Outpatient scheduling using Simulations

A.R. Kahali
(2542764)

Vrije Universiteit Amsterdam (FEWEB)

Simulation and Stochastic Systems - MSc Econometrics and Operations Research
Lecturer: Dr. A.A.N. Ridder,
February 21, 2016

Contents

1 Assignment: Distribution Fitting

1.1 M-estimation and other methods

We are given data i.i.d $\{x_i\}_{i=1}^n \sim F(x_i;\boldsymbol{\theta})$, of which we know that $F(x_i;\boldsymbol{\theta})$ is either the lognormal distribution with pdf as in equation 1 or it is the Weibull distribution with the pdf as in equation 2.

$$f_{LN}(x_i,\boldsymbol{\theta}) = \frac{1}{x_i\sqrt{\sigma^2 2\pi}}e^{-\frac{1}{2}(\log(x)-\mu)^2/\sigma^2} \text{ with } \boldsymbol{\theta} = (\mu,\sigma^2) \tag{1}$$

$$f_W(x_i,\boldsymbol{\theta}) = \alpha\lambda(\lambda x_i)^{\alpha-1}e^{-(\lambda x_i)^\alpha}, \text{ with } \boldsymbol{\theta} = (\alpha,\lambda) \tag{2}$$

We fit the data to both distributions by estimating $\boldsymbol{\theta}$ using Maximum Likelihood. Given the pdf's in equation 1 and 2 and the fact the $\{x_i\}_{i=1}^n$ is i.i.d. we can factorize the joint density function $f(x_1,\ldots,x_n;\boldsymbol{\theta}) = \prod_{i=1}^n f(x_i;\boldsymbol{\theta})$ where we can get the log-likelihood by taking the log of the factorized joint density as in equations 4 and 6.

$$L_n(\boldsymbol{X}_n,\boldsymbol{\theta})_{LN} = \sum_{i=1}^n -\log(x_i) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\frac{(\log(x_i)-\mu)^2}{\sigma^2} \tag{3}$$

$$\propto \sum_{i=1}^n -\log(x_i) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2}\frac{(\log(x_i)-\mu)^2}{\sigma^2} \tag{4}$$

$$L_n(\boldsymbol{X}_n,\boldsymbol{\theta})_W = \sum_{i=1}^n \log(\alpha) + \log(\lambda) + (\alpha-1)\log(\lambda) + (\alpha-1)\log(x_i) - (\lambda x_i)^\alpha \tag{5}$$

$$= \sum_{i=1}^n \log(\alpha) + \alpha\log(\lambda) + (\alpha-1)\log(x_i) - (\lambda x_i)^\alpha \tag{6}$$

We can write the log likelihoods of equations 4 and 6 as average loglikelihoods for our M-estimator $\hat{\boldsymbol{\theta}}_n \in \arg\max_{\boldsymbol{\theta}\in\Theta} Q_n(\boldsymbol{X}_n,\boldsymbol{\theta})$ as in equations 7 and 8. Using the average loglikelihood has the added numerical benefit that it is more robust for maximization.

$$Q_n(\boldsymbol{X}_n,\boldsymbol{\theta})_{LN} = \frac{1}{n}\sum_{i=1}^n -\log(x_i) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2}\frac{(\log(x_i)-\mu)^2}{\sigma^2} \tag{7}$$

$$Q_n(\boldsymbol{X}_n,\boldsymbol{\theta})_W = \frac{1}{n}\sum_{i=1}^n \log(\alpha) + \alpha\log(\lambda) + (\alpha-1)\log(x_i) - (\lambda x_i)^\alpha \tag{8}$$

Assuming without proof that:

$$\hat{\boldsymbol{\theta}}_n \overset{approx}{\sim} \mathcal{N}(\boldsymbol{\theta}_0, \Omega\Sigma\Omega^\top/n) \quad \text{as} \quad n\to\infty \tag{9}$$

where $\boldsymbol{\theta}_0$ is the pseudo-true parameter. We can estimate the standard errors. Σ is the asymptotic variance of the standardized derivative of the log-likelihood, $\sqrt{n}\nabla Q_n(\boldsymbol{X}_n,\boldsymbol{\theta}_0)$. When this derivative function in $\boldsymbol{\theta}_0$ is a *martingale difference sequence*, then Σ is just the variance of the terms in the sequence(note that the data is i.i.d.). Hence, under the null hypothesis $H_0 : \boldsymbol{\theta}_0 = \boldsymbol{\theta}_0^*$, we know that we can estimate Σ by

$$\hat{\Sigma}_T = \frac{1}{n}\sum_{i=1}^{n} \nabla\ell(x_i, \boldsymbol{\theta}_0^*)^2 \tag{10}$$

and, in addition, that $\Omega = \Sigma^{-1}$. For solving the maximisation we employ the non-linear MAXBFGS solver of OX. Past experiences have shown this solver to be more robust than the solvers of Python and even MATLAB.

For the log-normal distribution alternatively we can solve for the estimators analytically, either using maximum likelihood or the method of moments. Analytical solution are easily found, for the MLE we omit the derivations due to time constraints and present the results(derivations are available on request):

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \log(x_i) \tag{11}$$

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=n}^{n} (\log(x_i) - \hat{\mu})^2 \tag{12}$$

We furthermore estimate the parameter using the method of moments (exactly as in the distributions handout). The results of the estimates of the log-normal distribution are found in table 1. As a verification we simulated data: $\{x_i\}_{i=1}^{100000} \overset{d}{\sim} \log\mathcal{N}(2, 0.5)$, we observe that the all estimates are close the original values (note that this is but one expriment, but for our purposes of verifying the program it suffices). We observe for the σ^2 that the estimates between the MLE and the MoM are quite apart. We choose to employ the MLE based on it supremacy according to statistical theory.

For the MLE estimation of the Weibull distribution it is also possible to take analytical derivatives and find a solution for the shape parameter using univariate fault-finding. Then the scale parameter is found by plugging the shape parameter estimate into the analytical scale parameter estimate.(See (Balakrishnan and Kateri, 2008) for example). We omit including this method due to time constraints. In table 2 we have tabulated the estimation results using MAXBFGS in Ox.

1.2 Testing the goodness of fit

To statistically test which fitted distribution better suits the data we employ the Kolmogorov-Smirnov Test. The null hypothesis is that $\{x_i\}_{i=1}^{n} \sim F(\boldsymbol{\theta}), i.i.d.$ where $F(\boldsymbol{\theta})$ is either the Weibull or the log-normal distribution (two-sided test). The statistic is simple:

$$d = \sup_x |F(\boldsymbol{\theta})^{(emp)}(x) - F(x)| \tag{13}$$

Where $F(\boldsymbol{\theta})^{(emp)}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{x_i \leq x}$ is the empirical distribution function, which is random, hence we obtain the statistic D_n. For this test the Python function `scipy.stats.kstest` is readily available. To compute the statistic we have input the CDF of the Weibull and the Lognormal (`scipy.stats.frechet_r.cdf` and `stats.lognorm.cdf`) as a callable lambda function (as in Ox the parametrization of the Weibull is a bit different, see Python code).

To obtain p-values we look at the distribution of D_n. Therefore if we look at the source code for the KS-test(with the options we have used for the function) we find that the p-values are obtained using the complementary cumulative distribution function of the limiting distribution of the two-sided test. (similar p-value approximations to (Wang et al., 2003)).

The results are found in table 3, which indicate that the Weibull distribution is a better fit to the data than the Lognormal (with the used parameters).

2 Assignment: Schedule Simulation

In the previous section we concluded that the consulting times $\{C_i\}_{i=1}^n$ are best fitted by the Weibull distribution, i.e., $\{C_i\}_{i=1}^n \sim \mathcal{W}(\alpha, \lambda)$, where $\alpha = 1.2544$ and $\lambda = 0.0957$. And the consulting times are independent of each other.

2.1 First Schedule

To perform simulations for the $GI/G/1$ queue we first set up a deterministic schedule. The idea is simple we assume that each visitor will need $\mathbb{E}[C_1] = \frac{1}{\lambda}\Gamma(1 + \frac{1}{\alpha}) = 9.7245$ time of consulting, and leaves immediately afterwards. Furthermore the number of patients for the day will be $n = 20$, all patients will be on time, and there will be no no-shows, and the doctor is available from 9:00 AM ($t = 0$), and has no due time of ending his working day. We can then set up the schedule:

$$T = (0 \cdot \mathbb{E}[C_1], 1 \cdot \mathbb{E}[C_1], \ldots, 19 \cdot \mathbb{E}[C_1]) \tag{14}$$

2.2 Simulation

We can then simulate $\{C_i\}_{i=1}^n \sim \mathcal{W}(\alpha, \lambda)$ and using the Lindley equations below we get the waiting times $\{W_i\}_{i=1}^n$ of the visitors and the idle times $\{I_i\}_{i=1}^n$ of the doctor.

$$W_1 = I_1 = 0 \tag{15}$$
$$W_{k+1} = [(W_k + C_k) - A_{k+1}]^+ \tag{16}$$
$$I_{k+1} = [A_{k+1} - (W_k + C_k)]^+ \quad k \in \{1, \ldots, n\} \tag{17}$$

Note that given the schedule T we can easily obtain the inter-arrival times $\{A_i\}_{i=1}^n$ by $(0, t_2 - t_1, \ldots, t_20 - t_19)$. Let $\mathbf{C}$ be a vector of consulting times and $H_W(\mathbf{C}) = \sum_{i=1}^n W_i$, and $H_I(\mathbf{C}) = \sum_{i=1}^n I_i$ and aim to estimate $\mathbb{E}[H(\mathbf{C})] = \int_{\mathcal{C}} H(\mathbf{C}) f(x) d\mathbf{C}$ using crude Monte-Carlo as well as antithetic variables. Denote estimated $\mathbb{E}[H(\mathbf{C})_W]$ and $\mathbb{E}[H(\mathbf{C})_I]$ as $\hat{W}_N$ and $\hat{I}_N$ respectively. In figure 1 we have plotted the average waiting times per visitor.

2.3 Crude Monte-Carlo

Using CMC we can estimate $\mathbb{E}[H(\mathbf{C})] = \int_{\mathcal{C}} H(\mathbf{C}) f(\mathbf{C}) dC$ by $\mathbb{E}[\widehat{H(\mathbf{C})}] = \frac{1}{N}\sum_{i=1}^N H(\mathbf{C}^{(i)})$ for some large N. Here $\{H(\mathbf{C}^{(i)})\}_{i=1}^N$ is a sequence obtained by applying $H(\cdot)$ to N draws of the random vector $\{C_i\}_{i=1}^n$. This is justified by noting that the $\mathbf{C}^{(i)}$ are i.i.d. and assuming that $E[H(\mathbf{C})] < \infty$ using Theorem 2.2.

For a finite sample size N, we have that $\hat{W}_N$ and $\hat{I}_N$ are random and we wish to construct a confidence interval with a certain α level (we set $\alpha = 0.05$). For this we use Theorem 2.1 for i.i.d. sequences (Assuming (2) is satisfied, and (1) is by noting that the $\mathbf{C}^{(i)}$ are i.i.d.) which implies that $\hat{W}_N \overset{d}{\sim} \mathcal{N}(\mathbb{E}[H(\mathbf{C})_W], \sigma^2/N)$ and $\frac{\hat{W}_N - \mathbb{E}[H(\mathbf{C})_W]}{\sigma/\sqrt{N}} \overset{d}{\sim} \mathcal{N}(0, 1)$ if σ would be known. Then using this distribution we can construct a $1 - \alpha$ confidence interval as in equation 21.

Let Z_q be the q-th quantile of the normal distribution.

$$\mathbb{P}(Z_{\frac{\alpha}{2}} \leq \frac{\hat{W}_N - \mathbb{E}[H(\mathbf{C})_W]}{\sigma/\sqrt{N}} \leq Z_{1-\frac{\alpha}{2}}) \approx 1-\alpha \tag{18}$$

$$\text{by symmetry} \Leftrightarrow \mathbb{P}(-Z_{1-\frac{\alpha}{2}}\sigma/\sqrt{N} \leq \hat{W}_N - \mathbb{E}[H(\mathbf{C})_W] \leq Z_{1-\frac{\alpha}{2}}\sigma/\sqrt{N}) \approx 1-\alpha \tag{19}$$

$$\Leftrightarrow \mathbb{P}(\hat{W}_N - Z_{1-\frac{\alpha}{2}}\sigma/\sqrt{N} \leq \mathbb{E}[H(\mathbf{C})_W] \leq \hat{W}_N + Z_{1-\frac{\alpha}{2}}\sigma/\sqrt{N}) \approx 1-\alpha \tag{20}$$

$$\text{thus we get the } 1-\alpha \text{ confidence interval: } [\hat{W}_N - Z_{1-\frac{\alpha}{2}}\sigma/\sqrt{N}, \hat{W}_N + Z_{1-\frac{\alpha}{2}}\sigma/\sqrt{N}] \tag{21}$$

However because σ is unknown we estimate it by $\hat{\sigma} := \sqrt{\frac{N}{N-1}\left(\frac{1}{N}\sum_{i=1}^{N} H(\mathbf{C}^{(i)})_W^2 - \hat{W}_N^2\right)}$, which by the property of the sample variance implies that $\frac{\hat{W}_N - \mathbb{E}[H(\mathbf{C})_W]}{\hat{\sigma}/\sqrt{N}} \overset{d}{\sim} t_{N-1}$. This analytically changes the confidence interval from the one in equation 21 to the one in equation 22. However in practice this doesnt matter much because (as we will see later on) we use a sufficiently large N where the student-t distribution is itself approximately normal.

$$[\hat{W}_N - t_{N-1,1-\frac{\alpha}{2}}\hat{\sigma}/\sqrt{N}, \hat{W}_N + t_{N-1,1-\frac{\alpha}{2}}\hat{\sigma}/\sqrt{N}] \tag{22}$$

In exactly the same way we get standard errors and confidence interval for $\hat{I}_N$.

The estimated price $\hat{W}_N$ depends on N the chosen sample size. In order to make a calculated choice with regards to this decision we make use of the estimator for the relative error $\mathbf{RE}[\hat{\mathbf{W}}_\mathbf{N}]$ and the relative width of the confidence interval $\mathbf{RW}[\hat{\mathbf{W}}_\mathbf{N}]$ defined in the equations below.

$$\mathbf{RE}[\hat{\mathbf{W}}_\mathbf{N}] = \frac{\hat{\sigma}/\sqrt{N}}{\hat{W}_N} \tag{23}$$

$$\mathbf{RW}[\hat{\mathbf{W}}_\mathbf{N}] = 2 \times t_{N-1,1-\frac{\alpha}{2}} \times \mathbf{RE}[\hat{\mathbf{W}}_\mathbf{N}] \tag{24}$$

We use the criterion $\mathbf{RE}[\hat{\mathbf{W}}_\mathbf{N}] < \epsilon$ to determine N, by for each $n \in N$ checking whether the condition is fulfilled or not and stopping when it is the case. Given that $\mathbf{RE}[\hat{\mathbf{W}}_\mathbf{N}]$ is stochastic itself we repeat the experiment 500 times (with $\epsilon = 0.01$)and take the ceiling of the average of these as our N. We do this for both $\hat{W}_N$ and $\hat{I}_N$ (using the same simulation each time).
In figure 2 we see a histogram of the found sample sizes for both $\hat{W}_N$ and $\hat{I}_N$. For $\hat{W}_N$ we get $N = 7331$ and for $\hat{I}_N$ we find $N = 6489$. The reason the algorithm stops earlier for $\hat{I}_N$ is because the relative error depends on the the standard deviation of the estimator. If we look at the graph in figure 1 we see that indeed the idle times are much less variable. This is because due to the large values for $\hat{W}_N$ we get for the schedule we use, the idle times often become negative in the Lindley equation, which makes it zero. And because the idle times are often zero it is less variable. Because the waiting times and idle times go hand in hand we choose the maximum of the two found values, 7331, and round it up to $N = 7500$. In the last subsection of this assignment we discuss the results.

Theorem 2.1. *Central Limit Theorem for i.i.d. sequences*
(1) : *if $Y_1, Y_2, \ldots, Y_N$ be i.i.d. replications of $Y \overset{d}{\sim} f$*

(2) : *if* $\mathbb{V}ar[Y_i] := \sigma^2 < \infty$ $\forall i$ *which implies a finite first moment* $\mathbb{E}[Y]$
Then: $\sqrt{N}(\hat{C}_N - \mathbb{E}[Y]) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$

Theorem 2.2. *Strong Law of Large Numbers:*
(1) : *if* $Y_1, Y_2, \ldots, Y_N$ *be i.i.d. replications of* $Y \overset{d}{\sim} f$
(2) : *if* $\mathbb{E}[Y_i] := \mathbb{E}[Y] < \infty,\ \forall i$
Then : $\frac{1}{N}\sum_{i=1}^{N} Y_i \xrightarrow{a.s.} \mathbb{E}[Y]$ *as* $N \to \infty$

2.4 Antithetic Variables

We also estimate $\mathbb{E}[H(\mathbf{C})] = \int_{\mathcal{C}} H(\mathbf{C}) f(x) d\mathbf{C}$ using Antithetic variables. We explain the procedure for $\hat{W}_N$ and note that the $\hat{I}_N$ is obtained in the exact same way (using W_i, I_i form the Lindley equations). The main difference here is that instead of drawing $\{C_i\}_{i=1}^{n} \sim \mathcal{W}(\alpha, \lambda)$ we draw $\{(C_i, C_j^a)\}_{1=1,j=1}^{n,n}$ where $n = 20$ (number of patients to be scheduled). To specify the distributions of $\{(C_i, C_j^a)\}_{1=1,j=1}^{n,n}$ we employ the inverse transform sampling method. We first draw a vector of 20 uniform random numbers $\mathbf{U}$ on the interval $(0, 1)$ then we define the random vectors $\mathbf{C} = F^{-1}(\mathbf{U})$ and $\mathbf{C}^a = F^{-1}(1 - \mathbf{U})$. In order draw form the Weibull distribution we let $F^{-}1(\mathbf{X}) = \frac{1}{\lambda}(-\log(\mathbf{X}))^{\frac{1}{\alpha}}$. Then using the Lindley equations we get $H(\mathbf{C})_W$ and $H(\mathbf{C}^a)_W$. Letting $M = N/2$ we estimate $\hat{W}_N = \frac{1}{M}\sum_{i=1}^{M} \frac{H(\mathbf{C})_W + H(\mathbf{C}^a)_W}{2}$. Using this method we call the estimator $\hat{W}_N^{anth}$. The variance in this method $\sigma_{anth}^2 = \frac{1}{4}\mathbb{V}\text{ar}H(\mathbf{C})_W + H(\mathbf{C}^a)_W$ can then be shown to be $\frac{1}{2}\sigma^2(1 + \rho)$ where ρ is $\mathbf{C}orr[H(\mathbf{C})_W, H(\mathbf{C}^a)_W]$ and σ^2 is the variance in the CMC method. Hence if we have a large correlation coefficient the variance will be reduced. We estimate $\hat{\sigma^2}_{anth} = \frac{1}{4}(\mathbb{V}\text{ar}[\widehat{H(\mathbf{C}})_W] + \mathbb{V}\text{ar}[\widehat{H(\mathbf{C}^a})_W]) + 2\mathbf{C}orr[H\widehat{(\mathbf{C})_W, H}(\mathbf{C}^a)_W]$, where the estimators for the variances and the covariances/correlations are the usual unbiased estimators (as in slide 22 lecture 5).Then filling in $\sqrt{\hat{\sigma^2}_{anth}} = \hat{\sigma}_{anth}$ and $\hat{W}_N^{anth}$ in equation 22 as well as the equations of the relative error and relative width we get the needed statistics as before for the antithetic variables method.
If we now apply a similar N finding algorithm as applied for the CMC method (see code section) we get the results as visible in figure 3. The ceilings of the averages found are 1128 for $\hat{W}_N^{anth}$ and 497 for $\hat{I}_N^{anth}$. Then for this method we would said $N = 1200$ by the same reasoning as before.

2.5 Comparison CMC and Antithetic Variables

If we set $N = 7500$ for both methods and compute the reduction factor and efficiency as in slide 29 we get the results as as in table 4. We observe that the reduction factors are quite high, and the efficiency obtained for $\hat{W}_N^{anth}$ is 6.3439 and for $\hat{I}_N^{anth}$, it is 12.8469 which shows that the method is indeed better than CMC in reducing the variance even if we take into account the computing time. Note the $\hat{\rho}$ for both estimators which is negative and towards -1, which as stated before constitutes the variance reduction.

2.6 Estimation results for the schedule

In table 5 we can see the results of the estimation with the aforementioned methods and schedule. We clearly see that with this schedule the waiting times for the patients are much larger than the idles times of the doctor. This could have also been seen in figure 1. This can be desirable if we value the time of the doctor more than the time of the patient, we discuss this further in assignment 3. Note furthermore the reduction in variance using the Antithetic variables.

3 Assignment: Advice

3.1 Other Schedules

We turn to previous research on scheduling rules for instance the outpatient appointment rule according to Welch and Bailey (1952) is defined as follows:

$$T = (t_1, t_2, \ldots, t_{20}) \tag{25}$$
$$t_1 = t_2 = \ldots = t_k = 0 \textit{ then for } i > k \ t_i = t_{i-1} + \mathbb{E}[C_1] \tag{26}$$

Welch and Bailey (1952) found that using k = 2 best find the best compromise between the conflicting needs of minimizing patients and the doctors idle time. We note that the schedule used in assignment one is quivalent to the Bailey-Welch rule where k= 1. We thus consider the following schedule:

$$T = (t_1, t_2, \ldots, t_{20}) \tag{27}$$
$$t_1 = t_2 = 0 \textit{ then for } i > k \ t_i = t_{i-1} + \mathbb{E}[C_1] \textbf{ Rule 1} \tag{28}$$

A set of rules tested which is employed by Ho and Lau (1992) in a simulation study are:

$$T = (t_1, t_2, \ldots, t_{20}) \tag{29}$$
$$t_1 = 0 \tag{30}$$
$$t_i = (i-1)\mathbb{E}[C_1] - k\sqrt{\mathbb{V}\mathrm{ar}(C_1)} \tag{31}$$

where $\sqrt{\mathbb{V}\mathrm{ar}(C_1)} = \sqrt{60.8755} = 7.8023$ and Ho and Lau (1992) found $k = 0.1$ to work best. Note that for $k = 0$ this rule reduced to the one used in assignment 2 and for $k > 0$ we have that patients arrive earlier. With the $k = 0.1$ we expect this schedule to perform similar to Rule 1. For these values in equation 31 we call the rule **Rule 2**. Another set of rules employed by Ho and Lau (1992) were the following:

$$T = (t_1, t_2, \ldots, t_{20}) \tag{32}$$
$$t_1 = 0 \tag{33}$$
$$\textit{first set } t_i = (i-1)\mathbb{E}[C_1] \tag{34}$$
$$\textit{modify for } i \leq K \ t_i = t_i - k_1(K-i)\sqrt{\mathbb{V}\mathrm{ar}(C_1)} \tag{35}$$
$$\textit{modify for } i > K \ t_i = t_i - k_2(K-i)\sqrt{\mathbb{V}\mathrm{ar}(C_1)} \tag{36}$$

One of the best found combinations is $K = 5, k_1 = 0.15, k_2 = 0.3$ which we call **Rule 3**. This configuration increases the idle times of the doctor and distributed less waiting time to the patients. The other best found parameter configuration was $K = 5, k_1 = 0.25, k_2 = 0.5$, which distributes waiting times evenly. This configuration performs better for the **Sum** criterion, which with a value of 183.1450 is slightly below the **Sum** of rule 3 (see next section for criteria and table 8 for rule 3 results). Only for the weighted sum criterion **Sum*** it performs less well, **Sum*** = 151.7348. For **Sum**** it performs better too with **Sum**** = 170.5809. In another study by Soriano (1966) it was found that using a two-patient appointment per slot performed well which we also experiment with.

$$T = (t_1 = t_2,\ \textit{empty}\ , t_3 = t_4,\ \textit{empty}\ , \ldots, t_{19} = t_{20}) \tag{37}$$

$$t_1 = t_2 = 0 \tag{38}$$

$$t_i = t_i + 1 = (i-1)\mathbb{E}[C_1]\ \textbf{Rule 4} \tag{39}$$

Lastly as a last schedule that would decrease the idle time of the doctor we simply let the first 5 patients come in at t_1 then at $t = 6*\mathbb{E}[C_1]$ we let schedule the next 5 patients. And so one untill the last 5 patients. We call this **Rule 5**. Another option would have been to follow Kaandorp and Koole (2007), who employ a heuristic local search algorithm to optimize wait time, idle time, and overtime with homogeneous patients with equal slot lengths. However they find the optimal appointment rule to be very similar to the Bailey-Welch rule with particular weighting values.Because that is out of the scope of this assignment and the optimal appointment rule to is similar to the Bailey-Welch rule, we omit that experiment.

3.2 Sum and Weighted sum criteria

As a criterion one can aim to minimize the sum of the (estimated) expected total waiting times and the (estimated) expected total idle times. However, it is widely accepted that the doctors idle time comes at a higher cost than the patients waiting time, but it is also recognized that the doctors time is not infinitely valuable. Thus besides weighing each estimated expected total equally we also use two other sums. We use the criterion **Sum*** (see tables), which correspond to weighing the doctors waiting time as 75% and the patient as 25% as well **Sum**** which corresponds to weighing the doctors waiting time as 60% patient as 40%. The **Sum** criterion corresponds to an equally weighted sum.

3.3 Results

In tables 6, 7, 8, 9, 10 the resulting estimates are tabulated for each scheduling rule. We find that rule 3 find for all criteria the the lowest total estimated expected waiting time. Looking at the estimates it is a schedule that more evenly divides the waiting time between the doctor and the patients, yet is performs best on the **Sum*** and **Sum**** criteria. The estimates according to Rules 1 and 2 are found to be similar to the schedule used in assignment 2. Rule 5 performs (as expected) the worst.Furthermore rules 1,2,4,5 all perform less well than the schedule in assignment 2. Using the parameters of rule 3 (k_1, k_2, K) a possibility would be to minimize either of the criteria using some search heuristic, however due to time constraints we were unable to implement this.

In figures 4, 5, 6, 7, 8, we graphically see the average waiting times per visitor for the patients and the doctor. These figures are in a clear line with the found results.

4 Tables

4.1 Assignment 1

Table 1: Parameter estimation results for the LogNormal fit

		$\boldsymbol{MLE_{num}}$	$\boldsymbol{MLE_{exact}}$	$\boldsymbol{MoM}$	$\mathbf{SE_{MLE}}$
Simulation	$\hat{\boldsymbol{\mu}}$	2.0022	2.0022	2.0014	0.0022
	$\hat{\boldsymbol{\sigma^2}}$	0.50029	0.50029	0.50199	0.0022
	$L_n(\boldsymbol{X}_n, \hat{\boldsymbol{\theta}})$	-215595			
Data	$\hat{\boldsymbol{\mu}}$	1.8970	1.8970	2.0152	0.030664
	$\hat{\boldsymbol{\sigma^2}}$	0.94028	0.94122	0.51602	0.042051
	$L_n(\boldsymbol{X}_n, \hat{\boldsymbol{\theta}})$	-2366.22			

Note: The simulated data is $\{x_i\}_{i=1}^{100000} \overset{d}{\sim} \log\mathcal{N}(2, 0.5)$, we observe that the all estimates are close the original values (note that this is but one expriment, but for our purposes of verifying the program it suffices). We observe for the σ^2 that the estimates between the MLE and the MoM are quite apart. We choose to employ the MLE based on its supremacy according to statistical theory.

Table 2: Parameter estimation results for the Weibull fit

		$\boldsymbol{MLE_{num}}$	$\mathbf{SE_{MLE}}$
Simulation	$\hat{\boldsymbol{\alpha}}$	2.0048	0.0049472
	$\hat{\boldsymbol{\lambda}}$	0.0501	8.3122×10^{-05}
	$L_n(\boldsymbol{X}_n, \hat{\boldsymbol{\theta}})$	-358873	
Data	$\hat{\boldsymbol{\alpha}}$	1.2544	0.0305
	$\hat{\boldsymbol{\lambda}}$	0.0957	0.0025
	$L_n(\boldsymbol{X}_n, \hat{\boldsymbol{\theta}})$	-3234.61	

Note: The simulated data is $\{x_i\}_{i=1}^{100000}\mathcal{W}(2, 0.05)$, we observe that the all estimates are close the original values (note that this is but one expriment, but for our purposes of verifying the program it suffices)..

Table 3: KS-Test results

Distribution	Test Stat.	P-Value
$\boldsymbol{\mathcal{W}(1.2544, 0.0957)}$	0.0188	0.8706
$\boldsymbol{log\mathcal{N}(1.8970, 0.9412)}$	0.0688	0.0001

Note: We see the for the $\log\mathcal{N}(1.8970, 0.9412)$ H_0 is rejected for any reasonable α while for $\mathcal{W}(1.2544, 0.0957)$ the null is not rejected at any reasonable confidence level.

4.2 Assignment 2

Table 4: Performance of the Antithetic variables method relative to the CMC

	$\hat{W}_N^{anth}$	$\hat{I}_N^{anth}$
RED	84.1876	92.1917
Efficiency	6.3439	12.8469
$\hat{\rho}$	-0.3868	-0.6966

Note: We see that the estimated correlation coefficient is indeed negative towards -1 and by a significant amount, which explains the efficiency gains.

Table 5: Estimation results with $N = 7500$

	Estimate	**95%CI**		**RE**	**RW**	**SE**
W, CMC	277.0859	271.7820	282.3898	0.0098	0.0383	2.7057
I, CMC	22.7660	22.3555	23.1764	0.0092	0.0361	0.2094
W, Anth	277.9750	275.8659	280.0841	0.0039	0.0152	1.0759
I, Anth	22.7411	22.6264	22.8558	0.0026	0.0101	0.0585
Sum (Anth)	300.7161					
Sum* (Anth)	173.0992					
Sum (Anth)**	249.66937					

Note: Again it is visible that the Antithetic Variables method constituted a reduction in the variance. Notice also that with the exception of Rule 3 it performs better than the schedules of assignment 3.

4.3 Assignment 3

Table 6: Estimation results with $N = 7500$, for **Rule 1**

	Estimate	**95%CI**		**RE**	**RW**	**SE**
W, CMC	332.4706	326.4284	338.5127	0.0093	0.0363	3.0823
I, CMC	14.8237	14.4553	15.1921	0.0127	0.0497	0.1879
W, Anth	333.9368	331.6656	336.2079	0.0035	0.0136	1.1586
I, Anth	14.8394	14.7217	14.9572	0.0040	0.0159	0.0601
Sum (Anth)	348.7762					
Sum* (Anth)	189.2275					
Sum (Anth)**	284.9567					

Note: This schedule reduces the estimated expected total idle time of the physician and increases the estimated expected total waiting time for the patients.

Table 7: Estimation results with $N = 7500$, for **Rule 2**

	Estimate	**95%CI**		**RE**	**RW**	**SE**
W, CMC	335.1693	328.9560	341.3826	0.0095	0.0371	3.1696
I, CMC	14.9119	14.5374	15.2865	0.0128	0.0502	0.1911
W, Anth	340.0553	337.7848	342.3257	0.0034	0.0134	1.1582
I, Anth	14.3715	14.2531	14.4900	0.0042	0.0165	0.0604
Sum (Anth)	354.4268					
Sum* (Anth)	191.5849					
Sum (Anth)**	289.2900					

Note: The results of this schedule are quite similar to that of Rule 1

Table 8: Estimation results with $N = 7500$, for **Rule 3**

	Estimate	**95%CI**		**RE**	**RW**	**SE**
W, CMC	164.6117	160.7065	168.5168	0.0121	0.0474	1.9921
I, CMC	42.6593	42.0566	43.2619	0.0072	0.0283	0.3074
W, Anth	161.6441	160.0560	163.2323	0.0050	0.0197	0.8102
I, Anth	42.2242	42.0859	42.3625	0.0017	0.0066	0.0706
Sum (Anth)	203.8683					
Sum* (Anth)	144.1583					
Sum (Anth)**	179.9843					

Note: This schedule makes the doctor wait in total wait longer (estimation of expectation) but in the weighted sums it reduces the estimated expected waiting times.

Table 9: Estimation results with $N = 7500$, for **Rule 4**

	Estimate	**95%CI**		**RE**	**RW**	**SE**
W, CMC	340.6431	335.0781	346.2081	0.0083	0.0327	2.8389
I, CMC	20.7974	20.3911	21.2037	0.0100	0.0391	0.2072
W, Anth	339.5388	337.4787	341.5990	0.0031	0.0121	1.0509
I, Anth	20.6960	20.5810	20.8111	0.0028	0.0111	0.0587
Sum (Anth)	360.2349					
Sum* (Anth)	200.8135					
Sum (Anth)**	296.4663					

Note: This schedule shows similar performance as rule 1 and 2.

Table 10: Estimation results with $N = 7500$, for **Rule 5**

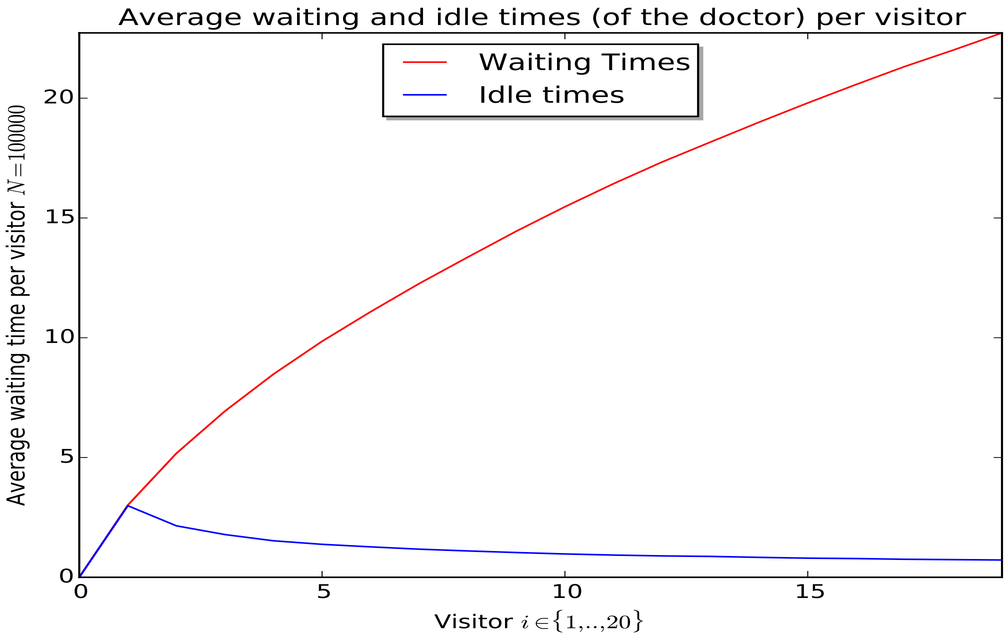
	Estimate	**95%CI**		**RE**	**RW**	**SE**
W, CMC	475.8966	470.3782	481.4150	0.0059	0.0232	2.8151
I, CMC	17.5659	17.1959	17.9360	0.0107	0.0421	0.1888
W, Anth	478.2037	476.1726	480.2347	0.0022	0.0085	1.0361
I, Anth	17.5012	17.3917	17.6107	0.0032	0.0125	0.0559
Sum (Anth)	495.7048					
Sum* (Anth)	265.3536					
Sum (Anth)**	403.5643					

Note: This schedule performs worst

5 Figures

5.1 Assignment 2

Figure 1: Average waiting time per visitor for N = 100000



Note: It is visible that on average the idle times are much lower after each visitor. This is because given the high waiting times the idle times as computed in the Lindley equations often becomes negative and hence zero.

Figure 2: Found N with the stopping algorithm for CMC

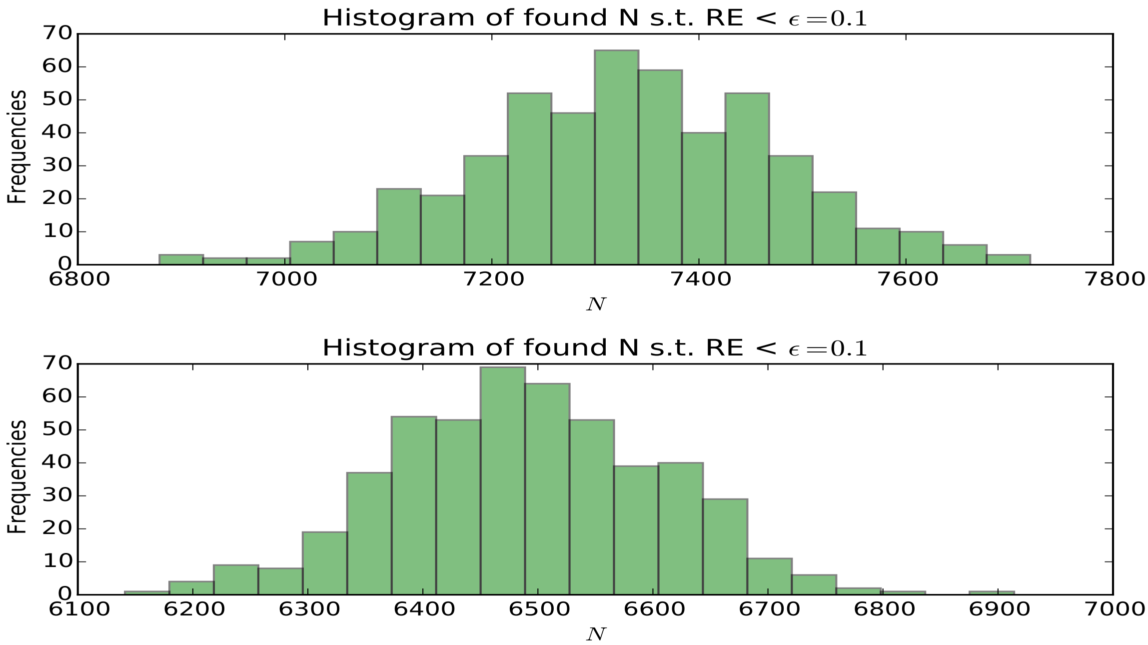
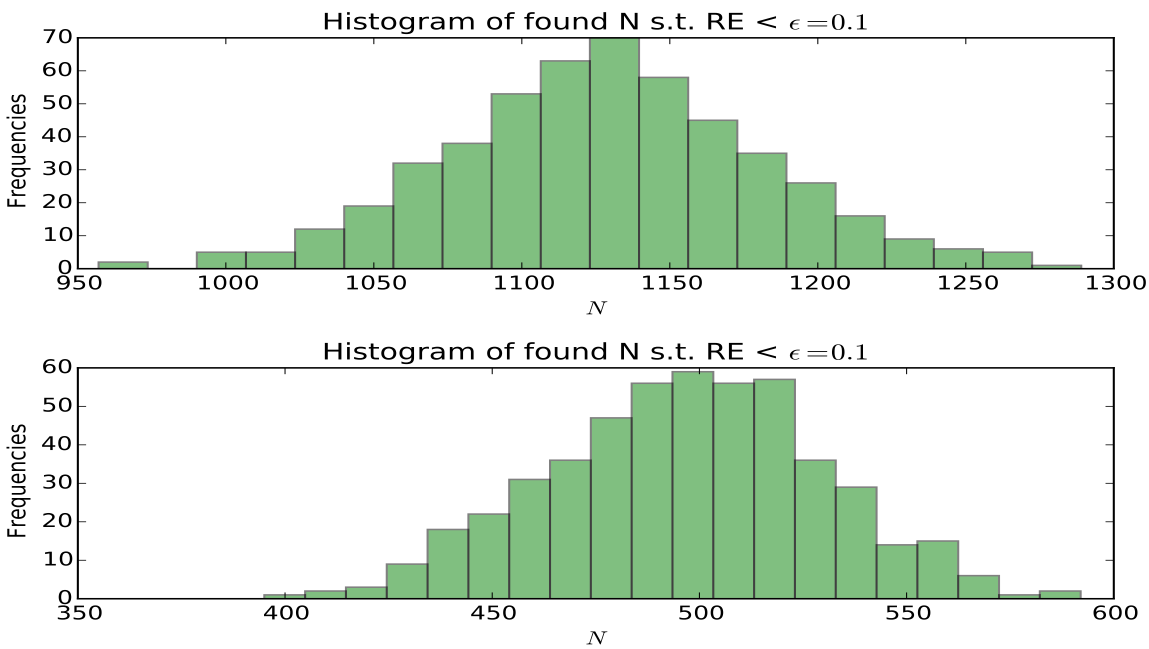


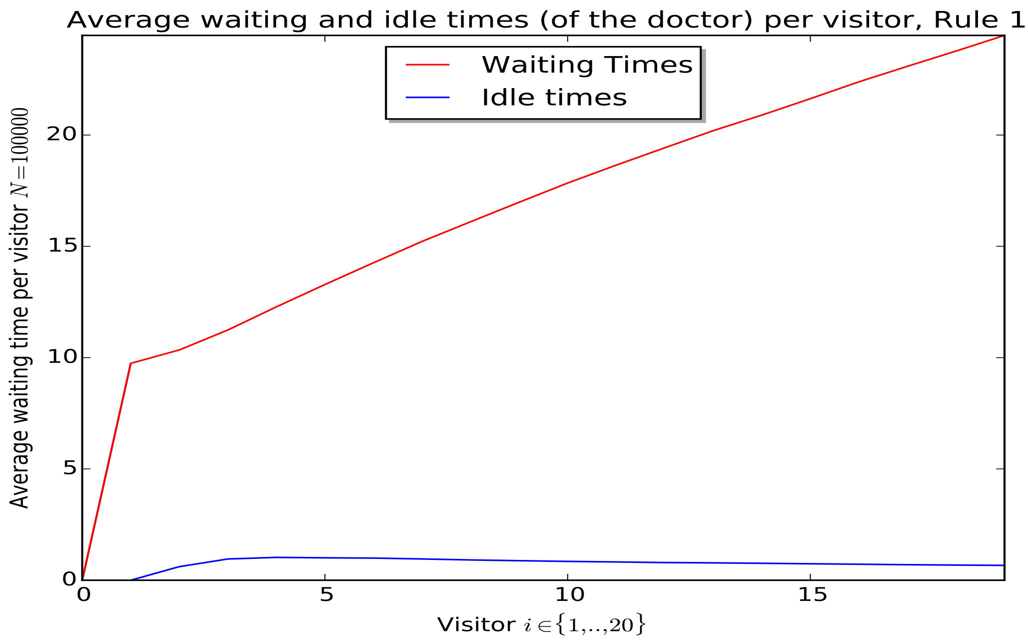
Figure 3: Found N with the stopping algorithm for the Antithetic Method



Note: We see that much lower N is needed to obtain the same relative errors below $\epsilon < 0.01$

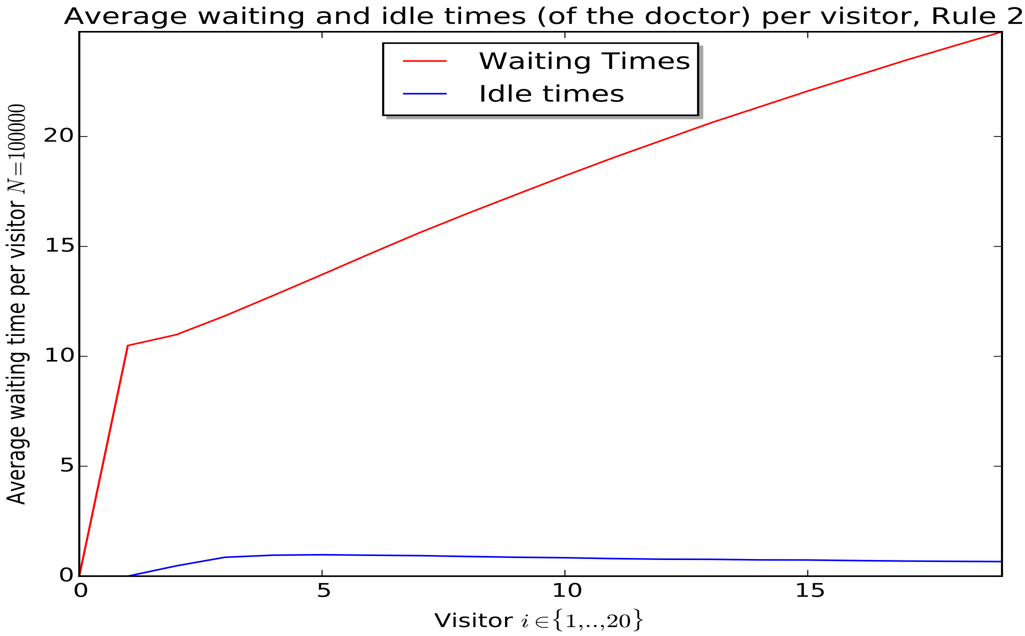
5.2 Assignment 3

Figure 4: Average waiting time per visitor for N = 100000, Rule 1



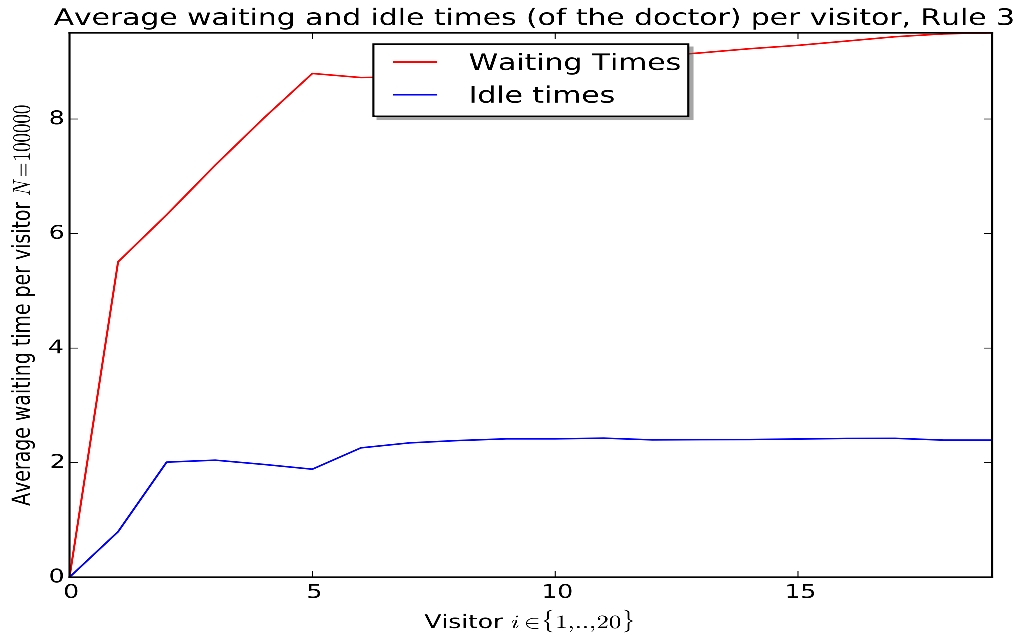
Note: We see that with this rule the average waiting time per patient peaks at the beginning (the first two come at the same time) and then moves upward.

Figure 5: Average waiting time per visitor for N = 100000, Rule 2



Note: We see that this rule (as expected) is similar to rule 2.

Figure 6: Average waiting time per visitor for N = 100000, Rule 3



Note: It is visible that the waiting times are distributed differently here.

Figure 7: Average waiting time per visitor for N = 100000, Rule 4

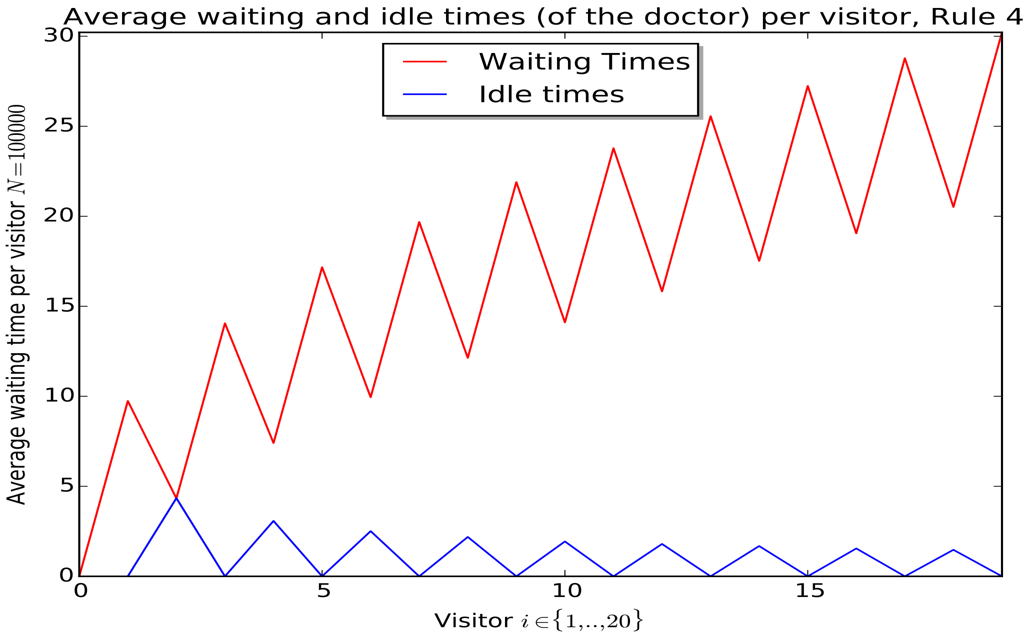
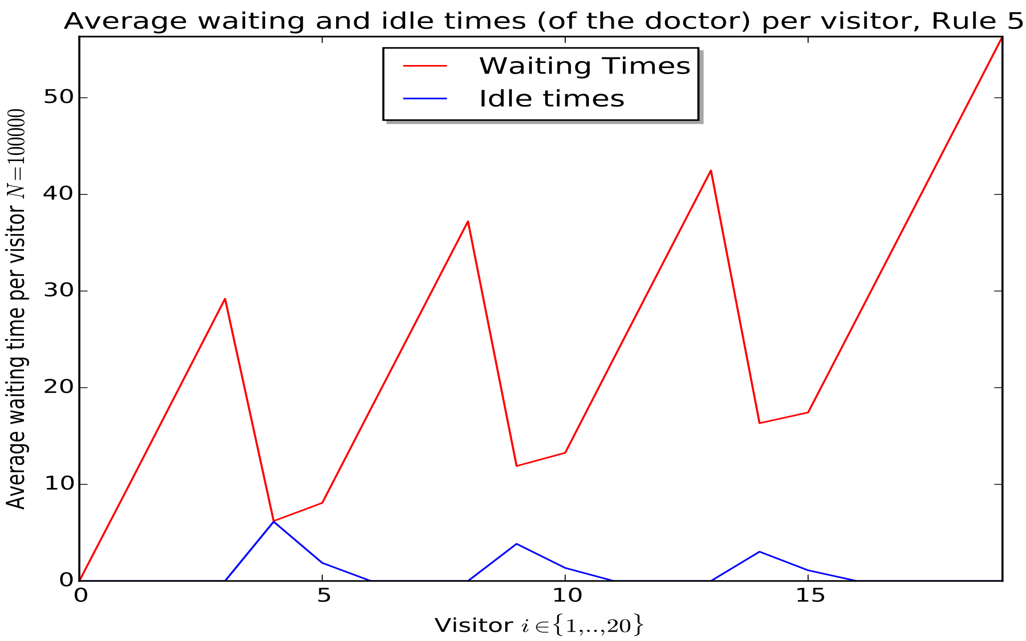


Figure 8: Average waiting time per visitor for N = 100000, Rule 5



References

Balakrishnan, N. and Kateri, M. (2008). On the maximum likelihood estimation of parameters of weibull distribution based on complete and censored data. *Statistics & Probability Letters*, 78(17):2971–2975.

Ho, C.-J. and Lau, H.-S. (1992). Minimizing total cost in scheduling outpatient appointments. *Management science*, 38(12):1750–1764.

Kaandorp, G. C. and Koole, G. (2007). Optimal outpatient appointment scheduling. *Health Care Management Science*, 10(3):217–229.

Soriano, A. (1966). Comparison of two scheduling systems. *Operations Research*, 14(3):388–397.

Wang, J., Tsang, W. W., and Marsaglia, G. (2003). Evaluating kolmogorov's distribution. *Journal of Statistical Software*, 8(18).

Welch, J. and Bailey, N. J. (1952). Appointment systems in hospital outpatient departments. *The Lancet*, 259(6718):1105–1108.