{2}}(a - \hat{a})\| \leq \delta\}$$

于是，鲁棒优化问题：

$$\hat{a}^T x \leq b$$

可以被化成：

$$\hat{a}^T x + k \|\Sigma^{\frac{1}{2}} x\| \leq b$$

这就转化成了一个可以求解的规划类问题了。

8 资产多样化

分散风险(Diversification) 在证券投资上，是指将资金分配在多种资产上，而这些资产的回报率相互之间的关联性比较低，以达分散风险的目的。这样做既可以降低风险，又不会损及收益。推出的各种银行、理财和基金系统的QDII产品为中国投资者提供了投资全球、分散风险和资产配置의更多选择。

8.1 资产分散的例子

假设有一个投资者计划投资两只股票（股票1和股票2），两种股票的回报率为：

$$E(r_1) = \mu_1 = 9.1\%$$

$$E(r_2) = \mu_2 = 12.1\%$$

两种股票的标准差：

$$\sigma_1 = 16.5\%$$

$$\sigma_2 = 15.8\%$$

很明显，第二只股票的表现要优于第一只股票，因为第二只股票的收益率比较高并且方差比较小。

现在我们假定两只股票的相关系数 $\rho_{12} = -0.22$, 设 w_1, w_2 为投资与两只股票的权重。我们总的投资回报率为：

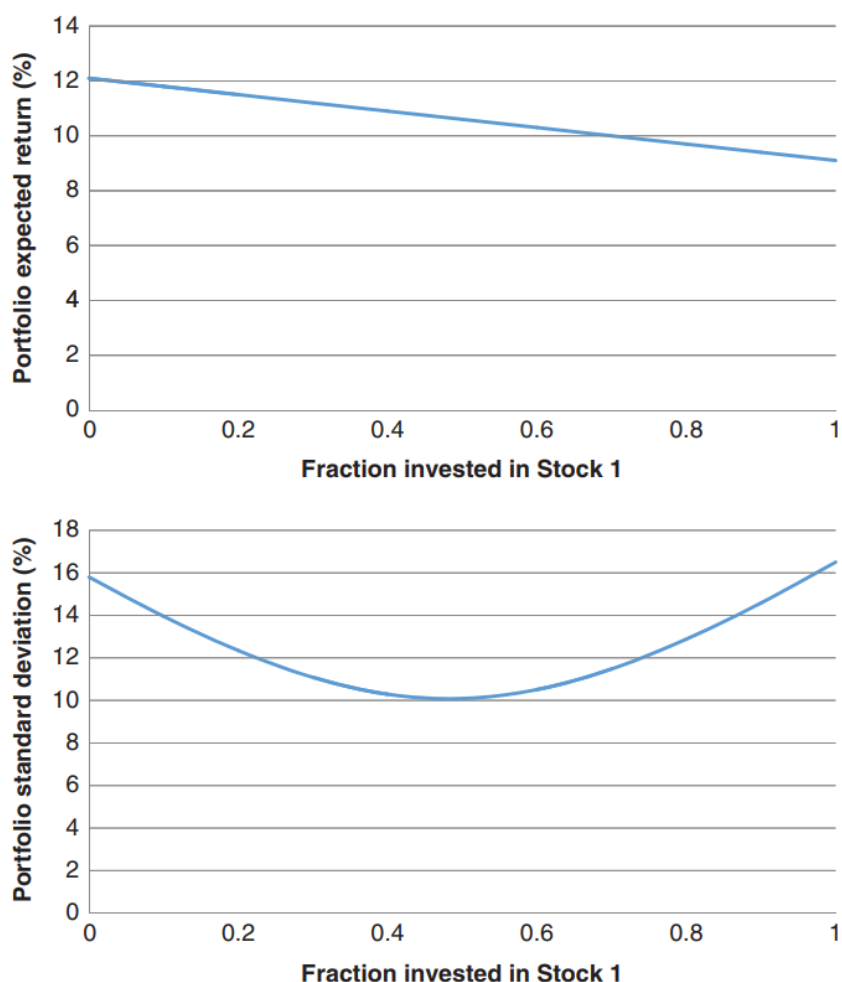
$$r_p = w_1 r_1 + w_2 r_2$$

容易得到：

$$E(r_p) = w_1 \mu_1 + w_2 \mu_2$$

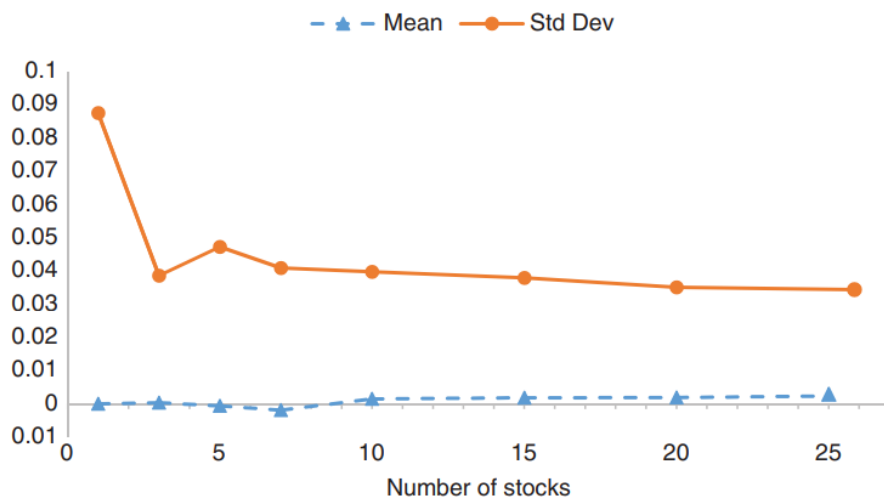
$$\sigma_p^2 = Var(w_1 r_1 + w_2 r_2) = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1 w_2 \sigma_1 \sigma_2 \rho_{12}$$

显然，我们的投资组合的均值和方差是随着 w_1, w_2 而变化的，我们可以从下图看到这一现象。



在我们的这个例子中，具有负相关的两只股票显著地降低了投资的风险。然而，当两只股票不相关或者弱相关的时候这一效应就没有那么显著了。我们选取了标普500中的25只股票计算了这一个投资组合在过去12个月的表现（先选1只，依次增

加) 可以看到, 随着股票数目的增加, 投资的方差在减少, 但是其减少的速度是递减的。



8.2 经典均值方差优化模型

假设一个投资者计划投资于 N 个标的资产。投资的权重 $w = (w_1, \dots, w_N)$ 。并且 $\sum_{i=1}^N w_i = 1$ 。那么预期的投资回报:

$$E(r_p) = \sum_{i=1}^N w_i E(r_i) = \mu^T w$$

方差:

$$\sigma_p^2 = w^T \Sigma w$$

其中 Σ 是协方差矩阵。

如果 N 只股票的相关系数矩阵为:

$$C = \begin{bmatrix} 1 & \dots & \rho_{1N} \\ \vdots & & \vdots \\ \rho_{N1} & \dots & 1 \end{bmatrix}$$

那么协方差可以写成:

$$\sigma_p^2 = (w^s)^T C w^s$$

其中 $w^s = (w_1 \sigma_1, \dots, w_N \sigma_N)$

对于 $N = 2$ 的情况:

$$E(r_p) = w_1 \mu_1 + w_2 \mu_2$$

$$\sigma_p^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1 w_2 \sigma_1 \sigma_2 \rho_{12}$$

经典的均值-方差优化问题可以写成：

$$\begin{aligned} \min_w \quad & w^T \Sigma w \\ \text{s.t.} \quad & w^T \mu = r_{target} \\ & w^T \mathbf{1} = 1 \end{aligned}$$

其中 $\mathbf{1} = (1, \dots, 1)_{1 \times N}$ 这是一个凸二次规划问题。构造拉格朗日乘法：

$$L = w^T \Sigma w + 2\lambda(\mu - w^T r_{target}) + 2\eta(1 - w^T \mathbf{1}),$$

令 $\frac{\partial L}{\partial w} = 0, \frac{\partial L}{\partial \lambda} = 0, \frac{\partial L}{\partial \eta} = 0$ 解得：

$$w^{opt} = \frac{1}{\Delta} [(\mu^T \Sigma^{-1} \mu) \Sigma^{-1} \mathbf{1} - (\mathbf{1}^T \Sigma^{-1} \mu) \Sigma^{-1} \mu] + \frac{r_{target}}{\Delta} [(\mathbf{1}^T \Sigma^{-1} \mathbf{1}) \Sigma^{-1} \mu - (\mu^T \Sigma^{-1} \mathbf{1}) \Sigma^{-1} \mathbf{1}]$$

其中：

$$\Delta = (\mu^T \Sigma^{-1} \mu)(\mathbf{1}^T \Sigma^{-1} \mathbf{1}) - (\mu^T \Sigma^{-1} \mathbf{1})^2 > 0.$$

8.3 有效边界理论

有效边界是在收益—风险约束条件下能够以最小的风险取得最大的收益的各种证券的集合。如图所示。

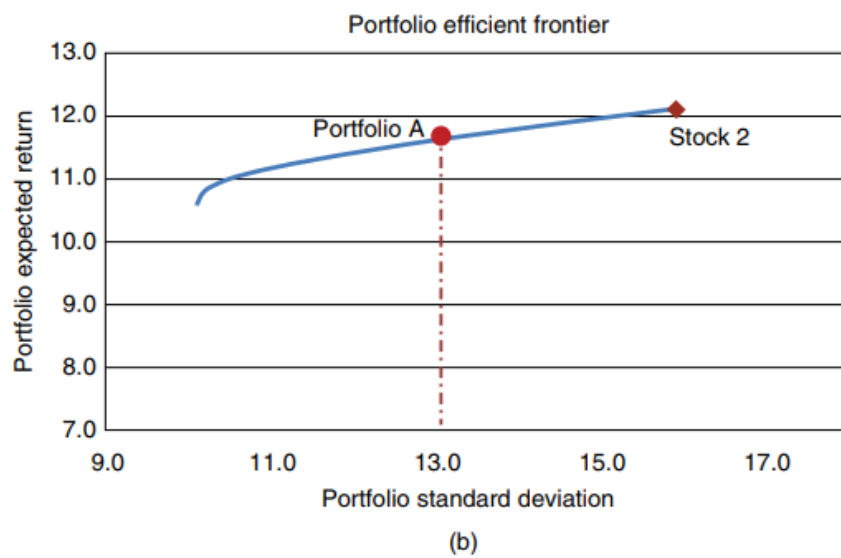
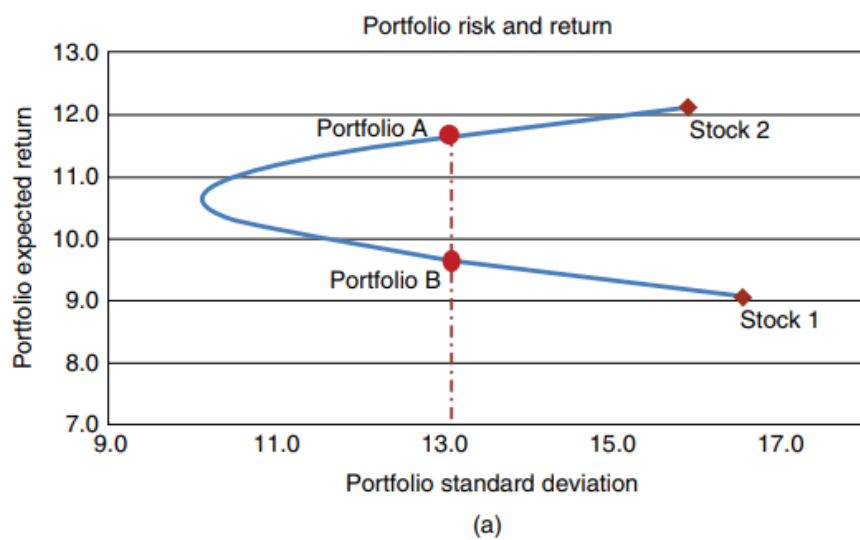
图中，横坐标表示风险，纵坐标代表收益，阴影中的任意一点代表一种可行的组合证券，每一组合证券所提供的风险—收益组合可以通过横、纵坐标上相应的两点予以确定。很明显，图中曲线上的各点代表的各种组合证券所对应的风险—收益组合远较阴影部分中其他各点代表的组合证券为优，因为它们能够在风险最小的情况下取得最大的收益；或者说，为取得一定收益而承受的风险最小，承受一定风险所获得的收益最大。曲线代表着证券投资的有效边界

8.4 经典均值方差优化模型

8.4.1 预期回报公式

如果我们想要最大化投资组合的回报率，我们可以把优化问题写成：

$$\begin{aligned} \min_w \quad & w^T \mu \\ \text{s.t.} \quad & w^T \Sigma w = \sigma_{target}^2 \\ & w^T \mathbf{1} = 1 \end{aligned}$$



这是经典的均值方差优化模型，它将我们的风险限制在一定的水平下。我们会在第10章详细讨论。

8.4.2 风险厌恶公式

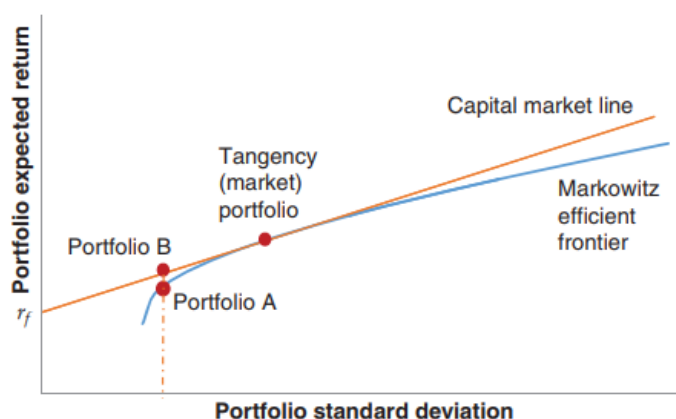
对于一个风险厌恶的投资者，他更加希望其投资组合不应该承受太多的风险。我们常常采用惩罚函数的方式来实现这一目标。

$$\begin{aligned} \min_w \quad & w^T \mu - \lambda w^T \Sigma w \\ \text{s.t.} \quad & w^T \mathbf{1} = 1 \end{aligned}$$

其中的 λ 被称为风险厌恶系数。

8.5 资本市场线

资本市场线(Capital Market Line, 简称CML)是指表明有效组合的期望收益率和标准差之间的一种简单的线性关系的一条射线。它是沿着投资组合的有效边界，由风险资产和无风险资产构成的投资组合。



设 r_f 表示无风险收益率，令 $w = (w_1, \dots, w_N)$ 表示投资于 N 件风险资产的权重， $r = (r_1, \dots, r_N)$ 表示投资预期收益率。容易知道我们总的资产回报率为：

$$w^T r + (1 - w^T) r_f$$

方差为：

$$w^T \Sigma w$$

均值方差优化模型可以写成：

$$\begin{aligned} \min_w \quad & w^T \Sigma w \\ \text{s.t.} \quad & w^T r + (1 - w^T) r_f = r_{\text{target}} \end{aligned}$$

可以得到最优解为:

$$w = C\Sigma^{-1}(r - r_f\mathbf{1})$$

其中:

$$C = \frac{r_{target} - r_f}{(r - r_f\mathbf{1})^T \Sigma^{-1} (r - r_f\mathbf{1})}$$

如果我们的目标是最大化夏普比率,那么我们的目标函数可以写成:

$$\begin{aligned} \max_w \quad & \frac{w^T r - r_f}{w^T \Sigma w} \\ \text{s.t.} \quad & w^T \mathbf{1} = 1 \end{aligned}$$

上述问题的解为:

$$w^M = \frac{1}{\mathbf{1}^T \Sigma (r - r_f\mathbf{1})} \Sigma^{-1} (r - r_f\mathbf{1})$$

8.6 期望效用函数理论

如果某个随机变量 W 以概率 p_i 取值 $w_i, i = 1, 2, \dots, n$ 而某人在确定地得到 w_i 时的效用为 $u(w_i)$ 那么, 该随机变量给他的效用便是: $E[u(W)]$ 。对于投资类问题, 设 u 是投资者的效用函数, W_0 表示期初的资产, W 表示期末的资产, 投资者的目标是最大化 $E[u(W)]$, 那么投资者的效用优化问题可以写成:

$$\begin{aligned} \max_w \quad & E(u(W_0(1 + w^T r))) \\ \text{s.t.} \quad & w^T \mathbf{1} = 1 \end{aligned}$$

下面给出一些常见的效用函数。

8.6.1 二次效用函数

二次效用函数的形式为:

$$u(x) = x - \frac{b^2}{2}x^2, b > 0$$

容易计算得到:

$$E(u(W_0(1 + w^T r))) = u(W_0) + W_0\mu_P(1 - bW_0) - \frac{b}{2}W_0^2(\sigma_P^2 - \mu_P^2)$$

其中 μ_P, σ_P 表示收益的均值和方差。

8.6.2 线性效用函数

$$u(x) = a + bx$$

$$E(u(W_0(1 + w^T r))) = a + bE(W)$$

8.6.3指数效用函数

$$u(x) = -\frac{1}{a}e^{-ax}$$

虽然指数效用函数是负值函数，但是这没有关系，因为他是严格递增的函数。

8.6.4幂效用函数

$$u(x) = ax^a, 0 < a \leq 1$$

8.6.5logistic效用函数

$$u(x) = \ln(x)$$

8.6.6再定义多样化

马科维茨的资产多样化理论在实际的场景中很难被使用，主要的原因是预期回报率通常是难以确定的。Choueifaty在2008年提出了如下的指标用于衡量资产的多样化比例。

资产多样化比例(DR)

$$DR(w) = \frac{w^T \sigma}{\sqrt{w^T \Sigma w}}$$

其中 σ 是投资标的的标准差。当 $DR(w) = 1$ 时表示投资于单一品种。多样化的投资应该使得 $DR(w) > 1$ 。例如，等权重的投资组合 $DR(w) = \sqrt{N}$ 。第 i 件投资标的的风险占总投资风险的比例可以采用如下的计算公式：

$$RC_i(w) = \frac{\partial \sigma(w)}{\partial w} = \frac{w_i(\Sigma w)_i}{\sigma(w)}$$

其中 $(\sigma w)_i$ 是 σw 的第 i 个元素。这样的定义保证了总的投资风险为 $\sigma(w)$ 。一个风险平价投资组合相当于一个等权重的投资组合：

$$RC_i(w) = \frac{\sigma(w)}{N}$$

Maillard在2010年考虑了如下的优化问题：

$$\min_{w>0} f(w) = w^T \Sigma w - \sum_{i=1}^N \ln(w_i)$$

这是一个优化领域的著名问题，采用内点法可以求解。近年来，多种多样的资产优化问题被提出，读者可以参考相关文献。

9 因子模型

9.1 金融经济学文献中的因子模型

线性因子模型代表了应用于金融中的绝大多数因子模型：

$$r_i = \alpha_i + f_1\beta_{i1} + \cdots + f_p\beta_{ip} + \varepsilon_i \quad i = 1, \dots, n,$$

r_i :第*i*份资产收益率

f_k : 影响因子

β_{ik} :回归系数（灵敏度）

α :常数

ε_i 随机扰动

最简单的线性因子模型—单一指数模型,是Markowitz(1959)为简化均值一方差投资组合分析的目标而提出的。虽然马科维茨在其著作的脚注中提到了这个方法,但Shape(1963)进一步研究了这一方法。这个模型包含一个称为指数的单一因子。理论上,该指数是经济中所有资产的价值加权投资组合,其中“价值加权”意味着投资组合中每种资产回报的权重代表了该资产的市值(发行股数乘以每股价格)在经济中所有资产的总市值的份额。在实践中,几个宽基指数(如标普500指数)都得到使用。假定因子是市场指数(类如标普500指数)回报的模型称为单一指数市场模型。

在单一指数市场模型中,相对于市场变化的证券的敏感度(证券的贝塔值)是利用简单线性回归从一组观察到的证券回报和市场回报中估计得到的:

$$r_i = \alpha_i + \beta_i + r_M + \varepsilon_i$$

r_i :第*i*份资产收益率

r_M : 市场组合的回报

β_{ik} :回归系数（灵敏度）

α_i :证券*i*特有的回报

ε_i 随机扰动

单一指数模型有三个假设:

- (1) 模型中的所有随机变量都有有限的均值和方差
- (2) 市场指数的随机回报与证券的随机误差项之间的协方差为零。
- (3) 任何两个证券*i, j*回报的特有部分的协方差是零

我们可以得到,对 $\forall i \neq j$:

$$Cov(r_i, r_j) = Cov(\beta_i r_M, \beta_j r_M) = \beta_i \beta_j \sigma_M^2$$

上面的式子表明如果我们的假设是正确的,那么投资组合的风险仅仅由贝塔系数决定!

本章的后续讨论了Capm模型,但是笔者认为假设条件过于强,所以没有特定写出。

9.2 因子模型的均值方差优化

考虑一个具有 N 个资产的投资组合,其期望回报向量 $\mu = (\mu_1, \dots, \mu_N)$, 回报协方差矩阵为 Σ 。如果资产的权重是 $W = (w_1, w_2, \dots, w_N)$ 那么回忆第8章,投资组合期望回报和方差可以分别写成 $\mu^T w$ 和 $w^T \Sigma w$ 。

为了在不使用因子模型的情况下,计算出必要的输入信息以估计一个具有 N 个证券的投资组合的风险(定义为投资组合方差),那么人们需要发现 N 个证券的协方差矩阵。这将需要估计 N 个方差(N 个证券中的每一个证券都有一个方差)和 $N(N-1)/2$ 个协方差。人们还需要估计 N 个期望回报。总计需要 $2N + N(N-1)/2$ 个估计。

假设投资组合经理正在考虑构建一个投资组合,并且投资组合有1000个候选证券(即 $N = 1000$)。为了计算投资组合,投资组合经理需要估计1000个方差和499 500协方差。加上1000个期望回报的估计,总共需要对501500个值进行估计,这是一个非常艰难的任务。除了准备用于估计的数据是困难的之外,估计也将是不稳定的。在时间上,一个证券对特定因子的敏感性可能比该证券对投资组合中其他证券的敏感性更加稳定。如果使用基于证券之间直接关系的协方差估计,那么来自投资组合优化流程的输出值将是不可靠且不稳定的。输入参数的一个微小变化可能导致最优投资组合配置的大幅波动。

现在假设投资组合经理希望从 N 个证券中构建一个投资组合,并且 N 个证券回报的向量可以通过因子模型表示:

$$r = \alpha^T w + Bf + \varepsilon$$

r : 预期收益率

f : K 维因子向量

B : $N \times K$ 维因子载荷矩阵

ε : N 维残差

投资组合收益率的期望:

$$E(r) = \alpha^T w + BE(f)$$

投资组的方差：

$$w^T (B \Sigma_f B^T + \Sigma_\varepsilon) w$$

其中 Σ_f 是因子的协方差矩阵。 Σ_ε 是对角残差矩阵。

假设投资组合使用单一指数模型来估计输入信息。那么，一个含有N个证券的投资组将需要估计以下参数以便进行均值一方差优化：

N个阿尔法来计算期望回报。

N个贝塔来计算期望方差和协方差。

1个指数方差 σ_M^2 的估计来计算方差和协方差。

N个残差的方差

总共是3N+1个估计。对于一个具有1000个候选证券的投资组合，这转换成3001个估计，比不使用因子模型的501500个估计更加易于操作。假设投资组合经理使用具有K因子的模型，逻辑是类似的，只不过需要估计：

N个矩阵中每一个证券的K个因子载荷

K(K-1)/2个因子协方差

K个因子方差

N个残差的方差

9.3 实践中的因子选择

一般地，因子模型中使用的经验因子分为以下三类：

(1) 外部宏观经济因子，如国内生产总值(GDP)、消费者价格指数(CPI)、失业率、债券信用息差，以及收益率曲线的陡峭度。

(2) 基本面因子（公司特征），如市盈率，股息支付率盈利增长预测和财务杠杆

(3) 提取的统计因子，例如市场投资组合的回报率特定行业(公用事业、交通、航空航天)股票回报率的平均值等。

下图展示了将资产类型分解成为因子的案例

9.4 阿尔法构建中的因子模型

几十年来，构建阿尔法一直是主动量化投资管理的核心.主要思路是识别带来出色绩效的回报（因子)的经验动因。



Exhibit 9.1 An example of breaking down asset classes into factors.
Source: Hogan, Hodges, Potts, and Ransenberg (2015). Reprinted with permission.

在实践中，构建阿尔法的过程开始于观察到的现象或信号，例如“小盘股优于大盘股”“低波动股票优于高波动股票”“每股收益高的公司优于每股收益低的公司”。在每一个这些观察到的例子背后是一个因子(规模、波动性、每股收益)，投资组合经理试图用因子模拟投资组合来描绘这个因子。

我们来抽象化这一问题，令 B 为因子载荷矩阵， α 是包含阿尔法信号的向量，那么优化问题可以写成：

$$\begin{aligned} \min_w w^T \Sigma w \\ s.t. B^T w = 0 \\ \alpha^T w = 1 \end{aligned}$$

9.5 因子模型用于风险评估

9.5.1 宏观经济因子模型

逻辑上，资产回报率应该收到宏观经济的影响。例如，如果美联储削减利率，债券和股票市场都会做出反应。原因是当利率降低时，投资者需要较低的投资回报率。较低的必要回报率一般带来证券价值的升值。

宏观经济模型中的每个因子的每个贝塔系数通常根据时间序列回归方程来估计：

$$r_{it} = \beta_{i1}f_{1t} + \cdots + \beta_{iK}f_{Kt} + \varepsilon$$

9.5.2 基本面因子模型

基本面因子模型依赖于基本面研究——梳理行业和个体公司数据以找到具有市场价值的公司特征。这样的公司特征可能与公司估值或投资者期望有关，并且通常称为风格因子。包括公司规模、财务比率(如市盈率)、股息率、动量、增长、财务杠杆、收益变动和波动性。行业因子(例如，证券的发行人属于哪个行业类别)是组成基本面因子模型部分的另一组因子。通常，出于建模目的，为了处理多重共线性效应并提高模型的解释力，常常使用风险指数。这些风险指数是基本面因子的汇集，而不是单一财务比率。

基本面因子模型中的每个因子的每个贝塔系数通常根据时间序列回归方程来估计：

$$r_{it} = \beta_{i1}f_{1t} + \cdots + \beta_{iK}f_{Kt} + \varepsilon$$

9.5.3统计因子模型

在数学上，证券收益率的协方差矩阵的主成分通过分解协方差矩阵 Σ 来提取：

$$\Sigma = PAP^T$$

其中 P 是正交矩阵。

主成分分析的结果可以被认为是不相关的 N 个投资组合，此外，主成分分析可以帮助降低维度。如果最终选取了 K 个因子，那么模型解释的方差占比：

$$\frac{\sum_{k=1}^K \lambda_k}{\sum_{k=1}^N \lambda_k}$$

一个确定隐含的统计因子的替代技术是因子分析。当基本因子模型的理论基础存在时，因子分析比主成分分析更为合适。因子分析通常使用回归来估计因子载荷矩阵，因此证券回报可以表示为：

$$r = \alpha^T w + Bf + \varepsilon$$

其中 α 是回报的 N 维向量， f 是因子的 K 维向量， B 是因子载荷的 $K \times N$ 矩阵， ε 是残差的 N 维向量。需要估计的协方差矩阵可以假设为：

$$\Sigma = BB^T + \Sigma_\varepsilon$$

估计过程分为两个步骤。第一步中，通过线性回归或最大似然估计法来估计协方差矩阵和因子载荷。第二步中，使用协方差矩阵和因子载荷来估计因子。在每个时间点 t 的因子估计为：

$$f_t = (B^T \Sigma_\varepsilon^{-1} B)^{-1} B^T \Sigma_\varepsilon^{-1} (r_t - \alpha)$$

9.5.4混合因子模型

如今几个因子模型供应商都提供混合因子模型。这些模型使用传统的因子组(货币、行业、国家等)来尽可能多地解释回报的不定性，然后使用统计因子(通过主成分分析获得)吸收因子模型中的剩余风险，即因子模型误差项中剩余的不定性。

一方面，使用基本面宏观经济因子模型为风险和投资组合管理提供了背景，另一方面使用统计因子模型使得捕获残差回报中的任何剩余隐藏因子得以可能。因此，混合因

子模型试图做到两者最好:它对模型强加了一些结构设置,但也从数据中挖掘那些在理论模型中无法得到的附加信息。

9.5.5 选择“适合的”因子模型

如前所述,许多资产管理公司使用软件供应商提供的因子模型。“基本面”因子模型仅使用公司的基本面信息。“全球”因子模型纳入了关于世界货币和利差的信息。“混合”因子模型将基本面和/或宏观经济因子模型与统计因子模型相结合。

那么哪个因子模型是“最好的”呢?这没有单一的答案,但投资组合经理的投资策略细节可以缩小选择。

首先,投资经理应选择一个能涵盖其投资领域的因子模型:单一国家、地区或全球。这是因为全球因子模型考虑了区域投资时不那么重要的货币和其他因子。

第二,投资经理应该考虑投资组合的投资时限。供应商因子模型通常具有目标时限——短期、中期或长期,其中投资时限是投资组合的平均周转时间。不同的因子在短期内和长期内是不同的。短期因子模型更适食期限为一天至两个月的投资者,例如对冲基金。长期时限是指长于一年的投资期,中期时限是指介于短期和长期之间的任何时限。

9.6 数据管理与质量问题

9.6.1 数据调整

数据通常存储在多个数据库中需要合并才可以进行分析。合并来自不同数据库的数据可能是非常具有挑战性的。例如,有时不同数据库为同一公司使用不同的标识符,使得难以对应记录。即使使用诸如CUS1P或股票代码之类的公共标识符,后者也可能随时间而改变,并且有时会因为公司不再存在于数据库中而被重新使用,因此可能难以跨数据库正确地连接所有公司。

数据调整中的一个更微妙的问题是,因为数据的差异,但基于不同变量的数据记录,用不同的方法来计算模型输入信息,可能会导致不一致的估计。Faboi、Focardi和Kolm展示了两个替代但等价的方法来计算一个基本面因子(EBITDA/EV),数据来源是Capital IQ Compustat Point-In-Time数据库中的记录两种方法得出非常不同的EBITDA/EV估计。这两种方法的不同在于用于计算分母(EBITDA)的数据不同:

方法一:EBITDA=销货-销货成本-销货和一般管理费用

方法二:EBITDA=未计折旧前的经营收入

二者的关系:

未计折旧前的经营=销货-销货成本-销货和一般管理费用

然而，从数据库中提取数据不是一祥的Fabozzi列出了这种数据不一致的多种原因。首先，可能存在四舍五入和轻微误差。第二，记录中可能有错误。第三，有时不同的数据项被结合在一起。有时旧和摊销费用不是损益表上的单独项目；而是包括在销售商品的成本中。第四，数据项可能在不同的公司、部门或行业中报告不一致当财务数据提供者不正确地将财务指标从公司报告映射到特定数据库项目时，也会发生这种情况。

9.6.2幸存者偏差

如果当公司不再存在而从数据库中删除，并且只有还运行的公司的数据被用于分析时，幸存者偏差(survival bias)便会发生。（公司可能因不同的原因而不再存在，例如破产或合并。）幸存者偏差导致错误的分析。只关注幸存者和忽略那些不再存在的公司，都可能会在分析中忽略重要的信息。通过下面的例子可以理解这个问题。假设一个人将在比较2005年3 000家最大(最高市值)的公司与2015，3 000家最大的公司后发现，3 000家最大公司的价值上升了40%。有人可能会认为，包括这些公司在内的投资组合的回报率是40%。然而，如果一些在2005年存在的公司不再存在(或不在2015年3000家最大公司的名单上)，那么实际上是无法构建包含最大公司的投资组合的！！

9.6.3前视偏差

在预测模型中使用在预测发生时还没有的数据，就会发生前视偏差。例如，假设将年终收益用作1月份回报预测模型中的一个因子这些间报在年度结束后至少一个月后才会公布。因此，它们不能用于预测1月份的收益。它们不可能在1月初获得。如果使用回填和重述，前视偏差可能变得更糟。回填(backfilling)是将先前丢失的数据在收到后输入数据库，这可能发生在比必须使用该信息以用于预测的时间晚很多。重述(restatement)发生于(例如)当公司修改了其最初发布的财报时。许多数据库公司(例如，Capital IQ)会覆盖最初记录的数字。然而，如果将信息用于预测模型，则只有原始数字可获有。使用更新的信息来构建预测模型引入了偏差，因为它不能正确地反映在需要做出投资决定时可用的信息！

9.6.4数据迁就

数据迁就是在数据中发现“模式”，即使这些模式实际上并不存在。当人们不能进行控制实验而必须依靠记录的数据和统计推断来寻找模式时，这个问题尤其突出。这正是金融数据的情况。

拥有足够的时间和足够的数据，人们可以找到想找到的任何数据模式。因此，确

保投资组合经理使用的任何分析模型都建立在合理的经济理论和常识基础上，而不仅仅是经验证据，是非常重要的。

9.7 风险分解、风险归因和业绩归因

风险分解就是将投资组合方差根据是否持有或风险类型分解为个体贡献在使用因子模型进行风险分解的情况下，人们确定投资组合对因子模型中不同因子的风险敞口。

9.8 因子投资

因子模型自20世纪70年代以来就一直存在，并用于阿尔法构建和风险管理环境

因子投资策略——有时称为聪明贝塔、科学贝塔、基于因子的投资或基本面指数化有巨大的多样性，但它们所共有的是对历史表现良好或投资经理认为可能继续增加风险调整后回报的因子的持续敞口。因子投资已经改变了特定资产类别中的传统资产分配和投资组合选择。

在传统的资产配置策略中，投资组合在不同资产类别之间是多元化的(例如，60/40投资组合将包含投资于股票的资金为60%和债券为40%)。然而，这样的投资组合不一定能够很好地多元化——资产类别之间的相关性往往很高，并且实际上市场困难时会爆发。例如，在“正常”的市场，如2009—2014年期间，美国大盘股与标普500指数完全相关;美国中盘股与标300指数的相关系数为0.96;美国小盘股与标普500指数的相关系数为0.93。根据晨星机构的数据，在2008 2009年金融危机后，这些相关性分别为1.00、0.97和0.96。同样高的相关和相同的现象也可见于发达和新兴的股票市场。

更缜密的投资方法可能需要将资产类别分解成可以解释其相关性和回报/风险特征的共同因子。例如，贝莱德(BlackRock)——一个聪明贝塔产品的供应商，识别出6个跨越不同资产类别的因子:经济(与经济增长不确定性相关的风险信用(违约风险)、实际利率(暴露于实际利率变化的风险)、通货膨胀(暴露于价格变化的风险)、新兴市场(政治动荡的风险)和流动性(在特定时间无法卖出投资的风险通过将投资配置到这些因子而不是资产类别，投资可以避免因资产类别的因子敞口重叠而导致的风险集中。因子本身显示较低的相关性，投资组合更好的多元化。下图显示贝莱德所确定的因子之间的相关性。

ECONOMIC	ECONOMIC	CREDIT	REAL RATES	INFLATION	EMERGING MARKETS	LIQUIDITY
1						
CREDIT	0.58	1				
REAL RATES	0.16	-0.09	1			
INFLATION	-0.37	-0.34	-0.14	1		
EMERGING MARKETS	0.52	0.51	0.08	-0.28	1	
LIQUIDITY	0.44	0.44	0.09	-0.31	0.38	1

Exhibit 8.2 Correlations of major factors that explain asset class returns and co-movements. Correlations measured over the period August 1988 through September 2014.
Source: Hogan, Hodges, Potts, and Ransenberg (2015). Reprinted with permission.