Pacific Journal of Mathematics

A SIMPLE PROOF OF THE MOY-PEREZ GENERALIZATION OF THE SHANNON-MCMILLAN THEOREM

JOHN CRONAN KIEFFER

Vol. 51, No. 1 November 1974

PACIFIC JOURNAL OF MATHEMATICS
Vol. 51, No. 1, 1974

A SIMPLE PROOF OF THE MOY-PEREZ GENERALIZATION OF THE SHANNON-MCMILLAN THEOREM

J. C. KIEFFER

The Shannon-McMillan Theorem of Information Theory has been generalized by Moy and Perez. The purpose of this paper is to give a simple proof of this generalization.

1. Introduction. Let T be either the semigroup of nonnegative integers N or nonnegative real numbers R^+. Let $\mathscr{U} = \{U^t : t \in T\}$ be a semigroup of measurable mappings from a given measurable space $(\Omega, \mathscr{F})$ to itself. (We suppose U^0 is the identity map.) If X_0 is a measurable mapping from the space $(\Omega, \mathscr{F})$ to another space, let $(X_t : t \in T)$ be the process generated by $\mathscr{U}$; that is, $X_t = X_0 \cdot U^t$, $t \in T$. If $a, b \in T$, $a \leqq b$, let $\mathscr{F}_{ab}$ denote the sub-sigmafield of $\mathscr{F}$ generated by the mappings $\{X_t : t \in T, a \leqq t \leqq b\}$.

Let P, Q be probability measures on $\mathscr{F}$; let $P_{ab}(Q_{ab})$ be the restriction of $P(Q)$ to $\mathscr{F}_{ab}$. We suppose that P_{0t} is absolutely continuous with respect to Q_{0t}, $t \in T$. Then the Radon-Nikodym derivatives $f_{st} = dP_{st}/dQ_{st}$ exist, $s \leqq t$. We assume that the entropies $H_{st} = \int_\Omega \log f_{st} dP$, $s \leqq t$, are all finite. (We use natural logarithms.) It is known that

(1) H_{0t} is a nonnegative, nondecreasing function of t ([6], p. 54); and

(2) If $\|\cdot\|$ denotes the $L^1(P)$ norm, then

$$\|\log (f_{ru}/f_{st})\| \leqq H_{ru} - H_{st} + 1,\ r \leqq s \leqq t \leqq u\ ,$$

([6], inequality (2.4.10), and p. 54).

The Moy-Perez result. The following generalization of the Shannon-McMillan Theorem was proved by Moy [4] for the case $T = N$, and by Perez [5], for the case $T = R^+$.

THEOREM. *Let $(X_t : t \in T)$ be a stationary process with respect to P and a Markov process with stationary transition probabilities with respect to Q. If the sequence $\{n^{-1} H_{0n} : n = 1, 2, \cdots\}$ is bounded above, then the functions $\{t^{-1} \log f_{0t} : t > 0, t \in T\}$ converge as $t \to \infty$ in $L^1(P)$ to a function h which is invariant with respect to $\mathscr{U}$; that is, $h \cdot U^t = h$, $t \in T$.*

To prove this theorem, Moy and Perez embedded the process (X_t) in a bilateral process $(X_t : -\infty < t < \infty)$, stationary with respect to P and Markov with respect to Q; Doob's Martingale Convergence

Theorem was then used. We present a simple proof which requires no such embedding and no martingale theory. The method of proof is a generalization of the method used by Gallager ([2], p. 60) to prove the Shannon-McMillan Theorem, and uses the L^1 version of the Mean Ergodic Theorem.

Proof of the Moy-Perez result. The assumptions made on P and Q imply that:

(3) The sequence $\{H_{0n}: n = 1, 2, \cdots\}$ is convex ([4], Theorem 2); (A sequence $c_1, c_2, \cdots$ is convex if $c_{n+2} - 2c_{n+1} + c_n \geqq 0$, $n=1, 2, \cdots$.)

(4) $f_{0t} \cdot U^s = f_{s,s+t}$ a.e. $[P]$, and therefore $H_{0t} = H_{s,s+t}$ ([4], Theorem 1); and

(5) $E_Q(f_{rt} \mid \mathscr{F}_{0s}) = f_{rs}$, $r \leqq s \leqq t$.

Because of (3), $H_{0n} - H_{0,n-1}$ is an increasing sequence and so has a limit H, possibly infinite. Since

$$n^{-1}H_{0n} = n^{-1}\sum_{i=1}^{n}(H_{0i} - H_{0,i-1}) + n^{-1}H_{00}, \text{ and } \left\{\frac{1}{n}H_{0n}\right\}$$

is bounded,

$$\lim_{n\to\infty} n^{-1}H_{0n} = \lim_{n\to\infty}(H_{0n} - H_{0,n-1}) = H < \infty .$$

From (1), we have

$$[t]^{-1}H_{0[t]}[t]t^{-1} \leqq t^{-1}H_{0t} \leqq ([t] + 1)^{-1}H_{0,[t]+1}([t] + 1)t^{-1} ,$$

which implies that $\lim_{t\to\infty} t^{-1} H_{0t} = H$.

Also, since

$$\| t^{-1}\log f_{0t} - [t]^{-1}f_{0[t]} \| \leqq \| t^{-1}\log f_{0[t]} - [t]^{-1}\log f_{0[t]} \| + \| t^{-1}\log (f_{0t}/f_{0[t]}) \| ,$$

and by (2)

$$\| t^{-1}\log (f_{0t}/f_{0[t]}) \| \leqq t^{-1}(H_{0t} - H_{0[t]} + 1) ,$$

we see that the convergence of $n^{-1}\log f_{0n}$ in $L^1(P)$ as $n \to \infty$ would imply the convergence of $t^{-1}\log f_{0t}$ in $L^1(P)$ as $t \to \infty$ to the same limit.

Now, for fixed $s \in T$, we have for $t \geqq s$,

$$\| t^{-1}\log f_{0t} - t^{-1}\log f_{s,s+t} \| \leqq \| t^{-1}\log (f_{0t}/f_{st}) \| + \| t^{-1}\log (f_{s,s+t}/f_{st}) \| \leqq \frac{2}{t}(H_{0t} - H_{0,t-s} + 1) ,$$

using (2) and (4). Consequently if $\lim_{t\to\infty} t^{-1}\log f_{0t} = h$, then

$$\lim_{t\to\infty} t^{-1}\log f_{s,t+s} = h .$$

It follows then that $h = h \cdot U^s$ because

$$\lim_{t\to\infty} t^{-1} \log f_{s,s+t} = \lim_{t\to\infty} (t^{-1} \log f_{0t}) \cdot U^s = h \cdot U^s ,$$

where we used (4).

These considerations show that it suffices to prove the $L^1(P)$ convergence as $n \to \infty$ of $\{n^{-1} \log f_{0n} : n = 1, 2, \cdots\}$. This we now do.

Given $\varepsilon > 0$, pick N to be a positive integer so large that $| N^{-1} H_{0N} - H | < \varepsilon$, and $| H_{0,N+1} - H_{0N} - H | < \varepsilon$. Define the sequence of functions h_n, $n = N + 1, N + 2, \cdots$, as follows:

$$h_n = f_{0N} \prod_{i=0}^{n-N-1} (f_{i,N+i+1}/f_{i,N+i}) I(f_{i,N+i}) ,$$

where for a given function f, $I(f)$ we define to be the function such that $I(f) = 1$ if $f > 0$, and $I(f) = 0$, otherwise.

Now, using (5), we have

$$E_Q(h_n \mid \mathscr{F}_{0,n-1}) = h_{n-1}[I(f_{n-N-1,n-1})/f_{n-N-1,n-1}] E_Q(f_{n-N-1,n} \mid \mathscr{F}_{0,n-1}) \leqq h_{n-1} .$$

Since h_n is $\mathscr{F}_{0n}$-measurable, it follows that

$$E_P(h_n/f_{0n}) \leqq E_Q(h_n) \leqq E_Q(h_{N+1}) \leqq E_Q(f_{0,N+1}) = 1 .$$

Now

$$| \log x | = 2 \log^+ x - \log x \leqq 2x - \log x;$$

therefore,

$$| n^{-1} \log (h_n/f_{0n}) | \leqq 2n^{-1}(h_n/f_{0n}) - n^{-1} \log (h_n/f_{0n}) ,$$

a.e. $[P]$. Integrating with respect to P, we obtain

$$\| n^{-1} \log f_{0n} - n^{-1} \log h_n \| \leqq 2n^{-1} - n^{-1} E_P [\log (h_n/f_{0n})].$$

However,

$$\begin{aligned} -E_P[\log (h_n/f_{0n})] &= -H_{0N} - (n - N)(H_{0,N+1} - H_{0N}) + H_{0n} \\ &\leqq -N(H - \varepsilon) - (n - N)(H - \varepsilon) \\ &\quad + H_{0n} = -n(H - \varepsilon) + H_{0n} , \end{aligned}$$

and so $\overline{\lim}_{n\to\infty} \| n^{-1} \log f_{0n} - n^{-1} \log h_n \| \leqq \varepsilon$.

Using (4), we have, a.e. $[P]$,

$$n^{-1} \log h_n = n^{-1} \log f_{0N} + n^{-1} \sum_{i=0}^{N-n-1} \log (f_{0,N+1}/f_{0N}) \cdot U^i ,$$

which converges as $n \to \infty$ in $L^1(P)$ to a function h_ε by the Mean Ergodic Theorem ([1], p. 667). This gives

$$\overline{\lim_{n\to\infty}} \| n^{-1} \log f_{0n} - h_\varepsilon \| \leqq \varepsilon , \text{ for every } \varepsilon > 0 ,$$

which makes $n^{-1}\log f_{0n}$ a Cauchy sequence in $L^1(P)$, and therefore a convergent sequence.

Final Remark. For the reader who may wish to consult [5], we point out that the proof of the Moy-Perez Theorem given in [5] is erroneous. The Theorem 2.3 of [5] states that the Moy-Perez result holds as well for the case when $(X_t: t \in T)$ is stationary with respect to P and Q, with no Markov assumption made. This is false; a counterexample is given in [3].

REFERENCES

1. N. Dunford and J. T. Schwartz, *Linear Operators Part I*, 1958.
2. R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, 1968.
3. J. C. Kieffer, *A counterexample to Perez's generalization of the Shannon-McMillan theorem*, Annals of Probability, Vol. **1**, No. 2, April, 1973.
4. Shu-Teh C. Moy, *Generalizations of Shannon-McMillan theorem*, Pacific J. Math., **11** (1961), 705–714.
5. A. Perez, *Extensions of Shannon-McMillan's Limit Theorem to More General Stochastic Processes*, Transactions of the Third Prague Conference on Information Theory, Prague, (1964), 545–574.
6. M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*, Holden-Day, 1964.

Received October 3, 1972.

UNIVERSITY OF MISSOURI-ROLLA

Printed in Japan by Intarnational Academic Printing Co., Ltd., Tokyo, Japan

Pacific Journal of Mathematics

Vol. 51, No. 1 November, 1974