The System

We are analysing calls to a taxi company. This is a multiple server system with unknown interarrival time and service time distributions.

The company already uses queuing theory to optimise performance. This business process is commercially sensitive, so we will be modeling a reduced system with single stage service.

The simplified system is still complex because there is more than one type of job and many of the system's characteristics are unknown.

Data

The arrival time (timestamp), answer time (seconds after timestamp) and completion time (seconds after timestamp) are kept thusly:

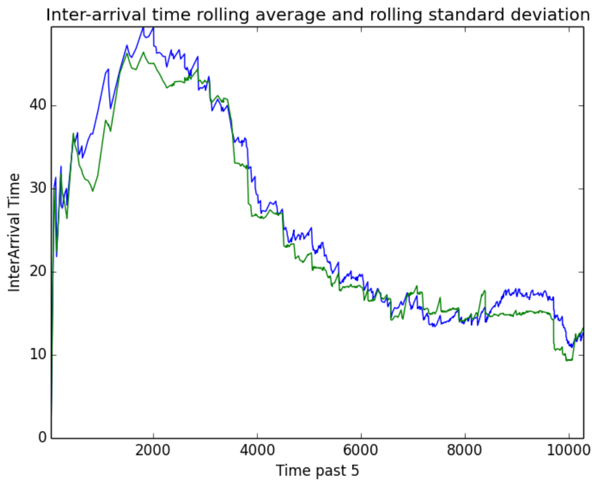
Call Received:	Call Answered: 7:51:21 PM	Call Completed: 7:51:50 PM
7:51:11 PM	10	39

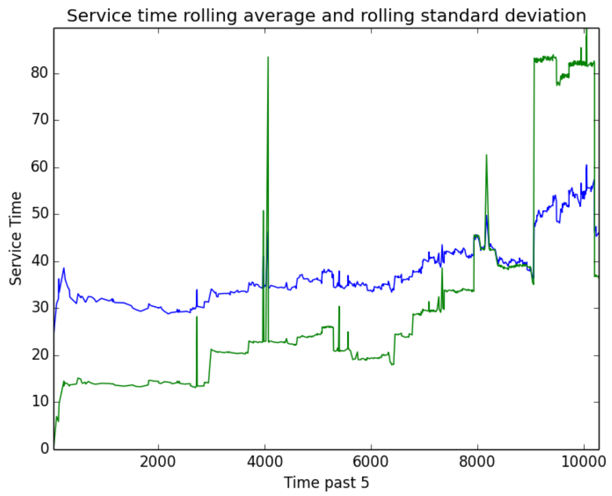
Business process

The business process and the data it produces are shown below.

Timestamp	The center receives a call	Inter-arrival time
-	The call enters the queue	
Call answered	The call is transferred to an available operator	L_q/W_q
-	The customer requests a taxi; they are served	Service time
Call completed	The call leaves the system	L_s/W_s

Estimating Distributions





Our first step in analyzing our dataset was to look at rolling averages and standard deviations. We did two graphs; inter-arrival times and service times. The graphs show running statistics for 50 observations. Averages are shown in blue and standard deviations are in green.

The graphs show:

- Inter-arrival times are not constant.
- Service times are consistent except for a short period of less than half an hour at the end of the observation period.
- The standard deviation of service times spikes at 4000 seconds past 5pm (about 6.05pm) and again at 9000 seconds (7.30pm).

We started our data analysis looking to estimate the most consistent period in the system.

Estimation of parameters

Inter-Arrival Times:
We know the arrival rate is not constant. To test the deviation of the inter-arrival times we regressed them against the hour of day the call was received (5, 6 or 7pm). The mean for 5pm was 40 seconds. At 6pm the mean was 19 seconds and at 7pm it was 14 seconds. It is clear the first hour of the data cannot be used to model the later hours. The full regression output is in appendix B.

Service Time Outliers:
The spike in standard deviation at 6.05pm is a cause for concern. At 6.07pm a call is received that lasts for over 10 minutes. We looked for more outliers.

Checking our exploratory distribution, we have a mean of 40.5 and a standard deviation of 53.7. Using a 99% confidence interval we classed outliers as having service times greater than 177 seconds (2 minutes, 57 seconds). There are 12 observations, listed in appendix C.

9 of the outliers occur in the last half of the observation period. 7 of them are in the last quarter of the observation period.

The Final Distribution

Finalising the subsets:

Our final dataset removed calls before 6pm and calls with service times over 177 seconds. This subset has much lower kurtoses and skewness for both time distributions. The full results are in appendix D.

We tested the beta, Poisson/exponential and gamma distributions for both data sets. The most accurate fit was a gamma distribution. The full output for the gamma distribution is in appendix D with distribution characteristics in appendices E and F.

Parameters

The best fit for both inter-arrival and service times is the gamma distribution. The parameters are shown below.

Gamma Distributions

Inter-arrival times		Service times	
shape	0.70437192	shape	1.76935387
rate	0.03439354	rate	0.04071883

Density Graphs:

Appendix A *Exploratory analysis of inter-arrival and service times.*

We analysed inter-arrival and service times using R and the distrfitplus library.

First, we did an exploratory command called 'descdist'. This command tests the data for appropriate distributions and generates a Cullen-Frey graph. The graph shows estimated distributions based on the data's skewness and kurtosis. Skewness indicates the horizontal symmetry of the distribution's probability density function. Kurtosis measures the probability density function's height and sheerness (Brown, 2015).

Inter-arrival times	Service times
descdist(a$Interarrival.time) summary statistics ------ min: 0 max: 177 median: 13 mean: 20.488 estimated sd: 23.28489 estimated skewness: 2.714051 estimated kurtosis: 13.23815	descdist(a$Service.time) summary statistics ------ min: 0 max: 608 median: 30 mean: 43.45 estimated sd: 53.69818 estimated skewness: 6.145044 estimated kurtosis: 52.98694
summary(fitdist(a$Interarrival.time,'gamma')) Fitting of the distribution ' gamma ' by maximum likelihood Parameters : estimate Std. Error shape 0.70437192 0.037958558 rate 0.03439354 0.002603604 Loglikelihood: -1986.589 AIC: 3977.178 BIC: 3985.607 Correlation matrix: shape rate shape 1.0000000 0.7108808 rate 0.7108808 1.0000000	summary(fitdist(a$Service.time, 'gamma')) Fitting of the distribution ' gamma ' by maximum likelihood Parameters : estimate Std. Error shape 1.76935387 0.102963703 rate 0.04071883 0.002734107 Loglikelihood: -2344.617 AIC: 4693.233 BIC: 4701.662 Correlation matrix: shape rate shape 1.0000000 0.8657891 rate 0.8657891 1.0000000

Appendix B *Interarrival time regression*

Regression showed that the hour the call received had a significant impact on the expected inter-arrival time. Below is a linear model using hour to predict inter-arrival time.

lm(formula = a$Interarrival.time ~ factor(a$Hour))
Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
5pm	39.753	2.270	17.511	< 2e-16 ***
6pm	-20.857	2.746	-7.594	1.55e-13 ***
7pm	-25.698	2.692	-9.545	< 2e-16 ***

Appendix C *Service time outliers*

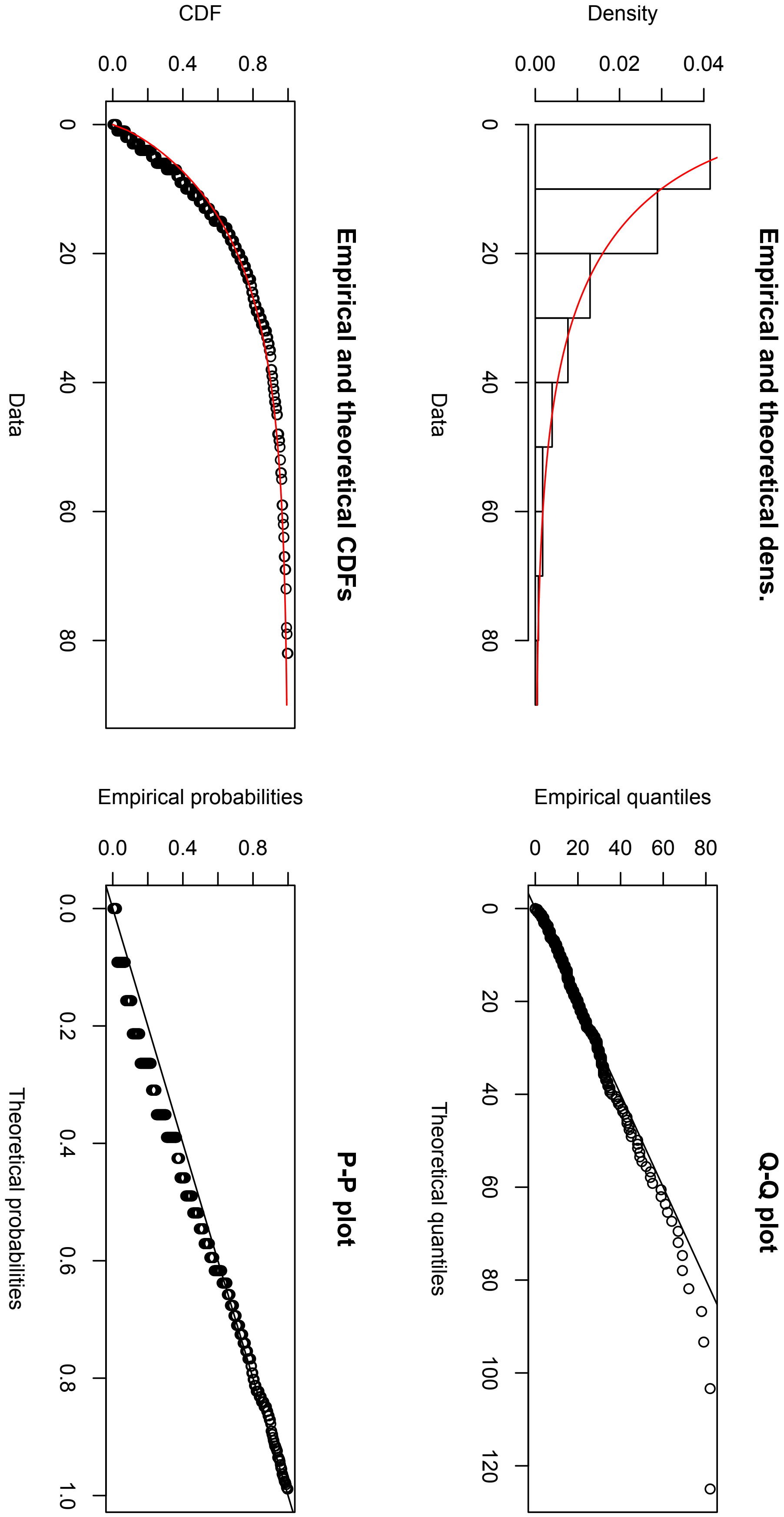
```
> a[a$Service.time>177,]
    Received.time Answer.seconds Completion.seconds Service.time Interarrival.time
66           2730             10                218          208                12
106          3975             22                381          359                11
110          4066             12                620          608                44
176          5399              8                202          194                11
292          7335             10                226          216                16
329          7936             20                277          257                 5
342          8169             22                395          373                32
397          9064             12                582          570                11
447          9719             22                212          190                 1
468          9944             12                241          229                 2
484         10047             63                354          291                 5
486         10051             80                404          324                 3
```

Appendix D *The finalized subsets*

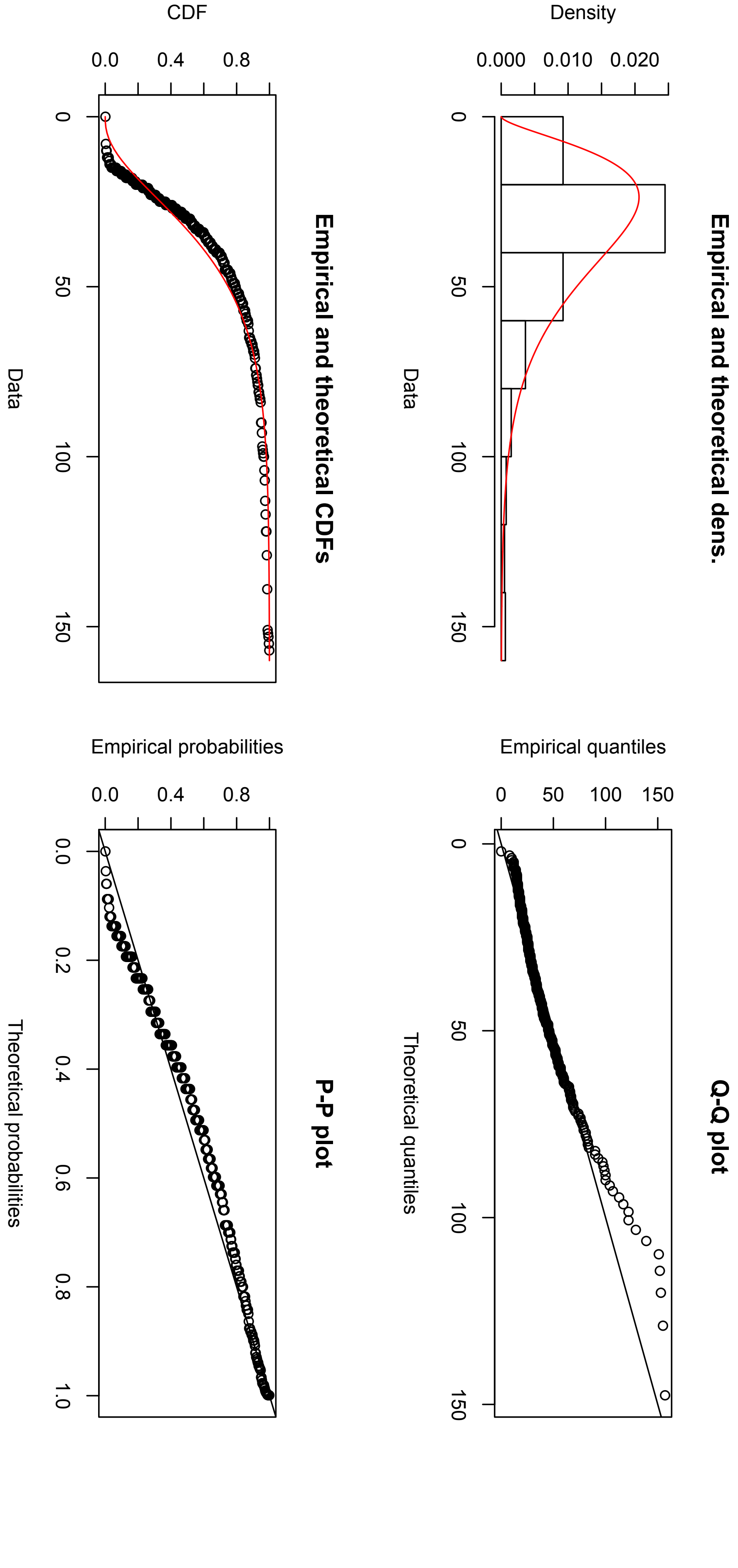
There are 89 calls before 6pm (1 Service time outlier).
There are 12 service time outliers (1 before 6pm).
The final subset has 400 observations.

Inter-arrival times	Service times
descdist(a$Interarrival.time) summary statistics ------ min: 0 max: 82 median: 12 mean: 16.4125 estimated sd: 15.68814 estimated skewness: 1.749913 estimated kurtosis: 6.360868	descdist(a$Service.time) summary statistics ------ min: 0 max: 157 median: 30 mean: 37.98 estimated sd: 25.49371 estimated skewness: 2.157215 estimated kurtosis: 8.74182
summary(fitdist(a$Interarrival.time,'gamma')) Fitting of the distribution ' gamma ' by maximum likelihood Parameters : estimate Std. Error shape 0.81018103 0.049469967 rate 0.04936669 0.004072817 Loglikelihood: -1512.929 AIC: 3029.858 BIC: 3037.841 Correlation matrix: shape rate shape 1.0000000 0.7396077 rate 0.7396077 1.0000000	summary(fitdist(a$Service.time,'gamma')) Fitting of the distribution ' gamma ' by maximum likelihood Parameters : estimate Std. Error shape 2.64961522 0.176720944 rate 0.06974988 0.005120118 Loglikelihood: -1772.312 AIC: 3548.624 BIC: 3556.607 Correlation matrix: shape rate shape 1.0000000 0.9082801 rate 0.9082801 1.0000000

Appendix E *Inter-arrival distribution characteristics graph*



Appendix F *Service rate distribution characteristics graph*



Appendix G *Inter-arrival distribution characteristics graph*

There is an automated message at the start of every call that lasts 7 seconds. Those seconds are included in the time a job spends in the system but are potentially misleading for measuring the efficiency of the server.