ARKIV FÖR MATEMATIK Band 4 nr 18

Communicated 25 May 1960 by EINAR HILLE and OTTO FROSTMAN

Contributions to information theory for abstract alphabets

By ALEXANDRA IONESCU TULCEA[1]

In this paper we give some results centering around B. McMillan's theorem in information theory. The paper is presented in the abstract setting developed by A. Perez in [9] and [10]. In § 1 we prove some preliminary results. The main theorems are given in § 2. They deal with mean convergence of order p (where $1 \leqslant p < \infty$), domination by a function in L^p, as well as convergence almost everywhere. § 3 is devoted to various remarks and comments. In particular, it is shown here that various classical theorems from information theory for finite alphabets are particular cases of the results proved in this paper. The appendix contains a short and direct proof of the fact that (essentially, this follows also from the general theorem 1), in the case of a finite alphabet, the almost-everywhere convergence holds in McMillan's theorem.

1. Preliminary results

Let X be a set. Let $\mathcal{B}$ be a σ-algebra of subsets of X, λ and ϱ two probabilities on $\mathcal{B}$. We say that λ is absolutely continuous with respect to ϱ, and we write $\lambda \prec \varrho$, whenever the relations $E \in \mathcal{B}$, $\varrho(E) = 0$ imply $\lambda(E) = 0$; we denote by $d\lambda/d\varrho$ the corresponding Radon–Nikodym density.

Throughout this paper we shall use the following notations:

$$\log^+ t = \begin{cases} \log t & \text{if } t > 1 \\ 0 & \text{if } t \leqslant 1 \end{cases} \quad \text{and} \quad \log^- t = \begin{cases} \log t & \text{if } t < 1 \\ 0 & \text{if } t \geqslant 1. \end{cases}$$

Let $(\mathcal{C}_n)_{1 \leqslant n < \infty}$ be a sequence of σ-algebras on X, such that $\mathcal{C}_n \subset \mathcal{C}_{n+1}$ for each $n \geqslant 1$; denote with $\mathcal{C}_\infty$ the σ-algebra generated by $\bigcup_{n=1}^{\infty} \mathcal{C}_n$. If $\lambda = \lambda_\infty$ is a probability on $\mathcal{C}_\infty$, we shall denote with λ_n the restriction of λ to $\mathcal{C}_n$.

Proposition 1. *Let $(\mathcal{C}_n)_{1 \leqslant n < \infty}$ be a sequence of σ-algebras on X such that $\mathcal{C}_n \subset \mathcal{C}_{n+1}$ for each $n \geqslant 1$. Let $\mathcal{C}_\infty$ be the σ-algebra generated by $\bigcup_{n=1}^{\infty} \mathcal{C}_n$, and $\lambda = \lambda_\infty$, $\varrho = \varrho_\infty$ two probabilities on $\mathcal{C}_\infty$ such that $\lambda_n \prec \varrho_n$ for each $n \geqslant 1$. Then:*

(i) *For each $t > 0$ we have*

[1] This paper was sponsored by the Office of Ordnance Research under contract No. DA-19-020-ORD-4912.

$$\lambda\left(\left\{x \mid \inf_{1\leqslant n<\infty} \frac{d\lambda_n}{d\varrho_n}(x) < t\right\}\right) \leqslant t. \tag{1}$$

(ii) *There is* $1<q<\infty$ *such that* $\sup_{1\leqslant n<\infty}\int (d\lambda_n/d\varrho_n)^q\, d\varrho_n$ *is finite, if and only if there are two constants* $C>0$, $\delta>0$ *verifying the inequality*

$$\lambda\left(\left\{x \mid \sup_{1\leqslant n<\infty}\left|\log\frac{d\lambda_n}{d\varrho_n}(x)\right| > t\right\}\right) \leqslant C e^{-t\delta} \tag{2}$$

for every $t>0$.

(i) Let $t>0$ and define the sets

$$A(t) = \left\{x \mid \inf_{1\leqslant n<\infty} \frac{d\lambda_n}{d\varrho_n}(x) < t\right\} \tag{3}$$

and

$$\left.\begin{aligned} A_1(t) &= \left\{x \mid \frac{d\lambda_1}{d\varrho_1}(x) < t\right\}, \\ A_j(t) &= \left\{x \mid \inf\left(\frac{d\lambda_1}{d\varrho_1}(x), .., \frac{d\lambda_{j-1}}{d\varrho_{j-1}}(x)\right) \geqslant t, \frac{d\lambda_j}{d\varrho_j}(x) < t\right\}, \quad j\geqslant 2. \end{aligned}\right\} \tag{4}$$

It is easily seen that for each $t>0$ we have

$$A_j(t) \cap A_k(t) = \phi \quad \text{if} \quad j\neq k,\ 1\leqslant j,\ k<\infty; \tag{5}$$

$$A(t) = \bigcup_{j=1}^{\infty} A_j(t); \tag{6}$$

$$\lambda(A_j(t)) \leqslant t\,\varrho(A_j(t)) \quad \text{for each } 1\leqslant j<\infty. \tag{7}$$

Using (5), (6) and (7), we deduce immediately the inequality (1).

(ii) For each $t>0$ define

$$B(t) = \left\{x \mid \sup_{1\leqslant n<\infty} \frac{d\lambda_n}{d\varrho_n}(x) > t\right\}. \tag{8}$$

It is clear that ($A(e^{-t})$ is defined by (3)):

$$\left\{x \mid \sup_{1\leqslant n<\infty}\left|\log\frac{d\lambda_n}{d\varrho_n}(x)\right| > t\right\} = A(e^{-t}) \cup B(e^t) \tag{9}$$

for every $t>0$. Assume now that there exists $1<q<\infty$ such that

$$\sup_{1\leqslant n<\infty}\int\left(\frac{d\lambda_n}{d\varrho_n}\right)^q d\varrho_n$$

is finite. In this case $\lambda_\infty \prec \varrho_\infty$ and $[(d\lambda_n/d\varrho_n)^q]_{1\leqslant n\leqslant\infty}$ is a ϱ-semi-martingale[1]; it then follows that $\int(\sup_{1\leqslant n\leqslant\infty}(d\lambda_n/d\varrho_n)^q\,d\varrho$ is finite (see [2], p. 319, p. 317). For $t>0$ we have

$$\begin{aligned}
e^{t(q-1)}\lambda(B(e^t)) &\leqslant \int_{B(e^t)}\left(\sup_{1\leqslant n\leqslant\infty}\frac{d\lambda_n}{d\varrho_n}\right)^{q-1}d\lambda \\
&\leqslant \int_{B(e^t)}\left(\sup_{1\leqslant n\leqslant\infty}\frac{d\lambda_n}{d\varrho_n}\right)^{q-1}\cdot\frac{d\lambda_\infty}{d\varrho_\infty}\,d\varrho \\
&\leqslant \int_{B(e^t)}\left(\sup_{1\leqslant n\leqslant\infty}\frac{d\lambda_n}{d\varrho_n}\right)^{q}d\varrho.
\end{aligned}\tag{10}$$

Using (1), (9) and (10), we deduce immediately (2). Conversely, assume that there exist two constants $C>0$, $\delta>0$ verifying the inequality (2) for every $t>0$. Let $1<q<\delta+1$. Then, for each $t>1$ we have

$$\begin{aligned}
&\lambda\left(\left\{x\,\middle|\left(\sup_{1\leqslant n<\infty}\frac{d\lambda_n}{d\varrho_n}(x)\right)^{q-1}>t\right\}\right)\\
&\quad\leqslant\lambda\left(\left\{x\,\middle|\sup_{1\leqslant n<\infty}\left|\log\frac{d\lambda_n}{d\varrho_n}(x)\right|>\log\sqrt[q-1]{t}\right\}\right)\\
&\quad\leqslant C/t^{\delta/(q-1)}.
\end{aligned}$$

It follows that $\int[\sup_{1\leqslant n<\infty}(d\lambda_n/d\varrho_n)]^{q-1}\,d\lambda$ is finite, and hence that

$$\sup_{1\leqslant n<\infty}\int\left(\frac{d\lambda_n}{d\varrho_n}\right)^q d\varrho_n$$

is finite. Thus the proposition is proved.

Let $(\mathsf{C}_n)_{1\leqslant n<\infty}$ be a sequence of σ-algebras on X such that $\mathsf{C}_n\subset\mathsf{C}_{n+1}$ for each $n\geqslant 1$, C_∞ the σ-algebra generated by $\bigcup_{n=1}^\infty \mathsf{C}_n$, and $\lambda=\lambda_\infty$, $\varrho=\varrho_\infty$ two probabilities on C_∞. We may now give the following consequences of proposition 1:

Corollary 1. *Suppose that* $\lambda_n\prec\varrho_n$ *for each* $n\geqslant 1$. *Then:*

(i) $\sup_{1\leqslant n<\infty}\left(-\log^-\dfrac{d\lambda_n}{d\varrho_n}\right)\in L^p(X,\mathsf{C}_\infty,\lambda)$ *for each* $1\leqslant p<\infty$;

(ii) $\liminf_{n\to\infty}\dfrac{1}{n}\log\dfrac{d\lambda_n}{d\varrho_n}(x)\geqslant 0$ *λ-almost everywhere.*

(i) Using (1) ((i), proposition 1), we can write

$$\lambda\left(\left\{x\,\middle|\sup_{1\leqslant n<\infty}\left(-\log^-\frac{d\lambda_n}{d\varrho_n}(x)\right)>t\right\}\right)=\lambda\left(\left\{x\,\middle|\inf_{1\leqslant n<\infty}\frac{d\lambda_n}{d\varrho_n}(x)<e^{-t}\right\}\right)\leqslant e^{-t}\tag{11}$$

[1] Relative to the sequence $(\mathsf{C}_n)_{1\leqslant n<\infty}$ of σ-algebras.

for each $t>0$. From (11) follows immediately that the function

$$\sup_{1\leqslant n<\infty}\left[-\log^{-}\frac{d\lambda_n}{d\varrho_n}\right]$$

belongs to $L^p(X, C_\infty, \lambda)$ for each $1\leqslant p<\infty$.

(ii) From (i) follows in particular that

$$\lim_{n\to\infty}\frac{1}{n}\log^{-}\frac{d\lambda_n}{d\varrho_n}(x)=0$$

λ-almost everywhere. Therefore,

$$\liminf_{n\to\infty}\frac{1}{n}\log\frac{d\lambda_n}{d\varrho_n}(x)\geqslant\liminf_{n\to\infty}\frac{1}{n}\log^{+}\frac{d\lambda_n}{d\varrho_n}(x)\geqslant 0$$

λ-almost everywhere, and hence (ii) is proved.

Corollary 2. *Suppose that* $\lambda_\infty\prec\varrho_\infty$. *Then for each* $1\leqslant p<\infty$ *the following assertions are equivalent:*

(a) $\sup_{1\leqslant n<\infty}\int|\log(d\lambda_n/d\varrho_n)|^p\,d\lambda_n$ *is finite;*

(b) $\sup_{1\leqslant n<\infty}\int[\log^{+}(d\lambda_n/d\varrho_n)]^p\,d\lambda_n$ *is finite;*

(c) $\log(d\lambda_\infty/d\varrho_\infty)\in L^p(X, C_\infty, \lambda)$;

(d) $\log^{+}(d\lambda_\infty/d\varrho_\infty)\in L^p(X, C_\infty, \lambda)$;

(e) $\lim_{n\to\infty}\|\log(d\lambda_n/d\varrho_n)-\log(d\lambda_\infty/d\varrho_\infty)\|_p=0$.

Let us remark that

$$\lim_{n\to\infty}\log^{+}\frac{d\lambda_n}{d\varrho_n}(x)=\log^{+}\frac{d\lambda_\infty}{d\varrho_\infty}(x),\tag{12}$$

$$\lim_{n\to\infty}\log^{-}\frac{d\lambda_n}{d\varrho_n}(x)=\log^{-}\frac{d\lambda_\infty}{d\varrho_\infty}(x),\tag{13}$$

λ-almost everywhere $\left(\text{since } \lim_{n\to\infty}\frac{d\lambda_n}{d\varrho_n}(x)=\frac{d\lambda_\infty}{d\varrho_\infty}(x)\ \varrho\text{-almost everywhere and } \frac{d\lambda_\infty}{d\varrho_\infty}(x)\neq 0\ \lambda\text{-almost everywhere}\right)$. On the other hand, $\left(\frac{d\lambda_n}{d\varrho_n}\cdot\log^{+}\frac{d\lambda_n}{d\varrho_n}\right)_{1\leqslant n\leqslant\infty}$ is a ϱ-semi-martingale, provided that $\int\frac{d\lambda_\infty}{d\varrho_\infty}\cdot\log^{+}\frac{d\lambda_\infty}{d\varrho_\infty}\,d\varrho$ is finite; equivalently $\left[\log^{+}\frac{d\lambda_n}{d\varrho_n}\right]_{1\leqslant n\leqslant\infty}$ is a λ-semi-martingale, provided that $\int\log^{+}\frac{d\lambda_\infty}{d\varrho_\infty}\,d\lambda$ is finite.

Using (i), corollary 1, the relations (12) and (13), and a classical result on semi-martingales of non-negative functions, (see [2], p. 325), we can easily establish the equivalence of (a), (b), (c), (d) and (e).

2. Main theorems

Let Y be a set, $\mathcal{C}$ a σ-algebra of subsets of Y, and $X=\prod_{n\in Z} X_n$ [1], where $X_n=Y$ for all $n\in Z$. For each $n\in Z$, denote by pr_n the projection of X onto X_n, and by $\mathcal{B}_n$ the σ-algebra $\{pr_n^{-1}(E)\,|\,E\in\mathcal{C}\}$. For each part $I\subset Z$, denote by $\mathcal{B}_I$ the σ-algebra generated by $\bigcup_{n\in I}\mathcal{B}_n$. If $I=\{n\}$, then $\mathcal{B}_{\{n\}}=\mathcal{B}_n$; for $I=Z$, we shall write $\mathcal{B}$ instead of $\mathcal{B}_Z$. Denote with τ the mapping $(x_n)_{n\in Z}\to(x_{n+1})_{n\in Z}$ of X onto X. The mapping τ is $\mathcal{B}$-measurable and $\tau(\mathcal{B}_n)=\mathcal{B}_{n-1}$ for each $n\in Z$. Let now λ be a probability on $\mathcal{B}$. For each part $I\subset Z$, we shall denote with λ_I the restriction of λ to $\mathcal{B}_I$. If $I=\{n\}$, we shall write λ_n instead of $\lambda_{\{n\}}$. The probability λ is stationary if for every $E\in\mathcal{B}$, $\lambda(\tau^{-1}(E))=\lambda(E)$.

Let $(I(s))_{s\in T}$ be a family of disjoint parts of Z; for each $s\in T$, let $\nu_{I(s)}$ be a probability on $\mathcal{B}_{I(s)}$. In what follows we shall denote with $\otimes_{s\in T}\nu_{I(s)}$ the (probability on $\mathcal{B}_{\bigcup_{s\in T}I(s)}$) direct product of the $\nu_{I(s)}$.

A probability ν on $\mathcal{B}$ has the *property* (A) if ν is *stationary* and $\nu=\otimes_{n\in Z}\nu_n$. We shall say that a system $\{\mu,\nu\}$ of probabilities defined on $\mathcal{B}$ has the *property* (B) if $\mu_{(0,\dots,n-1)}\prec\nu_{(0,\dots,n-1)}$ for each $n\geq 1$. A system $\{\mu,\nu\}$ of probalities defined on $\mathcal{B}$ has the *property* (PE) if $\{\mu,\nu\}$ has the property (B) and if

$$\lim_{n\to\infty}\frac{1}{n}\log\frac{d\mu_{(0,\dots,n-1)}}{d\nu_{(0,\dots,n-1)}}(x)=h(x)$$

exists, is finite, and $h(\tau(x))=h(x)$ μ-almost everywhere. We say that $\{\mu,\nu\}$ has the *property* (ME_q), where $1\leq q<\infty$, if $\{\mu,\nu\}$ has the property (B) and if there exist two functions $h\in L^q(X,\mathcal{B},\mu)$, $G^*\in L^q(X,\mathcal{B},\mu)$ such that:

(α) h is invariant under τ ($h\circ\tau=h$);

(β) $\left|\frac{1}{n}\log\frac{d\mu_{(0\dots n-1)}}{d\nu_{(0\dots n-1)}}\right|\leq G^*$ for all $n\geq 1$;

(γ) $\lim_{n\to\infty}\left\|\frac{1}{n}\log\frac{d\mu_{(0\dots n-1)}}{d\nu_{(0\dots n-1)}}-h\right\|_q=0.$

Let now $\{\mu,\nu\}$ be a system of probabilities on $\mathcal{B}$ such that (see [10]):

(C) μ is stationary;

(D) $\mu_{(\dots,-1,0)}\prec\mu_{(\dots,-2,-1)}\otimes\mu_0$;

(E) $\mu_0\prec\nu_0$.

If ν has the property (A), then, using (C), (D), (E), we deduce:

[1] $Z=\{\dots,-1,0,1,\dots\}$.

$$\mu_{(\ldots,-1,0)} \prec \mu_{(\ldots,-2,-1)} \otimes \nu_0; \tag{14}$$

$$\mu_{(-n,\ldots,-1,0)} \prec \mu_{(-n,\ldots,-1)} \otimes \nu_0 \text{ for all } n \geqslant 1; \tag{15}$$

$$\text{the system } \{\mu, \nu\} \text{ has the property } (B). \tag{16}$$

Let now

$$\left.\begin{aligned} f_n &= \frac{d\,\mu_{(0,\ldots,n-1)}}{d\,\nu_{(0,\ldots,n-1)}} \quad \text{for} \quad 1 \leqslant n < \infty; \\ g_0 &= \frac{d\,\mu_0}{d\,\nu_0}, \quad g_k = \frac{d\,\mu_{(-k,\ldots,-1,0)}}{d\mu_{(-k,\ldots,-1)} \otimes \nu_0} \quad \text{for} \quad 1 \leqslant k < \infty; \\ g_\infty &= \frac{d\,\mu_{(\ldots,-1,0)}}{d\mu_{(\ldots,-1)} \otimes \nu_0}. \end{aligned}\right\} \tag{17}$$

We then have

$$\frac{1}{n} \log f_n = \frac{1}{n} \sum_{k=0}^{n-1} \log g_k \circ \tau^k \tag{18}$$

for all $n \geqslant 1$.

Theorem 1. *Let $\{\mu, \nu\}$ be a system of probabilities on $\mathfrak{B}$ satisfying (C), (D) and (E). Suppose that ν has the property (A). Then:*

(i) *The sequence $[(1/n) \log f_n]_{1 \leqslant n < \infty}$ converges in $L^1(X, \mathfrak{B}, \mu)$ if and only if $\sup_{0 \leqslant k < \infty} \int \log^+ g_k \, d\mu$ is finite; if h is the limit of the sequence $[(1/n) \log f_n]_{1 \leqslant n < \infty}$ in $L^1(X, \mathfrak{B}, \mu)$, then $h \geqslant 0$, μ-almost everywhere.*

(ii) *If $1 < q < \infty$ and $\sup_{0 \leqslant k < \infty} \int (\log^+ g_k)^q \, d\mu$ is finite, then the system $\{\mu, \nu\}$ has the properties (ME_q) and (PE).*

(iii) *If, in particular, there exists $1 < p < \infty$ such that $\sup_{0 \leqslant k < \infty} \int g_k^p \, d\mu_{(-k,\ldots,-1)} \otimes \nu_0$ is finite, then the system $\{\mu, \nu\}$ has the properties (PE) and (ME_q) for all $1 \leqslant q < \infty$. In this case, there exist a function G^*, dominating the sequence $[(1/n) \log f_n]_{1 \leqslant n < \infty}$, and two constants $C_1 > 0$, $\delta_1 > 0$, verifying the inequality*

$$\mu(\{x \mid G^*(x) > t\}) \leqslant C_1 e^{-t\delta_1} \tag{19}$$

for each $t > 0$.

(i) Assume that $\sup_{0 \leqslant k < \infty} \int \log^+ g_k \, d\mu$ is finite. Using corollary 2, we deduce that $\log g_\infty \in L^1(X, \mathfrak{B}, \mu)$, $\log g_k \in L^1(X, \mathfrak{B}, \mu)$ for each $k \geqslant 0$, and that

$$\lim_{k \to \infty} \| \log g_k - \log g_\infty \|_1 = 0.$$

The argument of B. McMillan (see [8]; see also A. Perez [10]) shows then that the sequence $[(1/n) \log f_n]_{1 \leqslant n < \infty}$ converges in $L^1(X, \mathfrak{B}, \mu)$; if h is the limit function, then (use (ii), corollary 1) $h(x) \geqslant 0$ μ-almost everywhere. Conversely, assume that the sequence $[(1/n) \log f_n]_{1 \leqslant n < \infty}$ converges in $L^1(X, \mathfrak{B}, \mu)$. By (i), corollary 1

$$G' = \sup_{0 \leqslant k < \infty} (-\log^- g_k) \in L^q(X, \mathfrak{B}, \mu) \quad \text{for each} \quad 1 \leqslant q < \infty. \tag{20}$$

Using (18), we obtain for each $n \geqslant 1$

$$\frac{1}{n}\sum_{k=0}^{n-1} \log^+ g_k \circ \tau^k \leqslant \left|\frac{1}{n}\log f_n\right| + \frac{1}{n}\sum_{k=0}^{n-1} G' \circ \tau^k,$$

whence
$$\frac{1}{n}\sum_{k=0}^{n-1} \int \log^+ g_k \, d\mu \leqslant \left\|\frac{1}{n}\log f_n\right\|_1 + \|G'\|_1 \leqslant M, \tag{21}$$

where M is a constant independent of n. But $\int \log^+ g_k \, d\mu \leqslant \int \log^+ g_{k+1} \, d\mu$ for each $k \geqslant 0$, since the sequence $(\log^+ g_k)_{0 \leqslant n < \infty}$ is a μ-semi-martingale (see for instance the proof of corollary 2). We deduce then from (21) that $\int \log^+ g_k \, d\mu \leqslant M$ for each $k \geqslant 0$, and hence that $\sup_{0 \leqslant k < \infty} \int \log^+ g_k \, d\mu$ is finite.

(ii) Assume that $1 < q < \infty$ and that $\sup_{0 \leqslant k < \infty} \int (\log^+ g_k)^q \, d\mu$ is finite. Since the sequence $(\log^+ g_k)_{0 \leqslant k \leqslant \infty}$ is a μ-semi-martingale, we have then (see [2], p. 317):

$$G'' = \sup_{0 \leqslant k < \infty} \log^+ g_k \in L^q(X, \mathcal{B}, \mu). \tag{22}$$

From (20) and (22) we get

$$G = \sup_{0 \leqslant k < \infty} |\log g_k| \leqslant G' + G''. \tag{23}$$

Now $G \in L^q(X, \mathcal{B}, \mu)$, since $G' \in L^q(X, \mathcal{B}, \mu)$ and $G'' \in L^q(X, \mathcal{B}, \mu)$. On the other hand, $\lim_{k \to \infty} \log g_k(x) = \log g_\infty(x)$ μ-almost everywhere. Hence we can apply a generalized ergodic theorem (see [1] and [7]) and deduce the existence of a function $h \in L^q(X, \mathcal{B}, \mu)$, invariant under τ, such that

$$\lim_{n \to \infty} \frac{1}{n}\sum_{k=0}^{n-1} \log g_k(\tau^k(x)) = h(x) \tag{24}$$

μ-almost everywhere. Therefore the system $\{\mu, \nu\}$ has the property (PE). To complete the proof of (ii), we have to show that the system $\{\mu, \nu\}$ has also the property (ME_q). By the "dominated ergodic theorem", we have

$$G^* = \sup_{1 \leqslant n < \infty} \frac{1}{n}\sum_{k=0}^{n-1} G \circ \tau^k \in L^q(X, \mathcal{B}, \mu). \tag{25}$$

Using (24), (25) and the obvious fact that G^* dominates the sequence $[(1/n)\log f_n]_{1 \leqslant n < \infty}$, we deduce that

$$\lim_{n \to \infty} \left\|\frac{1}{n}\log f_n - h\right\|_q = 0. \tag{26}$$

Hence the system $\{\mu, \nu\}$ has the property (ME_q), and thus (ii) is proved.

(iii) From the hypothesis and from (ii), proposition 1, we deduce that the function G defined in (23) satisfies the inequality

$$\mu(\{x \mid G(x) > t\}) \leqslant C e^{-t\delta} \tag{27}$$

for each $t>0$, $C>0$ and $\delta>0$ being suitable constants. It follows that $G \in L^q(X, \mathcal{B}, \mu)$ for every $1 \leqslant q < \infty$; hence $\sup_{0 \leqslant k < \infty} \int (\log^+ g_k)^q \, d\mu$ is finite for each $1 \leqslant q < \infty$, and so, by (ii) above, the system $\{\mu, \nu\}$ has the properties (PE) and (ME_q) for each $1 \leqslant q < \infty$. The inequality (19) follows from (27) and the "maximal ergodic theorem" (see [3], p. 676) applied to the functions G, G^* defined by (23) and (25), respectively. In fact, it is clear that the inequality (19) is satisfied for suitable $C_1 > 0$, $\delta_1 > 0$, when $0 < t \leqslant 1$. For $t \geqslant 1$, the inequality (19) can be proved, for instance, [1] using the following relations:

$$\mu(\{x \mid G^*(x) > t\}) \leqslant \frac{2}{t} \int\limits_{\{x \mid G(x) > t/2\}} G \cdot d\mu \leqslant \frac{2}{t} \|G\|_2 \cdot \mu\left(\left\{x \mid G(x) > \frac{t}{2}\right\}\right)^{\frac{1}{2}}.$$

Hence the theorem is completely proved.

Theorem 2. *Let ν be a probability on $\mathcal{B}$ having the property (A), and λ, μ two probabilities on $\mathcal{B}$ such that $\lambda \prec \mu$. Assume that the system $\{\mu, \nu\}$ has the property (PE). Then:*

(i) *The system $\{\lambda, \nu\}$ has the property (PE) and we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{d\lambda_{(0,\ldots,n-1)}}{d\nu_{(0,\ldots,n-1)}}(x) = \lim_{n\to\infty} \frac{1}{n} \log \frac{d\mu_{(0,\ldots,n-1)}}{d\nu_{(0,\ldots,n-1)}}(x) = h(x) \geqslant 0 \tag{28}$$

λ-almost everywhere.

(ii) *If $1 \leqslant q < \infty$ and if the sequence*

$$\left(\left(\frac{1}{n} \log^+ \frac{d\lambda_{(0,\ldots,n-1)}}{d\nu_{(0,\ldots,n-1)}}\right)^q\right)_{1 \leqslant n < \infty}$$

is λ-uniformly integrable, then

$$\lim_{n\to\infty} \left\| \frac{1}{n} \log \frac{d\lambda_{(0,\ldots,n-1)}}{d\nu_{(0,\ldots,n-1)}} - h \right\|_q = 0.$$

(iii) *In particular, if there exist constants $1 < p < \infty$, $a \geqslant 1$, such that*

$$\int \left(\frac{d\lambda_{(0,\ldots,n-1)}}{d\nu_{(0,\ldots,n-1)}}\right)^p d\nu_{(0,\ldots,n-1)} = O(a^n) \quad \text{when } n \to \infty$$

then the system $\{\lambda, \nu\}$ has the property (ME_q) for every $1 \leqslant q < \infty$, and $0 \leqslant h(x) \leqslant \log a/(p-1)$ λ-almost everywhere.

(i) Let us remark first that for $n \geqslant 1$ we have

[1] This simple argument was shown to the author by C. Ionescu Tulcea; it replaces the initial argument of the author.

$$\frac{d\,\lambda_{(0,\,..\,,\,n-1)}}{d\,\nu_{(0,\,..\,,\,n-1)}}(x)=\frac{d\,\lambda_{(0,\,..\,,\,n-1)}}{d\mu_{(0,\,..\,,\,n-1)}}(x)\cdot\frac{d\mu_{(0,\,..\,,\,n-1)}}{d\nu_{(0,\,..\,,\,n-1)}}(x) \tag{29}$$

$\nu_{(0,\,..\,,\,n-1)}$-almost everywhere, and hence also $\lambda_{(0,\,..\,,\,n-1)}$-almost everywhere. It follows that

$$\frac{1}{n}\log\frac{d\,\lambda_{(0,\,..\,,\,n-1)}}{d\,\nu_{(0,\,..\,,\,n-1)}}(x)=\frac{1}{n}\log\frac{d\,\lambda_{(0,\,..\,,\,n-1)}}{d\mu_{(0,\,..\,,\,n-1)}}(x)+\frac{1}{n}\log\frac{d\,\mu_{(0,\,..\,,\,n-1)}}{d\,\nu_{(0,\,..\,,\,n-1)}}(x) \tag{30}$$

λ-almost everywhere (obviously, $(d\,\lambda_{(0,\,..\,n-1)}/d\,\nu_{(0,\,..\,,\,n-1)})\,(x)\neq 0$ λ-almost everywhere). Since the system $\{\mu,\,\nu\}$ has the property (PE) and since $\lambda\prec\mu$, there exists a function $h\geqslant 0$ (use (ii), corollary 1) and finite λ-almost everywhere, such that $h\,(\tau\,(x))=h\,(x)$ and

$$\lim_{n\to\infty}\frac{1}{n}\log\frac{d\mu_{(0,\,..\,,\,n-1)}}{d\,\nu_{(0,\,..\,,\,n-1)}}(x)=h\,(x) \tag{31}$$

λ-almost everywhere. Hence to prove (i), it will be enough (in view of (30) and (31)) to show that $\lim_{n\to\infty}(1/n)\log\,(d\,\lambda_{(0,\,..\,n-1)}/d\,\mu_{(0,\,..\,n-1)})\,(x)=0$ λ-almost everywhere. But the sequence $(d\,\lambda_{(0,\,..\,,\,n-1)}/d\,\mu_{(0,\,..\,,\,n-1)})_{1\leqslant n<\infty}$ is a μ-martingale, and since $\lambda_{(0,\,1,\,..\,)}\prec\mu_{(0,\,1,\,..\,)}$, we deduce that

$$\lim_{n\to\infty}\frac{d\,\lambda_{(0,\,..\,,\,n-1)}}{d\mu_{(0,\,..\,,\,n-1)}}(x)=\frac{d\,\lambda_{(0,\,1,\,..\,)}}{d\mu_{(0,\,1,\,..\,)}}(x)\neq 0$$

λ-almost everywhere. It follows that

$$\lim_{n\to\infty}\frac{1}{n}\log\frac{d\,\lambda_{(0,\,..\,,\,n-1)}}{d\,\mu_{(0,\,..\,,\,n-1)}}(x)=0$$

λ-almost everywhere. Hence (i) is proved.

(ii) follows immediately from (i) and corollary 1. In fact, it is enough to remark that

$$\sup_{1\leqslant n<\infty}\left(-\frac{1}{n}\log^{-}\frac{d\,\lambda_{(0,\,..\,,\,n-1)}}{d\,\nu_{(0,\,..\,,\,n-1)}}\right)\in L^{q}\,(X,\,\mathcal{B},\,\lambda)\ \text{ for each }\ 1\leqslant q<\infty. \tag{32}$$

(iii) By hypothesis there are constants $1<p<\infty$, $a\geqslant 1$, $M>0$ such that $\int(d\,\lambda_{(0,\,..\,,\,n-1)}/d\,\nu_{(0,\,..\,,\,n-1)})^{p-1}\,d\,\lambda_{(0,\,..\,,\,n-1)}\leqslant M\,a^{n}$ for each $n\geqslant 1$. It follows that for each $n\geqslant 1$ and $t>0$

$$\lambda\left(\left\{x\,\Big|\,\frac{1}{n}\log^{+}\frac{d\,\lambda_{(0,\,..\,,\,n-1)}}{d\,\nu_{(0,\,..\,,\,n-1)}}(x)>t\right\}\right)\leqslant M\,a^{n}\,e^{-n\,(p-1)\,t}. \tag{33}$$

It is then easily seen that for $t>\log a/(p-1)$ we have

$$\lambda\left(\left\{x \,\middle|\, \sup_{1\leqslant n<\infty}\left(\frac{1}{n}\log^{+}\frac{d\lambda_{(0,\ldots,n-1)}}{d\nu_{(0,\ldots,n-1)}}(x)\right)>t\right\}\right)$$

$$\leqslant \sum_{n=1}^{\infty}\lambda\left(\left\{x \,\middle|\, \frac{1}{n}\log^{+}\frac{d\lambda_{(0,\ldots,n-1)}}{d\nu_{(0,\ldots,n-1)}}(x)>t\right\}\right)\leqslant M a/(e^{(p-1)t}-a).$$

We deduce that

$$\sup_{1\leqslant n<\infty}\left(\frac{1}{n}\log^{+}\frac{d\lambda_{(0,\ldots,n-1)}}{d\nu_{(0,\ldots,n-1)}}\right)\in L^{q}(X,\mathcal{B},\lambda) \text{ for each } 1\leqslant q<\infty. \tag{34}$$

From (32) and (34) follows

$$\sup_{1\leqslant n<\infty}\left|\frac{1}{n}\log\frac{d\lambda_{(0,\ldots,n-1)}}{d\nu_{(0,\ldots,n-1)}}\right|\in L^{q}(X,\mathcal{B},\lambda) \text{ for each } 1\leqslant q<\infty. \tag{35}$$

Hence, by (ii) above, the system $\{\lambda, \nu\}$ has the property $(M E_q)$ for every $1\leqslant q<\infty$. To complete the proof of (iii), we have to show that $h(x)\leqslant\log a/(p-1)$ λ-almost everywhere. Let $T>\log a/(p-1)$ and define

$$E_n(T)=\left\{x \,\middle|\, \frac{1}{n}\log^{+}\frac{d\lambda_{(0,\ldots,n-1)}}{d\nu_{(0,\ldots,n-1)}}(x)>T\right\} \text{ for } n\geqslant 1,$$

and $E_\infty(T)=\{x\,|\,h(x)>T\}$; let $\varphi_{E_n(T)}$, $\varphi_{E_\infty(T)}$ be the characteristic functions of the sets $E_n(T)$ and $E_\infty(T)$, respectively. Since

$$\lim_{n\to\infty}\frac{1}{n}\log^{+}\frac{d\lambda_{(0,\ldots,n-1)}}{d\nu_{(0,\ldots,n-1)}}(x)=h(x)$$

λ-almost everywhere (see corollary 1), it follows that $\varphi_{E_\infty(T)}(x)\leqslant\liminf_{n\to\infty}\varphi_{E_n(T)}(x)$ λ-almost everywhere. From (33) we deduce that $\lim_{n\to\infty}\lambda(E_n(T))=0$, and hence, applying Fatou's theorem, that $\lambda(E_\infty(T))=0$. Since $T>\log a/(p-1)$ was arbitrary, we conclude that $\lambda(\{x\,|\,h(x)>\log a/(p-1)\})=0$. This completes the proof of the theorem.

3. Remarks

(1) Concerning corollary 2, see also [9]. (2) Concerning (i), theorem 1, see also [10]. (3) The condition

$$\int\left(\frac{d\lambda_{(0,\ldots,n-1)}}{d\nu_{(0,\ldots,n-1)}}\right)^{p} d\nu_{(0,\ldots,n-1)}=O(a^{n})$$

when $n\to\infty$ $(1<p<\infty,\ a\geqslant 1)$ of (iii), theorem 2, was suggested by the reading of [5]. (4) Let Y, $\mathcal{C}$, X and τ be the objects introduced in the beginning of § 2, and assume that: Y is a finite set (= alphabet), $\mathcal{C}$ is the set of all subsets of Y. Let a be the number of elements (= letters) of Y. Define the probability ν on $\mathcal{B}$ as follows: $\nu = \otimes_{n \in Z} \nu_n$, where $\nu_n [pr_n^{-1}(\{y\})] = 1/a$ for each $y \in Y$ and $n \in Z$.

(I) Let μ be a stationary probability on $\mathcal{B}$. It is clear that the system $\{\mu, \nu\}$ satisfies (C), (D), (E) and also that ν has the property (A). Let us also remark that $g_k(x) \leqslant a$, $\mu_{(-k,\ldots,-1)} \otimes \nu_0$-almost everywhere, for each $k \geqslant 0$ (obviously, the g_k's are defined by (17)). From (iii), theorem 1, it then follows that the system $\{\mu, \nu\}$ has the properties (PE) and (ME_q) for all $1 \leqslant q < \infty$. These results contain in particular those given in [5] for discrete stationary sources.

(II) Let λ be an almost periodic probability on $\mathcal{B}$ in the sense of [5]. There exists then a stationary probability μ on $\mathcal{B}$, such that $\lambda \prec \mu$ (see [5]). Let us remark now that $(d\lambda_{(0,\ldots,n-1)}/d\nu_{(0,\ldots,n-1)})(x) \leqslant a^n$, $\nu_{(0,\ldots,n-1)}$-almost everywhere, for each $n \geqslant 1$. As we saw in (I), the system $\{\mu, \nu\}$ has the property (PE); from (i) and (iii), theorem 2, it then follows that the system $\{\lambda, \nu\}$ has the properties (PE) and (ME_q) for all $1 \leqslant q < \infty$, and that $0 \leqslant h(x) \leqslant \log a$ λ-almost everywhere (since $\|h\|_\infty \leqslant (\log a^p)/(p-1) = (p/(p-1)) \log a$ for each $1 < p < \infty$). These results contain in particular those given in [5] for discrete almost periodic sources.

APPENDIX

Let $(X, \mathcal{B}, P)$ be a probability space. A measurable measure-preserving transformation of $(X, \mathcal{B}, P)$ is a mapping τ of X into X such that $\tau^{-1}(E) \in \mathcal{B}$ and $P(\tau^{-1}(E)) = P(E)$ for every $E \in \mathcal{B}$. For any set $E \in \mathcal{B}$ and σ-algebra $\mathcal{C} \subset \mathcal{B}$, we denote with $P(E \mid \mathcal{C})$ the conditional probability of E relative to $\mathcal{C}$. For any finite σ-algebra $\mathcal{A} \subset \mathcal{B}$, we denote with $\pi(\mathcal{A})$ the (uniquely determined) partition of the space X such that the σ-algebra generated by $\pi(\mathcal{A})$ coincides with $\mathcal{A}$, and with $C(\mathcal{A})$ the number of elements of $\pi(\mathcal{A})$. If $(\mathcal{C}_\iota)_\iota$ is a family of σ-algebras contained in $\mathcal{B}$, we denote with $\bigvee_\iota \mathcal{C}_\iota$ the σ-algebra generated by $\bigcup_\iota \mathcal{C}_\iota$.

Let $\mathcal{A} \subset \mathcal{B}$ be a finite σ-algebra, $\mathcal{C} \subset \mathcal{B}$ an arbitrary σ-algebra. The information of $\mathcal{A}$ and the conditional information of $\mathcal{A}$ relative to $\mathcal{C}$ are defined by the equations (we write $0 \log 0 = 0$, $-\log 0 = +\infty$ and we denote with φ_A the characteristic function of the set (A):

$$I(\mathcal{A})(x) = -\sum_{A \in \pi(\mathcal{A})} \varphi_A(x) \log P(A)$$

for every $x \in X$, and

$$I(\mathcal{A} \mid \mathcal{C})(x) = -\sum_{A \in \pi(\mathcal{A})} \varphi_A(x) \log P(A \mid \mathcal{C})(x)$$

for almost every $x \in X$, respectively.

LEMMA. *Let $\mathcal{A} \subset \mathcal{B}$ be a finite σ-algebra, $(\mathcal{C}_n)_{0 \leqslant n < \infty}$ a sequence of σ-algebras such that $\mathcal{C}_n \subset \mathcal{C}_{n+1} \subset \mathcal{B}$ for each $n \geqslant 0$. For every $t > 0$ we have*

$$P(\{x \mid \sup_{0 \leqslant n < \infty} I(\mathcal{A} \mid \mathcal{C}_n)(x) > t\}) \leqslant C(\mathcal{A}) e^{-t}.$$

The proof is straightforward and similar to that of (i), proposition 1 [1]. It will be enough to show that for each $A \in \pi(\mathcal{A})$ and $t > 0$

$$(\ast) \quad P(\{x \mid \sup_{0 \leqslant n < \infty} I(\mathcal{A} \mid \mathcal{C}_n)(x) > t\} \cap A) \leqslant e^{-t}.$$

Fix now $A \in \pi(\mathcal{A})$ and define the sets

$$F(t) = \{x \mid \sup_{0 \leqslant n < \infty} I(\mathcal{A} \mid \mathcal{C}_n)(x) > t\} \cap A$$

$$F_0(t) = \{x \mid P(A \mid \mathcal{C}_0)(x) < e^{-t}\}$$

$$F_k(t) = \{x \mid \inf_{0 \leqslant j \leqslant k-1} P(A \mid \mathcal{C}_j)(x) \geqslant e^{-t},\ P(A \mid \mathcal{C}_k)(x) < e^{-t}\},\ k \geqslant 1$$

for every $t > 0$. It is easily seen that for each $t > 0$ we have:

(α) $F_n(t) \cap F_m(t) = \phi$ if $n \neq m$, $0 \leqslant n, m < \infty$;

(β) $F_n(t) \in \mathcal{C}_n$ for every $n \geqslant 0$;

(γ) $P(F_n(t) \cap A) \leqslant e^{-t} P(F_n(t))$ for every $n \geqslant 0$;

(δ) $F(t) = \bigcup_{n=0}^{\infty} F_n(t) \cap A$.

From (α), (γ) and (δ) the inequality ($\ast$) follows immediately; hence the lemma is proved.

Remarks. (1) Let u be a strictly decreasing (not necessarily convex or concave) mapping of $[0, 1]$ onto $[\delta, +\infty]$ ($\delta \geqslant 0$) and let $v = \overset{-1}{u}$. The above lemma holds for every $t > \delta$ (and the method of proof is the same) if we replace everywhere $-\log$ by u and e^{-t} by $v(t)$. (2) Let $\mathcal{A} \subset \mathcal{B}$ be a finite σ-algebra, and $(\mathcal{C}_n)_{0 \leqslant n < \infty}$ a sequence of σ-algebras such that $\mathcal{C}_n \subset \mathcal{C}_{n+1} \subset \mathcal{B}$ for each $n \geqslant 0$. Then: (2 a) $\sup_{0 \leqslant n < \infty} I(\mathcal{A} \mid \mathcal{C}_n)$ belongs to $L^q(X, \mathcal{B}, P)$ for every $1 \leqslant q < \infty$; (2 b) $\lim_{n \to \infty} I(\mathcal{A} \mid \mathcal{C}_n)(x)$ exists and is finite P-almost everywhere. The assertion (2 a) is a consequence of the above lemma, and (2 b) follows from the fact that for every $A \in \pi(\mathcal{A})$, $(P(A \mid \mathcal{C}_n))_{0 \leqslant n < \infty}$ is a martingale. (3) Let τ be a measurable measure-preserving transformation of $(X, \mathcal{B}, P)$, $\mathcal{A} \subset \mathcal{B}$ a finite σ-algebra, and $(\mathcal{C}_n)_{0 \leqslant n < \infty}$ the sequence of σ-algebras defined as follows: $\mathcal{C}_0 = \{X, \phi\}$, $\mathcal{C}_k = \mathsf{V}_{i=1}^{k} \tau^{-i} \mathcal{A}$ for $k \geqslant 1$. It is clear that $\mathcal{C}_n \subset \mathcal{C}_{n+1} \subset \mathcal{B}$ for each $n \geqslant 0$. Let us remark now that for every $n \geqslant 1$

$$I(\mathsf{V}_{i=0}^{n-1} \tau^{-i} \mathcal{A}) = I(\mathcal{A}) \circ \tau^{n-1} + \sum_{k=1}^{n-1} I(\mathcal{A} \mid \mathsf{V}_{i=1}^{k} \tau^{-i} \mathcal{A}) \circ \tau^{n-1-k}$$

(see [4]). Since the sequence $(I(\mathcal{A}), I(\mathcal{A} \mid \tau^{-1} \mathcal{A}), \ldots, I(\mathcal{A} \mid \mathsf{V}_{i=1}^{k} \tau^{-i} \mathcal{A}), \ldots)$ satisfies (2 a) and (2 b) (see remark (2) above), we can apply a generalized ergodic theorem (see [1] and [7]) and deduce the existence of a function h, belonging to $L^q(X, \mathcal{B}, P)$ for every $1 \leqslant q < \infty$, invariant under τ, such that

[1] The lemma can in fact be deduced from (i), proposition 1; however, the proof given below is direct and shorter.

$$\lim_{n\to\infty} \| I(\bigvee_{i=0}^{n-1} \tau^{-i} \mathcal{A})/n - h \|_q = 0 \text{ for every } 1 \leqslant q < \infty,$$

and
$$\lim_{n\to\infty} I(\bigvee_{i=0}^{n-1} \tau^{-i} \mathcal{A})(x)/n = h(x) \ P\text{-almost everywhere.}$$

In this way we obtain the classical form of McMillan's theorem in information theory (see [4], [5], [6], [8]), as well as the assertion that almost everywhere convergence holds in McMillan's theorem. See also [1] and the remarks made in [4]. (4) The results contained in this appendix were presented on February 9, 1960, in a seminar on information theory held at Yale University under the direction of Professor S. Kakutani.

REFERENCES

1. L. Breiman, The individual ergodic theorem of information theory, Ann. Math. Stat. *28* (1957).
2. J. L. Doob, Stochastic processes, Wiley, New York, 1953.
3. N. Dunford & J. T. Schwartz, Linear operators, I, New York, 1958.
4. P. R. Halmos, Entropy in ergodic theory, Lecture notes, University of Chicago, September 1959.
5. K. Jacobs, Die Übertragung diskreter Informationen durch periodische und fastperiodische Kanäle, Math. Ann., *137* (1959).
6. A. Khintchin, On the fundamental theorems of information theory, Uspehi Mat. Nauk, vol. XI, no. 1, 1956.
7. P. Maker, The ergodic theorem for a sequence of functions, Duke Math. J. *6* (1940).
8. B. McMillan, The basic theorems of information theory, Ann. Math. Stat. *24* (1953).
9. A. Perez, Notions généralisées d'incertitude, d'entropie et d'information du point de vue de la théorie des martingales, Trans. First Prague Conference on Information Theory, 1956 (1957).
10. A. Perez, Sur la théorie de l'information dans le cas d'un alphabet abstract, Trans. First Prague Conference on Information Theory, 1956 (1957).

Tryckt den 10 december 1960

Uppsala 1960. Almqvist & Wiksells Boktryckeri AB