ANNALES
POLONICI MATHEMATICI
LIII.2 (1991)

Relative entropy and stability of stochastic semigroups

by KRZYSZTOF ŁOSKOT and RYSZARD RUDNICKI (Katowice)

Abstract. For any convex function η and densities f and g the η-entropy of f relative to g is defined by $H_\eta(f|g) = \int g\eta(f/g)d\mu$. It is proved that $H_\eta(f_n|g_n) \to \eta(1)$ implies $f_n - g_n \to 0$ in L^1. This theorem is applied to stochastic semigroups generated by partial differential equations. A sufficient condition for convergence of densities to the Maxwellian is also given.

Introduction. The main object in the statistical description of a dynamical system is a probabilistic measure defined on the phase space. Such a measure may be interpreted as a state of the system. In the present paper we consider the case when the phase space is a σ-finite measure space $(X, \mathcal{F}, \mu)$ and states are absolutely continuous with respect to μ. Each state λ can be identified with its density, i.e., the Radon–Nikodym derivative $f = d\lambda/d\mu$. In statistical physics the functional

$$H_0(f|g) = \int (f \ln f - f \ln g) d\mu \tag{0.1}$$

measures how easy it is to distinguish two states given by the densities f and g. This functional is called the *Kullback–Leibler entropy* (or *information*) of f relative to g [8].

In this paper we use the notion of the relative entropy (also called the *statistical distance* [10]) introduced by Csiszár [4], [5]. For every continuous convex function η: $[0, \infty) \to \boldsymbol{R}$ we define the η-entropy of f relative to g by

$$H_\eta(f|g) = \int g\eta(f/g) d\mu. \tag{0.2}$$

The Kullback–Leibler entropy is a special case of the η-entropy for $\eta(u) = u \ln u$. The main goal of this paper is to investigate connections between the η-entropy and stochastic operators.

The organization of the paper is as follows. The precise definition of the η-entropy is given in Section 1. Section 2 contains an elementary proof of the Csiszár inequality

$$H_\eta(Uf \mid Ug) \leqslant H_\eta(f \mid g) \tag{0.3}$$

for every stochastic operator U. Voigt [13] rediscovered this inequality in the special case of the Kullback–Leibler entropy. In his proof he used methods of Gelfand–Naimark representation theory. In the special case when U is a double-stochastic operator and $g = 1$ inequality (0.3) appears in the paper of Misra and Prigogine [11]. In Section 3 the convergence of $H_\eta(f_n | g_n)$ to $\eta(1)$ is proved to imply strong L^1-convergence of the difference $f_n - g_n$ to 0. Using the functional (0.2) to examine convergence of Markov chains to equilibrium was proposed by Rényi [12]. Developing the idea of Rényi, Fritz [6] applied the Kullback–Leibler entropy to limit theorems for reversible Markov processes. The idea of using conditional entropy to examine ergodic properties of dynamical systems was exploited by Lasota and Mackey [9]. Section 3 also contains a simple proof of Elmroth's theorem [3] on convergence to the Maxwellian. In Section 4 we show an application of our theorem to stochastic semigroups generated by the heat equation and the Dirichlet problem on the half space. Our proofs are based on simple geometrical properties of convex functions [1].

1. The η-entropy. Let $(X, \mathcal{F}, \mu)$ be a measure space and let $L^1(\mu)$ denote the space of all real μ-integrable functions on X. By $L^1_+(\mu)$ we denote the positive cone in $L^1(\mu)$, i.e., $L^1_+(\mu) = \{f \in L^1(\mu)\colon\ f \geqslant 0 \text{ a.e.}\}$, and let $D(\mu) = \{f \in L^1_+(\mu)\colon \int f\, d\mu = 1\}$ be the set of all densities. Let $\eta\colon [0, \infty) \to \mathbf{R}$ be a continuous convex function. In order to give the precise definition of the η-entropy we introduce an auxiliary function $\varphi\colon [0, \infty) \times [0, \infty) \to \mathbf{R} \cup \{\infty\}$ by setting

$$\varphi(u, v) = \begin{cases} v\eta(u/v), & v > 0,\ u \geqslant 0, \\ 0, & v = 0,\ u = 0, \\ u\eta'(\infty), & v = 0,\ u > 0, \end{cases} \tag{1.1}$$

where

$$\eta'(\infty) = \lim_{v \to \infty} \eta'(v) = \lim_{v \to \infty} \eta(v)/v. \tag{1.2}$$

Further, we shall need some properties of φ. It is well known that for a given convex function η there exist sequences (a_n) and (b_n) of real numbers such that

$$\eta(u) = \sup\{a_n + b_n u\colon\ n \in N\}. \tag{1.3}$$

From (1.3) it follows immediately that

$$\varphi(u, v) = \sup\{a_n v + b_n u\colon\ n \in N\}. \tag{1.4}$$

In particular, the last formula implies that φ is a positively homogeneous convex function.

Now for $f,\ g \in L^1_+(\mu)$ we define the *η-entropy* of f relative to g by

$$H_\eta(f | g) = \int \varphi(f, g)\, d\mu. \tag{1.5}$$

The definition is correct: Indeed, from (1.4) it follows that $\varphi(f, g) \geqslant a_1 g + b_1 f$, which implies that the integral (1.5) exists and $-\infty < H_\eta(f|g) \leqslant \infty$. From the convexity of φ it follows that $H_\eta(f|g)$ is convex with respect to (f, g). Moreover, it is a positively homogeneous functional on $L^1_+(\mu) \times L^1_+(\mu)$. The η-entropy distinguishes two densities f and g. In fact, using Jensen's inequality we obtain

$$H_\eta(f|g) \geqslant \eta\left(\int \frac{f}{g} g\, d\mu\right) = \eta\left(\int f\, d\mu\right) = \eta(1). \tag{1.6}$$

If in addition η is a strictly convex function, then inequality (1.6) is strict for $f \neq g$. Therefore the difference $H_\eta(f|g) - \eta(1)$ may be regarded as a measure of the similarity between the densities f and g.

Now we give two examples of η-entropies different from the Kullback–Leibler entropy.

(i) The norm $\|\cdot\|$ in $L^1(\mu)$ is an η-entropy. Indeed, if we put $\eta_1(u) = |1-u|$ then $H_{\eta_1}(f|g) = \|f-g\|$.

(ii) Another interesting and useful example is

$$H_\alpha(f|g) = -\int f^\alpha g^{1-\alpha}\, d\mu, \tag{1.7}$$

where $\alpha \in (0, 1)$. Here $H_\alpha = H_{\eta_\alpha}$ for $\eta_\alpha(u) = -u^\alpha$.

Inequality (1.6) for the H_α-entropy is just the Minkowski inequality. The H_α-entropy with $\alpha = \frac{1}{2}$ will be used in Section 4.

2. Stochastic operators and decreasing of entropy. Let $(X_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$, be σ-finite measure spaces and let $U: L^1(\mu_1) \to L^1(\mu_2)$ be a linear operator. The operator U is called *positive* if $U(L^1_+(\mu_1)) \subset L^1_+(\mu_2)$. If additionally $\int Uf\, d\mu_2 = \int f\, d\mu_1$ for $f \in L^1(\mu_1)$, then U is called *stochastic*. It is well known [4] that each positive operator U can be extended to the space of all measurable functions $f\colon X_1 \to \bar{R}$ for which $f^- \in L^1(\mu_1)$, where $f^- = \max(0, -f)$. This extension is given by $\tilde{U}f = \sup\{Ug\colon g \in L^1(\mu_1),\ g \leqslant f\}$. The operator $\tilde{U}$ preserves monotonicity, i.e., for $f \leqslant g$ we have $\tilde{U}f \leqslant \tilde{U}g$. Moreover, if $f^- \in L^1(\mu_1)$ then $(\tilde{U}f)^- \in L^1(\mu_2)$. If U is stochastic then $\tilde{U}$ also preserves the integral.

As in Section 1 we shall assume that $\eta\colon [0, \infty) \to R$ is a continuous convex function and φ is given by (1.1).

THEOREM 2.1. *Let $U\colon L^1(\mu_1) \to L^1(\mu_2)$ be a stochastic operator. Then*

$$H_\eta(f|g) \geqslant H_\eta(Uf|Ug) \quad \text{for all } f, g \in L^1_+(\mu_1). \tag{2.1}$$

In order to prove this theorem we need the following generalization of Jensen's inequality.

PROPOSITION 2.2. *Let $U\colon L^1(\mu_1) \to L^1(\mu_2)$ be a positive linear operator. Then*

$$\tilde{U}\varphi(f, g) \geqslant \varphi(Uf, Ug) \quad \text{for all } f, g \in L^1_+(\mu_1). \tag{2.2}$$

Proof of Proposition 2.2. From (1.4) it follows that $\varphi(f, g) \geqslant a_n g + b_n f$. Monotonicity of $\tilde{U}$ implies

$$\tilde{U}\varphi(f, g) \geqslant \tilde{U}(a_n g + b_n f) = a_n Ug + b_n Uf.$$

Using once more (1.4) we obtain (2.2). ∎

Proof of Theorem 2.1 follows immediately from inequality (2.2) and stochasticity of U. ∎

3. The η-entropy and L^1-convergence. Let $(X, \mathscr{F}, \mu)$ be a fixed measure space and let D denote the set of all densities on X. A convex function η: $[0, \infty) \to \boldsymbol{R}$ will be called *strictly convex for* $u = 1$ if there exists $c \in \boldsymbol{R}$ such that

$$\eta(u) > \eta(1) + c(1-u) \quad \text{for } u < 1, \tag{3.1}$$

$$\eta(u) \geqslant \eta(1) + c(1-u) \quad \text{for } u \geqslant 1. \tag{3.2}$$

If, for example, η has a second derivative at $u = 1$ and $\eta''(1) \neq 0$, then η is strictly convex for $u = 1$.

THEOREM 3.1. *Let η: $[0, \infty) \to \boldsymbol{R}$ be a continuous convex function, strictly convex for $u = 1$. Let (f_n) and (g_n) be sequences of densities on X such that $H_\eta(f_n | g_n)$ converges to $\eta(1)$. Then $f_n - g_n$ converges to 0 in L^1-norm.*

Proof. Let c be a real number such that (3.1) and (3.2) hold. Define a new function $\bar{\eta}$ by $\bar{\eta}(u) = \eta(u) - \eta(1) - c(1-u)$. Then $\bar{\eta}$ is nonnegative and

$$\bar{\eta}(u) > 0 \quad \text{for } u < 1. \tag{3.3}$$

Moreover, $\bar{\eta}$ is convex and we can define the $\bar{\eta}$-entropy. We have $H_{\bar{\eta}}(f_n | g_n) = H_\eta(f_n | g_n) - \eta(1)$, which implies

$$\lim_{n \to \infty} H_{\bar{\eta}}(f_n | g_n) = 0. \tag{3.4}$$

Now, fix $r \in (0, 1)$ and define

$$A_n = \{x \in X\colon f_n(x) < g_n(x)\}, \quad B_{r,n} = \{x \in X\colon f_n(x) < r g_n(x)\}, \quad C_{r,n} = A_n \setminus B_{r,n}.$$

From (3.3) it follows that there exists $\lambda_r > 0$ such that $\bar{\eta}(u) > \lambda_r(1-u)$ for $u \in (0, r)$. Since f_n and g_n are densities, this implies that

$$\begin{aligned}\|g_n - f_n\|_{L^1(\mu)} &= 2 \int_{A_n} (g_n - f_n)\, d\mu \\ &= 2 \int_{B_{r,n}} (1 - f_n/g_n) g_n\, d\mu + 2 \int_{C_{r,n}} (g_n - f_n)\, d\mu \\ &\leqslant 2\lambda_r^{-1} H_{\bar{\eta}}(f_n | g_n) + 2(1-r).\end{aligned}$$

Now from (3.4) it follows that

$$\limsup_{n \to \infty} \|g_n - f_n\|_{L^1} \leqslant 2 - 2r. \tag{3.5}$$

Since this holds for every $r \in (0, 1)$, we finally obtain $\lim_{n \to \infty} \|g_n - f_n\| = 0$. ∎

Now using Theorem 3.1 we prove the following result by Elmroth [3].

COROLLARY 3.2. *Let $X = \boldsymbol{R}^3$ and let dx be the Lebesgue measure on $\boldsymbol{R}^3$. Let $E(x)$ be the Maxwellian on X, i.e., $E(x) = \exp(a|x|^2 + b\cdot x + c)$, $a < 0$, $b \in \boldsymbol{R}^3$, $c \in \boldsymbol{R}$, and let (f_n) be a sequence of densities. Assume that*

$$\int_{\boldsymbol{R}^3} g(x) f_n(x)dx = \int_{\boldsymbol{R}^3} g(x)E(x)dx \tag{3.6}$$

for $g(x) = 1$, x_1, x_2, x_3, $|x|^2$. If $H(f_n) = \int_{\boldsymbol{R}^3} f_n(x)\ln f_n(x)dx$ converges to $H(E)$, then f_n converges to E in L^1-norm.

Proof. Using (3.6) we calculate the Kullback–Leibler entropy $H_0(f_n|\,E)$:

$$H_0(f_n|\,E) = \int \big(f_n(x)\ln f_n(x) - f_n(x)(a|x|^2 + b\cdot x + c)\big)dx$$
$$= H(f_n) - \int E(x)(a|x|^2 + b\cdot x + c)dx = H(f_n) - H(E).$$

Thus $H_0(f_n|\,E) \to 0$ and consequently $f_n \to E$ in L^1. ∎

4. Applications to stochastic semigroups. Let $(X, \mathscr{F}, \mu)$ be a measure space. A family $\{U^t\}_{t\geqslant 0}$ of stochastic operators on $L^1(\mu)$ will be called a *continuous stochastic semigroup* if it satisfies the following conditions:

(i) $U^0 = \mathrm{Id}$ (Id = identity),
(ii) $U^{t+s} = U^t U^s$ for $t, s \geqslant 0$,
(iii) for every $f \in L^1(\mu)$ the function $t \to U^t f$ is continuous.

We shall consider two examples of stochastic semigroups. In both cases $X = \boldsymbol{R}^d$, and μ is the Lebesgue measure on $\boldsymbol{R}^d$.

EXAMPLE 4.1. Consider the heat equation

$$\partial u/\partial t = \Delta u(t, x), \quad t \geqslant 0,\ x \in \boldsymbol{R}^d, \tag{4.1}$$

with the initial condition

$$u(0, x) = v(x), \quad x \in \boldsymbol{R}^d, \tag{4.2}$$

where $\Delta = \sum_{j=1}^d \partial^2/\partial x_j^2$ is the d-dimensional Laplacian. It is known that problem (4.1), (4.2) generates a continuous stochastic semigroup $\{U_1^t\}_{t\geqslant 0}$ on $L^1(\boldsymbol{R}^d)$ given by

$$U_1^t v(x) = u(t, x) = \int_{\boldsymbol{R}^d} v(y) K_1(t, x, y)dy,$$

where

$$K_1(t, x, y) = (4\pi t)^{-d/2}\exp\big(-|x-y|^2/(4t)\big).$$

EXAMPLE 4.2. Consider the Dirichlet problem on the half space (cf. [2])

$$\partial^2 u/\partial t^2 = -\Delta u(t, x), \quad t \geqslant 0,\ x \in \boldsymbol{R}^d, \tag{4.3}$$

$$u(0, x) = v(x), \quad x \in \boldsymbol{R}^d. \tag{4.4}$$

This problem generates a continuous semigroup $\{U_2^t\}_{t\geqslant 0}$ on $L^1(\boldsymbol{R}^d)$ given by

$$U_2^t v(x) = u(t, x) = \int_{\boldsymbol{R}^d} v(y) K_2(t, x, y)dy,$$

where

$$K_2(t, x, y) = \frac{2t}{\omega_{d+1}[t^2+|x-y|^2]^{(d+1)/2}}$$

and ω_{d+1} is the surface area of the unit sphere in $\boldsymbol{R}^{d+1}$.

The functions $K_1(t, 0, x)$ and $K_2(t, 0, x)$ are the fundamental solutions of the corresponding problems. For every $t > 0$ we have $K_1(t, 0, \cdot)\in D(\boldsymbol{R}^d)$ and $K_2(t, 0, \cdot)\in D(\boldsymbol{R}^d)$. The functions $K_i(t, x, y)$ may be written in the form

$$K_i(t, x, y) = t^{-\alpha d} k_i\big(t^{-\alpha}(x-y)\big), \tag{4.5}$$

where $k_i(x) = K_i(1, 0, x)$ is a continuous function and $\alpha = i/2$ for $i = 1, 2$. Moreover, for every $t \geqslant 0$ we have

$$(U_i^t k_i)(x) = K_i(t+1, 0, x). \tag{4.6}$$

THEOREM 4.3. *For every $v\in D(\boldsymbol{R}^d)$ and $i = 1, 2$, we have*

$$\lim_{t\to\infty} \|U_i^t v - K_i(t, 0, \cdot)\| = 0. \tag{4.7}$$

Proof. The proof in both cases $i = 1, 2$ is the same and we shall omit the index i. Set $w = U^1 v$. According to Theorem 3.1 and (4.6) it is sufficient to verify that for $\eta(u) = -\sqrt{u}$ we have

$$\lim_{t\to\infty} H_\eta(U^t w\,|\,U^t k) = \eta(1) = -1. \tag{4.8}$$

Since the η-entropy is convex and $w, k\in D(\boldsymbol{R}^d)$, we have

$$\begin{aligned} H_\eta(U^t w\,|\,U^t k) &= H_\eta\big(\textstyle\int w(y) K(t, \cdot, y)\,dy\,\big|\,\int k(z) K(t, \cdot, z)\,dz\big) \\ &\leqslant \textstyle\int\int w(y)k(z) H_\eta\big(K(t, \cdot, y)\,|\,K(t, \cdot, z)\big)\,dy\,dz. \end{aligned}$$

Using (4.5), (1.7) we obtain

$$H_\eta(U^t w\,|\,U^t k) \leqslant -\int\int\int w(y)k(z)\big(k(x-t^{-\alpha}y)k(x-t^{-\alpha}z)\big)^{1/2}dx\,dy\,dz.$$

In this inequality the integrand is nonnegative and becomes $w(y)k(z)k(x)$ as $t\to\infty$. This implies, according to the Fatou lemma,

$$\limsup_{t\to\infty} H_\eta(U^t w\,|\,U^t k) \leqslant -\int\int\int w(y)k(z)k(x)\,dx\,dy\,dz = -1 = \eta(1).$$

This and (1.6) give (4.8). ∎

References

[1] N. Bourbaki, *Fonctions d'une variable réelle*, Hermann, Paris 1958.

[2] R. Courant, D. Hilbert, *Methods of Mathematical Physics*, Vol. II. *Partial Differential Equations*, Interscience, New York 1962.

[3] T. Elmroth, *On the H-function and convergence towards equilibrium for a space-homogeneous molecular density*, SIAM J. Appl. Math. 44 (1984), 150–159.

[4] I. Csiszár, *Information-type measure of difference of probablity distributions and indirect observations*, Studia Sci. Math. Hungar. 2 (1967), 299–318.

[5] —, *Generalized entropy and quantization problems*, in: Trans. Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Prague, September 19–25, 1971, Academia, Prague 1973, 159–174.

[6] J. Fritz, *An information-theoretical proof of limit theorems for reversible Markov processes*, ibidem, 183–197.

[7] S. H. Foguel, *The Ergodic Theory of Markov Processes*, Van Nostrand, New York 1969.

[8] S. Kullback, *Information Theory and Statistics*, Wiley, New York 1969.

[9] A. Lasota and M. C. Mackey. *Probabilistic Properties of Deterministic Systems*, Cambridge University Press, Cambridge 1985.

[10] F. Liese and I. Vajda, *Convex Statistical Distances*, Teubner-Texte, Teubner-Verlag, Leipzig 1987.

[11] B. Misra and I. Prigogine, *On the foundations of kinetic theory*, in: Stochastic Nonlinear Systems in Physics, Chemistry and Biology, Proc. Workshop, Bielefeld, October 5–11, 1980, L. Arnold, R. Lefever (eds.), Springer, Berlin 1981.

[12] A. Rényi, *On measures of entropy and information*, in: Proc. 4th Berkeley Symposium on Math. Statist. and Probability, Vol. I, Berkeley 1961, 547–561.

[13] J. Voigt, *Stochastic operators, information and entropy*, Commun. Math. Phys. 81 (1981), 31–38.

INSTITUTE OF MATHEMATICS
SILESIAN UNIVERSITY
Bankowa 14, 40-007 Katowice, Poland

Reçu par la Rédaction le 15.05.1989