Vignette spINAR

Introduction to spINAR

The spINAR package:

- It provides an efficient semiparametric estimation of the autoregressive parameters and the innovation distribution for the integer-valued autoregressive model of order p (INAR(p)) for $p \in \{1, 2\}$.
- The estimation of the parameters is conducted by maximizing the conditional likelihood of the model.

```
library(spINAR)
set.seed(1234)
```

INAR(1)

First, we create the function `pinar1` to generate INAR(1) data with poisson distributed innovations, where we use $x_0 = 0$ as initialization. The function `pinar1` take three arguments:

1. `n`: sample size.
2. `alpha`: true INAR(1) coefficient.
3. `lambda`: true parameter of the poisson innovation distribution.

```
pinar1 <- function(n, alpha, lambda) {
  err <- rpois(n, lambda)
  x <- numeric(n)
  x[1] <- err[1]
  for (i in 2:n) {
    x[i] <- rbinom(1, x[i - 1], alpha) + err[i]
  }
  return(x)
}
```

Fixing the parameters $n = 1000$, $\alpha = 0.5$ and $\lambda = 1$ in the `pinar1` function, we generate $n = 1000$ INAR(1) observations with poisson distributed innovations denoted by `data_p1`.

```
n <- 1000
alpha <- 0.5
lambda <- 1
data_p1 <- pinar1(n, alpha, lambda)
```

```
barplot(table(data_p1), xlab = "data", ylab = "count")
```
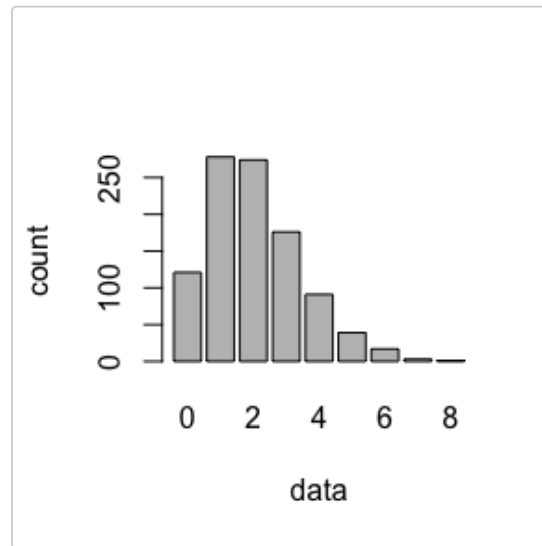
The `spinar` function in the `spINAR` package returns a vector jointly containing the following two entities:

- estimated autoregressive coefficients $\hat{\alpha}_1, \ldots, \hat{\alpha}_p$, where $\hat{\alpha}_i \in (0, 1)$, $\forall i \in \{1, \ldots, p\}$.
- estimated entries of the probability mass function of the innovation distribution denoted with $\hat{g}(0), \hat{g}(1), \ldots$ with $\hat{g}(i) \in [0, 1] \, \forall i \in \{0, 1, \ldots\}$ and $\sum_i \hat{g}(i) = 1$.

```
est_p1 <- spinar(data_p1, 1)
str(est_p1)
#>  num [1:10] 0.4716 0.3303 0.4052 0.1492 0.0845 ...
```

Here, we get (rounded to four decimal places) $\hat{\alpha}_1 = 0.4716$ and $\hat{g}(0) = 0.3303$, $\hat{g}(1) = 0.4052$, $\hat{g}(2) = 0.1492$, $\hat{g}(3) = 0.0845$, , $\hat{g}(4) = 0.0282$, $\hat{g}(5) = 0.0026$, $\hat{g}(6) = 0$, $\hat{g}(7) = 0$ and $\hat{g}(8) = 0$.
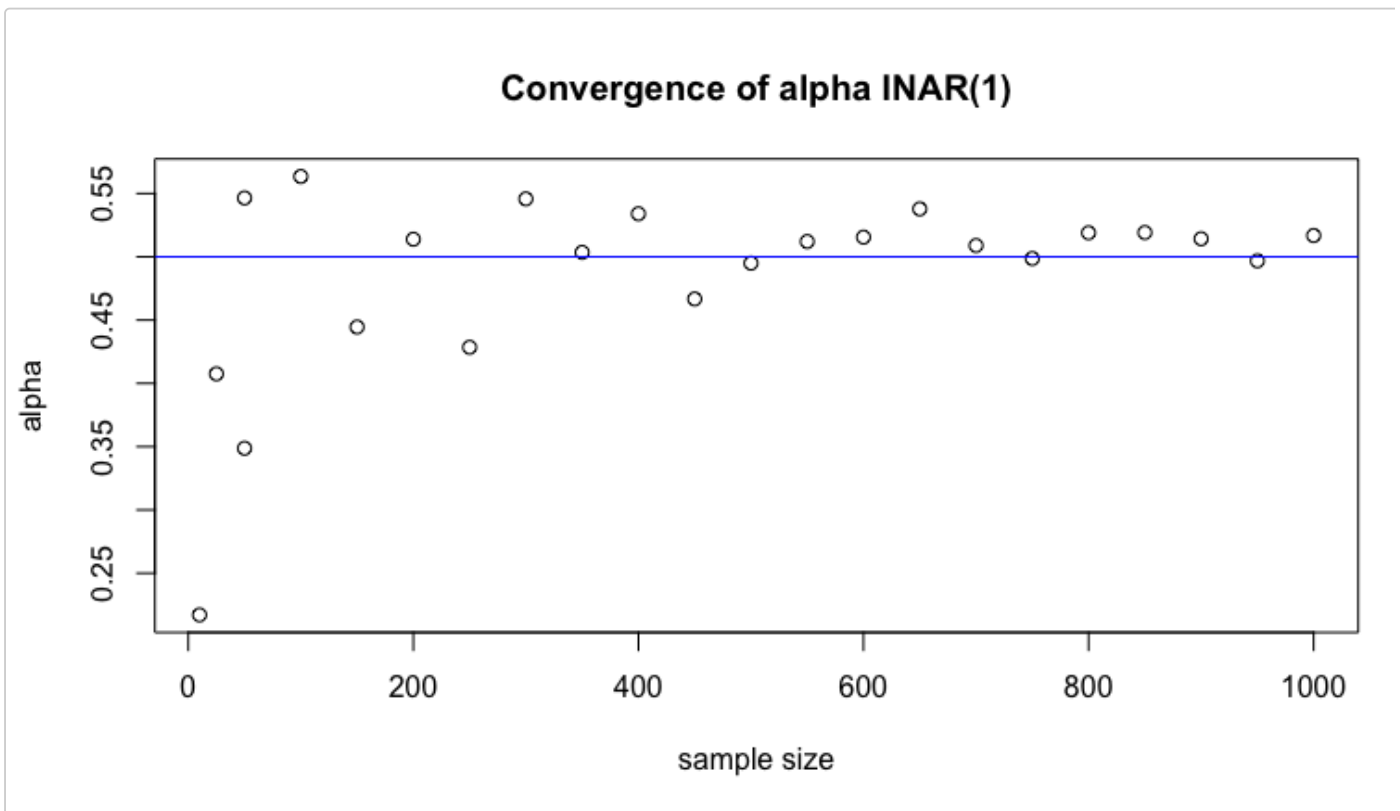
The squared L_2 distance between the estimated and the true innovation distribution is given as follows.

```
sum((est_p1-c(alpha, dpois(0:(length(est_p1)-2),lambda)))^2)
#> [1] 0.005526475
```

The convergence to the true parameter α is presented in the following graph.

```
alphaList_inar1 <- c()
Yaxis_inar1 <- c(c(10, 25, 50), seq(50,n, 50))
for (m in Yaxis_inar1){
  alphaList_inar1 <- append(alphaList_inar1, spinar(pinar1(m, alpha, lambda), 1)[1])
}
```

```
plot(Yaxis_inar1, alphaList_inar1, xlab="sample size", ylab="alpha", main = "Convergence of
  alpha INAR(1)")
abline(h=0.5, col="blue")
```

Convergence of alpha INAR(1)

These results show that:

1. The small resulting value for the squared L_2 distance of approximately 0.0055 indicates a good estimation performance.

2. The estimated values resulting from using the `spinar` function are (rounded to four decimal places):

   - $\hat{\alpha}_1 = 0.4716$, which is close to the true value $\alpha_1 = 0.5$.

   - The estimated entries of the probability mass function of the innovation distribution denoted by $\{g(i), i \in \{0, 1, \ldots\}\}$, where $\hat{g}(0) = 0.3303$, $\hat{g}(1) = 0.4052$, $\hat{g}(2) = 0.1492$, $\hat{g}(3) = 0.0845$, $\hat{g}(4) = 0.0282$, $\hat{g}(5) = 0.0026$, $\hat{g}(6) = 0$, $\hat{g}(7) = 0$, $\hat{g}(8) = 0$, with $\sum_{i=0}^{8} \hat{g}(i) = 1$.

3. About the convergence in parameter α. We see that on average, the bigger the sample size, the better is the estimation.
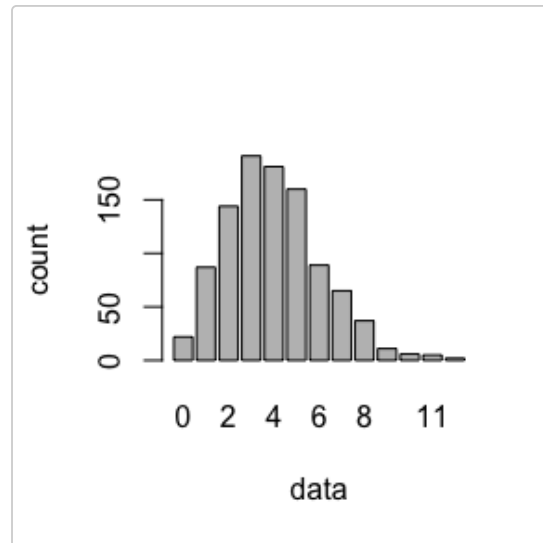
INAR(2)

Analogously we create the function `pinar2` to generate INAR(2) data with poisson distributed innovations.

```
pinar2 <- function(n, alpha1, alpha2, lambda2) {
  err <- rpois(n, lambda2)
  x <- numeric(n)
  x[1] <- err[1]
  x[2] <- x[1] + err[2]
  for (i in 3:n) {
    x[i] <-
      rbinom(1, x[i - 1], alpha1) + rbinom(1, x[i - 2], alpha2) + err[i]
  }
  return(x)
}
```

We set the true parameters as follows and generate the data.

```
n <- 1000
alpha1 <- 0.4
alpha2 <- 0.2
lambda2 <- 1.5
data_p2 <- pinar2(n, alpha1, alpha2, lambda2)
```

```
barplot(table(data_p2), xlab = "data", ylab = "count")
```



```
est_p2 <- spinar(data_p2, 2)
```

```
sum((est_p2-c(alpha1, alpha2, dpois(0:(length(est_p2)-3),lambda2)))^2)
#> [1] 0.01439354
```

The convergence to the true parameter α is presented in the following graph.

```
alpha1List_inar2 <- c()
alpha2List_inar2 <- c()
Yaxis_inar2 <- c(c(10, 25, 50), seq(50, n, 50))
for (m in Yaxis_inar2){
  input_pinar2 <- pinar2(m, alpha1, alpha2, lambda2)
  input_inar2 <- spinar(input_pinar2,2)
  alpha1List_inar2 <- append(alpha1List_inar2, input_inar2[1])
  alpha2List_inar2 <- append(alpha2List_inar2, input_inar2[2])
}
```

```
par(mfrow=c(1,2))
```

```
plot(Yaxis_inar2, alpha1List_inar2, xlab="sample size", ylab="alpha1")
abline(h = alpha1, col = "blue")
```

```
title("Convergence of alpha INAR(2)", line = 1, adj = 0.2)
plot(Yaxis_inar2, alpha2List_inar2, xlab="sample size", ylab="alpha2")
abline(h = alpha2, col = "blue")
```
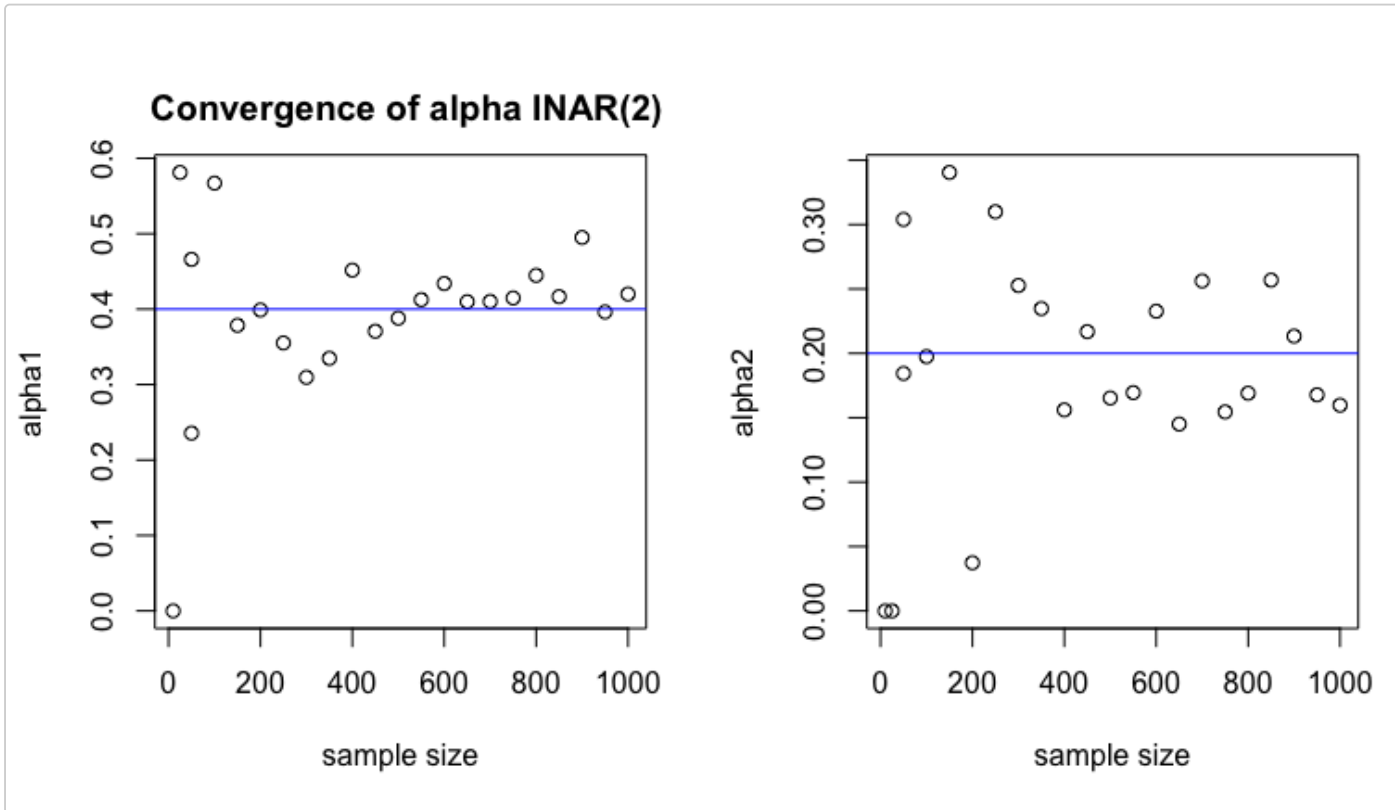
Convergence of alpha INAR(2)

These results show that:

1. The small resulting value for the squared L_2 distance of approximately 0.0144 indicates a good estimation performance.

2. The estimated values resulting from using the `spinar` function are (rounded to four decimal places):

   - $\hat{\alpha}_1 = 0.4146$, and $\hat{\alpha}_2 = 0.1631$ which are close to the true values $\alpha_1 = 0.4$ and $\alpha_2 = 0.2$.

   - The estimated entries of the probability mass function of the innovation distribution denoted by $\{g(i), i \in \{0, 1, \dots\}\}$, where $\hat{g}(0) = 0.208$, $\hat{g}(1) = 0.2732$, $\hat{g}(2) = 0.2929$, $\hat{g}(3) = 0.0949$, $\hat{g}(4) = 0.1239$, $\hat{g}(5) = 10^{-4}$, $\hat{g}(6) = 0.0026$, $\hat{g}(7) = 0.0042$, $\hat{g}(8) = 10^{-4}$, $\hat{g}(9) = 0$, $\hat{g}(10) = 0$, $\hat{g}(11) = 0$ with $\sum_{i=0}^{11} \hat{g}(i) = 1$.

3. About the convergence in parameter α. We see that on average, the bigger the sample size the better is the estimation for α_1 and α_2.