# Likelihood estimation for the INAR($p$) model by saddlepoint approximation

K. Fokianos
fokianos@ucy.ac.cy

(joint work with A.C. Davison and X. Pedeli)

UNIVERSITY OF CYPRUS

Recent developments in statistics for complex dependent data

29th August 2015

# Outline

# Past research

- Introduced into statistics in a seminal paper by Daniels (1954).
- Studied extensively during the last two decades, see e.g. Field and Ronchetti (1990), Jensen (1995) and Butler (2007).
- Quite complex theory but fairly straightforward use in practice.
- Main advantage: The high accuracy with which they can approximate intractable densities and tail probabilities, even for extremely small sample sizes (Davison, 2003).

# Basic idea: From Taylor series to Saddlepoints

(Goutis and Casella, 1999)

- Consider a positive function $f(x)$ and suppose that our aim is to approximate its value at some point $x_0$.
- Let $h(x) = \log f(x) \Rightarrow f(x) = \exp h(x)$
- Taylor series expansion:
$$f(x) \approx \exp\left\{h(x_0) + (x - x_0)h'(x_0) + \frac{(x - x_0)^2}{2}h''(x_0)\right\}.$$

# Basic idea: From Taylor series to Saddlepoints

(Goutis and Casella, 1999)

- Consider a positive function $f(x)$ and suppose that our aim is to approximate its value at some point $x_0$.
- Let $h(x) = \log f(x) \Rightarrow f(x) = \exp h(x)$
- Taylor series expansion:

$$f(x) \approx \exp\left\{ h(x_0) + (x - x_0)h'(x_0) + \frac{(x - x_0)^2}{2} h''(x_0) \right\}.$$

- If $x_0 = \hat{x}$, then $h'(\hat{x}) = 0$ and the second terms disappears, i.e.

$$f(x) \approx \exp\left\{ h(x_0) + \frac{(x - x_0)^2}{2} h''(x_0) \right\} \tag{1}$$

# Basic idea: From Taylor series to Saddlepoints

(Goutis and Casella, 1999)

- Consider a positive function $f(x)$ and suppose that our aim is to approximate its value at some point $x_0$.
- Let $h(x) = \log f(x) \Rightarrow f(x) = \exp h(x)$
- Taylor series expansion:
$$f(x) \approx \exp\left\{h(x_0) + (x - x_0)h^{'}(x_0) + \frac{(x - x_0)^2}{2}h^{''}(x_0)\right\}.$$
- If $x_0 = \hat{x}$, then $h^{'}(\hat{x}) = 0$ and the second terms disappears, i.e.

$$f(x) \approx \exp\left\{h(x_0) + \frac{(x - x_0)^2}{2}h^{''}(x_0)\right\} \tag{1}$$

- Equation (1) is exact if $h(x)$ is a quadratic equation.

# Basic idea: From Taylor series to Saddlepoints (con't)

## Laplace approximation

- Equation (1) can be useful for computing integrals of positive functions:

$$\int f(x)dx \approx \int \exp\left\{h(\hat{x}) + \frac{(x-\hat{x})^2}{2}h''(\hat{x})\right\}dx \qquad (2)$$

# Basic idea: From Taylor series to Saddlepoints (con't)

## Laplace approximation

- Equation (1) can be useful for computing integrals of positive functions:

$$\int f(x)dx \approx \int \exp\left\{h(\hat{x}) + \frac{(x-\hat{x})^2}{2} h^{''}(\hat{x})\right\} dx \qquad (2)$$

- If $\hat{x}$ is a maximum, $h^{''}(\hat{x}) < 0$ and (2) is equivalent to

$$\int f(x)dx \approx \exp\{h(\hat{x})\} \left(-\frac{2\pi}{h^{''}(\hat{x})}\right)^{1/2} \qquad (3)$$

# Basic idea: From Taylor series to Saddlepoints (con't)

## Saddlepoint approximation

- Combine (1) and (3), i.e. try to approximate a function by a Laplace type approximation of an integral:

# Basic idea: From Taylor series to Saddlepoints (con't)

## Saddlepoint approximation

- Combine (1) and (3), i.e. try to approximate a function by a Laplace type approximation of an integral:
- Write the function $f$ as $f(x) = \int m(x,u)du$, for some positive $m(x,u)$.

# Basic idea: From Taylor series to Saddlepoints (con't)

## Saddlepoint approximation

- Combine (1) and (3), i.e. try to approximate a function by a Laplace type approximation of an integral:
- Write the function $f$ as $f(x) = \int m(x,u)du$, for some positive $m(x,u)$.
- By defining $k(x,u) = \log m(x,u)$, consider the Laplace approximation of the integral of $\exp k(x,u)$ with respect to $u$.

# Basic idea: From Taylor series to Saddlepoints (con't)

## Saddlepoint approximation

- Combine (1) and (3), i.e. try to approximate a function by a Laplace type approximation of an integral:
- Write the function $f$ as $f(x) = \displaystyle\int m(x,u)du$, for some positive $m(x,u)$.
- By defining $k(x,u) = \log m(x,u)$, consider the Laplace approximation of the integral of $\exp k(x,u)$ with respect to $u$.
- Hence, for any fixed $x$,

$$\begin{aligned} f(x) &\approx \int \exp\left\{ k(x,\hat{u}(x)) + \frac{(u-\hat{u}(x))^2}{2} \frac{\partial^2 k(x,u)}{\partial u^2}\bigg|_{\hat{u}(x)} \right\} du \\ &= \exp\{k(x,\hat{u}(x))\} \left( - \frac{2\pi}{\frac{\partial^2 k(x,u)}{\partial u^2}\Big|_{\hat{u}(x)}} \right)^{1/2} \end{aligned} \quad (4)$$

# Basic idea: From Taylor series to Saddlepoints (con't)

## Saddlepoint approximation

- Combine (1) and (3), i.e. try to approximate a function by a Laplace type approximation of an integral:
- Write the function $f$ as $f(x) = \displaystyle\int m(x,u)du$, for some positive $m(x,u)$.
- By defining $k(x,u) = \log m(x,u)$, consider the Laplace approximation of the integral of $\exp k(x,u)$ with respect to $u$.
- Hence, for any fixed $x$,

$$\begin{aligned} f(x) &\approx \int \exp\left\{ k(x,\hat{u}(x)) + \frac{(u-\hat{u}(x))^2}{2} \frac{\partial^2 k(x,u)}{\partial u^2}\Big|_{\hat{u}(x)} \right\} du \\ &= \exp\{k(x,\hat{u}(x))\} \left( -\frac{2\pi}{\frac{\partial^2 k(x,u)}{\partial u^2}\Big|_{\hat{u}(x)}} \right)^{1/2} \end{aligned} \quad (4)$$

- Equation (4) is a saddlepoint approximation of $f(x)$ and $\hat{u}(x)$ is the saddlepoint that maximizes $k(x,u)$.

# Basic idea: From Taylor series to Saddlepoints (con't)

- For a density $f(x)$, using the inversion formula, it can be proved that
$$k(x,u) = K_X(u) - ux$$
where $K_X(u)$ is the cumulant generating function, i.e.
$$K_X(u) = \log M_X(u) = \log \int_{-\infty}^{+\infty} \exp(ux) f(x) dx,$$
whith $M_X(u)$ denoting the moment generating function.
- The saddlepoint is the point $\hat{u}(x)$ that satisfies
$$K_X'(u) = x.$$
- After some tedious calculations, the saddlepoint approximation to $f(x)$ proves to be
$$f_X(x) \approx \left( \frac{1}{2\pi K_X''(\hat{u}(x))} \right)^{1/2} \exp\left\{ K_X(\hat{u}(x)) - \hat{u}(x)x \right\}.$$

# Basic idea: The saddlepoint method in a time series framework

Let $\{X_t, t \in \mathbb{Z}\}$ be a discrete–valued time series with available conditional pmf $f_{X_t|X_{t-1},\ldots,X_{t-p}}(x_t)$ and cumulant generating function (cgf)

$$K_t(u) = \log E(\exp(uX_t)|X_{t-1}, \ldots, X_{t-p}),$$

for some $p \geq 1$. Then, the saddlepoint approximation (SPA) to the conditional pmf of $X_t$ given $X_{t-1}, \ldots, X_{t-p}$, is given by

$$\tilde{f}_{X_t|X_{t-1},\ldots,X_{t-p}}(x_t) = \left\{ \frac{1}{2\pi K_t^{''}(\tilde{u})} \right\}^{1/2} \exp\{K_t(\tilde{u}) - \tilde{u}_t x_t\}, \qquad (5)$$

where $K_t^{'}(u)$ and $K_t^{''}(u)$ denote the first and second order derivatives of $K_t(u)$ with respect to $u$, and $\tilde{u}_t$ is the unique value of $u$ satisfying $K_t^{'}(u) = x_t$, for $t = p+1, p+2, \ldots$

# Basic idea: The saddlepoint method in a time series framework (con't)

- Expression (5) is the leading term of a saddlepoint density approximation, but further terms can be included.
- For instance, a second-order approximation is

$$\tilde{f}_{X_t|X_{t-1},\ldots,X_{t-p}}(x_t) = \left\{\frac{1}{2\pi K_t''(\tilde{u})}\right\}^{1/2} \exp\{K_t(\tilde{u}) - \tilde{u}_t x_t\} \left\{1 + \left(\frac{1}{8}\tilde{\kappa}_t^{(4)} - \frac{5}{24}[\tilde{\kappa}_t^{(3)}]^2\right)\right\},$$

where $\tilde{\kappa}_t^{(r)} = K_t^{(r)}(\tilde{u})/\{K_t''(\tilde{u})\}^{r/2}$, $r \geq 3$ is the $r$-th order standardized cumulant of $X_t$ (Butler, 2007).

- In the classical setting, $X_t$ is taken to be an average of $m$ independent observations, and then it may be shown that its true and approximate densities $f$ and $\tilde{f}$ are related by $f(x_t) = \tilde{f}(x_t)\{1 + \mathcal{O}(m^{-1})\}$.

# Remarks

- $\tilde{f}_{X_t|X_{t-1},\ldots,X_{t-p}}(x_t)$ is always positive, but it does not sum to one exactly. Hence, it is often renormalized, a practice which also improves the relative order of the approximation, see e.g. Field and Ronchetti (1990) and Kolassa (2006). Renormalization is not possible when considering conditional probabilities.

# Remarks

- $\tilde{f}_{X_t|X_{t-1},\ldots,X_{t-p}}(x_t)$ is always positive, but it does not sum to one exactly. Hence, it is often renormalized, a practice which also improves the relative order of the approximation, see e.g. Field and Ronchetti (1990) and Kolassa (2006). Renormalization is not possible when considering conditional probabilities.
- Approximation (5) is based on the normal distribution but alternative approximations based on other distributions, such as the Gamma and inverse Gaussian distributions, are also available (Wood et al., 1993).

# Remarks

- $\tilde{f}_{X_t|X_{t-1},\ldots,X_{t-p}}(x_t)$ is always positive, but it does not sum to one exactly. Hence, it is often renormalized, a practice which also improves the relative order of the approximation, see e.g. Field and Ronchetti (1990) and Kolassa (2006). Renormalization is not possible when considering conditional probabilities.
- Approximation (5) is based on the normal distribution but alternative approximations based on other distributions, such as the Gamma and inverse Gaussian distributions, are also available (Wood et al., 1993).
- The saddlepoint approximation requires that the cgf exists.

# Model

## Definition

A sequence of random variables $\{X_t, t \in \mathbb{Z}\}$ is an integer–valued autoregressive, INAR($p$), process if it satisfies a difference equation of the form

$$X_t = \sum_{i=1}^{p} \alpha_i \circ X_{t-i} + \epsilon_t, \tag{6}$$

where $\alpha_i \in [0, 1)$ are fixed constants and $\{\epsilon_t\}$ is a sequence of uncorrelated non–negative integer–valued random variables having mean $\mu_\epsilon$ and finite variance $\sigma_\epsilon^2$. The sequence $\{\epsilon_t\}$ is called the innovation process (McKenzie, 1985; Al–Osh and Alzaid, 1987).

# Model (con't)

- Definition (6) is based on the notion of binomial thinning.
- The binomial thinning operator '$\circ$' is defined by

$$\alpha \circ X = \sum_{j=1}^{X} Y_j$$

  where $Y_j$ are i.i.d. Bernoulli random variables with $P(Y_j = 1) = 1 - P(Y_j = 0) = \alpha$.
- The binomial thinning operations involved in (6) are independent, so the INAR($p$) process has the classical AR($p$) correlation structure.
- A unique strictly stationary and ergodic solution of (6) exists if $\sum_{i=1}^{p} \alpha_i < 1$.

# Example: $p = 1$

Let $\alpha \in [0, 1)$ and let $\{\epsilon_t\}$ be a sequence of i.i.d. non–negative integer valued random variables with $\mathrm{E}[\epsilon_t] = \mu_\epsilon$ and $\mathrm{Var}[\epsilon_t] = \sigma_\epsilon^2$. The integer–valued autoregressive process of order 1, INAR(1), $\{X_t, t \in \mathbb{Z}\}$ is defined by the equation

$$X_t = \alpha \circ X_{t-1} + \epsilon_t,$$

where $\alpha \circ X_{t-1}$ is the sum of $X_{t-1}$ Bernoulli random variables all of which are independent of $X_{t-1}$.

# Example: $p = 1$ (con't)

It can be shown that

- $\mathrm{E}\left[X_t\right] = \mu_\epsilon/(1-\alpha)$,
- $\mathrm{Var}\left[X_t\right] = (\alpha\mu_\epsilon + \sigma_\epsilon^2)/(1-\alpha^2)$
- the autocovariance function evaluated at lag $k$, $c(k)$, is given by

$$c(k) \equiv \mathrm{Cov}\left[X_t, X_{t-k}\right] = \alpha^k c(0). \tag{7}$$

- Consequently, the autocorrelation function, $\rho(k)$, is

$$\rho(k) = \frac{c(k)}{c(0)} = \alpha^k, \tag{8}$$

so that $\rho(k)$ decays exponentially with the lag $k$ as in AR(1), but unlike the autocorrelation of a stationary AR(1) process, it is always positive for $\alpha \in (0,1)$.

# Saddlepoint Approximation for INAR processes

## Proposition

*Let $\{X_t\}$ be the INAR(p) process defined by (6). For large values of $x_{t-i}$ and if $0 < \sum_{i=1}^{p} \alpha_i < 1$, the exact and approximate densities are related by*

$$
\begin{aligned}
f_{X_t|X_{t-1},\ldots,X_{t-p}}(x_t) &= \tilde{f}_{X_t|X_{t-1},\ldots,X_{t-p}}(x_t)\left\{1+\mathcal{O}\left(\frac{1}{\sum_{i=1}^{p} x_{t-i}+1}\right)\right\} \\
&= \left\{\frac{1}{2\pi K_t''(\tilde{u}_t)}\right\}^{1/2} \exp\left\{K_t(\tilde{u}_t)-\tilde{u}_t x_t\right\} \\
&\times \left\{1+\mathcal{O}\left(\frac{1}{\sum_{i=1}^{p} x_{t-i}+1}\right)\right\}.
\end{aligned} \tag{9}
$$

# Inference

## Maximum likelihood estimation (MLE)

Let $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \ldots, \alpha_p, \mu_\epsilon)^T$ be the unknown parameter vector. The maximum likelihood estimator of $\boldsymbol{\theta}$ is calculated by maximizing the conditional log likelihood function

$$\ell(\boldsymbol{\theta}) = \sum_{t=p+1}^{n} \log P(X_t = x \mid X_{t-1} = x_{t-1}, \ldots X_{t-p} = x_{t-p}),$$

that is $\hat{\boldsymbol{\theta}}_{\text{MLE}} = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta})$.

# Inference (con't)

## Example

An important special case is that of Poisson INAR(1) process $X_t = \alpha \circ X_{t-1} + \epsilon_t$, with $\{\epsilon_t\}$ a sequence of i.i.d. Poisson random variables with mean $\mu_\epsilon$. The conditional distribution of $X_t$ given $X_{t-1}$ is

$$P(X_t = x | X_{t-1} = x_{t-1}) = x_{t-1}! \exp(-\mu_\epsilon) \sum_{i=1}^{m} \frac{\alpha^i (1-\alpha)^{x_{t-1}-i} \mu_\epsilon^{x-i}}{i!(x_{t-1}-i)!(x-i)!},$$

where $m = \min(x_{t-1}, x)$.

# Inference (con't)

## Example

Under a Poisson INAR($p$) process,

$$P(X_t = x | X_{t-1} = x_{t-1}, \ldots X_{t-p} = x_{t-p}) =$$

$$\sum_{i_1=0}^{\min(x_{t-1},x)} \binom{x_{t-1}}{i_1} \alpha_1^{i_1} (1-\alpha_1)^{x_{t-1}-i_1}$$

$$\times \sum_{i_2=0}^{\min(x_{t-2},x-i_1)} \binom{x_{t-2}}{i_2} \alpha_2^{i_2} (1-\alpha_2)^{x_{t-2}-i_2}$$

$$\ldots \sum_{i_p=0}^{\min(x_{t-p},x-(i_1+\ldots+i_p))} \binom{x_{t-p}}{i_p} \alpha_1^{i_p} (1-\alpha_p)^{x_{t-p}-i_p}$$

$$\times \frac{\exp(-\mu_\epsilon)\mu_\epsilon^{x-(i_1+\ldots+i_p)}}{(x-(i_1+\ldots+i_p))!}.$$

# Motivation

- Maximization of $\ell(\boldsymbol{\theta})$ is quite cumbersome, owing to the nested summations involved in its computation, and the numerical difficulties that can arise when summing up many small probabilities.
- Alternative approach: Approximate the log–likelihood of the INAR($p$) model by the saddlepoint method.
- The saddlepoint approximation (SPA),
  - circuvments computational difficulties,
  - provides simple and highly accurate approximations to the conditional maximum likelihood estimators even for complicated INAR models.

# Generalities

## Key observation

Conditional on $X_{t-p} = x_{t-p}, \ldots, X_{t-1} = x_{t-1}$, $X_t$ is a sum of independent binomial variates with parameters $(x_{t-i}, \alpha_i)$, plus the innovation term $\epsilon_t$. Therefore, its conditional cgf is

$$K_t(u) = \sum_{i=1}^{p} x_{t-i} \log\{1 - \alpha_i + \alpha_i \exp(u)\} + K_{\epsilon_t}(u),$$

where $K_{\epsilon_t}(u)$ denotes the cumulant-generating function of $\epsilon_t$.

# Generalities

## Key observation

Conditional on $X_{t-p} = x_{t-p}, \ldots, X_{t-1} = x_{t-1}$, $X_t$ is a sum of independent binomial variates with parameters $(x_{t-i}, \alpha_i)$, plus the innovation term $\epsilon_t$. Therefore, its conditional cgf is

$$K_t(u) = \sum_{i=1}^{p} x_{t-i} \log\{1 - \alpha_i + \alpha_i \exp(u)\} + K_{\epsilon_t}(u),$$

where $K_{\epsilon_t}(u)$ denotes the cumulant-generating function of $\epsilon_t$.

Remarks:

- The domain of $K_t(u)$ is the same set as the domain of $K_{\epsilon_t}(u)$. This is the entire real line for Poisson innovations, and an interval of form $(-\infty, c)$ for negative binomial innovations.
- $K_t(u)$ is a sum of convex functions and thus itself convex.

# Generalities (con't)

## Approximate log–likelihood function

$$\tilde{\ell}(\boldsymbol{\theta}) = \sum_{t=p+1}^{n} \log \tilde{f}_{X_t|X_{t-1},\ldots,X_{t-p}}(x_t),$$

where $\tilde{f}_{X_t|X_{t-1},\ldots,X_{t-p}}(x_t) = \left\{\frac{1}{2\pi K_t''(\tilde{u})}\right\}^{1/2} \exp\left\{K_t(\tilde{u}) - \tilde{u}_t x_t\right\}$, is the SPA to the conditional pmf of $X_t$ (Daniels, 1954). The value $\tilde{\boldsymbol{\theta}}$ maximizing $\tilde{\ell}(\boldsymbol{\theta})$ is the saddlepoint maximum likelihood estimator (SPMLE).

# Generalities (con't)

## Evaluation of $\tilde{\ell}(\boldsymbol{\theta})$

- It requires the computation of $n-p$ values of $\tilde{u}$, one for each term in the sum.
- For $x_t = 0$ the pmf need not to be approximated, since

$$P(X_t = 0 \mid X_{t-1} = x_{t-1}, \ldots X_{t-p} = x_{t-p}) = f_{\epsilon_t}(0) \prod_{i=1}^{p} (1-\alpha_i)^{x_{t-i}}.$$

- If the saddlepoint equation $K_t^{'}(u) = x_t$ cannot be solved analytically, then $\tilde{u}_t$ can be approximated numerically, e.g. using the Newton-Raphson method of Lieberman (1994): linear expansion of the saddlepoint equation in the neighborhood of an initial value $u_0$, i.e. $K_t^{'}(u_0) + K_t^{''}(u_0)(\tilde{u}_t - u_0) \approx x_t$, to provide the update

$$\tilde{u}_t = \frac{x_t - K_t^{'}(u_0)}{K_t^{''}(u_0)} + u_0,$$

which is iterated until the approximation is adequate.

# Asymptotic properties of the SPMLE

- Established result (Bu et al., 2008):

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N\{0, I^{-1}(\boldsymbol{\theta}_0)\}, \quad n \to \infty, \tag{10}$$

  where $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}_\mathbf{0}$ based on data $X_1, \ldots, X_n$ from a stationary INAR($p$) process, and $I(\boldsymbol{\theta}_0)$ is the Fisher information in a single observation.

- We have proved that

$$\tilde{f}(x_t \mid x_{t-1}, \ldots, x_{t-p}) = \left\{\frac{1}{2\pi\sigma_t^2(\boldsymbol{\theta})}\right\}^{1/2} \exp\left[-\frac{\{x_t - \mu_t(\boldsymbol{\theta})\}^2}{2\sigma_t^2(\boldsymbol{\theta})}\right], \tag{11}$$

  which is the probability density function of a Gaussian random variable with

$$\begin{aligned} \mu_t(\boldsymbol{\theta}) &= E(X_t \mid X_{t-1}, \ldots, X_{t-p}) = \sum_{i=1}^{p} \alpha_i x_{t-i} + \mu, \\ \sigma_t^2(\boldsymbol{\theta}) &= \mathrm{Var}(X_t \mid X_{t-1}, \ldots, X_{t-p}) = \sum_{i=1}^{p} \alpha_i(1-\alpha_i) x_{t-i} + \sigma_\epsilon^2. \end{aligned}$$

# Asymptotic properties of the SPMLE (con't)

- The conditional mean and variance can be used to construct a quasilikelihood function for the estimation of $\boldsymbol{\theta}$, and the resulting estimator is asymptotically normally distributed; see Theorem 3.2.23 of Taniguchi and Kakizawa (2000).
- Empirical results indicate close agreement between the MLE $\hat{\boldsymbol{\theta}}$ and the SPMLE $\tilde{\boldsymbol{\theta}}$.
- Proposition 2.1 can be used to show that in the usual asymptotic setting, with $n \to \infty$ and the parameters fixed, the SPMLE is inconsistent. However, when the process assumes large values, i.e., when the mean of process is growing, then it can be shown that that the SPMLE is a consistent estimator of $\theta$.

## SPA for Poisson INAR($p$) processes

Assume that the error sequence $\{\epsilon_t\}$ consists of i.i.d. Poisson random variables with mean $\lambda$. Then,

- $K_t(u) = \sum_{i=1}^{p} x_{t-i} \log\left(1 - \alpha_i + \alpha_i e^u\right) + \lambda(e^u - 1)$,
- $K_t^{'}(u) = \sum_{i=1}^{p} \left[ x_{t-i} \frac{\alpha_i e^u}{\alpha_i e^u + (1 - \alpha_i)} \right] + \lambda e^u$,
- $K_t^{''}(u) = \sum_{i=1}^{p} \left[ x_{t-i} \frac{\alpha_i (1 - \alpha_i) e^u}{(\alpha_i e^u + (1 - \alpha_i))^2} \right] + \lambda e^u$.

# SPA for Poisson INAR($p$) processes (con't)

The SPA to the log–likelihood is given by

$$
\begin{aligned}
\tilde{\ell}(\boldsymbol{\theta}) = &-\frac{1}{2}\sum_{t=p+1}^{n}\log\left\{\sum_{i=1}^{p}\left[x_{t-i}\frac{\alpha_i(1-\alpha_i)e^{\tilde{u}_t}}{(\alpha_i e^{\tilde{u}_t}+(1-\alpha_i))^2}\right]+\lambda e^{\tilde{u}_t}\right\} \\
&+\sum_{t=p+1}^{n}\left\{\left[\sum_{i=1}^{p}\left[x_{t-i}\log(\alpha_i e^{\tilde{u}_t}+(1-\alpha_i))\right]+\lambda(e^{\tilde{u}_t}-1)\right]-\tilde{u}_t\right\},
\end{aligned}
$$

evaluated for each time point $t=p+1,\ldots,n$, at $u=\tilde{u}_t$.

# Existence of a finite saddlepoint $\tilde{u}_t$

## Theorem

*Let $X$ be a random variable supported on the possibly infinite interval $[-x_1, x_2]$, whose cumulant generating function $K_X(u)$ is defined on an interval with endpoints $-c_1$ and $c_2$, and $x_i < \infty$ or $c_i = \infty$ for $i = 1, 2$. Then a solution $\tilde{u}_t$ of the saddlepoint equation exists for all values of $x \in (-x_1, x_2)$ if (Kolassa, 2006)*

$$\lim_{u \to -c_1} K_X^{'}(u) = -x_1, \quad \lim_{u \to c_2} K_X^{'}(u) = x_2.$$

# Existence of a finite saddlepoint $\tilde{u}_t$

**Theorem**

*Let $X$ be a random variable supported on the possibly infinite interval $[-x_1, x_2]$, whose cumulant generating function $K_X(u)$ is defined on an interval with endpoints $-c_1$ and $c_2$, and $x_i < \infty$ or $c_i = \infty$ for $i = 1, 2$. Then a solution $\tilde{u}_t$ of the saddlepoint equation exists for all values of $x \in (-x_1, x_2)$ if (Kolassa, 2006)*

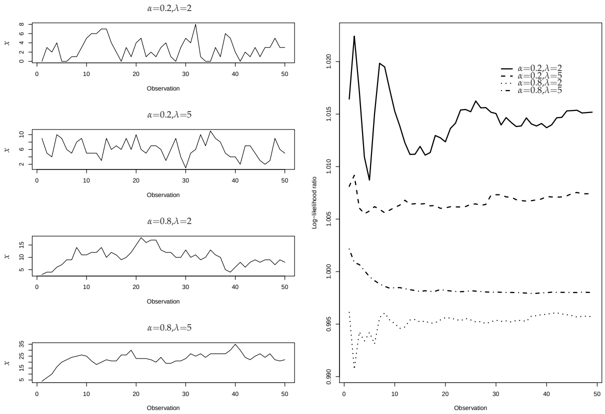
$$\lim_{u \to -c_1} K_X'(u) = -x_1, \quad \lim_{u \to c_2} K_X'(u) = x_2.$$

- For an INAR($p$) process, $X_t \in [0, \infty)$ and hence $x_1 = 0$ and $x_2 = \infty$.
- For the Poisson INAR($p$) model, the existence of a unique solution to the equation $K_t'(u) = x_t$ is easily established, since $K_t'(u)$ is strictly monotone increasing in $u$, and

$$\lim_{u \to -\infty} K_t'(u) = 0, \quad \lim_{u \to \infty} K_t'(u) = \infty.$$

# Quality of the approximation

Quality of the proposed approximation for four simulated series of length $n = 50$, shown in the left panel, generated from an INAR(1) process with $\alpha = 0.2$ or 0.8, $\lambda = 2$ or 5. The right panel shows the ratio of the true log–likelihood function to the approximate log–likelihood obtained by the saddlepoint method. The relative error is at most $\pm 2.5\%$. All ratios fluctuate around 1 indicating the high quality of the saddlepoint approximation. The accuracy increases for large mean values of the generated process, when the convolutions involved in the transition probabilities become more awkward.



# Design

- Simulation experiments under the assumptions of INAR(2) and INAR(4) processes with Poisson and negative binomial innovations.
-

| | Estimation methods | | | Sample sizes | |
|---|---|---|---|---|---|
| | CLS | MLE | SPA | $n$ | |
| INAR(2) | ✓ | ✓ | ✓ | 50 | 500 |
| INAR(4) | ✓ | | ✓ | 100 | 500 |

- Logit and log transformations to avoid inadmissible parameter estimates:

$$\xi = \log(\alpha/(1-\alpha)) \quad \rightsquigarrow \quad \alpha = \exp(\xi)/(1+\exp(\xi))$$
$$\eta = \log(\mu) \quad \rightsquigarrow \quad \mu = \exp(\eta)$$

- 1000 replicate simulations in each setting.
- Use of the `optim()` function in `R` for the minimization of the least squares and negative log likelihoods.

# Results for INAR(2) processes

Poisson INAR(2) with $(\alpha_1, \alpha_2, \lambda) = (0.3, 0.4, 5)$

| | | Bias×100 | | | MSE×100 | | |
|---|---|---|---|---|---|---|---|
| | | CLS | MLE | SPA | CLS | MLE | SPA |
| $n = 50$ | $\hat{\alpha}_1$ | -3.0 | -1.6 | -1.6 | 1.9 | 2.4 | 2.2 |
| | $\hat{\alpha}_2$ | -5.4 | -3.6 | -3.8 | 1.9 | 2.2 | 2.1 |
| | $\hat{\lambda}$ | 126.7 | 76.7 | 79.3 | 577.5 | 379.5 | 390.0 |
| $n = 500$ | $\hat{\alpha}_1$ | -0.4 | -0.2 | 0.1 | 0.2 | 0.2 | 0.2 |
| | $\hat{\alpha}_2$ | -0.6 | -0.3 | -0.6 | 0.2 | 0.1 | 0.1 |
| | $\hat{\lambda}$ | 15.8 | 7.9 | 10.9 | 47.8 | 32.9 | 34.2 |

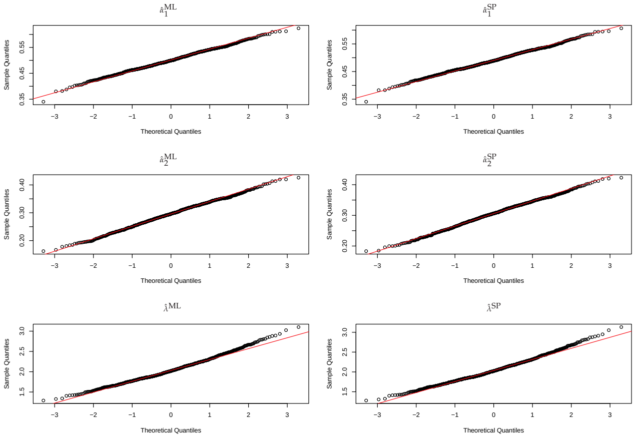
INAR(2)–NegBin innovations with $(\alpha_1, \alpha_2, \mu, r^*) = (0.5, 0.3, 2, 3)$

| | | Bias×100 | | | MSE×100 | | |
|---|---|---|---|---|---|---|---|
| | | CLS | MLE | SPA | CLS | MLE | SPA |
| $n = 50$ | $\hat{\alpha}_1$ | 4.1 | 0.8 | 1.7 | 2.3 | 0.9 | 1.0 |
| | $\hat{\alpha}_2$ | -5.7 | -1.2 | -0.9 | 2.1 | 1.0 | 1.2 |
| | $\hat{\mu}$ | 87.7 | 2.1 | -8.7 | 280.8 | 37.0 | 42.2 |
| $n = 500$ | $\hat{\alpha}_1$ | -0.2 | 0.1 | 0.9 | 0.2 | 0.1 | 0.1 |
| | $\hat{\alpha}_2$ | -0.9 | -0.2 | 0.6 | 0.5 | 0.1 | 0.1 |
| | $\hat{\mu}$ | 10.1 | 0.1 | -14.9 | 15.5 | 4.0 | 6.6 |

* $r$ is kept fixed

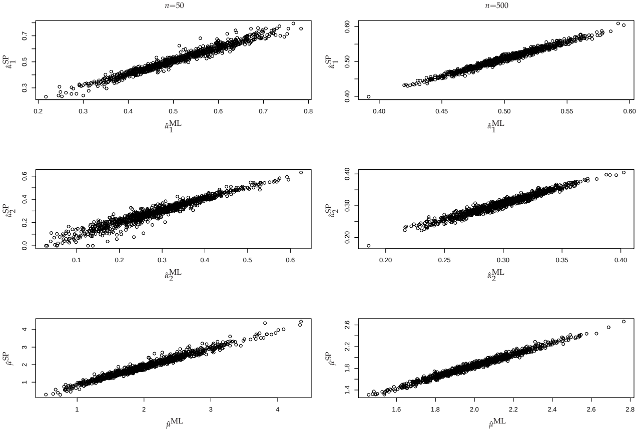
# Results for INAR(2) processes (con't)

Normal Q–Q plots for the MLE and SPA estimates obtained from a Poisson INAR(2) model ($n = 500$) with true parameter values $(\alpha_1, \alpha_2, \lambda) = (0.3, 0.4, 5)$. Results have been obtained after 1000 simulation replicates.



# Results for INAR(2) processes (con't)

Scatterplots between the MLE and SPA estimates obtained from an INAR(2) model with negative binomial innovations and true parameter values $(\alpha_1, \alpha_2, \mu, r) = (0.5, 0.3, 2, 3)$.

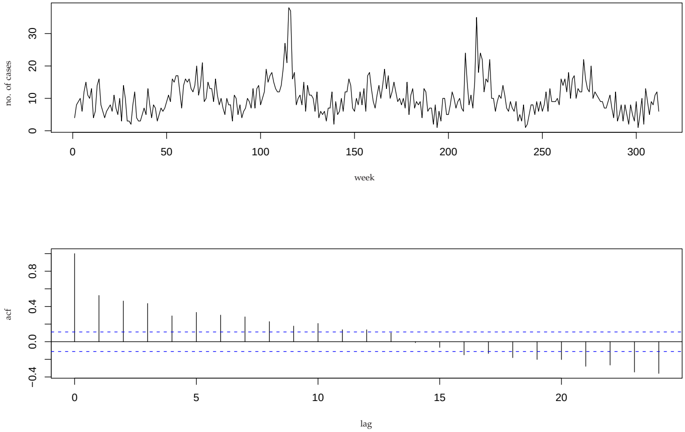


# Data

- Weekly number of meningococcal disease cases in Germany for the years 2001–2006 ($n = 312$).
- Source: the German national surveillance system for notifiable diseases, administered by the Robert Koch Institute (RKI).
- Significant overdispersion present in the data (mean 10.09, variance 27.83)$\rightsquigarrow$Need for a more flexible distribution than the Poisson distribution, e.g. negative binomial.

# Data (con't)

Time series and acf plots for the weekly number of meningococcal disease cases in Germany, 2001–2006.



# Modelling

- Let $X_t$ denote the number of meningococcal disease cases at time $t = 1, \ldots, 312$.
- $X_t$ is defined as an INAR($p$) process of the form,

$$X_t = \sum_{i=1}^{p} \alpha_i \circ X_{t-i} + \epsilon_t, \quad \epsilon_t \sim NegBin(\mu_t, r).$$

- To account for seasonality,

$$\log(\mu_t) = \beta_0 + \sum_{s=1}^{S} (\beta_{1;s} \sin(\omega_s t) + \beta_{2;s} \cos(\omega_s t)),$$

  where $S$ is the number of harmonics to include and $\omega_s$ are Fourier frequencies of the form $\omega_s = 2\pi s/52$ (see also Paul et al., 2008).
- Challenges: Model selection & parameter estimation through the application of the SPA technique.

# Results

INAR($p$) models fitted to the meningococcal disease data. $\log L$ denotes the approximate maximized log–likelihood obtained from the saddlepoint approximation.

| Model | $p$ | $r$ | $S$ | $\log L$ | AIC |
|---|---|---|---|---|---|
| 1. | 1 | 0.5 | 1 | −858.11 | 1724.2 |
| 2. | 1 | 0.5 | 2 | −857.73 | 1727.5 |
| 3. | 1 | 1 | 1 | −862.11 | 1732.2 |
| 4. | 1 | 1 | 2 | −861.80 | 1735.6 |
| 5. | 2 | 0.5 | 1 | −851.51 | 1713.0 |
| 6. | 2 | 0.5 | 2 | −851.21 | 1716.4 |
| 7. | 2 | 1 | 1 | −869.01 | 1748.0 |
| 8. | 2 | 1 | 2 | −868.62 | 1751.2 |
| 9. | 3 | 0.5 | 1 | −844.95 | 1701.9 |
| 10. | 3 | 0.5 | 2 | −844.49 | 1705.0 |
| 11. | 3 | 1 | 1 | −859.08 | 1730.2 |
| 12. | 3 | 1 | 2 | −859.09 | 1734.2 |
| 13. | 4 | 0.5 | 1 | −845.46 | 1704.9 |
| 14. | 4 | 0.5 | 2 | −858.95 | 1731.9 |
| 15. | 4 | 1 | 1 | −845.55 | 1709.1 |
| 16. | 4 | 1 | 2 | −862.46 | 1742.9 |

# Results

INAR($p$) models fitted to the meningococcal disease data. log$L$ denotes the approximate maximized log–likelihood obtained from the saddlepoint approximation.

| Model | $p$ | $r$ | $S$ | log$L$ | AIC |
|---|---|---|---|---|---|
| 1. | 1 | 0.5 | 1 | −858.11 | 1724.2 |
| 2. | 1 | 0.5 | 2 | −857.73 | 1727.5 |
| 3. | 1 | 1 | 1 | −862.11 | 1732.2 |
| 4. | 1 | 1 | 2 | −861.80 | 1735.6 |
| 5. | 2 | 0.5 | 1 | −851.51 | 1713.0 |
| 6. | 2 | 0.5 | 2 | −851.21 | 1716.4 |
| 7. | 2 | 1 | 1 | −869.01 | 1748.0 |
| 8. | 2 | 1 | 2 | −868.62 | 1751.2 |
| 9. | 3 | 0.5 | 1 | −844.95 | 1701.9 |
| 10. | 3 | 0.5 | 2 | −844.49 | 1705.0 |
| 11. | 3 | 1 | 1 | −859.08 | 1730.2 |
| 12. | 3 | 1 | 2 | −859.09 | 1734.2 |
| 13. | 4 | 0.5 | 1 | −845.46 | 1704.9 |
| 14. | 4 | 0.5 | 2 | −858.95 | 1731.9 |
| 15. | 4 | 1 | 1 | −845.55 | 1709.1 |
| 16. | 4 | 1 | 2 | −862.46 | 1742.9 |

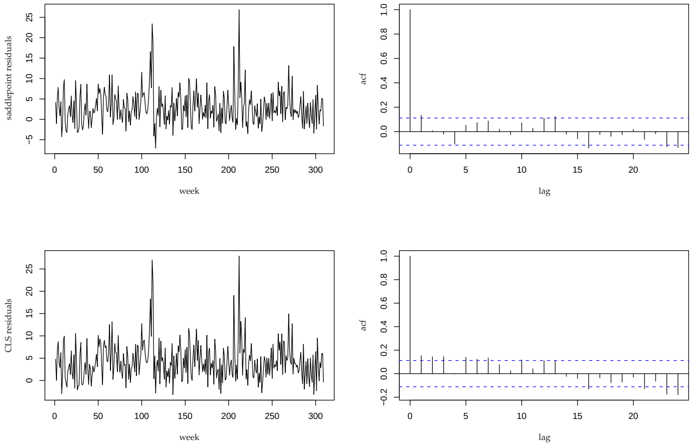
# Results (con't)

Parameter estimates and standard errors (s.e.) obtained after fitting model 9 to the meningococcal disease data.

| | Saddlepoint approximation | | CLS estimation | |
|---|---|---|---|---|
| | estimate | s.e. | estimate | s.e. |
| $\alpha_1$ | 0.189 | 0.052 | 0.216 | 0.057 |
| $\alpha_2$ | 0.196 | 0.051 | 0.096 | 0.059 |
| $\alpha_3$ | 0.185 | 0.049 | 0.092 | 0.057 |
| $\beta_0$ | 1.419 | 0.092 | 1.749 | 0.140 |
| $\beta_{1;1}$ | 0.334 | 0.083 | 0.284 | 0.061 |
| $\beta_{2;1}$ | 0.324 | 0.079 | 0.380 | 0.058 |

# Results (con't)

Time series and acf plots for the SPA and CLS residuals, $e_t = X_t - \sum_{i=1}^{3} \hat{\alpha}_i X_{t-i} - \hat{\mu}_t$, obtained after fitting model 9 to the meningococcal disease data.



# References I


M.A. Al–Osh and A.A. Alzaid.
First–Order Integer–Valued Autoregressive Process.
*Journal of Time Series Analysis*, 8: 261–275, 1987.

R. Bu, B. McCabe and K. Hadri.
Maximum likelihood estimation of higher–order integer–valued autoregressive processes.
*Journal of Time Series Analysis*, 29: 973–994, 2008.

R. Butler.
*Saddlepoint Approximations with Applications*.
Cambridge: Cambridge University Press, 2007.

H.E. Daniels.
Saddlepoint approximations in statistics.
*The Annals of Mathematical Statistics*, 25:631–650, 1954.

A.C. Davison.
*Statistical Models*.
Cambridge: Cambridge University Press, 2003.

C. Field and E. Ronchetti.
*Small Sample Asymptotics. Institute of Mathematical Statistics Lecture Notes–Monograph Series*.
Hayward, CA: Institute of Mathematical Statistics, 1990.

C. Goutis and G. Casella.
Explaining the Saddlepoint Approximation.
*The American Statistician*, 53:216–224, 1999.

# References II


J. Jensen.
*Saddlepoint Approximations*.
London: Oxford University Press, 1995.

J.E. Kolassa.
*Series approximation methods in statistics (3rd. ed.)*.
Springer, New York, 1994.

O. Lieberman.
On the approximation of saddlepoint expansions in statistics.
*Econometric Review*, 10: 900–916.

E. McKenzie.
Some Simple Models for Discete Variate Time Series.
*Water Resources Bulletin*, 21: 645–650, 1985.

M. Paul, L. Held and A. Tonchke.
Multivariate modelling of infectious disease surveillance data.
*Statistics in Medicine*, 27: 6250–6267, 2008.

A.T.A. Wood, J.G. Booth and R.W. Butler.
Saddlepoint approximations to the CDF for some statistics with non–normal limit distributions.
*Journal of the American Statistical Association*, 88:680–686, 1993.