**2022年7月A股所有股票市值变化率统计分析**

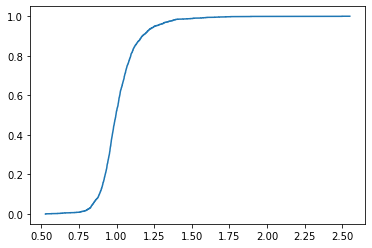
**摘要**

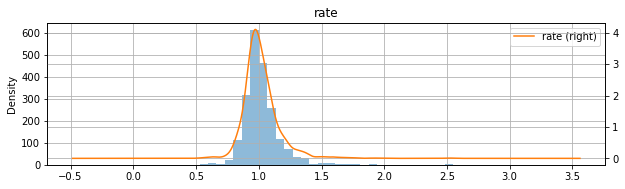
本文通过python程序借助开源网站baostock.com爬取了2022年7月份A股所有股票的市值，从而计算出在这个月内他们的变化率作为数据集。根据经济学基本知识，猜测应当服从对数正态分布，于是进行了一系列统计学分析，包括计算数字特征、估计参数与分布、检验、方差分析、回归分析等。同时由于具有良好的可扩展性，例如日期、时间跨度和股票过滤，故作为项目发布于github.com

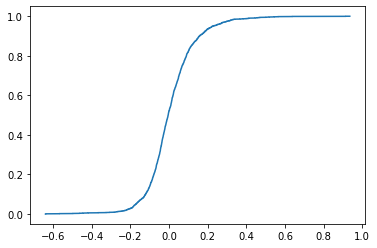
正文

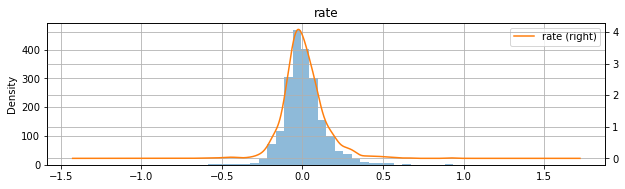
本文数据附于文章末尾。

变化率的均值为1.0181518541654626，方差为0.02050602513579442。

 经验分布函数为：

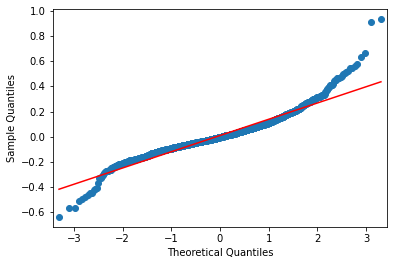
 阶梯函数为（近似为经验密度函数）：

对数化之后经验分布函数为：

 对数化之后阶梯函数为：

接下来进行分布估计，对对数化之后的变化率进行分析，根据对数化之后的经验函数可进行假设检验：

其中F(x)是总体的分布函数，是已知的分布函数，即正态分布

 绘制Q-Q图如下：

K-S检验p值为2.552216429613483e-294，明显小于0.05，故拒绝原假设，认为变化率不服从对数正态分布。但实际上查阅相关文献可以指出，在面对样本量较大时，K-S检验对于样本过于敏感，并且与数据可视化图表的观察结果不一致。

利用W检验得出的p值也为1.5558938019348248e-29 < 0.05，故仍然拒绝原假设。但由于图像工具显示更具有说服力，故仍然相信本月变化率服从对数正态分布，并以对数化之后的变化率估计参数。

参数估计：

可得

区间估计，在置信度0.95的情况下，得出a的0.95置信区间为

(0.0036895621609148764, 0.014781695655875098)

的0.95置信区间为

(0.015816159567639192, 0.017851539757211463)

从而的0.95置信区间为

(0.1257623137813518, 0.13360965443115053)

做参数假设检验：

做方差分析：

做回归分析：