RT-MAE 2009-07

A NOTE ON SEQUENCES OF NON CORRELATED DEPENDENT RANDOM VARIABLES

by

José Carlos Simon de Miranda

Palavras-Chave: Non correlated sequences, discrete random variables.

Classificação AMS: 60E05.

- Outubro de 2009 -

A note on Sequences of Non Correlated Dependent Random Variables

José Carlos Simon de Miranda

Institute of Mathematics and Statistics
University of São Paulo
e-mail: simon@ime.usp.br

Abstract: In this note we present a constructive proof of the existence of sequences, and finite dimensional vectors as a particular case, of non correlated dependent discrete random variables with arbitrary, possibly different, one dimensional marginals under the only requirement that each random variable in the sequence, or in the vector, presents at least four distinct values at its image.

keywords:. non correlated sequences, discrete random variables.
Mathematics Subject Classification: AMS 60E05

1. Introduction

In this short note we present a proof of the existence of sequences of discrete real random variables with the following property: Every pair of random variables in the sequence is dependent but zero correlated. Let X be a real discrete random variable. We will denote by $ImX = \{x \in \mathbb{R} : \mathbb{P}(X = x) > 0\}$ the image of this random variable. We will also write $ImX = \{x_i \in \mathbb{R} : i \in I \subset \mathbb{N}\}$. The probability function of X is denoted by f_X. If $(X_i)_{1\leq i\leq k}$ is a discrete random vector then its joint probability function is denoted either by $f_{(X_i)_{1\leq i\leq k}}$ or by $f_{X_1...X_k}$. We will denote $(X|A)$ the random variable whose probability function is $f_{(X|A)}(x) = \frac{\mathbb{P}(X=x,A)}{\mathbb{P}(A)}$, where A is an event with positive probability. Of special interest will be the case $(X_k|X_{k-1} = x_{k-1}, .., X_1 = x_1)$. We use $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$.

2. Main Result

Theorem 2.1. *Let, for all $n \in \mathbb{N}^*$, Y_n be a discrete real random variable such that $\#ImY_n \geq 4$. Then, there exists a sequence $(X_n)_{n\in\mathbb{N}^*}$, such that for all $n \in \mathbb{N}^*$, $X_n \sim Y_n$, with the following property: For every pair of distinct indexes, i and j, X_i and X_j are dependent but $Cov(X_i, X_j) = 0$.*

Proof Write $I(n) = \{k \in \mathbb{N} : 0 \leq k \leq \#ImY_n - 1\}$ if ImY_n is finite and $I(n) = \mathbb{N}$ otherwise. Let us denote $ImY_n = \{y_{n,m} : m \in I(n)\}$. Since the property of pairwise dependence and zero correlation for all pairs in the sequence is invariant under the addition of constants to the random variables, we may assume without loss of generality that for all $n \in \mathbb{N}^*$ $0 \in ImY_n$. We will also write, for all $n \in \mathbb{N}^*$, $y_{n,1} < y_{n,2} < y_{n,3}$.

For all $k \in \mathbb{N}^*$, let $\{e_i : 1 \leq i \leq k\}$ be the canonical basis of $\mathbb{R}^k$ and, for $1 \leq i \leq k$, $\pi_i : \mathbb{R}^k \to \mathbb{R}$, $\pi_i(x_1, ..., x_k) = x_i$, be the canonical projection. Let $h_k(y_1, ..., y_k) = \prod_{i=1}^k f_{Y_i}(y_i)$.

Now, for all $k \in \mathbb{N}$, $k \geq 2$, we can construct a vector $(X_1, ..., X_k)$ with the required property of pairwise dependence and zero correlation. This can be done by taking, for all i, $1 \leq i \leq k$, $ImX_i = \{x_{i,m} = y_{i,m} : 0 \leq m \leq I(i)\} = ImY_i$ and $f_{X_1...X_k}$ defined on $\prod_{i=1}^k ImX_i$ given by the following relations:

For all i and j, $1 \leq i < j \leq k$,

$$f_{X_1...X_k}(x_{i,1}e_i + x_{j,1}e_j) = h_k(x_{i,1}e_i + x_{j,1}e_j) - \delta_{k,i,j},$$
$$f_{X_1...X_k}(x_{i,2}e_i + x_{j,1}e_j) = h_k(x_{i,2}e_i + x_{j,1}e_j) + \delta_{k,i,j},$$
$$f_{X_1...X_k}(x_{i,1}e_i + x_{j,2}e_j) = h_k(x_{i,1}e_i + x_{j,2}e_j) + \delta_{k,i,j},$$
$$f_{X_1...X_k}(x_{i,2}e_i + x_{j,2}e_j) = h_k(x_{i,2}e_i + x_{j,2}e_j) - \delta_{k,i,j} + \epsilon_{k,i,j},$$
$$f_{X_1...X_k}(x_{i,3}e_i + x_{j,2}e_j) = h_k(x_{i,3}e_i + x_{j,2}e_j) - \epsilon_{k,i,j},$$
$$f_{X_1...X_k}(x_{i,2}e_i + x_{j,3}e_j) = h_k(x_{i,2}e_i + x_{j,3}e_j) - \epsilon_{k,i,j},$$
$$f_{X_1...X_k}(x_{i,3}e_i + x_{j,3}e_j) = h_k(x_{i,3}e_i + x_{j,3}e_j) + \epsilon_{k,i,j},$$

where $\delta_{k,i,j}$ and $\epsilon_{k,i,j}$ are such that

$$0 < \delta_{k,i,j} < h_k(x_{i,1}e_i + x_{j,1}e_j),$$
$$0 < \epsilon_{k,i,j} < min\{h_k(x_{i,3}e_i + x_{j,2}e_j), h_k(x_{i,2}e_i + x_{j,3}e_j)\},$$
$$0 \leq h_k(x_{i,2}e_i + x_{j,2}e_j) - \delta_{k,i,j} + \epsilon_{k,i,j},$$

and

(1) $$(x_{i,2}-x_{i,1})(x_{j,2}-x_{j,1})(-\delta_{k,i,j})+(x_{i,3}-x_{i,2})(x_{j,3}-x_{j,2})\epsilon_{k,i,j}=0,$$

and

$$f_{X_1\dots X_k}(w)=h_k(w)$$

for all $w\in\prod_{i=1}^k ImX_i\setminus\bigcup_{(i,j):1\le i<j\le k}\{(x_{i,1}e_i+x_{j,1}e_j),(x_{i,2}e_i+x_{j,1}e_j),(x_{i,1}e_i+x_{j,2}e_j),(x_{i,2}e_i+x_{j,2}e_j),(x_{i,3}e_i+x_{j,2}e_j),(x_{i,2}e_i+x_{j,3}e_j),(x_{i,3}e_i+x_{j,3}e_j)\}$.

It is easy to verify that there exists $\Delta_{k,i,j}>0$ such that for every $\delta_{k,i,j}\in(0,\Delta_{k,i,j})$ the inequalities above are fulfilled by $\delta_{k,i,j}$ and $\epsilon_{k,i,j}$ given by equation (1).

Clearly, for all $i\neq j$, X_i and X_j are dependent.

Denote $\langle S\rangle$ the subspace of $\mathbb{R}^k$ generated by the set $S\subset\mathbb{R}^k$. To simplify notation, denote $f_k=f_{X_1\dots X_k}$.

Now, for all $m\notin\{1,2,3\}$,

$$f_{X_i}(x_{i,m})=\sum_{w\in\pi_i^{-1}(x_{i,m})}f_k(w)=\sum_{w\in\pi_i^{-1}(x_{i,m})}h_k(w)=f_{Y_i}(x_{i,m}).$$

We also have

$$f_{X_i}(x_{i,1})=\sum_{w\in\pi_i^{-1}(x_{i,1})\setminus(\cup_{j:j\neq i,1\le j\le k}(\{e_i,e_j\}))}f_k(w)+\sum_{w\in\pi_i^{-1}(x_{i,1})\cap(\cup_{j:j\neq i,1\le j\le k}(\{e_i,e_j\}))}f_k(w)=$$

$$\sum_{w\in\pi_i^{-1}(x_{i,1})\setminus(\cup_{j:j\neq i,1\le j\le k}(\{e_i,e_j\}))}h_k(w)+\sum_{w\in\pi_i^{-1}(x_{i,1})\cap(\cup_{j:j\neq i,1\le j\le k}(\{e_i,e_j\}))}h_k(w)+\sum_{j:j\neq i,1\le j\le k}(-\delta_{k,i,j}+\delta_{k,i,j})=f_{Y_i}(x_{i,1}),$$

and, analogously,

$$f_{X_i}(x_{i,2})=\sum_{w\in\pi_i^{-1}(x_{i,2})\setminus(\cup_{j:j\neq i,1\le j\le k}(\{e_i,e_j\}))}h_k(w)+\sum_{w\in\pi_i^{-1}(x_{i,2})\cap(\cup_{j:j\neq i,1\le j\le k}(\{e_i,e_j\}))}h_k(w)+$$

$$\sum_{j:j\neq i,1\le j\le k}(\delta_{k,i,j}+(-\delta_{k,i,j}+\epsilon_{k,i,j})-\epsilon_{k,i,j})=f_{Y_i}(x_{i,2}),$$

and

$$f_{X_i}(x_{i,3})=\sum_{w\in\pi_i^{-1}(x_{i,3})\setminus(\cup_{j:j\neq i,1\le j\le k}(\{e_i,e_j\}))}h_k(w)+\sum_{w\in\pi_i^{-1}(x_{i,3})\cap(\cup_{j:j\neq i,1\le j\le k}(\{e_i,e_j\}))}h_k(w)+$$

$$\sum_{j:j\neq i,1\le j\le k}(-\epsilon_{k,i,j}+\epsilon_{k,i,j})=f_{Y_i}(x_{i,3}).$$

Thus, for all i, $1\le i\le k$, $f_{X_i}=f_{Y_i}$, i.e., $X_i\sim Y_i$.

Now, for all i and j, $1\le i<j\le k$,

$$\mathbb{E}(X_iX_j)=\sum_{(x_1,\dots,x_k)\in(\prod_{l=1}^k ImX_l)\setminus(\cup_{m,n:1\le m<n\le k}(\{e_m,e_n\}))}x_ix_jh_k(x_1,\dots,x_k)+$$

$$\sum_{(x_1,\dots,x_k)\in(\prod_{l=1}^k ImX_l)\cap(\cup_{m,n:1\le m<n\le k}(\{e_m,e_n\}))}x_ix_jf_k(x_1,\dots,x_k)=$$

$$\sum_{(x_1,\dots,x_k)\in(\prod_{l=1}^k ImX_l)\setminus(\cup_{m,n:1\le m<n\le k}(\{e_m,e_n\}))}x_ix_jh_k(x_1,\dots,x_k)+\sum_{(x_1,\dots,x_k)\in(\prod_{l=1}^k ImX_l)\cap(\{e_i,e_j\})}x_ix_jf_k(x_1,\dots,x_k)=$$

$$\sum_{(x_1,\dots,x_k)\in(\prod_{l=1}^k ImX_l)\setminus(\cup_{m,n:1\le m<n\le k}(\{e_m,e_n\}))}x_ix_jh_k(x_1,\dots,x_k)+\sum_{(x_1,\dots,x_k)\in(\prod_{l=1}^k ImX_l)\cap(\{e_i,e_j\})}x_ix_jh_k(x_1,\dots,x_k)+$$

$$x_{i,1}x_{j,1}(-\delta_{k,i,j})+x_{i,2}x_{j,1}\delta_{k,i,j}+x_{i,1}x_{j,2}\delta_{k,i,j}+x_{i,2}x_{j,2}(-\delta_{k,i,j}+\epsilon_{k,i,j})+x_{i,3}x_{j,2}(-\epsilon_{k,i,j})+x_{i,2}x_{j,3}(-\epsilon_{k,i,j})+x_{i,3}x_{j,3}\epsilon_{k,i,j}=$$

$$\sum_{(x_1,\dots,x_k)\in\prod_{l=1}^k ImX_l}x_ix_jh_k(x_1,\dots,x_k)+0=\mathbb{E}X_i\mathbb{E}X_j,$$

by relation (1). Thus, $\mathbb{E}(X_iX_j)-\mathbb{E}X_i\mathbb{E}X_j=0$.

Now, to generate the sequence we proceed inductively. Take x_1 a realization of the random variable $X_1\sim Y_1$. Then, for all $k\ge1$ we take, independently from all previous realizations,$x_1,\dots x_k$, a realization x_{k+1} of the random variable $X_{k+1}\sim(Z_{k+1}|Z_k=x_k,\dots,Z_1=x_1)$ where $(Z_1,\dots,Z_{k+1})$ has joint probability function f_{k+1}. Let us prove that the sequence $(X_n)_{n\in\mathbb{N}}$ has the following property: For all $m \in \mathbb{N}$, $m \geq 2$, $(X_n)_{1\leq n\leq m}$ has probability function f_m. We will prove this by induction. Observe also, that $f_{X_1} = f_{Y_1}$. Denote $f_1 = f_{X_1}$.

For $m = 2$, we have

$$f_{X_1X_2}(x_1,x_2) = \mathbb{P}(X_2 = x_2, X_1 = x_1) = \mathbb{P}((Z_2|Z_1 = x_1) = x_2)\mathbb{P}(X_1 = x_1) =$$
$$f_{(Z_2|Z_1=x_1)}(x_2)f_1(x_1) = \frac{f_2(x_1,x_2)}{f_1(x_1)}f_1(x_1) = f_2(x_1,x_2).$$

If the law of $(X_n)_{1\leq n\leq k}$ is f_k we have

$$f_{X_1\ldots X_{k+1}}(x_1,\ldots,x_{k+1}) = \mathbb{P}((Z_{k+1}|Z_k = x_k,\ldots,Z_1 = x_1) = x_{k+1})\mathbb{P}(X_k = x_k,\ldots,X_1 = x_1) =$$
$$f_{(Z_{k+1}|Z_k=x_k,\ldots,Z_1=x_1)}(x_{k+1})f_k(x_1,\ldots,x_k) =$$
$$\frac{f_{k+1}(x_1,\ldots,x_{k+1})}{f_k(x_1,\ldots,x_k)}f_k(x_1,\ldots,x_k) = f_{k+1}(x_1,\ldots,x_{k+1}),$$

that is, the law of $(X_n)_{1\leq n\leq k+1}$ is f_{k+1}.

Thus, since their laws are f_m, all vectors $(X_n)_{1\leq n\leq m}$ in the sequence $(X_n)_{n\in\mathbb{N}^*}$ have the desired properties $f_{X_n} = f_{Y_n}$ and the random variables X_i and X_j are dependent but obey $Cov(X_i,X_j) = 0$, for all n, i, j, and m, $1 \leq n \leq m$, $1 \leq i < j \leq m$, $m \geq 2$. So, the sequence $(x_n)_{n\in\mathbb{N}^*}$ is a realization of $(X_n)_{n\in\mathbb{N}^*}$, a sequence of dependent random variables with one dimensional marginals distributed, for each n, like Y_n and satisfying $Cov(X_i,X_j) = 0$ whenever $i \neq j$.

■

The following corollary is important in the construction of some non internally correlated point processes. See [1].

Corollary 2.1. *For every sequence of strictly positive real numbers, $(\lambda_n)_{n\in\mathbb{N}}$, there exists a sequence of dependent Poisson random variables $(X_n)_{n\in\mathbb{N}}$, $X_n \sim Poisson(\lambda_n)$, such that, for all i and j in $\mathbb{N}$, if $i \neq j$ then $Cov(X_i,X_j) =$* 0.

3. Final Remark

Observe that for every partition of the natural numbers, $\{I_j\}_{j\in J}$, we can form a sequence of random variables $(X_n)_{n\in\mathbb{N}}$, $X_n \sim Y_n$ with the following properties: for all i and j in $\mathbb{N}$, if $i \neq j$ we have $Cov(X_i,X_j) = 0$; the subsequences $(X_n)_{n\in I_j}$, $j \in J$, are independent; there is dependence within all the variables that belong to the same subsequence. To form such sequence we can proceed in the following ways: First, for all $j \in J$, generate independently the sequences, or vectors, $(X_n)_{n\in\mathbb{N}}$, if $\#I_j = \#\mathbb{N}$, and $(X_n)_{0\leq n\leq \#I_j-1}$, if $\#I_j < \infty$, then apply the necessary reordering of indexes; alternatively, proceed as in the proof of Theorem 2.1, but choose $\delta_{k,i,j} = 0$ whenever i and j belong to different sets that form the partition of $\mathbb{N}$.

Acknowledgement. The author thanks our Lord and Saviour Jesus Christ for His Love and Mercy.

References

[1] de Miranda, J.C.S. (2009). Non Internally Correlated Point Processes Technical Report N. 6, Department of Statistics, Institute of Mathematics and Statistics, University of São Paulo.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2009-01 - ALVAREZ, N.G.G., BUENO, V.C. Optimal burn-in time under an extended general failure model. 2009. 17p.(RT-MAE-2009-01)

2009-02 - ALVAREZ, N.G.G., BUENO, V.C. Estimating a warranty discounted cost of a minimally repaired coherent system. 2009.26p.(RT-MAE-2009-02)

2009-03 - BELITSKY, V., PEREIRA, A.L. Stability analysis with applications of a two-dimensional dynamical system arising from a stochastic model for an asset market. (RT-MAE-2009-03)

2009-04 - BELITSKY, V., DAWID, P.E., PRADO, F.P.A. When and how heterogeneity of social susceptibility of consumers causes multiplicity of population relative excess demands. 34p. (RT-MAE-2009-04)

2009-05 - PORTO, R.F., MORETTIN, P.A., PERCIVAL, D.B., AUBIN, E.C.Q. Wavelet Shrinkage for Regression Models with Random Design and Correlated Errors. 28p. (RT-MAE-2009-05)

2009-06 - MIRANDA J.C.S. Non internally correlated point process. 07 p. (RT-MAE-2009-06)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05314-970 - São Paulo, Brasil