TIME SERIES WITH NON-POSITIVE AUTOCORRELATIONS

ŠÁRKA DOŠLÁ

dosla@karlin.mff.cuni.cz

Department of Statistics, Charles University, Prague

ROBUST 2008

Abstract: We deal with time series models with non-positive autocorrelations. For a strictly stationary process with an autocorrelation function $\{r_k\}_{k=1}^{\infty}$ such that $r_k \le 0$ for all $k \ge 1$ the lower boundary for the sum $\sum_{k=1}^{\infty} r_k$ is investigated. Some results are introduced for general stationary random processes, but the main part of the work is devoted to dependent Bernoulli variables.

Introduction

We deal with strictly stationary time series of dependent Bernoulli variables $\{\xi_t, t \in \mathbb{Z}\}$ with the success probability $p \in (0,1)$

$$\mathsf{P}(\xi_t = 1) = p, \quad \mathsf{P}(\xi_t = 0) = 1 - p \quad \text{for all } t \in \mathbb{Z}$$

and non-positive autocorrelations

$$r_k = \mathsf{corr}\,(\xi_t, \xi_{t+k}) \le 0 \quad \text{for all } k \ge 1.$$

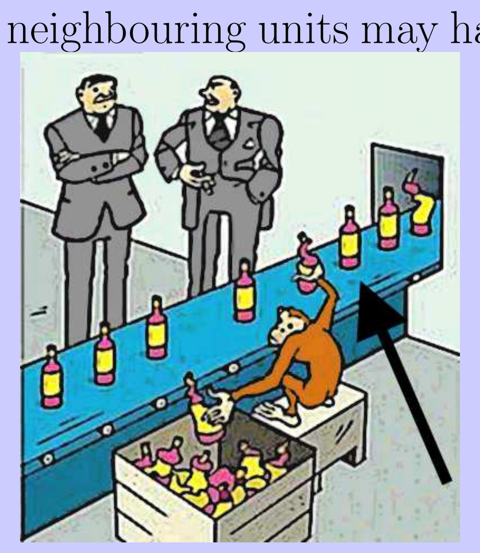
For such processes we investigate the lower bound for the sum of autocorrelations $\sum_{k=1}^{\infty} r_k$.

Motivation

• Processes of dependent Bernoulli variables appear often in practice ⇝ it is worth investigating the properties of their autocorrelation sequences.

• In some situations it is desirable to have statistical models for generating artificial correlated binary data with a specified correlation structure.

Example: Bernoulli sampling is a well-known real time sampling method where the population units are selected independently of each other. However, the neighbouring units may have similar properties and one would like to avoid getting units close to each other too often. Thus, sampling with negative correlations might be **more efficient**. For instance, Meister (2002) describes situations in which some gain is achieved in the efficiency comparing this approach with the classical Bernoulli sampling.



➔ Processes with **non-positive autocorrelations** are of our interest. The following question arises.

Question: How to construct variables ξ_t with autocorrelations $r_k \le 0$ for $k \ge 1$ such that r_k would be **"as negative as possible"**?

➔ We are interested in the lower bound for $\sum_{k=1}^{\infty} r_k$ and in models attaining this bound.

General results

Let $\{r_k\}$ be a sequence of real numbers such that $r_0 = 1$, $r_{-k} = r_k$ for $k \in \mathbb{Z}$ and $r_k \le 0$ for all $k \ge 1$. What are the necessary and sufficient conditions for $\{r_k\}$ to be an autocorrelation sequence of a stationary process?

► Simple for the **Gaussian** case. Bondesson (2003): the sequence $\{r_k\}$ is the autocorrelation function of a stationary Gaussian process $\{X_t\}$ if and only if

$$\sum_{k=1}^{\infty} r_k \ge -\frac{1}{2}. \tag{1}$$

► Anděl and Došlá (2008): The property (1) can be further extended to autocorrelation functions of **more general** processes, as follows.

Theorem 1. *Let $\{X_t, t \in \mathbb{Z}\}$ be a real stationary process with $\mathsf{E}X_t = 0$, $\mathsf{var}\,X_t > 0$ and let the autocorrelation function $\{r_t\}$ be such that $r_t \le 0$ for all $t \ge 1$ and $\sum_{t=-\infty}^{\infty} |r_t| < \infty$. Then the inequality (1) holds.*

The bound $-1/2$ is reached if and only if $f(0) = 0$, where f is the continuous spectral density of the proces $\{X_t\}$, see Anděl and Došlá (2008). This gives a guideline how the "optimal" value $-1/2$ can be reached by an ARMA process.

► **Bernoulli case:** it seems that the bound $-1/2$ for $\sum_{k=1}^{\infty} r_k$ might be **improved**.

Bernoulli variables

The following problem was proposed in Bondesson (2003). It stays **unsolved** even in the simplest case of 1-dependent Bernoulli variables.

Problem: Let $\{r_k\}$ be an autocorrelation sequence of a strictly stationary Bernoulli process and $r_k \le 0$ for all $k \ge 1$. Which value of α is the smallest possible such that $\sum_{k=1}^{\infty} r_k \ge -\alpha$?

1-dependent variables

Consider two Bernoulli variables Y_1 and Y_2 with $p = \mathsf{P}(Y_i = 1)$, $i = 1, 2$, $p \in (0,1)$ and set $q = 1 - p$. Then

$$\mathsf{corr}\,(Y_1, Y_2) = \frac{\mathsf{P}(Y_1 = 1, Y_2 = 1) - p^2}{pq} \ge \frac{-p^2}{pq} = -\frac{p}{q}.$$

Since Y_1, Y_2 and $Y_1^* = 1 - Y_1, Y_2^* = 1 - Y_2$ have the same correlation ⇝ we get $\mathsf{corr}\,(Y_1, Y_2) \ge -q/p$ as well.

➔ This and Theorem 1 give us **the lower bound for the 1 lag autocorrelation** r_1 of a 1-dependent Bernoulli process $\{\xi_t\}$ with the success probability p

$$r_1 \ge \max\left\{-\frac{p}{1-p}, -\frac{1-p}{p}, -\frac{1}{2}\right\}. \tag{2}$$

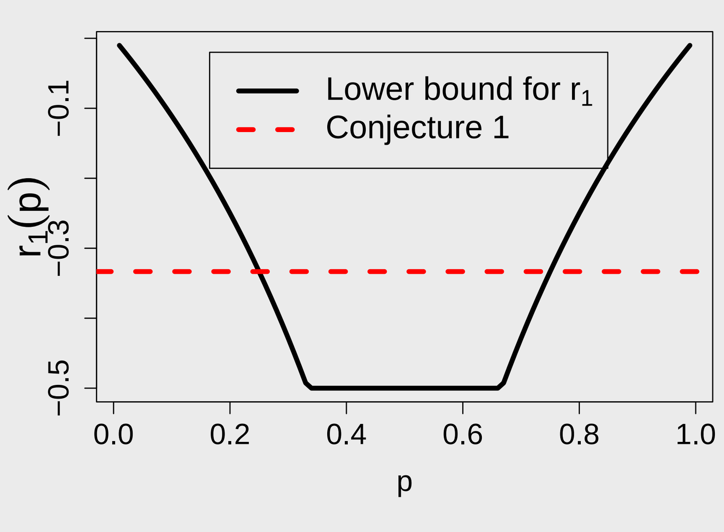
Joe (1997) shows how the bound (2) can be attained for $p \le 1/4$ and $p \ge 3/4$. However, a model with r_1 reaching the lower bound in (2) for $p \in (1/4, 3/4)$ has not been found ➔ it seems that **sharper bounds** for r_1 for $p \in (1/4, 3/4)$ **might be derived.**

Bondesson (2003) stated the following conjecture about r_1 of 1-dependent variables $\{\xi_t\}$.

Conjecture 1. *Bondesson (2003)*

$$r_1 \ge -1/3 \textit{ for all } p \in (0,1). \tag{3}$$

Conjecture 1 has not been proved in the general case of 1-dependent Bernoulli variables, but the inequality (1) **holds for some broad classes of processes**. See Bondesson (2003) and Anděl and Došlá (2008).



General Bernoulli variables

Although the problem stays unsolved in general, it is possible to find the lower bound for $\sum_{t=1}^{\infty} r_t$ for some classes of processes. A common model for generating dependent Bernoulli variables is based on clipping a sequence of correlated Gaussian variables.

Theorem 2. *Let $\{Z_t\}$ be a stationary time series of Gaussian random variables with $\mathsf{E}\,Z_t = 0$ and the autocorrelation function $\{\rho_t\}$ such that $\rho_t \le 0$ for all $t \ge 1$. Define Bernoulli variables $\{\xi_t\}$ as $\xi_t = I[Z_t \ge 0]$ and denote $\{\rho_t^*\}$ the autocorrelation function of the process $\{\xi_t\}$. Then $\rho_t^* \le 0$ for all $t \ge 1$ and*

$$\sum_{t=1}^{\infty} \rho_t^* \ge -\frac{1}{3}. \tag{4}$$

Remarks

• The inequality (4) holds for $\mathsf{E}Z_t \ne 0$ as well. The proof of Theorem 2 can be found in Došlá (2008).

• The bound $-1/3$ is reached for m-dependent variables if and only if $\rho_k = -1/2$ for some $k \in \{1, \dots, m\}$ and $r_t = 0$ otherwise. For any other setting, the sum of correlations lies above $-1/3$.

Conjecture 1 and Theorem 2 may lead us to the following question.

Question: Is $\sum_{k=1}^{\infty} r_k$ always above $-1/3$ for general dependent Bernoulli variables?

However, the answer is no. The following Example 1 shows a model of m-dependent variables, $m \ge 2$, with $\sum_{k=1}^{\infty} r_k$ below $-1/3$.

Example 1. Let X_t be iid with the uniform distribution on $[0,1]$ and for $m \in \mathbb{N}$, $c = 1/(m+1)$ define

$$\xi_t = I[X_t < c, X_{t-1} > c, \dots, X_{t-m} > c].$$

Then the process $\{\xi_t\}$ is m-dependent stationary with autocorrelations $r_k < 0$ for $k = 1, \dots, m$ and $r_k = 0$ for $k > m$. Furthermore,

$$\lim_{m\to\infty} \sum_{k=1}^{\infty} r_k = \lim_{m\to\infty} \sum_{k=1}^{m} r_k = \lim_{m\to\infty} m r_1 = -\frac{1}{e}.$$

The sequence $\{\sum_{k=1}^{m} r_k\}_{m=1}^{\infty}$ is decreasing and it dips below $-1/3$ already for $m = 2$. See Bondesson (2003), cf. Došlá (2008).

Hence, for any fixed $m \ge 2$ we are able to construct a sequence $\{\xi_t\}$ of Bernoulli variables with non-positive autocorrelations $\{r_k\}$ and $\sum_{k=1}^{\infty} r_k < -1/3$.

Theorem 2 + Example 1 ➔ Generating correlated Bernoulli variables by **clipping a Gaussian process** (e.g. package `binary` in program `R`) is **not suitable** for our situation because a more efficient approach exists.

The following problem stays unsolved.

Open problem: Is $\alpha = 1/e$ the lowest possible value for a general process with Bernoulli variables?

Conclusion

The lower bound for the sum of autocorrelations $\sum_{t=1}^{\infty} r_t$ is investigated for stationary time series.

► For general stationary processes we show that the inequality $\sum_{t=1}^{\infty} r_t \ge -1/2$ holds and the lower bound $-1/2$ can be easily attained by an ARMA process.

► The situation is more complex for series of dependent Bernoulli variables. The sharp lower bound for $\sum_{t=1}^{\infty} r_t$ stays an open problem in the general case.

➔ There are still topics for further research.

Acknowledgments

This poster is based on a joint work with Professor Jiří Anděl. He is gratefully thanked for his advice, comments and suggestions.

This work is supported by grant 56808 of the Grant Agency of Charles University in Prague and grant 201/05/H007 of the Grant Agency of the Czech Republic.

References

Anděl, J. and Došlá, Š., Bernoulli processes with negative correlations, *submitted*, 2008.

Bondesson, L., On a minimum correlation problem, *Stat. and Probab. Letters*, *62*, 361–370, 2003.

Došlá, Š., Processes with negative correlations, *submitted*, 2008.

Joe, H., *Multivariate Models and Dependence Concepts*, Chapman and Hall, London, 1997.

Meister, K., Asymptotic considerations concerning real time sampling methods, *Statistics in Transition*, *5*, 1037–1050, 2002.