66. *The Information Theoretic Proof of Kac's Theorem*

By Yoshiaki ITOH
Tokyo Institute of Technology

(Comm. by Kinjirô KUNUGI, M. J. A., March 12, 1970)

1. Introduction.

The object of this paper is to give information theoretic proofs of the following two well known theorems.

Theorem 1. *Let* X_1 *and* X_2 *be independent random variables and let*

$$(1) \qquad Y_1 = X_1 \cos\gamma + X_2 \sin\gamma$$

$$(2) \qquad Y_2 = -X_1 \sin\gamma + X_2 \cos\gamma.$$

If Y_1 *and* Y_2 *are independent of each other for sufficiently small neighbor of some* γ. *Then the variables* X_1 *and* X_2 *are normally distributed.*

Theorem 2. *Let* $F(x)$ *be a distribution function with mean zero and variance one. If for any positive* σ_1 *and* σ_2 *there exists* $\sigma > 0$ *satisfying the following relation*

$$(3) \qquad F\left(\frac{x}{\sigma_1}\right) * F\left(\frac{x}{\sigma_2}\right) = F\left(\frac{x}{\sigma}\right).$$

Then $F(x)$ *is normal distribution, where the notation* $*$ *denotes the convolution of distribution functions.*

Theorem 1 was proved by M. Kac 1 in a general form. And Theorem 2 was first proved by G. Pólya. We must assume appropriate conditions, as our proof is based on the information measures of C. E. Shannon, R. A. Fisher and YU. V. Linnik 1.

2. Notations and Lemmas.

We consider one dimensional random variable X with continuous probability density $p(x)$ and satisfying the conditions

$$(4) \qquad \sup p(x) < \infty, \; E(X) = \int_{-\infty}^{\infty} x p(x) dx = 0, \quad \text{and} \quad D(X) = \int_{-\infty}^{\infty} x^2 p(x) dx.$$

And we put

$$(5) \qquad I(X) = H(X) - \frac{1}{2} \log D(X)$$

following YU. V. Linnik 1 where

$$(6) \qquad H(X) = -\int_{-\infty}^{\infty} p(x) \log p(x) dx$$

introduced by C. Shannon.

Lemma 1. *$I(X)$ is invariant with respect to a homothetic transformation, i.e., for any $\alpha > 0$*

(7) $$I(\alpha X) = I(X).$$

The proof is in the paper of Linnik 1.

Lemma 2. *Let X and Y be mutually independent random variables with probability densities $p(x)$ and $q(y)$ and variances $D(X)$ and $D(Y)$ respectively. Then*

(8) $$I(X+\beta Y) - I(X) = \frac{1}{2}\beta^2 D(Y) f(X) + o(\beta^2)$$

for sufficiently small $\beta > 0$, where

(9) $$f(X) = \int_{-\infty}^{\infty} \left(\frac{p'(x)}{p(x)}\right)^2 p(x)dx - \frac{1}{D(X)}$$

Proof. Let $\pi(x)$ be the probability density of $X+\beta Y$, then

$$\pi(x) = \int_{-\infty}^{\infty} p(x-\beta y)q(y)dy = \int_{-\infty}^{\infty} \left\{p(x) + \frac{1}{2}\beta^2 y^2 p''(x)\right\} q(y)dy + o(\beta^2)$$
$$= p(x) + \frac{1}{2}\beta^2 D(Y) p''(x) + o(\beta^2)$$

Hence

$$I(X+\beta Y) = -\int_{-\infty}^{\infty} \pi(x)\log \pi(x)dx - \frac{1}{2}\log(D(X) + \beta^2 D(Y))$$

and

$$I(X) = -\int_{-\infty}^{\infty} p(x)\log p(x)dx - \frac{1}{2}\log D(X)$$

And so

$$I(X+\beta Y) - I(X) = \frac{1}{2}\beta^2 D(Y) f(X) + o(\beta^2).$$

The detailed estimation of $o(\beta^2)$ can be done analogously to the paper of Linnik 1.

Lemma 3. *Let $f(X)$ be the value defined by (9), then we have*

$$f(X) \geqq 0$$

and

$$f(X) = 0$$

if and only if X is normally distributed.

It is easy to see that the inequality $f(X) \geqq 0$ is equivalent to Rao-Cramér inequality. And so the proof is known. For example see Linnik 1 or H. P. Mckean 1.

3. Proofs of Theorem.

Proof of Theorem 1.

For sufficiently small ε, we consider

(10) $$Z_1 = X_1 \cos(\gamma - \varepsilon) + X_2 \sin(\gamma - \varepsilon)$$

(11) $$Z_2 = -X_1 \sin(\gamma - \varepsilon) + X_2 \cos(\gamma - \varepsilon).$$

Then from the assumption, Z_1 and Z_2 are mutually independent.

By (1), (2), (10) and (11)

$$Y_1 = Z_1 \cos \varepsilon + Z_2 \sin \varepsilon \tag{12}$$

$$Y_2 = -Z_1 \sin \varepsilon + Z_2 \cos \varepsilon. \tag{13}$$

By (12) we have

$$\frac{Y_1}{\cos \varepsilon} = Z_1 + \tan \varepsilon Z_2. \tag{14}$$

Hence from Lemma 1, Lemma 2 and Lemma 3, we see

$$\lim_{\varepsilon \to 0} \frac{I(Y_1) - I(Z_1)}{\tan^2 \varepsilon} = \lim_{\varepsilon \to 0} \frac{I\left(\frac{Y_1}{\cos \varepsilon}\right) - I(Z_1)}{\tan^2 \varepsilon} = \frac{1}{2} D(Z_2) f(Z_1) \geqq 0. \tag{15}$$

By the same reason

$$\lim_{\varepsilon \to 0} \frac{I(Y_2) - I(Z_2)}{\tan^2 \varepsilon} = \frac{1}{2} D(Z_1) f(Z_2) \geqq 0. \tag{16}$$

As (12) and (13) are respectively equivalent to the following

$$Z_1 = Y_1 \cos \varepsilon - Y_2 \sin \varepsilon \tag{17}$$

$$Z_2 = Y_1 \sin \varepsilon + Y_2 \cos \varepsilon\ , \tag{18}$$

we can apply the same argument to (17) and (18). Then we can obtain following relations

$$\lim_{\varepsilon \to 0} \frac{I(Z_1) - I(Y_1)}{\tan^2 \varepsilon} = \frac{1}{2} D(Y_2) f(Y_1) \geqq 0 \tag{19}$$

$$\lim_{\varepsilon \to 0} \frac{I(Z_2) - I(Y_2)}{\tan^2 \varepsilon} = \frac{1}{2} D(Y_1) f(Y_2) \geqq 0. \tag{20}$$

From (15), (16), (19) and (20)

$$f(Y_1) = f(Y_2) = f(Z_1) = f(Z_2) = 0.$$

By Lemma 3, Z_1 and Z_2 are normally distributed.

So X_1 and X_2 are normally distributed.

Proof of Theorem 2. Let X, Y and Z be the independent random variables related to the distribution $F(x)$. By the assumption we have

$$\sigma_1 X + \sigma_2 Y = \sigma Z. \tag{21}$$

Now we may take $\sigma_1 = 1$ and sufficiently small $\sigma_2 > 0$. From Lemma 2

$$I(X + \sigma_2 Y) - I(X) = \frac{1}{2} \sigma_2^2 f(X) + o\sigma_2^2).$$

On the other hand by (21) and Lemma 1

$$I(X + \sigma_2 Y) = I(\sigma Z) = I(Z).$$

As Z and σX have the same distribution $F(x)$, we have

$$I(Z) = I(X).$$

So it is necessary that $f(X)$ is zero. Hence $F(x)$ is a normal distribution function.

Acknowledgement. The author is much indebted to Prof. K. Kunisawa for his valuable advice.

References

[1] Kac, M. 1.: On a characterization of normal distribution. Amer. J. Math., **61** (1939).

[2] Linnik, YU. V. 1.: An information theoretic proof of the central limit theorem with the Lindeberg condition. Theory of Probability and its Applications, **4**, 288–299 (1959).

[3] Mckean, H. P. 1.: Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas. Arch. Rational Mech. Anal., **21**, 343–367 (1966).

[4] Pólya, G. 1: Herleitung des Gauss'schen Gesetzes aus einer Funktionalgleichung. Math. Zeit., **18**, 96–103 (1923).

[5] Shannon, C. 1.: A mathematical theory of communication. Bell Syst. Tech. J., **27**, 379–423, 623–656 (1948).