Sankhyā : The Indian Journal of Statistics
1985, Volume 47, Series A, Pt. 2, pp. 166-173.

A CHARACTERIZATION OF INFINITE GAUSSIAN SEQUENCES BY CONDITIONAL MOMENTS

By WLODZIMIERZ BRYC and AGNIESZKA PLUCIŃSKA
Instytut Matematyki, Politechnika Warszawska, Poland

SUMMARY. In this paper we study random sequences which have linear regression and non-random conditional variance.

I. Introduction

It is well known, that each gaussian sequence $(X_i)_{i \in N}$ for every integer n has the following properties : regression is linear, i.e.

LR. $$E(X_i \mid X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$$

is almost everywhere a linear function of

$$X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n \ (i = 1, \ldots, n);$$

conditional variance is a.e. constant, i.e.

CV. $$\operatorname{var}(X_i \mid X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$$

is nonrandom $(i = 1, \ldots, n)$.

In this note we prove, that under some additional assumptions on the covariance function of the sequence $(X_i)_{i \in N}$, conditions LR and CV characterize infinite gaussian sequences. We prove also, that for two linearly independent and correlated random variables, X_1, X_2 conditions LR and CV imply p-integrability for each $p \geqslant 0$.

Conditions close to LR and CV (but weaker) were used by Laha (1957) for the characterization of the gaussian distribution of random vector (X_1, X_2) in the case, when X_1 and X_2 are linear functions of the fixed finite sequence of square integrable independent random variables.

Conditions LR and CV were used by Plucińska (1983) for the characterization of gaussian processes with sufficiently smooth covariance. Our main result, Theorem 1, complements the paper of Plucińska (1983) in the purely discontinuous case.

AMS (1980) *subject classification :* 62E10; secondary ; 60E15.

Key words and phrases : Characterization problems, conditional moments, gaussian sequences, existence of moments, regression.

In this paper we assume that all random variables in consideration are defined on the same fixed probability space. By $\rho(\cdot, \cdot)$ we denote the correlation coefficient, $E(\cdot \mid X)$ as usual stands for the conditional expectation with respect to random variable X. Recall also that random variables $X_1, X_2, \ldots$ are linearly independent, if for each m functions $X_1, \ldots, X_m : \Omega \to R$ are linearly independent. Note that linear independence is a restriction on the covariance function of $(X_i)_{i \in N}$, provided that second moments exist. To simplify the formulation of Theorem 1 we introduce the following abreviation.

We say that the sequence $(X_i)_{i \in N}$ of integrable random variables satisfies condition LR*, if each finite subsequence of $(X_i)_{i \in N}$ satisfies LR and for each $n \geqslant 1$ in the expression

$$E(X_{n+1} \mid X_1, \ldots, X_n) = a_0 + \sum_{k=1}^{n} a_k X_k$$

we have

$$(*) \qquad a_n \neq 0.$$

Note that by formula (6) below, condition (*) is a property of the covariance function of sequence $(X_i)_{i \in N}$, provided that LR holds for each finite subsequence and $X_1, X_2, \ldots$, are linearly independent with finite second moments.

We say, that infinite sequence $(X_i)_{i \in N}$ of square-integrable random variables satisfies condition CV, if each finite subsequence of $(X_i)_{i \in N}$ satisfies CV.

Remark: It can be checked that, if for some $n > 1$, $X_1, \ldots, X_n$ are linearly independent and satisfy conditions LR and CV, then each subsequence of $(X_1, \ldots, X_n)$ satisfies conditions LR and CV.

Theorem 1 : *Let $(X_i)_{i \in N}$ be a sequence of square-integrable random variables and assume that $X_1, X_2, \ldots$ are linearly independent and satisfy LR* and CV conditions. Then $(X_i)_{i \in N}$ is the gaussian sequence.*

The following result of independent interest will be used as the lemma in the proof of Theorem 1.

Theorem 2 : *Let X_1, X_2 be square-integrable random variables satisfying LR and such that $0 < |\rho(X_1, X_2)| < 1$. If conditional variances $var(X_1 \mid X_2)$ and $var(X_2 \mid X_1)$ are a.e. bounded by some constant, then X_1 and X_2 are p-integrable for each $p \geqslant 0$.*

Without losing generality in the proofs of both theorems we assume $E(X_i) = 0$, $E(X_i^2) = 1$, $i = 1, 2, \ldots$, or $i = 1, 2$.

Proof of Theorem 2 : Consider function

$$N(t) = P(|X_1| > t)+P(|X_2| > t), \ t \geqslant 0.$$

It is enough to prove that there exist constants C, K, $M > 0$ such that for each $t > M$

$$N(Kt) \leqslant C/t^2 \, N(t). \quad \ldots \quad (1)$$

Indeed, by integration by parts for each $p \geqslant 1$

$$\sum_i E(|X_i|^{p+2}) = \sum_i (p+2) \int_0^\infty t^{p+1} P(|X_i| > t)dt = (p+2) \int_0^\infty t^{p+1} N(t)dt.$$

By simple substitution, trivial inequality $N(t) \leqslant 2$ and (1) imply

$$\sum_i E(|X_i|^{p+2}) \leqslant (p+2)K^{p+2} \int_0^M 2t^{p+1}dt + (p+2)K^{p+2} \int_M^\infty Ct^{p-1}N(t)dt$$

$$\leqslant K^{p+2} \left[2M^{p+2} + C\frac{p+2}{p} \sum_i E(|X_i|^p) \right],$$

which implies the finiteness of the moments of all orders, because by assumption $E(X_1^2)+E(X_2^2) < \infty$.

It remains to prove (1). First let us observe that, if $X_i' = \pm X_i$, then (X_1', X_2') also satisfies the assumptions of the theorem and it suffices to prove (1) for any of the pairs (X_1', X_2'). The later will be often written without primes.

Let $\rho = \rho(X_1, X_2)$. Considering $X_1' = -X_1$, if necessary, we may assume $0 < \rho < 1$. Then by LR

$$E(X_i | X_j) = \rho X_j, \ i, j = 1, 2, \ i \neq j. \quad \ldots \quad (2)$$

By (2) and the assumption on conditional variance for some constant C_1

$$E((X_i - \rho X_j)^2 | X_j) \leqslant C_1, \ i, j = 1, 2. \quad \ldots \quad (3)$$

Let $K = 2/\rho$, $P = P(X_1 > Kt)$. Then

$$P \leqslant P(X_1 > t, X_2 > t) + P(X_1 > Kt, X_2 \leqslant t) = P_1 + P_2 \text{ (say)}.$$

It is easily seen that

$$P_2 \leqslant P(X_1 > Kt, X_2 - \rho X_1 < -t) \leqslant P(X_1 > Kt, |X_2 - \rho X_1| > t)$$
$$= \int_{X_1 > Kt} P(|X_2 - \rho X_1| > t | X_1)dP.$$

By the conditional version of Chebyshev inequality applied to the right hand side of the above inequality and by (3) we have

$$P_2 \leqslant C_1/t^2 \, P,$$

Hence $P(1-C_1/t^2) \leqslant P_1$ and for $t > M = (2C_1)^{1/2}$

$$P \leqslant 2P_1. \qquad \ldots \quad (4)$$

To bound P_1 we use the following decomposition.

$$\begin{aligned} P_1 &= P(X_1 > t,\, X_2-\rho X_1 > (1-\rho)t,\, X_2 > t) \\ &\quad + P(X_2 > t,\, X_1-\rho X_2 > (1-\rho)t,\, X_2-\rho X_1 \leqslant (1-\rho)t,\, X_1 > t) \\ &\quad + P(X_1 > t,\, X_2 > t,\, X_1-\rho X_2 \leqslant (1-\rho)t,\, X_2-\rho X_1 \leqslant (1-\rho)t). \end{aligned}$$

Note that the last term on the right hand of this equality vanishes. Indeed, since $0 < \rho < 1$, hence inequalities

$$X_2-\rho X_1 \leqslant (1-\rho)t,$$
$$X_1-\rho X_2 \leqslant (1-\rho)t,$$

after multiplying one of them by ρ and adding to the other, imply that $X_1 \leqslant t$ and $X_2 \leqslant t$. Therefore

$$P_1 \leqslant P(X_1 > t,\, X_2-\rho X_1 > (1-\rho)t) + P(X_2 > t,\, X_1-\rho X_2 > (1-\rho)t)$$

and since $1-\rho > 0$ we have

$$P_1 \leqslant P(X_1 > t,\, |X_2-\rho X_1| > (1-\rho)t) + P(X_2 > t,\, |X_1-\rho X_2| > (1-\rho)t).$$

Once more using the conditional Chebyshev inequality and (3) we get the bound

$$P_1 \leqslant \frac{C_1}{(1-\rho)^2}\, \frac{N(t)}{t^2}.$$

This jointly with (4) gives for each $t > M$

$$P(X_1 > Kt) \leqslant \frac{C_2}{t^2}\, N(t), \qquad \ldots \quad (5)$$

where C_2 is some constant.

Inequality (5) and the same inequality applied to the pair $(-X_1, -X_2)$ instead of (X_1, X_2) gives $P(|X_1| > Kt) \leqslant 2C_2\, N(t)/t^2$. By the symmetry the reasoning can be repeated for the pair (X_2, X_1) proving (1) with $C = 4C_2 = 4C_1(1-|\rho|)^{-2}$.

2. Notation and lemmas for the proof of theorem 1

Let $k \leqslant n$. By LR condition for $(X_1, \ldots, X_n)$ and by mean zero assumption we have

$$E(X_k \mid X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n) = \sum_{\substack{i=1 \\ i\neq k}}^{u} a_{i,k,n} X_i.$$

It is known (see Plucińska, 1983, formula (5)) that if $X_1, \ldots, X_n$ are linearly independent, then

$$a_{i,k,n} = (-1)^{i+k+1} \frac{\det[\rho_{\mu,\gamma}]_{\mu,\gamma=1,\ldots,n;\ \mu\neq i,\ \gamma\neq k}}{\det[\rho_{\mu,\nu}]_{\mu,\gamma=1,\ldots,n;\ \mu,\gamma\neq k}} \quad \ldots \quad (6)$$

where $\rho_{\mu,\gamma} = \rho(X_\mu, X_\gamma)$, $\mu, \gamma = 1, 2, \ldots, n$.

Lemma 1: *If $Y_1, \ldots, Y_n$ are linearly independent square-integrable random variables such that for some $k < n$ both sequences $(Y_1, \ldots, Y_k)$ and $(Y_1, \ldots, Y_n)$ satisfy LR condition, then*

$$E(Y_1 \mid Y_2, \ldots, Y_n) - E(Y_1 \mid Y_2, \ldots, Y_k) = \sum_{i=k+1}^{n} a_{i,1,n}(Y_i - E(Y_i \mid Y_2, \ldots, Y_k)),$$

where $a_{i,1,n}$ are defined by formula (6) applied to $Y_1, \ldots, Y_n$ instead of $X_1, \ldots, X_n$.

This lemma is known in the gaussian case. For $k = n-1$ the lemma was proved by Plucińska (1983), see her Lemma 2 formula (6). The general case of arbitrary k follows by induction.

In the proof of Theorem 1 we will use only very special conditionings. Therefore we denote by $\mathcal{E}_{n-1}(\cdot)$ the conditional expectation with respect to $X_1, \ldots, X_{n-1}$, $n = 1, 2, \ldots$ ($\mathcal{E}_0(\cdot) = E(\cdot)$ by definition) and let $\mathcal{E}_{n-1, X_{n+1}}(\cdot)$ denote conditional expectation with respect to $X_1, \ldots, X_{n-1}, X_{n+1}$. We need the following corollary to the well known theorem, that gaussian distribution is uniquely determined by its moments.

Lemma 2: *Let $(X_i)_{i \in N}$ be a sequence of random variables with finite moments of all orders and such that $X_1, \ldots, X_n$ for each n are linearly independent and satisfy LR condition. If centered conditional moments $\mathcal{E}_{n-1}(X_n - \mathcal{E}_{n-1}(X_n))^k$, $k, n = 1, 2, \ldots$ are a.e. equal to the corresponding centered conditional moments of a gaussian sequence with the same covariance function as $(X_i)_{i \in N}$, then $(X_i)_{i \in N}$ is the gaussian sequence.*

Proof: Let $(Y_i)_{i \in N}$ be a mean zero gaussian sequence with the same covariance function as $(X_i)_{i \in N}$. By the well known properties of the gaussian distribution, it suffices to prove, that for every $k, n_1, \ldots, n_k \in N$ there holds

$$E(X_1^{n_1} \ldots X_k^{n_k}) = E(Y_1^{n_1} \ldots Y_k^{n_k}). \quad \ldots \quad (7)$$

We will prove (7) by induction. Since $E(X_i) = 0$ and $\mathcal{E}_0 = E$, (7) holds for $k = 1$ by assumption. If (7) holds for some $k \in N$, then

$$m = E(X_1^{n_1} \ldots X_k^{n_k} X_{k+1}^{n_{k+1}})$$

$$= \sum_{i=1}^{n_{k+1}} \binom{n_{k+1}}{i} E(X_1^{n_1} \ldots X_k^{n_k}) \mathcal{E}_k[X_{k+1} - \mathcal{E}_k(X_{k+1})]^i [\mathcal{E}_k(X_{k+1})]^{n_{k+1}-i},$$

By assumption $\mathcal{E}_k(X_{k+1}-\mathcal{E}_k(X_{k+1}))^i$ is a number equal to $\mathcal{E}_k(Y_{k+1}-\mathcal{E}_k(Y_{k+1}))^i$ (we use here the same abreviation E_k from conditional expectations with respect to Y's and X's) and from LR condition it follows that $\mathcal{E}_k(X_{k+1})$ is a linear combination of $X_1, \ldots, X_k$ with uniquely determined coefficients. Hence $m = \Sigma\, \alpha_{i_1 \ldots i_k} E(X_1^{i_1} \ldots X_k^{i_k})$. Since the same reasoning can be repeated for $m' = E(Y_1^{n_1} \ldots Y_{k+1}^{n_{k+1}})$ and since coefficients $\alpha_{i_1 \ldots i_k}$ are determined uniquely by covariance, by induction assumption we have $m = m'$, which ends the proof.

Proof of Theorem 1: First let us observe, that X_n, $n = 1, 2, \ldots$ have finite moments of all orders. Indeed, by (*) and (6) applied to $a_{1,2,2}$ it follows that $\rho(X_1, X_2) \neq 0$ and for each $n > 1$ there exists $m < n$ such that $\rho(X_n, X_m) \neq 0$ (if there were no such m, then by (6) $a_{n-1,n,n} = 0$ which contradicts (*)). Since X_n, X_m are linearly independent, by Theorem 2 we obtain that X_n $(n = 1, 2, \ldots)$ have finite moments of all orders.

By Lemma 2 it remains therefore to determine all centered conditional moments $\mathcal{E}_k(X_{k+1}-\mathcal{E}_k(X_{k+1}))^i$, $k, i = 0, 1, \ldots$. We proceed by induction. Clearly all centered conditional moments of order 1 are equal 0 and it follows from CV condition that all centered conditional moments of order 2 are as in the gaussian case. Assume, that for some $m \geqslant 2$ centered conditional moments $\mathcal{E}_{n-1}(X_n-\mathcal{E}_{n-1}(X_n))^k$, $\mathcal{E}_{n-1}(X_{n+1}-\mathcal{E}_{n-1}(X_{n+1}))^k$ for $0 \leqslant k \leqslant m$ and $n = 1, 2, \ldots$ are as in the gaussian case (in particular we assume they are nonrandom). Consider expressions

$$W_k = \mathcal{E}_{n-1}[(X_n-\mathcal{E}_{n-1}(X_n))^k (X_{n+1}-\mathcal{E}_{n-1}(X_{n+1})^{m+1-k}].$$

By Lemma 1

$$\begin{aligned} W_1 &= \mathcal{E}_{n-1}[(X_{n+1}-\mathcal{E}_{n-1}(X_{n+1}))^m \mathcal{E}_{n-1,X_{n+1}}(X_n-\mathcal{E}_{n-1}(X_n))] \\ &= a_{n+1,n,n+1}\mathcal{E}_{n-1}[X_{n+1}-\mathcal{E}_{n-1}(X_{n+1})]^{m+1}. \end{aligned} \qquad \ldots (8)$$

Similarly using Lemma 1 we get

$$\begin{aligned} W_1 &= \mathcal{E}_{n-1}[(X_n-\mathcal{E}_{n-1}(X_n))\mathcal{E}_n(X_{n+1}-\mathcal{E}_n(X_{n+1})+\mathcal{E}_n(X_{n+1}-\mathcal{E}_{n-1}(X_{n+1})))^m] \\ &= \mathcal{E}_{n-1}[(X_n-\mathcal{E}_{n-1}(X_n))\mathcal{E}_n(X_{n+1}-\mathcal{E}_n(X_{n+1})+a_{n,n+1,n+1}(X_n-\mathcal{E}_{n-1}(X_n)))^m] \\ &= (a_{n,n+1,n+1})^m \mathcal{E}_{n-1}(X_n-\mathcal{E}_{n-1}(X_n))^{m+1}+R, \end{aligned} \qquad \ldots (9)$$

where $R = \sum_{r=0}^{m-1} \binom{m}{r} (a_{n,n+1,n+1})^r \mathcal{E}_{n-1}(X_n-\mathcal{E}_{n-1}(X_n))^{r+1}\mathcal{E}_n(X_{n+1}-\mathcal{E}_n(X_{n+1}))^{m-r}$ is the number depending only on the centered conditional moments of orders strictly less then $m+1$. Denote

$$\mathcal{E}_{n-1}(X_n-\mathcal{E}_{n-1}(X_n))^{m+1} = x, \quad \mathcal{E}_{n-1}(X_{n+1}-\mathcal{E}_{n-1}(X_{n+1}))^{m+1} = y.$$

Comparing right hand sides of (8) and (9) we obtain the following linear equation with unknown x, y

$$-a^{m}_{n,\,n+1,\,n+1}x+a_{n+1,\,n,\,n+1}y = R. \quad \ldots \quad (10)$$

The similar reasoning applied to W_2 gives

$$-a^{m-1}_{n,\,n+1,\,n+1}\,x+a^{2}_{n+1,\,n,\,n+1}y = R', \quad \ldots \quad (11)$$

where R' is a number depending only on centered conditional moments of order strictly less than $m+1$.

Note that the determinant of the system (10), (11) is non-zero. Indeed,

(i) $a_{n,n+1,n+1} \neq 0$ by condition $(*)$;

(ii) $a_{n+1,n,n+1} \neq 0$ by (6) and $(*)$;

(iii) $a_{n,n+1,n+1}\, a_{n+1,n,n+1} \neq 1$, because by (6) and the formula for determinants quoted by Plucińska (1983) as formula (10)

$$a_{n,n+1,n+1}\, a_{n+1,n,n+1} = 1-\frac{\det[\rho_{\mu,\,\gamma}]_{\mu,\,\gamma\,=\,1,\,\ldots\,,\,n+1}}{\det[\rho_{\mu,\,\gamma}]_{\mu,\,\gamma\,=\,1,\,\ldots\,,\,n}}$$

and (iii) follows by linear independence of $X_1, X_2, \ldots$. Therefore the centered conditional moments x, y are determined uniquely. Let $(Y_i)_{i\in N}$ be mean zero gaussian sequence with the same covariance function as $(X_i)_{i\in N}$. Since the gaussian sequence satisfies LR* and CV conditions and (6) holds, thus the corresponding centered conditional moments of $(Y_i)_{i\in N}$ also satisfy equations (10) and (11). Hence by uniqueness they are equal to x and y respectively. Since n is arbitrary, by induction and Lemma 2 the proof of the Theorem 1 is completed.

3. Concluding remarks

(1) The assumptions of Theorem 1 have particularly simple form in the case of Markov chains. The Markov version of Theorem 1 is as follows :

Let $(X_i)_{i\in N}$ be square-integrable Markov chain such that

(i) $E(X_i \mid X_{i-1}, X_{i+1}) = A_iX_{i-1}+B_iX_{i+1}+C_i$

for some constants A_i, B_i, C_i $(i = 1, 2, \ldots; X_0 = 0)$;

(ii) $\mathrm{var}(X_i \mid X_{i-1}, X_{i+1})$ is nonrandom $(i = 1, 2, \ldots)$;

(iii) $0 < |\rho(X_i, X_{i+1})| < 1$ $(i = 1, 2, \ldots)$.

Then $(X_i)_{i\in N}$ is the gaussian sequence.

Indeed, from (i) follows in the Markov case that each finite subsequence satisfies LR and (ii) implies CV. Then (iii) implies linear independence and condition $(*)$ by the form of the covariance function in the Markov case.

(2) Theorem 1 is not true for finite sequences. The simple counter-example provides vector (X_1, X_2) with the joint distribution given by characteristic function $\varphi(t_1, t_2) = p\cos(t_1+t_2)+(1-p)\cos(t_1-t_2)$, where $0 < p < 1/2$; (X_1, X_2) satisfies LR, CV and $(*)$ conditions and evidently is non-gaussian.

(3) Theorem 2 can be formulated in a slightly more general form (with the same proof): Let $\varepsilon > 0$ and let X_1, X_2 be p_0-integrable r.v., where $p_0 = \max(1; \varepsilon)$, satisfying LR and $(*)$ (for $n = 2$). Assume that X_1, X_2 are linearly independent and $E(|X_i - E(X_i|X_j)|^{\varepsilon}|X_j)$ is a.e. bounded by some constant $(i, j = 1, 2)$. Then X_1 and X_2 are p-integrable for every $p \geqslant 0$.

ACKNOWLEDGEMENT

The authors benefited from the conversation with Professor Stanisław Kwapień, especially on the question of the analog of Theorem 1 for finite sequences. The example quoted in Remark 2 is due to him. The authors thank to the referee for his remarks which improved the presentation of the paper.

REFERENCES

KAGAN, A. M. LINNIK, JU. V. RAO S. R. (1972): *Characterization Problems in Mathematical Statistics*, Moscow (in Russian).

LAHA R. G. (1957): On some characterization problems connected with linear structural relations. *Ann. Math. Stat.*, **28**, 405-414.

PLUCIŃSKA A. (1983): On a stochastic process determined by the conditional expectation and the conditional variance. *Stochastics*, **10**, 115–129.

Paper received : December, 1983.
Revised : August, 1984.