

系统工程期末项目报告：针对上交所 上市企业发行审批流程的排队研究

第一部分 项目介绍

绪论：

十一届三中全会以后，我国经济体制开始了由计划经济至市场经济的过渡，社会主义市场经济体制也得到了确立与实施。1990 年末，我国资本市场方兴未艾，为进一步发展和完善社会主义市场经济体制，提升国内市场法制化、国际化、专业化水平，经中国证券监督管理委员会等有关部门批准，上海证券交易所正式成立。建成 30 年以来，上交所秉持初心，已发展成为市场结构较为完整的证券交易所：截至 2022 年末，沪市上市公司家数达 2174 家，总市值 46.4 万亿元；2022 年全年股票累计成交金额 96.3 万亿元，股票市场筹资总额 8477 亿元——上交所已经成为全球第三大证券交易所和全球最活跃的证券交易所之一，股票总市值、IPO 筹资额分别位居全球第 3 名、第 1 名。而对于企业来说，获得权威证券交易所的上市资格能够帮助其拓展更多的金融与用户资源，有助于在一定程度上改善企业的经营状况，对企业扩大规模、改善经营状况、实现创业目标等方面有着积极意义。因此，在这个规模庞大、体系完整的证券交易市场中，上市企业的审批流程自然成为了学者关心的话题之一。此外，对于准备申请 IPO 的企业来说，了解审核流程耗时对其运营决策与经营情况具有重大影响。综上所述，本项目致力于根据相关企业上市申请的受理时间与通过时间，建立一个一般化的排队模型，并根据相关数据分析企业上市申请的审核时间与申请间隔时间分布，归纳出上市公司申请审核的部分规律，为即将在上交所或其他同类型证券交易所上市的企业提供有效参考。

研究方法：

排队论

K-S (Kolmogorov-Smirnov) 检验

第二部分 模型介绍

2.1 企业 IPO

IPO 指 Initial Public Offering（首次公开募股），即企业首次通过证券交易所向公众出售股份，以期募集用于企业发展资金的过程。通常来说，上市公司的股份是根据向相应证券会出具的招股书或登记声明中约定的条款通过经纪商或做市商进行销售的。一旦首次公开上市完成后，提出申请的公司就可以在证券交易所或报价系统挂牌进行交易。

2.2 IPO 审核流程（以上海证券交易所为例）：

提交申请：具有公开募股需求的企业，可以向证券交易所提出首次公开募股申请，而发行人应当向证券交易所提交发行上市申请文件；

受理：以上海证券交易所为例，其在受到发行上市申请文件后 5 个工作日内会做出是否予以受理的决定。若发行人申请文件不符合要求，则应在 30 个工作日内补交申请材料；

审核：上交所审核机构将在受理之日其 20 个工作日内发出审核问询，问询可进行多轮。证券交易所的审核流程与中国证监会的注册时间总计不超过三个月，审核机构认为问询有效后将出具审核报告提交上市委；

上市委会议：上市委在受到审核报告后将召开会议对审核报告及发行上市申请文件进行审议，就其提出的初步审核意见进行分析，并给出审议意见，此流程没有时间限制；

报送证监会：证券交易所结合上市委审议意见，将出具进一步审核意见。证券交易所审核通过的，将审核意见、相关审核资料和发行人的上市申请文件报送中国证监会履行注册程序。中国证监会认为存在需进一步说明或落实事项的，需要求证券交易所进行进一步问询；

证监会注册：中国证监会在 20 个工作日内对发行人的注册申请做出统一或不予注册的决定。

发行上市：若得到中国证监会的同意批复，申请企业在 1 年内均有发行上市的权利。发行人应按照相关规定在有效期内发行股票，发行时间点由发行人自主选择。

IPO审核流程



图表 1 上交所 IPO 企业申请审核流程

2.3 审核流程分析：

整体来看，虽然大部分审核程序都规定了一定的处理时间上限，但上市委审议流程并未单独规定处理时间。另一方面，上市委作为决定企业是否具备上市资格的关键机构之一，其决定将承担承上启下的角色，直接影响到企业后续的证监会审核流程，因此上市委的审议流程将在 IPO 企业申请审核流程中发挥重要作用。此外，上市委审议时间的无规定性与不确定性为上市企业申请时间的随机性创造了条件。基于该审核流程的随机性，我们可以将证券交易所抽象为一个随机模型，并假设 IPO 企业的审核流程为一个随机过程。

第三部分 模型介绍与假设

根据先前的假设,我们选取了国内最大的证券交易所之一——上海证券交易所的公开项目动态数据,建立了相应的分析框架并设计了一个模型:我们考虑一个一般化的上市审核系统,其具备一定规模,且在固定时间内营业,而提交 IPO 申请的企业的到达过程是随机的。此外,上市审核系统的审核过程也使随机的。若将系统抽象为服务台的话,我们假设系统为无穷服务台系统。

需要特别指出的是,我们假设上市审核系统的容量是无穷的,申请企业也无需排队,即同一时间或时刻上申请上市的企业数目没有限制,所有上市的企业的应用均能第一时间受理并开始接受服务,不存在等待队列。原因是证券交易所作为一个专门处理上市请求的机构,通常具备高度并行处理的能力,他们可能配置了大量的工作人员和专门的系统来处理递交的上市申请。这样的配置使得交易所能够同时处理多个企业的请求,而不需要等待前一个请求完成。

对于审核的具体内容,我们认定提交审核申请的企业均为具备相应上市资质且均能通过审核的企业,中途终止上市或补充上市的企业均不纳入本模型考虑及分析范畴。

为方便研究,根据上交所的申请审核流程与真实数据,我们将上市审核系统的服务流程简化为:系统受理申请、系统通过申请,分别对应上交所审核流程的“受理申请”与“上市委审议通过”两个环节。如此对流程进行简化能够在保证原有假设随机性的前提下最大限度地进行模拟分析并且使实验结果具有显著性与可参考性。

第四部分 模型分析

4.1 输入分析

4.1.1 数据收集:

根据模型与假设,我们收集了 2019 年 3 月 22 日至 2022 年 6 月 30 日在上海证券交易所申请上市并审核通过的 542 家企业的相关数据,并根据其“受理日期”和“状态更新日期”(即审核状态通过的更新)获得了其申请的审核时间和不同企业到达时间间隔。原始数据见附录。

4.1.2 数据处理:

在到达时间数据中,总体上到达间隔都比较短,到达速率较快,故我们不能忽略非工作日的间隔对到达间隔的影响,所以将两个企业之间的到达间隔处理为工作日数。由于非工作日不会有审核得到受理,所以也对审核时间进行了同样的处理,去除所有非工作日。

4.1.3 数据分析结果：

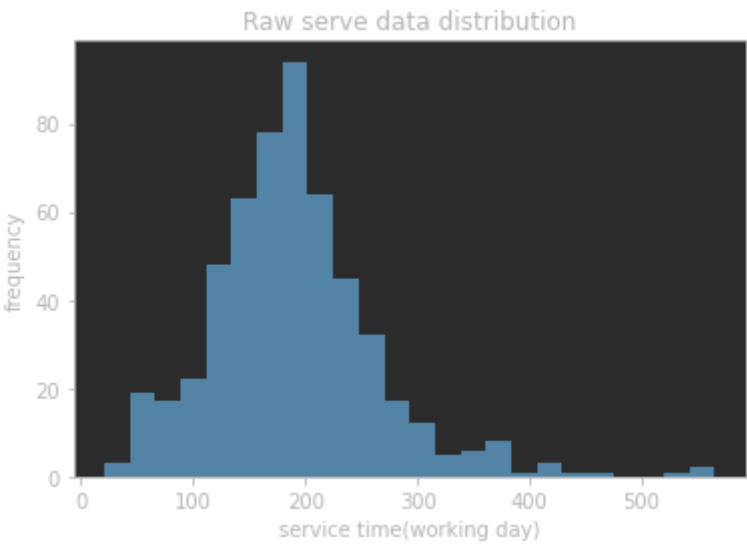
我们对到达间隔和审核时间进行了数据分析，得出了两组数据的描述性统计：

审核时间		到达间隔	
count	542	count	542
mean	189.8321	mean	1.573801
std	74.27132	std	3.764448
min	21	min	0
25%	144	25%	0
50%	185	50%	1
75%	224	75%	1
max	565	max	45

图表 2 原始数据基本统计量

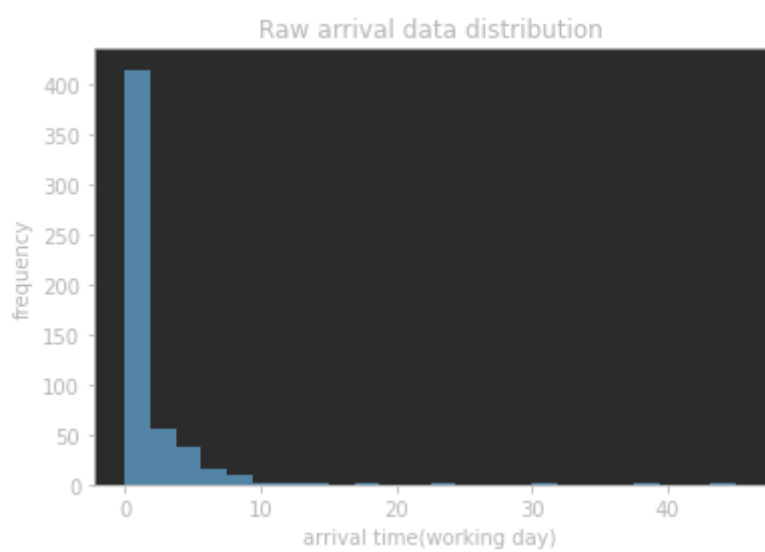
4.1.4 数据分组原则：

为获得较为合理的组数与组距，我们对样本数目（542）开平方根，并向上取整，得到组数为 24 组。以上操作是通过 Excel 软件的 *Roundup* 函数实现的。随后，为得到组距，我们用数据的最大值减去最小值，得到审核时间数据的极差，再用此极差除以组数，得到组距为 31.9。以上结果与使用 Sturge's Rule 得到的组数相比，数据点分布更加分散合理。使用统计软件分析后，审核时间数据的频数分布直方图如下：



图表 3 服务时间频数分布直方图

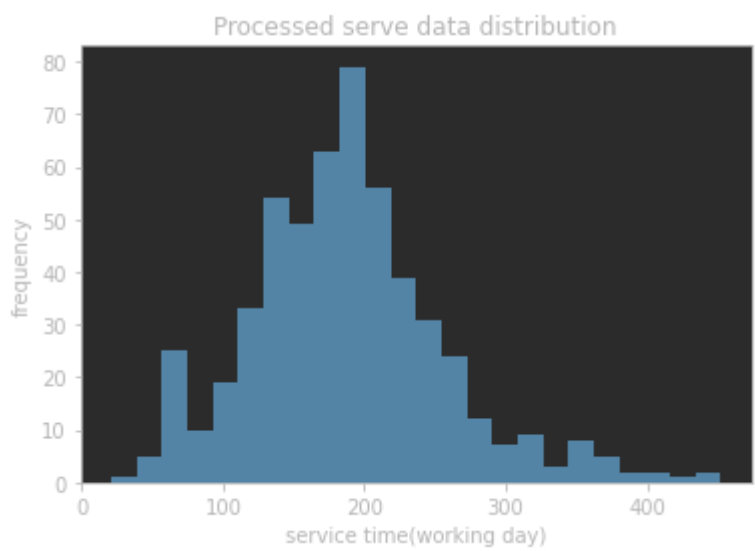
用相同数据处理办法，我们得到申请 IPO 企业到达间隔的频数分布直方图：



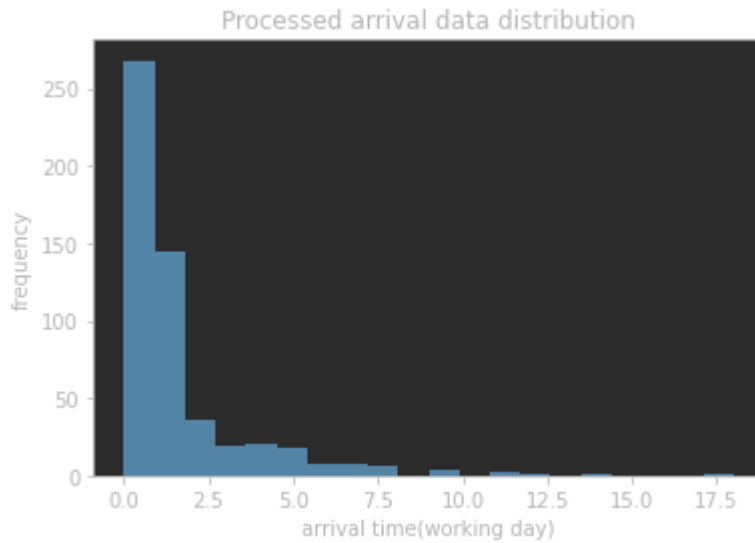
图表 4 到达时间间隔频数分布直方图

从原始数据可以看出，存在极少数审核时间特别长（大于 500 天），或是到达时间特别长（大于 20 天）的异常极端值。这可能是由于企业本身存在问题延长了审核时间，政策调整等异常事件导致了长时间没有企业到达。这些极端值不代表典型情况，因此为了更好地拟合主要的分布，我们忽略审核时间大于 500 天，到达时间大于 20 天的极端值。

处理后的数据分布直方图：



图表 5 处理后的服务时间频数分布直方图



图表 6 处理后的到达时间间隔频数分布直方图

4.1.5 选择分布：

通过 K-S (Kolmogorov-Smirnov) 检验，我们确定了到达间隔与审核时间服从的分布与相应参数。K-S 检验是通过 Python 编程软件的 *distfit* 包完成的。*distfit* 库主要用于进行数据的拟合与分析，我们设置假设检验的置信水平 $\alpha=0.05$ ，并根据常见连续性随机变量设置需要 *distfit* 对象检验的分布，然后将数据应用于 *distfit* 对象进行拟合和转换操作，尝试用给定分布拟合数据，找到与数据最佳匹配的分布。相应代码如下：

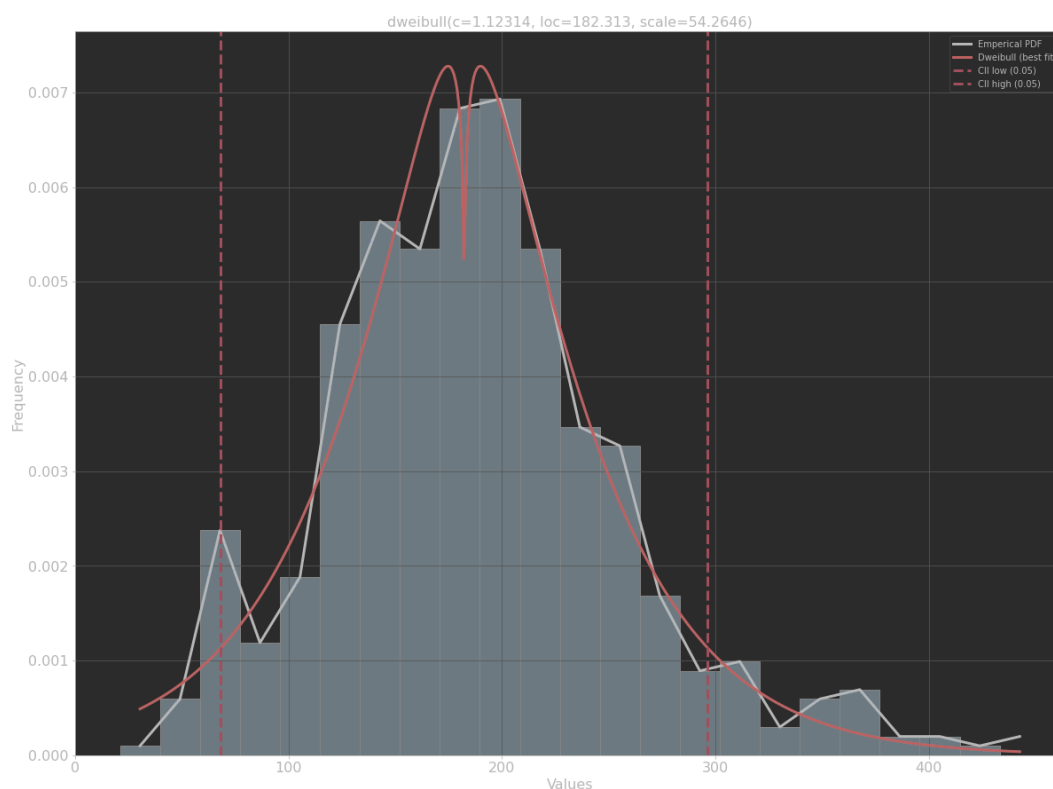
审核间隔检验：

```
dist = distfit()
dist.alpha = 0.05 # 设置 CII KS 检验的显著性水平为 1%
dist.distributions = [
    'norm', 'expon', 'pareto', 'dweibull', 't', 'gamma', 'lognorm',
    'beta',
    'uniform', 'loggamma', 'genextreme', 'weibull', 'triang',
    'cauchy', 'chi2'
] # 添加要测试的分布
dist.fit_transform(X)
dist.model
# 绘制分布图像
dist.plot()
```

代码运行结果如下：

```
[distfit] >INFO> fit
[distfit] >INFO> transform
[distfit] >INFO> [norm      ] [0.00 sec] [RSS: 1.08059e-05] [loc=187.816 scale=69.296]
[distfit] >INFO> [expon     ] [0.0 sec] [RSS: 0.000148102] [loc=21.000 scale=166.816]
[distfit] >INFO> [pareto    ] [0.01 sec] [RSS: 0.00036343] [loc=-0.058 scale=21.058]
[distfit] >INFO> [dweibull  ] [0.01 sec] [RSS: 6.31899e-06] [loc=182.313 scale=54.265]
[distfit] >INFO> [t         ] [0.04 sec] [RSS: 1.06569e-05] [loc=187.331 scale=69.118]
[distfit] >INFO> [genextreme] [0.12 sec] [RSS: 4.1204e-05] [loc=146.674 scale=69.087]
[distfit] >INFO> [gamma     ] [0.01 sec] [RSS: 9.68936e-06] [loc=-107.635 scale=16.053]
[distfit] >INFO> [lognorm   ] [0.05 sec] [RSS: 9.4237e-06] [loc=-230.192 scale=412.428]
[distfit] >INFO> [beta      ] [0.06 sec] [RSS: 9.7739e-06] [loc=-103.187 scale=79619168.488]
[distfit] >INFO> [uniform   ] [0.00 sec] [RSS: 0.000119936] [loc=21.000 scale=431.000]
[distfit] >INFO> [loggamma  ] [0.04 sec] [RSS: 1.18009e-05] [loc=-25156.612 scale=3298.633]
[distfit] >INFO> Compute confidence intervals [parametric]
[distfit] >INFO> Create pdf plot for the parametric method.
[distfit] >INFO> Estimated distribution: Dweibull(loc:182.313041, scale=54.264593)
```

运行结果表示，审核间隔可以被认为服从尺度参数（Scale Parameter）为 54.3、位置参数（Location Parameter）为 182.3 和形状参数（Shape Parameter）为 1.12 的 double weibull 分布。



图表 7 拟合分布函数与数据直方图对比

对数据进行 KS Test，dweibull 拟合分布成功通过了 95%置信度的检验。

```
[distfit] >INFO> Bootstrap: [dweibull ] > Score: 0.35 > Pass 95% CII KS-test: True
```

根据相同的数据分析方法，我们对到达时间的分布进行拟合，输出结果如下：


```
[distfit] >INFO> fit
[distfit] >INFO> transform
[distfit] >INFO> [norm      ] [0.00 sec] [RSS: 4.80969] [loc=1.291 scale=2.200]
[distfit] >INFO> [expon     ] [0.00 sec] [RSS: 3.4349] [loc=0.000 scale=1.291]
[distfit] >INFO> [pareto    ] [0.04 sec] [RSS: 5.53959] [loc=-0.000 scale=0.000]
[distfit] >INFO> [dweibull  ] [0.02 sec] [RSS: 4.55635] [loc=1.000 scale=1.560]
[distfit] >INFO> [t         ] [0.03 sec] [RSS: 5.60956] [loc=0.000 scale=0.000]
[distfit] >INFO> [genextreme] [0.09 sec] [RSS: 5.61237] [loc=0.000 scale=0.000]
[distfit] >INFO> [gamma     ] [0.04 sec] [RSS: 3.61219] [loc=-0.000 scale=4.263]
[distfit] >INFO> [lognorm   ] [0.07 sec] [RSS: 4.70991] [loc=-0.000 scale=0.432]
[distfit] >INFO> [beta      ] [0.06 sec] [RSS: 2.77387] [loc=-0.000 scale=50.355]
[distfit] >INFO> [uniform   ] [0.00 sec] [RSS: 5.38398] [loc=0.000 scale=18.000]
[distfit] >INFO> [loggamma  ] [0.03 sec] [RSS: 4.82448] [loc=-474.932 scale=69.594]
```

我们选取 lognormal 分布作为拟合分布，并进行 KS Test。

```
[distfit] >INFO> Bootstrap: [lognorm   ] > Score: 0.95 > Pass 95% CII KS-test: True
```

结果表明，到达间隔可以被认为服从形状参数(scale)为 0.432，位置参数 (loc) 为 0 的 lognormal 分布

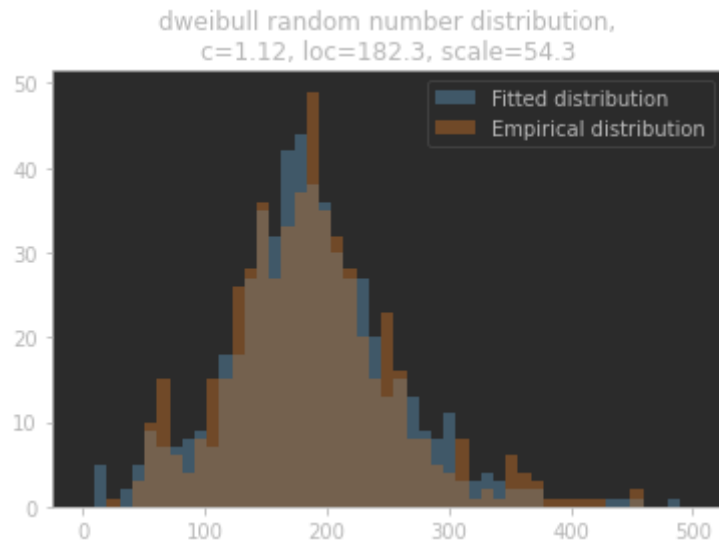
4.1.6 检验分布:

为检验分布的拟合程度, 我们通过生成符合对应分布随机数并绘制相应频数分布直方图的方法对审核时间与到达间隔分别进行了检验。

生成审核时间分布随机数:

```
from scipy.stats import dweibull
size = 550
rvs = dweibull.rvs(c=1.12314, loc=182.313, scale=54.2646, size=size)
rvs = [x for x in rvs if x >= 0]
bins = np.linspace(0, 500, 50)
plt.title('dweibull random number distribution,\nc=1.12, loc=182.3, scale=54.3')
plt.hist(rvs,bins,alpha=0.5,label='Fitted distribution')
plt.hist(X,bins,alpha=0.5,label='Empirical distribution')
plt.legend(loc='upper right')
plt.show()
```

输出结果:



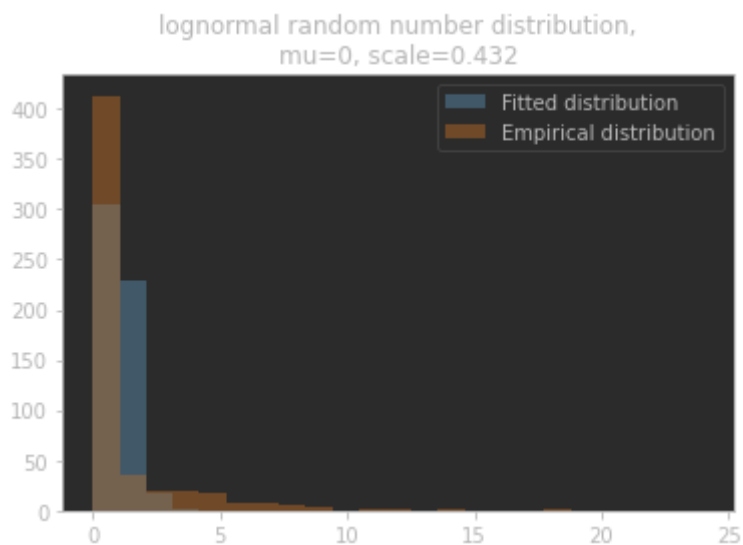
图表 8 拟合分布函数生成随机数的直方图与原数据直方图的对比

可以看出整体拟合效果较好，比较符合经验分布。

生成到达时间分布随机数：

```
from scipy.stats import lognorm
size = 550
rvs = lognorm.rvs(0.432,0, size=size)
bins = np.linspace(0, 20, 20)
plt.title('lognormal random number distribution,\nmu=0, scale=0.432')
plt.hist(rvs,bins,alpha=0.5,label='Fitted distribution')
plt.hist(X,bins,alpha=0.5,label='Empirical distribution')
plt.legend(loc='upper right')
plt.show()
```

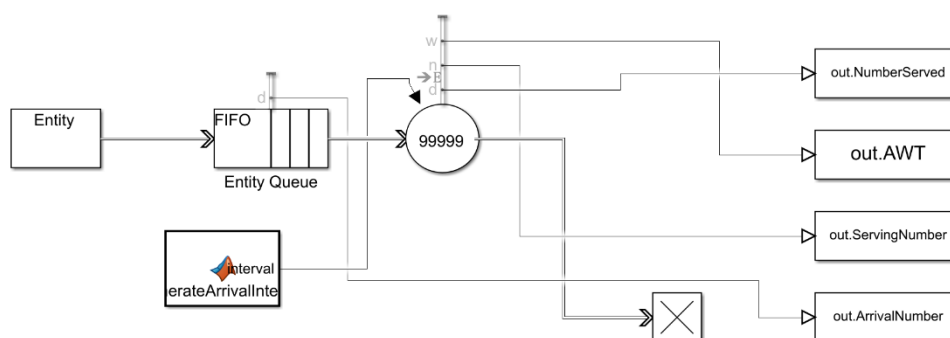
输出结果：



从图中可以看出，到达间隔的经验分布结果更加偏向两边的极端值，而 Lognormal 分布将两边的极端值集中在一起，但拟合分布的随机性可以通过 95% 的 KS Test，我们推测可能原因：企业在上市前一般会寻找券商协作，而券商可能同时帮助多个企业一起向交易所递交上市申请。这意味着在某些时间点，可能会有多个企业同时到达交易所的上市申请。这些同时到达的申请可能导致了到达间隔的极端值，即较短的到达间隔；券商在协助企业进行 IPO 过程中可能需要进行调度和处理。这包括准备和审核文件、与企业进行沟通和协商等。这些过程可能需要一定的时间，而这些时间可能会在到达间隔中体现出来。较长的到达间隔可能是由于券商在处理前一个企业的申请之后，才能处理下一个企业的申请。

4.1.7 模型构建：

利用 MATLAB 的 Simulink 软件来搭建 G/G/inf 排队模型，其中 Entity 是生成器，用于生成企业，其到达间隔服从参数为 (0,0.432) 的 lognormal 分布；Entity Queue 是队列，由于本模型的 Server 容量无限，故企业会直接进入 Server，队列始终为空；Server 是无限容量的服务台，连接一个函数发生器，使得服务台的处理时间服从参数为 (1.12, 182.31, 54.26) 的 dweibull 分布。模型最终输出已服务人数、平均等待时间、正在服务台中的人数、到达人数。



图表 9 基于 Simulink 的模拟程序

为每次模拟选取不同的随机数种子，选取模拟时间 500 个工作日，重复模拟 30 次，获得输出结果。

```
mdl = 'untitled';
isModelOpen = bdIsLoaded(mdl);
open_system(mdl);
NumSims = 30;
all = [];
len = [];

for i = 1:1:NumSims
    seed = i;
    in(i) = Simulink.SimulationInput(mdl);

end

out=sim(in);
```

4.2 输出分析

首先观察 500 个工作日中的排队情况，关注已经完成服务的人数（Number Served）、进入队列的总人数（Arrival Number）、正在排队的人数（Serving Number）三个量，并输出到 excel 中。将单次数据输出到 excel 的代码如下：

```
for i = 1:1:NumSims
    a1 = out(i).NumberServed.Data;
    a2 = out(i).ArrivalNumber.Data(end);
    a3 = out(i).ServingNumber.Data(end);
    a4 = out(i).AWT.Data(end);
    all = [a1,a2,a3,a4];

    filename = 'output.xlsx';

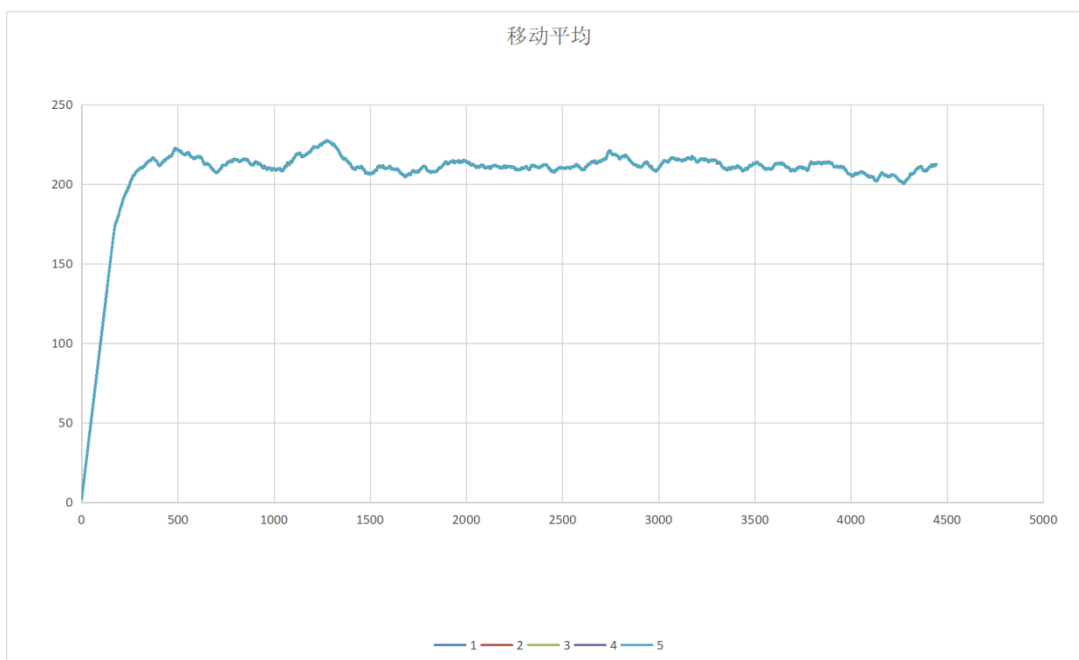
    writematrix(all, filename);
```

共进行了 30 次模拟，在 excel 里输出了 30 行。经过统计发现，500 个工作日内平均有 459.53 ± 2.72 个人加入，其中 248.16 ± 2.23 人被服务， 211.37 ± 2.06 个人仍在等待。具体数据如图：

平均值	248.167	459.53	211.37
标准差	6.2427	7.6055	5.7565
置信区间	2.23388	2.7215	2.0599

然后通过 Serving Number，分析该模型能否到达稳态。我们通过 Welch's procedures，求移动平均，发现在 1000 工作日后可以认为系统内的 serving number 到达稳态。过程如下：

先将 $t=1000$ 时的所有数据写入 excel 中，本次采用了 10 次 $t=5000$ 时的模拟，然后将这些数据求 1-5 项移动平均，把小于 $1/4$ 总数（按照理论中要求的 $w \leq m/4$ ）的各移动平均值量在图中画出，发现该图像在 $t=1000$ 后基本平稳，并且五条曲线基本重合。



值得一提的是，此处仅为“基本”重合。经过模拟，我们发现队列长度实际上在以非常缓慢的速率增加。T=1000 时约为 210，t=5000 时约为 211，t=20000 时接近 212。但在实际应用角度，可以认为队列不再增长，在 1000 工作日后较为平稳，在 210 附近上下波动。也就是说，Warm-up Period 可选择 1000 工作日，即四年。如果继续放宽条件，从图中可以认为 t=500 即两个工作年后，队列长度就较为稳定了。

第五部分 讨论与总结

5.1 系统仿真结果

根据输出分析结果，IPO 企业排队系统将在约 1000 个工作日后达到相对稳定的状态。在此状态下，系统中同时接受审批的企业数目在 210 家左右。

5.2 模型的不足

5.2.1 数据选取还有提升空间：

本研究的数据均来源于上海证券交易所“发行上市审核”模块，由于数据源的限制，我们只能得到 2019 年 3 月 22 日以后的企业审核时间数据。除此之外，考虑到 2023 年企业 IPO 流程进行了较大程度的调整，因此数据的有效性值得商榷。在此情况下，这一时间区间的数据既无法很好地代表过去，亦无法相对精确地预测未来，因此数据的选取具有相当的局限性。这样的研究成果当然具备一定的实际价值，但仍有相当大的提升空间。

另一方面,考虑到中华人民共和国大陆地区还有深圳证券交易所和北京证券交易所两家规模较大的证券交易所,虽然上海证券交易所在三者中结构、IPO 筹资额、所含上市公司总市值均处于领先地位,但片面地选取上海证券交易所的数据依然无法很好地代表中国大陆地区的 IPO 企业审核时间结果。

此外,本研究只选取了审核通过的 IPO 企业进行研究,对于终止上市、补充审核的企业则未予考虑。以上交所“发行上市审核”模块披露的信息为例,2019 年 3 月 22 日以来,在全部记录在案的 1052 家 IPO 企业中,申请成功的企业数目仅为 579 家,占企业总数的 55.03%。由此可见,还有近一半的企业未在模型中加以分析。必须要指出的是,对于审核失败的企业,虽然其审核时间对研究没有特别积极的贡献,但其到达时间依然具备相当的研究价值——因为这与交易所模型的容量等关键指标息息相关。

5.2.2 输入、输出分析检验方法具备一定局限性:

正如前文所述,本研究的输入分析的 k-s 检验等部分是基于 python 编程软件中的 scipy、distfit 等代码库完成的。水平与时间所限,我们未能深入透彻地研究 python 库中每一个算法的底层逻辑,因而对 K-S 检验等步骤的认识依然停留在 python 的输出结果上。另一方面,由于不同软件的兼容性问题,对于 IPO 企业到达间隔的分布,我们在输出分析中未能使用拟合效果最佳的经验分布,而是退而求其次,选择了对数正态分布,这可能对输出结果的实用性造成一定影响。

没有充分考虑 IPO 企业上市的相关因素:

除了前文提及的券商因素外,还有其他可能影响 IPO 企业审核流程的因素可能会影响审核流程,如突发公共卫生事件、总需求总供给发生波动、行业环境发生变化等。对于其他可能影响上市流程的因素,我们未能将上述潜在影响因素所发生的时间与审核时间的波动一一对应,从而未能发现审核时间变化的原因。

5.3 后续研究展望

5.3.1 细化研究方向:

其一,可以将 IPO 企业按照行业、规模等属性进行划分:

由于企业行业背景不同,其融资方式、上市需求、申请材料可能都存在着许多差异。如上交所科创板上市的九号有限公司,由于其股权结构、主营业务的特殊性,其上市时间整整花费了 514 天,但其他行业或规模的企业审核时间就相对较短。由此可见,不同行业、规模

的企业在审核时间方面具备较大的差异。另一方面，对某一行业或规模的企业进行针对性的研究将更有利于更具代表性研究成果与解决方案的诞生。综上所述，将 IPO 企业按照行业、规模等属性进行划分、进行更有针对性的研究将是后续研究的重要组成部分。

其二，可以将 IPO 企业审批流程进行更加精确的划分：

正如前文所述，IPO 申请流程是包括提交申请、交易所受理、交易所审核、上市委员会、证监会审议、证监会注册、发行等七个流程的复杂过程。本研究为简化研究过程，只选取了“上交所受理”和“报送证监会”两个流程进行分析。虽然其他流程的时间相对确定，但仍有一定的随机性，因此仍然具备相当的研究价值。在后续的研究中，可利用排队网络的研究方法，针对整个系统进行总体的、更加完善的研究，亦可选取若干个其他关键指标进行模拟分析。

5.3.2 完善数据处理过程：

其一，选取更大时间范围内的 IPO 企业数据

由于 IPO 审核政策更新时间较晚，无法产生足够的数据量，因此现阶段内很难对改革后的 IPO 审核流程进行研究。对于改革前的 IPO 审核流程，充足的数据量依然是实验结果科学性的必要保证，因此仍应在选取更长时间区间内的 IPO 审核时间数据进行分析。

其二，将工作日与节假日同时算入审核时间的分析当中

在输入分析中，我们将节假日的时间都予以省略。这对到达间隔来说的确是一个相对合理的决策，因为证券交易所受理与否取决于其是否处于工作状态，跨越节假日的到达间隔不能很好地代表工作状态下证券交易所的受理情况。但对于审核时间而言，审核时间是一个时间跨度相对较大的过程，以上交所数据为例，550 组数据的审核时间均高于 25 天。这说明审核时间必然包括周末等节假日，因此节假日对审核时间而言仍然具有相当的价值，不应忽略。

其三，同时考虑其他证券交易所的 IPO 企业审核情况

若需扩大研究范围。以中华人民共和国境内的 IPO 审核流程为例，为了得到更具显著性的统计数据，后续研究还应纳入深交所和北交所的相关数据。如有必要，也可将港股等相关数据一并考虑。若针对上海证券交易所的 IPO 流程，则也应将不同证券交易所的数据进行对比，比对差异寻找原因，或许也能发现有价值的结论。