**利用协方差解决生活中最优组合问题**

作者：罗浩

学院：计算机学院

专业：计算机科学与技术

学号：2023211088

# 内容摘要：

从日常角度出发，对日常生活中组合收益最大值进行提取和研究，通过金融投资领域出现投资组合优化问题的提取与分析，总结其中较为传统且行之有效的均值方差分析方法，并从风险衡量的角度将此种方法进行移植以及数据处理上的再优化，扩宽其泛用性。

# 关键字：

协方差矩阵、拉格朗日乘数法、t分布、t检验、风险、信息量

# 目录：

1. **问题引入**
2. **相关知识再拓展**
3. **维度降低**
4. **传统组合优化策略**
5. **应用以解决生活中的组合风险优化问题**
6. **再优化**

### 投资组合优化问题的引入

投资组合优化问题：即识别满足以下三个条件的投资组合： 最小化风险代理、匹配或超过收益代理。、满足基本的可行性要求的问题。

在实际生活中，仅仅通过单一事物的期望比较而推断它的好坏往往是片面而理论的。实际生活中，摆在人们面前的不仅仅是单一选择，而寻找不同选择的组合的最优值——如使用一定的预算去购置电脑配件，从主机到显卡，如何将有限的预算合理分配入不同种类不同等级不同价格的电脑配件从而使性能最大化，以及股民炒股需要购置多只股票在一定的期望回报下风险最小化等等，这些都是长期以来人们迫切想要解决的问题。这篇文章将重点通过协方差、正态分布以及t分布等相关概念的视角去尝试解决这些问题。

### 关于协方差的拓展补充

协方差定义为两个变量的总体误差的衡量，用数学语言表达式即为

虽然方差是特殊的协方差，但两者意义不尽相同，方差表示单个变量的离散程度，而协方差表示两者的变化趋势，如果两个变量的变化趋势一致，那么两个变量之间的协方差就是正值；如果两个变量的变化趋势相反，那么两个变量之间的协方差就是负值。

然而需要注意的是，协方差只能用来描述两随机变量的**线性**相关性，不能通过其值大小判断相关强弱。

协方差矩阵对角线用于表示方差，非对角线位置用于任意两两随机变量的协方差。

### 关于t分布的拓展补充

定义：当随机变量服从正态分布，然而样本量过少，无法准确估计均值方差，这时采用区别于中心极限变换的另一种变换，t变换，使得新的随机变量服从一种t分布，从而容忍样本的变异性。

令随机变量Y等于前n项随机变量的平方和,则新的随机变量服从t分布，t分布在样本量向无穷增长时趋近于正态分布。

通过配合t检验以及根据置信度与自由度查表得来的值判断是否具有相关性，其中t检验为**,**其中Xa表示随机变量x的均值，表示随机变量y的均值。

如果查表得到的t值小于t检验得到的结果，则意味着X与Y具有一定的相关性。

例如，当样本数量n=5时，则自由度df=4，可以查找表中以4开头的行。该行第5列值为2.132，对应的单侧值为95%（双侧值为90%）。这也就是说，T小于2.132的概率为95%（即单侧），此时如果作出假设，H0：其均值为w,若此时根据单样本t检验得到t值大于2.132，那么便有较强理由拒绝原假设，其均值一定不为w，甚至大于w，因此这种差异是有因可究的。

### 利用协方差矩阵特征分解降低维度

在投资组合构建中，常常面临多种选择多种维度，会面临过多的分析以及噪音。通过查询文献，我们可以通过降低维度同时最大程度保留原有信息来简化我们的分析过程，一次完成更高效的组合分析。我们发现，那些在新坐标轴下的数据维度方差越大，那么它们所蕴含的信息量越大，这样能在降低维度的过程中最大保留原有信息。因此可以通过保留信息量最大的坐标轴或者称为新维度摒弃原来方差较小的维度来减低维度。通过将资产投向这些方差最大的主成分，投资者可以实现更高的回报和更低的风险。

事实上，通过计算数据矩阵的协方差矩阵，然后得到协方差矩阵的特征值特征向量，选择特征值最大(即方差最大)的k个特征所对应的特征向量组成的矩阵。这样就可以将数据矩阵转换到新的空间当中，实现数据特征的降维。下面本人提供一种证明。

首先，假设有n个指标，每个指标采集了m个数据，每一个指标的均值为。

对数据进行中心化处理，即保留方差特征，简化接下来的计算：

即要找到某个方向上数据的方差总和最大。假设最大值对应方向为，假设第i个指标对应样本值为yi，那么对应样本的坐标为每一维度的数据在该方向上的投影为。再假设p为单位向量，则满足

那么该维度的在该方向上的总方差计算为（向量表示为每个维度的第i个数据组成的向量）

**那么方差总和为：**

其中

相当于在条件下求解的最大值，构造拉格朗日函数

对p中的每个分量求解偏导得到并进行综合：

即

可知p是对应的特征值为的特征向量，得证。

之后，对特征值从大到小排序，选择其中最大的k个。然后将其对应的k个特征向量分别作为行向量组成特征向量矩阵P。

将数据转换到k个特征向量构建的新空间中，即Y=PX。

因此，我们通过这种方式可以将多种不同的风险投资产品提取为一种主要收益成分，将其归为不同的类进行投资组合的优化。

### 采用协方差的组合优化策略

组合的总和期望收益率：

假设现在已经通过数据收集以及主要成分提取得到市面上n种投资产品为m种主要收益成分，将原始收益数据投影到这m个主成分上（m<n）。

m种主要收益成分的期望收益率为其中，我们采用权重比例来表示对每种收益成分的投资比重向量那么我们可以算出期望的投资收益率（假设为）此时已知每个产品的投资比重总和一定为1，即，那么问题便转化为满足约束条件的期望的总和的投资收益率最大值。此时只需要遵循贪心的思想，只需要将所持有的所有资金投入到期望收益最大的投资产品即可。这点无可厚非。

然而，更加现实的情况是，我们一般不会将所有鸡蛋都放进一个篮子里，更行要实现一个综合的收益最大化，于是考虑采用协方差矩阵近似计算并寻找最优解。

问题的升级——在确保一定的期望收益条件下将投资风险降到最低。

首先是如何衡量风险，一般来说只要将鸡蛋放入多个篮筐就是一个相对稳健的策略，如竞争中同时押注竞争双方等等。所以采用如果投资的产品相关性较低那么就可以考虑买入，如果相关性太高了，那就尽量避免。考虑使用协方差衡量买入的多种投资产品的相关性。

因此可以直接使用期望收入率的方差进行风险的表征。

下面进行公式的推导，

整理得到：

至此，我们使用方差与协方差完成了对风险的表示

使用表示协方差矩阵：

可将风险表示为：

问题转化为在条件1.达成预期收益 2.以及所有的权重比重之和为1

之下求解风险的最小值之后采用高数课本中拉格朗日乘数法构造多元函数可求解最小值。

设函数为

若令每一个变量的偏导数等于0，求出驻点，可以得到

的形式的解，与综合收益率呈线性关系，因此可以得出固定期望收益最小风险的最优解。

### 组合优化问题的实际运用

假设手中有10000元用于购入不同电脑器件，假设需要买同一品牌的主板、显卡、CPU这三样事物，已知1.不同物件的迭代升级给我们的满足度是不一样的 2.我们想要三种物品的评价尽量高且买入其中一样物品的人对另一种物品的评价不要过于低，假设参考他人的的反馈与社会调查，我们计算得知三种物品的期望性价比为a1,a2,a3设三种事物的分配比例为

定义性价比算法为心理价格除以现实价格。

其中则期望性价比为

此时我们使用用于表示每一件配件的购入对其他配件购入的影响以及每种配件自身的评价割裂来衡量我们买入每一个价格位的电脑配件的风险。

假设我们期望性价比为e（e是一个常数）

则令函数等于

得到求出偏导后的方程

此时假设三者期望性价比分别为0.6，0.8，0.8时，同时假设已求出协方差矩阵具体值，并假设预期性价比为e=0.9：

*可解得现实意义是，尽量先购置性能不错的的CPU，在考虑其他的配件。（数据构造不严谨，仅供参考）*

### 解决非线性关系无法探查问题以及协方差的再优化

在统计数据时，以上若以上数据接遵循正态分布，表示此时收益值均匀分布可以采用中心极限化化为标准正态分布求解概率，结果会比较接近真实情况；同时一般而言大量数据都遵循这一分布。

然而再特殊情况下，例如处理特殊数据或数据量过小时，数据分布可能呈现较为极端的情况；

考虑采用t分布，通过双总体t检验分析每两只股票之间是否相互具有影响，从而在之后可以具体的分析两者的线性关系或者非线性关系。

对于只能衡量非线性关系的情况，可以**图形方法**——通过绘制散点图或其他可视化工具，可以直观地识别非线性关系。

例如，对于X-Y类指数分布（如图1），可以利用e的指数分布进行拟合，最简单的方法可以是设随机变量,这样就相当于分析Z与X的线性关系,从而很好的避免了协方差的局限性。以此类推，可以在二阶或者更高的关系上对不同的组合风险进行表达。

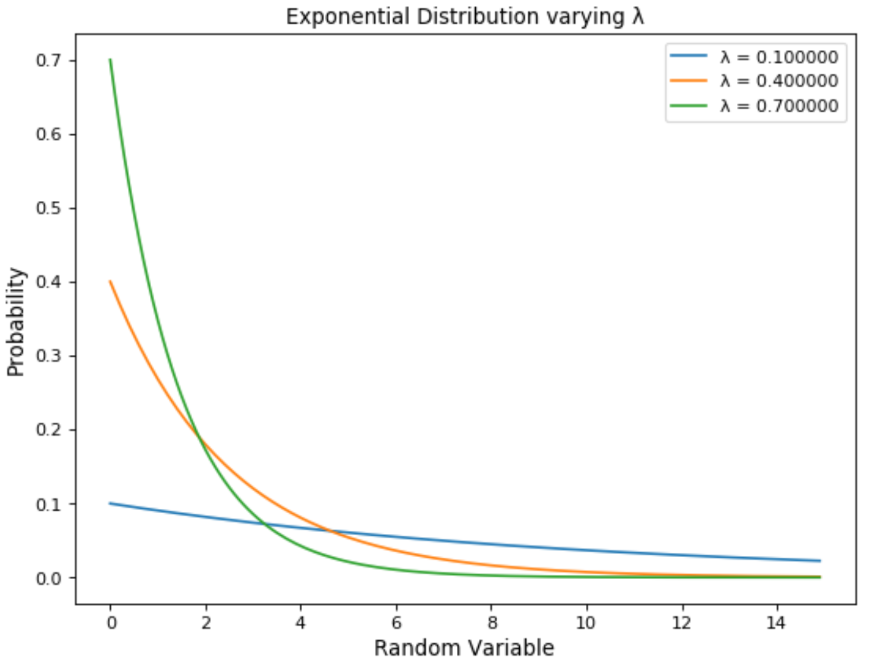


图1

### 结论

传统的组合优化方法适用于风险分散以及多变量的组合的期望值最大化。通过适当的数据统计以及数据处理的优化，使得该方法实用性更广，使之适用于非线性关系以及样本量更少的数据。

参考文献：

维基百科——t检验/t分布

概率论与统计学在经济学中的应用 洪永淼(1)(,)(2)1. 中国科学院数学与系统科学研究院, 北京 100190 2. 中国科学院大学经济与管理学院, 北京 100190

《投资学》 （美国）滋维•博迪（Zvi Bodie） （美国）亚历克斯•凯恩（Alex Kane） （美国） 艾伦 J.马库斯（Alan J.Marcus） 译者：汪昌云 张永冀等