Introduction to Count Time Series

Alejandra Cabaña

Universitat Autònoma de Barcelona

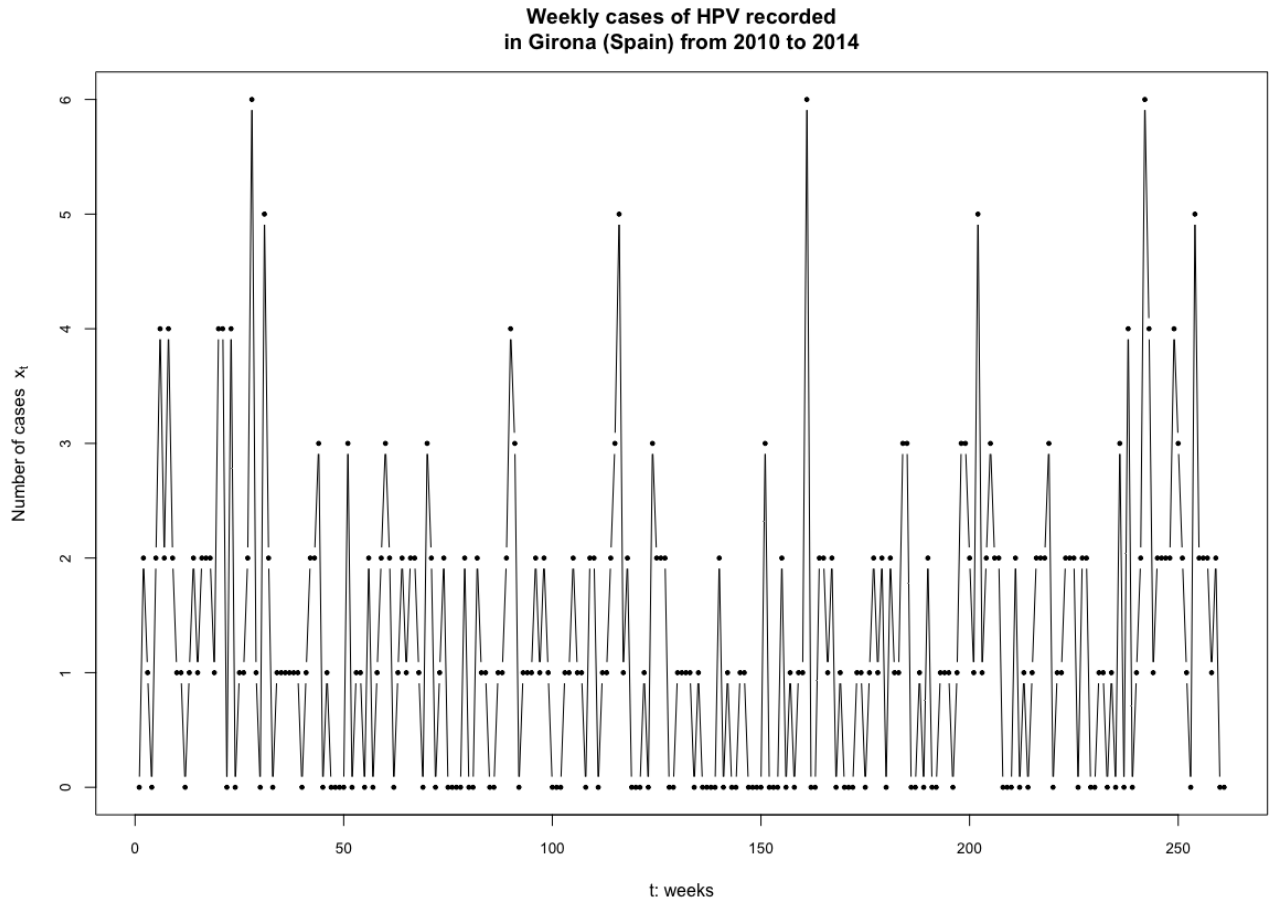
Mini-course ECODEP, February 8 and 15, 2021

Overview

1. Stationary count time series
2. Thinning based methods
3. Conditional Regression Models (INGARCH)

Stationary count time series

Assume that $X_1, \ldots X_T$ is a univariate time series originating from a stationary count process.



We can check its marginal characteristics, and based on these, define a set of candidate models for the subsequent model fitting and diagnostics.

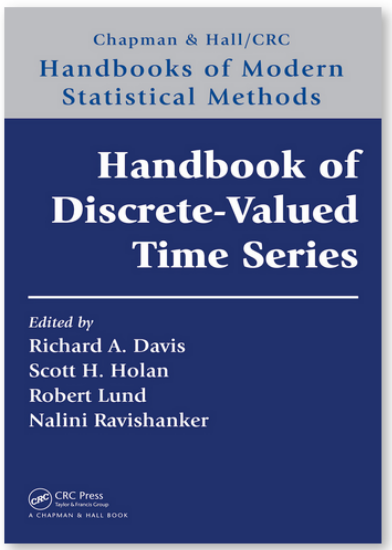
In general, the theoretical aspects of count time series are quite technical.

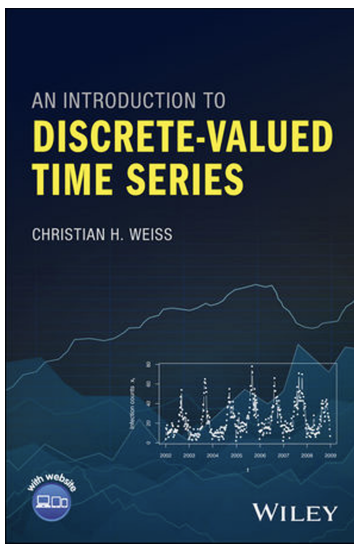
The main problem is that the observed response, being discrete valued, might not be strong-mixing.

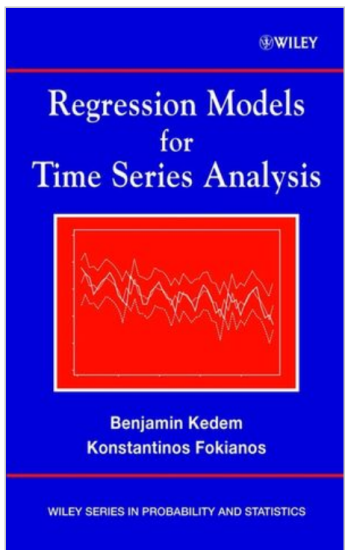
Many work on weak dependence[1] and mixing results via coupling[2] has been done.

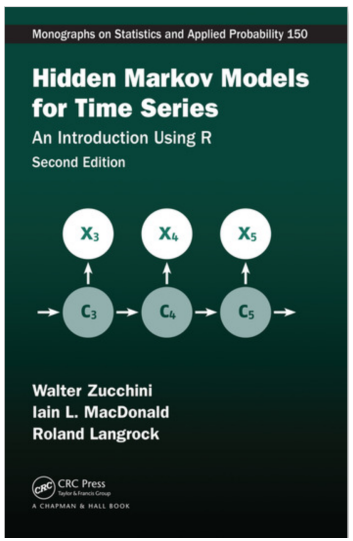
But in this lectures, we will confine our attention to data analysis tools.

References:









[1] Doukhan, Fokianos and Tjostheim,(2012) S&PL
[2] Neumann (2011), Bernoulli

Models for Stationary Count time Series

There are different approaches for modelling univariate (unbounded) count time series, the most popular ones are:

1. Thinning-based methods (INARMA)

   McKenzie(19859, Al-Osh and Alzaid (1987), Du and Li (1991) , Weiss (2009), Scotto et.al (2015), Weiss (2018), ...

2. Conditional (linear or non-linear) regression models (INGARCH)

   Ferland, Laatour and Oraichi (2006), Fokianos and Tjotstheim (2011)...

3. Hidden-Markov models (HMM)

   Zucchini and Mc Donald (2009)

The former two are ARMA-like approaches.

Thinning-based Methods

INARMA models

In data applications with stationary count time series, one often observes an autocorrelation structure similar to that of the ordinary autorregressive moving average (ARMA) models (Box and Jenkins, 1970),

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots \theta_q \varepsilon_{t-q}$$

where Y_t and ε_t are real-valued random variables.
This model cannot be applied to count processes, because multiplications do not preserve the integer range.

The idea is to modify the basic ARMA in such a way that the resulting model generates only count data, but still with an ARMA-like ACF.

Substitute multiplications by an integer-valued counterpart, a so-called "thinning operation".

The resulting models are then referred to as integer-valued ARMA (INARMA) models.

Binomial thinning

There exists a large variety of possible thinning operations[3], but we shall concentrate on the **binomial thinning operator**[4].

For $\alpha \in [0,1]$, and applied to the count random variable X, define the **α-thinning of X** by

$$\alpha \circ X = \sum_{i=1}^{X} Z_i$$

where Z_i are i.i.d Bernoulli(α).
In contrast to the ordinary multiplication by α, the thinning constitutes a random operation that generates a count value from $\{0, \ldots, X\}$. But since both operations have the same mean, we can use binomial thinning as an integer-valued substitute of the multiplication.

$\alpha \circ X$ can be interpreted as the random number of *survivors* from a population of size X, where each individual survives with probability α.

[3] see the survey by Scotto, Weiß, and Gouveia (2015)
[4] Steutel and van Harn (1979)

The expectation and variance of the *binomial thinning* operator on X, with $\mathbf{E}X = \mu$ and $\mathbf{Var}X = \sigma^2$ are

$$\mathbf{E}\left(\alpha \circ X | X\right) = \alpha X \Rightarrow \mathbf{E}\alpha \circ X = \alpha\mu$$

$$\mathbf{Var}\left(\alpha \circ X | X\right) = \alpha(1-\alpha)X$$

so that

$$\mathbf{Var}\alpha \circ X = \mathbf{Var}\left(\mathbf{E}\alpha \circ X | X\right) + \mathbf{E}(\alpha(1-\alpha)X) = \alpha^2\sigma^2 + \quad \alpha(1-\alpha)\mu$$

INAR(1)[5]

Consider a process X_n

$$X_n = \alpha \circ X_{n-1} + \epsilon_n,$$

where $0 < \alpha < 1$ is a fixed parameter, ϵ_n are non-negative integer valued RVs, i.i.d., independent of X_n and $\circ$ is the binomial thinning operator:

$$\alpha \circ X_{n-1} = \sum_{i=1}^{X_{n-1}} Z_i$$

where Z_i are i.i.d Bernoulli(α).
The expectation and variance of the *binomial thinning* operator are

$$\mathbf{E}\left(\alpha \circ X_{n-1} | X_{n-1} = x_{n-1}\right) = \alpha x_{n-1}$$

$$\mathbf{Var}\left(\alpha \circ X_{n-1} | X_{n-1} = x_{n-1}\right) = \alpha(1-\alpha) x_{n-1}$$

[5] McKenzie (1985), *Some simple models for discrete variate time series* inTime Series Analysis in Water Resources and Al-Osh and Alzaid (1987), JTSA.

The INAR(1) process is a homogeneous Markov chain with transition probabilities

$$\mathbf{P}(X_n = i | X_{n-1} = j) = \sum_{k=0}^{i \wedge j} \binom{j}{k} \alpha^k (1-\alpha)^{j-k} \mathbf{P}(\epsilon_n = i - k)$$

The conditional expectation and variance are both linear in X_{n-1}

$$\mathbf{E}\left(X_n | X_{n-1}\right) = \alpha X_{n-1} + \mathbf{E}(\epsilon_n)$$

$$\mathbf{Var}\left(X_n | X_{n-1}\right) = \alpha(1-\alpha) X_{n-1} + \mathbf{Var}(\epsilon_n)$$

hence INAR(1) is conditionally heteroscesdastic, as opposed to AR(1).

Branching processes with immigration (BPI)

Branching processes with immigration (BPI) $(X_t)_{\mathbf{N}_0}$**:**

$$X_t = \underbrace{Z_{t;1} + \cdots + Z_{t;X_{t-1}}}_{=0 \text{ if } X_{t-1}=0} + \epsilon_t,$$

where $X_0, Z_{t;r}, \epsilon_s$ independent count r.v.

Offspring $Z_{t;r}$ i.i.d. with pgf $g_Z(z) = \sum_{k=0}^{\infty} a_k z^k$.

Immigration ϵ_s i.i.d. with pgf $g_\epsilon(z) = \sum_{k=0}^{\infty} b_k z^k$.

If offspring mean $\mu_Z = g_Z'(1) < 1$, then **subcritical** BPI.

If $a_k = 0$ for $k > 2$, i.e. if $Z_{t;r} \sim \text{Bin}(1, a_1)$, then $Z_{t;1} + \cdots + Z_{t;X_{t-1}}$ is a binomial thinning $a_1 \circ X_{t-1}$ (interpretation of survivors).

$\Rightarrow$ INAR(1) model is *a* special type of subcritical BPI.

Stationary distributions of INAR(1)

Results for subcritical BPIs are applicable to INAR(1) process:

Subcritical BPIs which are irreducible and aperiodic Markov chains that satisfy $\mathbf{E}(h(\epsilon_s)) < 1$, where $h(k) = \sum_{j=1}^{k} j^{-1}$ have a proper stationary marginal distribution[6] for $X_t, t = 0, 1, \ldots$.

$\mathbf{E}(h(\epsilon_s)) < 1$ holds if ϵ_s has finite mean.

Geometric ergodicity of such subcritical BPIs, allows to derive mixing properties for INAR(1) models[7].

In fact, INAR(1) models are α-mixing, so there's a CLT.

---

[6] Heathcote (1966)

[7] Pakes(1971)

Marginal Distribution

It is straightforward to see from $X_t = \alpha \circ X_{t-1} + \epsilon_t$ that

$$\mu_X = \alpha\mu_X + \mu_\epsilon \quad \Rightarrow \quad \mu_X = \frac{\mu_\epsilon}{1-\alpha}$$

and

$$\sigma_X^2 = \alpha^2\sigma_X^2 + \alpha(1-\alpha)\mu_X + \sigma_\epsilon^2 \quad \Rightarrow \quad \sigma_X^2 = \frac{\alpha\mu_\epsilon + \sigma_\epsilon^2}{1-\alpha^2}.$$

And also, the probability generating function $pgf(z) = g_X(z) == \mathbf{E}(z^X)$ allows the computation of moments quite easily. Observe that, if there is a stationary distribution for the observations, its *pfg* satisfies

$$\begin{aligned}\mathbf{E}\left(z^{\alpha\circ X_{t-1}+\epsilon_t}|X_{t-1}\right) &= \mathbf{E}\left(z^{\alpha\circ X_{t-1}}|X_{t-1}\right)\mathbf{E}\left(z^{\epsilon_t}|X_{t-1}\right)\\ &= (1-\alpha+\alpha z)^{X_{t-1}} g_\epsilon(z).\end{aligned}$$

Poisson INAR(1) model

The most popular of the INAR(1) family is the Po-INAR(1) model, where the innovations $\epsilon_t \sim$Poiss(λ).
Since all $P(\epsilon_t = j) > 0$, all the transition probabilities of the Markov Chain defining the INAR are positive, consequently Po-INAR is an irreducible and aperiodic Markov Chain, and being a subcritical BPI, it has a unique stationary marginal distribution.
Moreover, since the Poisson distribution is invariant w.r.t. binomial thinning, and it is additive, then the stationary distribution of an INAR(1) process X_n with Poisson(λ) innovations is Poisson with mean and variance

$$\mu_X = \sigma_X^2 = \frac{\lambda}{1-\alpha}$$

Its auto-covariance and auto-correlation functions are

$$\gamma_X(k) = \alpha^{|k|}\lambda \qquad \rho_X(k) = \alpha^{|k|}.$$

Parameter Estimation

- **Method of moments**
  - Marginal mean μ (sample mean $\bar{x}$);
  - first-order autocorrelation $\rho(1)$ for MM estimator of α:

    $\hat{\alpha}_{\text{MM}} := \hat{\rho}(1) := \hat{\gamma}(1)/\hat{\gamma}(0)$ with
    $\hat{\gamma}(k) = \frac{1}{T}\sum_{t=k+1}^{T}(x_t - \bar{x})(x_{t-k} - \bar{x})$ for $k \in \mathbb{N}_0$.
  - Possibly further moment relations, e. g.,

    sample variance $s^2 := \hat{\gamma}(0) = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2$.

- **Conditional Least Squares**

  for **Poisson INAR(1) model**.

  Idea: Choose α and λ to minimize squared deviations

  between x_t and $E[X_t \,|\, x_{t-1}] = \alpha \cdot x_{t-1} + \lambda$, i. e.,

$$(\hat{\alpha}_{\text{CLS}}, \hat{\lambda}_{\text{CLS}}) := \arg\min_{(\alpha,\lambda)} CSS(\alpha, \lambda),$$
$$\text{where} \quad CSS(\alpha, \lambda) := \sum_{t=2}^{T}(x_t - \alpha \cdot x_{t-1} - \lambda)^2.$$

- **Maximum Likelihood estimation**

INAR(1) is a stationary Markov chain, so, the log-likelihood is simply

$$\ell(\boldsymbol{\theta}) = \log p_{X_1}(\boldsymbol{\theta}) + \sum_{t=2}^{T} \log p_{X_t|X_{t-1}}(\boldsymbol{\theta})$$

The marginal probability $p_{X_1}(\boldsymbol{\theta})$ can be computed by MonteCarlo or we can simply maximize the conditional log-likelihood

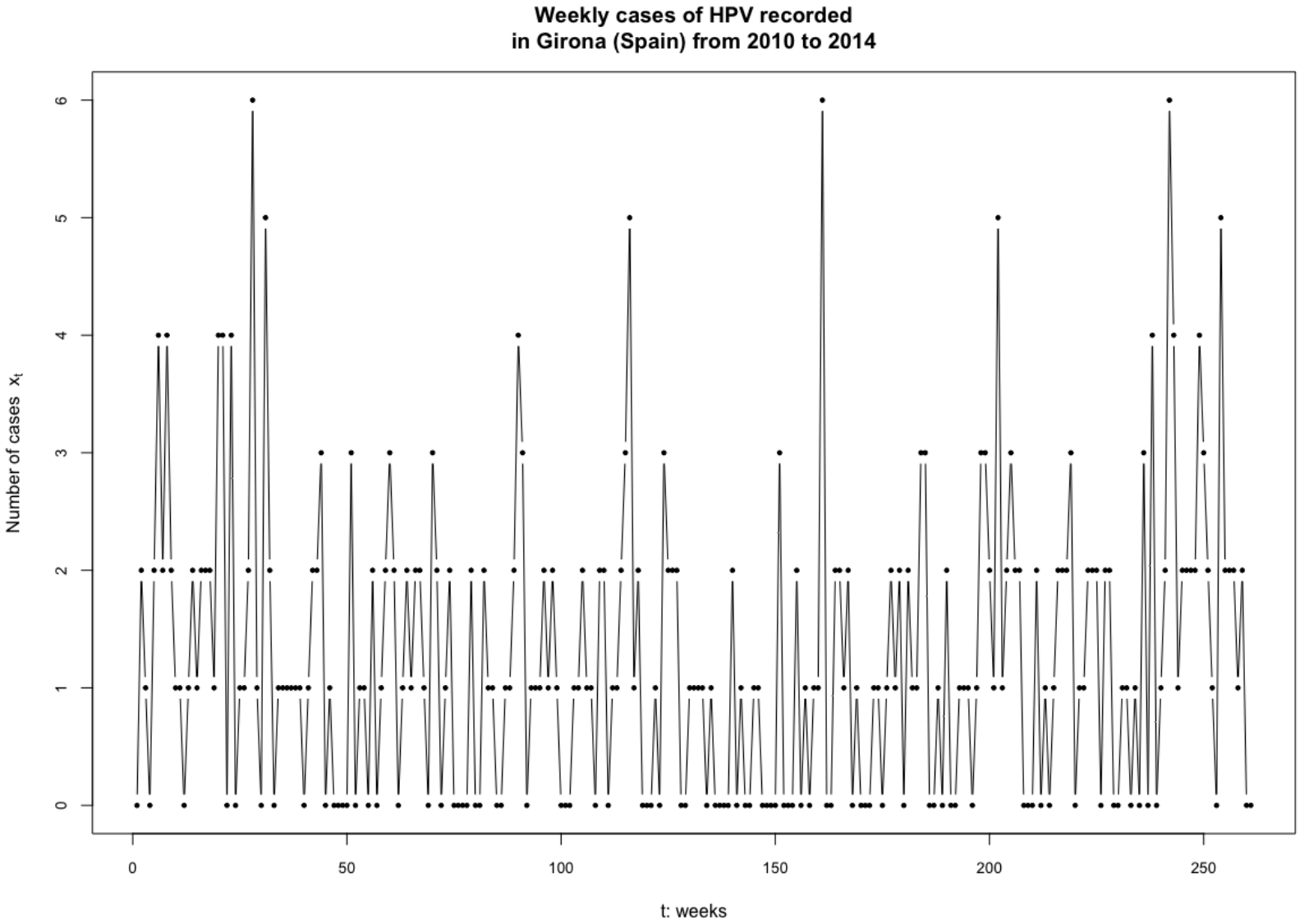
$$\ell(\boldsymbol{\theta}|X_1) = \sum_{t=2}^{T} \log p_{X_t|X_{t-1}}(\boldsymbol{\theta})$$

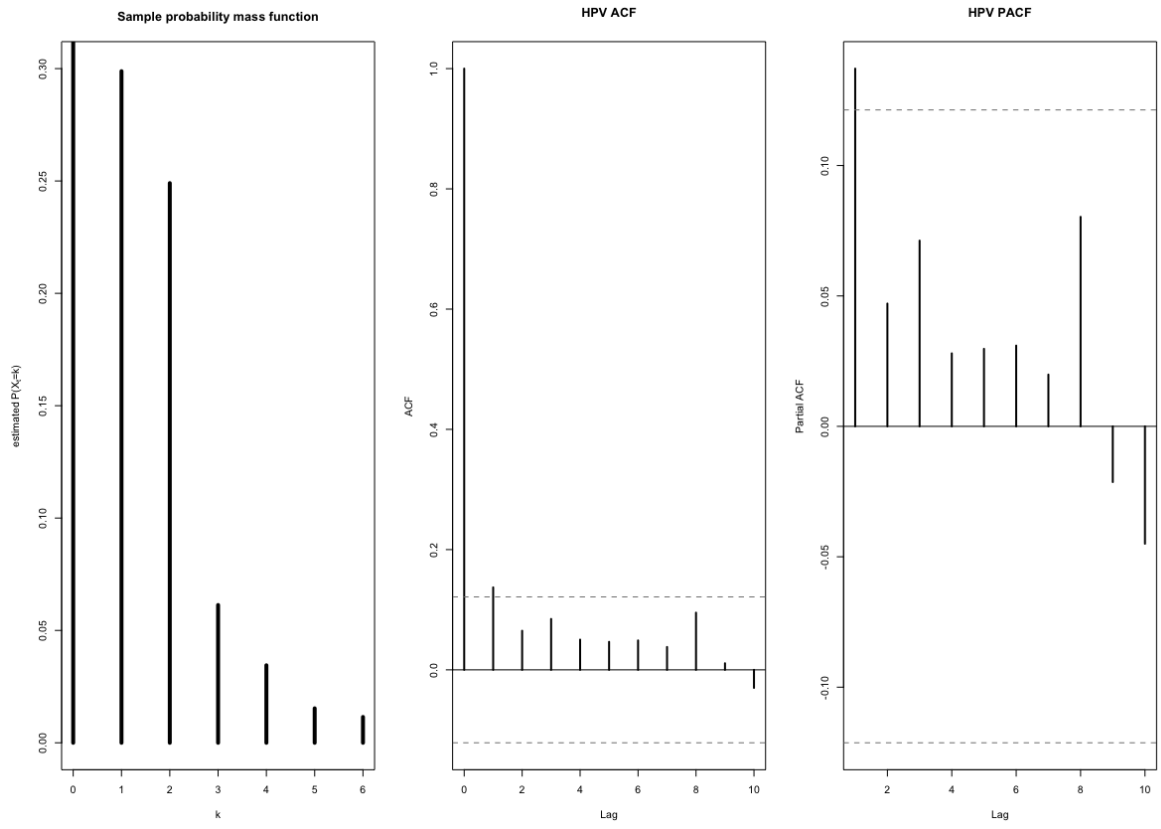
Under regularity conditions[8] $\sqrt{T}\,(\widehat{\boldsymbol{\theta}}_{CML} - \boldsymbol{\theta})$ is asymptotically Normal with zero mean and the variance is the inverse of the (expected) Fisher Information, estimated with the observed information.

[8] Billingsley (1961)

Example: HPV data

Modelling the number of weekly cases of human papillomavirus in Girona (Spain) from 2010 to 2014:





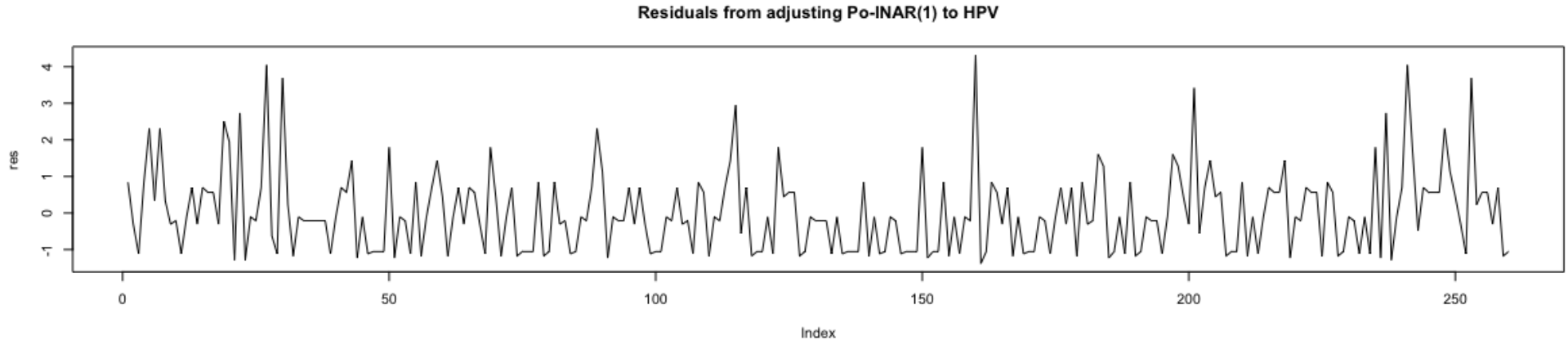
$\bar{X} = 1.26$ $\quad$ $\mathbf{Var}X = 1.60$ $\quad$ and lag-1 autocorrelation $\rho_1 = 0.14$

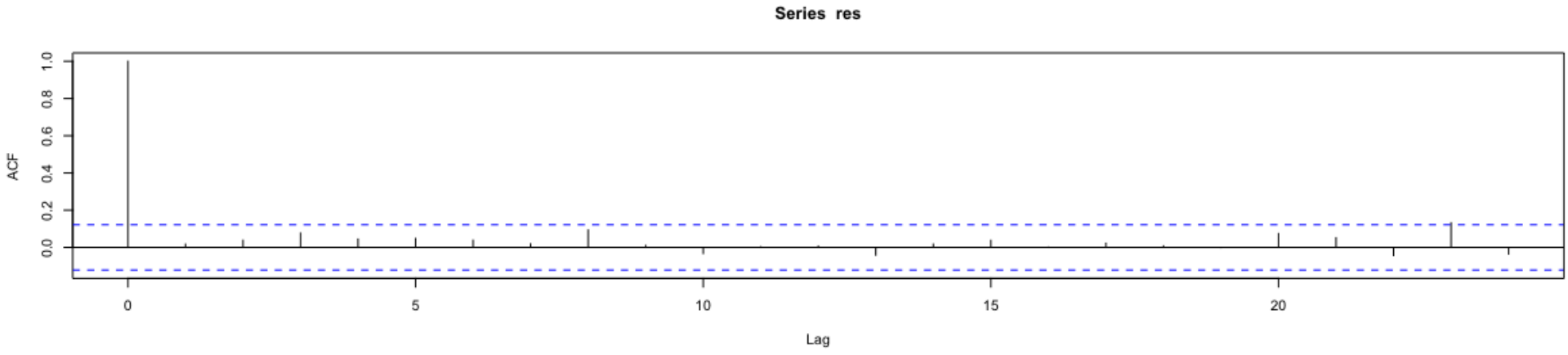
- Bootstrap confidence interval for the index of dispersion $(1.073, 1.560)$, hence the data are (slightly) over-dispersed.

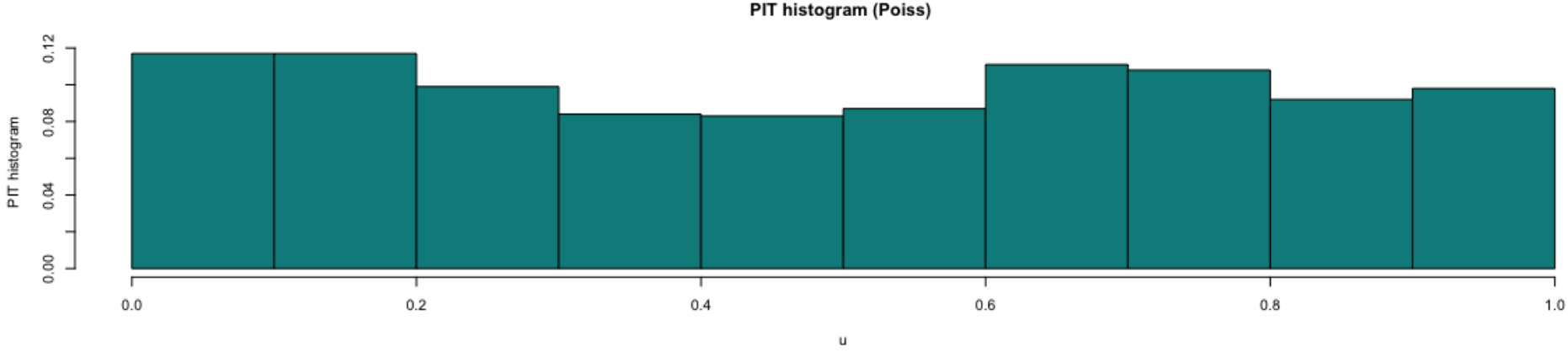
- The series can be considered stationary.
- PACFas a significant value only at lag 1, indicating that the data-generating process might have 1st-order autoregressive (AR) structure.

- Series ranges from 0 to 6 weekly cases, with a mean of 1.27 and a median of 1 case per week. The variance is 1.60. The dispersion index is 1.26 which is statistically different of 1 (p-value=0.0018): **overdispersed series**.
- Recall that a Poisson mix is always overdispersed.

```
> #Estimates:
> c(lambdaestml,alphaestml)
[1] 0.1840517 0.7623560
> #Observations' Poisson parameter:
> muestml <- lambdaestml/(1-alphaestml)
> muestml
[1] 0.7744849
> exp(-muestml) #                    probability of 0
[1] 0.4609412
> #Estimated standard errors:
> c(sqrt(diag(estcov)))
[1] 0.19099337 0.20294772 0.06345492 1.22723097
> #AIC and BIC:
> AIC <- 2*neglmax+2*2
> BIC <- 2*neglmax+log(Tlen)*2
> c(neglmax, AIC, BIC)
[1] 387.4377 778.8755 786.0045
```


Residuals from adjusting Po-INAR(1) to HPV
res
-1
0
1
2
3
4
0
50
100
150
200
250
Index


Series res
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
Lag


PIT histogram (Poiss)
PIT histogram
0.00
0.04
0.08
0.12
0.0
0.2
0.4
0.6
0.8
1.0
u

Compound Poisson innovations

The **compound Poisson** is the probability distribution of the sum of a Poisson number of iid random variables.

- CP distributions are infinitely divisible.
- When $X_1, X_2, X_3, \ldots$ are non-negative integer-valued i.i.d random variables with $\mathrm{P}(X_1 = k) = \alpha_k$, $(k = 1, 2, \ldots)$ and $N \sim$Poiss(λ), independent of X_i, define $S_N = \sum_{k=1}^{N} X_i$, then $\mathbf{E}S_N = \mathbf{E}N\mathbf{E}X$ and $\mathbf{Var}(S_N) = \mathbf{E}(N)\mathbf{E}(X^2)$ and the probability generating function of S_N is

$$g_{S_N}(z) = g_N(g_X(z)) = \exp\left(\lambda(g_X(z) - 1) = \exp\left(\sum_{k=1}^{\infty} \lambda\alpha_k(z^k - 1)\right)\right.$$

We say that $S_N \sim$CP$(\lambda\alpha_1, \lambda\alpha_2, \ldots)$.

- We call $\mathrm{CP}_\nu(\lambda, g)$ the members of the family for which $\alpha_k = 0$ for $k > \nu$, and $g(z) = \mathbf{E}z^X$.
- CP_1 is Poisson Distribution, and CP_2 Hermite Distribution.
- CP_ν-distributions with $\nu > 1$ are overdispersed.

Negative Binomial as a CP

Negative-Binomial $\mathrm{NB}(r, p)$ with $\Pr(X = k) = \binom{k+r-1}{r-1}(1-p)^k p^r$ is a Compound Poisson distribution.

Let $\{X_n, n \in \mathbb{N}\}$ denote a sequence of independent and identically distributed random variables, each one having the logarithmic distribution, with probability mass function

$$\frac{-p^k}{k \ln(1-p)}, \quad k \in \mathbb{N}$$

Let N be a random variable, independent of the sequence, and $N \sim$Poisson distribution with mean $\lambda = -r \log(1-p)$. Then the random sum

$$S_N = \sum_{n=1}^{N} X_n \sim \mathrm{NB}(r, p)$$

Compound Poisson-INAR(1)

We can define overdispersed INAR(1) processes with $\mathrm{CP}_\nu(\lambda, H)$ innovations.

The marginal distributions belong to the CP-family.

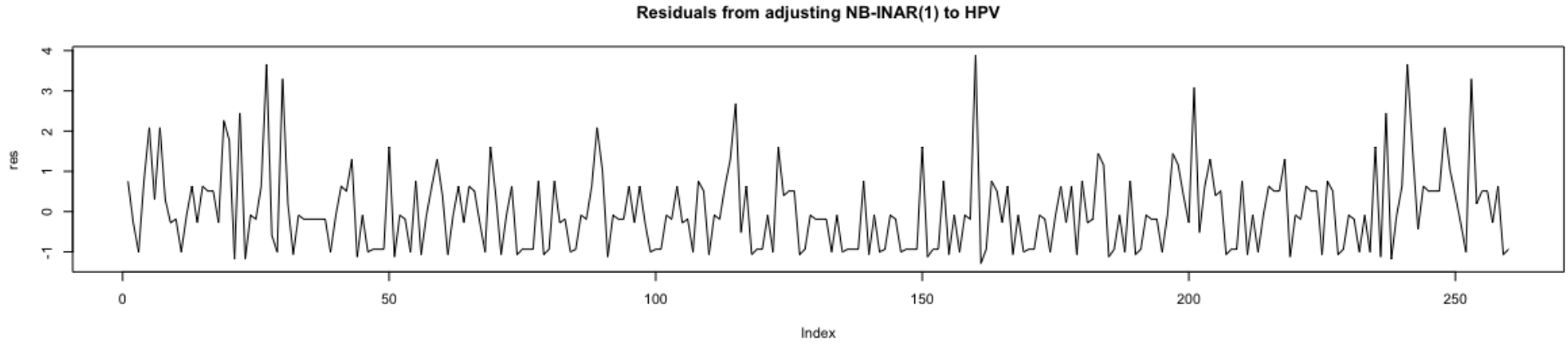
CP-INAR(1) processes are irreducible and aperiodic. They have finite mean if and only if $H'(1) < 1$, and in this case they have a unique stationary marginal distribution.

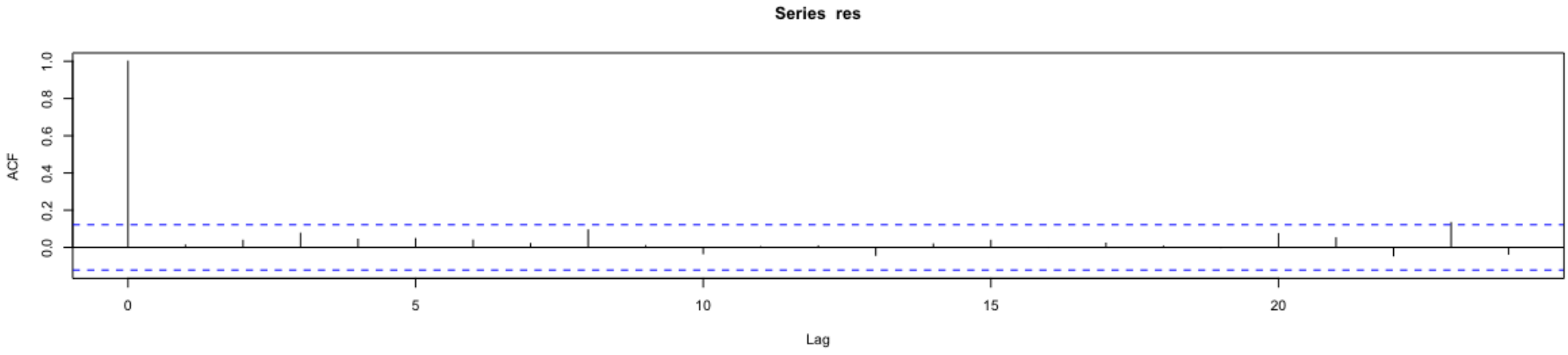
The relation to BPIs is also used to prove that CP-INAR(1) process with $H'(1) < 1$ is α-mixing with geometrically decreasing weights.

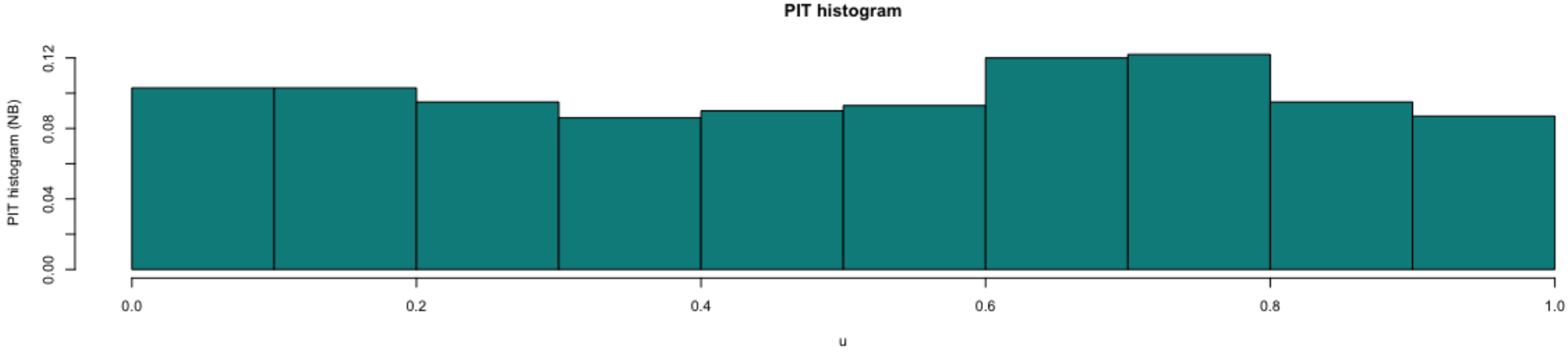
Important example: NB-INAR(1).

Let us fit a NB-INAR(1) to the HPV data, and see whether the residuals behave better than the residuals of the Po-INAR(1):

```
> #Estimates:
> c(nestml,pestml,alphaestml)
[1] 4.2157268 0.7921317 0.1239965
> #Estimated standard errors:
> c(sqrt(diag(estcov)))
[1] 2.06947335 0.07985233 0.05978602
> #AIC and BIC:
> AIC <- 2*neglmax+2*3
> BIC <- 2*neglmax+log(Tlen)*3
> c(neglmax, AIC, BIC)
[1] 388.1040 782.2080 792.9016
> #Properties of this model:
> #Innovations' mean
> nestml*(1-pestml)/pestml
[1] 1.106276
> #Observations' mean
> nestml*(1-pestml)/pestml/(1-alphaestml)
[1] 1.262867
> #Innovations' index of Dispersion:
> 1/pestml
[1] 1.262416
> #Observations' index of Dispersion:
> (1/pestml+alphaestml)/(1+alphaestml)
[1] 1.233467
> #Innovations' zero probability:
> pestml^nestml
[1] 0.3744186
> #Observations' zero probability:
> p0estml <- pestml
> for(k in c(1:nmax)){
+       p0estml <- p0estml*pestml/(1-(1-pestml)*(1-alphaestml^k))
+ }
> p0estml <- p0estml^nestml
> p0estml
[1] 0.3208593
```


Residuals from adjusting NB-INAR(1) to HPV
res
Index


Series res
ACF
Lag


PIT histogram
PIT histogram (NB)
u

INMA(q)

Counterpart to usual MA(q) model:
INMA(q) model with recursion

$$X_t \;=\; \beta_0 \circ_t \epsilon_t \;+\; \beta_1 \circ_t \epsilon_{t-1} \;+\ldots+\; \beta_{\mathsf{q}} \circ_t \epsilon_{t-\mathsf{q}}, \qquad \mathsf{q} \geq 1.$$

Time index t below '$\circ$': thinnings performed at each epoch t.

***** joint distribution of
$(\beta_0 \circ_t \epsilon_t, \beta_1 \circ_{t+1} \epsilon_t, \ldots, \beta_{\mathsf{q}} \circ_{t+\mathsf{q}} \epsilon_t)$ given ϵ_t has to be considered, leading to different models of same model order.

Four different INMA(q) models in literature,
all having different interpretations and probabilistic properties
For example,

- **INMA(q) independence model** by McKenzie (1988) assumes conditional independence, and
- **INMA(q) sale model** by Brännäs & Hall (2001) assumes conditional MULT$^*(\epsilon_t;\ \beta_0, \ldots, \beta_{\mathsf{q}})$.

Same marginal properties for all models

$$\text{pgf}(z) \;=\; \prod_{j=0}^{\text{q}} \text{pgf}_\epsilon(1-\beta_j+\beta_j z),$$

$$\mu \;=\; \mu_\epsilon\,\beta_\bullet \qquad \sigma^2 \;=\; \mu_\epsilon\,\beta_\bullet \;+\; (\sigma_\epsilon^2-\mu_\epsilon)\sum_{j=0}^{\text{q}} \beta_j^2,$$

Joint distributions differ between different INMA(q) models.

For all models, $\gamma(k)=0$ for $k>\text{q}$, in analogy to usual MA(q).

Poisson INMA(q) model:

If $\epsilon_t \sim \text{Poi}(\lambda)$,

then observations $X_t \sim \text{Poi}(\mu)$ with mean $\mu=\lambda\,\beta_\bullet$.

INAR(p)

Define INAR(p) models based on the recursion

$$X_t = \alpha_t \circ_t X_{t-1} + \cdots + \alpha_p \circ_t X_{t-p} + \varepsilon_t \quad \alpha_\bullet = \sum_{i=1}^{p} \alpha_i < 1$$

Since the thinning operations are random, the joint distribution of

$$(\alpha_1 \circ_{t+1} X_t, \ldots, \alpha_p \circ_{t+p} X_t)$$

has to be considered, leading to different types of INAR(p) models:
Alzaid and Al-Osh[9] assume a conditional multinomial distribution given X_t, **Du and Li**[10] require conditional independence of all the thinnings of X_t.
INAR(p) models are quite complex and difficult to interpret for $p \geq 2$. In addition, the choice of appropriate marginal distributions for X_t and ϵ_t is problematic.

[9] *An integer-valued pth-prder AR structure process*,JAP(1990)
[10] *The integer valued autorregressive process INAR(p)*, JTSA (1991)

- The stationary marginal mean is (in both cases) given by $\mu = \frac{\mu_\epsilon}{1-\alpha_\bullet}$.
- For the DL-INAR(p) the variance satisfies

$$\sigma^2(1-\sum_{i=1}^{p}\alpha_i\rho(i)) = \mu\sum_{i=1}^{p}\alpha_i(1-\alpha_i)+\sigma_\epsilon^2$$

and the ACF can be obtained from the (conventional) Yule-Walker equations.

- The ACF for the AA-INAR(p) is an ARMA($p, p-1$)-like correlation.
- For AA-INAR Poisson innovations lead to Poisson observations, which is not the case for DL-INAR.
- The Du-Li model is usually preferred, mixing and weak dependence properties were proven by Doukhan (2012,2013), but it seems we are far from having a definitive proposal.

Conditional Regression Models

INGARCH

Conditional Auto-regressions

INGARCH(p, q) models circumvent the multiplication problem to define ARMA-like models for stationary count time series in a different way than INARMA's thinnings. They use a **linear regression for the conditional mean**

$$M_t = \mathbf{E}(X_t | X_{t-1}, X_{t-2}, \dots)$$

For instance, an AR(1)-like structure is transferred to the conditional mean as

$$M_t = \beta_0 + \alpha_1 X_{t-1}$$

Then, the count time series is generated by using a - for instance - Poisson distributions, *i.e.*

$$X_t \sim \text{Poiss}(M_t)$$

thus, the outcomes are always integers.

Po-INARCH(p)

Po-INARCH(p) model for X_t assume that the conditional distribution of X_t given its past is Poiss(M_t), where

$$M_t = \beta_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}$$

with $\alpha_\bullet = \sum_i \alpha_i < 1$.
It is a $p-$th order Markov model. In particular, for $p = 1$, $M_t = \beta + \alpha X_{t-1}$ the transition probabilities (needed for ML estimation) are

$$p_{k|l} = exp\{-\beta - \alpha l\} \frac{(\beta + \alpha l)^k}{k!}$$

Conditional mean and variance coincide, and are both linear on the previous observation: $\beta + \alpha X_{t-1}$.

INARCH(1) is a stationary ergodic and $\alpha-$mixing Markov Chain[11], all its moments are finite, and in particular

$$\mu = \frac{\beta}{1-\alpha}, \quad \sigma^2 = \frac{\beta}{(1-\alpha)(1-\alpha^2)} \Rightarrow I = \frac{1}{1-\alpha^2} > 1, \quad \rho(k) = \alpha^k$$

No closed formula for the stationary marginal distribution.

INAR(1) and INARCH(1) are very similar: choosing the same α and $\beta = \lambda$, they have the same marginal mean and autocorrelations.

BUT, Po-INAR(1) is unconditionally equidispersed, while PO-INARCH(1) has overdispersion increasing with α. And usually sample paths show more extreme counts.

[11] Neumann (2011)

INGARCH(p, q)

Just as GARCH models extend ARCH by adding information on previous conditional variances, INGARCH models are defined with feedback from the previous conditional means:
$(X_t)_{t\in\mathbb{Z}}$ is a Po-INGARCH(p, q) model[12], with $p \geq 1$, $q \geq 0$ if

1. X_t conditioned on $X_{t-1}, X_{t-2}, \dots$ is Poisson distributed, with mean M_t
2. $M_t = \mathbf{E}(X_t | X_{t-1}, X_{t-2}, \dots) = \sum_{i=1}^{p} \alpha_i X_{t-i} + \beta_0 + \sum_{j=1}^{q} \beta_j X_{t-j}$, with $\beta_0 > 0$, $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q \geq 0$.

Even though the conditional distribution is Poisson,

$$\mathbf{E}X_t = \mathbf{E}M_t \quad \text{and} \quad \mathbf{Var}X_t = \mathbf{E}X_t + \mathbf{Var}M_t > \mathbf{E}X_t$$

so, Po-INGARCH are suitable models for overdispersed data.

[12] Ferland et.al. (2006), Fokianos et.al (2009)

It can be shown[13] that if $\alpha_\bullet + \beta_\bullet < 1$, there is a strictly stationary solution to $M_t = \beta_o + \sum_i \alpha_i X_{t-1} + \sum_j \beta_j M_{t-j}$ with mean $\mu = \frac{\beta_0}{1 - \alpha_\bullet - \beta_\bullet}$ and variances and covariances determined by a set of Yule-Walker-like equations.

$$\gamma(k) = \sum_{i=1}^{p} \alpha_i \, \gamma(|k-i|) + \sum_{j=1}^{\min\{k-1,q\}} \beta_j \, \gamma(k-j) + \sum_{j=k}^{q} \beta_j \, \gamma_M(j-k),$$

$$\gamma_M(l) = \sum_{i=1}^{\min\{l,p\}} \alpha_i \, \gamma_M(|l-i|) + \sum_{i=l+1}^{p} \alpha_i \, \gamma(i-l) + \sum_{j=1}^{q} \beta_j \, \gamma_M(|l-j|),$$

where $\gamma(k) = \text{Cov}(X_t, X_{t-k})$ and $\gamma_M(k) = \text{Cov}(M_t, M_{t-k})$

[13] Ferland et.al.(2006)

NB-INGARCH models

The INGARCH model can also be combined with non Poisson conditional distributions.
For example, more overdispersion can be achieved by using NB(n_t, p) with $n_t = M_t \frac{p}{1-p}$.
A NB-INGARCH(1,1) satisfies

$$\sigma^2 = \frac{\mu}{p(1-\alpha)}$$

Comparing INARMA and INGARCH

- Both are ARMA-like approaches for modelling stationary count time series.

- Its pure AR members have identical autocorrelation structure
- Marginal properties are easier to derive for INARMA.
- Conditional distributions are simpler for INGARCH. For instance, transition probabilities of a Po-INGARCH(p) are Poisson probabilities, while for INAR(p), they involve cumbersome convolution computations.
- INARMA include pure MA-type models, while INGARCH don't. In fact, the MA part in INGARCH causes some sort of **long memory**, because the current observation is affected by all previous information. For INGARCH(1,1)[14]

$$M_t = \alpha_1 \sum_{k=1}^{t} \beta_1^{k-1} X_{t-k} + \beta_1^t M_0 + \beta_0 \frac{1-\beta_1^t}{1-\beta_1}$$

- Both methodologies can be adapted to non-stationary data.

[14] Fokianos (2012)

HPV data

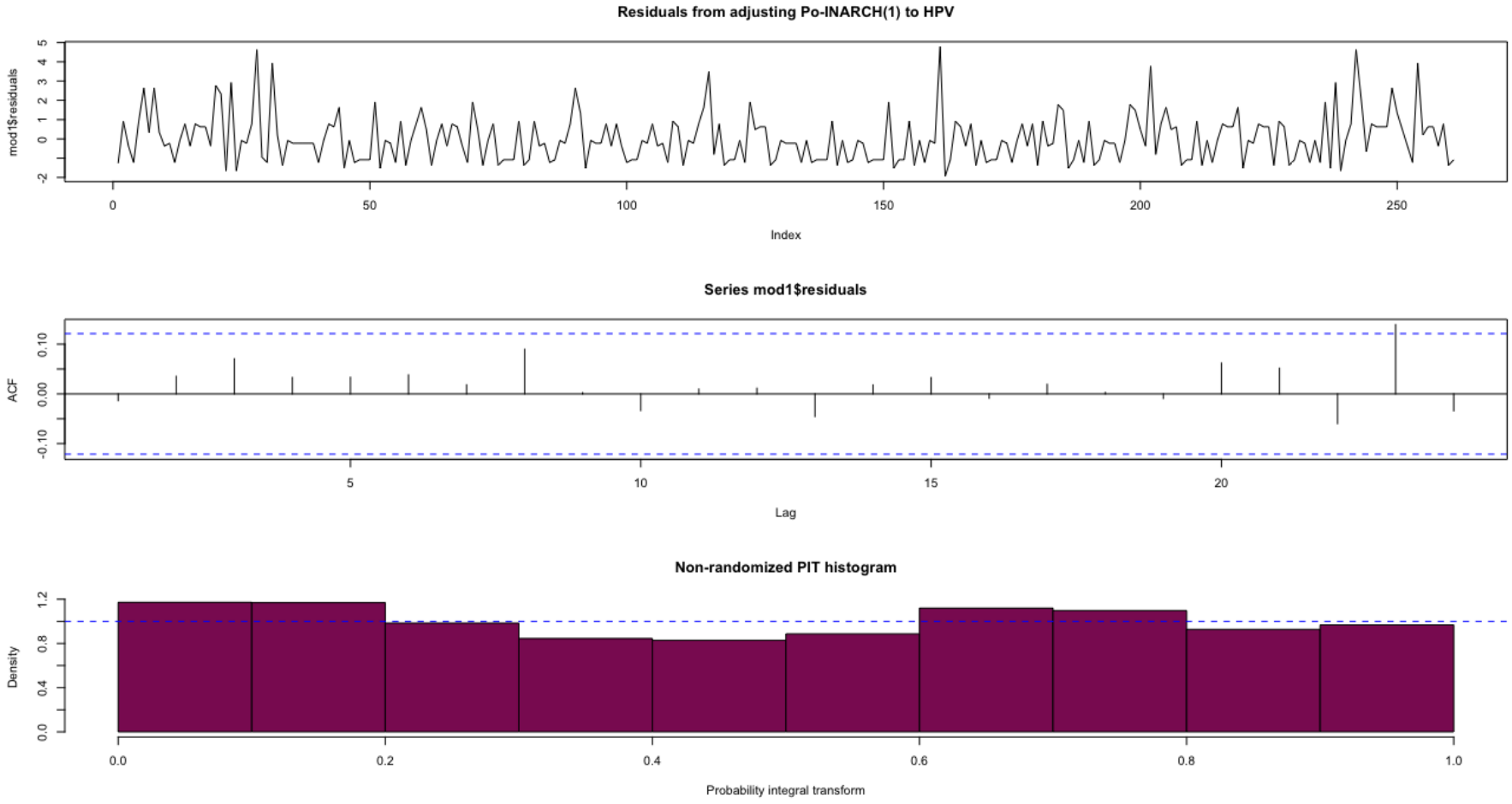
INGARCH and in general GLM- conditional regressions can be fitted using the R library `tscount`, but beware, the parameters have different names! Coming back to the HPV data we have analized before, the adjusted po-INARCH(1) results

```
> mod1 <- tsglm(data, model = list(past_obs = 1), distr="poisson")
> summary(mod1)

Call:
tsglm(ts = data, model = list(past_obs = 1), distr = "poisson")

Coefficients:
             Estimate  Std.Error  CI(lower)  CI(upper)
(Intercept)     1.081     0.0950     0.8949       1.27
beta_1          0.144     0.0589     0.0291       0.26
Standard errors and confidence intervals (level =  95 %) obtained
by normal approximation.

Link function: identity
Distribution family: poisson
Number of coefficients: 2
Log-likelihood: -390.6732
AIC: 785.3463
BIC: 792.4754
QIC: 785.2738
```


Residuals from adjusting Po-INARCH(1) to HPV
mod1$residuals
-2
0
1
2
3
4
5
0
50
100
150
200
250
Index
Series mod1$residuals
ACF
-0.10
0.00
0.10
5
10
15
20
Lag
Non-randomized PIT histogram
Density
0.0
0.4
0.8
1.2
0.0
0.2
0.4
0.6
0.8
1.0
Probability integral transform

An INGARCH(1,1) with NB innovations produces smaller AIC/BIC

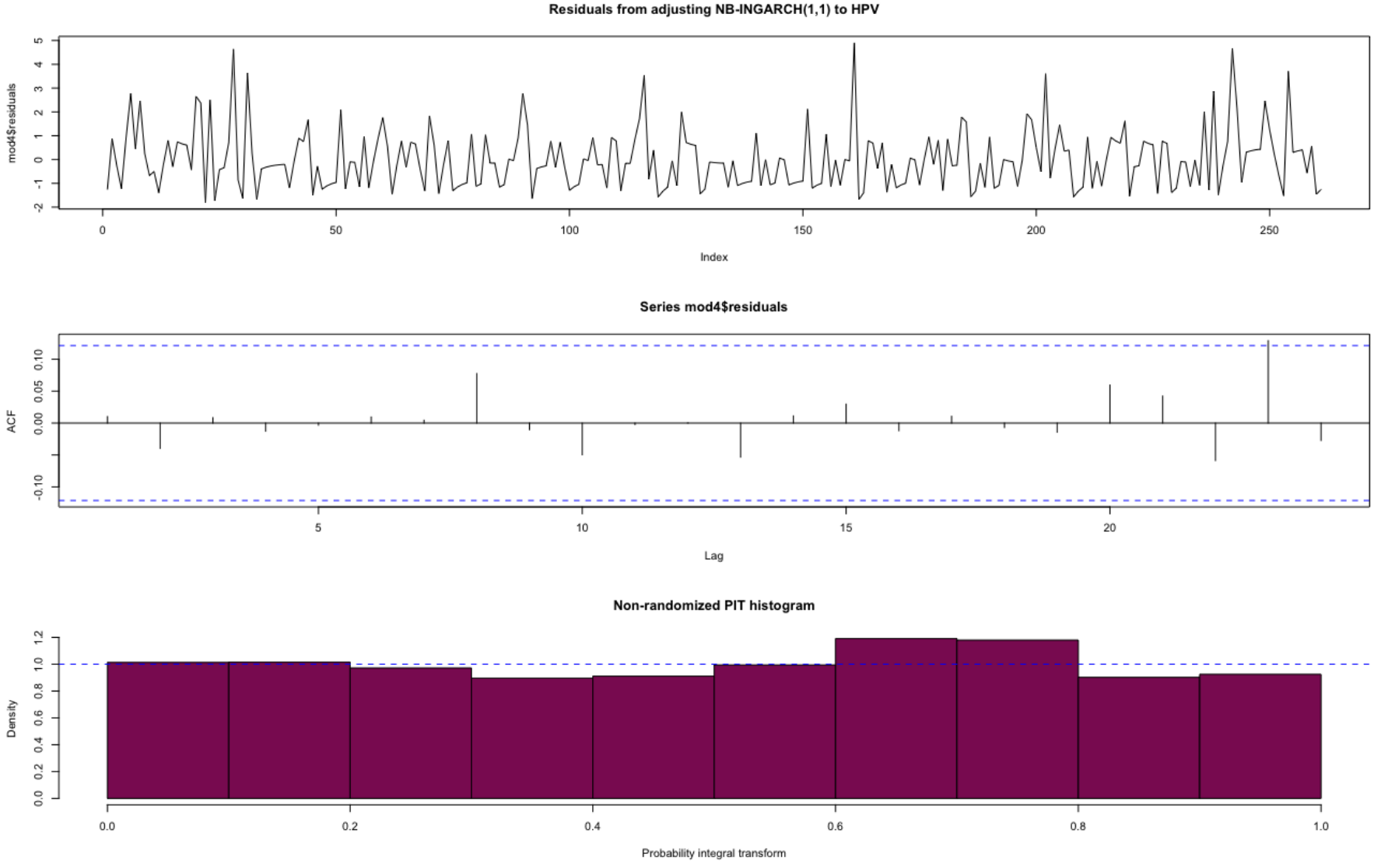
```
> summary(mod4)

Call:
tsglm(ts = data, model = list(past_obs = 1, past_mean = 1), distr = "nbinom")

Coefficients:
             Estimate  Std.Error  CI(lower)  CI(upper)
(Intercept)     0.276      0.218  -0.150092      0.703
beta_1          0.112      0.057  -0.000098      0.223
alpha_1         0.669      0.207   0.264180      1.074
sigmasq         0.186         NA         NA         NA
Standard errors and confidence intervals (level =  95 %) obtained
by normal approximation.

Link function: identity
Distribution family: nbinom (with overdispersion coefficient 'sigmasq')
Number of coefficients: 4
Log-likelihood: -386.4464
AIC: 780.8928
BIC: 795.1508
QIC: 784.7853
```

In this case, a Negative Binomial conditional distribution parametrized with $M_t = n\frac{1-p_t}{p_t}$ is used.


Residuals from adjusting NB-INGARCH(1,1) to HPV
mod4$residuals
-2
-1
0
1
2
3
4
5
0
50
100
150
200
250
Index
Series mod4$residuals
ACF
-0.10
0.00
0.05
0.10
5
10
15
20
Lag
Non-randomized PIT histogram
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0.0
0.2
0.4
0.6
0.8
1.0
Probability integral transform

Generalizations

The previous models have many generalizations[15].

Log-linear Poisson autorregression[16]

$$\log(M_t) = \beta_0 + \alpha_1 \log(X_{t-1} + 1) + \beta_1 \log(M_{t-1})$$

The parameters can be negative, but some restrictions apply in order to guarantee the existence of a stationary distribution.

Seasonal log-linear model[17]

$$\log(M_t) = \overbrace{\gamma_0 + \gamma_1 t/T + \sum_{s=1}^{2}(\gamma_{2s}\cos(s\omega t) + \gamma_{2s+1}\sin(s\omega t))}^{=:\log \mu_{0,t}(\gamma)} + \sum_{i=1}^{2} \alpha_i(\log(X_{t-i} + 1) - \log(\mu_{0,t-i}(\gamma) + 1)$$

where $\omega = 2\pi/T$, T is the period.

---

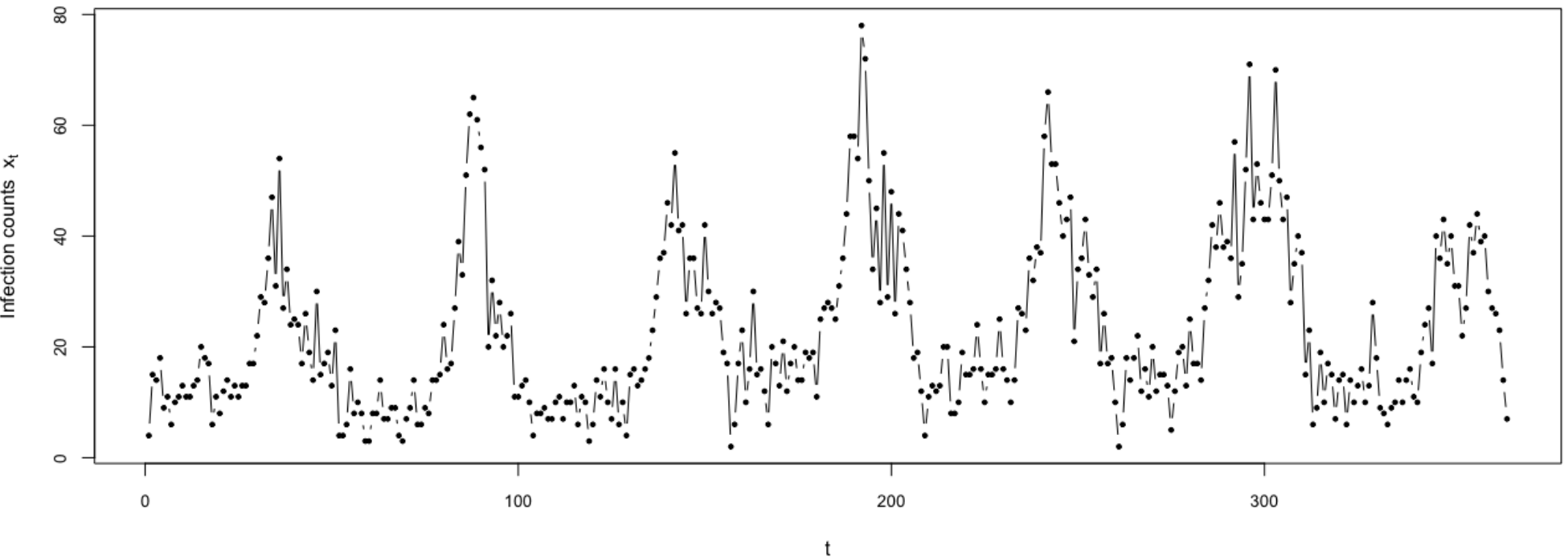
[15] see the books by C. Weiß, (2018), and (forthcoming) by K. Fokianos

[16] Fokianos and Tjotsheim (2011)

[17] Zeger & Qaqish (1988)

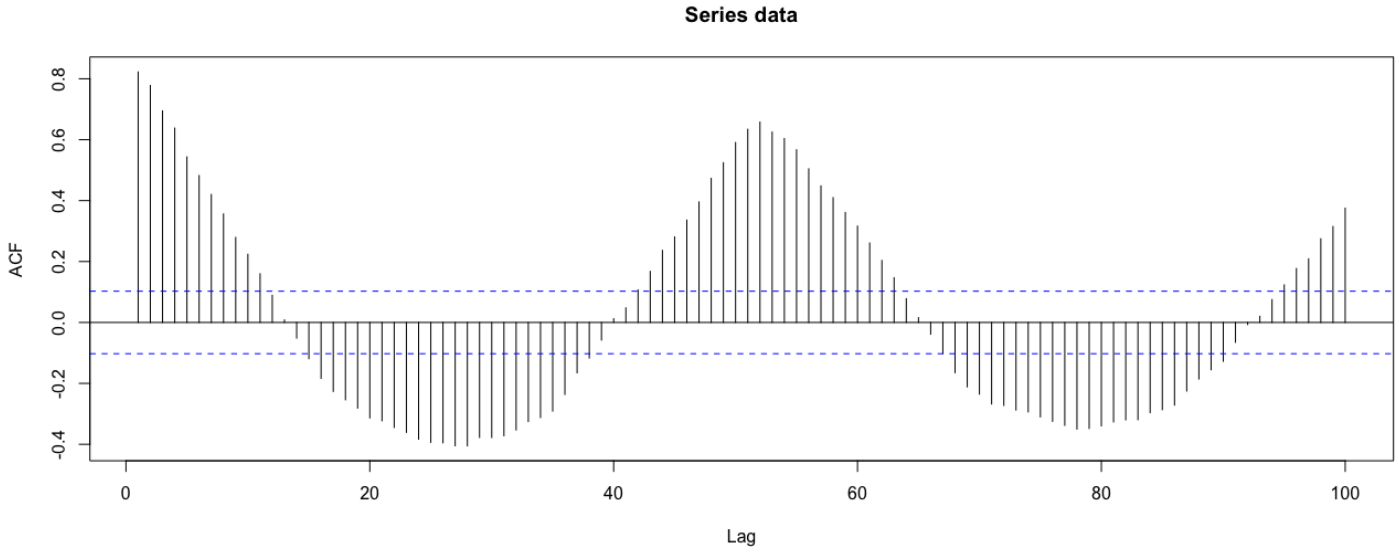
Cryptosporidiosis data

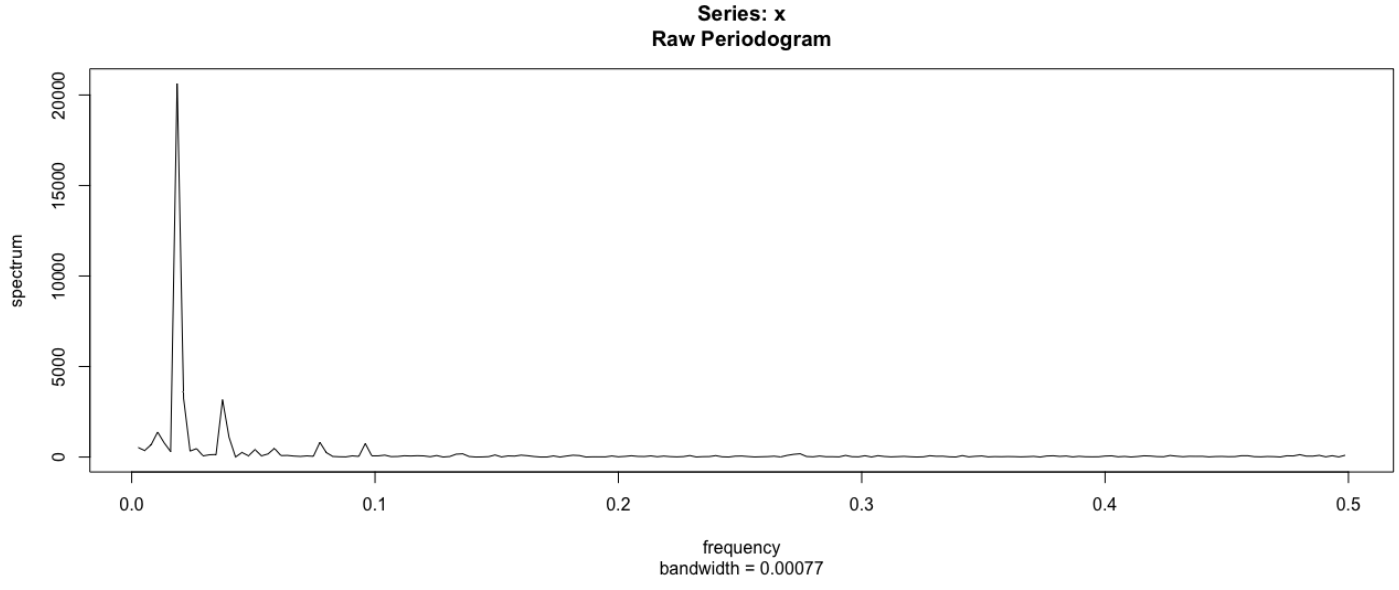
Weekly counts of new infections by cryptosporidiosis (causing watery diarrhea) in Germany in 2002-2008 [18].



[18] Source: Robert-Koch-Institut (2016) & C.Weiß

Cryptosporidiosis data





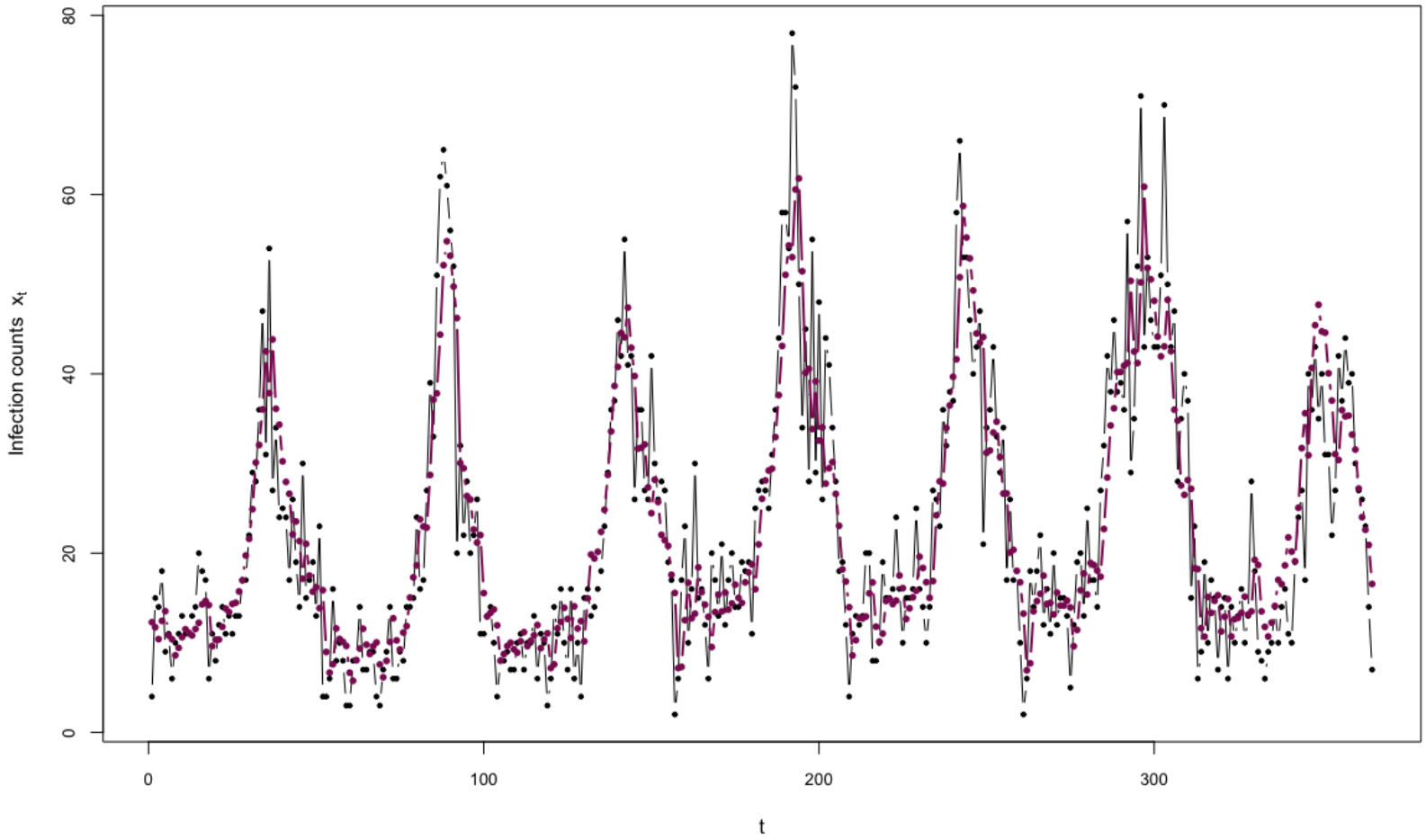
A Poisson Seasonal log-linear model is fitted,

$$\ln M_t = \overbrace{\gamma_0 + \gamma_1\, t/T + \sum_{s=1}^{2} \left(\gamma_{2s}\cos(s\,\omega t) + \gamma_{2s+1}\sin(s\,\omega t)\right)}^{=:\ \ln \mu_{0,t}(\boldsymbol{\gamma})} + \sum_{i=1}^{2} \alpha_i \left(\ln\left(X_{t-i}+1\right) - \ln\left(\mu_{0,t-i}(\boldsymbol{\gamma})+1\right)\right).$$

with

parameter	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5	α_1	α_2
estimate	2.83	0.42	-0.11	-0.64	-0.15	0.05	0.41	0.18
se	0.06	0.09	0.04	0.04	0.03	0.03	0.04	0.03

Cryptosporidiosis data and fitted means




```
> #Pearson residuals
> res <- (data-fittedmeans)/sqrt(fittedmeans)
> Box.test(res,lag=52)

        Box-Pierce test

data:  res
X-squared = 58.41, df = 52, p-value = 0.2517
```