HIROSHIMA MATH. J.
4 (1974), 75–81

An Inequality for Certain Functional of Multidimensional Probability Distributions

Hiroshi MURATA and Hiroshi TANAKA

(Received September 10, 1973)

§ 1. Introduction and the results

Denote by $\mathscr{P}$ the class of all probability distributions f in R^d such that $\int |x|^2 f(dx) < \infty$ and $\int (x_i - \mu_i)^2 f(dx) > 0 \ (1 \leqq i \leqq d)$, where $\mu = (\mu_1, \dots, \mu_d)$ is the mean vector of f. For each $f \in \mathscr{P}$, denote by g_f the Gaussian distribution with the same mean vector and variance matrix as those of f. We introduce a functional $\mathfrak{e}$ on $\mathscr{P}$ by

$$\mathfrak{e}[f] = \inf E\{|X - Y|^2\}, \qquad f \in \mathscr{P},$$

where the infimum is taken over all pairs of R^d-valued random variables X and Y defined on a probability space $(\Omega, \mathscr{F}, P)$ and distributed according to f and g_f respectively. We also write $\mathfrak{e}[X]$ for $\mathfrak{e}[f_X]$, where f_X is the probability distribution of a random variable X.

In the one dimensional case, the functional $\mathfrak{e}$ was introduced and its basic properties were studied in [4] with an application to Kac's one-dimensional model of a Maxwellian gas. The purpose of this paper is to extend some results in [4] to the multi-dimensional case, that is, we will prove the following theorems.

THEOREM 1. *Let X and Y be random variables with probability distributions $f \in \mathscr{P}$ and g_f respectively, and assume that $\mathfrak{e}[f] = E\{|X - Y|^2\}$. Then, X is equal to some Borel function of Y almost surely.*

THEOREM 2. *Let X_1 and X_2 be independent random variables with probability distributions belonging to $\mathscr{P}$. Then,*

$$\mathfrak{e}[X_1 + X_2] < \mathfrak{e}[X_1] + \mathfrak{e}[X_2]$$

unless both X_1 and X_2 are Gaussian. In other words, the functional equation

$$\mathfrak{e}[f_1 * f_2] = \mathfrak{e}[f_1] + \mathfrak{e}[f_2], \qquad f_1, f_2 \in \mathscr{P}$$

gives a characterization of Gaussian distributions.

§ 2. Proof of the theorems

The proof of Theorem 1 will be given in a series of lemmas. In what follows, $<\cdot, \cdot>$ denotes the usual inner product in R^d.

LEMMA 1. *From the same assumption as in Theorem* 1, *it follows that*

$$<X(\omega)-X(\omega'),\ Y(\omega)-Y(\omega')> \geqq 0$$

for almost all (ω, ω') *with respect to* $P \otimes P$.

PROOF. In proving this lemma, we may assume that the basic probability space $(\Omega, \mathscr{F}, P)$ is chosen as follows: Ω is the unit interval $[0, 1)$, $\mathscr{F}$ is the class of Borel sets of Ω and P is the Lebesgue measure in Ω. Suppose the conclusion of the lemma is false. Then, there exists $\varepsilon>0$ such that the set

$$\tilde{A}=\{(\omega, \omega')\in\Omega\times\Omega:\ <X(\omega)-X(\omega'),\ Y(\omega)-Y(\omega')> < -\varepsilon\}$$

has strictly positive $P\otimes P$-measure. Now for integers n, $N\geqq 1$ and for any lattice point $\mathbf{m}=(m_1, \ldots, m_d)\in Z^d$, we set

$$A_{\mathbf{m}}^n=\prod_{i=1}^d [m_i 2^{-n}, (m_i+1)2^{-n})$$

$$X_n(\omega)=\mathbf{m}2^{-n} \quad \text{for} \quad \omega\in X^{-1}(A_{\mathbf{m}}^n)$$

$$Y_n(\omega)=\mathbf{m}2^{-n} \quad \text{for} \quad \omega\in Y^{-1}(A_{\mathbf{m}}^n)$$

$$\tilde{A}_{n,N}=\left\{(\omega, \omega')\in\Omega\times\Omega: \begin{array}{l} <X_n(\omega)-X_n(\omega'),\ Y_n(\omega)-Y_n(\omega')> < -\varepsilon \\ |X_n(\omega)|,\ |X_n(\omega')|,\ |Y_n(\omega)|,\ |Y_n(\omega')|<N \end{array}\right\}.$$

Then, there exists N such that $P\otimes P(\tilde{A}_{n,N})>0$ for all sufficiently large n. Fixing such an N, we choose an n so that $P\otimes P(\tilde{A}_{n,N})>0$ and

$$2^{-n+3}N\sqrt{d}+2^{-2n+2}d<\varepsilon. \tag{2.1}$$

Since

$$\tilde{A}_{n,N}=\bigcup(X^{-1}(A_{\mathbf{m}_1}^n)\cap Y^{-1}(A_{\mathbf{m}_2}^n))\times(X^{-1}(A_{\mathbf{m}'_1}^n)\cap Y^{-1}(A_{\mathbf{m}'_2}^n))$$

where the union is taken over all quartets $(\mathbf{m}_1, \mathbf{m}'_1, \mathbf{m}_2, \mathbf{m}'_2)$ satisfying

$$\begin{cases} <\mathbf{m}_1 2^{-n}-\mathbf{m}'_1 2^{-n},\ \mathbf{m}_2 2^{-n}-\mathbf{m}'_2 2^{-n}> < -\varepsilon \\ |\mathbf{m}_1 2^{-n}|,\ |\mathbf{m}'_1 2^{-n}|,\ |\mathbf{m}_2 2^{-n}|,\ |\mathbf{m}'_2 2^{-n}|<N, \end{cases} \tag{2.2}$$

there exist $\mathbf{m}_1, \mathbf{m}'_1, \mathbf{m}_2, \mathbf{m}'_2\in Z^d$ (satisfying (2.2)) such that

$$P(A)>0,\ A=X^{-1}(\Lambda^n_{\mathbf{m}_1})\cap Y^{-1}(\Lambda^n_{\mathbf{m}_2}),$$

$$P(A')>0,\ A'=X^{-1}(\Lambda^n_{\mathbf{m}'_1})\cap Y^{-1}(\Lambda^n_{\mathbf{m}'_2}).$$

By (2.1) and (2.2), we see that

(2.3) $<x-x',\ y-y'><0$ for any $x\in\Lambda^n_{\mathbf{m}_1},\ x'\in\Lambda^n_{\mathbf{m}'_1},\ y\in\Lambda^n_{\mathbf{m}_2},\ y'\in\Lambda^n_{\mathbf{m}'_2}$.

Next, we take an irrational number λ and denote by T the (ergodic) Weyl automorphism $\omega\in\Omega\to\omega+\lambda$ (mod. 1). Then there exists an integer k such that $P(A\cap T^{-k}A')>0$. We set $U=T^k$, $B=A\cap U^{-1}A'$, $B'=UB$. Since $B\cap B'=\phi$ and U: $B\to B'$ is measure-preserving, we can define a new random variable $X^{\#}$ with probability distribution f by

$$X^{\#}(\omega)=\begin{cases}X(U(\omega)) & \text{for}\quad \omega\in B\\ X(U^{-1}(\omega)) & \text{for}\quad \omega\in B'\\ X(\omega) & \text{for}\quad \omega\notin B\cup B'.\end{cases}$$

From (2.3), we see that for $\omega\in B$

$$|X(U(\omega))-Y(\omega)|^2+|X(\omega)-Y(U(\omega))|^2$$
$$<|X(\omega)-Y(\omega)|^2+|X(U(\omega))-Y(U(\omega))|^2,$$

and this inequality combined with the fact that U is measure-preserving gives us $E\{|X^{\#}-Y|^2\}<E\{|X-Y|^2\}$. This is a contradiction, and the proof is finished.

LEMMA 2. *Let X and Y be R^d-valued random variables defined on a probability space $(\Omega, \mathscr{F}, P)$, and assume that Y has a non-degenerate Gaussian distribution g. If*

$$<X(\omega)-X(\omega'),\ Y(\omega)-Y(\omega')>\geqq 0$$

holds for almost all (ω, ω') with respect to $P\otimes P$, then there exist a regular conditional probability distribution $P_y(\cdot)$ of X given Y and a set $A(\subset R^d)$ of Lebesgue measure 0 such that $P_y\otimes P_{y'}(\Gamma_{y,y'})=1$ holds for all $y,\ y'\notin A$, where

$$\Gamma_{y,y'}=\{(x, x')\in R^{2d}:\ <x-x',\ y-y'>\geqq 0\}.$$

PROOF. Let $\Lambda^n_{\mathbf{m}}$ and Y_n be the same as in the proof of the preceding lemma, and let $P^{(n)}_y(\cdot)$ be a regular conditional probability distribution of X given Y_n; it is given by

$$P^{(n)}_y(\Gamma)=P\{X^{-1}(\Gamma)\cap Y^{-1}(\Lambda^n_{\mathbf{m}})\}/g(\Lambda^n_{\mathbf{m}}),$$

for $\Gamma\in\mathscr{B}(R^d)$, $y\in\Lambda_{\mathbf{m}}^n$. If we set

$$\Psi_n(y)=\int_{R^d}\psi(x)P_y^{(n)}(dx)$$

for a bounded continuous function ψ, then $\{\Psi_n(y), \mathscr{B}_n, g\}$ is a martingale, where $\mathscr{B}_n$ is the σ-field generated by $\{\Lambda_{\mathbf{m}}^n, \mathbf{m}\in Z^d\}$. Therefore, by the convergence theorem of martingales the set

$$B_\psi=\{y\in R^d: \lim_{n\to\infty}\Psi_n(y) \text{ exists}\}$$

has full g-measure. Take a coutable family $\{\psi_k\}_{k\geqq 1}$ which is dense in $C_0(R^d)$, the space of real valued continuous functions on R^d vanishing at infinity, and let B be the intersection of all B_{ψ_k}, $k\geqq 1$. Then $g(B)=1$. Moreover, it is easy to see that for each $y\in B$ the limit $L_y(\psi)$ of $\Psi_n(y)$ as $n\to\infty$ exists for any $\psi\in C_0(R^d)$ and defines a unique measure $P_y(\cdot)$, that is,

$$\lim_{n\to\infty}\int\psi(x)P_y^{(n)}(dx)=\int\psi(x)P_y(dx),\ \psi\in C_0(R^d).$$

Now we define $P_y(\cdot)$ for $y\notin B$ to be an arbitrary probability measure on R^d and put $A=B^c\cup\{y: P_y(R^d)\neq 1\}$. We also redefine $P_y(\cdot)$ for y such that $P_y(R^d)\neq 1$ to be an arbitrary probability measure. Then $g(A)=0$ and $P_y(\cdot)$ is a regular conditional probability distribution of X given Y. To show that A and $\{P_y(\cdot)\}$ have the desired property, we first notice that

$$<X(\omega)-X(\omega'),\ Y_n(\omega)-Y_n(\omega')>\ \geqq -\sqrt{d}2^{-n+1}|X(\omega)-X(\omega')|$$

holds for almost all (ω, ω') with respect to $P\otimes P$ and hence

$$P_y^{(n)}\otimes P_{y'}^{(n)}(\Gamma_{y,y'}^{(n)})=1 \tag{2.4}$$

for almost all (y, y') with respect to $g\otimes g$, where

$$\Gamma_{y,y'}^{(n)}=\{(x, x')\in R^{2d}:\ <x-x',\ y-y'>\ \geqq -\sqrt{d}2^{-n+1}|x-x'|\}.$$

But, since $P_y^{(n)}(\cdot)$ is constant on each $\Lambda_{\mathbf{m}}^n$, the equality (2.4) holds for *all* (y, y'). Because $\Gamma_{y,y'}^{(n)}\downarrow\Gamma_{y,y'}$ as $n\uparrow\infty$, we have $P_y^{(n)}\otimes P_{y'}^{(n)}(\Gamma_{y,y'}^{(n_0)})=1$ for $n\geqq n_0$; letting $n\uparrow\infty$ and using the facts that $P_y^{(n)}$ converges to P_y for $y\notin A$ and that $\Gamma_{y,y'}^{(n_0)}$ is closed, we obtain $P_y\otimes P_{y'}(\Gamma_{y,y'}^{(n_0)})=1$ for $y, y'\notin A$. Since n_0 is arbitrary, the lemma is proved.

By definition a set valued function S: $y\in R^d\to S(y)\subset R^d$ is said to be *monotone*, if there exists a set $A(\subset R^d)$ of Lebesgue measure 0 such that the inequality

$$<x-x',\ y-y'>\ \geqq 0 \quad \text{for} \quad x\in S(y),\ x'\in S(y')$$

holds whenever $y, y'\notin A$.

LEMMA 3. *Let* $S\colon y\in R^d \to S(y)\subset R^d$ *be monotone. Then,* $S(y)$ *consists of a single point for almost all* y.

PROOF. First we consider the case $d=1$, and let $I(y)$ be the smallest closed interval containing $S(y)$. Then, the monotone property of S implies that $I(y)$ and $I(y')$ are non-overlapping if $y\neq y'$, $y, y'\notin A$ (a null set in the definition of monotonicity). Therefore, $I(y)$ consists of a single point for almost all y, and hence so does $S(y)$. Next, we consider the case $d>1$. Given $k(1\leqq k\leqq d)$ and $z=(z_1, \dots, z_{d-1})\in R^{d-1}$, we define a set valued function S_k^z on R^1 by

$$S_k^z(\eta)=\left\{\begin{array}{l}\xi\in R^1\colon (w_1, \dots, w_{k-1}, \xi, w_k, \dots, w_{d-1})\in S(y)\\ \quad\text{for some}\quad w=(w_1, \dots, w_{d-1})\in R^{d-1}\end{array}\right\}$$

where $y=(z_1, \dots, z_{k-1}, \eta, z_k, \dots, z_{d-1})$. We put

$$A_k^z=\{\eta\in R^1\colon (z_1, \dots, z_{k-1}, \eta, z_k, \dots, z_{d-1})\in A\}$$

$$B_k=\{z\in R^{d-1}\colon A_k^z \text{ is a null set}\}.$$

Then, by Fubini's theorem B_k^c and A_k^z for each $z\in B_k$ are null sets, and from the monotone property of S it follows that S_k^z is monotone for each $z\in B_k$. So, the result for the case $d=1$ implies that, for each $z\in B_k$, $S_k^z(\eta)$ is a single point for almost all η. Let D_k be the set of all $y\in R^d$ such that the projection to the k-th coordinate reduces $S(y)$ to a single point, and put $D=\cap D_k$. Then, D^c is a null set, and $S(y)$ is a single point for each $y\in D$, as was to be proved.

The proof of Theorem 1 is now completed as follows. From the first two lemmas, it follows that there exist a null set A and a regular conditional probability distribution $P_y(\cdot)$ of X given Y with the property stated in Lemma 2. If we define $S(y)$, $y\in R^d$, as the smallest closed set of full P_y-measure, then $S(y)\times S(y')\subset \Gamma_{y,y'}$ provided $y, y'\notin A$, or what is the same, the mapping $S\colon y\in R^d\to S(y)$ is monotone. Therefore, by Lemma 3 $S(y)$ is a single point for almost all y; this means that X is equal to some Borel function of Y almost surely.

We give the proof of Theorem 2. We remark that Theorem 1 implies the following: if $f\in\mathscr{P}$ and Y is g_f-distributed, then there exists some Borel function φ from R^d into itself such that $\mathfrak{e}[f]=E\{|\varphi(Y)-Y|^2\}$, since there exists some pair of random variables (with distributions f and g_f) which gives the infimum value $\mathfrak{e}[f]$. Now we take independent Gaussian random variables Y_1 and Y_2 whose mean vectors and variance matrices are the same as those of X_1 and X_2 respectively. Then by the above remark, there exist Borel functions φ_1 and φ_2 such that $\mathfrak{e}[X_1]=E\{|\varphi_1(Y_1)-Y_1|^2\}$ and $\mathfrak{e}[X_2]=E\{|\varphi_2(Y_2)-Y_2|^2\}$. We have

$$\mathfrak{e}[X_1]+\mathfrak{e}[X_2]=E\{|(\varphi_1(Y_1)+\varphi_2(Y_2))-(Y_1+Y_2)|^2\}. \tag{2.5}$$

Since $\varphi_1(Y_1)+\varphi_2(Y_2)$ has the same distribution as that of X_1+X_2 and Y_1+Y_2 has the same mean vector and variance matrix as those of X_1+X_2, the right hand side of (2.5) (and hence $\mathfrak{e}[X_1]+\mathfrak{e}[X_2]$) dominates $\mathfrak{e}[X_1+X_2]$. Next, we suppose that $\mathfrak{e}[X_1]+\mathfrak{e}[X_2]=\mathfrak{e}[X_1+X_2]$. Then, by Theorem 1 there exists a Borel function φ such that

$$\varphi_1(Y_1)+\varphi_2(Y_2)=\varphi(Y_1+Y_2) \quad \text{a.s.}$$

This equation implies that φ_1, φ_2 and φ must be linear and hence X_1 and X_2 must have Gaussian distributions.

§ 3. Applications

1. Let $X_1, X_2, \ldots$ be R^d-valued independent random variables with common distribution $f(\in\mathscr{P})$ of mean vector 0. Then, by the same arguments as in [4], we can prove that $\mathfrak{e}[n^{-1/2}(X_1+\cdots+X_n)]\to 0$ as $n\to\infty$ and hence the probability distribution of $n^{-1/2}(X_1+\cdots+X_n)$ converges to g_f as $n\to\infty$; this is the well-known central limit theorem.

2. Let X_1 and X_2 be real-valued independent random variables, and assume that

$$\tilde{X}_1=X_1\cos\theta+X_2\sin\theta,\ \tilde{X}_2=-X_1\sin\theta+X_2\cos\theta$$

are independent for some θ which is not an integral multiple of $\pi/2$. Then, X_1 and X_2 are Gaussian. This is known as a theorem of M. Kac [3]. There are several proofs (for example, see [1], [2]); here we give a proof based upon Theorem 2 assuming that the probability distributions of X_1 and X_2 are in $\mathscr{P}$.

By Theorem 2, we have

$$\begin{cases} \mathfrak{e}[\tilde{X}_1]\leqq\mathfrak{e}[X_1]\cos^2\theta+\mathfrak{e}[X_2]\sin^2\theta, \\ \mathfrak{e}[\tilde{X}_2]\leqq\mathfrak{e}[X_1]\sin^2\theta+\mathfrak{e}[X_2]\cos^2\theta. \end{cases} \tag{3.1}$$

Let

$$X=\begin{pmatrix} X_1 \\ X_2 \end{pmatrix},\quad A=\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

Then we can prove that $\mathfrak{e}[AX]=\mathfrak{e}[X]$, $\mathfrak{e}[AX]=\mathfrak{e}[\tilde{X}_1]+\mathfrak{e}[\tilde{X}_2]$ and $\mathfrak{e}[X]=\mathfrak{e}[X_1]+\mathfrak{e}[X_2]$; here we have used the orthogonality of the matrix A for the first equality and the independence of the components for the last two equalities. Therefore (3.1) holds with "=", and hence X_1 and X_2 are Gaussian by Theorem 2.

References

[1] W. Feller, An Introduction to Probability Theory and Its Applications, Vol. II, 2nd ed., John Wiley & Sons, Inc., New York, 1971.

[2] Y. Itoh, The information theoretic proof of Kac's theorem, Proc. Japan Acad., **46** (1970), 283–286.

[3] M. Kac, On a characterization of the normal distribution, Amer. J. Math. **61** (1939), 726–728.

[4] H. Tanaka, An inequality for a functional of probability distributions and its application to Kac's one-dimensional model of a Maxwellian gas, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete. **27** (1973), 47–52.

Department of Mathematics,
Faculty of Science,
Hiroshima University