

# Fitting Poisson HMMs

Simone

21/05/2021

## Generate Data from a 3-State Poisson HMM

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

```
T=10000
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(0.4583296, 0.3914683, 0.1502021))
sample <- pois.HMM.generate_sample(T, pois3s)
poisdata = data.frame(Time = c(1:T),
                      Observation = sample$observ,
                      State = factor(sample$state))
x <- poisdata$Observation
```

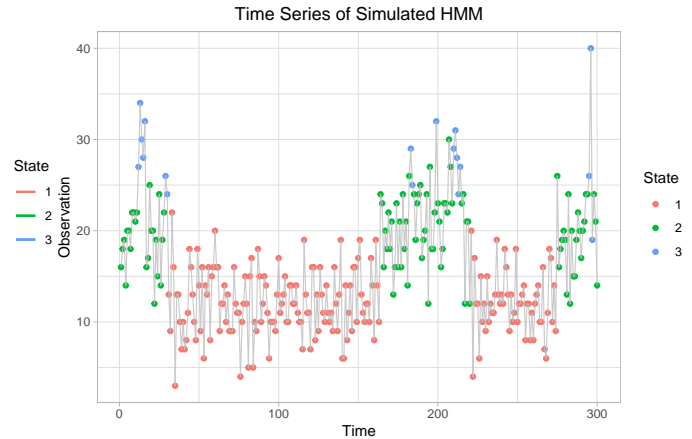
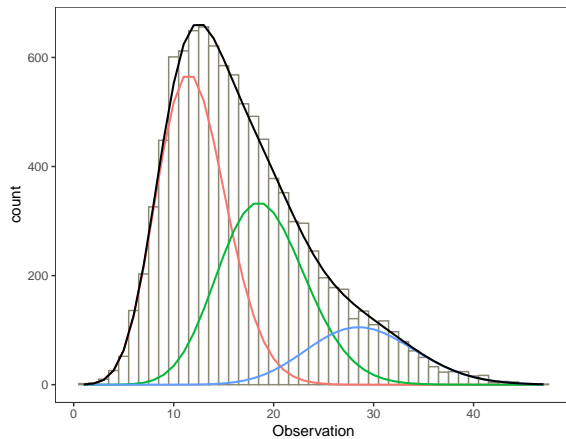


Figure 1: LEFT: Histogram of observations with state dependent poisson distributions overlaid. Black line is the marginal distribution. RIGHT: Timeseries of first 300 observations with colors corresponding hidden states

## Fit the Generated Data with a 3-State Poisson HMM

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

```
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.
##      1.00   12.00   16.00   16.93   21.00   47.00
```

I choose  $\lambda_0$  according to the 1st quartile, median, and 3rd quartile. I choose  $\Gamma_0$  so that the diagonals 0.9 and the off-diagonals are 0.05 (common practice in book).

```
modpois3s <-pois.HMM.mle(x, m, lambda0, gamma0, c(1/3,1/3,1/3), stationary=FALSE)
modpois3s
```

```
## $m
## [1] 3
##
## $lambda
## [1] 11.95563 19.00236 29.01963
##
## $gamma
##           [,1]      [,2]      [,3]
## [1,] 0.95849617 0.02152453 0.01997931
## [2,] 0.05023318 0.90625454 0.04351228
## [3,] 0.01228184 0.17529989 0.81241828
##
## $delta
## [1] 0.4903408 0.3713038 0.1383554
##
## $code
## [1] 1
##
## $mllk
## [1] 30011.18
##
## $AIC
## [1] 60044.36
##
## $BIC
## [1] 60123.68
```

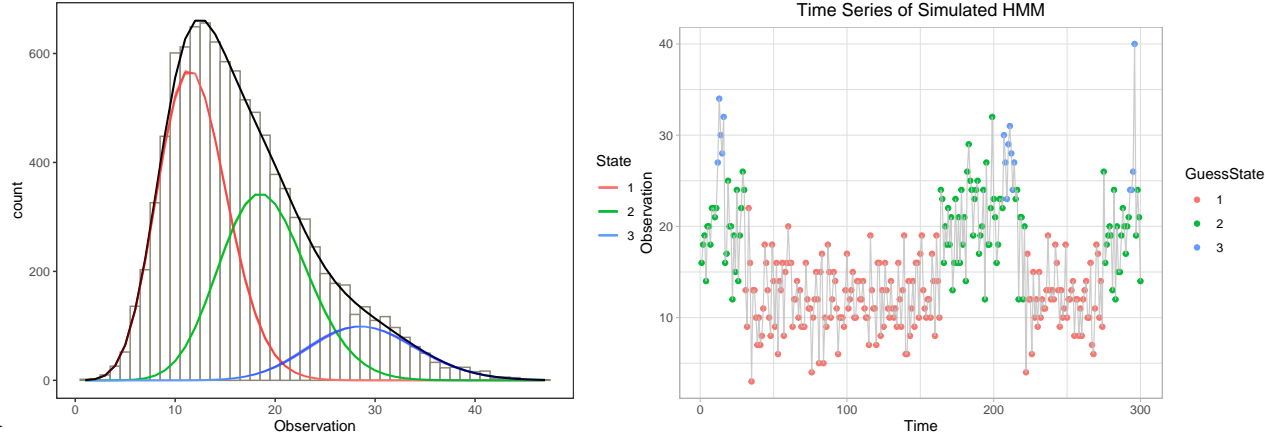
Get confidence intervals for fitted parameters

```
CI3 = pois.HMM.CI(x, 20, modpois3s, stationary = TRUE)
CI3
```

```
## $lambda
## [1] 11.95563 19.00236 29.01963
##
## $lambda.upper.conf
## [1] 12.07267 19.14102 29.35079
##
## $lambda.lower.conf
## [1] 11.83860 18.86370 28.68847
##
## $gamma
##           [,1]      [,2]      [,3]
## [1,] 0.95849617 0.02152453 0.01997931
## [2,] 0.05023318 0.90625454 0.04351228
## [3,] 0.01228184 0.17529989 0.81241828
##
## $gamma.upper.conf
##           [,1]      [,2]      [,3]
## [1,] 0.96452822 0.02647039 0.02631424
## [2,] 0.05851771 0.91771778 0.05167342
## [3,] 0.01944644 0.19409051 0.83014240
##
## $gamma.lower.conf
##           [,1]      [,2]      [,3]
## [1,] 0.952464112 0.01657867 0.01364437
## [2,] 0.041948641 0.89479131 0.03535114
## [3,] 0.005117226 0.15650927 0.79469415
##
## $delta
## [1] 0.4903408 0.3713038 0.1383554
##
## $delta.upper.conf
## [1] 0.5392038 0.4160130 0.1554508
##
## $delta.lower.conf
## [1] 0.4414779 0.3265945 0.1212600
```

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)
```



\begin{figure}  
\caption{LEFT: Histogram of observations with state dependent poisson distributions and their 95% confidence intervals overlayed. The black line is the marginal distribution. RIGHT: Timeseries of first 300 observations with colors corresponding fitted hidden states decoded by Viterbi algorithm.} \end{figure}

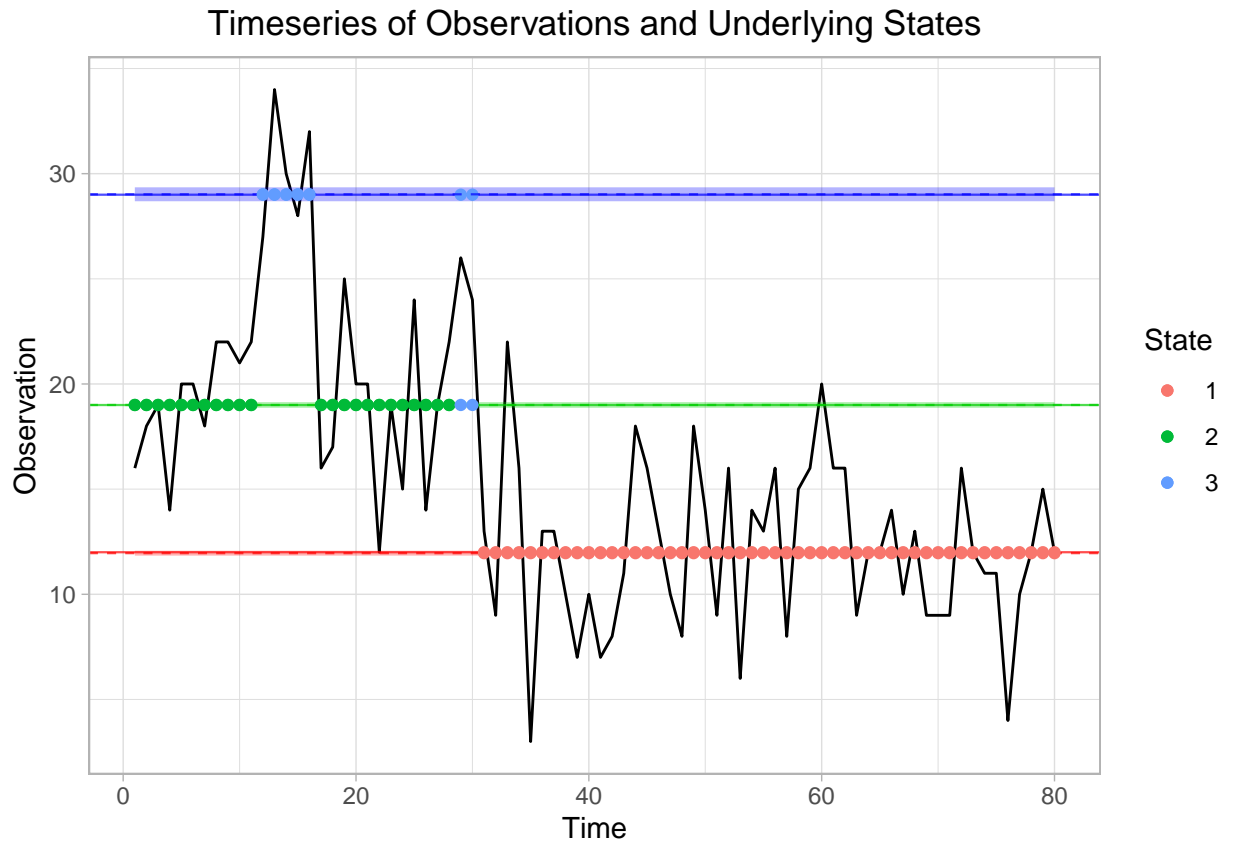


Figure 2: Timeseries of first 80 observations. The solid horizontal lines are the true means ( $\lambda$ ) of each state dependent poisson distribution and the dashed lines are the fitted means. For each time interval there are two points, one falling on the true state mean line and one on the fitted state mean line. The colors of the points correspond to the true states of the observation at that time. The 95% confidence intervals for the fitted  $\lambda$ s are the shaded regions.

```
pr <- pois.HMM.pseudo_residuals(x, modpois3s)
```

```

poisdata3s$PR_lo <- pr$lo
poisdata3s$PR <- pr$mi
poisdata3s$PR_hi <- pr$hi

index3 <- pr.plot.discr(poisdata3s)
hist3 <- pr.hist(poisdata3s)
qq3 <- pr.qq(poisdata3s)
acf3 <- pr.acf(poisdata3s$PR)
grid.arrange(index3,hist3,qq3,acf3, ncol=2)

```

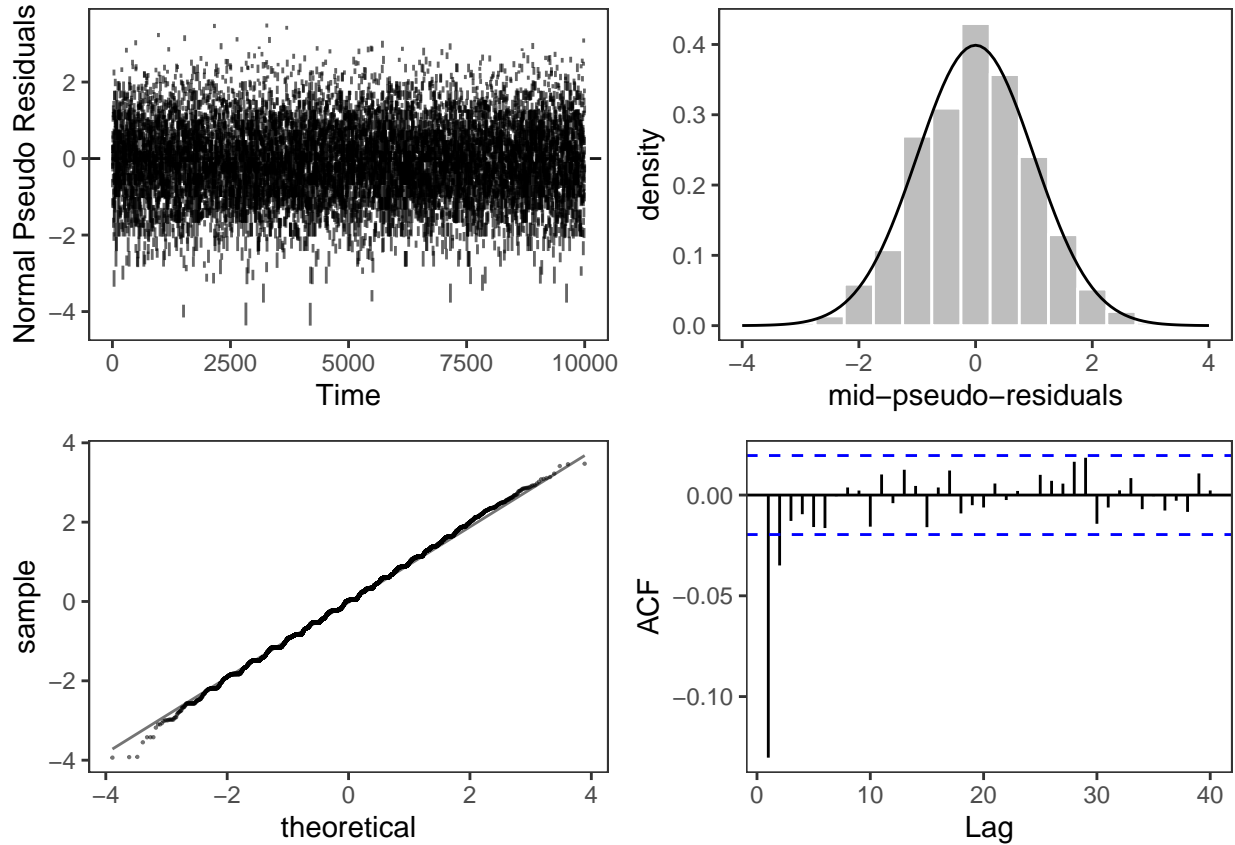


Figure 3: Plots for the normal pseudo-residuals of the fitted 3-state poisson HMM. TOP LEFT: Index plot of the normal pseudo residuals. TOP RIGHT: Histogram of the normal pseudo-residuals with  $N(0,1)$  overlaid. BOTTOM LEFT: Quantile-quantile plot of the normal pseudo-residuals. BOTTOM RIGHT: Autocorrelation 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.
##      1.00   12.00   16.00   16.93   21.00   47.00
```

I choose  $\lambda_0$  according to the 1st quartile, and 3rd quartile. I choose  $\Gamma_0$  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, delta0 = c(1/3,1/3), stationary=FALSE)
modpois2s
```

```
## $m
## [1] 2
##
## $lambda
## [1] 12.23434 22.43904
##
## $gamma
##           [,1]      [,2]
## [1,] 0.94838880 0.0516112
## [2,] 0.06049222 0.9395078
##
## $delta
## [1] 0.5396109 0.4603891
##
## $code
## [1] 1
##
## $mllk
## [1] 31158.43
##
## $AIC
## [1] 62326.85
##
## $BIC
## [1] 62362.9
```

Get 95% confidence intervals for fitted parameters

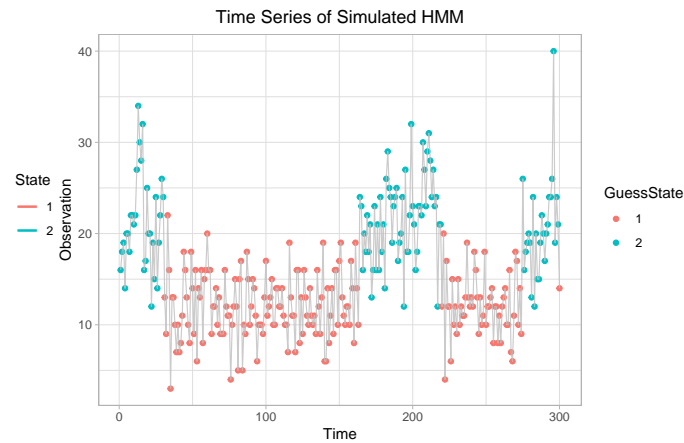
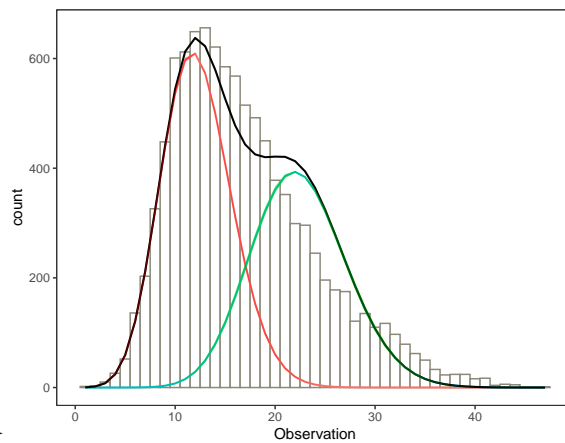
```
CI2 = pois.HMM.CI(x, 20, modpois2s, stationary = TRUE)
CI2
```

```
## $lambda
## [1] 12.23434 22.43904
##
## $lambda.upper.conf
## [1] 12.30460 22.59332
##
## $lambda.lower.conf
```

```
## [1] 12.16409 22.28477
##
## $gamma
##      [,1]      [,2]
## [1,] 0.94838880 0.0516112
## [2,] 0.06049222 0.9395078
##
## $gamma.upper.conf
##      [,1]      [,2]
## [1,] 0.95551513 0.05873752
## [2,] 0.06898791 0.94800346
##
## $gamma.lower.conf
##      [,1]      [,2]
## [1,] 0.94126248 0.04448487
## [2,] 0.05199654 0.93101209
##
## $delta
## [1] 0.5396109 0.4603891
##
## $delta.upper.conf
## [1] 0.5818832 0.5026615
##
## $delta.lower.conf
## [1] 0.4973385 0.4181168
```

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)
```



\begin{figure}  
\caption{LEFT: Histogram of observations with state dependent poisson distributions and their 95% confidence intervals overlaid. Black line is the marginal distribution. RIGHT: Timeseries of first 300 observations with colors corresponding fitted hidden states decoded by Viterbi algorithm.} \end{figure}

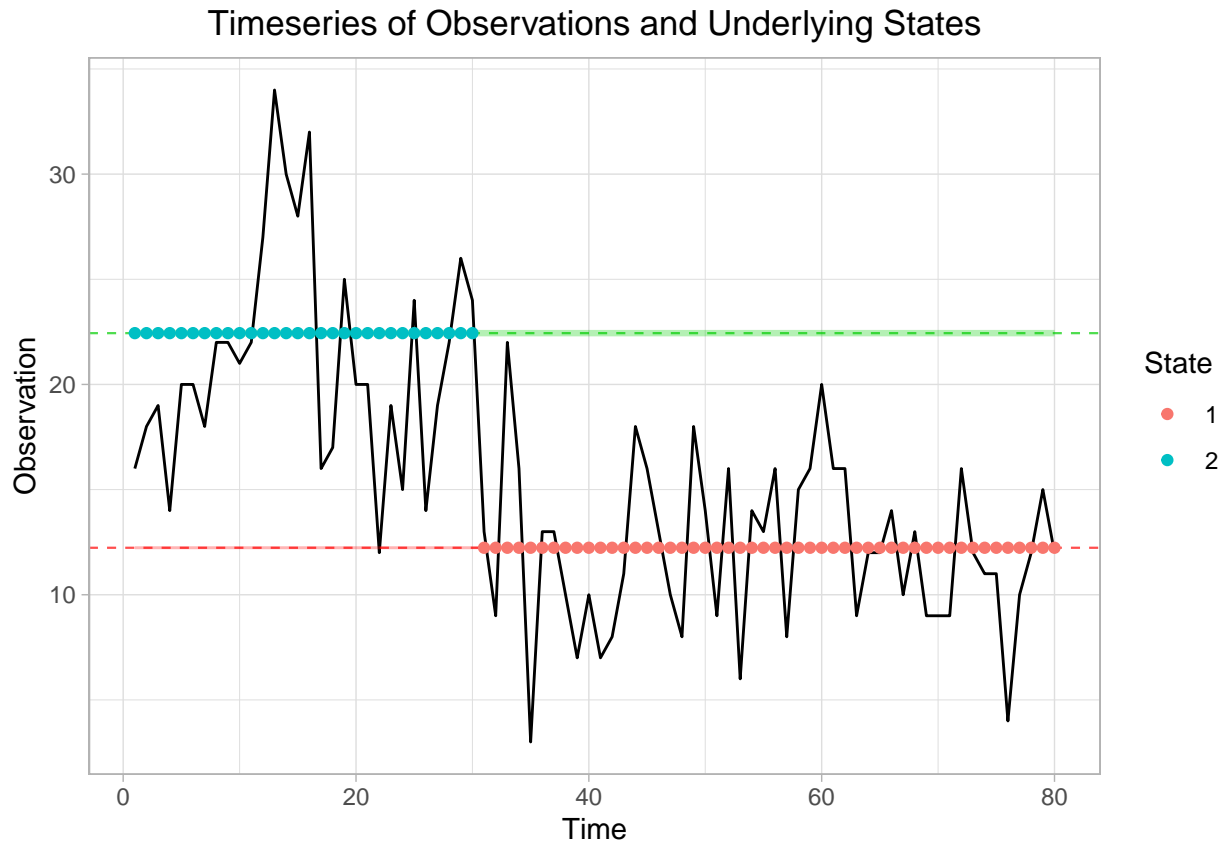


Figure 4: Timeseries of first 80 observations. The dashed lines are the fitted means ( $\lambda$ ) for each state dependent poisson distribution. The colors of the points correspond to the fitted state of the observation at that time.

```
pr <- pois.HMM.pseudo_residuals(x, modpois2s)

poisdata2s$PR_lo <- pr$lo
poisdata2s$PR <- pr$mi
poisdata2s$PR_hi <- pr$hi

index2 <- pr.plot.discr(poisdata2s)
hist2 <- pr.hist(poisdata2s)
qq2 <- pr.qq(poisdata2s)
acf2 <- pr.acf(poisdata2s$PR)
grid.arrange(index2, hist2, qq2, acf2, ncol=2)
```



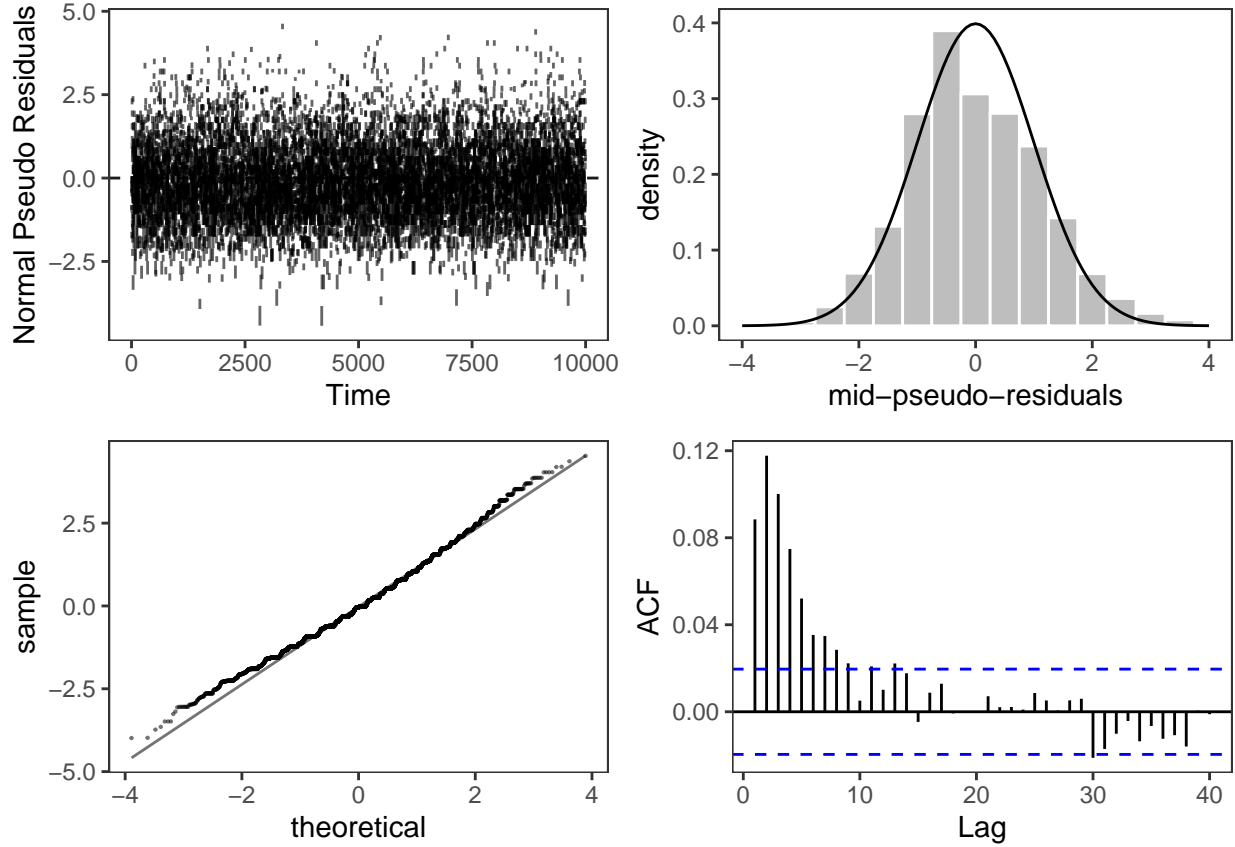


Figure 5: Plots for the normal pseudo-residuals of the fitted 2-state poisson HMM. TOP LEFT: Index plot of the normal pseudo residuals. TOP RIGHT: Histogram of the normal pseudo-residuals with  $N(0,1)$  overlaid. BOTTOM LEFT: Quantile-quantile plot of the normal pseudo-residuals. BOTTOM RIGHT: Autocorrelation 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.
##	1.00	12.00	16.00	16.93	21.00	47.00

I choose  $\lambda_0$  based on the 1st quartile, median, and 3rd quartile with some variation. I choose  $\Gamma_0$  so that the diagonals 0.91 and the off-diagonals are 0.03.

```
modpois4s <-pois.HMM.mle(x, m, lambda0, gamma0, c(1/3,1/3,1/3,1/3), stationary=FALSE)
modpois4s
```

```
## $m
## [1] 4
##
## $lambda
## [1] 11.95396 18.98512 29.89970 28.10162
##
## $gamma
##           [,1]      [,2]      [,3]      [,4]
## [1,] 0.95850724 0.02177583 0.01238713 0.007329799
## [2,] 0.05051773 0.90421992 0.01062718 0.034635171
## [3,] 0.02201914 0.08865364 0.04809125 0.841235968
## [4,] 0.00275594 0.25606240 0.71033473 0.030846933
##
## $delta
## [1] 0.49029681 0.37003106 0.06570621 0.07396592
##
## $code
## [1] 4
##
## $mllk
## [1] 30010.41
##
## $AIC
## [1] 60058.82
##
## $BIC
## [1] 60195.81
```

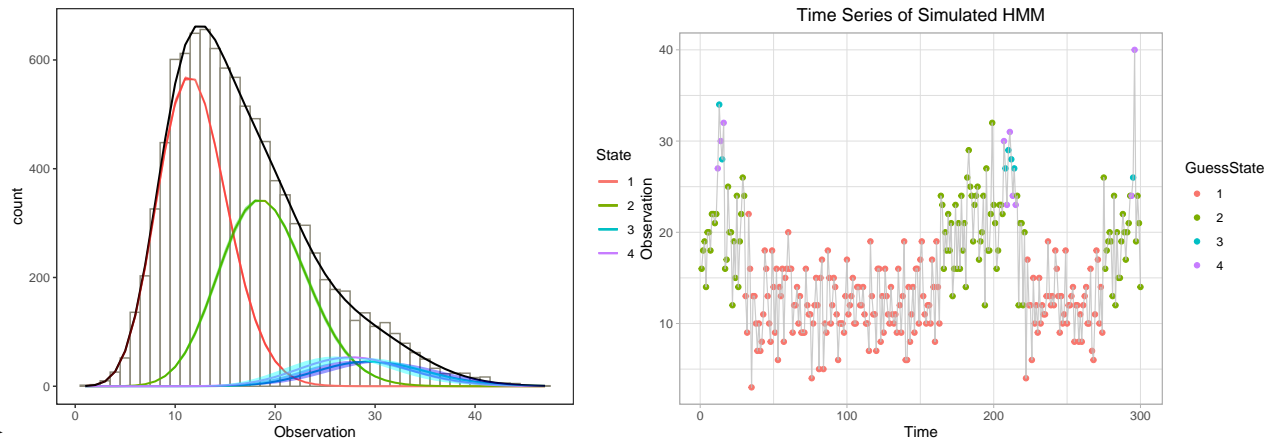
```
CI4 = pois.HMM.CI(x, 20, modpois4s, stationary = TRUE)
CI4
```

```
## $lambda
## [1] 11.95396 18.98512 29.89970 28.10162
##
## $lambda.upper.conf
## [1] 12.05012 19.16244 31.45609 29.69341
##
## $lambda.lower.conf
## [1] 11.85780 18.80780 28.34330 26.50983
##
## $gamma
##           [,1]      [,2]      [,3]      [,4]
## [1,] 0.95850724 0.02177583 0.01238713 0.007329799
## [2,] 0.05051773 0.90421992 0.01062718 0.034635171
## [3,] 0.02201914 0.08865364 0.04809125 0.841235968
## [4,] 0.00275594 0.25606240 0.71033473 0.030846933
##
## $gamma.upper.conf
##           [,1]      [,2]      [,3]      [,4]
## [1,] 0.965012591 0.02924791 0.0228051 0.01840584
```

```
## [2,] 0.059034747 0.91703179 0.0290880 0.05734650
## [3,] 0.044323045 0.22908705 0.1320966 1.02791291
## [4,] 0.009565335 0.38866754 0.8464090 0.06387981
##
## $gamma.lower.conf
##      [,1]      [,2]      [,3]      [,4]
## [1,] 0.95200188 0.01430375 0.001969172 0.00000000
## [2,] 0.04200072 0.89140804 0.000000000 0.01192384
## [3,] 0.00000000 0.00000000 0.000000000 0.65455903
## [4,] 0.00000000 0.12345727 0.574260479 0.00000000
##
## $delta
## [1] 0.49029681 0.37003106 0.06570621 0.07396592
##
## $delta.upper.conf
## [1] 0.53671195 0.40592790 0.07852726 0.08737663
##
## $delta.lower.conf
## [1] 0.44388167 0.33413421 0.05288516 0.06055520
```

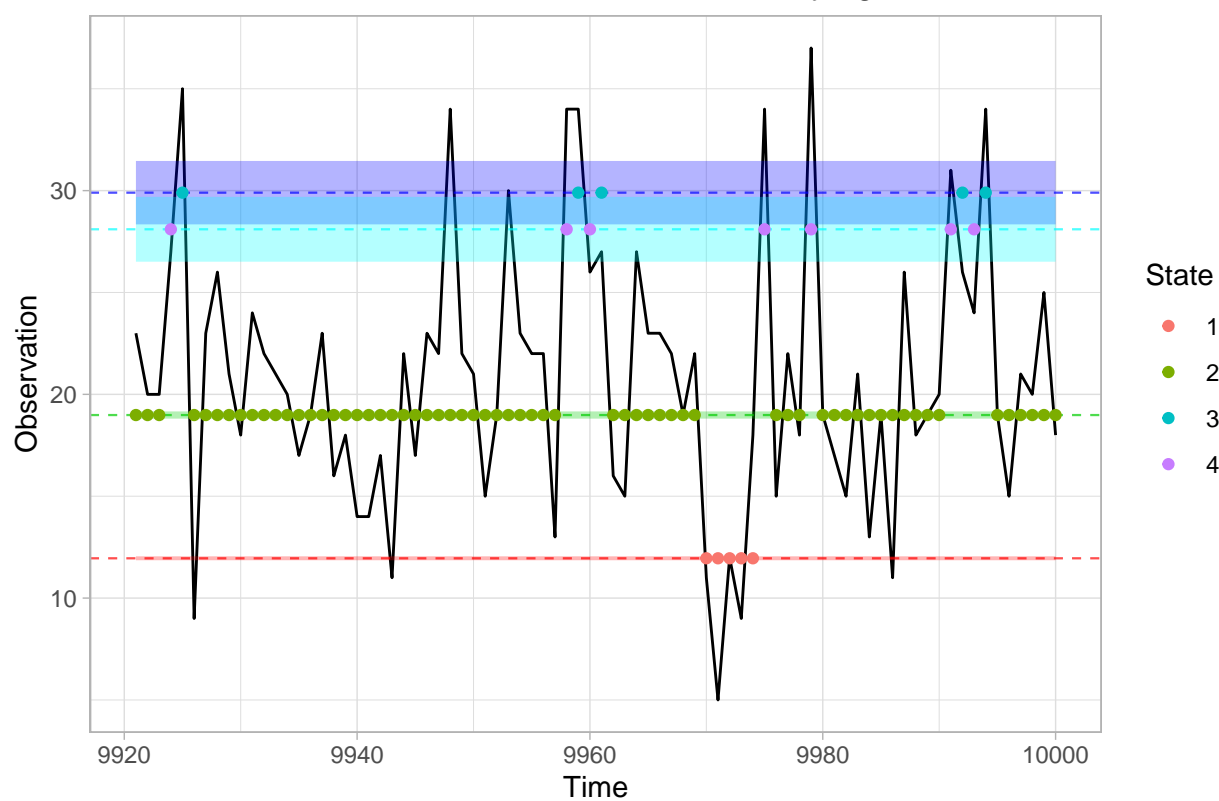
Now I find the underlying state sequence using global decoding by the Viterbi algorithm

```
state_seq <- pois.HMM.viterbi(x, modpois4s)
poisdata4s <- poisdata
poisdata4s$GuessState <- as.factor(state_seq)
```



\begin{figure}  
\caption{LEFT: Histogram of observations with state dependent poisson distributions and their 95% confidence intervals overlaid. Black line is the marginal distribution. RIGHT: Timeseries of first 300 observations with colors corresponding fitted hidden states decoded by Viterbi algorithm.} \end{figure}  
We can visualize the state sequence and observations the same way as before (without the true states)

## Timeseries of Observations and Underlying States



```
pr <- pois.HMM.pseudo_residuals(x, modpois4s)
```

```
poisdata4s$PR_lo <- pr$lo
poisdata4s$PR <- pr$mi
poisdata4s$PR_hi <- pr$hi
```

```
index4 <- pr.plot.discr(poisdata4s)
hist4 <- pr.hist(poisdata4s)
qq4 <- pr.qq(poisdata4s)
acf4 <- pr.acf(poisdata4s$PR)
grid.arrange(index4, hist4, qq4, acf4, ncol=2)
```

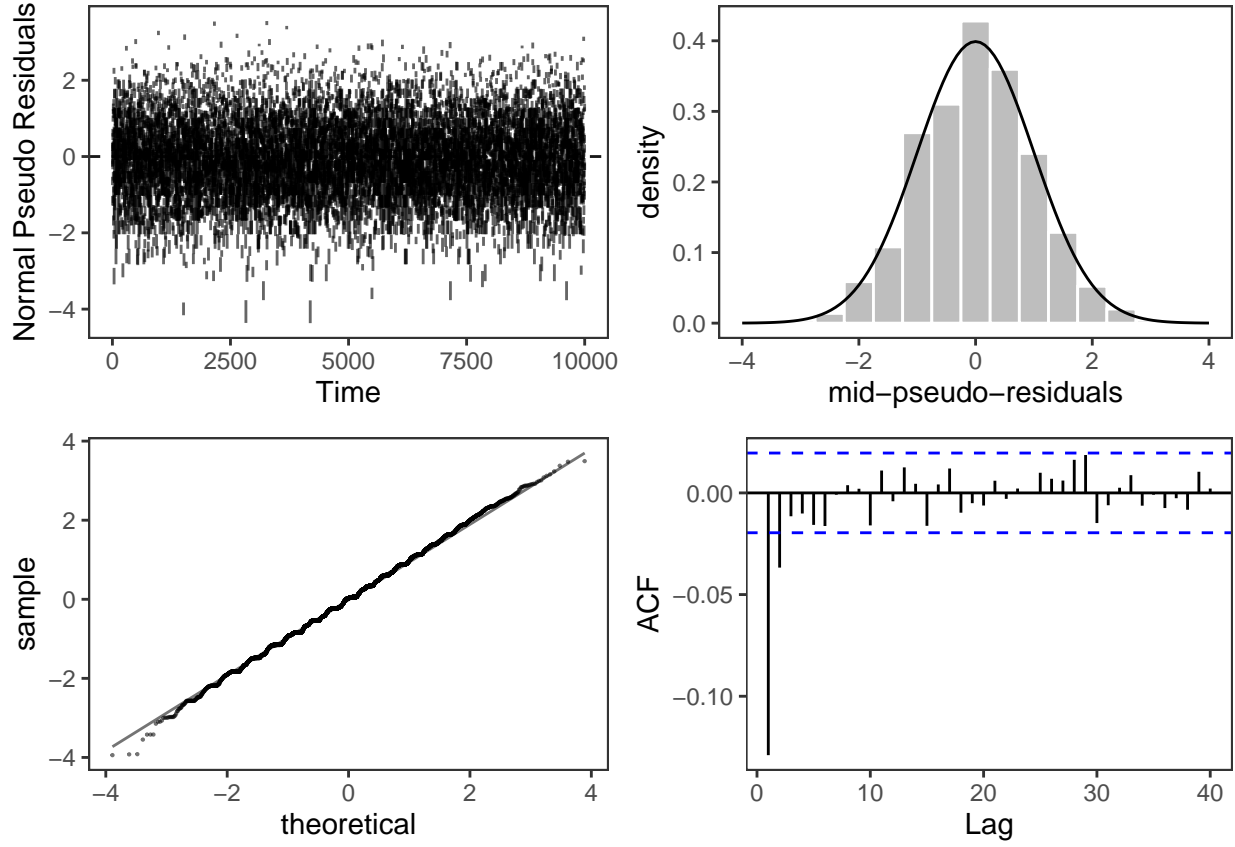


Figure 6: Plots for the normal pseudo-residuals of the fitted 4-state poisson HMM. TOP LEFT: Index plot of the normal pseudo residuals. TOP RIGHT: Histogram of the normal pseudo-residuals with  $N(0,1)$  overlaid. BOTTOM LEFT: Quantile-quantile plot of the normal pseudo-residuals. BOTTOM RIGHT: Autocorrelation function of the normal pseudo-residuals.

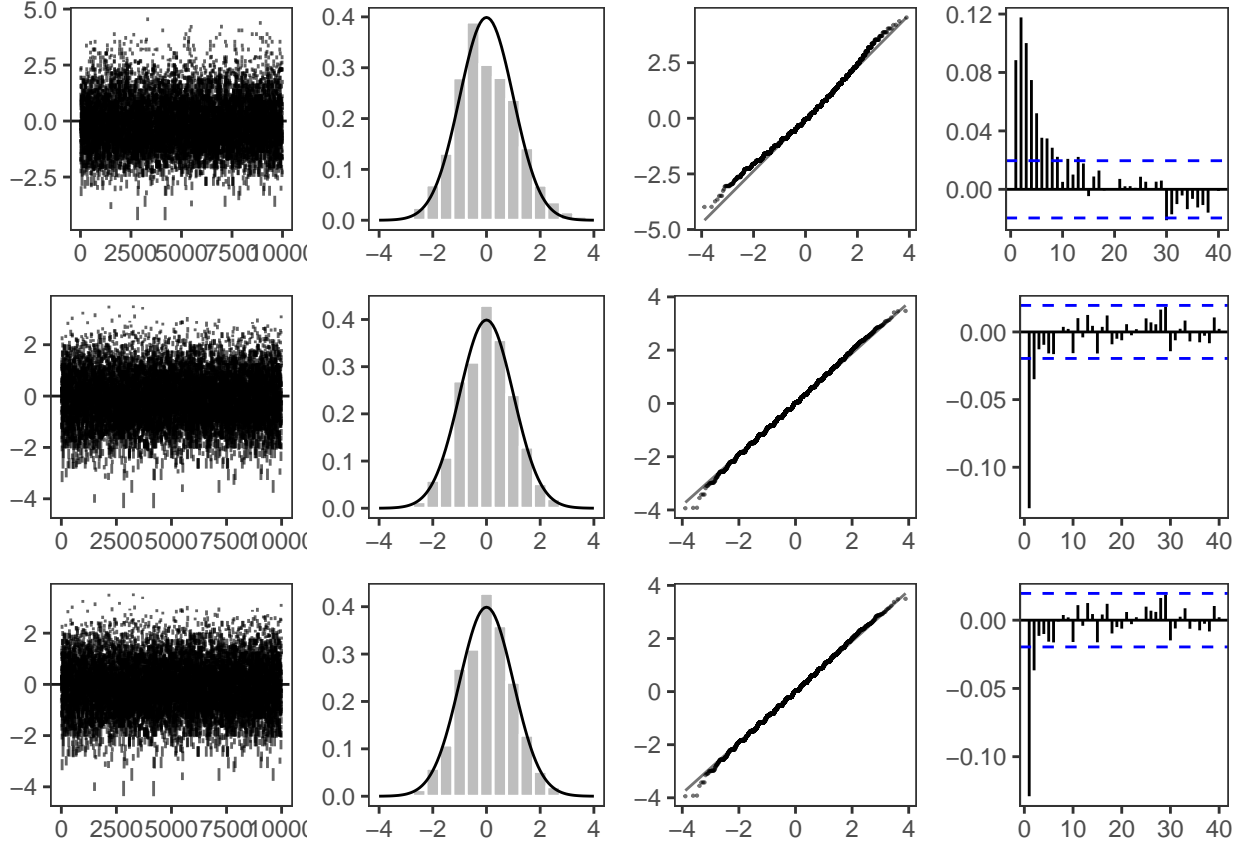


Figure 7: Each row shows plots of the pseudo residuals for different fitted HMMs. The first row is for a 2-state HMM with poisson state dependent distributions. The second is a 3-State poisson distributed HMM and the third is a 4-state poisson distributed HMM. The first column shows index plots of the 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