

Written exercise 1

Assignment 1 – 02433 Hidden Markov Models – Anders Hørsted (s082382)

In this assignment a data set with sales figures for a soap product is analysed. The data set contains sales counts for a soap product for 242 consecutive weeks.

Question a: Choosing a model

To start the analysis of the soap sales series, a plot of the data set is created and is shown in figure 1.

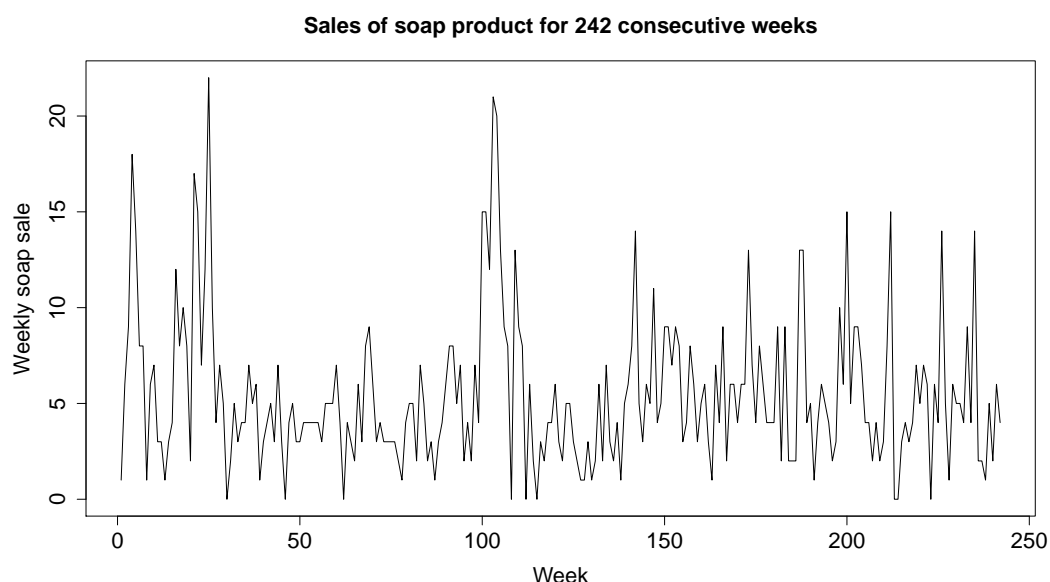


Figure 1: Weekly sales of a soap product in a supermarket for 242 consecutive weeks

From the plot it is seen that some distinct periods can be found in the data set. E.g. from week 25 to week 100 both the mean and variance seems to be smaller than for the whole data set. This could be a sign that a simple Poisson model isn't adequate to describe the data. Also the mean and variance for the whole data set is calculated, and gives

$$\bar{x} = 5.44 \quad s^2 = 15.40$$

Since the mean equals the variance in the Poisson distribution, the data is overdispersed relative to the Poisson distribution and this further shows that a simple Poisson model will not fit the data.

One way to model data that is overdispersed relative to the simple Poisson model, is to use an independent mixture of Poisson distributions, but plotting the ACF (see figure 2) of the data, shows that consecutive data points are positive correlated. By definition, this isn't true for an independent mixture of Poisson distributions and therefore this type of model isn't considered further.

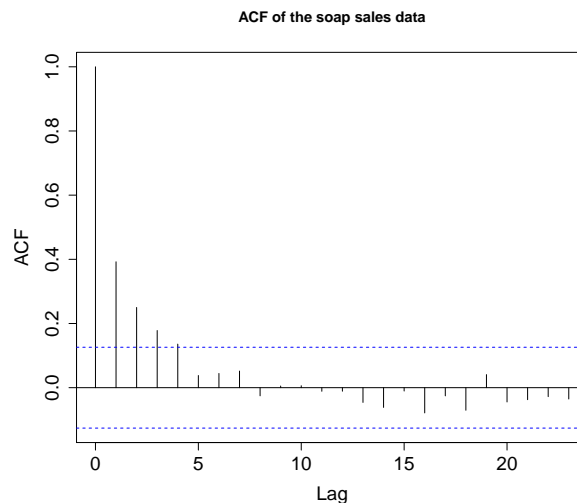


Figure 2: ACF plot of the soap data set

To handle both the overdispersed data and the correlation in the data a Poisson-HMM is instead fitted to the data.

Question b: Fitting Poisson-HMMs by direct maximization of the likelihood

In this section a 2-, 3-, and 4-state Poisson-HMM is fitted to the soap data set, by maximizing the likelihood function. A couple of details need to be handled to get a usable fit. The number of operations needed to calculate the likelihood needs to be held low. This is done by recursively calculating the forward probabilities. Also to avoid underflow problems the forward probabilities need to be scaled. And to be able to use an unconstrained optimization algorithm the parameters (the Poisson means and the transition probabilities) need to be transformed to some unconstrained working parameters. All these concerns are handled by the R-function `pois.HMM.mle` in appendix A.1.4 in [1] and therefore that function is used.

Fitting a 2-state Poisson-HMM

To fit the model initial values for the parameters should be chosen. Using the advice in section 3.4.2 in [1] a single value is chosen for the off-diagonal elements in the initial t.p.m. The initial mean values are chosen symmetrically around the sample mean. As mentioned in section 3.4.1 in [1] the likelihood function will frequently have multiple local maxima so different initial values* need to be tested to minimize the chance of choosing a local maxima that is far from the global maxima.

For the 2-state Poisson-HMM the different initial parameters all gave the same maxima and it is therefore likely that it is indeed the global maxima. Using the initial parameters

$$\mathbf{\Gamma}_0 = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} \quad \boldsymbol{\lambda}_0 = \begin{pmatrix} 2 & 8 \end{pmatrix}$$

gave the following parameter estimates

$$\mathbf{\Gamma} = \begin{pmatrix} 0.912 & 0.088 \\ 0.370 & 0.630 \end{pmatrix} \quad \boldsymbol{\lambda} = \begin{pmatrix} 4.022 & 11.371 \end{pmatrix}$$

From the estimate $\mathbf{\Gamma}$ the stationary distribution can then be calculated as

$$\boldsymbol{\delta} = \begin{pmatrix} 0.809 & 0.191 \end{pmatrix}$$

A plot of the marginal distribution of the 2-state Poisson model and a plot of the two estimated means are shown in figure 3

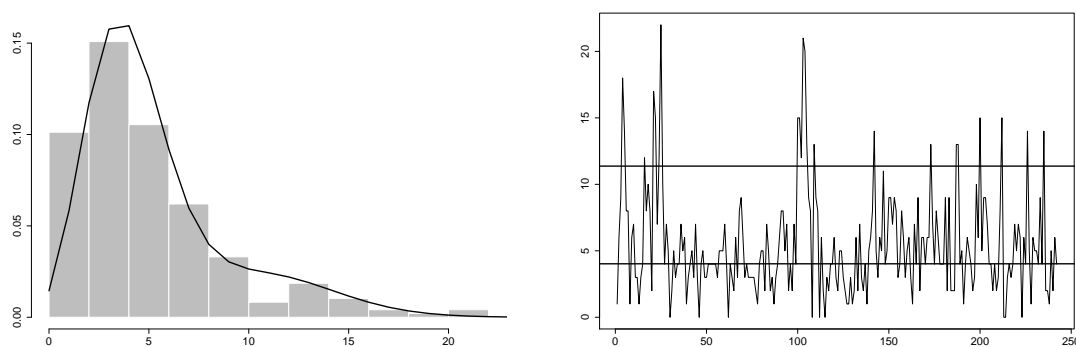


Figure 3: Left plot is a plot of the marginal distribution of the 2-state Poisson-HMM above the histogram for the soap data set. Right plot is the soap data set with the estimated means of the two state conditional distributions of the 2-state Poisson-HMM.

It is seen that the marginal distribution is matching the sample distribution rather well, and from the data plot with means included it is seen that state 1 corresponds to

*See appendix A.1 for the specific initial values tested

the “normal” sales level and that state 2 corresponds to the weeks with very high sale figures.

Fitting a 3-state Poisson-HMM

As with the 2-state Poisson-HMM different initial parameters[†] was used to fit a 3-state Poisson-HMM to the soap data set. Two different local maximas was found depending on the initial values of the λ parameter. For the lowest of the two maxima one of the state dependent means was close to 0 which gives a model that is almost a 2-state HMM. For the other maxima all state dependent means were different from zero, and this maxima was found by eg. using the intial parameters

$$\Gamma_0 = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{pmatrix} \quad \lambda_0 = (2 \quad 5 \quad 8)$$

Which gave the estimates

$$\Gamma = \begin{pmatrix} 0.864 & 0.117 & 0.019 \\ 0.445 & 0.538 & 0.017 \\ 0.000 & 0.298 & 0.702 \end{pmatrix} \quad \lambda = (3.736 \quad 8.443 \quad 14.927)$$

and the stationary distribution

$$\delta = (0.722 \quad 0.220 \quad 0.058)$$

It is seen that the process is only in state 3 in little more than 5% of the time. With a mean of $\lambda_3 = 14.927$ this state handles only the few weeks with exceptionally high sale figures. This is also seen in the data plot with state dependent means shown in figure 4.

[†]See appendix A.2 for the initial values that was tested

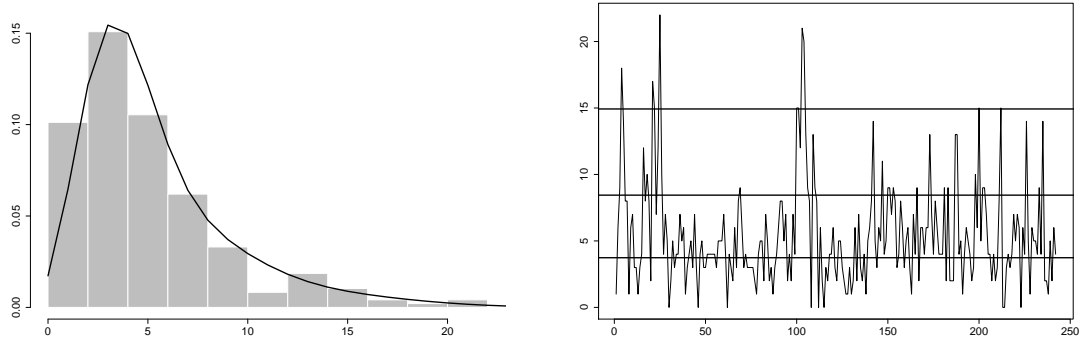


Figure 4: Left plot is a plot of the marginal distribution of the 3-state Poisson-HMM above the histogram for the soap data set. Right plot is the soap data set with the estimated means of the two state conditional distributions of the 3-state Poisson-HMM.

Fitting a 4-state Poisson-HMM

As for the 2- and 3-state models different initial values are used when fitting the 4-state model, but unlike the 2- and 3-state models, the maximas obtained for the 4-state model is quite sensitive to the initial parameters. This is seen in table 1 where 9 different initial values are used to fit a 4-state model.

$\gamma_{ij, i \neq j}$	λ_0	λ	$-\ell$
0.01	$\begin{pmatrix} 1 & 4 & 6 & 9 \end{pmatrix}$	$\begin{pmatrix} 0.008 & 8.226 & 3.886 & 14.359 \end{pmatrix}$	604.26
0.05	$\begin{pmatrix} 1 & 4 & 6 & 9 \end{pmatrix}$	$\begin{pmatrix} 0.000 & 3.873 & 8.197 & 14.690 \end{pmatrix}$	604.19
0.10	$\begin{pmatrix} 1 & 4 & 6 & 9 \end{pmatrix}$	$\begin{pmatrix} 0.000 & 8.197 & 3.873 & 14.690 \end{pmatrix}$	604.19
0.01	$\begin{pmatrix} 2 & 4 & 6 & 8 \end{pmatrix}$	$\begin{pmatrix} 3.063 & 3.995 & 6.069 & 12.034 \end{pmatrix}$	612.28
0.05	$\begin{pmatrix} 2 & 4 & 6 & 8 \end{pmatrix}$	$\begin{pmatrix} 4.243 & 3.490 & 8.952 & 14.788 \end{pmatrix}$	608.80
0.10	$\begin{pmatrix} 2 & 4 & 6 & 8 \end{pmatrix}$	$\begin{pmatrix} 3.911 & 0.415 & 8.192 & 14.627 \end{pmatrix}$	604.49
0.01	$\begin{pmatrix} 1 & 4 & 8 & 12 \end{pmatrix}$	$\begin{pmatrix} 0.019 & 7.594 & 3.789 & 13.188 \end{pmatrix}$	605.68
0.05	$\begin{pmatrix} 1 & 4 & 8 & 12 \end{pmatrix}$	$\begin{pmatrix} 0.000 & 3.873 & 8.197 & 14.690 \end{pmatrix}$	604.19
0.10	$\begin{pmatrix} 1 & 4 & 8 & 12 \end{pmatrix}$	$\begin{pmatrix} 0.000 & 3.873 & 8.197 & 14.690 \end{pmatrix}$	604.19

Table 1: Maximum likelihood for 9 different starting values for the 4-state Poisson-HMM

As with the 3-state model some of the found maximas have $\lambda_1 < 10^{-10}$ which effectively makes it a 3-state model. Ignoring the maximas with λ_1 very small the model is found with the initial parameters

$$\mathbf{\Gamma}_0 = \begin{pmatrix} 0.97 & 0.01 & 0.01 & 0.01 \\ 0.01 & 0.97 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.97 & 0.01 \\ 0.01 & 0.01 & 0.01 & 0.97 \end{pmatrix} \quad \boldsymbol{\lambda}_0 = (1 \quad 4 \quad 6 \quad 9)$$

The estimates found are

$$\mathbf{\Gamma} = \begin{pmatrix} 0.173 & 0.000 & 0.640 & 0.187 \\ 0.082 & 0.580 & 0.338 & 0.000 \\ 0.000 & 0.092 & 0.882 & 0.025 \\ 0.002 & 0.317 & 0.000 & 0.682 \end{pmatrix} \quad \boldsymbol{\lambda} = (0.008 \quad 8.226 \quad 3.886 \quad 14.359)$$

and it is seen 4 of 16 elements in $\mathbf{\Gamma}$ is smaller than $5 \cdot 10^{-4}$ which isn't surprising as explained on page 52 in [1]. Also $\lambda_1 = 0.008$ is mathematically different from 0, but statistically it is probably not significantly different from 0[‡]. Ignoring state 1 it is seen that the means of state 2,3,4 matches the means of state 1,2,3 in the 3-state model. From $\mathbf{\Gamma}$ the stationary distribution is found as

$$\boldsymbol{\delta} = (0.021 \quad 0.207 \quad 0.704 \quad 0.068)$$

Which further confirms that state 1 is handling special cases. Plots of the marginal distribution and the state dependent means are shown in figure 5.

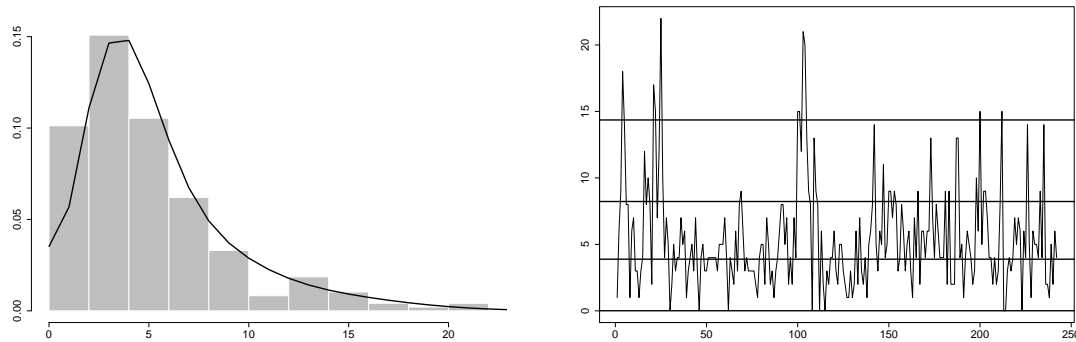


Figure 5: Left plot is a plot of the marginal distribution of the 4-state Poisson-HMM above the histogram for the soap data set. Right plot is the soap data set with the estimated means of the two state conditional distributions of the 4-state Poisson-HMM.

The marginal distribution of the 4-state model is very much like the marginal distribution of the 3-state model and the question is whether the extra parameters in the 4-state model are needed

[‡]No error estimates of the parameters are calculated. See the section “Remarks”

Model selection

To choose between the 2-, 3- and 4-state model the AIC and the BIC of the models can be compared. Furthermore the mean, variance and correlation of the models can be compared with the sample mean, variance and correlation. To calculate the mean, variance and correlations for the models, the function `pois.HMM.moments` from exercise 2.9, is used.

	μ	σ^2	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7
Sample	5.442	15.401	0.392	0.250	0.178	0.136	0.038	0.044	0.052
2-states	5.429	13.784	0.329	0.178	0.097	0.052	0.028	0.015	0.008
3-states	5.421	14.721	0.407	0.268	0.178	0.120	0.081	0.055	0.037
4-states	5.418	14.779	0.380	0.241	0.157	0.103	0.067	0.044	0.029

Table 2: Mean, variance and correlation coefficients for the data and the three models. Note that $\rho_k = \text{Corr}(X_t, X_{t+k})$

	AIC	BIC
2-states	1245.337	1259.292
3-states	1239.043	1270.444
4-states	1240.528	1296.351

Table 3: AIC and BIC scores for the three models

From the results in table 2 it is seen that all three model means are close to the sample mean. The variance of the 2-state model is lower than the sample variance and the correlation coefficients of the 2-state model are all lower than the sample correlation coefficients. On the other hand both the 3- and 4-state models seems to match both the variance and correlation coefficients of the sample pretty well. To choose between the 3- and 4-state model the AIC and BIC score are therefore used.

From table 3 it is seen that the 3-state model scores better than the 4-state model in both AIC and BIC score. This indicates that a 4-state model might overfit the data. It is worth noting that the 2-state model has the lowest BIC of the three models. This is due to the fact that the BIC penalizes models with many parameters more[§] than the AIC score does. The low variance and correlation coefficients of the 2-state model, compared with the sample, rules out the 2-state model though and the final choice of model is the 3-state model.

Question c: Fitting Poisson-HMMs using the EM algorithm

In this section 2-, 3- and 4-state Poisson-HMM models are fitted to the soap data set using the EM algorithm. The markov chain is assumed to not be stationary. As mentioned in section 4.2.5 in [1] this makes the implementation of the EM algorithm easier.

[§]Whenever the number of observations is larger than $e^2 \approx 7$

The distribution δ_0 of the first state then needs to be determined; either by including the distribution in the parameters to estimate or by maximizing the likelihood conditioned on the Markov chain starting in a particular state (δ_0 being a unitvector). As mentioned in section 4.2.4 the easiest approach is to maximize the likelihood conditioned on the Markov chain starting in a particular state, but since the implementation in appendix A.2.3 in [1] estimates δ_0 that is the approach taken in this assignment. As in the examples in section 4.3 in [1] the starting values for δ_0 is set as the uniform distribution.

Fitting a 2-state Poisson-HMM

Using the same 9 initial parameter values as in the direct maximization of the likelihood the 2-state Poisson-HMM parameters was estimated by the EM algorithm. All 9 starting points gave the same maxima, so it is likely to be a global maxima. The estimated parameters was found as

$$\mathbf{\Gamma} = \begin{pmatrix} 0.911 & 0.089 \\ 0.367 & 0.633 \end{pmatrix} \quad \boldsymbol{\lambda} = \begin{pmatrix} 4.019 & 11.354 \end{pmatrix}$$

$$\boldsymbol{\delta} = \begin{pmatrix} 1.000 & 0.000 \end{pmatrix}$$

A Appendices

All R code used for this assignment is included here. All source code incl. latex code for this report can be found at <https://github.com/alphabits/dtu-spring-2012/tree/master/02433/assignment-1>

A.1 Fitting results for 2-state Poisson-HMM

$\gamma_{ij,i \neq j}$	λ_0	λ	$-\ell$
0.01	$\begin{pmatrix} 2 & 8 \end{pmatrix}$	$\begin{pmatrix} 4.022 & 11.371 \end{pmatrix}$	618.67
0.05	$\begin{pmatrix} 2 & 8 \end{pmatrix}$	$\begin{pmatrix} 4.022 & 11.371 \end{pmatrix}$	618.67
0.10	$\begin{pmatrix} 2 & 8 \end{pmatrix}$	$\begin{pmatrix} 4.022 & 11.371 \end{pmatrix}$	618.67
0.01	$\begin{pmatrix} 3 & 7 \end{pmatrix}$	$\begin{pmatrix} 4.022 & 11.371 \end{pmatrix}$	618.67
0.05	$\begin{pmatrix} 3 & 7 \end{pmatrix}$	$\begin{pmatrix} 4.022 & 11.371 \end{pmatrix}$	618.67
0.10	$\begin{pmatrix} 3 & 7 \end{pmatrix}$	$\begin{pmatrix} 4.022 & 11.371 \end{pmatrix}$	618.67
0.01	$\begin{pmatrix} 4 & 6 \end{pmatrix}$	$\begin{pmatrix} 4.022 & 11.371 \end{pmatrix}$	618.67
0.05	$\begin{pmatrix} 4 & 6 \end{pmatrix}$	$\begin{pmatrix} 4.022 & 11.371 \end{pmatrix}$	618.67
0.10	$\begin{pmatrix} 4 & 6 \end{pmatrix}$	$\begin{pmatrix} 4.022 & 11.371 \end{pmatrix}$	618.67

Table 4: Maximum likelihood for 9 different starting values for the 2-state Poisson-HMM

A.2 Fitting results for 3-state Poisson-HMM

$\gamma_{ij,i \neq j}$	λ_0	λ	$-\ell$
0.01	$\begin{pmatrix} 1 & 5 & 9 \end{pmatrix}$	$\begin{pmatrix} 0.000 & 4.134 & 11.442 \end{pmatrix}$	612.14
0.05	$\begin{pmatrix} 1 & 5 & 9 \end{pmatrix}$	$\begin{pmatrix} 0.000 & 4.134 & 11.442 \end{pmatrix}$	612.14
0.10	$\begin{pmatrix} 1 & 5 & 9 \end{pmatrix}$	$\begin{pmatrix} 0.000 & 4.134 & 11.442 \end{pmatrix}$	612.14
0.01	$\begin{pmatrix} 2 & 5 & 8 \end{pmatrix}$	$\begin{pmatrix} 3.736 & 8.443 & 14.927 \end{pmatrix}$	610.52
0.05	$\begin{pmatrix} 2 & 5 & 8 \end{pmatrix}$	$\begin{pmatrix} 3.736 & 8.443 & 14.927 \end{pmatrix}$	610.52
0.10	$\begin{pmatrix} 2 & 5 & 8 \end{pmatrix}$	$\begin{pmatrix} 3.736 & 8.443 & 14.927 \end{pmatrix}$	610.52
0.01	$\begin{pmatrix} 3 & 5 & 7 \end{pmatrix}$	$\begin{pmatrix} 8.443 & 3.736 & 14.927 \end{pmatrix}$	610.52
0.05	$\begin{pmatrix} 3 & 5 & 7 \end{pmatrix}$	$\begin{pmatrix} 3.736 & 8.443 & 14.927 \end{pmatrix}$	610.52
0.10	$\begin{pmatrix} 3 & 5 & 7 \end{pmatrix}$	$\begin{pmatrix} 3.736 & 8.443 & 14.927 \end{pmatrix}$	610.52

Table 5: Maximum likelihood for 9 different starting values for the 3-state Poisson-HMM

References

- [1] Walter Zucchini and Iain L. MacDonald. *Hidden Markov Models for Time Series*. Chapman & Hall/CRC, 1st edition, 2009.