Fitting Poisson HMMs

Simone

21/05/2021

Generate Data from a 3-State Poisson HMM

$$\boldsymbol{\lambda} = (12,\ 19,\ 29), \quad \boldsymbol{\Gamma} = \begin{pmatrix} 0.955 & 0.024 & 0.021 \\ 0.050 & 0.899 & 0.051 \\ 0.007 & 0.190 & 0.803 \end{pmatrix}, \quad \boldsymbol{\delta} = (0.4436,\ 0.4045,\ 0.1519)$$

```
T=1000
pois3s <- list(m = 3,
               lambda = c(12, 19, 29),
               gamma = matrix(c(0.955, 0.024, 0.021,
                                0.050, 0.899, 0.051,
                                0.007, 0.190, 0.803), nrow=3, ncol=3, byrow = TRUE),
               delta = c(0.4436, 0.4045, 0.1519))
sample <- pois.HMM.generate_sample(T, pois3s)
poisdata = data.frame(Time = c(1:T),
                      Observation = sample$observ,
                      State = factor(sample$state))
x <- poisdata$Observation
```
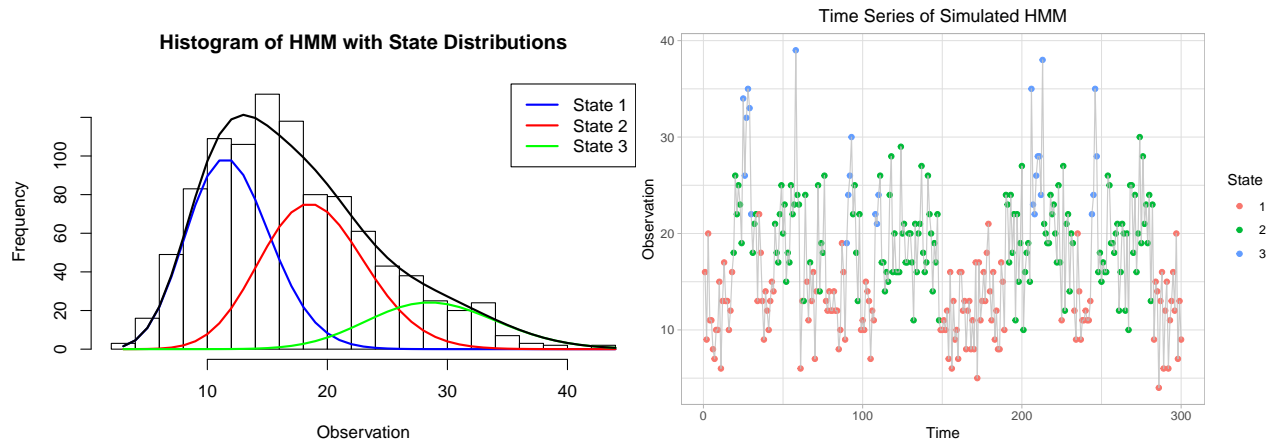

Figure 1: LEFT: Histogram of observations with state dependent poisson distributions overlayed. Black line is the marginal distribution. RIGHT: Timeseries of first 300 observations with colors corresponding hidden states

Fit the Generated Data with a 3-State Poisson HMM

To start, I summarize the data so that I can pick reasonable lambdas.

```r
summary = summary(x)
Q1 = as.numeric(summary[2])
med = as.numeric(summary[3])
Q2 = as.numeric(summary[5])

m       <-3
lambda0 <-c(Q1,med,Q2)
gamma0  <-matrix(c(0.9 ,0.05 ,0.05,
                   0.05 ,0.9 ,0.05,
                   0.05 ,0.05 ,0.9),m,m,byrow=TRUE)
summary
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3.00   12.00   17.00   17.57   22.00   44.00
```

I choose $\boldsymbol{\lambda}_0$ according to the 1st quartile, median, and 3rd quartile. I choose $\boldsymbol{\Gamma}_0$ so that the diagonals 0.9 and the off-diagonals are 0.5 (common practice in book).

```r
modpois3s <-pois.HMM.mle(x, m, lambda0, gamma0, stationary=TRUE)
modpois3s
```

```
## $m
## [1] 3
##
## $lambda
## [1] 12.08966 19.17823 28.40414
##
## $gamma
##            [,1]       [,2]       [,3]
## [1,] 0.94124636 0.05202894 0.00672470
## [2,] 0.04914964 0.88607189 0.06477847
## [3,] 0.03220076 0.15410532 0.81369392
##
## $delta
## [1] 0.4306115 0.4109564 0.1584321
##
## $code
## [1] 1
##
## $mllk
## [1] 3029.499
##
## $AIC
## [1] 6076.997
##
## $BIC
## [1] 6121.167
```

Now I find the underlying state sequence using global decoding by the Viterbi algorithm

```
state_seq <- pois.HMM.viterbi(x, modpois3s)
poisdata3s <-poisdata
poisdata3s$GuessState <- as.factor(state_seq)
```
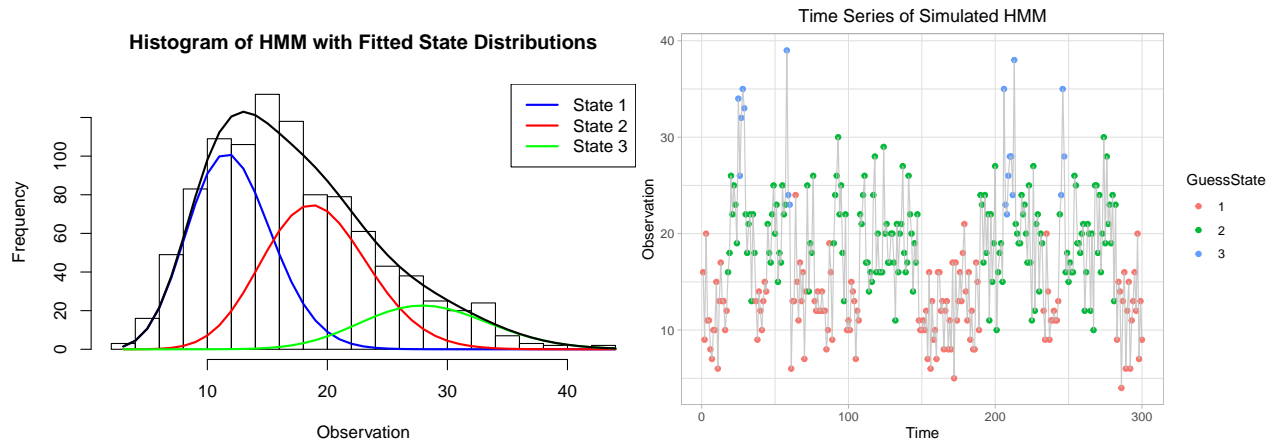

Figure 2: LEFT: Histogram of observations with fitted state dependent poisson distributions overlayed. Black line is the marginal distribution. RIGHT: Timeseries of first 300 observations with colors corresponding fitted hidden states decoded by Viterbi algorithm.
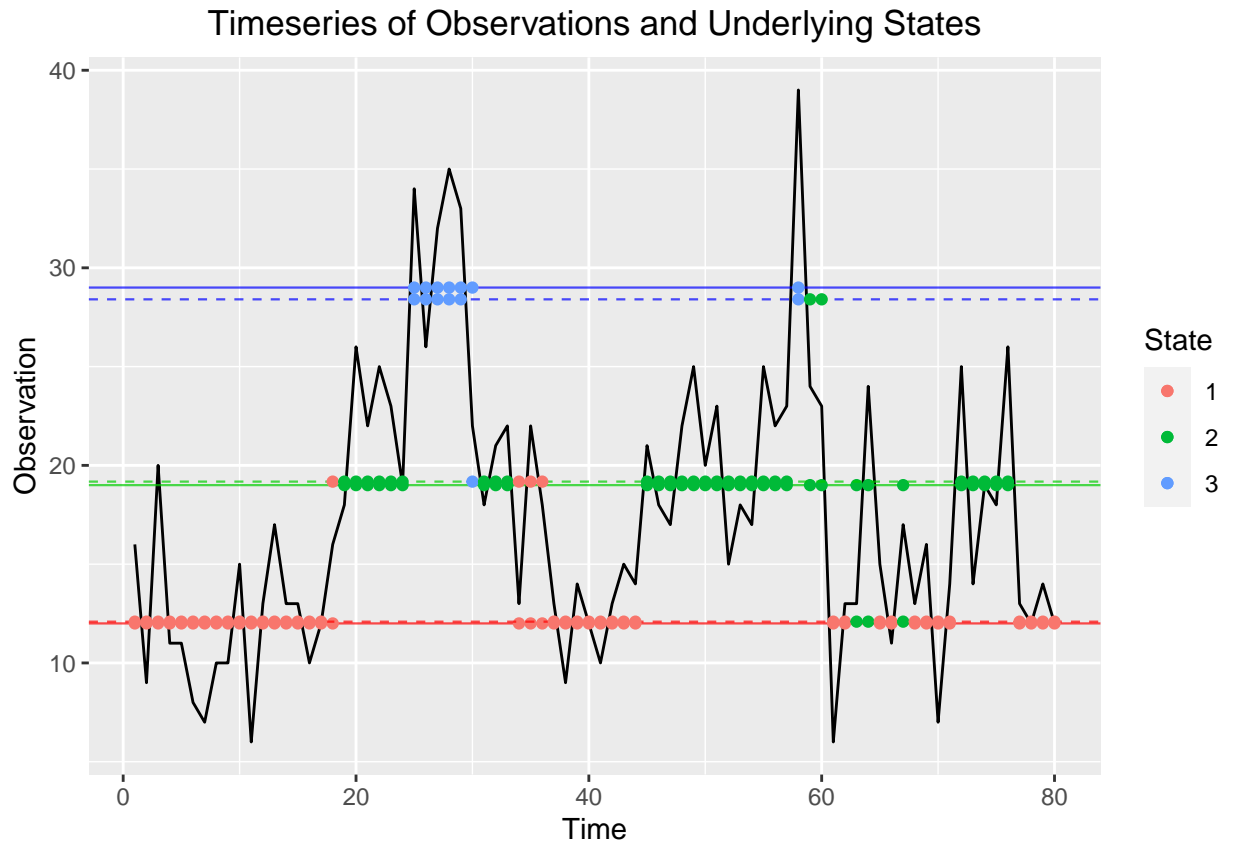
Figure 3: Timeseries of first 80 observations. The solid horizontal lines are the true means (lambda) of each state dependent poisson distribution and the dashed lines are the fitted means. For each time interval there are two points, one falling on the true state mean line and one on the fitted state mean line. The colors of the points correspond to the true states of the observation at that time.
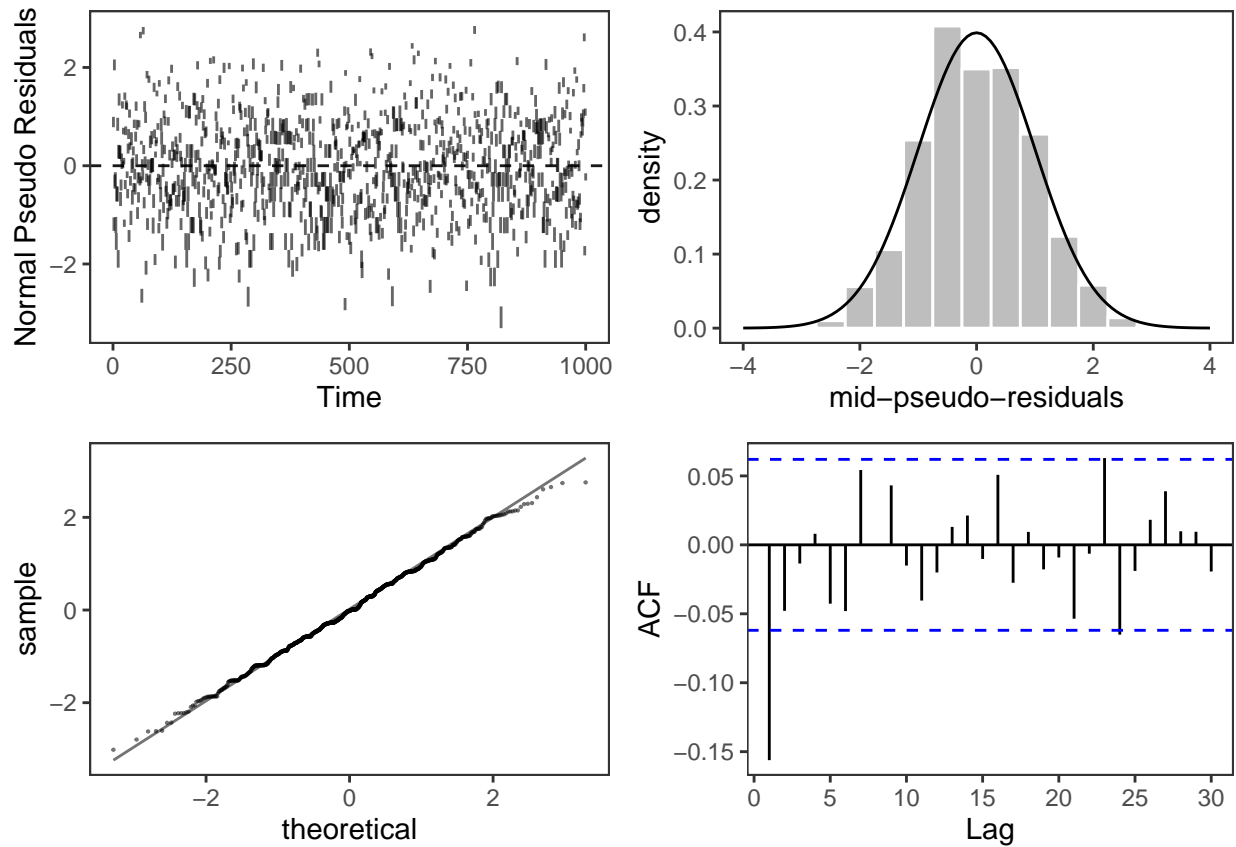
Figure 4: Plots for the normal pseudo-residuals of the fitted 3-state poisson HMM. TOP LEFT: Index plot of the normal pseudo residuals. TOP RIGHT: Histogram of the normal pseudo-residuals with N(0,1) overlayed. BOTTOM LEFT: Quantile–quantile plot of the normal pseudo-residuals. BOTTOM RIGHT: Autocorrelation function of the normal pseudo-residuals.

Fit the Generated Data with a 2-State Poisson HMM

Summarize the data so that I can pick reasonable lambdas.

```
Q1 = as.numeric(summary[2])
Q2 = as.numeric(summary[5])

m       <-2
lambda0 <-c(Q1,Q2)
gamma0  <-matrix(c(0.9 ,0.1,
                   0.1 ,0.9),m,m,byrow=TRUE)
summary
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3.00   12.00   17.00   17.57   22.00   44.00
```

I choose $\boldsymbol{\lambda}_0$ according to the 1st quartile, and 3rd quartile. I choose $\boldsymbol{\Gamma}_0$ so that the diagonals 0.9 and the off-diagonals are 0.1 (common practice in book).

```
modpois2s <-pois.HMM.mle(x, m, lambda0, gamma0, stationary=TRUE)
modpois2s
```

```
## $m
## [1] 2
##
## $lambda
## [1] 12.37481 22.30013
##
## $gamma
##              [,1]        [,2]
## [1,] 0.9330938 0.06690623
## [2,] 0.0608973 0.93910270
##
## $delta
## [1] 0.4764915 0.5235085
##
## $code
## [1] 1
##
## $mllk
## [1] 3127.282
##
## $AIC
## [1] 6262.565
##
## $BIC
## [1] 6282.196
```

Now I find the underlying state sequence using global decoding by the Viterbi algorithm

```
state_seq <- pois.HMM.viterbi(x, modpois2s)
poisdata2s <- poisdata
poisdata2s$GuessState <- as.factor(state_seq)
```
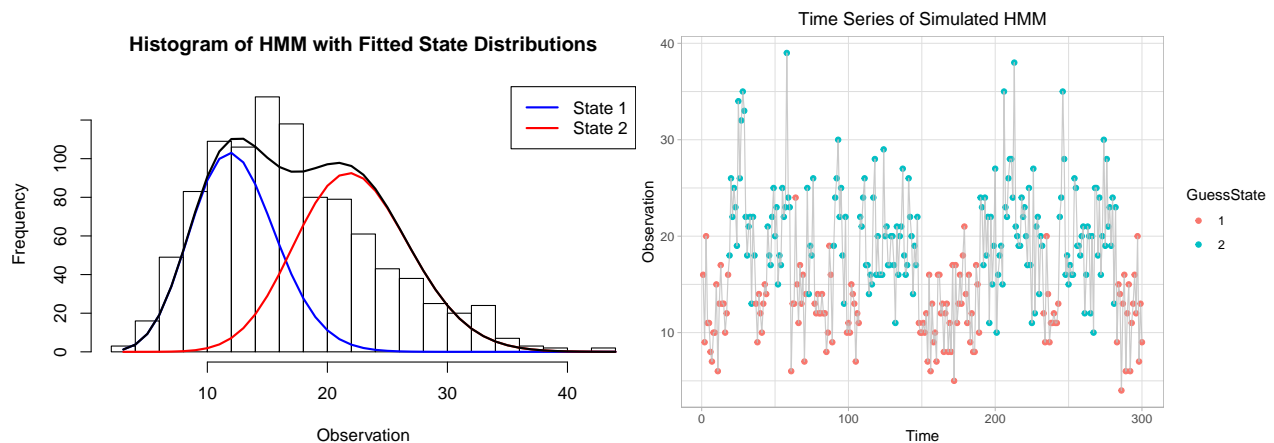

Figure 5: LEFT: Histogram of observations with fitted state dependent poisson distributions overlayed. Black line is the marginal distribution. RIGHT: Timeseries of first 300 observations with colors corresponding fitted hidden states decoded by Viterbi algorithm.
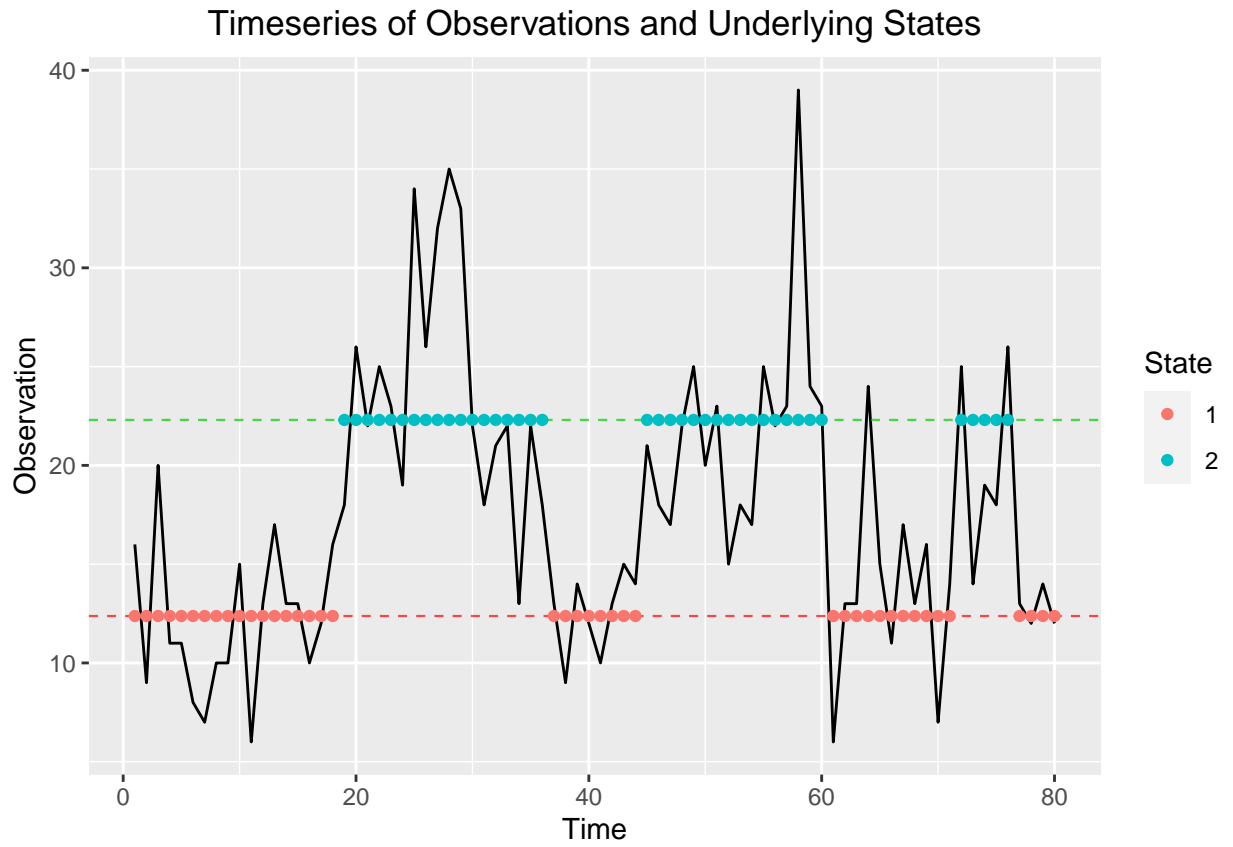
6

Figure 6: Timeseries of first 80 observations. The dashed lines are the fitted means (lambda) for each state dependent poisson distribution. The colors of the points correspond to the fitted state of the observation at that time.
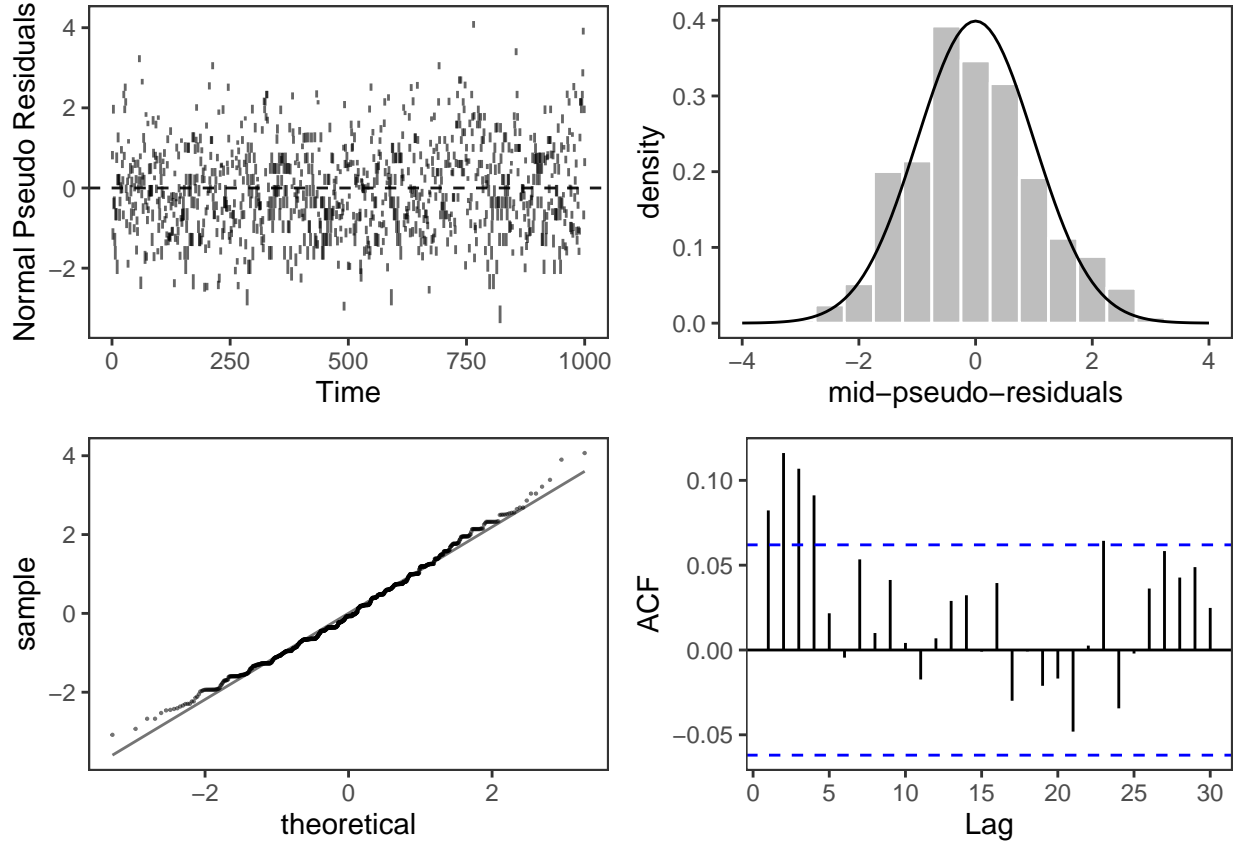
Figure 7: Plots for the normal pseudo-residuals of the fitted 2-state poisson HMM. TOP LEFT: Index plot of the normal pseudo residuals. TOP RIGHT: Histogram of the normal pseudo-residuals with N(0,1) overlayed. BOTTOM LEFT: Quantile–quantile plot of the normal pseudo-residuals. BOTTOM RIGHT: Autocorrelation function of the normal pseudo-residuals.

Fit the Generated Data with a 4-State Poisson HMM

Summarize the data so that I can pick reasonable lambdas.

```
Q1 = as.numeric(summary[2])-2
med = as.numeric(summary[3])-2
Q2 = as.numeric(summary[5])-2
w = as.numeric(summary[5])+4


m        <-4
lambda0 <-c(Q1,med,Q2,w)
gamma0  <-matrix(c(0.91 ,0.03 ,0.03, 0.03,
                   0.03 ,0.91 ,0.03, 0.03,
                   0.03 ,0.03, 0.03 ,0.91),m,m,byrow=TRUE)
summary
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3.00   12.00   17.00   17.57   22.00   44.00
```

I choose $\boldsymbol{\lambda}_0$ based on the 1st quartile, median, and 3rd quartile with some variation. I choose $\boldsymbol{\Gamma}_0$ so that the diagonals 0.91 and the off-diagonals are 0.03.

```
modpois4s <-pois.HMM.mle(x, m, lambda0, gamma0, stationary=TRUE)
modpois4s
```

```
## $m
## [1] 4
##
## $lambda
## [1] 12.08554 19.22702 29.57468 27.57881
##
## $gamma
##               [,1]         [,2]         [,3]         [,4]
## [1,] 9.409693e-01 5.262015e-02 6.409973e-03 5.845064e-07
## [2,] 5.121868e-02 8.919983e-01 5.678243e-02 6.330263e-07
## [3,] 6.619869e-11 3.792881e-12 3.158800e-08 1.000000e+00
## [4,] 4.858451e-02 2.607815e-01 5.095653e-01 1.810686e-01
##
## $delta
## [1] 0.43016347 0.41505426 0.06968682 0.08509545
##
## $code
## [1] 1
##
## $mllk
## [1] 3028.118
##
## $AIC
## [1] 6088.237
##
## $BIC
## [1] 6166.761
```

Now I find the underlying state sequence using global decoding by the Viterbi algorithm

```
state_seq <- pois.HMM.viterbi(x, modpois4s)
poisdata4s <- poisdata
poisdata4s$GuessState <- as.factor(state_seq)
```
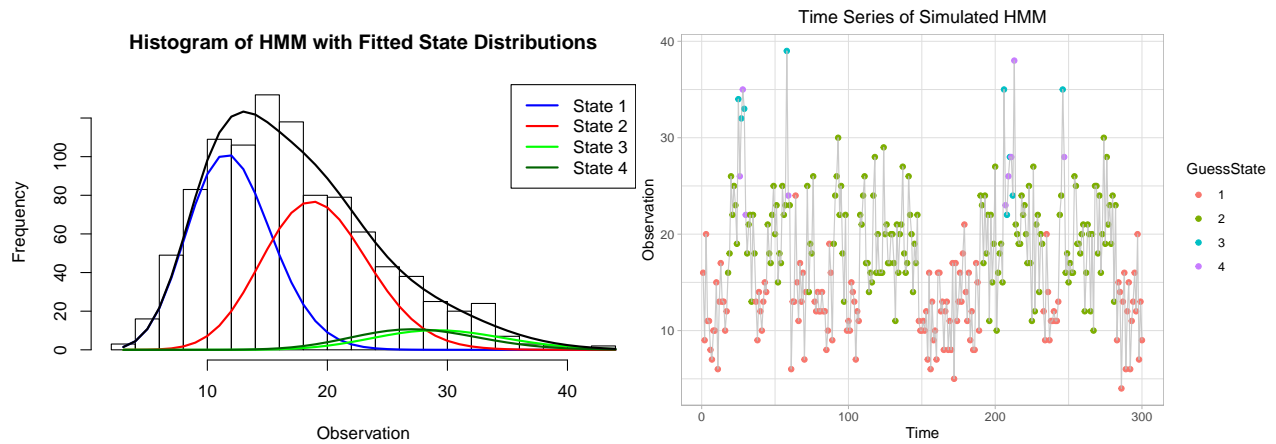
Figure 8: LEFT: Histogram of observations with fitted state dependent poisson distributions overlayed. Black line is the marginal distribution. RIGHT: Timeseries of first 300 observations with colors corresponding fitted hidden states decoded by Viterbi algorithm.

We can visualize the state sequence and observations the same way as before (without the true states)
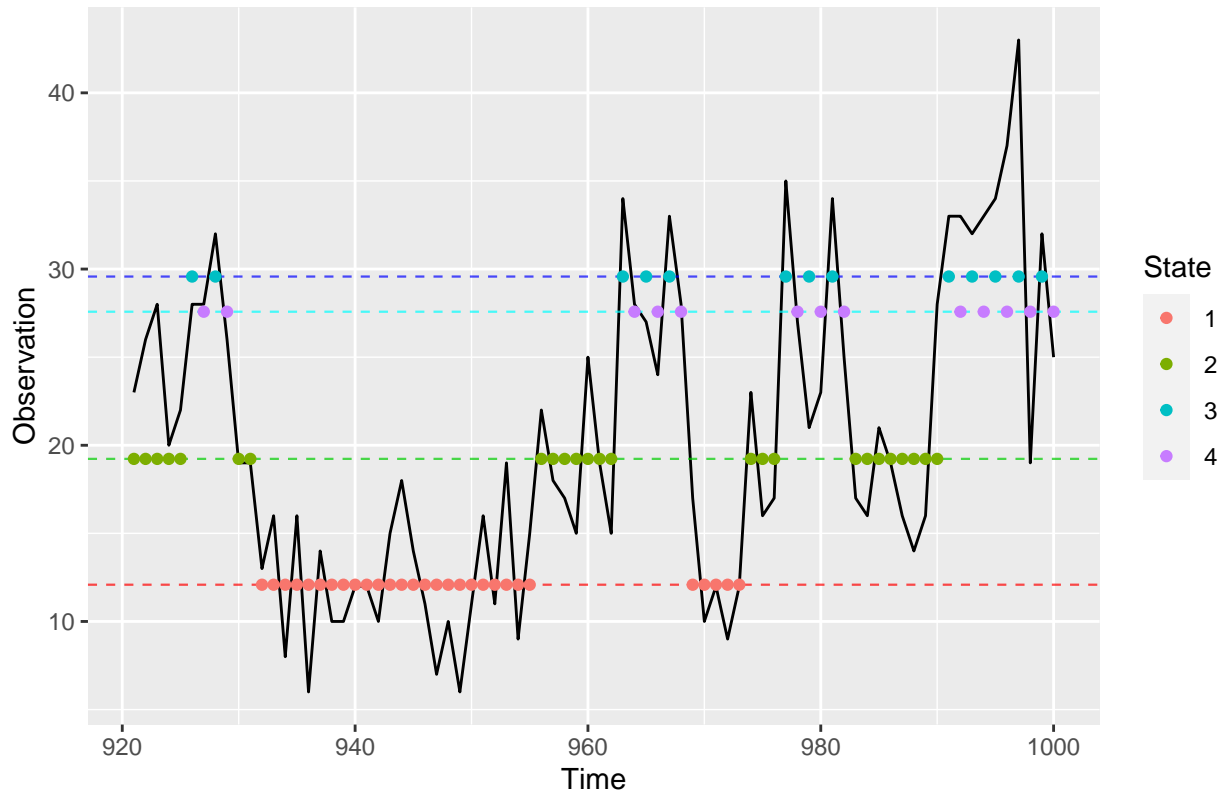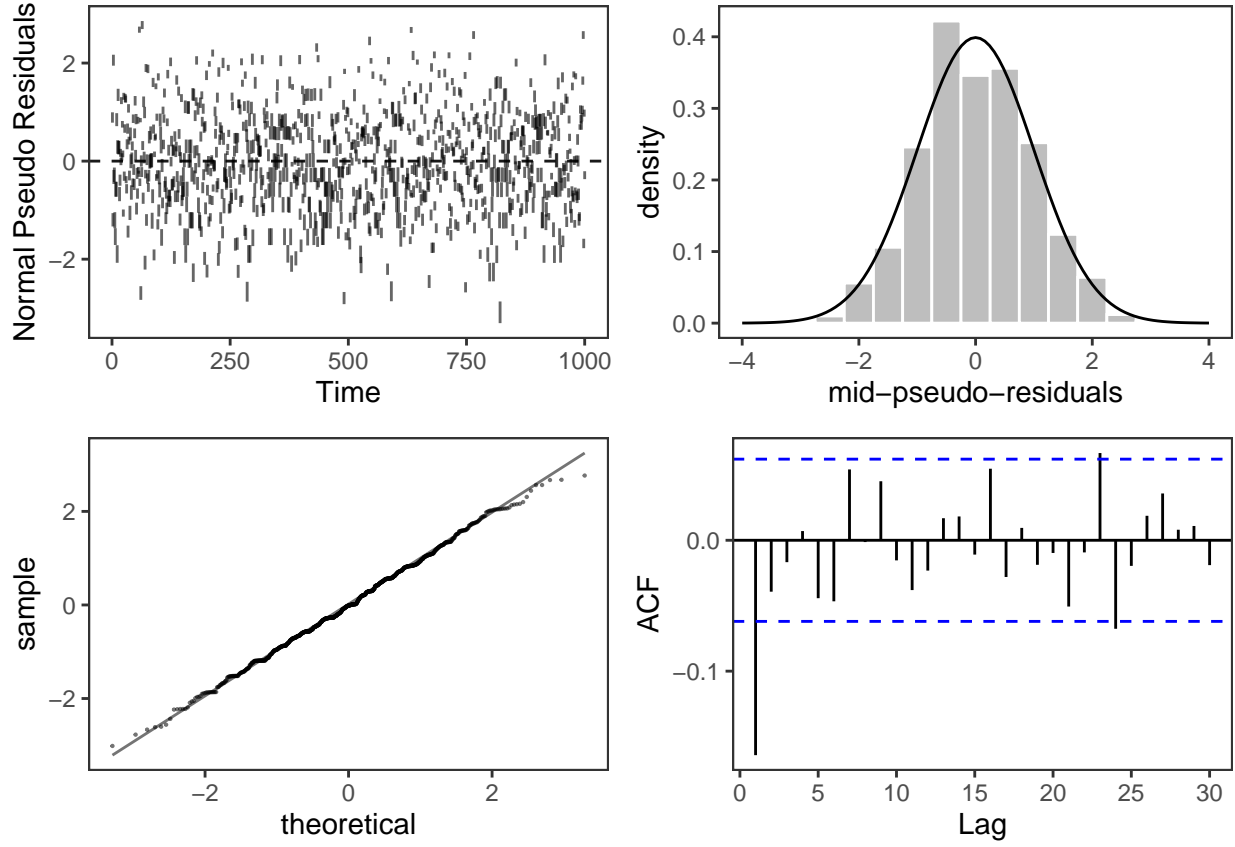
Figure 9: Plots for the normal pseudo-residuals of the fitted 4-state poisson HMM. TOP LEFT: Index plot of the normal pseudo residuals. TOP RIGHT: Histogram of the normal pseudo-residuals with N(0,1) overlayed. BOTTOM LEFT: Quantile–quantile plot of the normal pseudo-residuals. BOTTOM RIGHT: Autocorrelation function of the normal pseudo-residuals.
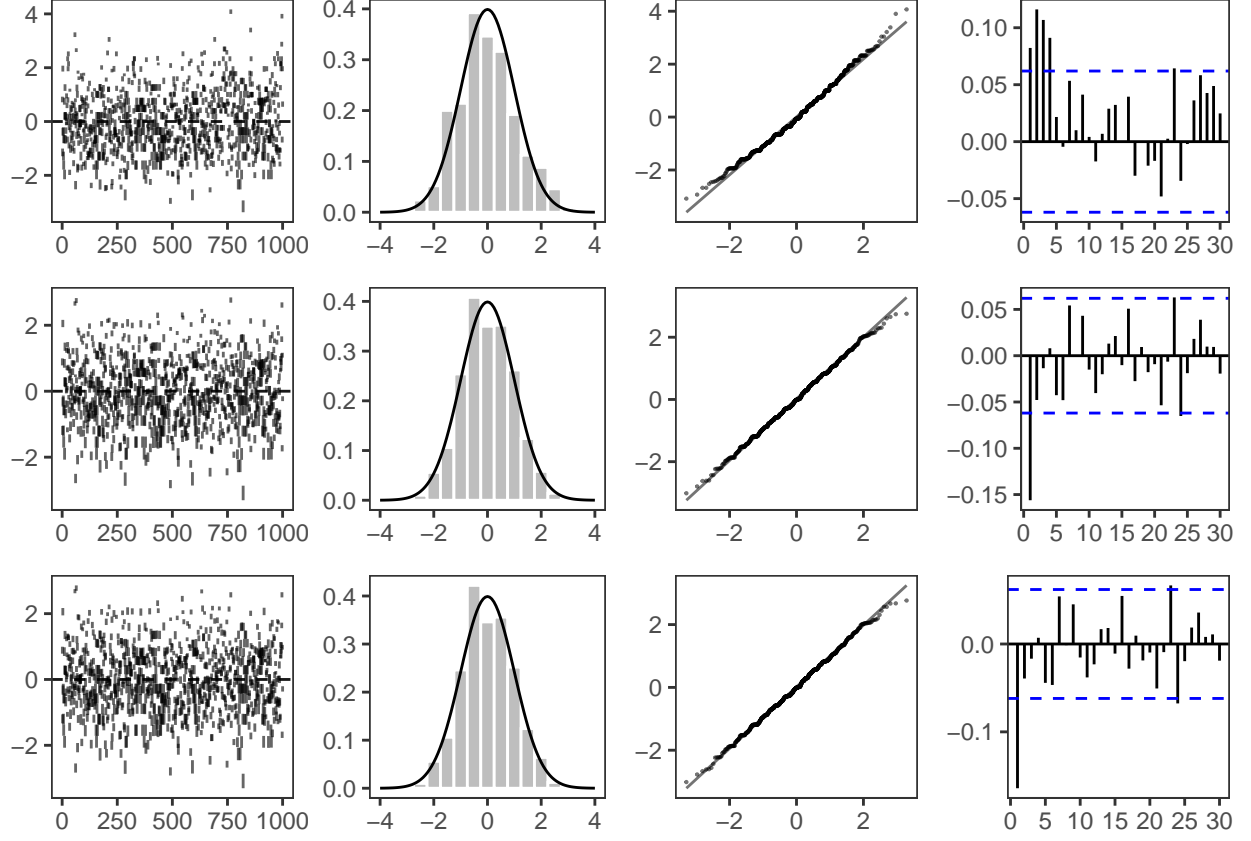
Figure 10: Each row shows plots of the pseudo residuals for different fitted HMMs. The first row is for a 2-state HMM with poisson state dependent distributions. The second is a 3-State poisson distributed HMM and the third is a 4-state poisson distributed HMM. The first column shows index plots of the normal pseudo-residuals. The second column shows the histograms of the normal pseudo-residuals with N(0,1) overlayed. The third column shows the quantile–quantile plots of the normal pseudo-residuals. The fourth column shows the autocorrelation functions of the normal pseudo-residuals