Dynamic INAR Processes[1]

Hedibert F. Lopes, Insper

Workshop on Risk Analysis and Applications

IME-USP, September 2025

[1] Joint with Refik Soyer and Di Zhang, GWU. FAPESP grant 2023/02538-0.

Outline

Time series of counts

Time series of counts arise in many areas such as business, economics, engineering, and medicine. Applications include, among others, the modeling of the

- Number of deaths from a specific disease in a given month (Schmidt and Pereira, 2011)
- Number of arrivals to a call center (Aktekin and Soyer, 2011)
- Number of monthly mortgage defaults (see Aktekin et al., 2013)
- Crash counts in different regions (Hu et al., 2013)
- Number of accidents in a given time interval (Serhiyenko et al., 2014)
- Number of weekly shopping trips of households (Aktekin et al., 2018)
- Network flows (Chen et al., 2019).

Recent advances in Bayesian modeling and computation for count time series are presented and discussed in Soyer and Zhang (2021).

Our contributions

- Dynamic version of the integer autoregressive (INAR) processes for count data.
- Unification of models to describe temporal correlations in univariate count time series.
- Derivation of a customized MCMC scheme.
- Derivation of a customized particle filtering scheme.

Static INAR(1) models

For a stationary count time series Y_t, the INAR(1) model is defined as

$$Y_t = \alpha \circ Y_{t-1} + \epsilon_t \tag{1}$$

where "$\circ$" is the binomial thinning operation represented by

$$M_t \alpha \circ Y_{t-1} = \sum_{j=1}^{Y_{t-1}} B_{jt} \tag{2}$$

where B_{jt}'s are Bernoulli random variables with probability α. It is easy to see that

$$M_t | Y_{t-1}, \alpha \sim Binom(Y_{t-1}, \alpha).$$

B_{jt}'s are independent of the ϵ_t's.

Connection to branching process

As pointed out by Weiß (2008), the INAR(1) process can be interpreted as a special case of immigration branching processes.

Y_t is the population at time t which consists of two components:

- M_t, those who survive from time $(t-1)$ with probability α, and
- ϵ_t, those who arrive at the beginning of time t.

Basic INAR(1)

If $\{\epsilon_t\}$ is a sequence of i.i.d. Poisson random variables with rate θ in (1), then the model is referred to as an INAR(1) process with Poisson errors.

Alternatively, the error distribution can be assumed as geometric or negative binomial as discussed in Weiß (2008).

Finally, one can consider a mixture of Poisson and geometric as in Marques, Graziadei, and Lopes (2022).

If Y_0 is assumed to be Poisson with rate $\theta/(1-\alpha)$, then it can be shown that Y_t's is a stationary Poisson series with parameter $\theta/(1-\alpha)$.

Autocorrelation functions

The autocorrelation function of the process (McKenzie, 1988) is given by

$$\rho_Y(k) = \alpha^k \quad \text{for} \quad k > 0,$$

Convolution of a binomial and a Poisson

$$p(Y_t|Y_{t-1},\theta,\alpha) = \sum_{j=0}^{min(Y_{t-1},Y_t)} \frac{e^{-\theta}\,\theta^{Y_t-j}}{(Y_t-j)!}\binom{Y_{t-1}}{j}\alpha^j(1-\alpha)^{Y_{t-1}-j}, \tag{3}$$

with $E[Y_t|Y_{t-1},\theta,\alpha] = \alpha Y_{t-1} + \theta$.

The k-step ahead forecast distribution, $p(Y_{t+h}|Y_t,\,\theta,\alpha)$, can be obtained. Also,

$$E[Y_{t+h}|Y_t,\theta,\alpha] = \alpha^h\left(Y_t - \frac{\theta}{(1-\alpha)}\right) + \frac{\theta}{(1-\alpha)}. \tag{4}$$

Dynamic arrival rates

INAR(1) process with Poisson errors by allowing rates of the ϵ_t's to vary over time in (1):

$$Y_t - M_t | M_t, \theta_t \sim Poisson(\theta_t), \tag{5}$$

with

$$M_t | Y_{t-1}, \alpha \sim Bin(Y_{t-1}, \alpha).$$

Markov evolution: Scaled beta over $(0, \theta_{t-1}/\gamma)$

$$\theta_t | \theta_{t-1}, M^{(t-1)}, Y^{(t-1)} \sim Beta(\gamma a_{t-1}, (1-\gamma)a_{t-1}, \theta_{t-1}/\gamma), \tag{6}$$

Alternatively,

$$\theta_t = \frac{\theta_{t-1}}{\gamma} w_t \tag{7}$$

where $w_t | Y^{(t-1)}, M^{(t-1)} \sim Beta[\gamma a_{t-1}, (1-\gamma)a_{t-1}]$.

West and Harrison's discount factor

The quantity $0 < \gamma < 1$ is a discount factor in the sense of West and Harrison (1997).

w_t can be considered as a system error term with mean γ and (7) provides a random walk-type evolution for θ_t.

The Markov evolution (7) was also considered in Aktekin and Soyer (2011) and in Gamerman et al. (2013) for parameter driven non-Gaussian time series.

In our proposed dynamic INAR(1) process, equations (5) and (7) will play the roles of the observation equation and the system equation, respectively.

Dynamic evolution and updating

We complete the definition of our model by assuming that, at time $t-1$, the rate θ_{t-1} follows a gamma distribution

$$\theta_{t-1}|M^{(t-1)},\ Y^{(t-1)} \sim Gamma(a_{t-1}, b_{t-1}).$$

It follows from the above that the forecast distribution of θ_t at $t-1$ is given by

$$\theta_t|M^{(t-1)}, Y^{(t-1)} \sim Gamma(\gamma a_{t-1}, \gamma b_{t-1}). \tag{8}$$

Using the Bayes' rule

$$p(\theta_t|M^{(t)}, Y^{(t)}) \propto p(\theta_t|M^{(t-1)}, Y^{(t-1)})p(Y_t - M_t|M_t, \theta_t),$$

it can be shown that the posterior (or filtering) distribution of θ_t can be obtained as

$$\theta_t|M^{(t)}, Y^{(t)} \quad \sim \quad Gamma(a_t, b_t), \tag{9}$$

for $a_t = \gamma a_{t-1} + (Y_t - M_t)$ and $b_t = \gamma b_{t-1} + 1$.

Additional properties

We note that starting at time 0 with prior $\theta_0 \sim Gamma(a_0, b_0)$, the proposed model provides us with a conjugate update of the dynamic rates, which is attractive for developing posterior and predictive Bayesian inferences.

Other attractive features of the model are its ability to deal with over-dispersed count time series and the inclusion of some of the previously models as special cases.

$$
\begin{aligned}
p(\epsilon_t|\gamma, M^{(t)}, Y^{(t-1)}) &= \int_0^\infty p(\epsilon_t|\theta_t, M_t)\; p(\theta_t|\gamma, M^{(t)}, Y^{(t-1)})d\theta_t, \\
&= \frac{\Gamma(\gamma a_{t-1} + \epsilon_t)}{\Gamma(\epsilon_t + 1)\,\Gamma(\gamma a_{t-1})}\left(1 - \frac{\gamma b_{t-1}}{\gamma b_{t-1} + 1}\right)^{\epsilon_t}\left(\frac{\gamma b_{t-1}}{\gamma b_{t-1} + 1}\right)^{\gamma a_{t-1}},
\end{aligned}
$$

which is a negative binomial for integer values of γa_{t-1}, which can handle over-dispersed counts.

Gibbs sampler

Bayesian inference is possible via a customized Gibbs sampler.

Let the count time series be $Y^{(T)} = (Y_1, \ldots, Y_T)$ for T time periods.

We will derive the posterior full conditionals:

- α
- γ
- $\theta^{(T)} = (\theta_1, \ldots, \theta_T)$
- $M^{(T)} = (M_1, \ldots, M_T)$

The trickiest one is $p(M^{(T)}|\alpha, \gamma, \theta^{(T)}, Y^{(T)})$.

Learning $M^{(T)}$

Let $M_1 = 0$, the joint distribution $p(\theta_1, Y_1, \theta_2, Y_2, M_2, \ldots, \theta_T, Y_T, M_T)$ is

$$p(\theta_1|D_0)p(Y_1|\theta_1)\prod_{t=2}^{T} p(\theta_t|\theta_{t-1}, M^{(t-1)}, Y^{(t-1)})p(M_t|Y_{t-1})p(Y_t|M_t, \theta_t), \tag{10}$$

so the full conditional for each M_t can be obtained as

$$\begin{aligned} p(M_t|M^{(-t)}, \theta^{(T)}, Y^{(T)}) &\propto \left\{\prod_{s=t+1}^{T} p(\theta_s|\theta_{s-1}, M^{(s-1)}, Y^{(s-1)})\right\} \\ &\times p(M_t|Y_{t-1})p(Y_t|M_t, \theta_t), \end{aligned} \tag{11}$$

where $t = 2, \ldots, T-1$ and $M^{(-t)} = \{M_s; s \neq t\}$ and

$$(\theta_s|\theta_{s-1}, M^{(s-1)}, Y^{(s-1)}) \sim Beta[\gamma a_{s-1}, (1-\gamma)a_{s-1}]\mathcal{I}(\theta_s < \theta_{s-1}/\gamma). \tag{12}$$

Computational burden

For $s = (t+1), (t+2), \ldots, T$, the values a_{s-1} should be evaluated at the given value of M_t.

For example, when $s = (t+1)$, we have $a_t = \gamma a_{t-1} + (Y_t - M_t)$. Therefore, for each value of M_t, we need to evaluate a_t, a_{t+1},. . ., a_T and these values will be different than the values in updating of θ_t's.

For $t = T$, we can write

$$p(M_T | M^{(-T)}, \theta^{(T)}, Y^{(T)}) \propto p(M_T | Y_{T-1}) p(Y_T | M_T, \theta_T).$$

Learning $\theta^{(T)}$ via FFBS

We will use the forward filtering backward sampling (FFBS) algorithm of Frühwirth-Schnatter (1994) to jointly sample from

$$p(\theta_1, \ldots, \theta_T | M^{(T)}, \gamma, Y^{(T)}).$$

The joint full conditional can be written as

$$p(\theta_T | M^{(T)}, Y^{(T)}, \gamma) p(\theta_{T-1} | \theta_T, M^{(T-1)}, Y^{(T-1)}, \gamma) \cdots \; p(\theta_1 | \theta_2, Y_1, \gamma), \tag{13}$$

where $p(\theta_{t-1} | \theta_t, M^{(t-1)}, Y^{(t-1)}, \gamma)$ is a shifted gamma density over $(\gamma\theta_t, \infty)$, denoted as

$$Gamma((1-\gamma)a_{t-1},\, b_{t-1})\mathcal{I}(\theta_{t-1} > \gamma\theta_t).$$

Learning α and γ

The full conditional of α can be obtained as

$$\alpha|M^{(T)}, Y^{(T)} \sim Beta\left(a_\alpha + \sum_{t=2}^{T} M_t, b_\alpha + \sum_{t=2}^{T}(Y_{t-1} - M_t)\right). \tag{14}$$

It is straightforward to see that the full conditional of γ can be obtained as

$$p(\gamma|\theta^{(T)}, M^{(T)}, Y^{(T)}) \propto \left\{\prod_{t=2}^{T} p(\theta_t|\theta_{t-1}, M^{(t-1)}, Y^{(t-1)}, \gamma)\right\} p(\gamma), \tag{15}$$

where $p(\theta_t|\theta_{t-1}, M^{(t-1)}, Y^{(t-1)}, \gamma)$ is the scaled beta distribution given by (6). Then, we can use a Metropolis step or by using a discrete prior for γ.

Particle filtering

MCMC schemes are computationally expensive for online Bayesian learning and forecasting.

Sequential Monte Carlo (SMC) methods have been proposed to alleviate such computational inefficiencies.

An important class of SMC methods are particle filters (PF), that was originally proposed by Gordon et al. (1993); see Lopes and Tsay (2011) for a review of PF and Singpuwalla et al. (2018) for a historical perspective.

Learning about static parameters in PF is not trivial, but if (conditional) sufficient statistics are available, one can develop an efficient recursive updating scheme called *Particle Learning* (PL). See Carvalho et al. (2010) and Lopes et al. (2011) for further details.

We develop PL for the dynamic INAR(1) model.

Particle learning

Particles at $t-1$:

$$\{\theta_{t-1}^i, M_{t-1}^i, \alpha^i\}_{i=1}^N$$

Sample from $p(M_t|Y^{(t-1)})$ by sampling M_t^i from $Bin(Y_{t-1}, \alpha^i)$, for $i = 1, \ldots, N$.

Resample θ_{t-1} with weights given by the predictive

$$p(Y_t|M_t, \theta_{t-1}, M^{(t-1)}, Y^{(t-1)}) = \int p(Y_t|M_t, \theta_t)p(\theta_t|\theta_{t-1}, M^{(t-1)}, Y^{(t-1)})d\theta_t$$

where

$$\begin{array}{rcl}(Y_t|M_t, \theta_t) & \sim & Poi(\theta_t)\mathcal{I}(Y_t \geq M_t) \\ (\theta_t|\theta_{t-1}, M^{(t-1)}, Y^{(t-1)}) & \sim & Beta(\gamma a_{t-1}, (1-\gamma)a_{t-1}, \theta_{t-1}/\gamma).\end{array}$$

The confluent hypergeometric negative binomial distribution

$p(Y_t|M_t, \theta_{t-1}, M^{(t-1)}, Y^{(t-1)})$ is a confluent hypergeometric negative binomial distribution

$$\int_0^{\theta_{t-1}/\gamma} \frac{\theta_t^{(Y_t-M_t)}\, e^{-\theta_t}}{(Y_t-M_t)!}\xi(\gamma, a_{t-1})(\gamma/\theta_{t-1})^{a_{t-1}-1}\, \theta_t^{\gamma a_{t-1}-1} \left(\frac{\theta_{t-1}}{\gamma} - \theta_t\right)^{(1-\gamma)a_{t-1}-1} d\theta_t,$$

where $\xi(\gamma, a_{t-1}) = Beta\,(\gamma a_{t-1}, (1-\gamma)a_{t-1})$.

Also, by using change of variable $u_t = (\gamma/\theta_{t-1})\theta_t$, we can write the above integral as

$$\kappa(Y_t, M_t, \theta_{t-1}, \gamma, a_{t-1}) \int_0^1 e^{-\frac{\theta_{t-1}}{\gamma}u_t}\, u_t^{\gamma a_{t-1}+(Y_t-M_t)-1}(1-u_t)^{(1-\gamma)a_{t-1}-1} du_t,$$

where

$$\kappa(Y_t, M_t, \theta_{t-1}, \gamma, a_{t-1}) = \frac{(\theta_{t-1}/\gamma)^{(Y_t-M_t)}}{(Y_t-M_t)!}\xi(\gamma, a_{t-1}).$$

The integral in the above expression represented as

$$Beta(Y_t - M_t + \gamma a_{t-1}, (1-\gamma)a_{t-1}) CHF(a, a+b, -c),$$

where $CHF(a, a+b, -c)$ is the confluent hyper-geometric function of Abramowitz and Stegun (1968) with $a = (Y_t - M_t) + \gamma a_{t-1}, b = (1-\gamma)a_{t-1}$ and $c = \theta_{t-1}/\gamma$. To evaluate the CHF function, we can use the R package `gsl` of Hankin (2006).

Thus, the weights $p(Y_t | M_t^i, \theta_{t-1}^i, M^{(t-1)}, Y^{(t-1)})$ are proportional to

$$\begin{aligned} \omega^i &\propto \kappa(Y_t, M_t^i, \theta_{t-1}^i, \gamma, a_{t-1}) \\ &\times Beta(Y_t - M_t^i + \gamma a_{t-1}, (1-\gamma)a_{t-1}) \\ &\times CHF((Y_t - M_t^i) + \gamma a_{t-1}, (Y_t - M_t^i) + a_{t-1}, -\theta_{t-1}^i/\gamma), \end{aligned} \tag{16}$$

which are used to obtain the resampled values of $(M_t^{k(i)}, \theta_{t-1}^{k(i)})$.

Given $M_t^{k(i)}$, we can sample α by updating sufficient statistics of the beta distribution.

Propogating θ_t

Propogate θ_t conditionally on $(M_t^{k(i)}, \theta_{t-1}^{k(i)})$, from

$$\begin{aligned} p(\theta_t|\theta_{t-1}, M_t, Y^{(t)}, M^{(t-1)}) &\propto p(Y_t|M_t, \theta_t)p(\theta_t|M_t, \theta_{t-1}, M^{(t-1)}, D_{t-1}) \\ &\propto e^{-\theta_t}\left(\frac{\theta_{t-1}}{\gamma} - \theta_t\right)^{(1-\gamma)a_{t-1}-1} \theta^{\gamma a_{t-1}+(Y_t-M_t)-1}, \end{aligned}$$

which is proportional to a scaled hypergeometric beta density as discussed in Gordy (1998).

We can us a Metropolis-Hastings step or some form of rejection sampling step.

Alternative: Blind propagation

Alternatively, in propogating θ_t, if generating from the hypergeometric beta density is not computationally efficient, we can use a sequential importance sampling type step as suggested in Aktekin et. al (2018).

More specifically, θ_t can be (blind) propagated via

$$p(\theta_t|\theta_{t-1}, M^{(t-1)}, Y^{(t-1)}),$$

which is a scaled Beta distribution, $Beta(\gamma a_{t-1}, (1-\gamma)a_{t-1}, \theta_{t-1}/\gamma)$. The particles are then resampled with weights

$$p(Y_t|\theta_t, M_t),$$

which is a shifted $Poisson(\theta_t)$, with $Y_t \geq M_t$.

Finally, we can update γ offline or alternatively using a discrete prior with the marginal likelihood.

Collapsed particle filtering

To avoid the evaluation of CHF, we can integrate θ_t out and use the predictive distribution of $\epsilon_t = Y_t - M_t$ given $M^{(t-1)}$, $Y^{(t-1)}$ for resampling, the *collapsed particle filtering*.

Particles at $t-1$: (M^i_{t-1},α^i), for $i = 1, \ldots, N$.

Sample M_t from $Bin(Y_{t-1}, \alpha^i)$.

Resample M_t using the predictive distribution

$$p(\epsilon_t|\gamma, M^{(t-1)}, Y^{(t-1)}) = \frac{\Gamma(\gamma a_{t-1} + \epsilon_t)}{\Gamma(\epsilon_t + 1)\,\Gamma(\gamma a_{t-1})}\Big(1 - \frac{\gamma b_{t-1}}{\gamma b_{t-1} + 1}\Big)^{\epsilon_t}\Big(\frac{\gamma b_{t-1}}{\gamma b_{t-1} + 1}\Big)^{\gamma a_{t-1}}.$$

Finally, we can update the sufficient statistics in the posterior distribution of

$$(\theta_t|\gamma, M_t, M^{(t-1)}, Y^{(t-1)}) \sim Gamma(a_t, b_t),$$

where $a_t = \gamma a_{t-1} + (Y_t - M_t^{k(i)})$ and $b_t = \gamma b_{t-1} + 1$.

PL algorithm

Let $\{(\theta_{t-1}, M_{t-1}, \alpha)^i\}_{i=1}^N$ be the particle set at time $t-1$.

1. Draw $M_t^i \sim Binomial(Y_{t-1}, \alpha^i)$.
2. Compute resampling weights from Equation (18)
3. Resampling particles $\{(\theta_{t-1}, M_t)^{k(i)}\}_{i=1}^N$ with weights from 2.
4. Update (a_t, b_t): $a_t^i = \gamma a_{t-1}^{k(i)} + (Y_t - M_t^{k(i)})$ and $b_t^i = \gamma b_{t-1}^{k(i)} + 1$.
5. Draw thinning: $\alpha | M^t, Y^t \sim Beta(S_{1t}^i, S_{2t}^i)$, where $S_{1t}^i = S_{1,t-1}^{k(i)} + M_t^{k(i)}$ and $S_{2t}^i = S_{2,t-1}^{k(i)} + (Y_{t-1} - M_t^{k(i)})$, where $S_{10} = a_\alpha$ and $S_{20} = b_\alpha$.
6. Propagate θ_t from $\theta_t \sim Beta(\gamma a_{t-1}, (1-\gamma) a_{t-1}, \theta_{t-1}/\gamma)$, a scaled Beta distribution.
7. Resample with weights $p(Y_t | \theta_t, M_t)$, a shifted $Poisson(\theta_t)$, with $Y_t \geq M_t$.

Collapsed PL algorithm

Let $\{(M_{t-1},\alpha)^i\}_{i=1}^N$ be the particle set at time $t-1$.

1. Draw $M_t^i \sim Binomial(Y_{t-1},\alpha^i)$.
2. Compute resampling weights from equation, for $\epsilon_t^i = Y_t - M_t^i$:

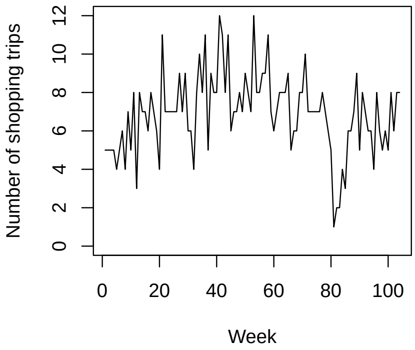
$$\omega^i = \frac{\Gamma(\gamma a_{t-1}+\epsilon_t^i)}{\Gamma(\epsilon_t^i+1)\,\Gamma(\gamma a_{t-1})}\left(1-\frac{\gamma b_{t-1}}{\gamma b_{t-1}+1}\right)^{\epsilon_t^i}\left(\frac{\gamma b_{t-1}}{\gamma b_{t-1}+1}\right)^{\gamma a_{t-1}},$$

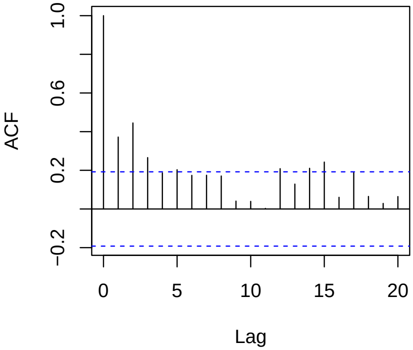
3. $M_t^{k(i)}$ are the resample draws.
4. Update SS and sample α^i: $S_{1t}^i = S_{1,t-1}^{k(i)} + M_t^{k(i)}$, $S_{2t}^i = S_{2,t-1}^{k(i)} + (Y_{t-1} - M_t^{k(i)})$, $\alpha^i \sim Beta(S_{1t}^i, S_{2t}^i)$, where $S_{10} = a_\alpha$ and $S_{20} = b_\alpha$.
5. Update a_t and b_t and sample θ_t from $(\theta_t|\gamma, M^{(t)}, Y^{(t)})$

$$\begin{aligned} a_t^i &= \gamma a_{t-1}^{k(i)} + (Y_t - M_t^{k(i)}) \\ b_t^i &= \gamma b_{t-1}^{k(i)} + 1 \\ \theta_t^i &\sim Gamma(a_t^i, b_t^i). \end{aligned}$$

Weekly shopping trips

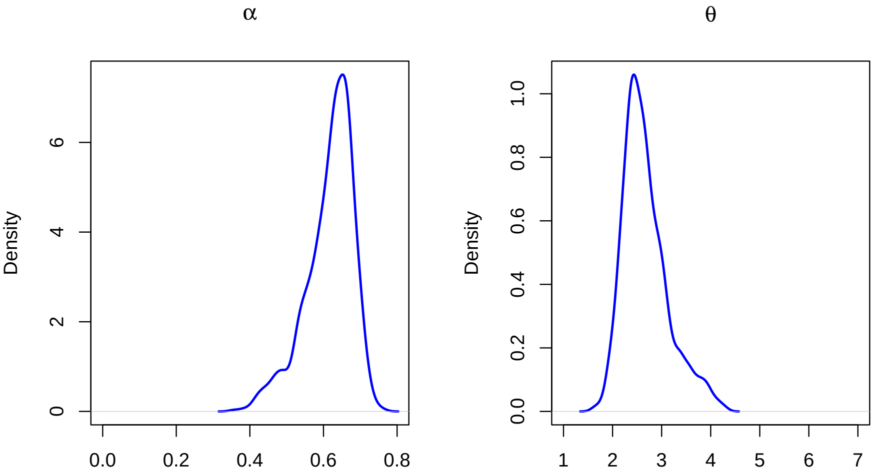
Data: number of weekly shopping trips of a household to the supermarket over 104 weeks. The data is a subset of a large set used in Kim (2013).





Static INAR(1)

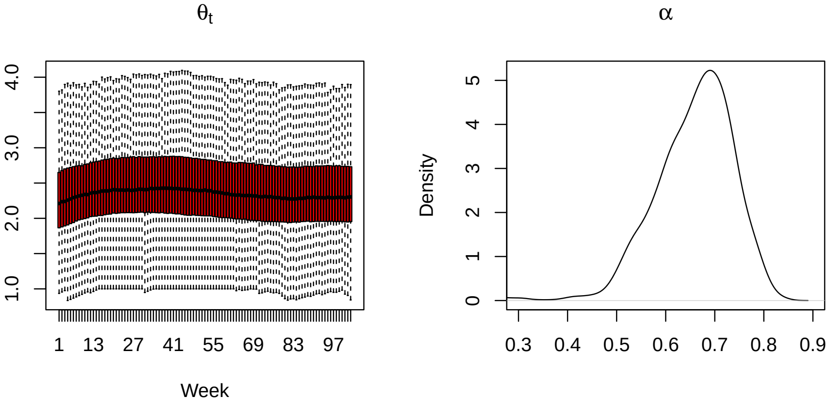
The posterior means of α and θ are 0.622 and 2.565, respectively.



Posterior distributions ($\gamma = 0.98$)

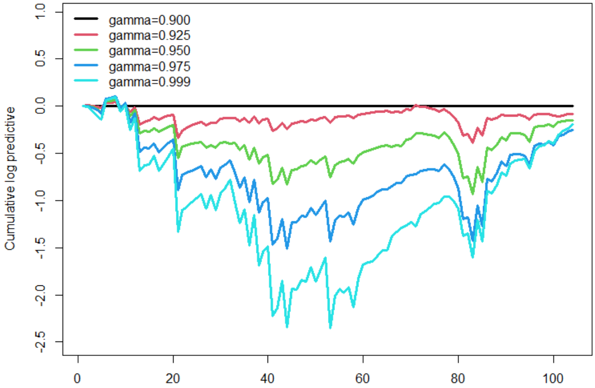
The smoothing distributions of θ_t agree with the posterior distribution of θ.

Similar agreement can also be noticed for posterior distributions of α.



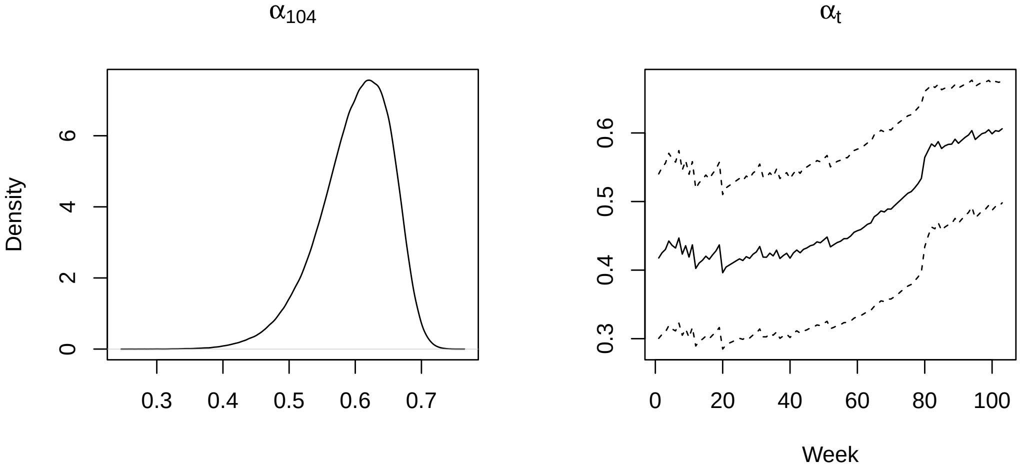
Learning γ

We consider 5 values of γ and evaluate the MC average of the marginal likelihood based on the particles of $M^{(t)}$ at each time point. The static INAR(1) model is not supported by the data. Based on the figure, we will use $\gamma = 0.9$ in our analysis of the dynamic INAR(1) model.



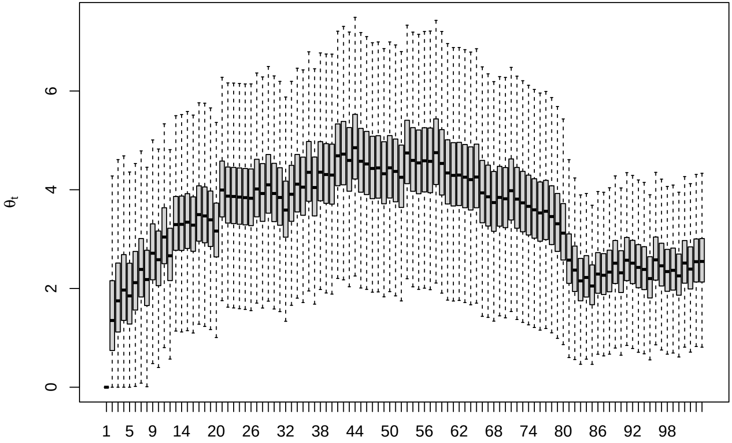
Collapsed PF

The mean probability of thinning stabilizes around the values of 0.4-0.5.

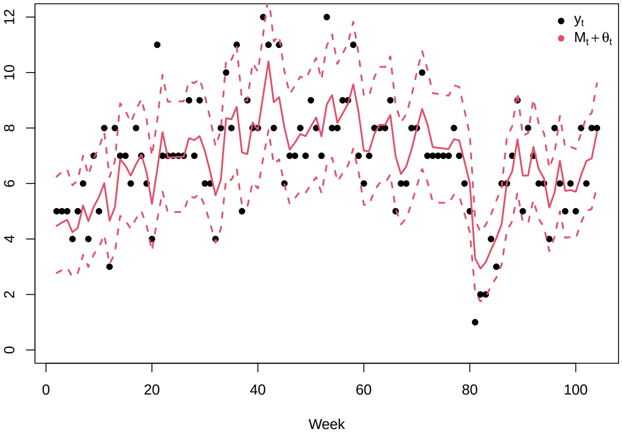


Filtering distributions of θ_t ($\gamma = 0.9$)

The filtering distributions exhibits more variability than their Gibbs counterparts ($\gamma = 0.98$).

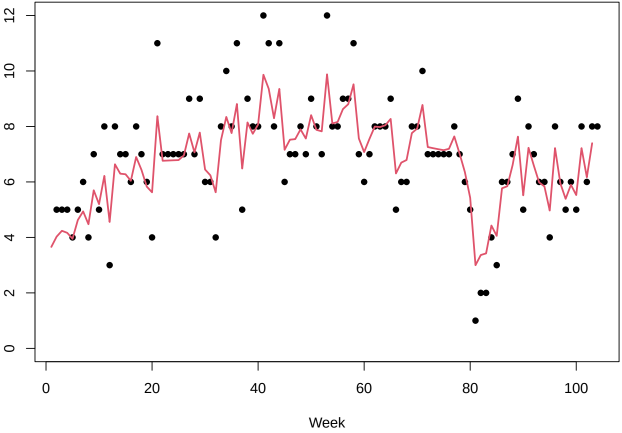


Number of trips vs $E(M_t + \theta_t | Y^t)$



Number of trips vs predicted values

$$E[Y_t|Y^{(t-1)}] = E[\alpha|Y^{(t-1)}]Y_{t-1} + E[\theta_t|Y^{(t-1)}]$$



Concluding remarks

- We introduced the class of dynamic INAR models for count data.
- Our proposal includes static INAR models and several alternatives.
- We developed a Gibbs sampler and particle filters for Bayesian inference.
- The dynamic model outperformed its static counterpart.

Future work

- Time-varying thinning.
- Dependent Dirichlet process vs Poisson.
- High-order dynamic INAR processes.
- Multivariate dynamic INAR processes.