Z. Wahrscheinlichkeitstheorie verw. Geb. 27, 47–52 (1973)

An Inequality for a Functional of Probability Distributions and Its Application to Kac's One-Dimensional Model of a Maxwellian Gas

Hiroshi Tanaka

1. Introduction

Let $\mathscr{P}$ be the class of 1-dimensional probability distributions f with

$$0<\alpha_2(f)<\infty,$$

where $\alpha_2(f)$ denotes the second moment of f. Taking a probability space (Ω, P) which is big enough to carry Gaussian random variables, we introduce a functional $\mathfrak{e}$ defined for $f\in\mathscr{P}$ by

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

where the infimum is taken over all pairs of random variables X and Y defined on (Ω, P) and distributed according to f and g respectively; here g is the Gaussian distribution with mean 0 and variance $\sigma^2=\alpha_2(f)$. $\mathfrak{e}[f]$ is sometimes denoted by $\mathfrak{e}[X]$ when X is a random variable with distribution f. It should be noticed that the value of $\mathfrak{e}[f]$ does not depend upon a choice of the probability space (Ω, P). The purpose of this paper is to present some basic properties of $\mathfrak{e}$ (especially, the inequality (2.2)) together with an application to the central limit theorem and then to show that the functional $\mathfrak{e}$ is monotone decreasing along Boltzmann solutions of Kac's one-dimensional model of a Maxwellian gas. Some of our results can be generalized to the case of R^3; for example, the functional $\mathfrak{e}$ similarly defined in R^3 decreases along solutions of Boltzmann's problem for the 3-dimensional Maxwellian gas, but this will be discussed in another occasion.

2. Basic Properties of $\mathfrak{e}$ and a Proof of the Central Limit Theorem

Theorem 1. *Let $f\in\mathscr{P}$, and denote by g the Gaussian distribution with mean* 0 *and variance $\sigma^2=\alpha_2(f)$. Let X and Y be random variables with distributions f and g, respectively. Then, $\mathfrak{e}[f]=E\{|X-Y|^2\}$ if and only if $X=F^{-1}(G(Y))$ almost surely, where F^{-1} is the right continuous inverse function of the distribution function F corresponding to f, and G is the distribution function corresponding to g.*

In the proof of this theorem the probability space is chosen as follows: Ω is the unit interval $[0, 1)$ and P is the Lebesgue measure in Ω. The proof is carried out in 3 steps.

Step 1. $(X(\omega)-X(\omega'))\,(Y(\omega)-Y(\omega'))\geqq 0$ for almost all (ω, ω') $(P\otimes P)$. Suppose the contrary holds. Then for some $\varepsilon>0$ at least one of the following events has positive $P\otimes P$-probability:

$$\{(\omega,\omega')\in\Omega\times\Omega:\ X(\omega)-X(\omega')<-3\varepsilon \text{ and } Y(\omega)-Y(\omega')>3\varepsilon\} \tag{1}$$

$$\{(\omega,\omega')\in\Omega\times\Omega:\ X(\omega)-X(\omega')>3\varepsilon \text{ and } Y(\omega)-Y(\omega')<-3\varepsilon\}. \tag{2}$$

We assume that the event (1) has positive probability for simplicity. Then, for some integers j_1, j_2, k_1, k_2 with $j_1+1<j_2$ and $k_1+1<k_2$, the event

$$\tilde{A}=\left\{(\omega,\omega'):\ \begin{matrix} j_1\varepsilon<X(\omega)\leqq(j_1+1)\varepsilon,\ j_2\varepsilon<X(\omega')\leqq(j_2+1)\varepsilon \\ k_1\varepsilon<Y(\omega')\leqq(k_1+1)\varepsilon,\ k_2\varepsilon<Y(\omega)\leqq(k_2+1)\varepsilon \end{matrix}\right\}$$

has positive $P\otimes P$-probability. If we set

$$A=\{\omega:\ j_1\varepsilon<X(\omega)\leqq(j_1+1)\varepsilon,\ k_2\varepsilon<Y(\omega)\leqq(k_2+1)\varepsilon\}$$

$$A'=\{\omega:\ j_2\varepsilon<X(\omega)\leqq(j_2+1)\varepsilon,\ k_1\varepsilon<Y(\omega)\leqq(k_1+1)\varepsilon\},$$

then $\tilde{A}=A\times A'$ and hence $P(A)>0$, $P(A')>0$. Next, we take an irrational number λ and denote by T the Weyl automorphism: $\omega\in\Omega\to\omega+\lambda(\text{mod } 1)$. Then there exists an integer $n\geqq 0$ such that $P(A\cap T^{-n}A')>0$. If we set $B=A\cap T^{-n}A'$, $B'=T^nB$, $\varphi=T^n$, then $P(B)=P(B')>0$ and $B\cap B'=\emptyset$. We now define

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

Since φ: $B\to B'$ is measure-preserving[1], $X^{\#}$ has still distribution f, and we have

$$\begin{aligned} E\{|X^{\#}-Y|^2\} &= \int_B |X(\varphi(\omega))-Y(\omega)|^2 P(d\omega)+\int_{B'} |X(\varphi^{-1}(\omega))-Y(\omega)|^2 P(d\omega) \\ &\quad + \int_{(B\cup B')^c} |X(\omega)-Y(\omega)|^2 P(d\omega) \\ &= \int_B \{|X(\varphi(\omega))-Y(\omega)|^2+|X(\omega)-Y(\varphi(\omega))|^2\} P(d\omega) \\ &\quad + \int_{(B\cup B')^c} |X(\omega)-Y(\omega)|^2 P(d\omega) \\ &< \int_B \{|X(\omega)-Y(\omega)|^2+|X(\varphi(\omega))-Y(\varphi(\omega))|^2\} P(d\omega) \\ &\quad + \int_{(B\cup B')^c} |X(\omega)-Y(\omega)|^2 P(d\omega)=E\{|X(\omega)-Y(\omega)|^2\}; \end{aligned}$$

the inequality part in the above employs the following elementary fact: if $a_1<a_2$ and $b_1<b_2$, then $(a_1-b_1)^2+(a_2-b_2)^2<(a_1-b_2)^2+(a_2-b_1)^2$. We thus arrive at a contradiction.

Step 2. Let $P(y,\cdot)$ be a regular conditional probability distribution of X given $Y=y$, and denote by S_y the smallest closed interval such that $P(y,S_y)=1$. We claim that

S_y and $S_{y'}$ are non-overlapping for almost all (y,y') with respect to $g\otimes g$. (2.1)

[1] I owe the use of the Weyl automorphism for constructing φ to Y. Takahashi.

Since $P(y,\cdot)\otimes P(y',\cdot)$ is a regular conditional probability distribution of $(X(\omega), X(\omega'))$ given $(Y(\omega), Y(\omega'))=(y,y')$, we have from Step 1

$$\iint g(dy)\,g(dy')\iint \chi(x,x',y,y')\,P(y,dx)\,P(y',dx') = E\{\chi(X(\omega),X(\omega'),Y(\omega),Y(\omega'))\}=1$$

where χ is the indicator function of the set

$$\Gamma=\{(x,x',y,y')\in R^4:\ (x-x')(y-y')\geqq 0\}.$$

Therefore, for almost all (y,y') with respect to $g\otimes g$, we have

$$\iint \chi(x,x',y,y')\,P(y,dx)\,P(y',dx')=1.$$

So, if we set $\Gamma_1=\{(x,x')\in R^2:\ x\geqq x'\}$, then $P(y,\cdot)\otimes P(y',\cdot)$ is supported by Γ_1 for almost all $(y,y')\in\Gamma_1$; but this is a complicated way of saying that (2.1) holds.

Step 3. From Step 2 one can prove easily that S_y is a single point for almost all y with respect to g. Now, this fact combined with the inequality of Step 1 implies that X is an increasing function of Y (a.s.); this is possible only when $X=F^{-1}(G(Y))$ almost surely. The "if" part is obvious, since the infimum in the definition of $\mathfrak{e}[f]$ is actually attained by some pair.

Theorem 2. *Let X and Y be independent random variables with distributions f_1 and $f_2\in\mathscr{P}$, respectively, and assume that $E\{X\}=E\{Y\}=0$. Then, for any real constants a, b such that $a\neq 0$, $b\neq 0$,*

$$\mathfrak{e}[aX+bY]<a^2\,\mathfrak{e}[X]+b^2\,\mathfrak{e}[Y], \tag{2.2}$$

unless both X and Y are Gaussian.

The proof of this theorem is based upon Theorem 1. It is obvious that

$$\mathfrak{e}[aX+bY]\leqq a^2\,\mathfrak{e}[X]+b^2\,\mathfrak{e}[Y]$$

holds, and so assuming the equality holds in the above, we will prove $f_i=g_i$, where g_i is the Gaussian distribution with mean 0 and variance $\sigma_i^2=\alpha_2(f_i)$, $i=1,2$. If X_1 and X_2 are independent random variables with distributions g_1 and g_2, respectively, then with the obvious notation it follows from Theorem 1 that

$$a^2\,\mathfrak{e}[X]+b^2\,\mathfrak{e}[Y]=a^2\,E\{|F_1^{-1}(G_1(X_1))-X_1|^2\}+b^2\,E\{|F_2^{-1}(G_2(X_2))-X_2|^2\}$$
$$=E\{|aF_1^{-1}(G_1(X_1))+bF_2^{-1}(G_2(X_2))-(aX_1+bX_2)|^2\}.$$

Since aX_1+bX_2 is also G-distributed, we have again from Theorem 1

$$aF_1^{-1}(G_1(X_1))+bF_2^{-1}(G_2(X_2))=F^{-1}(G(aX_1+bX_2))\ \text{a.s.}, \tag{2.3}$$

where F is the distribution function of $aX+bY$. By the right continuity of the functions involved, (2.3) yields

$$aF_1^{-1}(G_1(x))+bF_2^{-1}(G_2(y))=F^{-1}(G(ax+by))$$

for all $x,y\in R^1$. This functional equation for unknown F_1, F_2, F can easily be solved; the result is $F_1=G_1$, $F_2=G_2$, completing the proof.

We next list some simple properties of $\mathfrak{e}$ for later use.

1. If f_n converges to some $f \in \mathscr{P}$ as $n \uparrow \infty$ in such a way that

$$\lim_{N\to\infty} \sup_{n\geq 1} \int_{|x|>N} x^2 f_n(dx) = 0, \tag{2.4}$$

then $\lim_{n\to\infty} \mathfrak{e}[f_n] = \mathfrak{e}[f]$. The condition (2.4) is satisfied if for some $p>2$ the absolute p-th moments of f_n are bounded in n.

2. Let $f_\theta \in \mathscr{P}$ and $\alpha_2(f_\theta) = \sigma^2$ for $0 \leqq \theta < 1$, and assume that $\int \varphi(x) f_\theta(dx)$ is Borel measureable in θ for any bounded continuous function φ. Then for any probability measure μ on $[0, 1)$ we have

$$\mathfrak{e}[\int f_\theta \, \mu(d\theta)] \leqq \int \mathfrak{e}[f_\theta] \, \mu(d\theta).$$

3. By Theorem 1, $\mathfrak{e}$ admits the expression

$$\mathfrak{e}[f] = 2 \int \{x^2 - x\, G^{-1}(F(x))\} f(dx)$$

for a continuous probability distribution f in $\mathscr{P}$.

The inequality (2.2) will now be applied to give a simple proof of the central limit theorem. Let $\{X_n\}_{n=1,2,\ldots}$ be a sequence of independent identically distributed random variables with mean 0 and variance 1. Then the so-called central limit theorem states that the distribution of $\xi_n = n^{-\frac{1}{2}}(X_1 + \cdots + X_n)$ tends to a Gaussian distribution as $n \uparrow \infty$. *Here we prove that* $\mathfrak{e}[\xi_n] \to 0$ *as* $n \to \infty$ *assuming* $E\{X_1^4\} < \infty$[2]. This condition implies that

$$E\{\xi_n^4\} = \frac{1}{n} E\{X_1^4\} + 3\left(1 - \frac{1}{n}\right) < \text{const.} \quad \text{(independent of } n\text{)}. \tag{2.5}$$

Putting $\eta_k = \xi_{2^k}$, we first prove that $\mathfrak{e}[\eta_k] \downarrow 0$ as $k \uparrow \infty$. The decreasing property of $\mathfrak{e}[\eta_k]$ is obvious by the inequality (2.2), and so we denote by l the limit of $\mathfrak{e}[\eta_k]$ as $k \to \infty$. If f is a limit distribution of η_k as $k \to \infty$ via some subsequence $k_1 < k_2 < \cdots$, and if η and ζ are independent random variables with distribution f, then it follows from (2.5) and 1 of § 2 that

$$\mathfrak{e}[\eta] = \lim_{p\to\infty} \mathfrak{e}[\eta_{k_p}] = l \quad \text{and} \quad \mathfrak{e}\left[\frac{\eta+\zeta}{2}\right] = \lim_{p\to\infty} \mathfrak{e}[\eta_{2^{k_p}}] = l,$$

therefore by Theorem 2 the limit l must be 0. Next, we write an integer $n \geqq 1$ as $n = \sum_{k=0}^{m} n_k$ where $n_k = \varepsilon_k 2^k$ with $\varepsilon_k = 0$ or 1. Then, using the inequality (2.2) we have $\mathfrak{e}[\xi_n] \leqq \frac{1}{n} \sum_{k=0}^{m} n_k \mathfrak{e}[\eta_k]$, and hence $\mathfrak{e}[\xi_n] \to 0$ as was to be proved.

3. $\mathfrak{e}$ Decreases along Solutions of Boltzmann's Problem for Kac's Model of a Maxwellian Gas

Given $f_1, f_2 \in \mathscr{P}$ and $\theta \in [0, 2\pi)$, we denote by $B_\theta(f_1, f_2)$ the probability distribution of $X_1 \cos\theta + X_2 \sin\theta$, where X_1 and X_2 are random variables with distri-

[2] This condition is assumed just to simplify the proof. Without this $\mathfrak{e}[\xi_n]$ still tends to 0.

butions f_1 and f_2 respectively. We also put

$$B(f_1, f_2) = \frac{1}{2\pi} \int_0^{2\pi} B_\theta(f_1, f_2)\, d\theta.$$

In Kac's one-dimensional model of a Maxwellian gas, the distribution $u(t, dx)$ of molecular speeds at time $t > 0$ is determined by the solution of Boltzmann's problem

$$\frac{\partial u(t, \cdot)}{\partial t} = B(u(t), u(t)) - u(t, \cdot). \tag{3.1}$$

The solutions of this equation can be obtained by Wild's sum [2]. If the initial distribution has a density, then so does the solution, and it is known that the entropy increases along the solution with time, while the solution itself tends to a Gaussian distribution as $t \uparrow \infty$. McKean [1] gave detailed discussions on this subject; he gave also other functionals which are (or, at least are expected to be) monotone along the solutions of (3.1) together with an interesting conjecture about them. But, among these functionals, the entropy and Linnik's functional are the only ones which were used effectively in the investigation of the asymptotic properties of the solutions of (3.1). In this section, we prove that the functional $\mathfrak{e}$ decreases monotonically to zero along the solutions of (3.1); this statement itself implies automatically that the solutions of (3.1) tend to Gaussian distributions as $t \uparrow \infty$.

Theorem 3. *Let $u(t)$ be the solution of* (3.1) *with initial distribution $f \in \mathscr{P}$. Then,* (i) $\mathfrak{e}[u(t)]$ *is decreasing in t, and* (ii) *if f has finite fourth moment,* $\mathfrak{e}[u(t)]$ *decreases* to 0 *as $t \uparrow \infty$.*

The following corollary is an immediate consequence of the above theorem and 3 of § 2.

Corollary. *Let $\mathscr{P}_0$ be the subclass of $\mathscr{P}$ consisting of continuous probability distributions, and put $\mathfrak{e}_0[f] = \int x G^{-1}[F(x)] f(dx)$. Then the functional $\mathfrak{e}_0$ is increasing along the solutions of* (3.1) *with initial distributions $\in \mathscr{P}_0$.*

The proof of Theorem 3 will be given in several steps.

Proof. 1. Let $\mathscr{P}_n(f)$, $n \geqq 1$, be the (finite) set of probability measures from $\mathscr{P}$ defined inductively as follows: (i) $\mathscr{P}_1(f)$ consists of a single element f, and (ii) $\mathscr{P}_n(f)$ is the set of all probability measures of the form $B(f_1, f_2)$ with $f_1 \in \mathscr{P}_{n_1}(f)$, $f_2 \in \mathscr{P}_{n_2}(f)$, $n_1 + n_2 = n$. Then, the solution $u(t)$ of (3.1) with initial distribution f can be expressed as Wild's sum

$$u(t) = e^{-t} \sum_{n=1}^{\infty} (1 - e^{-t})^{n-1} p_n(f), \tag{3.2}$$

where $p_n(f)$ stands for a convex combination of elements in $\mathscr{P}_n(f)$, $n \geqq 1$ ([2], see also [1]).

2. If $\tilde{f}$ denotes the even part of f, say $\tilde{f}(dx) = \frac{1}{2}(f(dx) + f(-dx))$, then it is easy to see that $B(f_1, f_2) = B(\tilde{f}_1, \tilde{f}_2)$. Therefore, if f_1 and f_2 have the same second

moment, it follows from 2 of § 2 and Theorem 2 that

$$\begin{aligned} e[B(f_1,f_2)] = e[B(\tilde{f}_1,\tilde{f}_2)] &\leqq \frac{1}{2\pi}\int_0^{2\pi} e[B_\theta(\tilde{f}_1,\tilde{f}_2)]\,d\theta \\ &\leqq \frac{1}{2\pi}\int_0^{2\pi} \{e[\tilde{f}_1]\cos^2\theta + e[\tilde{f}_2]\sin^2\theta\}\,d\theta \leqq \frac{e[f_1]+e[f_2]}{2}, \end{aligned}$$

because $e[\tilde{f}] \leqq e[f]$. Therefore we have $e[p_n(f)] \leqq e[f]$, and hence by Wild's sum (3.2) and 2 of § 2 we see that

$$e[u(t)] \leqq e[f] \qquad (t>0); \tag{3.3}$$

the equality holds if and only if f is a Gaussian distribution. (3.3) implies the part (i) of the theorem.

3. If $\int x^4 f(dx) < \infty$, then by (3.1) the function $\alpha(t) = \int x^4 u(t, dx)$ satisfies the differential equation

$$\frac{d\alpha(t)}{dt} = \frac{3}{4}\sigma^4 - \frac{1}{4}\alpha(t), \qquad \sigma^2 = \alpha_2(f),$$

which implies that $\alpha(t) \to 3\sigma^4$ as $t \to \infty$, and hence $\alpha(t)$ is bounded. Next, let u_∞ be a limit distribution of $u(t)$ as $t \uparrow \infty$ via some subsequence $t_1 < t_2 < \cdots$. Since $\alpha(t)$ is bounded, we have $e[u_\infty] = \lim_{n\to\infty} e[u(t_n)] = \lim_{t\to\infty} e[u(t)]$ by 1 of § 2. If $u_\infty(t)$ denotes the solution of (3.1) with initial distribution u_∞, then an application of Wild's sum shows that $u_\infty(t) = \lim_{n\to\infty} u(t_n + t)$ and hence $e[u_\infty(t)] = \lim_{t\to\infty} e[u(t_n+t)] = e[u_\infty]$. Therefore u_∞ must be a Gaussian distribution from the preceding step, as was to be proved.

Note. After sending the manuscript to the editor, I was informed from T. Yanagimoto of a simple proof of Theorem 1 based upon the following Hoeffding's formula: if F denotes the joint and F_X and F_Y the marginal distribution functions of X and Y, then

$$E(XY) - E(X)E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [F(x,y) - F_X(x)F_Y(y)]\,dx\,dy$$

provided the expectations on the left hand side exist.

References

1. McKean, H. P.: Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas. Arch. Rat. Mech. Analysis **21**, 343–367 (1966)
2. Wild, E.: On Boltzmann's equation in the kinetic theory of gases. Proc. Cambridge Philos. Soc. **47**, 602–609 (1951)

Hiroshi Tanaka
Department of Mathematics
Hiroshima University
Hiroshima, Japan

(Received January 10, 1973)