Project 2 : Student number 759130, 472130 and 759875

Problem 1 : Insurance Claims

Let $N(t)$ denote the number of claims received by an insurance company from time 0 to time t. Assume that $N(t)$ is a Poisson process, i.e.

$$P[N(t) = n] = \frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

Here, the continuous time index $t \geq 0$ denotes days from January 1st, 0:00.00.

a)

With intensity $\lambda(t) = 3$, the probability that there are more than 175 claims before March 1st (i.e. $t = 59$) is (due to the law of total probability)

$$P[N(59) > 175] = 1 - \sum_{n=0}^{175} P[N(59) = n] \approx 54.0\%$$

By one simulation through the program `problem1a.m`, with 100 independent realization of insurance claims, the program gave $N(59) > 175$ for 57 realizations, i.e.

$$\frac{\#\text{ of realizations with } N(59) > 175}{\#\text{ of realizatons}} = 57.0\% \approx P[N(59) > 175],$$

thus giving a verification of the calculated probability by simulation. Figure 1 shows all the independent simulations $N^b(t), \quad b = 1, \ldots, 100$

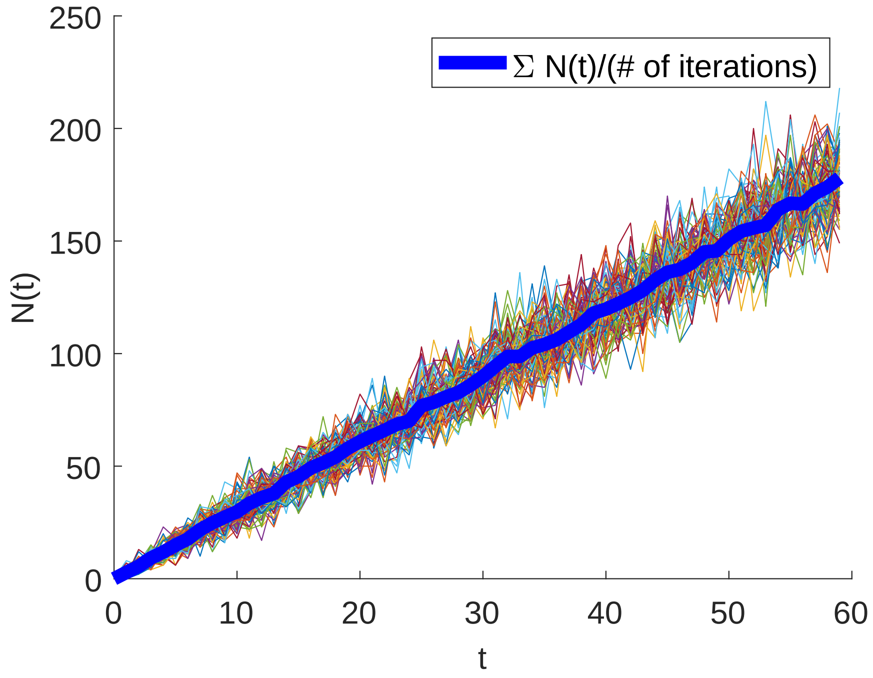

Figure 1: Plot of $N^b(t), \quad b = 1, \ldots, 100$, together with their cumulative average.

b)

With a inhomogeneous intensity $\lambda(t) = 2 + \cos(t\pi/182.5)$, the probability that there are more than 175 claims before March 1st (i.e. $t = 59$) is (due to the law of total probability)

$$P[N(59) > 175] = 1 - \sum_{n=0}^{175} P[N(59) = n] \approx 1.7\%$$

To verify the theoretical answer the code in `problem1b.m` simulates the inhomogeneous intensity by applying thinning of points. One instance of running the code gave $N(59) > 175$ for 2 out of 100 independent realization, thus

$$\frac{\#\text{ of realizations with } N(59) > 175}{\#\text{ of realizatons}} = 2.0\% \approx P[N(59) > 175]$$

This gives us a verification of the calculated probability by simulation.

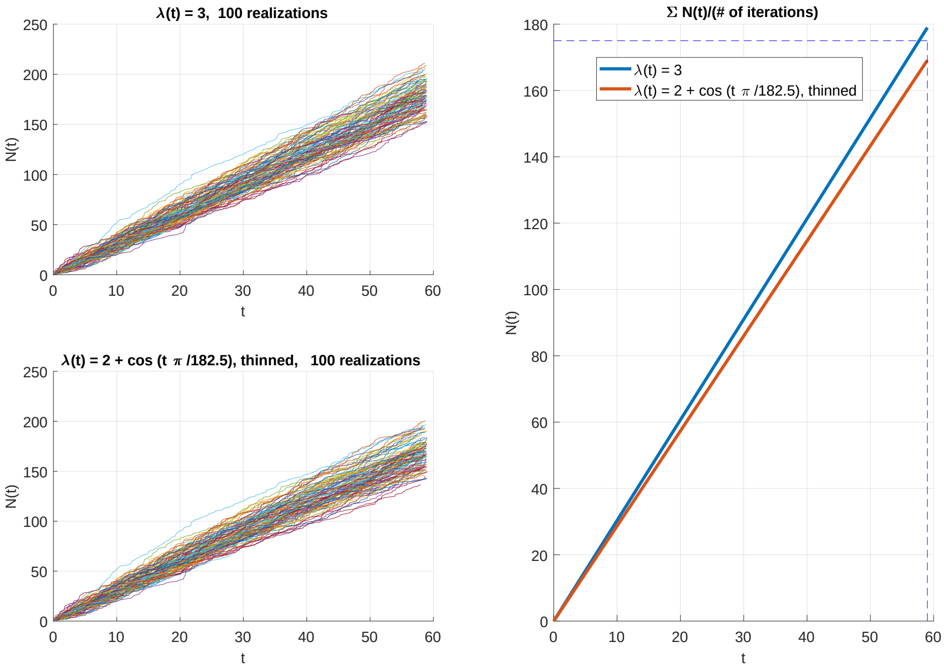

Figure 2: The two plots on the left side shows the number of claims with homogeneous intensity $\lambda = 3$ and inhomogeneous intensity $\lambda = 2 + \cos(t\pi/182.5)$ for $N^b(t), \quad b = 1, \ldots, 100$ realizations. The plot on the right shows a linear approximation of $N(t)$ for homogeneous and inhomogeneous intensity by using the average of end points at $N(59)$, together with dotted lines meeting in the point $(t = 59, N(t) = 175)$.

From Figure 2, we can tell that the difference in number of claims before March 1st i.e. $N(59)$ is relatively small, although the probability of exceeding 175 claims increases significantly from $P[N(59) > 175] \approx 1.7\%$ for the inhomogenous case to $P[N(59) > 175] \approx 54\%$ for homogeneous case. As visible from the endpoints at $t = 59$ and the dotted lines, the average clearly drops below N$(59) = 175$ for the inhomogenous case, and thus yielding such a low percentage of cases with $N(59) > 175$.

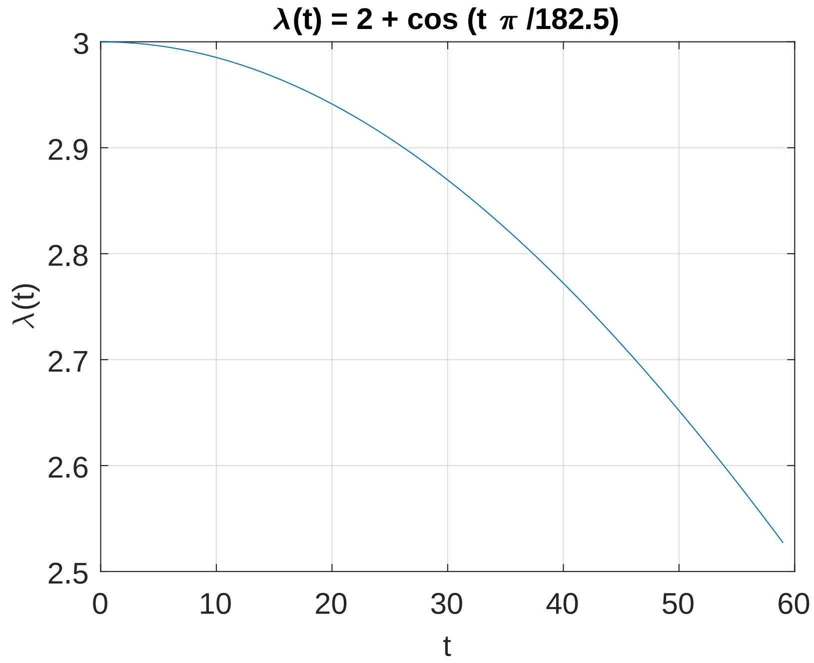

Figure 3: Plot for inhomogeneous intensity $\lambda = 2 + \cos(t\pi/182.5), \quad t = 0, \dots, 59$.

The small differences in number of claims $N(t)$ at $N(59)$ is due to the fact that is the intensity for the inhomogenous case is reduced compared to the homogenous case for $t = 0, \dots, 59$. This leads to few points being removed in the thinning algorithm used in code `problem1b.m` .

c)

From the Law of total expectation and double expectation we get $\mathrm{E}(Z_t) = \mathrm{E}(\mathrm{E}(Z_t|N(t) = n)$, with the number of claims $N(t)$ and size of claim each claim Y_i. This yields

$$\mathrm{E}_N\left(\mathrm{E}_Y\left(\sum_{i=1}^{N(t)} Y_i | N(t) = n\right)\right) = \mathrm{E}(N(t) \cdot \mathrm{E}(Y_i)) = \mathrm{E}(Y_i) \cdot \mathrm{E}(N(t)).$$

The claim amount is log-Gaussian distributed with parameters $\mu = -2$ and $\sigma = 1^2$. The number of insurance claims is Poisson distributed with intensity $\lambda = 3$. This gives us, for time $t = 59$:

$$\mathrm{E}(Z) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \cdot (\lambda t) \approx 39.5.$$

The variance is given by:

$$\begin{aligned}\mathrm{Var}(Z) &= \mathrm{Var}(\mathrm{E}(Z|N(t))) + \mathrm{E}(\mathrm{Var}((Z|N(t))) \\ &= \mathrm{Var}(N(t) \cdot \mathrm{E}(Y_i)) + \mathrm{E}(N(t) \cdot \mathrm{Var}(Y_i)).\end{aligned}$$

With Y_i being log-Gaussian distributed and $N(t)$ Poisson distributed the variance then becomes

$$\begin{aligned}\mathrm{Var}(Z) &= \mathrm{E}(Y_i)^2 \cdot \lambda t + \mathrm{Var}(Y_i) \cdot \lambda t \\ &= \exp{(2\mu + \sigma)} \cdot \lambda t + (\exp(\sigma^2) - 1)(\exp(2\mu + \sigma^2) \cdot \lambda t \approx 24.0.\end{aligned}$$

By running a simulation from `problem1c.m` for $b = 1, \ldots, 1000$ independent realizations, the expected value and variance of the total claim amount at time $t = 59$ ended up being

$$\begin{aligned}\mathrm{E}(Z(t)) &\approx 39.1, \\ \mathrm{Var}(Z(t)) &\approx 22.8,\end{aligned}$$

for the homogeneous case, thus giving a very good verification of the theoretical results mentioned earlier. For the homogeneous case, the same set of realizations gave

$$\begin{aligned}\mathrm{E}(Z(t)) &\approx 33.3, \\ \mathrm{Var}(Z(t)) &\approx 20.3.\end{aligned}$$

For the inhomogeneous case, the results are also as to be expected - both the expected value and variance is slightly lower than their homogeneous counterpart (they're both reduced by a factor of about 0.85). Any Poisson distributed variable has both variance and expected value equal to its intensity, so it's natural that they're both reduced by (almost) the same factor, as they are this case.

d)

The expected discounted amount for one year (i.e. $N(365)$) with discount rate $\alpha = 0.001$ is 204 mill. kr. for the homogeneous case with simulation by code given in `problem1d.m`, for $b = 1, \ldots, 100$ independent realizations. For the inhomogeneous case the simulation by code gave the expected discounted amount to be 136 mill. kr.

In the same simulation, the amount of money the insurance company should hold to be 95% certain to cover next years total amount, is also calculated. After every realization is completed, the top 5% of total amounts are removed and the next largest amount will be at least 95% certain to cover the total amount next year. This gives us approximately 231 mill. kr. and 158 mill. kr. needed for the homogeneous and the inhomogeneous case, respectively.

Problem 2 : Server Jobs

Assume that a computer server can handle 32 jobs in parallel, no matter their size. When all 32 units are on their tasks, an arriving job is forwarded to an- other server, no matter the size of that incoming job. Arrivals to the server follow a homogeneous Poisson process with rate $\lambda = 25$ per hour, and each job takes an exponential time, with rate parameter $\mu = 1$ per hour. This is then a birth-death process with finite bounds, and the number of jobs running on the server is $N(t) \in 0, 1, ..., 32$.

Thus, there are 32 possible states $N(t) \in 0, 1, ..., 32$ in this birth-death process, and for each of these there is a probability of the next state being one higher or one lower than the present one. These probabilities are given by the following relations, as shown on pp. 256 in Kirkwood:

$$P(n, n+1) = \frac{\lambda_n}{\lambda_n + \mu m}, P(n, n-1) = \frac{\mu_n}{\lambda_n + \mu m}. \tag{1}$$

For this server jobs process, $\lambda_0 = \lambda_1 = \ldots = \lambda_{31} = 25, \lambda_{32} = 0$, while $\mu_0 = 0, \mu_1 = 1, \ldots \mu_{31} = 31, \mu_{32} = 32$. This is due to the finite and bounded nature of the process. When there are 0 jobs on the server, the amount of jobs cannot go down. When the server is handling 32 jobs, the amount cannot go up.

These probabilities for all the states may be represented in the embedded Markov chain E and the infinitesimal generator G. For this process, the matrices will look like this:

$$G = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \ldots & 0 \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \ldots & 0 \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ldots & \vdots \\ 0 & 0 & \mu_{31} & -(\lambda_{31} + \mu_{31}) & \lambda_{31} & 0 \\ 0 & 0 & 0 & \mu_{31} & -(\lambda_{31} + \mu_{31}) & \lambda_{31} \\ 0 & 0 & 0 & 0 & \mu_{32} & -\mu_{32} \end{bmatrix}$$

$$
= \begin{bmatrix}
-25 & 25 & 0 & 0 & \dots & 0 \\
1 & -(25+1) & 25 & 0 & \dots & 0 \\
0 & 2 & -(25+2) & 25 & \dots & 0 \\
\vdots & \ddots & \ddots & \ddots & \dots & \vdots \\
0 & 0 & 30 & -(25+30) & 25 & 0 \\
0 & 0 & 0 & 31 & -(25+31) & 25 \\
0 & 0 & 0 & 0 & 32 & -32
\end{bmatrix},
$$

$$
E = \begin{bmatrix}
0 & \lambda_0 & 0 & 0 & \dots & 0 \\
\mu_1 & 0 & \lambda_1 & 0 & \dots & 0 \\
0 & \mu_2 & 0 & \lambda_2 & \dots & 0 \\
\vdots & \ddots & \ddots & \ddots & \dots & \vdots \\
0 & 0 & \mu_{30} & 0 & \lambda_{30} & 0 \\
0 & 0 & 0 & \mu_{31} & 0 & \lambda_{31} \\
0 & 0 & 0 & 0 & \mu_{32} & 0
\end{bmatrix}
$$

$$
= \begin{bmatrix}
0 & 25 & 0 & 0 & \dots & 0 \\
1 & 0 & 25 & 0 & \dots & 0 \\
0 & 2 & 0 & 25 & \dots & 0 \\
\vdots & \ddots & \ddots & \ddots & \dots & \vdots \\
0 & 0 & 30 & -0 & 25 & 0 \\
0 & 0 & 0 & 31 & 0 & 25 \\
0 & 0 & 0 & 0 & 32 & 0
\end{bmatrix}.
$$

The function `makeGE.m` generates the complete matrices for this birth-death process.

a)

To solve this problem, we need to look at the rates in and out of each state. As shown in the embedded Markov chain E the rates in are $\lambda_0 = \lambda_1 = \ldots = \lambda_{31} = 25, \lambda_{32} = 0$, while the rates out are $\mu_0 = 0, \mu_1 = 1, \ldots \mu_{31} = 31, \mu_{32} = 32$.

We exploit that this problem is finite and bounded Markov chain is *positive recurrent.* This holds if and only if

$$
\sum_{n=0}^{\infty} \frac{\lambda_0 \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \ldots \mu_n} = L < \infty. \tag{2}
$$

Then, by Theorem 5.5 on pp. 260 in Kirkwood, the equilibrium probability for each state n can be derived as follows:

$$\pi_n = \frac{\frac{\lambda_0\lambda_1 \ldots \lambda_{n-1}}{\mu_1\mu_2 \ldots \mu_n}}{L}. \tag{3}$$

The function p2ex2a.m exploits this relation to calculate the equilibrium probabilities. First, a value for L is calculated, which verifies the applicability of Theorem 5.5. For this problem, $L = 6.68601\dot{0}^{10} < \infty$. The function also generates a plot of the probability density function for the 33 states. This plot is shown in figure 4.

Note that this PDF has its maximum at $n = 25$, in which case $\lambda_{25} = 25 = \mu_{25}$. This is intuitive as the transition rate in (λ_{25}) is exactly equal to the transition rate out (μ_{25}) of the system at this state.

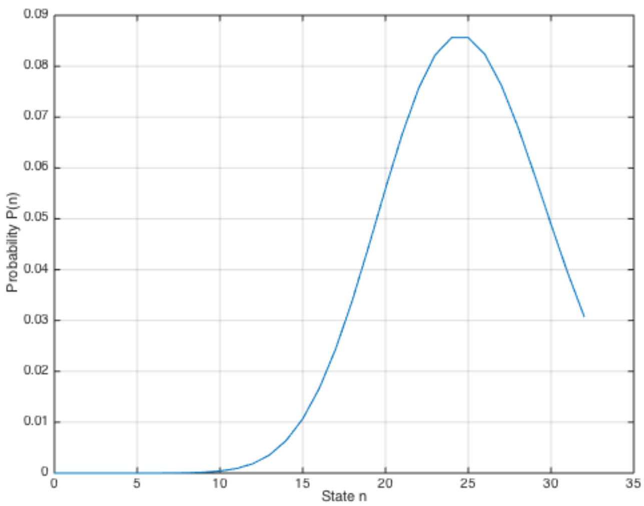

Figure 4: Plot of the probability density function for the 33 states generated by p2ex2a.m. Maximum at $n = 25$.

b)

The function p2ex2b.m simulates 100 realizations of the server job birth-death process. The births and deaths occur at a rate sampled from the exponential distribution, as the waiting time between events in a Poisson process are distributed exponentially.

In every state, the smallest waiting time until a birth or a death decides whether the state moves up or down. The function starts assuming number of jobs on the server $N(0) = 0$, storing every new state in a state vector, and the corresponding time spent in state n in element $i + 1$ of a time vector.

$n = 0$ and $n = 32$ present special cases, due to the bounded nature of the process. In $i = 0$, there can be no more deaths, and thus $P(n, n + 1) = 1$.

At $n = 32$, the server is at capacity, and any further jobs are forwarded to the other server. $P(n, n-1) = 1$. This process continues until $\sum_i t_i > 24 \cdot 7$ as specified in the assignment, at which point the birth-death process for the current iteration ends. The function then repeats this process for the number of realizations specified by the user.

Transient part of the birth-death process

To study the transient part of the process, `p2ex2b.m` calculates the mean state $\bar{n}$ at each state change in the process. This can then be plotted against the cumulative time of the process, as is done for one realization in figure 5. This shows some randomness, and seems to stabilize around $t = 40h$. This tendency can be verified by comparing plots of several realizations.

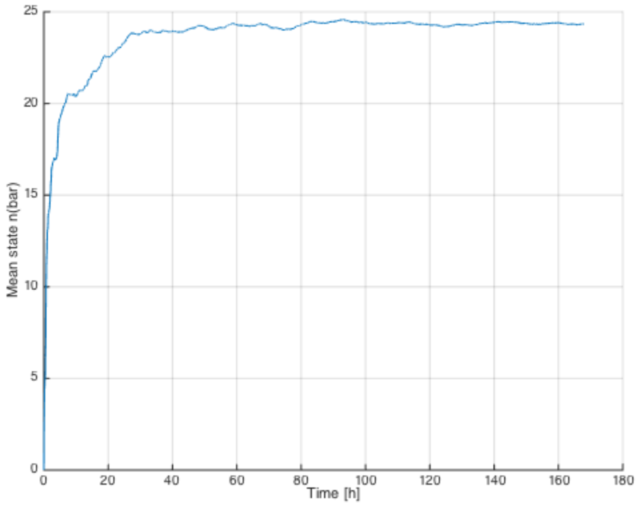

Figure 5: Plot of the mean state i as a function of time. Stabilizes around $t = 40h$

Verification of long term probabilities

`p2ex2b.m` calculates the π_n distribution for the states n by taking the ratio of the time spent in n divided by the total time of the process.

$$[H]\pi_n = \frac{\sum t(state = n)}{t_{total}}. \tag{4}$$

Figure 6 shows the plot generated by one realization of the process, and shows significant randomness compared with the theoretical PDF as calculated in problem 2a). Figure 7 shows the same simulated PDF, but averaged over 100 iterations. This result is very close to the one derived analytically, except for some anomalies in state 31 and 32. We have not been able to debug this, but due to the correctness for the other states, we assume the simulation is largely correct.

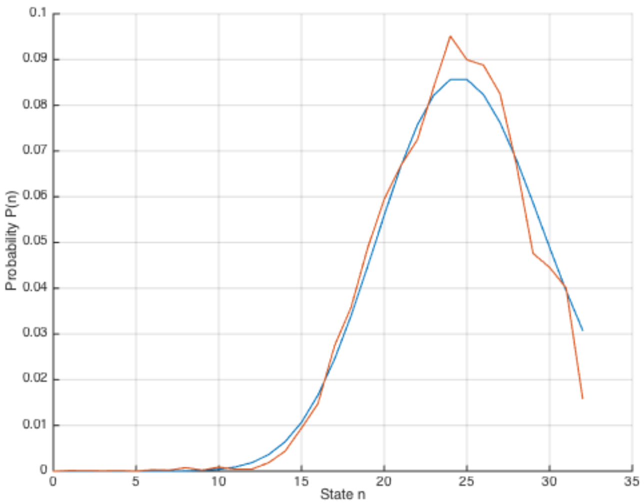

Figure 6: Plot of the probability density function for the 33 states simulated by only one iteration together with the one generated by `p2ex2a.m`.

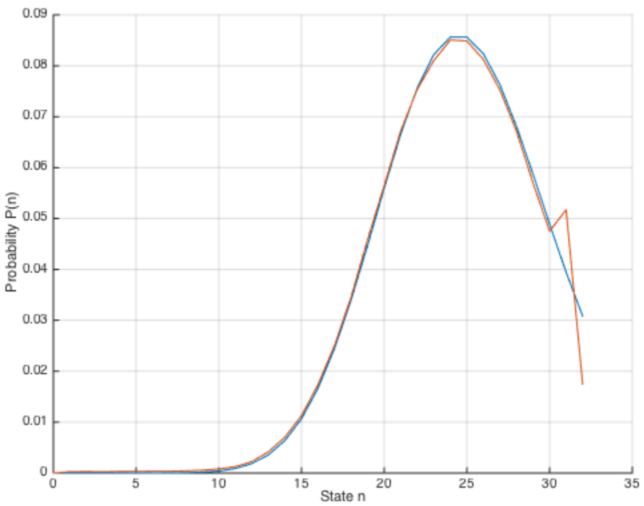

Figure 7: Plot of the simulated probability density function for 100 iterations for the 33 states together with the one generated by `p2ex2a.m`. The simulated graph has some obvious anomalies in states 31 and 32, possibly due to an index error in the function.

Jobs forwarded to the other server

When the system is in state $n = 32$, $\lambda = 0$, and no more births can occur. If one does indeed occur, that is if $t_{birth} < t_{death}$ as sampled from the exponential distribution, the job j is forwarded to the other server. `p2ex2b.m` counts every time this happens in each realization. One can then derive the rate of jobs forwarded per hour by dividing by t_{total}. Averaged over 100 iterations, $j_{forwarded} \approx 0.42$ per hour.