Z. Wahrscheinlichkeitstheorie verw. Gebiete
44, 191–211 (1978)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete
© by Springer-Verlag 1978

On the Entropy Rate of Stationary Point Processes and Its Discrete Approximation

F. Papangelou

Manchester-Sheffield School of Probability and Statistics, University of Manchester
Manchester M 13 9 PL England

§1. Introduction

The entropy of a point process on a finite interval of the real line was introduced by Rudemo [15] and McFadden [10] and later generalised by Fritz [4, 6] to point processes defined on finite measure spaces. McFadden calculated, under certain differentiability conditions involving infinitesimal birth equations, the rate of change of the entropy as a function of time in terms of the "conditional birth rate" (i.e. the conditional intensity). A notion of "long-term entropy rate" for a stationary point process defined on the whole real line is implicit in Fritz's generalisation of McMillan's theorem [5]. As is to be expected this is the infimum (over all $T>0$) of $\frac{1}{T} H_T$, where H_T is the entropy of the point process over $[0, T]$.

In the present paper we develop a direct approach to the assymptotic entropy rate of a stationary point process on the real line as "conditional entropy given the past". This is done under the assumption that the Palm probability P_0 of the process is absolutely continuous with respect to the conventional probability P on the σ-field $\mathscr{F}_0$ of events occurring in $(-\infty, 0)$. If $\frac{dP_0}{dP}$ denotes the corresponding Radon-Nikodym density and λ the intensity of the point process then, as will be seen, it is natural to define the entropy rate as $\lambda - E(\Lambda_0 \log \Lambda_0)$, where $\Lambda_0 = \lambda \frac{dP_0}{dP}$. It turns out that this is equal to the product of λ and the entropy rate of the discrete-parameter process of interpoint distances under the Palm probability (Theorems 3, 3a).

The paper's main results concern the local approximation of the conditional information function. For each $\varepsilon>0$ and $n=0, \pm 1, \pm 2, \dots$ let the random variable ξ_n^ε be 1 or 0 according as the point process has points in the interval $[n\varepsilon, (n+1)\varepsilon)$ or not, and define

$$p_\varepsilon(i|\mathscr{F}_0) = P(\xi_0^\varepsilon = i|\mathscr{F}_0), \qquad p_\varepsilon(i|\xi_{-1}^\varepsilon, \xi_{-2}^\varepsilon, \dots) = P(\xi_0^\varepsilon = i|\xi_{-1}^\varepsilon, \xi_{-2}^\varepsilon, \dots)$$

0044-3719/78/0044/0191/$04.20

($i=0,1$). Then, by Theorem 1 (which summarises certain auxiliary results proved in [11]) $\frac{1}{\varepsilon}p_\varepsilon(1|\mathscr{F}_0)\to\Lambda_0$ in L_1-mean as $\varepsilon\to 0$, i.e. Λ_0 is the conditional intensity at 0 (and therefore $\Lambda_0\,dt$ can be interpreted as the conditional probability of a point in $(0,0+dt)$ given the past $\mathscr{F}_0$), and it is natural to seek to establish the behaviour of $\log p_\varepsilon(i|\mathscr{F}_0)$, $\log p_\varepsilon(i|\xi^\varepsilon_{-1},\xi^\varepsilon_{-2},\ldots)$ and other related functions as $\varepsilon\to 0$. This is done in Theorems 4, 5 and 6 and their corollaries, which are proved under a minimum of hypotheses. The justification and focal point of these results is Theorem 6 which asserts that

$$\frac{1}{\varepsilon}\log p_\varepsilon(\xi^\varepsilon_0|\xi^\varepsilon_{-1},\xi^\varepsilon_{-2},\ldots)=\frac{1}{\varepsilon}\xi^\varepsilon_0\left(\log\Lambda_{\tau_1}-\log\frac{1}{\varepsilon}\right)-\Lambda_0+o(1)$$

in L_1 as $\varepsilon\to 0$, where τ_1 is the first nonnegative point of the process and Λ_t denotes for each t the conditional intensity at t.

In the last section we show the connexion between the above results and the notion of entropy over a finite interval. Using the likelihood function of the given point process with respect to the Poisson process ([14, 9]) one can calculate the entropy and the resulting formula generalises the one given by McFadden under his more restrictive conditions. It becomes apparent that the above results are the natural extensions to the whole real line of corresponding statements valid for finite intervals. As a by-product we obtain a direct proof of McMillan's theorem for point processes ([5]).

The author was prompted to do this work by an informal but stimulating discussion he had with P.M. Lee and D. Vere-Jones. He is especially indebted to the former for drawing his attention to the early literature on the subject.

§2. Notation and Basic Facts

The framework adopted here is mostly that of [11], with minor notational changes. The probability space on which our point processes are defined will be identified with the space of realisations; thus, for a point process on the whole real line R, Ω will be taken as the space of all subsets of R which have $+\infty$ and $-\infty$ as their only accumulation points. For any $\omega\in\Omega$ and any Borel set $Q\subset R$, $N(\omega,Q)$ will denote the number of elements of $\omega\cap Q$. The basic σ-field $\mathscr{F}$ in Ω is defined as the minimal σ-field such that $N(\omega,Q)$ is an $\mathscr{F}$-measurable function of ω for every bounded Borel set Q. If in this definition we restrict Q to range over the bounded Borel subsets of $(-\infty,t)$, we obtain the σ-field $\mathscr{F}_t$ of events occurring "before time t" ($-\infty<t<\infty$).

A point process on R is a probability measure P on $(\Omega,\mathscr{F})$ such that the expectation $E(N(Q))$ is finite for every bounded Borel set Q. Stationarity of a point process will be understood in the usual sense, i.e. as the invariance of P under the action of any translation on the elements of Ω. All point processes considered in the present paper, which are defined on the whole real line, will be assumed to be stationary.

The notion of Palm probability P_t (i.e. of the conditional probability measure given that the process has a point at t) will be assumed known. See [16] for instance. The intensity of a stationary point process is defined as $\lambda=E(N[0,1])$. To prevent confusion we will denote expectations with respect to P_0 by E_0. We can and will throughout choose regular (i.e. σ-additive) versions of conditional probabilities.

We label the points of each $\omega\in\Omega$

$$\cdots<\tau_{-1}(\omega)<\tau_0(\omega)<\tau_1(\omega)<\cdots$$

in such a way that $\tau_0(\omega)<0\leqq\tau_1(\omega)$, and define

$$\zeta_i(\omega)=\tau_{i+1}(\omega)-\tau_i(\omega), \qquad \theta(\omega)=-\tau_0(\omega).$$

The palm probability P_0 of a stationary P is carried by the set $\Omega_0=\{\omega\in\Omega: 0\in\omega\}$ and for $\omega\in\Omega_0$ it is convenient to introduce

$$\eta_i(\omega)=\zeta_{i+1}(\omega),$$

especially when we consider the distribution of these random variables under P_0. It is well-known (see [16]) that under P_0 the sequence $\ldots,\eta_{-1},\eta_0,\eta_1,\ldots$ is stationary and that $E_0\eta_i=\frac{1}{\lambda}$.

Notice that $\mathscr{F}_0$ is generated by the random variables $\theta,\zeta_{-1},\zeta_{-2},\zeta_{-3},\ldots$. Below we shall denote by ζ the vector $(\ldots,\zeta_{-3},\zeta_{-2},\zeta_{-1})$, by η the vector $(\ldots,\eta_{-3},\eta_{-2},\eta_{-1})$ and by x the element $(\ldots,x_{-3},x_{-2},x_{-1})$ of the space $C=\cdots\times(0,\infty)\times(0,\infty)$. Product spaces such as C will be furnished with their Borel product σ-fields. Following (as in [11]) the practice of denoting measures by their corresponding differentials we have the following lemmas.

Lemma 1. *The probability measure $P(\theta\in dt,\zeta\in dx)$ on $(0,\infty)\times C$ is absolutely continuous with respect to $dt\,P_0(\eta\in dx)$ and*

$$\frac{P(\theta\in dt,\zeta\in dx)}{dt\,P_0(\eta\in dx)}=\lambda\,P_0(\eta_0\geqq t\,|\,\eta=x)$$

Lemma 2. *If $0\leqq a<b$ then for $P(\theta\in dt,\zeta\in dx)$-almost all $(t,x)\in(0,\infty)\times C$*

$$P(a\leqq\tau_1<b\,|\,\theta=t,\zeta=x)=\frac{P_0(t+a\leqq\eta_0<t+b\,|\,\eta=x)}{P_0(\eta_0\geqq t\,|\,\eta=x)}.$$

Lemma 1 is essentially Lemma 3 in [11], while Lemma 2 is implied by Lemma 4 there. It follows from these Lemmas, from Lebesgue's theorem on the almost everywhere differentiability of monotone functions and from the assumed regularity of conditional probabilities that

$$M(t,x)=\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}P(N[0,\varepsilon)\geqq 1\,|\,\theta=t,\zeta=x) \tag{1}$$

exists for $P(\theta\in dt,\zeta\in dx)$-almost all (t,x) in $(0,\infty)\times C$ and hence:

Lemma 3. $\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}P(N[0,\varepsilon)\geqq 1\,|\,\mathscr{F}_0)$ *exists P-almost surely.*

We let $\Lambda_t = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} P(N[t, t+\varepsilon) \geqq 1 | \mathscr{F}_t)$. The process Λ_t, $-\infty < t < \infty$, can (and will) be chosen to be measurable. Clearly

$$\Lambda_0 = M(\theta, \zeta) \tag{2}$$

and it is easy to prove that

$$\Lambda_{\tau_1} = M(\theta + \tau_1, \zeta). \tag{3}$$

The next two theorems will be needed in the sequel.

Theorem 1. *For a stationary point process the following statements are equivalent.*

(i) *For some (and then for all) $t \in R$, P_t is absolutely continuous with respect to P on $\mathscr{F}_t$.*

(ii) *For some (and then for all) $t \in R$, $\frac{1}{\varepsilon} P(N[t, t+\varepsilon) \geqq 1 | \mathscr{F}_t)$ converges to Λ_t in L_1-mean as $\varepsilon \to 0$.*

(iii *There is a non-negative, stationary and measurable stochastic process Y_t, $-\infty < t < \infty$, adapted to the σ-fields $\mathscr{F}_t$, $-\infty < t < \infty$, with $EY_0 < \infty$ and such that for any $a \in R$ the process*

$$N[a, t) - \int_a^t Y_s \, ds, \quad t \geqq a$$

is an $\{\mathscr{F}_t, t \geqq a\}$-martingale.

(iv) *With P_0-probability one the conditional distribution under P_0 of η_0, given η, is absolutely continuous with respect to the Lebesgue measure.*

If the above statements are true and if X_t denotes the Radon-Nikodym density $\frac{dP_t}{dP}$ of P_t with respect to P on the σ-field $\mathscr{F}_t$ $(-\infty < t < \infty)$, then for every $t \in R$

$$\Lambda_t = Y_t = \lambda X_t \quad \text{a.s.} \tag{4}$$

Note that $E\Lambda_t = \lambda$. The random variable Λ_t is the "conditional intensity of the point process at t, given the past".

For the proof of Theorem 1 see Theorems 9, 10 and 11 in [11]. Notice that statement (ii) is equivalent to the L_1-convergence of $\frac{1}{\varepsilon} E(N[t, t+\varepsilon) | \mathscr{F}_t)$ to Λ_t, since $\frac{1}{\varepsilon} E(N[t, t+\varepsilon) - I_{\{N[t, t+\varepsilon) \geqq 1\}}) \to 0$ as $\varepsilon \to 0$[1]. Statement (iii) is somewhat different from the one in [11, Th. 10] but the proof is straightforward. The reader who is familiar with Meyer's theorem on the decomposition of submartingales will recognise $\int_a^t Y_s \, ds$ as the "predictable projection" of the point process after a.

For future reference note that (iii) and (4) imply that if Z_t, $0 \leqq t \leqq T$, is a measurable stochastic process on $(\Omega, \mathscr{F}, P)$, adapted to the σ-fields $\mathscr{F}_t$, $0 \leqq t \leqq T$,

[1] I_A denotes the indicator function of A

and such that

$$\int_0^T E(\Lambda_t|Z_t|)\,dt<\infty \tag{5}$$

then $\sum_{0\leqq\tau_i\leqq T} Z_{\tau_i}$ is integrable with respect to P and

$$E\Big(\sum_{0\leqq\tau_i\leqq T} Z_{\tau_i}\Big)=\int_0^T E(\Lambda_t Z_t)\,dt.$$

This is a special case of a theorem in stochastic integration. See Proposition 2 on p. 89 of [2] (compare with Theorem 5.1 in [3, p. 444]). For the theorem as used here see [1] and [9].

If statements (i)–(iv) of Theorem 1 are true, we denote by $f(t|x)$ $(t>0, x\in C)$ the conditional probability density (under P_0) of η_0 given $\eta=x$, i.e.

$$f(t|x)=\frac{P_0(\eta_0\in dt\,|\,\eta=x)}{dt}. \tag{7}$$

Theorem 2. *If statements* (i)–(iv) *of Theorem* 1 *are true then*

$$P(\theta\in dt,\zeta\in dx)=\lambda\left(\int_t^\infty f(s|x)\,ds\right)dt\,P_0(\eta\in dx) \tag{8}$$

and for $P(\theta\in dt,\zeta\in dx)$*-almost every* $(t,x)\in(0,\infty)\times C$

$$P(\tau_1\in ds\,|\,\theta=t,\zeta=x)=\left(\int_t^\infty f(u|x)\,du\right)^{-1} f(t+s|x)\,ds \tag{9}$$

and

$$M(t,x)=f(t|x)\left(\int_t^\infty f(s|x)\,ds\right)^{-1}. \tag{10}$$

Proof. (8) and (9) follow from Lemma 2 and (7). Notice that (9) implies

$$P(N[0,\varepsilon)\geqq 1\,|\,\theta=t,\zeta=x)=\left(\int_t^{t+\varepsilon} f(s|x)\,ds\right)\left(\int_t^\infty f(s|x)\,ds\right)^{-1} \tag{11}$$

and (10) is a consequence of this and (1).

Another implication of (9) is that if $\phi(x,t,s)$ is a Borel function on $C\times(0,\infty)\times(0,\infty)$ such that the random variable $\phi(\zeta,\theta,\tau_1)$ is integrable, then

$$E(\phi(\zeta,\theta,\tau_1)\,|\,\theta=t,\zeta=x)=\int_0^\infty \phi(x,t,s)\frac{f(t+s|x)}{\int_t^\infty f(u|x)\,du}\,ds. \tag{12}$$

§3. Two Preliminary Theorems

Throughout the present paper log will denote natural logarithm. The absolute continuity of a measure μ with respect to another ν will from now on be denoted by $\mu \ll \nu$.

Theorem 3. *If $P_0 \ll P$ on $\mathscr{F}_0$, then*

$$E(\Lambda_0 \log \Lambda_0) = \lambda + \lambda E_0 \left[\int_0^\infty f(t|\eta) \log f(t|\eta)\, dt \right].$$

Proof. By (2), (10), Lemma 1 and (7)

$$\begin{aligned}
E(\Lambda_0 \log \Lambda_0) &= E(M(\theta, \zeta) \log M(\theta, \zeta)) \\
&= \int_0^\infty \int_C M(t, x) \log M(t, x)\, P(\theta \in dt, \zeta \in dx) \\
&= \int_C \int_0^\infty f(t|x) \left(\int_t^\infty f(s|x)\, ds \right)^{-1} \log \left[f(t|x) \left(\int_t^\infty f(s|x)\, ds \right)^{-1} \right] \\
&\quad \cdot \left(\lambda \int_t^\infty f(s|x)\, ds \right) dt\, P_0(\eta \in dx) \\
&= \lambda \int_C \int_0^\infty f(t|x) \log f(t|x)\, dt\, P_0(\eta \in dx) \\
&\quad - \lambda \int_C \int_0^\infty f(t|x) \log \left(\int_t^\infty f(s|x)\, ds \right) dt\, P_0(\eta \in dx).
\end{aligned}$$

By the change of variable $u = \int_t^\infty f(s|x)\, ds$ we obtain

$$\int_0^\infty f(t|x) \log \left(\int_t^\infty f(s|x)\, dx \right) dt = \int_0^1 \log u\, du = -1$$

which proves the theorem.

Of course both sides in Theorem 3 may be ∞. As usual we define $0 \log 0 = 0$. In connexion with the hypothesis of the next theorem recall that $y \log y \geqq -e^{-1}$ for all $y \geqq 0$.

Theorem 4. *If $P_0 \ll P$ on $\mathscr{F}_0$ and $E(\Lambda_0 \log \Lambda_0) < \infty$* then, as $\varepsilon \to 0$

$$\frac{1}{\varepsilon} E(I_{[0,\varepsilon)}(\tau_1) \log \Lambda_{\tau_1} | \mathscr{F}_0) \xrightarrow[\text{a.s.}]{L_1} \Lambda_0 \log \Lambda_0 \tag{13}$$

$$\frac{1}{\varepsilon} P(\tau_1 < \varepsilon | \mathscr{F}_0) \log \left(\frac{1}{\varepsilon} P(\tau_1 < \varepsilon | \mathscr{F}_0) \right) \xrightarrow[\text{a.s.}]{L_1} \Lambda_0 \log \Lambda_0 \tag{14}$$

$$\frac{1}{\varepsilon} E\left(I_{[0,\varepsilon)}(\tau_1) \left| \log \left(\frac{1}{\varepsilon} P(\tau_1 < \varepsilon | \mathscr{F}_0) \right) - \log \Lambda_{\tau_1} \right| | \mathscr{F}_0 \right) \xrightarrow[\text{a.s.}]{L_1} 0 \tag{15}$$

$$\frac{1}{\varepsilon} I_{[0,\varepsilon)}(\tau_1)\left(\log\left(\frac{1}{\varepsilon} P(\tau_1<\varepsilon|\mathscr{F}_0)\right)-\log \Lambda_{\tau_1}\right) \xrightarrow[\text{a.s.}]{L_1} 0 \tag{16}$$

$$\frac{1}{\varepsilon} P(\tau_1 \geqq \varepsilon|\mathscr{F}_0)\log P(\tau_1 \geqq \varepsilon|\mathscr{F}_0) \xrightarrow[\text{a.s.}]{L_1} -\Lambda_0 \tag{17}$$

$$\frac{1}{\varepsilon} I_{\{\tau_1 \geqq \varepsilon\}} \log P(\tau_1 \geqq \varepsilon|\mathscr{F}_0) \xrightarrow[\text{a.s.}]{L_1} -\Lambda_0. \tag{18}$$

Proof. In the proof we shall make use of the following well-known facts. Firstly, if V and V_n, $n=1, 2, \ldots$ are random variables such that $V_n \geqq c$ for some constant c and all n, $V_n \to V$ in probability and $\limsup_{n\to\infty} EV_n \leqq EV < \infty$, then $V_n \to V$ in L_1-mean. Secondly, if g is a Lebesgue integrable function on $[0, \infty)$ then, as $\varepsilon\to 0$, $\frac{1}{\varepsilon}\int_t^{t+\varepsilon} |g(s)-g(t)|\,ds \to 0$ for almost all $t \geqq 0$, as well as in L_1-mean.

Recall that

$$\frac{1}{\varepsilon} P(\tau_1<\varepsilon|\mathscr{F}_0) \xrightarrow[\text{a.s.}]{L_1} \Lambda_0 \tag{19}$$

and that, by Theorem 3

$$-\frac{1}{\lambda e}-1 \leqq \int_C \int_0^\infty f(t|x)\log f(t|x)\,dt\,P_0(\eta\in dx)<\infty. \tag{20}$$

By (3), (10) and (12) we have for $P(\theta\in dt, \zeta\in dx)$-almost all (t, x)

$$\begin{aligned}
&\frac{1}{\varepsilon} E(I_{[0,\varepsilon)}(\tau_1)\log \Lambda_{\tau_1}|\theta=t, \zeta=x)\\
&\quad=\frac{1}{\varepsilon}\int_t^{t+\varepsilon} \frac{f(u|x)}{\int_t^\infty f(s|x)\,ds}\log\frac{f(u|x)}{\int_t^\infty f(s|x)\,ds}\,du\\
&\quad=\left(\int_t^\infty f(s|x)\,ds\right)^{-1}\left[\frac{1}{\varepsilon}\int_t^{t+\varepsilon} f(u|x)\log f(u|x)\,du\right.\\
&\qquad\left.-(\log\int_t^\infty f(s|x)\,ds)\frac{1}{\varepsilon}\int_t^{t+\varepsilon} f(u|x)\,du\right]
\end{aligned} \tag{21}$$

which converges, for $P(\theta\in dt, \zeta\in dx)$-almost all (t, x) to

$$\begin{aligned}
&\left(\int_t^\infty f(s|x)\,ds\right)^{-1}\left[f(t|x)\log f(t|x)-f(t|x)\log\int_t^\infty f(s|x)\,ds\right]\\
&\quad=M(t, x)\log M(t, x)
\end{aligned}$$

by (10). This proves the a.s. convergence in (13). To prove L_1-convergence note that the left-hand side of (21) is $\geqq -e^{-1}$ (as follows from the middle term of (21)) while its expectation is, by (8) and the third term of (21),

$$\lambda \int_C \int_0^\infty \frac{1}{\varepsilon} \int_t^{t+\varepsilon} f(u|x) \log f(u|x)\, du\, dt\, P_0(\eta \in dx)$$
$$+\lambda \int_C \int_0^\infty \left(\frac{1}{\varepsilon} \int_t^{t+\varepsilon} f(s|x)\, ds\right) \log \left(\int_t^\infty f(s|x)\, ds\right)^{-1} dt\, P_0(\eta \in dx). \tag{22}$$

The second integral is less than or equal to

$$\lambda \int_C \int_0^\infty f_\varepsilon(t|x) \log \left(\int_t^\infty f_\varepsilon(s|x)\, ds\right)^{-1} dt\, P_0(\eta \in dx) \tag{23}$$

where we have set $f_\varepsilon(t|x) = \frac{1}{\varepsilon} \int_t^{t+\varepsilon} f(s|x)\, ds$. This inequality follows from the fact that $\int_t^\infty f_\varepsilon(s|x)\, ds \leqq \int_t^\infty f(s|x)\, ds$, which is in turn a simple consequence of Fubini's theorem. The integral (23) can be calculated by a change of variable (as in the proof of Theorem 3) and is seen to be equal to

$$\lambda \int_C (a_\varepsilon(x) - a_\varepsilon(x) \log a_\varepsilon(x))\, P_0(\eta \in dx) \tag{24}$$

where $a_\varepsilon(x) = \int_0^\infty f_\varepsilon(t|x)\, dt$. Note that $a_\varepsilon(x) \leqq 1$. If we let $\varepsilon \to 0$, then $a_\varepsilon(x) \to 1$ and (24) converges to λ. It follows that the $\limsup_{\varepsilon \to 0}$ of (22) is less than or equal to

$$\lambda \int_C \int_0^\infty f(t|x) \log f(t|x)\, dt\, P_0(\eta \in dx) + \lambda = E(\Lambda_0 \log \Lambda_0).$$

This and the remark at the beginning of the proof establish L_1-convergence in (13).

The proof of (14) is similar. The left-hand side is trivially $\geqq -e^{-1}$ and converges a.s. to $\Lambda_0 \log \Lambda_0$ by (19). Its expectation on the other hand is, by (11),

$$\int_C \int_0^\infty \left(\frac{1}{\varepsilon} \int_t^{t+\varepsilon} f(s|x)\, ds\right) \log \frac{\frac{1}{\varepsilon} \int_t^{t+\varepsilon} f(s|x)\, ds}{\int_t^\infty f(s|x)\, ds} \lambda\, dt\, P_0(\eta \in dx)$$
$$= \lambda \int_C \int_0^\infty \left(\frac{1}{\varepsilon} \int_t^{t+\varepsilon} f(s|x)\, ds\right) \log \left(\frac{1}{\varepsilon} \int_t^{t+\varepsilon} f(s|x)\, ds\right) dt\, P_0(\eta \in dx)$$
$$+ \lambda \int_C \int_0^\infty \left(\frac{1}{\varepsilon} \int_t^{t+\varepsilon} f(s|x)\, ds\right) \log \left(\int_t^\infty f(s|x)\, ds\right)^{-1} dt\, P_0(\eta \in dx)$$

which is, by Jensen's inequality, less than or equal to (22) and L_1-convergence follows as above.

Turning to (15) note that by the triangle inequality

$$\frac{1}{\varepsilon} E(I_{[0,\varepsilon)}(\tau_1)\left|\log\left(\frac{1}{\varepsilon}P(\tau_1<\varepsilon|\mathscr{F}_0)\right)-\log\Lambda_{\tau_1}\right||\mathscr{F}_0)$$
$$\leqq\frac{1}{\varepsilon}P(\tau_1<\varepsilon|\mathscr{F}_0)\left|\log\left(\frac{1}{\varepsilon}P(\tau_1<\varepsilon|\mathscr{F}_0)\right)\right|$$
$$+\frac{1}{\varepsilon}E(I_{[0,\varepsilon)}(\tau_1)|\log\Lambda_{\tau_1}||\mathscr{F}_0)$$

and both terms on the right are uniformly integrable by (13) and (14). On the other hand the left-hand side converges to 0 a.s. This follows from the above inequality if $\Lambda_0=0$, while if $\Lambda_0>0$ (i.e. if θ and ζ take values t, x respectively such that $f(t|x)>0$) it follows from the fact that, by (12),

$$E\left(\frac{1}{\varepsilon}I_{[0,\varepsilon)}(\tau_1)\left|\log\left(\frac{1}{\varepsilon}P(\tau_1<\varepsilon|\theta,\zeta)\right)-\log\Lambda_{\tau_1}\right||\theta=t,\zeta=x\right)$$
$$=\frac{1}{\varepsilon}\int_t^{t+\varepsilon}\left|\log\frac{\frac{1}{\varepsilon}\int_t^{t+\varepsilon}f(s|x)\,ds}{\int_t^\infty f(s|x)\,ds}-\log\frac{f(u|x)}{\int_u^\infty f(s|x)\,ds}\right|\frac{f(u|x)}{\int_t^\infty f(s|x)\,ds}\,du$$

where we can obviously assume $\int_t^\infty f(s|x)\,ds>0$. By the triangle inequality the right-hand side is less than or equal to the product of $\left(\int_t^\infty f(s|x)\,ds\right)^{-1}$ and

$$\left|\log\frac{\frac{1}{\varepsilon}\int_t^{t+\varepsilon}f(s|x)\,ds}{\int_t^\infty f(s|x)\,ds}-\log\frac{f(t|x)}{\int_t^\infty f(s|x)\,ds}\right|\cdot\frac{1}{\varepsilon}\int_t^{t+\varepsilon}f(u|x)\,du$$
$$+\left|\log\frac{f(t|x)}{\int_t^\infty f(s|x)\,ds}\right|\cdot\frac{1}{\varepsilon}\int_t^{t+\varepsilon}|f(u|x)-f(t|x)|\,du$$
$$+\frac{1}{\varepsilon}\int_t^{t+\varepsilon}\left|f(t|x)\log\frac{f(t|x)}{\int_t^\infty f(s|x)\,ds}-f(u|x)\log\frac{f(u|x)}{\int_u^\infty f(s|x)\,ds}\right|du.$$

Clearly all three terms converge to 0. This proves (15).

Almost sure convergence in (16) is trivial while L_1-convergence follows from (15). Almost sure convergence in (17) and (18) follows from (19) and the inequalities

$$-\frac{y}{1-y}\leqq\log(1-y)\leqq-y \tag{25}$$

valid for $0 \leqq y < 1$, while L_1-convergence follows from the fact that the left-hand sides of (17) and (18) are $\leqq 0$ and both have expectation $\frac{1}{\varepsilon} E(P(\tau_1 \geqq \varepsilon | \mathscr{F}_0) \log P(\tau_1 \geqq \varepsilon | \mathscr{F}_0))$ which by (25) is

$$\geqq \frac{1}{\varepsilon} E(-P(\tau_1 < \varepsilon | \mathscr{F}_0)) = -\frac{1}{\varepsilon} P(N[0, \varepsilon) \geqq 1) \to -\lambda = E(-\Lambda_0).$$

This completes the proof of the theorem.

Now define the random rariable ξ_0^ε to be 1 or 0 according as $N[0, \varepsilon) \geqq 1$ or $N[0, \varepsilon) = 0$ and for any σ-field $\mathscr{G} \subset \mathscr{F}$ define

$$p_\varepsilon(i | \mathscr{G}) = P(\xi_0^\varepsilon = i | \mathscr{G}) \qquad (i = 0, 1).$$

Corollary. $-\frac{1}{\varepsilon} \log p_\varepsilon(\xi_0^\varepsilon | \mathscr{F}_0) + \frac{1}{\varepsilon} \xi_0^\varepsilon \left(\log \Lambda_{\tau_1} - \log \frac{1}{\varepsilon} \right) \xrightarrow[\text{a.s.}]{L_1} \Lambda_0$.

Proof. The left-hand side is equal to

$$-\frac{1}{\varepsilon} \xi_0^\varepsilon \left(\log \left(\frac{1}{\varepsilon} p_\varepsilon(1 | \mathscr{F}_0) \right) - \log \Lambda_{\tau_1} \right) - \frac{1}{\varepsilon} (1 - \xi_0^\varepsilon) \log p_\varepsilon(0 | \mathscr{F}_0)$$

and the result follows from (16) and (18).

§4. The Entropy Rate

If ξ is a random variable or random vector defined on some probability space $(\Omega, \mathscr{F}, P)$ and taking only countably many distinct values $x_1, x_2, \ldots$ $\left(\text{so that } \sum_{i=1}^{\infty} P(\xi = x_i) = 1 \right)$, then the entropy $H(\xi)$ of ξ is defined by

$$H(\xi) = -\sum_{i=1}^{\infty} P(\xi = x_i) \log P(\xi = x_i). \tag{26}$$

If ξ is a non-negative random variable with finite expectation, whose distribution has a probability density function $f(t)$, $t \geqq 0$, then its entropy is defined by

$$H(\xi) = -\int_0^\infty f(t) \log f(t)\, dt, \tag{27}$$

and as is well-known $-\infty \leqq H(\xi) \leqq 1 + \log E\xi$. Rényi [13] has given an interpretation of (27) in terms of (26). He proved that if (27) is finite and if for each $r = 1, 2, \ldots$ we define the random variable ξ_r by setting $\xi_r = \frac{k}{r}$ on $\left\{ \frac{k}{r} \leqq \xi < \frac{k+1}{r} \right\}$, $k = 0, 1, \ldots$, then

$$H(\xi_r) = \log r + H(\xi) + o(1) \qquad \text{as } r \to \infty. \tag{28}$$

A similar statement is true for N-dimensional vectors $\bar{\xi}$ with non-negative coordinates having finite means except that the dimension N enters (28) as follows

$$H(\bar{\xi}_r) = N \log r + H(\bar{\xi}) + o(1) \quad \text{as } r \to \infty. \tag{29}$$

These facts form the basis for the definition of the entropy of a (not necessarily stationary) point process on a finite interval, given independently by Rudemo [15] and McFadden [10] and later extended by Fritz [4] to point processes on finite measure spaces. (It is clear how the supporting probability spaces for such point processes can be introduced along the lines of §2). To be specific let $\tau_1 < \tau_2 < \cdots < \tau_N$ be the points of such a point process on the interval $[0, T]$ (where $N = N[0, T]$ is of course random) and assume that $EN < \infty$ and that the conditional distribution of $\tau = (\tau_1, \tau_2, \ldots, \tau_N)$ given $N = k$ is absolutely continuous for any $k \geqq 1$. Rudemo and McFadden defined the entropy as

$$-\sum_{k=0}^{\infty} P(N=k) \log P(N=k)$$
$$+\sum_{k=0}^{\infty} P(N=k) \cdot (\text{entropy of } \tau \text{ given } N=k).$$

Fritz proved that this entropy is equal to the generalised entropy

$$E_\nu \left(-\log \frac{d\nu}{d\mu} \right) \tag{30}$$

of the probability measure ν (in the space of realisations) of the given point process with respect to the measure $\mu = e^T \Pi$, where Π is the probability measure of the standard Poisson process on $[0, T]$. (E_ν denotes expectations with respect to ν). In fact [4] and [6] deal with point processes which are defined on any atomless finite measure space $(X, \mathfrak{A}, m)$ and satisfy $E(N(X)) < \infty$. It is shown in [6] that if for every partition $\Delta = \{G_1, G_2, \ldots, G_n\}$ of X with $m(G_i) > 0$ ($i = 1, 2, \ldots, n$) we set

$$\xi_\Delta = (N(G_1), \ldots, N(G_n))$$

and

$$\bar{\xi}_\Delta = (I_{\{N(G_1) \geqq 1\}}, \ldots, I_{\{N(G_n) \geqq 1\}}) \tag{31}$$

then

$$E_\nu \left(-\log \frac{d\nu}{d\mu} \right) = \inf_\Delta \{ H(\xi_\Delta) + E(N(X)) \cdot \log(\max_{1 \leqq i \leqq n} m(G_i)) \} \tag{32}$$
$$= \inf_\Delta \left\{ H(\bar{\xi}_\Delta) + \sum_{i=1}^{n} P(N(G_i) \geqq 1) \log m(G_i) \right\}.$$

(In this case ν is again the probability measure of the given point process and $\mu = e^{m(X)} \Pi$, where Π is the Poisson process on $(X, \mathfrak{A}, m)$. Note that if $\nu \not\ll \mu$ then all three expressions are equal to $-\infty$).

In the present paper we are concerned with the entropy "rate" of a stationary point process on the real line. Recall that if $\ldots, \xi_{-1}, \xi_0, \xi_1, \ldots$ is a stationary stochastic process such that each ξ_n takes only finitely many values then, as $n \to \infty$, $\frac{1}{n} H(\xi_0, \xi_1, \ldots, \xi_{n-1})$ converges to

$$\begin{aligned} &H(\xi_0 | \xi_{-1}, \xi_{-2}, \ldots) \\ &\quad = E[-\sum_x P(\xi_0 = x | \xi_{-1}, \xi_{-2}, \ldots) \log P(\xi_0 = x | \xi_{-1}, \xi_{-2}, \ldots)] \end{aligned} \tag{33}$$

and that if we let

$$p_n(x_0, \ldots, x_{n-1}) = P(\xi_0 = x_0, \ldots, \xi_{n-1} = x_{n-1}),$$

then

$$-\frac{1}{n} \log p_n(\xi_0, \ldots, \xi_{n-1})$$

converges a.s. and in L_1-mean to a random variable whose expectation is equal to (33).

For the rest of the paper we shall assume that P is a stationary point process on $(-\infty, \infty)$ with intensity λ and satisfying the following hypotheses.

(I) $P_0 \ll P$ on $\mathscr{F}_0$. $\left(\text{As in §2 we set } \Lambda_0 = \lambda \frac{dP_0}{dP}\right)$.

(II) $E(\Lambda_0 \log \Lambda_0) < \infty$.

Π will denote the standard Poisson process on $(-\infty, \infty)$ and Π_0 its Palm probability. E will always denote expectations with respect to P.

Employing the notation of §2 we see from Theorem 3 that $E_0\left[-\int_0^\infty f(t|\eta) \log f(t|\eta)\, dt\right]$ exists and is finite. We denote this expectation by $H(\eta_0 | \eta_{-1}, \eta_{-2}, \ldots)$ or $H(\eta_0 | \eta)$, since it is the analogue of (33) for the discrete-parameter process

$$\ldots, \eta_{-1}, \eta_0, \eta_1, \ldots \tag{34}$$

In fact we are in exactly the sort of situation investigated by Perez [12], since the distribution of (34) under Π_0 is the product measure $\cdots \otimes \pi \otimes \pi \otimes \cdots$ in $\cdots \times (0, \infty) \times (0, \infty) \times \cdots$, where π is the measure in $(0, \infty)$ with density e^{-t}, $t \geq 0$, with respect to the Lebesgue measure. Using Perez' generalisation of McMillan's theorem we see that $-\frac{1}{n} \log \frac{dP_0}{d\Pi_0}(\eta_0, \eta_1, \ldots, \eta_{n-1})$ converges in L_1-mean to a random variable with expectation

$$\begin{aligned} &E_0\left[-\int_0^\infty \frac{P_0(\eta_0 \in dt | \eta)}{\pi_0(dt)} \log \frac{P_0(\eta_0 \in dt | \eta)}{\pi_0(dt)} \pi_0(dt)\right] \\ &\quad = H(\eta_0 | \eta_{-1}, \eta_{-2}, \ldots) - E(\eta_0). \end{aligned} \tag{35}$$

$\left(\frac{dP_0}{d\Pi_0}(\eta_0, \ldots, \eta_{n-1})\right.$ denotes the Radon-Nikodym density on the σ-field $\left.\mathscr{F}(\eta_0, \eta_1, \ldots, \eta_{n-1}) \text{ generated by } \eta_0, \eta_1, \ldots, \eta_{n-1}\right)$. We call $H(\eta_0|\eta_{-1}, \eta_{-2}, \ldots)$ the *entropy rate* of (34) under P_0. Note that if we replace the measure Π_0, restricted to $\mathscr{F}(\eta_0, \eta_1, \ldots, \eta_{n-1})$, by the measure $e^{-\frac{n}{\lambda}}\Pi_0$ we can eliminate the term $E(\eta_0)$ in the right-hand side of (35).

Let us return to the point process under P. We treat such a process as the message from some source but assume that for technical reasons the receiver has only ε-accuracy, i.e. cannot distinguish between two points arriving within time ε of each other. To simplify matters further we subdivide the real line into the intervals $[n\varepsilon, (n+1)\varepsilon)$, $n=0, \pm 1, \pm 2, \ldots$ and assume that the receiver can only distinguish between the events $\{N[n\varepsilon, (n+1)\varepsilon) \geqq 1\}$ and $\{N[n\varepsilon, (n+1)\varepsilon)=0\}$. The message then becomes discrete with a two-letter alphabet and Shannon's classical theory applies. Let $\xi_n^\varepsilon = I_{\{N[n\varepsilon, (n+1)\varepsilon) \geqq 1\}}$, $n=0, \pm 1, \pm 2, \ldots$ (cf. (31)) and set $p_\varepsilon(i|\xi_{-1}^\varepsilon, \xi_{-2}^\varepsilon, \ldots) = p_\varepsilon(i|\mathscr{F}_0^\varepsilon)$ $(i=0, 1)$, where $\mathscr{F}_0^\varepsilon$ is the σ-field generated by $\xi_{-1}^\varepsilon, \xi_{-2}^\varepsilon, \ldots$.

Lemma 4. *If Y and Y_ε, $\varepsilon>0$, are integrable random variables such that $Y_\varepsilon \to Y$ in L_1-mean as $\varepsilon \to 0$, then $E(Y_\varepsilon|\xi_{-1}^\varepsilon, \xi_{-2}^\varepsilon, \ldots) \to E(Y|\mathscr{F}_0)$ in L_1-means as $\varepsilon \to 0$.*

Proof. If $A \in \mathscr{F}_0$ and $\delta>0$ then there is $\varepsilon_0>0$ such that for every ε with $0<\varepsilon \leqq \varepsilon_0$ there exists $A_\varepsilon \in \mathscr{F}_0^\varepsilon$ satisfying $P(A \triangle A_\varepsilon)<\delta$. If for instance A is of the form $\{N(J)<c\}$ for some interval $J \subset (-\infty, 0)$ and if we take $A_\varepsilon = \{N^\varepsilon(J)<c\}$ where $N^\varepsilon(J)$ is the number of intervals $[n\varepsilon, (n+1)\varepsilon) \subset J$ such that $\xi_n^\varepsilon = 1$, then $P(A \triangle A_\varepsilon) \leqq P(N(J) \neq N^\varepsilon(J)) \to 0$ as $\varepsilon \to 0$, and one can use standard arguments to extend this to arbitrary $A \in \mathscr{F}_0$. Now $P(A \triangle A_\varepsilon)<\delta$ implies $E|P(A|\mathscr{F}_0^\varepsilon) - I_A| < 2\delta$ and it follows easily that if Y is an integrable random variable, then $E(Y|\mathscr{F}_0^\varepsilon) \to E(Y|\mathscr{F}_0)$ in L_1-mean as $\varepsilon \to 0$. The lemma follows from this and the triangle inequality.

Lemma 5. *Let Y, Y_n $(n=1, 2, \ldots)$ be non-negative integrable random variables and $\mathscr{G}_n$ $(n=1, 2, \ldots)$ σ-fields of events in a probability space, and set $Z_n = E(Y_n|\mathscr{G}_n)$. If $Y_n \to Y$, $Z_n \to Y$ and $Y_n \log Y_n \to Y \log Y$ in L_1-mean, then $Z_n \log Z_n \to Y \log Y$ and $Y_n \log Z_n \to Y \log Y$ in L_1-mean.*

Proof. We can easily prove that $Z_n \log Z_n$ and $Y_n \log Z_n$ converge to $Y \log Y$ in probability by considering a.s. convergent subsequences of $\{Y_n\}$ and $\{Z_n\}$. Now $Z_n \log Z_n \geqq -e^{-1}$ while by Jensen's inequality

$$\begin{aligned} E(Z_n \log Z_n) &= E(E(Y_n|\mathscr{G}_n) \log E(Y_n|\mathscr{G}_n)) \\ &\leqq E(E(Y_n \log Y_n|\mathscr{G}_n)) = E(Y_n \log Y_n) \end{aligned} \tag{36}$$

which converges to $E(Y \log Y)$. It follows from the remark made at the beginning of the proof of Theorem 4 that $Z_n \log Z_n \to Y \log Y$ in L_1-mean. The L_1-convergence of $Y_n \log Z_n$ will follow if we show that $(Y_n - Z_n) \log Z_n \to 0$ in L_1-mean. To this end note that, for any $c>0$, $\{|\log Z_n|>c\} \in \mathscr{G}_n$ hence

$$E(|Y_n - Z_n|\,|\log Z_n|) \leqq c\,E|Y_n - Z_n| + \int_{\{|\log Z_n| > c\}} (Y_n + Z_n)\,|\log Z_n|\,dP$$
$$= c\,E|Y_n - Z_n| + 2 \int_{\{|\log Z_n| > c\}} Z_n |\log Z_n|\,dP$$
$$= c\,E|Y_n - Z_n| + \int_{\{Z_n < e^{-c}\}} |Z_n \log Z_n|\,dP + \int_{\{Z_n \log Z_n > c\}} Z_n \log Z_n\,dP$$

since $\{\log Z_n > c\} \subset \{Z_n \log Z_n > c\}$. The uniform integrability of $Z_n \log Z_n$ ($n = 1, 2, \ldots$) now implies that we can make the right-hand side arbitrarily small by first choosing c sufficiently large and then n sufficiently large.

Theorem 5. *Under* (I) *and* (II), *as* $\varepsilon \to 0$

$$\frac{1}{\varepsilon} p_\varepsilon(1|\xi^\varepsilon_{-1}, \ldots) \xrightarrow{L_1} \Lambda_0 \tag{37}$$

$$\frac{1}{\varepsilon} p_\varepsilon(1|\xi^\varepsilon_{-1}, \ldots) \log\left(\frac{1}{\varepsilon} p_\varepsilon(1|\xi^\varepsilon_{-1}, \ldots)\right) \xrightarrow{L_1} \Lambda_0 \log \Lambda_0 \tag{38}$$

$$\frac{1}{\varepsilon} p_\varepsilon(0|\xi^\varepsilon_{-1}, \ldots) \log p_\varepsilon(0|\xi^\varepsilon_{-1}, \ldots) \xrightarrow{L_1} -\Lambda_0 \tag{39}$$

$$\frac{1}{\varepsilon}(1 - \xi^\varepsilon_0) \log p_\varepsilon(0|\xi^\varepsilon_{-1}, \ldots) \xrightarrow{L_1} -\Lambda_0 \tag{40}$$

Proof. (37) follows from (19) and Lemma 4 if we take $Y = \Lambda_0$, $Y_\varepsilon = \frac{1}{\varepsilon} P(\tau_1 < \varepsilon|\mathscr{F}_0)$. (38) follows from (14) and Lemma 5, while (39) follows from (37) and (25). Again (37) and (25) imply that there is convergence in probability in (40). On the other hand the left-hand side of (40) is $\leqq 0$, while its expectation is

$$\frac{1}{\varepsilon} E[E((1 - \xi^\varepsilon{}_0) \log p_\varepsilon(0|\mathscr{F}^\varepsilon_0)|\mathscr{F}^\varepsilon_0)]$$
$$= \frac{1}{\varepsilon} E(E(1 - \xi^\varepsilon_0|\mathscr{F}^\varepsilon_0) \log p_\varepsilon(0|\mathscr{F}^\varepsilon_0))$$
$$= \frac{1}{\varepsilon} E(p_\varepsilon(0|\mathscr{F}^\varepsilon_0) \log p_\varepsilon(0|\mathscr{F}^\varepsilon_0))$$

which converges to $E(-\Lambda_0)$ by (39).

Theorem 6. $$-\frac{1}{\varepsilon} \log p_\varepsilon(\xi^\varepsilon_0|\xi^\varepsilon_{-1}, \ldots) + \frac{1}{\varepsilon} \xi^\varepsilon_0 \left(\log \Lambda_{\tau_1} - \log \frac{1}{\varepsilon}\right) \xrightarrow{L_1} \Lambda_0. \tag{41}$$

Proof. The left-hand side is

$$-\frac{1}{\varepsilon} \xi^\varepsilon_0 \left(\log\left(\frac{1}{\varepsilon} p_\varepsilon(1|\mathscr{F}^\varepsilon_0)\right) - \log \Lambda_{\tau_1}\right) - \frac{1}{\varepsilon}(1 - \xi^\varepsilon_0) \log p_\varepsilon(0|\mathscr{F}^\varepsilon_0).$$

By the corollary to Theorem 4 and (40) it is sufficient to show that

$$\frac{1}{\varepsilon} \xi^\varepsilon_0 \log\left(\frac{1}{\varepsilon} p_\varepsilon(1|\mathscr{F}_0)\right) - \frac{1}{\varepsilon} \xi^\varepsilon_0 \log\left(\frac{1}{\varepsilon} p_\varepsilon(1|\mathscr{F}^\varepsilon_0)\right) \xrightarrow{L_1} 0.$$

But clearly

$$E\left|\frac{1}{\varepsilon}\xi_0^\varepsilon\left(\log\left(\frac{1}{\varepsilon}p_\varepsilon(1|\mathscr{F}_0)\right)-\log\left(\frac{1}{\varepsilon}p_\varepsilon(1|\mathscr{F}_0^\varepsilon)\right)\right)\right|$$

$$=E\left[E\left(\frac{1}{\varepsilon}\xi_0^\varepsilon\left|\log\left(\frac{1}{\varepsilon}p_\varepsilon(1|\mathscr{F}_0)\right)-\log\left(\frac{1}{\varepsilon}p_\varepsilon(1|\mathscr{F}_0^\varepsilon)\right)\right|\,|\mathscr{F}_0\right)\right]$$

$$=E\left[\frac{1}{\varepsilon}p_\varepsilon(1|\mathscr{F}_0)\cdot\left|\log\left(\frac{1}{\varepsilon}p_\varepsilon(1|\mathscr{F}_0)\right)-\log\left(\frac{1}{\varepsilon}p_\varepsilon(1|\mathscr{F}_0^\varepsilon)\right)\right|\right]$$

$$=E\left|\frac{1}{\varepsilon}p_\varepsilon(1|\mathscr{F}_0)\log\left(\frac{1}{\varepsilon}p_\varepsilon(1|\mathscr{F}_0)\right)-\frac{1}{\varepsilon}p_\varepsilon(1|\mathscr{F}_0)\log\left(\frac{1}{\varepsilon}p_\varepsilon(1|\mathscr{F}_0^\varepsilon)\right)\right|$$

and the result follows from Lemma 5.

We now derive a number of corollaries from these theorems. Let

$$\chi(\xi_0^\varepsilon|\xi_{-1}^\varepsilon,\xi_{-2}^\varepsilon,\ldots)=-\sum_{i=0}^{1}p_\varepsilon(i|\xi_{-1}^\varepsilon,\xi_{-2}^\varepsilon,\ldots)\log p_\varepsilon(i|\xi_{-1}^\varepsilon,\xi_{-2}^\varepsilon,\ldots).$$

Clearly $E(\chi(\xi_0^\varepsilon|\xi_{-1}^\varepsilon,\ldots))=H(\xi_0^\varepsilon|\xi_{-1}^\varepsilon,\ldots)$.

Corollary 1. *Under* (I) *and* (II), *as* $\varepsilon\to 0$

$$\frac{1}{\varepsilon}\chi(\xi_0^\varepsilon|\xi_{-1}^\varepsilon,\ldots)-\left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right)p_\varepsilon(1|\xi_{-1}^\varepsilon,\ldots)\xrightarrow{L_1}\Lambda_0-\Lambda_0\log\Lambda_0.$$

This follows from (37) and (39).

Corollary 2. *Let* $n(\varepsilon)$ *be the greatest integer such that* $n(\varepsilon)\,\varepsilon\leqq 1$. *Under* (I) *and* (II), *as* $\varepsilon\to 0$.

$$\sum_{k=0}^{n(\varepsilon)-1}\chi(\xi_k^\varepsilon|\xi_{k-1}^\varepsilon,\ldots)-\left(\log\frac{1}{\varepsilon}\right)\sum_{k=0}^{n(\varepsilon)-1}P(\xi_k^\varepsilon=1|\xi_{k-1}^\varepsilon,\ldots)$$
$$\xrightarrow{L_1}\int_0^1(\Lambda_t-\Lambda_t\log\Lambda_t)\,dt.$$

In fact, by stationarity, the expectation of the absolute value of the difference between the two sides is

$$n(\varepsilon)\,\varepsilon E\left|\frac{1}{\varepsilon}\chi(\xi_0^\varepsilon|\xi_{-1}^\varepsilon,\ldots)-\left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right)P(\xi_0^\varepsilon=1|\xi_{-1}^\varepsilon,\ldots)\right.$$
$$\left.-\frac{1}{\varepsilon}\int_0^\varepsilon(\Lambda_t-\Lambda_t\log\Lambda_t)\,dt\right|+o(1)$$

and this converges to 0 by Corollary 1.

Corollary 3. *Under* (I) *and* (II), *as* $\varepsilon\to 0$

$$\frac{1}{\varepsilon}H(\xi_0^\varepsilon|\xi_{-1}^\varepsilon,\xi_{-2}^\varepsilon,\ldots)=\left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right)P(N[0,\varepsilon)\geqq 1)+\lambda-E(\Lambda_0\log\Lambda_0)+o(1).$$

From this we see that if $n(\varepsilon)$ is as in Corollary 2, then

$$\begin{aligned}H(\xi_0^\varepsilon, \xi_1^\varepsilon, \ldots, \xi_{n(\varepsilon)-1}^\varepsilon \mid \xi_{-1}^\varepsilon, \xi_{-2}^\varepsilon, \ldots) \\ = n(\varepsilon)\, H(\xi_0^\varepsilon \mid \xi_{-1}^\varepsilon, \xi_{-2}^\varepsilon, \ldots) \\ = \left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right) P(N[0, \varepsilon) \geqq 1) + \lambda - E(\Lambda_0 \log \Lambda_0) + o(1)\end{aligned} \tag{42}$$

as $\varepsilon \to 0$. We comment on the connexion between this and (32) at the end of the paper.

Corollary 4. *If in addition to* (I) *and* (II) *we assume*

$$\text{(III)} \quad \left(\int_0^\varepsilon E_0[f(t \mid \eta)]\, dt\right) \log \varepsilon \to 0 \qquad \text{as } \varepsilon \to 0$$

then

$$\frac{1}{\varepsilon} H(\xi_0^\varepsilon \mid \xi_{-1}^\varepsilon, \xi_{-2}^\varepsilon, \ldots) = \lambda \log \frac{1}{\varepsilon} + \lambda - E(\Lambda_0 \log \Lambda_0) + o(1).$$

To prove this we make use of the following lemma.

Lemma 6. $P(N[0, \varepsilon) \geqq 1) = \lambda\varepsilon - \lambda E_0\left(\int_0^\varepsilon (\varepsilon - t)\, f(t \mid \eta)\, dt\right)$.

Proof. By (8) and (11)

$$\begin{aligned}P(N[0, \varepsilon) \geqq 1) &= \int_C \int_0^\infty P(N[0, \varepsilon) \geqq 1 \mid \theta = t, \zeta = x)\, P(\theta \in dt, \zeta \in dx) \\ &= \int_C \int_0^\infty \int_t^{t+\varepsilon} f(s \mid x)\, ds\, \lambda\, dt\, P_0(\eta \in dx) \\ &= \lambda E_0\left(\int_0^\infty \int_t^{t+\varepsilon} f(s \mid \eta)\, ds\, dt\right).\end{aligned}$$

By Fubini's theorem this is equal to

$$\begin{aligned}&\lambda E_0\left(\int_0^\infty \left(\int_{\max(s-\varepsilon, 0)}^{s} dt\right) f(s \mid \eta)\, ds\right) \\ &= \lambda E_0\left(\int_0^\varepsilon s f(s \mid \eta)\, ds + \int_\varepsilon^\infty \varepsilon f(s \mid \eta)\, ds\right) \\ &= \lambda E_0\left(\int_0^\varepsilon s f(s \mid \eta)\, ds + \varepsilon\left[1 - \int_0^\varepsilon f(s \mid \eta)\, ds\right]\right)\end{aligned}$$

which implies the lemma.

This shows that

$$\left|\lambda \log\frac{1}{\varepsilon} - \frac{1}{\varepsilon} P(N[0,\varepsilon) \geqq 1) \log\frac{1}{\varepsilon}\right|$$
$$= \frac{\lambda}{\varepsilon} E_0\left(\int_0^{\varepsilon} (\varepsilon - t) f(t|\eta)\, dt\right) |\log\varepsilon|$$
$$\leqq \lambda E_0\left(\int_0^{\varepsilon} f(t|\eta)\, dt\right) |\log\varepsilon|$$

which, combined with Corollary 3 and (III) establishes Corollary 4.

Condition (III) is relatively mild and ensures that $f(t|\eta)$ does not "explode" too fast as $t \to 0$. It is satisfied for instance if $E_0(\sup_{0 \leqq t \leqq \varepsilon} f(t|\eta)) < \infty$ for some $\varepsilon > 0$. The function

$$f(t|x) = \begin{cases} t^{-1} (\log t)^{-2} & \text{if } 0 < t \leqq e^{-1} \\ 0 & \text{if } t > e^{-1} \end{cases}$$

is an example (renewal process) for which the conclusion of Corollary 4 is not true.

A glance at (29) justifies the following definition.

Definition. We call $\lambda - E(\Lambda_0 \log\Lambda_0) = E(\Lambda_0 - \Lambda_0 \log\Lambda_0)$ the *entropy rate* of the given point process and denote it by H.

We can now reformulate Theorem 3 as follows.

Theorem 3a. *If $P_0 \ll P$ on $\mathscr{F}_0$, then*

$$H = \lambda H(\eta_0 | \eta_{-1}, \eta_{-2}, \ldots)$$

where the conditional entropy on the right is taken with respect to the Palm probability P_0.

By Fubini's theorem and Jensen's inequality

$$H = -\lambda \int_0^{\infty} E_0(f(t|\eta) \log f(t|\eta))\, dt$$
$$\leqq -\lambda \int_0^{\infty} E_0(f(t|\eta)) \log E_0(f(t|\eta))\, dt.$$

Setting $F(t) = E_0(f(t|\eta))$ we have

Corollary 5. $H \leqq -\lambda \int_0^{\infty} F(t) \log F(t)\, dt$. *If* $\int_0^{\infty} F(t) \log F(t)\, dt < \infty$, *then there is equality if and only if for almost every* $t \geqq 0$ $f(t|\eta) = F(t)\, P_0$-a.s., i.e. *if and only if the point process is an equilibrium renewal process.*

Thus amongst the point processes with a given $F(t)$ the entropy rate is maximal for the renewal process. It is easy to prove that the entropy rate of the superposition of two independent point processes is less than or equal to the sum of their entropy rates.

Note that Corollaries 1–4 followed from Theorem 5. We now turn to Theorem 6 and look at one important consequence. First define

$$p_\varepsilon(i_0, i_1, \ldots, i_n | \xi^\varepsilon_{-1}, \xi^\varepsilon_{-2}, \ldots) = P(\xi^\varepsilon_0 = i_0, \ldots, \xi^\varepsilon_n = i_n | \xi^\varepsilon_{-1}, \xi^\varepsilon_{-2}, \ldots).$$

Corollary 6. *Let $n(\varepsilon)$ be the greatest integer such that $n(\varepsilon)\,\varepsilon \leqq 1$. Under* (I) *and* (II), *as $\varepsilon \to 0$*

$$-\log p_\varepsilon(\xi^\varepsilon_0, \ldots, \xi^\varepsilon_{n(\varepsilon)-1} | \xi^\varepsilon_{-1}, \xi^\varepsilon_{-2}, \ldots) - \left(\log \frac{1}{\varepsilon}\right) \sum_{k=0}^{n(\varepsilon)-1} \xi^\varepsilon_k$$

$$\xrightarrow{L_1} \int_0^1 \Lambda_t\, dt - \sum_{0 \leqq \tau_i \leqq 1} \log \Lambda_{\tau_i}.$$

Proof. Since $p_\varepsilon(\xi^\varepsilon_0, \ldots, \xi^\varepsilon_n | \xi^\varepsilon_{-1}, \ldots) = \prod_{k=0}^{n} p_\varepsilon(\xi^\varepsilon_k | \xi^\varepsilon_{k-1}, \ldots)$ we deduce by stationarity that the expectation of the absolute value of the difference between the two sides is less than or equal to

$$n(\varepsilon)\,\varepsilon\, E\left| -\frac{1}{\varepsilon} \log p_\varepsilon(\xi^\varepsilon_0 | \xi^\varepsilon_{-1}, \ldots) - \left(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon}\right) \xi^\varepsilon_0 \right.$$

$$\left. -\frac{1}{\varepsilon} \int_0^\varepsilon \Lambda_t\, dt + \frac{1}{\varepsilon} \sum_{0 \leqq \tau_i < \varepsilon} \log \Lambda_{\tau_i} \right| + o(1)$$

and the result will follow from Theorem 6 if we show that

$$\frac{1}{\varepsilon} \sum_{0 \leqq \tau_i < \varepsilon} \log \Lambda_{\tau_i} - \frac{1}{\varepsilon} \xi^\varepsilon_0 \log \Lambda_{\tau_1} \xrightarrow{L_1} 0.$$

Again by stationarity

$$E\left| \frac{1}{\varepsilon} \sum_{0 \leqq \tau_i < \varepsilon} \log \Lambda_{\tau_i} - \frac{1}{\varepsilon} \xi^\varepsilon_0 \log \Lambda_{\tau_1} \right|$$

$$\leqq \frac{1}{n(\varepsilon)\,\varepsilon}\, n(\varepsilon)\, E\left(I_{\{N[0,\varepsilon) \geqq 2\}} \sum_{0 \leqq \tau_i < \varepsilon} |\log \Lambda_{\tau_i}| \right)$$

$$\leqq \frac{1}{1-\varepsilon} E\left(\sum_{k=0}^{n(\varepsilon)-1} I_{\{N[k\varepsilon, (k+1)\varepsilon) \leqq 2\}} \sum_{k\varepsilon \leqq \tau_i < (k+1)\varepsilon} |\log \Lambda_{\tau_i}| \right)$$

and the last expectation converges to 0 since the integrand converges to 0 a.s. and is dominated by $\sum_{0 \leqq \tau_i \leqq 1} |\log \Lambda_{\tau_i}|$ which is integrable by (6).

§5. Connexion with the Likelihood Function

Let us now consider the entropy of our stationary point process over a finite interval, say $[0, T]$. We assume hypotheses (I) and (II). By Theorem 1, for any

$a \in R$ and any $t > a$

$$\frac{1}{\varepsilon} P(N[t, t+a) \geqq 1 | \mathscr{F}_{[a,t)}) \xrightarrow{L_1} E(\Lambda_t | \mathscr{F}_{[a,t)}). \tag{43}$$

Set $\Lambda_{[a,t)} = E(\Lambda_t | \mathscr{F}_{[a,t)})$. It is known that the existence of the above limit in L_1 for $a=0$ and every $t \in (0, T]$ implies $P \ll \Pi$ on $\mathscr{F}_{[0,T]}$ and that the corresponding Radon-Nikodym density (likelihood function) is

$$h_T = \prod_{0 \leqq \tau_i \leqq T} \Lambda_{[0,\tau_i)} \cdot \exp\left[\int_0^T (1 - \Lambda_{[0,t)})\, dt\right].$$

(See [14, 17, 1, 9]; cf. also [8] for the case of marked point processes.) If instead of the restriction of Π to $\mathscr{F}_{[0,T]}$ we consider as in (30) the measure $e^T \Pi$, we obtain the modified density $\bar{h}_T = e^{-T} h_T$, and the corresponding information function is

$$-\log \bar{h}_T = \int_0^T \Lambda_{[0,t)} dt - \sum_{0 \leqq \tau_i \leqq T} \log \Lambda_{[0,\tau_i)}. \tag{44}$$

If we denote by H_T the entropy of the point process in $[0, T]$ as defined by Fritz, then

$$H_T = E(-\log \bar{h}_T) = \int_0^T E \Lambda_{[0,t)} dt - E\Big(\sum_{0 \leqq \tau_i \leqq T} \log \Lambda_{[0,\tau_i)}\Big). \tag{45}$$

One can prove as in (36) that $E(\Lambda_{[0,t)} \log \Lambda_{[0,t)}) \leqq E(\Lambda_t \log \Lambda_t) < \infty$. Since $E(\Lambda_t \log \Lambda_{[0,t)}) = \mathrm{E}(\Lambda_{[0,t)} \log \Lambda_{[0,t)})$ it follows from (6) that

$$E\Big(\sum_{0 \leqq \tau_i \leqq T} \log \Lambda_{[0,\tau_i)}\Big) = \int_0^T E(\Lambda_{[0,t)} \log \Lambda_{[0,t)})\, dt. \tag{46}$$

From this and (45) we obtain

Proposition. $H_T = \int_0^T E(\Lambda_{[0,t)}(1 - \log \Lambda_{[0,t)}))\, dt.$

The reader is invited to check that the proof of this proposition goes through without the assumption of stationarity; we shall not go into this here. Under stricter assumptions involving the infinitesimal birth equations the above proposition follows from (3.6) in [10]. Similar calculations have been given by Brémaud [1] and Grigelionis [8] in connexion with the mutual information between two point processes. Cf. also [7].

Now notice that by Lemma 5

$$\Lambda_{[-t,0)} \log \Lambda_{[-t,0)} \xrightarrow{L_1} \Lambda_0 \log \Lambda_0 \tag{47}$$

$$\Lambda_0 \log \Lambda_{[-t,0)} \xrightarrow{L_1} \Lambda_0 \log \Lambda_0 \tag{48}$$

as $t \to \infty$. Since, by stationarity we have $H_T = \int_0^T E(\Lambda_{[-t,0)}(1 - \log \Lambda_{[-t,0)}))\, dt$, we obtain:

Corollary. $H=\lim_{T\to\infty}\frac{1}{T}H_T$.

This corollary shows that the entropy rate H as defined above agrees with the one which is implicit in Fritz's generalisation of McMillan's theorem (see [5]). In fact in the present case we can obtain this theorem directly from (44) and identify the limit as follows. Let $\mathscr{J}$ be the σ-field of shift invariant events in Ω.

Theorem 7. *Under hypotheses* (I) *and* (II)

$$\lim_{T\to\infty}\left(-\frac{1}{T}\log\bar{h}_T\right)=E(\Lambda_0-\Lambda_0\log\Lambda_0|\mathscr{J}) \text{ in } L_1\text{-mean.}$$

In fact, if we set

$$\bar{\bar{h}}_T=\prod_{0\leqq\tau_i\leqq T}\Lambda_{\tau_i}\cdot\exp\left[-\int_0^T\Lambda_t\,dt\right]$$

then, by (6),

$$\begin{aligned}
&E\left|\frac{1}{T}\log\bar{h}_T-\frac{1}{T}\log\bar{\bar{h}}_T\right|\\
&\leqq\frac{1}{T}E\Big(\sum_{0\leqq\tau_i\leqq T}|\log\Lambda_{[0,\tau_i)}-\log\Lambda_{\tau_i}|\Big)+\frac{1}{T}\int_0^T E|\Lambda_{[0,t)}-\Lambda_t|\,dt\\
&=\frac{1}{T}\int_0^T E(\Lambda_t|\log\Lambda_{[0,t)}-\log\Lambda_t|)\,dt+\frac{1}{T}\int_0^T E|\Lambda_{[0,t)}-\Lambda_t|\,dt\\
&=\frac{1}{T}\int_0^T E|\Lambda_0\log\Lambda_{[-t,0)}-\Lambda_0\log\Lambda_0|\,dt+\frac{1}{T}\int_0^T E|\Lambda_{[-t,0)}-\Lambda_0|\,dt
\end{aligned}$$

which converges to 0 as $T\to\infty$, by (48). To conclude the proof one then follows the standard step of applying the discrete ergodic theorem to the random variable $-\log\bar{\bar{h}}_1$ to deduce that $-\frac{1}{T}\log\bar{\bar{h}}_T$ converges in L_1. The limit is clearly shift invariant and its integral over any shift invariant event A agrees with that of $\Lambda_0-\Lambda_0\log\Lambda_0$. (Apply (6) with $Z_t=I_A\log\Lambda_t$, bearing in mind that A differs by a null set from a set in $\bigcap_{t\in R}\mathscr{F}_t$; see [3, p. 459]).

We conclude with some remarks on (42) and Corollary 6. It is easy to see that (32) implies

$$H(\xi_0^\varepsilon,\ldots,\xi_{n(\varepsilon)-1}^\varepsilon)=\left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right)P(N[0,\varepsilon)\geqq 1)+H_1+o(1).$$

Since $H_1=\int_0^1 E(\Lambda_{[0,t)}(1-\log\Lambda_{[0,t)}))\,dt$ we see that the analogue of (42) for the "unconditional" entropy follows from Fritz's result and the above Proposition. On the other hand if we let $\mathscr{A}_n$ $(n=1,2,\ldots)$ be the σ-field generated by

$\xi_0^{\frac{1}{n}}, \ldots, \xi_{n-1}^{\frac{1}{n}}$, then by Theorem 4 in [5] (applied to the measure $e\Pi$ on $\mathscr{F}_{[0,1]}$)

$$-\log E\left(\frac{dP}{d(e\Pi)}\middle|\mathscr{A}_n\right) \xrightarrow{L_1} -\log \bar{h}_1. \tag{49}$$

(This generalises a result of Perez. What is involved here is a submartingale with a directed parameter set). (49) and (44) easily imply

$$-\log p_{\frac{1}{n}}(\xi_0^{\frac{1}{n}}, \ldots, \xi_{n-1}^{\frac{1}{n}}) - (\log n) \sum_{k=0}^{n-1} \xi_k^{\frac{1}{n}} \xrightarrow{L_1} \int_0^1 \Lambda_{[0,t)}\, dt - \sum_{0 \leqq \tau_i \leqq 1} \log \Lambda_{[0,\tau_i)}$$

where we have put $p_{\frac{1}{n}}(i_0, \ldots, i_{n-1}) = P(\xi_0^{\frac{1}{n}} = i_0, \ldots, \xi_{n-1}^{\frac{1}{n}} = i_{n-1})$. This is the analogue of Corollary 6 for the "unconditional" information function. This statement is true without the assumption of stationarity, provided we replace (II) by $\int_0^1 E(\Lambda_{[0,t)}(1 - \log \Lambda_{[0,t)}))\, dt < \infty$.

References

1. Brémaud, P.M.: A martingale approach to point processes, Ph.D. thesis, Memorandum No. ERL–M345, Electronics Research Laboratory, University of California (1972)
2. Doleans-Dade, C., Meyer, P.A.: Intégrales stochastiques par rapport aux martingales locales, Séminaire de Probabilités IV, Lecture Notes in Mathematics 124, pp. 77–107, Berlin-Heidelberg-New York: Springer 1970
3. Doob, J.L.: Stochastic processes. New York: Wiley 1953
4. Fritz, J.: Entropy of point processes. Studia Sci. Math. Hungar. **4**, 389–399 (1969)
5. Fritz, J.: Generalization of McMillan's theorem to random set functions. Studia Sci. Math. Hungar. **5**, 369–394 (1970)
6. Fritz, J.: An approach to the entropy of point processes. Period. Math. Hungar. **3**, 73–83 (1973)
7. Grigelionis, B.: On mutual information for locally infinitely divisible stochastic processes (in Russian). Litovsk. Mat. Sb. **14**, No. 1, 5–11 (1974)
8. Grigelionis, B.: Stochastic point processes and martingales (in Russian). Litovsk. Mat. Sb. **15**, No. 3, 101–114 (1975)
9. Kabanov, Yu.M., Liptser, P.Š., Širyaev, A.N.: Martingale methods in the theory of point processes (in Russian). Proceedings of the School-Seminar on the theory of random processes, Part II, pp. 269–354. Institute of Physics and Mathematics, Vilnius 1975
10. McFadden, J.A.: The entropy of a point process. J. Soc. Industr. Appl. Math. **13**, 988–994 (1965)
11. Papangelou, F.: Integrability of expected increments of point processes and a related random change of scale. Trans. Amer. Math. Soc. **165**, 483–506 (1972)
12. Perez, A.: Information theory with an abstract alphabet. Generalized forms of McMillan's limit theorems for the case of discrete and continuous times. Theor. Probability Appl. **4**, 99–102 (1959)
13. Rényi, A.: On the dimension and entropy of probability distributions. Acta Math. Acad. Sci. Hungar. **10**, 193–215 (1959)
14. Rubin, I.: Regular point processes and their detection. IEEE Trans. Information Theory IT-**18**, 547–557 (1972)
15. Rudemo, M.: Dimension and entropy for a class of stochastic processes. Magyar Tud. Akad. Mat. Kutató Int. Közl **9**, 73–87 (1964)
16. Ryll-Nardzewski, C.: Remarks on processes of calls. Proc. 4th Berkeley Sympos. Math. Statist. Probab., Univ. Calif. **2**, 455–465 (1961)
17. Snyder, D.L.: Filtering and detection for doubly stochastic Poisson processes. IEEE Trans. Information Theory IT-**18**, 91–102 (1972)

Received April 22, 1977