

# Pacific Journal of Mathematics

## **A NOTE ON GENERALIZATIONS OF SHANNON-MCMILLAN THEOREM**

SHU-TEH CHEN MOY

# A NOTE ON GENERALIZATIONS OF SHANNON-MCMILLAN THEOREM

SHU-TEH C. MOY

**1. Introduction.** This paper is a sequel to an earlier paper [6]. All notations in [6] remain in force. As in [6] we shall consider two probability measures  $\mu, \nu$  on the infinite product  $\sigma$ -algebra of subsets of the infinite product space  $\Omega = \pi X$ .  $\nu$  is assumed to be *stationary* and  $\mu$  to be *Markovian with stationary transition probabilities*. Extensions to  $K$ -Markovian  $\mu$  are immediate.  $\nu_{m,n}$ , the contraction of  $\nu$  to  $\mathcal{F}_{m,n}$ , is assumed to be *absolutely continuous* with respect to  $\mu_{m,n}$ , the contraction of  $\mu$  to  $\mathcal{F}_{m,n}$ , and  $f_{m,n}$  is the Radon-Nikodym derivative. In [6] the following theorem is proved. If  $\int \log f_{0,0} d\nu < \infty$  and if there is a number  $M$  such that

$$(1) \quad \int (\log f_{0,n} - \log f_{0,n-1}) d\nu \leq M \text{ for } n = 1, 2, \dots$$

then  $\{n^{-1} \log f_{0,n}\}$  converges in  $L_1(\nu)$ . (1) is also a necessary condition for the  $L_1(\nu)$  convergence of  $\{n^{-1} \log f_{0,n}\}$ . We consider this theorem as a generalization of the Shannon-McMillan theorem of information theory. In the setting of [6] the Shannon-McMillan theorem may be stated as follows. Let  $X$  be a finite set of  $K$  points. Let  $\nu$  be any stationary probability measure of  $\mathcal{F}$ , and  $\mu$  the equally distributed independent measure on  $\mathcal{F}$ . Then  $\{n^{-1} \log f_{0,n}\}$  converges in  $L_1(\nu)$ . In fact, the  $P(x_0, x_1, \dots, x_n)$  of Shannon-McMillan is equal to  $K^{-(n+1)} f_{0,n}$ . The convergence with probability one of  $\{n^{-1} \log P(x_0, \dots, x_n)\}$  for a finite set  $X$  was proved by L. Breiman [1] [2]. K.L. Chung then extended Breiman's result to a countable set  $X$ . [3]. In this paper we shall prove that the convergence with  $\nu$ -probability one of  $\{n^{-1} \log f_{0,n}\}$  follows from the following condition.

$$(2) \quad \int \frac{f_{0,n}}{f_{0,n-1}} d\nu \leq L, n = 1, 2, \dots$$

(2) is a stronger condition than (1) since by Jensen's inequality

$$\log \int \frac{f_{0,n}}{f_{0,n-1}} d\nu \geq \int \log \frac{f_{0,n}}{f_{0,n-1}} d\nu.$$

An application to the case of countable  $X$  is also discussed.

---

Received November 28, 1960.

2. **The convergence theorem.** As was proved in [6], condition (1) implies the  $L_1(\nu)$  convergence of  $\{\log f_{-k,0} - \log f_{-k,-1}\}$  ([6] Theorem 1, 4). The convergence with  $\nu$ -probability one is automatically true ([6] Theorem 3). Applying a theorem (with obvious modification for  $T$  not necessarily ergodic) of Breiman ([1], Theorem 1) the convergence with  $\nu$ -probability one of  $\{n^{-1} \log f_{0,n}\}$  follows from the condition

$$(3) \quad \int \sup_{k \geq 1} |\log f_{-k,0} - \log f_{-k,-1}| d\nu < \infty .$$

We shall now investigate conditions under which (3) is valid.

Lemma 1. *The following inequality is always true.*

$$(4) \quad \int \sup_{k \geq 1} \log \frac{f_{-k,-1}}{f_{-k,0}} d\nu < \infty .$$

*Proof.* Let  $\nu'_{-k,0}$  be as in Lemma 1 [6]. Then

$$\nu_{-k,0} \ll \nu'_{-k,0} \ll \mu_{-k,0}$$

and

$$\frac{d\nu_{-k,0}}{d\nu'_{-k,0}} = \frac{f_{-k,0}}{f_{-k,-1}}, \quad \frac{d\nu'_{-k,0}}{d\mu_{-k,0}} = f_{-k,-1} .$$

Since  $\mu$  is Markovian,  $\nu'_{-k,0}$  are consistent for  $k = 1, 2, \dots$ . We shall prove (4) under the assumption that there is a probability measure  $\nu'$  on  $\mathcal{F}_{-\infty,0}$  which is an extension of  $\nu'_{-k,0}$  for  $k = 1, 2, \dots$ . We shall also prove Lemma 2 under this assumption. If no such  $\nu'$  exists, the usual procedure of representing  $\Omega$  into the space of real sequences may be used and the same conclusion follows (cf. the proof of Theorem 4[6]).

Let  $m$  be a nonnegative integer and

$$E(m) = \left[ \sup_{k \geq 1} \log \frac{f_{-k,-1}}{f_{-k,0}} > m \right] ,$$

$$E_k(m) = \left[ \sup_{1 \leq j < k} \log \frac{f_{-j,-1}}{f_{-j,0}} \leq m, \log \frac{f_{-k,-1}}{f_{-k,0}} > m \right] .$$

On  $E_k(m)$  we have

$$f_{-k,0} \leq 2^{-m} f_{-k,-1} .$$

Hence

$$\int_{E_k(m)} f_{-k,0} d\mu \leq 2^{-m} \int_{E_k(m)} f_{-k,-1} d\mu$$

so that

$$\nu[E_k(m)] \leq 2^{-m} \nu'[E_k(m)] .$$

Therefore

$$\nu[E(m)] \leq 2^{-m} \nu'[E(m)] \leq 2^{-m}$$

and

$$\int \sup_{k>1} \log \frac{f_{-k,-1}}{f_{-k,0}} d\nu \leq \sum_{m \geq 0} \nu[E(m)] \leq \sum_{m \geq 0} 2^{-m} < \infty .$$

Note that (4) is proved without assuming the integrability of either  $\log f_{-k,0}$  or  $\log f_{-k,-1}$  or  $\log \frac{f_{-k,0}}{f_{-k,-1}}$ .

LEMMA 2. *If there is a number  $L$  such that*

$$(5) \quad \int \frac{f_{-k,0}}{f_{-k,-1}} d\nu \leq L \text{ for } k = 1, 2, \dots$$

then

$$(6) \quad \int \sup_{k \geq 1} \log \frac{f_{-k,0}}{f_{-k,-1}} d\nu < \infty .$$

*Proof.* It is clear that

$$\int \frac{f_{-k,0}}{f_{-k,-1}} d\nu = \int \left( \frac{f_{-k,0}}{f_{-k,-1}} \right)^2 d\nu'$$

where  $\nu'$  is defined in the proof of Lemma 1.

Since  $\{f_{-k,0}/f_{-k,-1}, k = 1, 2, \dots\}$  is a  $\nu'$ -martingale,  $\{(f_{-k,0}/f_{-k,-1})^2, k = 1, 2, \dots\}$  is a  $\nu'$ -semi-martingale. Hence (5) implies that

$$\nu_{-\infty,0} \ll \nu', \int \left( \frac{d\nu_{-\infty,0}}{d\nu'} \right)^2 d\nu' < \infty, \left( \frac{f_{-k,0}}{f_{-k,-1}} \right)^2$$

are uniformly  $\nu'$ -integrable and  $\{(f_{-1,0}/f_{-1,-1})^2, (f_{-2,0}/f_{-2,-1})^2 \dots, (d\nu_{-\infty,0}/d\nu')^2\}$  is a  $\nu'$ -semi-martingale (Theorem 4.1s, pp. 324[5]).

Hence for any set  $F$  defined by  $x_0, x_{-1}, \dots, x_{-k}$

$$\int_F \left( \frac{f_{-k,0}}{f_{-k,-1}} \right)^2 d\nu \leq \int_F \left( \frac{f_{-(k+1),0}}{f_{-(k+1),-1}} \right)^2 d\nu \leq \int_F \left( \frac{d\nu_{-\infty,0}}{d\nu'} \right)^2 d\nu'$$

so that

$$(7) \quad \int_F \frac{f_{-k,0}}{f_{-k,-1}} d\nu \leq \int_F \frac{f_{-(k+1),0}}{f_{-(k+1),-1}} d\nu \leq \int_F \frac{d\nu_{-\infty,0}}{d\nu'} d\nu .$$

In fact, we have just proved that

$$\left\{ \frac{f_{-1,0}}{f_{-1,-1}}, \frac{f_{-2,0}}{f_{-2,-1}}, \dots, \frac{d\nu_{-\infty,0}}{d\nu'} \right\}$$

is a  $\nu$ -semi-martingale. Now let

$$F(m) = [\sup_{k \geq 1} \log \frac{f_{-k,0}}{f_{-k,-1}} > m]$$

and

$$F_k(m) = [\sup_{1 \leq j < k} \log \frac{f_{-j,0}}{f_{-j,-1}} \leq m, \log \frac{f_{-k,0}}{f_{-k,-1}} > m].$$

On  $F_k(m)$  we have

$$f_{-k,-1} \leq 2^{-m} f_{-k,0}.$$

Hence

$$\begin{aligned} \int_{F_k(m)} f_{-k,-1} \frac{f_{-k,0}}{f_{-k,-1}} d\mu &\leq 2^{-m} \int_{F_k(m)} \left( \frac{f_{-k,0}}{f_{-k,-1}} \right)^2 d\mu \\ &= 2^{-m} \int_{F_k(m)} \frac{f_{-k,0}}{f_{-k,-1}} d\nu. \end{aligned}$$

Applying (7), we obtain

$$\nu[F_k(m)] \leq 2^{-m} \int_{F_k(m)} \frac{d\nu}{d\nu'} d\nu,$$

therefore,

$$\nu[F(m)] \leq 2^{-m} \int_{F(m)} \frac{d\nu}{d\nu'} d\nu \leq 2^{-m} L.$$

Hence

$$\int_{k \geq 1} \sup \log \frac{f_{-k,0}}{f_{-k,-1}} d\nu \leq \sum_{m \geq 0} \nu[F(m)] \leq \sum_{m \geq 0} 2^{-m} L < \infty.$$

Combining Lemmas 1, 2 and noting that

$$\int \frac{f_{0,n}}{f_{0,n-1}} d\nu = \int \frac{f_{-n,0}}{f_{-n,-1}} d\nu$$

(cf. Theorem 1, [6]), we obtain the following theorem.

**THEOREM 1.** *If there is a number  $L$  such that*

$$\int \frac{f_{0,n}}{f_{0,n-1}} d\nu \leq L \text{ for } n = 1, 2, \dots \text{ then}$$

$$\int \sup_{k \geq 1} |\log f_{-k,0} - \log f_{-k,-1}| d\nu < \infty$$

and  $\{n^{-1} \log f_{0,n}\}$  converges with  $\nu$ -probability one.

Extensions of Lemma 1, Lemma 2 and Theorem 1 to  $K$ -Markovian  $\mu$  are immediate.

3. **The countable case.** Let  $X$  be countable with elements denoted by  $a$ . Let  $\nu$  be an arbitrary stationary probability measure on  $\mathcal{F}$ . Let

$$P(a_0, a_1, \dots, a_n) = \nu[x_0 = a_0, x_1 = a_1, \dots, x_n = a_n] .$$

Let

$$H_1 = -\sum_a P(a) \log P(a) = -\int \log P(x_n) d\nu .$$

Carleson showed that

$$(8) \quad H_1 < \infty$$

implies the  $L_1(\nu)$  convergence of  $\{n^{-1} \log P(x_0, x_1, \dots, x_n)\}$  [3]. Chung showed that (8) also implies the convergence with  $\nu$ -probability one of  $\{n^{-1} \log P(x_0, x_1, \dots, x_n)\}$  [4]. Let  $\mu$  be defined by

$$\mu[x_m = a_0, x_{m+1} = a_1, \dots, x_n = a_{n-m}] = P(a_0)P(a_1) \dots P(a_{n-m}) .$$

$\mu$  may be called the independent measure obtained from  $\nu$ . Then  $\nu_{m,n} \ll \mu_{m,n}$  with derivative

$$f_{m,n} = \frac{P(x_m, \dots, x_n)}{P(x_m) \dots P(x_n)}$$

and

$$(9) \quad \log \frac{f_{m,n}}{f_{m,n-1}} = \log \frac{P(x_m, \dots, x_n)}{P(x_m, \dots, x_{n-1})} - \log P(x_n) .$$

It follows from (9) that

$$\int (\log f_{0,n} - \log f_{0,n-1}) d\nu \leq \int -\log P(x_n) d\nu = H_1 .$$

Hence (8) implies that (1) is satisfied, therefore  $\{n^{-1} \log f_{0,n}\}$  converges in  $L_1(\nu)$  by Theorem 5 [6]. Since

$$\log f_{0,n} = \log P(x_0, \dots, x_n) + \sum_{k=0}^n \log P(x_k) ,$$

Carleson's theorem follows immediately. Furthermore, it follows from (9) and Lemma 1 that

$$\int \sup_{k \geq 1} [\log \frac{P(x_{-k}, \dots, x_{-1})}{P(x_{-k}, \dots, x_0)} + \log P(x_0)] d\nu < \infty .$$

Hence (8) implies

$$\int \sup_{k \geq 1} \log \frac{P(x_{-k}, \dots, x_{-1})}{P(x_{-k}, \dots, x_0)} d\nu < \infty$$

and Chung's theorem [4] follows.

By using a similar approach we shall give a sharpend version of Carleson's and Chung's theorems.

Let

$$P(a_0 | a_{-l}, \dots, a_{-1}) = \frac{P(a_{-l}, \dots, a_{-1}, a_0)}{P(a_{-l}, \dots, a_{-1})}$$

and let

$$\begin{aligned} H_l &= - \sum_{a_{-l}, \dots, a_{-1}} P(a_{-l}, \dots, a_0) \log P(a_0 | a_{-l}, \dots, a_{-1}) \\ &= - \int \log P(x_n | x_{n-l}, \dots, x_{n-1}) d\nu . \end{aligned}$$

$H_l$  is nonnegative but may be  $+\infty$ . It is known that

$$H_1 \geq H_2 \geq H_3 \geq \dots$$

Let

$$H = \lim_{l \rightarrow \infty} H_l .$$

The limit is taken to be  $+\infty$  if all  $H_l$  are  $+\infty$ .

**THEOREM 2.** *If  $H < \infty$  then  $\{n^{-1} \log P(x_0, \dots, x_n)\}$  converges both in  $L_1(\nu)$  and with  $\nu$ -probability one.*

*Proof.* There is an  $l$  such that  $H_l < \infty$ . We define an  $l$ -Markovian measure  $\mu$  on  $\mathcal{F}$  as follows.

$$\mu[x_m = a_0, x_{m+1} = a_1, \dots, x_n = a_{n-m}] = P(a_0, \dots, a_{n-m})$$

if  $n - m \leq l$ ,

$$\begin{aligned} &\mu[x_m = a_0, x_{m+1} = a_1, \dots, x_n = a_{n-m}] \\ &= P(a_0, \dots, a_l) P(a_{l+1} | a_1, \dots, a_l) \dots P(a_{n-m} | a_{n-m-l}, \dots, a_{n-m-1}) \end{aligned}$$

if  $n - m > l$ . It is easy to check that  $\mu$  is well defined and  $\nu_{m,n} \ll \mu_{m,n}$ . It is clear that, if  $n - m > l$ ,

$$\log \frac{f_{m,n}}{f_{m,n-1}} = \log \frac{P(x_m, \dots, x_n)}{P(x_m, \dots, x_{n-1})} - \log P(x_n | x_{n-l}, \dots, x_{n-1}) .$$

The rest of the proof goes in the same manner as for the case  $H_1 < \infty$  since Theorem 5 [6] and Lemma 1 of this paper remain true for  $l$ -Markovian  $\mu$ .

#### REFERENCES

1. L. Breiman, *The individual ergodic theorem of information theory*, Ann. Math. Stat. **28** (1957), 809-811.
2. ———, *A correction to the individual ergodic theorem of information theory*, Ann. Math. Stat., **31** (1960), 809-810.
3. L. Carleson, *Two remarks on the basic theorems of information theory*, Math. Scand., **6** (1958), 175-180.
4. K. L. Chung, *A note on the ergodic theorem of information theory*, Ann. Math. Stat., **32** (1961), 612-614.
5. J. L. Doob, *Stochastic Processes*, John Wiley and Sons Inc., New York, 1953.
6. Shu-Teh C. Moy, *Generalizations of Shannon-McMillan theorem*, Pacific J. Math., **11** (1961), 705-714.

SYRACUSE UNIVERSITY



# PACIFIC JOURNAL OF MATHEMATICS

## EDITORS

RALPH S. PHILLIPS  
Stanford University  
Stanford, California

A. L. WHITEMAN  
University of Southern California  
Los Angeles 7, California

F. H. BROWNELL  
University of Washington  
Seattle 5, Washington

L. J. PAIGE  
University of California  
Los Angeles 24, California

## ASSOCIATE EDITORS

E. F. BECKENBACH  
T. M. CHERRY

D. DERRY  
M. OHTSUKA

H. L. ROYDEN  
E. SPANIER

E. G. STRAUS  
F. WOLF

## SUPPORTING INSTITUTIONS

UNIVERSITY OF BRITISH COLUMBIA  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
UNIVERSITY OF CALIFORNIA  
MONTANA STATE UNIVERSITY  
UNIVERSITY OF NEVADA  
NEW MEXICO STATE UNIVERSITY  
OREGON STATE COLLEGE  
UNIVERSITY OF OREGON  
OSAKA UNIVERSITY  
UNIVERSITY OF SOUTHERN CALIFORNIA

STANFORD UNIVERSITY  
UNIVERSITY OF TOKYO  
UNIVERSITY OF UTAH  
WASHINGTON STATE COLLEGE  
UNIVERSITY OF WASHINGTON  
\* \* \*  
AMERICAN MATHEMATICAL SOCIETY  
CALIFORNIA RESEARCH CORPORATION  
HUGHES AIRCRAFT COMPANY  
SPACE TECHNOLOGY LABORATORIES  
NAVAL ORDNANCE TEST STATION

---

Mathematical papers intended for publication in the *Pacific Journal of Mathematics* should be typewritten (double spaced), and the author should keep a complete copy. Manuscripts may be sent to any one of the four editors. All other communications to the editors should be addressed to the managing editor, L. J. Paige at the University of California, Los Angeles 24, California.

50 reprints per author of each article are furnished free of charge; additional copies may be obtained at cost in multiples of 50.

---

The *Pacific Journal of Mathematics* is published quarterly, in March, June, September, and December. The price per volume (4 numbers) is \$12.00; single issues, \$3.50. Back numbers are available. Special price to individual faculty members of supporting institutions and to individual members of the American Mathematical Society: \$4.00 per volume; single issues, \$1.25.

Subscriptions, orders for back numbers, and changes of address should be sent to Pacific Journal of Mathematics, 103 Highland Boulevard, Berkeley 8, California.

Printed at Kokusai Bunken Insatsusha (International Academic Printing Co., Ltd.), No. 6, 2-chome, Fujimi-cho, Chiyoda-ku, Tokyo, Japan.

PUBLISHED BY PACIFIC JOURNAL OF MATHEMATICS, A NON-PROFIT CORPORATION

The Supporting Institutions listed above contribute to the cost of publication of this Journal, but they are not owners or publishers and have no responsibility for its content or policies.

Reprinted 1966 in the United States of America

A. V. Balakrishnan, <i>Prediction theory for Markoff processes</i> .....	1171
Dallas O. Banks, <i>Upper bounds for the eigenvalues of some vibrating systems</i> .....	1183
A. Białyński-Birula, <i>On the field of rational functions of algebraic groups</i> .....	1205
Thomas Andrew Brown, <i>Simple paths on convex polyhedra</i> .....	1211
L. Carlitz, <i>Some congruences for the Bell polynomials</i> .....	1215
Paul Civin, <i>Extensions of homomorphisms</i> .....	1223
Paul Joseph Cohen and Milton Lees, <i>Asymptotic decay of solutions of differential inequalities</i> .....	1235
István Fáry, <i>Self-intersection of a sphere on a complex quadric</i> .....	1251
Walter Feit and John Griggs Thompson, <i>Groups which have a faithful representation of degree less than <math>(p - 1/2)</math></i> .....	1257
William James Firey, <i>Mean cross-section measures of harmonic means of convex bodies</i> .....	1263
Avner Friedman, <i>The wave equation for differential forms</i> .....	1267
Bernard Russel Gelbaum and Jesus Gil De Lamadrid, <i>Bases of tensor products of Banach spaces</i> .....	1281
Ronald Kay Getoor, <i>Infinitely divisible probabilities on the hyperbolic plane</i> .....	1287
Basil Gordon, <i>Sequences in groups with distinct partial products</i> .....	1309
Magnus R. Hestenes, <i>Relative self-adjoint operators in Hilbert space</i> .....	1315
Fu Cheng Hsiang, <i>On a theorem of Fejér</i> .....	1359
John McCormick Irwin and Elbert A. Walker, <i>On <math>N</math>-high subgroups of Abelian groups</i> .....	1363
John McCormick Irwin, <i>High subgroups of Abelian torsion groups</i> .....	1375
R. E. Johnson, <i>Quotient rings of rings with zero singular ideal</i> .....	1385
David G. Kendall and John Leonard Mott, <i>The asymptotic distribution of the time-to-escape for comets strongly bound to the solar system</i> .....	1393
Kurt Kreith, <i>The spectrum of singular self-adjoint elliptic operators</i> .....	1401
Lionello Lombardi, <i>The semicontinuity of the most general integral of the calculus of variations in non-parametric form</i> .....	1407
Albert W. Marshall and Ingram Olkin, <i>Game theoretic proof that Chebyshev inequalities are sharp</i> .....	1421
Wallace Smith Martindale, III, <i>Primitive algebras with involution</i> .....	1431
William H. Mills, <i>Decomposition of holomorphs</i> .....	1443
James Donald Monk, <i>On the representation theory for cylindric algebras</i> .....	1447
Shu-Teh Chen Moy, <i>A note on generalizations of Shannon-McMillan theorem</i> .....	1459
Donald Earl Myers, <i>An imbedding space for Schwartz distributions</i> .....	1467
John R. Myhill, <i>Category methods in recursion theory</i> .....	1479
Paul Adrian Nickel, <i>On extremal properties for annular radial and circular slit mappings of bordered Riemann surfaces</i> .....	1487
Edward Scott O'Keefe, <i>Primal clusters of two-element algebras</i> .....	1505
Nelson Onuchic, <i>Applications of the topological method of Ważewski to certain problems of asymptotic behavior in ordinary differential equations</i> .....	1511
Peter Perkins, <i>A theorem on regular matrices</i> .....	1529
Clinton M. Petty, <i>Centroid surfaces</i> .....	1535
Charles Andrew Swanson, <i>Asymptotic estimates for limit circle problems</i> .....	1549
Robert James Thompson, <i>On essential absolute continuity</i> .....	1561
Harold H. Johnson, <i>Correction to "Terminating prolongation procedures"</i> .....	1571