COLLOQUIUM MATHEMATICUM

VOL. 97 2003 NO. 1

LARGE DEVIATION PRINCIPLE FOR EMPIRICAL MEASURES GENERATED BY COX POINT PROCESSES

BY

TOMASZ SCHREIBER (Toruń)

Abstract. Consider a sequence $(M_n)_{n=1}^{\infty}$ of random measures on $[0,\infty)$ such that $(M_n/n)_{n=1}^{\infty}$ satisfies the large deviation principle with a certain rate function I. Further, let N_n be the Cox point process (doubly stochastic Poisson point process) directed by M_n. The purpose of this paper is to establish the large deviation principle for the sequence of empirical measures $(N_n/n)_{n=1}^{\infty}$ with rate function expressed in terms of I.

1. Introduction and main results. The pioneering paper of Sanov (1957) on large deviations for empirical measures generated by independent random elements gave rise to an important number of refinements and applications both in probability theory (see e.g. Dembo and Zeitouni (1993), Deuschel and Stroock (1989), Dupuis and Ellis (1997), Groeneboom, Oosterhoff and Ruymgaart (1979) and the references therein, see also Proposition 3 in the Appendix) and in mathematical foundations of statistical mechanics (see e.g. Ellis (1985)). In the present paper we establish a large deviation principle for empirical measures corresponding to a stochastic mechanism of a different nature, namely to doubly stochastic Poisson point processes, usually also called Cox processes.

A Cox point process is constructed in a two-stage random procedure, by first taking at random a certain measure μ on the state space and then choosing the point configuration according to the Poisson process with intensity measure μ.

Cox processes arise naturally as a very general, flexible and yet analytically tractable model in a number of applications in statistics, physics and natural sciences. We refer the reader to Section 8.5 in Daley and Vere-Jones (1988) and to Chapter 7 in Karr (1991) for numerous examples of such applications. The latter reference reviews the rich set of existing mathematical tools for statistical inference for Cox point processes.

The large deviation principle for empirical measures generated by Cox point processes, as established below, apart from its intrinsic interest, seems

2000 *Mathematics Subject Classification*: 60F10, 60G55.

Key words and phrases: Cox point processes, large deviations, random measures, Sanov's theorem.

to provide a theoretical framework for statistical applications in the area of hypothesis testing, asymptotic discernibility and directing measure estimation for Cox point processes, much along the general lines sketched in Bucklew (1990), where a number of applications of the large deviation theory to statistical inference are discussed (see Chapters VI and IX there). The aforementioned applications, falling beyond the scope of the present article, are subject of the author's work in progress. Due to the variational form of the rate function in the general setting of our main Theorem 1, the anticipated results would also involve nonexplicit expressions of variational nature. Fortunately, in the important case of the mixed Poisson processes (see Section 1.1 in Karr (1991)) the explicit evaluation of the rate function is possible as stated in Corollary 1.

To proceed with the formal description of our setting, we denote by $\mathcal{M}_\infty$ the space of all nonnegative Borel measures on $[0,\infty)$ which are finite on compact sets, and we let $\mathcal{N}_\infty \subset \mathcal{M}_\infty$ be the subspace of all $\mathbb{N}\cup\{\infty\}$-valued (counting) nonnegative measures on $[0,\infty)$. We endow $\mathcal{M}_\infty$ with the so-called vague topology (see Definition A2.3.I in Daley and Vere-Jones (1988) or Section 1.9 in Kerstan, Matthes and Mecke (1982)) in which a sequence $(\mu_n)_{n=1}^\infty \subset \mathcal{M}_\infty$ converges to $\mu \in \mathcal{M}_\infty$ iff $\lim_{n\to\infty}\int f(x)\,\mu_n(dx) = \int f(x)\,\mu(dx)$ for all $f:[0,\infty)\to\mathbb{R}$ continuous with compact support. It is known that $\mathcal{N}_\infty$ is closed in $\mathcal{M}_\infty$ (see Proposition 1.9.5 in Kerstan, Matthes and Mecke (1982)). Further, it can be proved that $\mathcal{M}_\infty$ can be metrised so as to make it a Polish space (see Theorem A2.6.III in Daley and Vere-Jones (1988) or Proposition 1.9.1 in Kerstan, Matthes and Mecke (1982)). The Borel σ-field corresponding to that topology endows $\mathcal{M}_\infty$, and hence $\mathcal{N}_\infty$, with a measurable structure.

A random element taking values in $\mathcal{M}_\infty$ will be referred to as a *random measure* on $[0,\infty)$. Further, $\mathcal{N}_\infty$-valued random elements are called *point processes* on $[0,\infty)$. All random elements considered in this paper are assumed, without further mention, to be defined on a common probability space $(\Omega,\Im,\mathbb{P})$. For a random element X taking values in a measurable space $\mathcal{X}$ we write $\mathbb{P}^X$ for the distribution of X on $\mathcal{X}$.

For each random measure M we denote by Λ_M its intensity measure (see Section 1.1 in Kerstan, Matthes and Mecke (1982) or Definition 7.2.III in Daley and Vere-Jones (1988)), given by $\Lambda_M(A) := \mathbb{E}M(A)$ for each Borel $A \subseteq [0,\infty)$. Note that it may happen that $\Lambda_M \notin \mathcal{M}_\infty$. The operation of taking the intensity measure defines in the natural way the measure-valued mapping $\Lambda(\cdot)$ acting on $\mathcal{P}(\mathcal{M}_\infty)$ by $\Lambda(\mathbb{P}^M) := \Lambda_M$, with $\mathcal{P}(\mathcal{M}_\infty)$ standing for the space of all Borel probability measures on $\mathcal{M}_\infty$ endowed with the usual weak topology.

For a given random measure M the *Cox point process* (also called *doubly stochastic Poisson point process*) *directed by* M is defined to be the (unique in distribution) point process N such that its conditional distribution given by $M = \mu$ coincides for M-almost all $\mu \in \mathcal{M}_\infty$ with that of the Poisson point process Π_μ with intensity measure μ (see Daley and Vere-Jones (1988), Sections 2.4 and 8.5, or Karr (1991), Definitions 1.2 and 1.3).

Let $\mathcal{X}$ be a Polish metric space and $X_1, X_2, \ldots$ a sequence of $\mathcal{X}$-valued random elements. We say that $(X_n)_{n=1}^\infty$ satisfies the *large deviation principle on* $\mathcal{X}$ *with* (*good*) *rate function* $I : \mathcal{X} \to [0, \infty]$ if the following conditions are satisfied (see Section 1.1 in Dupuis and Ellis (1997) or Chapter 2 in Deuschel and Stroock (1989)):

- the level sets $\{x \in \mathcal{X} \mid I(x) \le A\}$ for $A \ge 0$ are compact (thus, in particular, I is lower semicontinuous),
- $\liminf_{n\to\infty} n^{-1} \log \mathbb{P}(X_n \in \mathcal{G}) \ge -\inf_{x\in\mathcal{G}} I(x)$ for each open set $\mathcal{G} \subseteq \mathcal{X}$,
- $\limsup_{n\to\infty} n^{-1} \log \mathbb{P}(X_n \in (X_n \mathcal{F}) \le -\inf_{x\in\mathcal{F}} I(x)$ for each closed set $\mathcal{F} \subseteq \mathcal{X}$.

To proceed consider a sequence $(M_n)_{n=1}^\infty$ of random measures on $[0, \infty)$ satisfying the following conditions:

(L) $(M_n/n)_{n=1}^\infty$ satisfies the large deviation principle on $\mathcal{M}_\infty$ with some rate function I,

(B) there exists a nonnegative measure $\kappa \in \mathcal{M}_\infty$ which is strictly positive on open sets and such that almost surely $M_n(A) \ge n\kappa(A)$ for each Borel set $A \subset [0, \infty)$ and $n \in \mathbb{N}$.

For all $n \in \mathbb{N}$ let N_n be the Cox point process directed by M_n. Dividing the random counting measure N_n by n we obtain an "empirical measure" whose behaviour for large n should, roughly speaking, resemble that of M_n/n. In other words, in view of condition (L) it is reasonable to expect that N_n/n satisfies the large deviation principle with some rate function closely related to I. This is the contents of our main theorem. To formulate it we need some additional notation.

For any two probability measures θ, γ defined on the same measurable space, by the *relative entropy of* θ *with respect to* γ (or, in classical terminology, the *Kullback–Leibler divergence*) we mean

$$R(\theta \,\|\, \gamma) = \begin{cases} \int \log(d\theta/d\gamma)\, d\theta \\ \qquad \text{if } \theta \text{ is absolutely continuous with respect to } \gamma, \\ \infty \quad \text{otherwise.} \end{cases} \tag{1}$$

For the details concerning this concept see e.g. Dupuis and Ellis (1997) and the references therein.

For our purposes it will be of particular importance to investigate the relative entropy with respect to the homogeneous Poisson point process.

Namely, for $\mu \in \mathcal{M}_\infty$ we set

$$(2) \qquad \mathcal{H}(\mu) = \inf\{R(\mathcal{Q} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) \mid \mathcal{Q} \in \mathcal{P}(\mathcal{N}_\infty),\, \Lambda(\mathcal{Q}) = \mu\},$$

where $\mathcal{P}(\mathcal{N}_\infty)$ denotes the space of all Borel probability measures on $\mathcal{N}_\infty$ (endowed with the usual weak topology) and Π_{λ_+} is the homogeneous Poisson point process on $[0,\infty)$ with intensity measure λ_+ defined as the restriction of the Lebesgue measure λ to $\mathbb{R}_+$.

Further, for two measures $\mu, \gamma \in \mathcal{M}_\infty$ define their composition $\mu \circ \gamma \in \mathcal{M}_\infty$ by

$$(3) \qquad \mu \circ \gamma([0,t]) = \mu([0, \gamma([0,t])])$$

for all $t \geq 0$. In other words, denoting by F_μ, F_γ and $F_{\mu\circ\gamma}$ the respective distribution functions, we have $F_{\mu\circ\gamma}(t) = F_\mu(F_\gamma(t))$. Our interest in the operation $\circ$ is motivated by the representation formula (33), to be established in the proof of our Theorem 1, according to which the empirical measure N_n/n coincides in distribution with the composition of $\Pi^{(n)}/n$ with M_n/n, where $\Pi^{(n)}$ is an intensity-n homogeneous Poisson point process on $[0,\infty)$ independent of M_n.

The main result of our article is:

Theorem 1. *Assume as above that N_n are Cox point processes directed by the random measures M_n and that conditions* (L) *and* (B) *hold. Then the sequence $(N_n/n)_{n=1}^\infty$ satisfies on $\mathcal{M}_\infty$ the large deviation principle with rate function J given for $\theta \in \mathcal{M}_\infty$ by*

$$(4) \qquad J(\theta) = \inf\{I(\gamma) + \mathcal{H}(\mu) \mid \mu, \gamma \in \mathcal{M}_\infty,\, \mu \circ \gamma = \theta\}$$

with I as in (L).

The relative entropy $\mathcal{H}(\mu)$ appearing in the expression for the rate function can be determined explicitly. Namely, we have

Lemma 1. *The Poisson point process Π_μ minimizes the relative entropy with respect to Π_{λ_+} among all point processes with a given intensity measure $\mu \in \mathcal{M}_\infty$, i.e.*

$$(5) \quad \mathcal{H}(\mu) = \inf\{R(\mathcal{Q} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) \mid \mathcal{Q} \in \mathcal{P}(\mathcal{N}_\infty),\, \Lambda(\mathcal{Q}) = \mu\} = R(\mathbb{P}^{\Pi_\mu} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}).$$

Further

$$(6) \qquad \mathcal{H}(\mu) = R(\mathbb{P}^{\Pi_\mu} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) = \begin{cases} \int_0^\infty \vartheta(d\mu/d\lambda_+)\, d\lambda_+, & \mu \ll \lambda_+, \\ \infty, & \textit{otherwise}, \end{cases}$$

where $\vartheta(u) := u \log u + 1 - u$ (note that $\vartheta(u) \geq 0$ for $u \geq 0$ so that the above integral is always well defined, though possibly infinite).

The variational formula (4) for the rate function in Theorem 1 may in general be extremely difficult to evaluate explicitly. However, if the nature of the directing measures M_n is not very complicated, it may happen rather often that M_n/n belongs almost surely to a certain narrow class $\mathcal{Y} \subseteq \mathcal{M}_\infty$.

If $\mathcal{Y}$ is chosen to be closed in $\mathcal{M}_\infty$, this yields immediately $I(\gamma) = \infty$ for $\gamma \notin \mathcal{Y}$. In particular, in (4) it is then enough to consider the infimum over $\gamma \in \mathcal{Y}$. Even further simplification can be obtained if it happens that a given $\theta \in \mathcal{M}_\infty$ admits only a very limited number of representations $\theta = \mu \circ \gamma$ with $\gamma \in \mathcal{Y}$ and $\mathcal{H}(\mu) < \infty$. In such cases we can hope to obtain an explicit expression for $J(\theta)$. As an example of such a situation we consider the so-called (*homogeneous*) *mixed Poisson point processes* defined to be Cox point processes directed by random multiplicities of the Lebesgue measure (see e.g. Section 1.1 in Karr (1991) or Exercise 2.1.8 in Daley and Vere-Jones (1988)). Clearly, in such a case we can set $\mathcal{Y} = \{\alpha\lambda_+ \mid \alpha \geq 0\}$. The details are given in the following corollary.

COROLLARY 1. *Consider a sequence of directing random measures M_n satisfying* (B) *and* (L) *and admitting the representation $M_n := \zeta_n \lambda_+$, where ζ_n are certain nonnegative random variables. Let N_n be the corresponding Cox* (*mixed Poisson*) *point processes. Further, denote by $\widehat{I}$ the unique rate function on $\mathbb{R}$ which governs the large deviations of ζ_n/n and is such that I satisfies*

$$I(\gamma) = \begin{cases} \widehat{I}(\alpha), & \gamma = \alpha\lambda_+, \\ \infty, & \text{otherwise.} \end{cases}$$

Then, for a given $\theta \in \mathcal{M}_\infty$, we have:

1. *If the limit $\alpha := \lim_{T\to\infty} \theta([0,T])/T$ exists and $0 < \alpha < \infty$ then $J(\theta) = \widehat{I}(\alpha) + \int_0^\infty \vartheta_\alpha(d\theta/d\lambda_+)\, d\lambda_+$ with $\vartheta_\alpha(u) := \alpha\vartheta(u/\alpha) = u\log u + \alpha - u(1+\log\alpha)$, provided θ is absolutely continuous with respect to λ_+,*
2. *$J(\theta) = \infty$ otherwise.*

2. Proofs. We will proceed as follows. First in Lemma 2 we investigate the properties of the intensity measure mapping on the level sets of the relative entropy. This technical result plays an important role in further proofs because in general, as mentioned before, the intensity measure mapping can exhibit quite an irregular behaviour being discontinuous or taking values outside of $\mathcal{M}_\infty$. Next, we prove Lemma 1. The most complex task to be accomplished in this paper is to prove the crucial Lemma 3 stating the large deviation principle for empirical measures generated by homogeneous Poisson point processes, which corresponds to the assertion of Theorem 1 for deterministic homogeneous directing measures $M_n = n\lambda_+$. The next step will be to deduce the main theorem from Lemma 3. Finally we establish Corollary 1.

When completing the paper we have learnt that results similar to our Lemma 3 have recently appeared in the literature (see Florens and Pham (1998) and Léonard (2000)). However, we present a proof using different methods, which we believe to be of independent interest as relying on a natural interpretation of the rate function $\mathcal{H}(\mu) = R(\mathbb{P}^{\Pi_\mu} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}})$.

2.1. *Lemma 2*

LEMMA 2. *The mapping*

$$\mathcal{P}(\mathcal{N}_\infty) \ni \mathcal{Q} \mapsto \Lambda(\mathcal{Q}) \in \mathcal{M}_\infty$$

is well defined and weakly continuous on the relative entropy level sets

$$\{\mathcal{Q} \in \mathcal{P}(\mathcal{N}_\infty) \mid R(\mathcal{Q} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) \leq C\}$$

for $C \geq 0$. *In particular*, *for each* $\mu \in \mathcal{M}_\infty$ *and* $C \geq 0$ *the set*

$$\{\mathcal{Q} \in \mathcal{P}(\mathcal{N}_\infty) \mid \Lambda(\mathcal{Q}) = \mu,\, R(\mathcal{Q} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) \leq C\}$$

is weakly compact.

Due to its technical character the proof of this lemma is postponed to the Appendix.

2.2. *Proof of Lemma 1.* Let us first concentrate on (5). Clearly we can assume without loss of generality that

$$\inf_{\mathcal{Q}} \{R(\mathcal{Q} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) \mid \mathcal{Q} \in \mathcal{P}(\mathcal{N}_\infty),\, \Lambda(\mathcal{Q}) = \mu\} < \infty \tag{7}$$

for otherwise our assertion becomes trivial. In view of (7) we conclude from Lemma 2 and from the lower semicontinuity of $R(\cdot \,\|\, \mathbb{P}^{\Pi_{\lambda_+}})$ (see e.g. Lemma 1.4.3(b) in Dupuis and Ellis (1997)) that there exists a point process $\widehat{\Psi}$ such that

$$R(\mathbb{P}^{\widehat{\Psi}} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) = \min_{\mathcal{Q}} \{R(\mathcal{Q} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) \mid \mathcal{Q} \in \mathcal{P}(\mathcal{N}_\infty),\, \Lambda(\mathcal{Q}) = \mu\}. \tag{8}$$

We will show that the relative entropy minimiser $\widehat{\Psi}$ inherits from Π_{λ_+} the independence property on disjoint domains and we will conclude that $\widehat{\Psi} \stackrel{\mathrm{d}}{=} \Pi_\mu$. To this end we consider an arbitrary partition $A, B \subset [0,\infty)$ with $A \cup B = [0,\infty)$ and $A \cap B = \emptyset$. Define the restricted point process $\widehat{\Psi}_{|A}$ by setting, for each $C \subseteq [0,\infty)$,

$$\widehat{\Psi}_{|A}(C) = \widehat{\Psi}(A \cap C)$$

almost surely and let $\widehat{\Psi}_{|B}$, $(\Pi_{\lambda_+})_{|A}$ and $(\Pi_{\lambda_+})_{|B}$ be defined analogously. For notational convenience we also agree to write $\mathbb{P}_A^{\widehat{\Psi}}$ instead of $\mathbb{P}^{\widehat{\Psi}_{|A}}$ etc. By general properties of the relative entropy (see Subsection 3.5 in the Appendix) we conclude that

$$\begin{aligned} R(\mathbb{P}^{\widehat{\Psi}} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) &= R(\mathbb{P}^{\widehat{\Psi}} \,\|\, \mathbb{P}_A^{\Pi_{\lambda_+}} \times \mathbb{P}_B^{\Pi_{\lambda_+}}) \\ &\geq R(\mathbb{P}_A^{\widehat{\Psi}} \times \mathbb{P}_B^{\widehat{\Psi}} \,\|\, \mathbb{P}_A^{\Pi_{\lambda_+}} \times \mathbb{P}_B^{\Pi_{\lambda_+}}) \\ &= R(\mathbb{P}_A^{\widehat{\Psi}} \,\|\, \mathbb{P}_A^{\Pi_{\lambda_+}}) + R(\mathbb{P}_B^{\widehat{\Psi}} \,\|\, \mathbb{P}_B^{\Pi_{\lambda_+}}) \end{aligned} \tag{9}$$

with equality iff

$$\mathbb{P}^{\widehat{\Psi}} = \mathbb{P}^{\widehat{\Psi}}_A \times \mathbb{P}^{\widehat{\Psi}}_B. \tag{10}$$

However, if the inequality in (9) were strict, the point process $\widetilde{\Psi}$ defined to be the sum of independent copies of $\widehat{\Psi}_{|A}$ and $\widehat{\Psi}_{|B}$ would have the property

$$R(\mathbb{P}^{\widetilde{\Psi}} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) = R(\mathbb{P}^{\widehat{\Psi}}_A \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}_A) + R(\mathbb{P}^{\widehat{\Psi}}_B \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}_B) < R(\mathbb{P}^{\widehat{\Psi}} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}),$$

which contradicts (8). This means that (10) has to be satisfied. In particular $\widehat{\Psi}_{|A}$ and $\widehat{\Psi}_{|B}$ are independent.

In addition $\widehat{\Psi}$ is a simple point process, i.e. it concentrates with probability 1 on the space of the counting measures whose atoms all have mass 1 (see Kerstan, Matthes and Mecke (1982), Section 1.2). This property of $\widehat{\Psi}$ follows immediately from $R(\mathbb{P}^{\widehat{\Psi}} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) < \infty$ and from the simplicity of Π_{λ_+} (ibid., Proposition 1.5.7). Since A and B were arbitrary, by Proposition 1.1.9 ibid. the independence of $\widehat{\Psi}_{|A}$ and $\widehat{\Psi}_{|B}$ yields the complete independence property for $\widehat{\Psi}$ (see Assumption 2.4.V in Daley and Vere-Jones (1988)), stating that for each finite family $A_1, \ldots, A_k$ of bounded disjoint Borel subsets of $[0, \infty)$ the random variables $\widehat{\Psi}(A_1), \ldots, \widehat{\Psi}(A_k)$ are mutually independent. Applying Theorem 2.4.VIII ibid. and taking into account the simplicity and the lack of fixed atoms of $\widehat{\Psi}$ we conclude that $\widehat{\Psi}$ is a certain Poisson point process. Recalling that $\Lambda_{\widehat{\Psi}} = \mu$ we see that $\widehat{\Psi}$ has the same law as Π_μ. This completes the proof of (5).

To proceed, we turn to (6). We begin by noting that if the measure μ is not absolutely continuous with respect to λ_+ then

$$R(\mathbb{P}^{\Pi_\mu} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) = \infty$$

because there exists some bounded $A \subset [0, \infty)$ such that $\mu(A) > 0$ and $\lambda_+(A) = 0$ and hence $\mathbb{P}(\Pi_\mu(A) > 0) > 0$ and $\mathbb{P}(\Pi_{\lambda_+}(A) > 0) = 0$. In particular (6) is satisfied because $R(\mu_{|[0,a]} \,\|\, \lambda_{|[0,a]}) = \infty$ for a such that $A \subset [0, a]$. Thus, henceforth we assume the absolute continuity of μ with respect to λ_+. Using the standard formulae for the density of one Poisson process with respect to another (see Kerstan, Matthes and Mecke (1982), Proposition 1.5.11) we arrive at

$$R(\mathbb{P}^{(\Pi_\mu)_{|[0,a]}} \,\|\, \mathbb{P}^{(\Pi_{\lambda_+})_{|[0,a]}}) = \int_0^a \vartheta\left(\frac{d\mu}{d\lambda_+}\right) d\lambda_+ \tag{11}$$

(recall that $\vartheta(u) = u \log u + 1 - u$). On the other hand, from the properties of the relative entropy (see Proposition 15.6 in Georgii (1988)) we conclude that

$$R(\mathbb{P}^{\Pi_\mu} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) = \lim_{a \to \infty} R(\mathbb{P}^{\Pi_{\mu|[0,a]}} \,\|\, \mathbb{P}^{\Pi_{\lambda|[0,a]}}).$$

Combining this with (11) yields (6). The proof of Lemma 1 is hence complete. ∎

2.3. *Lemma 3.* The following lemma is crucial to the proof of Theorem 1.

LEMMA 3. *For each* $n \in \mathbb{N}$ *let* $\Pi^{(n)}$ *denote the homogeneous Poisson point process on* $[0,\infty)$ *with intensity* n (*i.e. its intensity measure is* $n\lambda_+$). *Then the sequence* $(\Pi^{(n)}/n)_{n=0}^{\infty}$ *of random measures satisfies on* $\mathcal{M}_\infty$ *the large deviation principle with rate function* $\mathcal{H}$ *given by* (2) *and* (6).

Proof. Let $\Pi_{(1)}, \Pi_{(2)}, \ldots$ be a sequence of independent homogeneous Poisson point processes on $[0,\infty)$ with common intensity 1. It is clear that

$$\Pi^{(n)} \stackrel{\mathrm{d}}{=} \Pi_{(1)} + \ldots + \Pi_{(n)}. \tag{12}$$

This identity leads to the idea whose rough description is given below. For each $\omega \in \Omega$ the measure $(\Pi^{(n)}/n)(\omega) \in \mathcal{M}_\infty$ can be represented as the intensity measure of the distribution $\Theta_n(\omega) \in \mathcal{P}(\mathcal{N}_\infty)$ which assigns probability $1/n$ to $\Pi_{(1)}(\omega)$, probability $1/n$ to $\Pi_{(2)}(\omega)$ etc. More precisely, let Θ_n be the $\mathcal{P}(\mathcal{N}_\infty)$-valued random element defined by

$$[\Theta_n(\omega)](\mathcal{A}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{\Pi_{(i)}(\omega)\in\mathcal{A}\}}. \tag{13}$$

Then it is easily seen that $[\Pi^{(n)}/n](\omega) = \Lambda(\Theta_n(\omega))$ for almost all $\omega \in \Omega$. Hence, dropping the ω's,

$$\Pi^{(n)}/n = \Lambda(\Theta_n)$$

almost surely.

It is very convenient to regard Θ_n as the empirical measure generated on $\mathcal{N}_\infty$ by $\mathcal{N}_\infty$-valued random elements $\Pi_{(1)}, \ldots, \Pi_{(n)}$. We recall that the space $\mathcal{N}_\infty$ can be metrised to be Polish (see the discussion following its definition), and therefore, the same is true for $\mathcal{P}(\mathcal{N}_\infty)$ (see e.g. Lemma 3.2.2 in Deuschel and Stroock (1989)). Thus, applying Sanov's theorem (see e.g. Dupuis and Ellis (1997), Theorem 2.2.1, Deuschel and Stroock (1989), Theorem 3.2.17, or Proposition 3 in the Appendix) we conclude that the sequence $(\Theta_n)_{n=1}^{\infty}$ satisfies on $\mathcal{P}(\mathcal{N}_\infty)$ the large deviation principle with rate function

$$\mathcal{R}(\mathcal{Q}) = R(\mathcal{Q} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}), \qquad \mathcal{Q} \in \mathcal{P}(\mathcal{N}_\infty). \tag{14}$$

Therefore, if the intensity measure mapping Λ were well defined from $\mathcal{P}(\mathcal{N}_\infty)$ to $\mathcal{M}_\infty$ and if it were continuous, we could easily complete the proof by using the standard contraction principle (see e.g. Dupuis and Ellis (1997), Theorem 1.3.2), quoted in the Appendix as Proposition 2 for the convenience of the reader. However, the properties of this mapping are by no means that good: it is not continuous, and even worse, its values often fail to belong to $\mathcal{M}_\infty$. Nevertheless, we will proceed along the suggested lines, although a considerable additional effort will be needed to make things work.

For each increasing sequence $L = (L_0, L_1, \dots)$ of positive numbers, consider the space $\mathcal{M}_\infty(L)$ of all those measures $\mu \in \mathcal{M}_\infty$ which satisfy

$$\mu([k, k+1]) \le L_k + L_{k+1} \tag{15}$$

for every $k \in \mathbb{N}$. By the above boundedness condition, for each $\Xi \in \mathcal{P}(\mathcal{M}_\infty(L))$ the intensity measure $\Lambda(\Xi)$ is well defined and it belongs to $\mathcal{M}_\infty(L)$.

For each measure $\mu \in \mathcal{M}_\infty$ we construct the corresponding "rescaled" measure $\mu^{[L]} \in \mathcal{M}_\infty(L)$ by setting, for $k \in \mathbb{N}$ and $A \subseteq [k, k+1)$,

$$\mu^{[L]}(A) = \begin{cases} \mu(A) & \text{if } \mu([k,k+1)) \le L_k, \\ \dfrac{L_k}{\mu([k,k+1))}\,\mu(A) & \text{otherwise.} \end{cases} \tag{16}$$

An equivalent of Θ_n given by (13) with respect to this rescaling is $\Theta_n^{[L]}$ defined by

$$[\Theta_n^{[L]}(\omega)](\mathcal{A}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{\Pi_{(i)}^{[L]}(\omega) \in \mathcal{A}\}} \tag{17}$$

for $\mathcal{A} \subseteq \mathcal{M}_\infty(L)$. In other words, $\Theta_n^{[L]}$ is to be interpreted as the empirical measure on $\mathcal{M}_\infty(L)$ generated by the $\mathcal{M}_\infty(L)$-valued random elements $\Pi_{(1)}^{[L]}, \dots, \Pi_{(n)}^{[L]}$ (note that $\Pi_{(n)}^{[L]}$ does not have to be a point process). Clearly,

$$\Lambda(\Theta_n^{[L]}) = \frac{1}{n}\,(\Pi_{(1)}^{[L]} + \dots + \Pi_{(n)}^{[L]})$$

almost surely.

In analogy with what we did for Θ_n, also here we note that $\mathcal{M}_\infty(L)$, as a closed subset of $\mathcal{M}_\infty$, can be metrised as a Polish space, and therefore, we can apply Sanov's theorem to obtain the large deviation principle for $(\Theta_n)^{[L]}$ on $\mathcal{P}(\mathcal{M}_\infty(L))$ with rate function

$$\mathcal{R}^{[L]}(\mathcal{Q}) = R(\mathcal{Q} \,\|\, \mathbb{P}^{\Pi_{\lambda^+}^{[L]}}), \qquad \mathcal{Q} \in \mathcal{P}(\mathcal{M}_\infty(L)). \tag{18}$$

Applying the bounded convergence theorem it is easily checked that the intensity measure mapping $\Lambda : \mathcal{P}(\mathcal{M}_\infty(L)) \to \mathcal{M}_\infty(L)$ is continuous. Therefore, using (18) and the contraction principle (Theorem 1.3.2 in Dupuis and Ellis (1997), Proposition 2 in the Appendix) we conclude that the sequence $(\Lambda(\Theta_n^{[L]}))_{n=0}^\infty$ satisfies on $\mathcal{M}_\infty(L)$ the large deviation principle with rate function

$$\mathcal{H}^{[L]}(\mu) = \inf\{\mathcal{R}^{[L]}(\mathcal{Q}) \mid \mathcal{Q} \in \mathcal{P}(\mathcal{M}_\infty(L)),\, \Lambda(\mathcal{Q}) = \mu\}.$$

It is important to observe that $\mathcal{H}^{[L]}$ can be expressed in a more convenient form:

$$\mathcal{H}^{[L]}(\mu) = \inf\{\mathcal{R}(\mathcal{Q}) \mid \mathcal{Q} \in \mathcal{P}(\mathcal{N}_\infty),\ \Lambda(\mathcal{Q}^{[L]}) = \mu\}, \tag{19}$$

where $\mathcal{Q}^{[L]}$ denotes the image of the probability measure $\mathcal{Q} \in \mathcal{P}(\mathcal{N}_\infty)$ under the mapping $\nu \mapsto \nu^{[L]}$ and $\mathcal{R}$ is given by (14). Indeed, we know that the Poisson point process Π_{λ_+} almost surely takes its values in the space $\widehat{\mathcal{N}}_\infty$ of all simple counting measures (i.e. counting measures whose atoms all have mass 1; see Kerstan, Matthes and Mecke (1982), Section 1.2 and Proposition 1.5.7). Thus, almost surely $\Pi_{\lambda_+}^{[L]} \in (\widehat{\mathcal{N}}_\infty)^{[L]}$, where $(\widehat{\mathcal{N}}_\infty)^{[L]} = \{\nu^{[L]} \mid \nu \in \widehat{\mathcal{N}}_\infty\}$. Therefore, by (18), if for some $\mathcal{Q} \in \mathcal{P}(\mathcal{M}_\infty(L))$ we have $\mathcal{R}^{[L]}(\mathcal{Q}) < \infty$, then $\mathcal{Q}((\widehat{\mathcal{N}}_\infty)^{[L]}) = 1$. However, it is easily checked that the mapping $\widehat{\mathcal{N}}_\infty \ni \nu \mapsto \nu^{[L]} \in (\widehat{\mathcal{N}}_\infty)^{[L]}$ is a measurable bijection. Hence, $\mathcal{Q}$ is of the form $\widehat{\mathcal{Q}}^{[L]}$ for some $\widehat{\mathcal{Q}} \in \mathcal{P}(\mathcal{N}_\infty)$. Further, $\mathcal{R}^{[L]}(\widehat{\mathcal{Q}}^{[L]}) = \mathcal{R}(\widehat{\mathcal{Q}})$ for each $\widehat{\mathcal{Q}} \in \mathcal{P}(\widehat{\mathcal{N}}_\infty)$. Relation (19) now follows by standard arguments.

The idea underlying the remaining part of the proof is, roughly speaking, to show that letting $L_i \to \infty$ we can approximate Θ_n with $\Theta_n^{[L]}$, $\Pi^{(n)}/n$ with $\Lambda(\Theta_n^{[L]})$, and $\mathcal{H}$ with $\mathcal{H}^{[L]}$.

To proceed, we introduce an explicit metric ϱ on $\mathcal{M}_\infty$ with the vague topology. Let $\mathcal{D} = \{g_1, g_2, \ldots\}$ be a countable family of continuous functions defined on $[0,\infty)$ such that

(D1) the support of each g_i is contained in some interval of length 1,
(D2) the supremum norm $\|g_i\|_\infty$ equals 1 for each $i \in \mathbb{N}$,
(D3) the linear hull of $\mathcal{D}$ is uniformly dense in the space of all continuous functions on $[0,\infty)$ with compact supports.

The existence of such $\mathcal{D}$ is not difficult to establish (for instance we can take g_i's equal on their supports to appropriate polynomials or trigonometric polynomials). For $\mu, \nu \in \mathcal{M}_\infty$ we set

$$\varrho(\mu,\nu) = \sum_{n=1}^{\infty} \frac{1}{2^n} \Big| \int_{[0,\infty)} g_i\, d\mu - \int_{[0,\infty)} g_i\, d\nu \Big|. \tag{20}$$

Using standard arguments we easily prove that ϱ metrises the vague topology on $\mathcal{M}_\infty$ and that the resulting metric space $(\mathcal{M}_\infty, \varrho)$ is Polish.

In view of Corollary 1.2.5 in Dupuis and Ellis (1997) the proof of the lemma will be complete if we show that for each bounded Lipschitz function $h : (\mathcal{M}_\infty, \varrho) \to \mathbb{R}$ we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E} \exp(-nh(\Pi^{(n)}/n)) = -\inf_{\mu\in\mathcal{M}_\infty} (\mathcal{H}(\mu) + h(\mu)). \tag{21}$$

The proof of this identity is subdivided into three steps.

STEP (1): The limit in (21) exists. Let $\xi, \xi_1, \xi_2, \dots$ be a sequence of i.i.d. random variables with standard mean 1 Poisson distribution POISS(1). Take some $\varepsilon > 0$ and choose $L_0 > 0$ such that

$$\mathbb{E}(\xi \mathbf{1}_{\{\xi > L_0\}}) < \varepsilon. \tag{22}$$

Further, define $\mathcal{L}(l,t)$ to be the Laplace transform of the random variable $\mathbf{1}_{\{\xi>l\}}\xi$, i.e.

$$\mathcal{L}(l,t) = \sum_{j\le l} \frac{e^{-1}}{j!} + \sum_{j>l} \exp(tj)\, \frac{e^{-1}}{j!},$$

and let, for $\alpha \ge \varepsilon$ and $l \ge L_0$,

$$\mathcal{I}(l,\alpha) = \sup_{t\ge 0}(\alpha t - \log \mathcal{L}(l,t)).$$

Note that in view of (22),

$$\mathcal{I}(L_0, 2\varepsilon) > 0.$$

It is easily seen that $\lim_{l\to\infty} \mathcal{I}(l,\alpha) = \infty$ for fixed α, so it is possible to choose an increasing sequence $L_0 < L_1 < L_2 < \dots$ such that

$$\mathcal{I}(L_k, 2\varepsilon) > \mathcal{I}(L_0, 2\varepsilon) + k. \tag{23}$$

We will investigate the behaviour of the probabilities

$$p_{n,k} := \mathbb{P}\left(\left[\frac{1}{n} \sum_{i=1}^{n} \Pi_{(i)} - \frac{1}{n} \sum_{i=1}^{n} \Pi_{(i)}^{[L]} \right]([k, k+1)) > 2\varepsilon \right)$$

for $n, k \in \mathbb{N}$. It is easily verified, using (16) and the properties of a homogeneous Poisson point process, that

$$p_{n,k} \le \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{n} \xi_i \mathbf{1}_{\{\xi_i > L_k\}} > 2\varepsilon \right).$$

Hence, it follows by standard arguments (see e.g. Lemma 1.2.3 in Deuschel and Stroock (1989)) that

$$p_{n,k} \le \exp(-n\mathcal{I}(L_k, 2\varepsilon)).$$

Thus, by (23) we get

$$p_{n,k} \le \exp(-n(\mathcal{I}(L_0, 2\varepsilon) + k)).$$

Finally, taking the sum over k gives

$$\sum_{k=0}^{\infty} p_{n,k} \le \frac{1}{1-e^{-n}} \exp(-n\mathcal{I}(L_0, 2\varepsilon)).$$

Clearly $1/(1-e^{-n}) < 2$ for $n \ge 1$, so by (20) and in view of conditions (D1) and (D2) this bound yields

$$\mathbb{P}\biggl(\varrho\biggl(\frac{1}{n}\sum_{i=1}^{n}\Pi_{(i)},\frac{1}{n}\sum_{i=1}^{n}\Pi_{(i)}^{[L]}\biggr)>4\varepsilon\biggr)<2\exp(-n\mathcal{I}(L_0,2\varepsilon)).$$

Thus, by (12) and (17) we get

$$\mathbb{P}\biggl(\varrho\biggl(\frac{1}{n}\Pi^{(n)},\Lambda(\Theta_n^{[L]})\biggr)>4\varepsilon\biggr)<2\exp(-n\mathcal{I}(L_0,2\varepsilon)). \tag{24}$$

Now let C be the Lipschitz constant for h in (21) and let $\|h\|_\infty$ be the supremum norm of h. Then

$$\begin{aligned}\mathbb{E}\exp(-nh(\Pi^{(n)}/n))>\mathbb{E}\exp(-n(h(\Lambda(\Theta_n^{[L]}))+4C\varepsilon))\\ -\exp(n\|h\|_\infty)\mathbb{P}(\varrho(\Pi^{(n)}/n,\Lambda(\Theta_n^{[L]}))>4\varepsilon).\end{aligned}$$

Hence, by (24),

$$\begin{aligned}\mathbb{E}\exp(-nh(\Pi^{(n)}/n))>\ &\exp(-4nC\varepsilon)\mathbb{E}(\exp(-nh(\Lambda(\Theta_n^{[L]}))))\\ &-2\exp(n(\|h\|_\infty-\mathcal{I}(L_0,2\varepsilon))).\end{aligned}$$

Assuming without loss of generality that $\|h\|_\infty-\mathcal{I}(L_0,2\varepsilon)<-\|h\|_\infty$ (we can always guarantee it, possibly increasing L_0) we conclude that

$$\begin{aligned}&\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\biggl(-nh\biggl(\frac{1}{n}\Pi^{(n)}\biggr)\biggr)\\ &\qquad\ge\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp(-nh(\Lambda(\Theta_n^{[L]})))-4C\varepsilon.\end{aligned}$$

Thus, since the sequence $(\Lambda(\Theta_n^{[L]}))_{n=0}^{\infty}$ satisfies the large deviation principle on $\mathcal{M}_\infty(L)$ with rate function $\mathcal{H}^{[L]}$ (see (19)), applying Theorem 1.2.1 in Dupuis and Ellis (1997) we conclude that

$$\begin{aligned}&\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\biggl(-nh\biggl(\frac{1}{n}\Pi^{(n)}\biggr)\biggr)\\ &\qquad\ge-\inf_{\mu\in\mathcal{M}_\infty(L)}(\mathcal{H}^{[L]}(\mu)+h(\mu))-4C\varepsilon.\end{aligned} \tag{25}$$

In an analogous way we deduce that

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\biggl(-nh\biggl(\frac{1}{n}\Pi^{(n)}\biggr)\biggr)\\ &\qquad\le-\inf_{\mu\in\mathcal{M}_\infty(L)}(\mathcal{H}^{[L]}(\mu)+h(\mu))+4C\varepsilon.\end{aligned} \tag{26}$$

Since ε was arbitrary, this proves the existence of the limit in (21), and therefore, completes Step (1) of the proof.

STEP (2): We prove that

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\biggl(-nh\biggl(\frac{1}{n}\Pi^{(n)}\biggr)\biggr)\le-\inf_{\mu\in\mathcal{M}_\infty}(\mathcal{H}(\mu)+h(\mu)). \tag{27}$$

Take a decreasing sequence $\varepsilon_k \to 0$ and let $L(\varepsilon_k) = (L_0(\varepsilon_k), L_1(\varepsilon_k), \ldots)$ be chosen as in Step (1) so that (26) holds with $\varepsilon = \varepsilon_k$. Fix arbitrary $\delta > 0$ and choose a sequence $(\mathcal{Q}_k)_{k=0}^{\infty} \subset \mathcal{P}(\mathcal{N}_\infty)$ such that

$$\mathcal{R}(\mathcal{Q}_k) + h(\Lambda(\mathcal{Q}_k^{[L(\varepsilon_k)]})) \le \inf_{\mu \in \mathcal{M}_\infty(L(\varepsilon_k))} (\mathcal{H}^{[L(\varepsilon_k)]}(\mu) + h(\mu)) + \delta. \tag{28}$$

Note that the existence of such Ψ_k follows immediately from (19).

We claim that $(\Psi_k)_{k=0}^{\infty}$ contains a subsequence weakly convergent in $\mathcal{P}(\mathcal{N}_\infty)$. Indeed, since $n^{-1} \log \mathbb{E} \exp(-nh(\Pi^{(n)}/n)) \ge -\|h\|_\infty$, we conclude from (26) and (28) that

$$\mathcal{R}(\mathcal{Q}_k) + h(\Lambda(\mathcal{Q}_k^{[L(\varepsilon_k)]})) \le \|h\|_\infty + 4C\varepsilon_k + \delta,$$

so we obtain the uniform bound

$$\mathcal{R}(\mathcal{Q}_k) = R(\mathcal{Q}_k \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) \le 2\|h\|_\infty + 4C\varepsilon_1 + \delta. \tag{29}$$

We get our assertion because the level sets of the relative entropy $R(\cdot \,\|\, \mathbb{P}^{\Pi_{\lambda_+}})$ on the Polish space $\mathcal{P}(\mathcal{N}_\infty)$ are compact (see Dupuis and Ellis (1997), Lemma 1.4.3(c)).

Thus, we can assume without loss of generality that the sequence $\mathcal{Q}_k$ converges weakly in $\mathcal{P}(\mathcal{N}_\infty)$ to some $\mathcal{Q}_\infty$. Further, using (29) and applying Lemma 2 we conclude that $\mathcal{Q}_\infty$ has a boundedly finite intensity measure $\Lambda(\mathcal{Q}_\infty)$ and

$$\Lambda(\mathcal{Q}_k) \to \Lambda(\mathcal{Q}_\infty) \quad \text{vaguely.}$$

Hence, using the lower semicontinuity of $\mathcal{R}$, continuity of h and (28) we get

$$\liminf_{k\to\infty} \inf_{\mu \in \mathcal{M}_\infty(L(\varepsilon_k))} (\mathcal{H}^{[L(\varepsilon_k)]}(\mu) + h(\mu)) \ge \mathcal{R}(\mathcal{Q}_\infty) + h(\Lambda(\mathcal{Q}_\infty)) - \delta.$$

In view of (2) and (14) this yields

$$\liminf_{k\to\infty} \inf_{\mu \in \mathcal{M}_\infty(L(\varepsilon_k))} (\mathcal{H}^{[L(\varepsilon_k)]}(\mu) + h(\mu)) \ge \inf_{\mu \in \mathcal{M}_\infty} (\mathcal{H}(\mu) + h(\mu)) - \delta.$$

Using (26) and letting $\delta \to 0$ we complete the proof of (27).

Step (3): The final step of the proof. We establish the converse inequality to (27), namely

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E} \exp\left(-nh\left(\frac{1}{n}\Pi^{(n)}\right)\right) \ge -\inf_{\mu \in \mathcal{M}_\infty} (\mathcal{H}(\mu) + h(\mu)). \tag{30}$$

Fix some $\delta > 0$ and choose $\nu \in \mathcal{M}_\infty$ so that

$$\mathcal{H}(\nu) + h(\nu) = \mathcal{R}(\mathbb{P}^{\Pi_\nu}) + h(\nu) \le \inf_{\mu \in \mathcal{M}_\infty} (\mathcal{H}(\mu) + h(\mu)) + \delta. \tag{31}$$

Note that the first equality is a consequence of Lemma 1. Further, take a decreasing sequence $\varepsilon_k \to 0$ and let $L(\varepsilon_k) = (L_0(\varepsilon_k), L_1(\varepsilon_k), \ldots)$ be as in Step (1) so that (25) is satisfied for $\varepsilon = \varepsilon_k$ and $L_0(\varepsilon_k) \to \infty$ as $k \to \infty$. It is easily verified that

$$\Lambda_{\Pi_\nu^{[L(\varepsilon_k)]}} \to \nu \quad \text{vaguely.} \tag{32}$$

On the other hand, by (19),

$$\inf_{\mu\in\mathcal{M}_\infty(L(\varepsilon_k))} (\mathcal{H}^{[L(\varepsilon_k)]}(\mu) + h(\mu)) \le \mathcal{H}^{[L(\varepsilon_k)]}(\Lambda_{\Pi_\nu^{[L(\varepsilon_k)]}}) + h(\Lambda_{\Pi_\nu^{[L(\varepsilon_k)]}})$$
$$\le \mathcal{R}(\mathbb{P}^{\Pi_\nu}) + h(\Lambda_{\Pi_\nu^{[L(\varepsilon_k)]}}).$$

Thus, using (32) and the continuity of h we obtain

$$\limsup_{k\to\infty} \inf_{\mu\in\mathcal{M}_\infty(L(\varepsilon_k))} (\mathcal{H}^{[L(\varepsilon_k)]}(\mu) + h(\mu)) \le \mathcal{R}(\mathbb{P}^{\Pi_\nu}) + h(\nu).$$

Applying (31) we get

$$\limsup_{k\to\infty} \inf_{\mu\in\mathcal{M}_\infty(L(\varepsilon_k))} (\mathcal{H}^{[L(\varepsilon_k)]}(\mu) + h(\mu)) \le \inf_{\mu\in\mathcal{M}_\infty} (\mathcal{H}(\mu) + h(\mu)) + \delta.$$

Letting $\delta \to 0$ and taking into account (25) establishes (30). This completes the proof of Step (3).

Combining Steps (1)–(3) gives (21). The proof of Lemma 3 is complete. ∎

We are now ready to prove our main theorem.

2.4. *Proof of Theorem 1.* The first step of the proof is to establish a suitable representation for the Cox processes N_n. As in Lemma 3, we denote by $\Pi^{(n)}$ the homogeneous Poisson point process on $[0,\infty)$ with intensity n. In addition we require it to be independent of M_n.

Using (3) it can be verified that the point process $\Pi^{(n)} \circ (M_n/n)$ has the same distribution as N_n. Indeed, take some $\gamma \in \mathcal{M}_\infty$ and note that for $0 \le t_1 < t_2$,

$$\left[\Pi^{(n)} \circ \frac{\gamma}{n}\right]((t_1,t_2]) = \Pi^{(n)}\left(\left(\frac{\gamma}{n}([0,t_1]), \frac{\gamma}{n}([0,t_2])\right]\right)$$

and hence the integer-valued random variable $[\Pi^{(n)} \circ \gamma/n]((t_1,t_2])$ has the Poisson distribution $\textsc{Poiss}(\gamma((t_1,t_2]))$.

Further, observe also that for disjoint intervals $I_1 = (t_1^{(1)}, t_2^{(1)}], \ldots, I_k = (t_1^{(k)}, t_2^{(k)}]$ the random variables $[\Pi^{(n)} \circ \gamma/n](I_1), \ldots, [\Pi^{(n)} \circ \gamma/n](I_k)$ are jointly independent. This means that $\Pi^{(n)} \circ \gamma/n$ is the Poisson point process with intensity measure γ. Using the independence of $\Pi^{(n)}$ and M_n we finally obtain

$$N_n \stackrel{\mathrm{d}}{=} \Pi^{(n)} \circ (M_n/n)$$

and hence

$$(N_n/n) \stackrel{\mathrm{d}}{=} (\Pi^{(n)}/n) \circ (M_n/n). \tag{33}$$

The particular convenience of the representation given by the above formula is that it expresses the Cox process N_n in terms of two independent random elements: $\Pi^{(n)}$ and M_n. This suggests the following proof strategy. In Lemma 3 we have established the large deviation principle for $(\Pi^{(n)}/n)$ on $\mathcal{M}_\infty$ with the rate function $\mathcal{H}(\mu) = R(\mathbb{P}^{\Pi_\mu} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}})$. Further, in view of condition (B) the random measures (M_n/n) almost surely take their values in $\mathcal{M}_\infty^{(\kappa)}$ defined to be the space of all measures $\mu \in \mathcal{M}_\infty$ such that $\mu(A) \geq \kappa(A)$ for all Borel $A \subset [0,\infty)$ with κ as in (B). Since the vague convergence preserves the inequalities $\mu(K) \geq \kappa(K)$ for compact $K \subseteq [0,\infty)$ and we have $\mu(A) = \sup\{\mu(K) \mid K \subseteq A,\ K \text{ compact}\}$ for each Borel measure μ on $[0,\infty)$ and Borel set $A \subseteq [0,\infty)$, it follows that $\mathcal{M}_\infty^{(\kappa)}$ is closed in the Polish space $\mathcal{M}_\infty$ and hence it is Polish itself.

Thus, by condition (L) and by the independence of $\Pi^{(n)}$ and M_n we conclude, using Lemma 4.1.5(b), Exercise 4.2.7 and Exercise 4.1.10(c) in Dembo and Zeitouni (1993), that the sequence $((\Pi^{(n)}/n), (M_n/n))_{n=1}^\infty$ satisfies on $\mathcal{M}_\infty \times \mathcal{M}_\infty^{(\kappa)}$ the large deviation principle with the rate function

$$\mathcal{K}(\mu,\gamma) = \mathcal{H}(\mu) + I(\gamma). \tag{34}$$

In the Appendix, Proposition 1, we show that the mapping $\circ : \mathcal{M}_\infty \times \mathcal{M}_\infty^{(\kappa)} \to \mathcal{M}_\infty$ is continuous (note that in general $\circ : \mathcal{M}_\infty \times \mathcal{M}_\infty \to \mathcal{M}_\infty$ is not continuous). The assertion of Theorem 1 now follows from (34) by the standard contraction principle (see Theorem 1.3.2 in Dupuis and Ellis (1997) or Proposition 2 in the Appendix). The proof is complete. ■

2.5. *Proof of Corollary 1.* Pick arbitrary $\theta \in \mathcal{M}_\infty$ and suppose that $J(\theta) < \infty$. Then there exist $\gamma \in \mathcal{M}_\infty$ and $\alpha \geq 0$ such that

$$\theta = \mu \circ (\alpha\lambda_+). \tag{35}$$

We claim that under such assumptions we have

$$\alpha = \lim_{T\to\infty} \theta([0,T])/T, \tag{36}$$

so that in particular the limit exists. Indeed, observe that $\mathcal{H}(\mu) \leq J(\theta) < \infty$ and hence, recalling that $\vartheta(u) = u\log u + 1 - u$, we conclude from (6) that $\mu \ll \lambda_+$ and $\int_0^\infty \vartheta(d\mu/d\lambda_+)\,d\lambda_+ < \infty$. Taking into account the convexity of ϑ and applying Jensen's inequality, for $T > 0$ we get

$$T\vartheta(\mu([0,T])/T) \leq \int_0^T \vartheta\left(\frac{d\mu}{d\lambda_+}\right) d\lambda_+ < \infty.$$

Now, simple analysis yields $\lim_{T\to\infty} \mu([0,T])/T = 1$. Therefore

$$\theta([0,T])/T = \mu([0,\alpha T])/T \to \alpha,$$

which proves (36).

In particular, if $\alpha = 0$, we obtain $\theta([0,T]) = \mu(\{0\}) = 0$ for all T (the last equality follows from $\mu \ll \lambda_+$) so that $\theta \equiv 0$. Moreover, the uniqueness of α implies that $J(\theta) \geq \widehat{I}(0)$. However, by (B) we have $\widehat{I}(0) = \infty$, which contradicts our original assumption that $J(\theta) < \infty$. This means that $0 < \alpha < \infty$ and hence μ in (35) is uniquely determined by the relation $\mu = \theta \circ (\lambda_+/\alpha)$. In view of (4) and (6) the assertion of the corollary now follows by a straightforward calculation. ∎

3. Appendix. Below we give proofs of some technical details, omitted in the main body of the article. For the convenience of the reader we quote as well certain standard results of the large deviation theory, extensively used in the paper.

3.1. *Proof of Lemma 2.* Take an arbitrary sequence $(\Psi_k)_{k=1}^{\infty}$ of point processes on $[0,\infty)$ convergent in distribution to some point process Ψ_∞ and such that

$$R(\mathbb{P}^{\Psi_k} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) \leq C \tag{37}$$

for some constant $C \geq 0$. Fix some continuous function $f : [0,\infty) \to \mathbb{R}_+$ with compact support and for $k = 1,\ldots,\infty$ define the random variables

$$\eta_k := \int_{[0,\infty)} f(x)\, d\Psi_k(x).$$

Since Ψ_k converges in distribution to Ψ_∞, also η_k converges in distribution to η_∞ for $k \to \infty$. Further, it follows from (37) that

$$R\Big(\eta_k \,\Big\|\, \int_{[0,\infty)} f(x)\, d\Pi_{\lambda_+}(x)\Big) \leq C.$$

Thus, we deduce from Lemma 1.4.3(d) in Dupuis and Ellis (1997) that the sequence $(\eta_k)_{k=1}^{\infty}$ is uniformly integrable, i.e.

$$\lim_{l\to\infty} \sup_k \mathbb{E}(\mathbf{1}_{\{\eta_k > l\}}\eta_k) = 0.$$

In particular, $\mathbb{E}\eta_k < \infty$ and

$$\lim_{k\to\infty} \mathbb{E}\eta_k = \mathbb{E}\eta_\infty.$$

Taking into account that $\int_{[0,\infty)} f(x)\, d[\Lambda(\Psi_k)](x) = \mathbb{E}\eta_k$ for $k = 1,\ldots,\infty$ we see that $\int_{[0,\infty)} f(x)\, d[\Lambda_{\Psi_k}](x) < \infty$ and

$$\lim_{k\to\infty} \int_{[0,\infty)} f(x)\, d[\Lambda_{\Psi_k}](x) = \int_{[0,\infty)} f(x)\, d[\Lambda_{\Psi_\infty}](x).$$

Since $f : [0,\infty) \to \mathbb{R}_+$ was arbitrary we conclude that Ψ_k, $k = 1,\ldots,\infty$, have boundedly finite intensity measures and

$$\Lambda_{\Psi_k} \to \Lambda_{\Psi_\infty} \quad \text{vaguely,}$$

which proves the first assertion of the lemma.

The weak compactness of $\{\mathcal{Q} \in \mathcal{P}(\mathcal{N}_\infty) \mid \Lambda(\mathcal{Q}) = \mu,\ R(\mathcal{Q} \,\|\, \mathbb{P}^{\Pi_{\lambda^+}}) \le C\}$ now follows from the first part of the lemma and from the weak compactness of the entropy level set $\{\mathcal{Q} \in \mathcal{P}(\mathcal{N}_\infty) \mid R(\mathcal{Q} \,\|\, \mathbb{P}^{\Pi_{\lambda^+}}) \le C\}$ for $C \ge 0$ (see Lemma 1.4.3(c) in Dupuis and Ellis (1997)). The proof of Lemma 2 is complete. ∎

3.2. *Continuity properties of* $\circ$. Below we establish the continuity of the operation $\circ$ restricted to $\mathcal{M}_\infty \times \mathcal{M}_\infty^{(\kappa)}$ as required in the proof of Theorem 1.

PROPOSITION 1. *The mapping* $\circ : \mathcal{M}_\infty \times \mathcal{M}_\infty^{(\kappa)} \to \mathcal{M}_\infty$ *is continuous.*

Proof. Take arbitrary sequences $(\mu_n)_{n=1}^\infty \subset \mathcal{M}_\infty$ and $(\gamma_n)_{n=1}^\infty \subset \mathcal{M}_\infty^{(\kappa)}$ converging vaguely to $\mu_\infty \in \mathcal{M}_\infty$ and $\gamma_\infty \in \mathcal{M}_\infty^{(\kappa)}$ respectively. Let $\theta_n = \mu_n \circ \gamma_n$ for all $n \in \mathbb{N} \cup \{\infty\}$. Clearly, to prove our assertion it suffices to show that

$$\lim_{n\to\infty} \theta_n([0,\alpha]) = \theta_\infty([0,\alpha]) \tag{38}$$

for all α except for a countable set. Indeed, we then have, integrating by parts,

$$\int_{[0,\infty)} f(x)\, d\theta_n(x) = -\int_{[0,\infty)} f'(x)\theta_n([0,x])\, dx,$$

which converges as $n \to \infty$ to

$$-\int_{[0,\infty)} f'(x)\theta_\infty([0,x])\, dx = \int_{[0,\infty)} f(x)\, d\theta_\infty(x)$$

for each continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$ with compact support (not necessarily contained in $[0,\infty)$). Since the family of such functions is uniformly dense in the class of continuous functions on $[0,\infty)$ with compact support, the required vague convergence follows by the definition.

To proceed, denote by $\mathcal{A}$ and $\mathcal{B}$ the sets of atoms of γ_∞ and μ_∞ respectively and let $\mathcal{C}$ be the set of all $\alpha \in [0,\infty)$ for which $\gamma_\infty([0,\alpha]) \in \mathcal{B}$. Obviously, both $\mathcal{A}$ and $\mathcal{B}$ are at most countable. Further, by the definition of $\mathcal{M}_\infty^{(\kappa)}$, the function $\alpha \mapsto \gamma_\infty([0,\alpha])$ is strictly increasing and hence injective. This means that also $\mathcal{C}$ has to be at most countable (note that this is the moment where condition (B) intervenes in a relevant way, for in general there is no reason for $\mathcal{C}$ to be countable). Now take some $\alpha_0 \in (0,\infty)\setminus(\mathcal{A}\cup\mathcal{C})$. Then the assumption that $\alpha_0 \notin \mathcal{A}$ and the vague convergence $\gamma_n \to \gamma_\infty$ give

$$\lim_{n\to\infty} \gamma_n([0,\alpha_0]) = \gamma_\infty([0,\alpha_0]).$$

In addition, in view of (B), $\alpha_0 > 0$ yields $\beta_0 := \gamma_\infty([0,\alpha_0]) > 0$. Thus, by the vague convergence $\mu_n \to \mu_\infty$ we can choose arbitrarily small $0 < \varepsilon < \beta_0$

such that $\lim_{n\to\infty} \mu_n([0,\beta_0-\varepsilon]) = \mu_\infty([0,\beta_0-\varepsilon])$ and $\lim_{n\to\infty} \mu_n([0,\beta_0+\varepsilon]) = \mu_\infty([0,\beta_0+\varepsilon])$. Take n_0 such that

$$|\gamma_n([0,\alpha_0]) - \beta_0| < \varepsilon$$

for $n > n_0$. Hence, for $n > n_0$,

$$\mu_n([0,\beta_0 - \varepsilon]) \le \mu_n([0,\gamma_n([0,\alpha_0])]) = \theta_n(\alpha_0) \le \mu_n([0,\beta_0+\varepsilon]).$$

By our assumptions, letting $n \to \infty$ leads to

$$\mu_\infty([0,\beta_0-\varepsilon]) \le \liminf_{n\to\infty} \theta_n(\alpha_0) \le \limsup_{n\to\infty} \theta_n(\alpha_0) \le \mu_\infty([0,\beta_0+\varepsilon]).$$

Since $\alpha_0 \notin \mathcal{C}$, we have $\beta_0 \notin \mathcal{B}$ and hence the mapping $\beta \mapsto \mu_\infty([0,\beta])$ is continuous at β_0. Therefore, taking $\varepsilon \to 0$ we finally conclude that

$$\lim_{n\to\infty} \theta_n([0,\alpha_0]) = \mu_\infty([0,\beta_0]) = \theta_\infty(\alpha_0).$$

Thus, we have proved (38) for all $\alpha \in [0,\infty)$ outside the countable set $\mathcal{A} \cup \mathcal{C}$. This yields the required continuity of $\circ : \mathcal{M}_\infty \times \mathcal{M}_\infty^{(\kappa)} \to \mathcal{M}_\infty$. ∎

3.3. *Contraction principle.* A very useful property of the general large deviation principles is that they are covariant with respect to continuous mappings. This fact, formulated below as a proposition, is usually referred to as the *contraction principle*; see Dupuis and Ellis (1997), Theorem 1.3.2.

PROPOSITION 2. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish metric spaces and X_n a sequence of $\mathcal{X}$-valued random elements satisfying a large deviation principle with a* (*good*) *rate function I. Then for each continuous mapping $f : \mathcal{X} \to \mathcal{Y}$ the sequence of $\mathcal{Y}$-valued random elements $f(X_n)$ satisfies the large deviation principle with good rate function J given by $J(y) := \inf\{I(x) \mid x \in \mathcal{X},\ f(x) = y\}$ for $y \in \mathcal{Y}$.*

3.4. *Sanov's theorem.* Sanov's theorem is one of the classical results which gave rise to the large deviation theory in its present form. We quote it here for the sake of completeness (see Theorem 2.2.1 in Dupuis and Ellis (1997)).

PROPOSITION 3. *Let $X_1, X_2, \ldots$ be i.i.d. random elements*, *taking values in a Polish metric space $\mathcal{X}$*, *with common distribution γ on $\mathcal{X}$. Denote by $\widehat{\theta}_n$ the empirical measure corresponding to the sample $X_1,\ldots,X_n$*, *i.e. $\widehat{\theta}_n := n^{-1}\sum_{k=1}^n \delta_{X_k}$ with δ_x standing for the Dirac unit mass concentrated at x. Then $\widehat{\theta}_n$ satisfies on $\mathcal{P}(\mathcal{X})$ the large deviation principle with rate function given by the relative entropy $R(\cdot \,\|\, \gamma)$*, *where $\mathcal{P}(\mathcal{X})$ denotes the space of Borel probability measures on $\mathcal{X}$ endowed with the usual weak topology.*

3.5. *Proof of relation* (9). In order to shorten the formulae, we write $\sigma, \sigma_A, \sigma_B, \pi, \pi_A, \pi_B$ for the distributions on $\mathcal{N}_\infty$ of the point processes $\widehat{\Psi}, \widehat{\Psi}_{|A}, \widehat{\Psi}_{|B}, \Pi_{\lambda_+}, (\Pi_{\lambda_+})_{|A}, (\Pi_{\lambda_+})_{|B}$ respectively. Moreover, we use μ_A and μ_B

to denote generic measures on $[0,\infty)$ concentrated on A and B respectively. Finally, to streamline the proof, denote by $\sigma_{A\to B}(\cdot|\cdot)$ the regular conditional distribution of $\widehat{\Psi}_{|B}$ given $\widehat{\Psi}_{|A}$ and by $\pi_{A\to B}(\cdot|\cdot)$ the regular conditional distribution of $(\Pi_{\lambda_+})_{|B}$ given $(\Pi_{\lambda_+})_{|A}$.

By the standard properties of Poisson point processes we can choose a version of $\pi_{A\to B}$ so that

$$\pi_{A\to B}(\cdot|\mu_A) = \pi_B \tag{39}$$

for all $\mu_A \in \mathcal{N}_\infty$. Using the definition of the relative entropy we write

$$\begin{aligned}
R(\mathbb{P}^{\widehat{\Psi}} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) &= \int_{\mathcal{N}_\infty} \log\Big(\frac{d\sigma}{d\pi}(\mu)\Big) d\sigma(\mu) \\
&= \int_{\mathcal{N}_\infty}\int_{\mathcal{N}_\infty} \log\bigg(\frac{d\sigma}{d\pi}(\mu_A+\mu_B)\bigg)\, d\sigma_{A\to B}(\mu_B|\mu_A)\, d\sigma_A(\mu_A) \\
&= \int_{\mathcal{N}_\infty}\int_{\mathcal{N}_\infty} \log\bigg(\frac{d\sigma_{A\to B}}{d\pi_{A\to B}}(\mu_B|\mu_A)\frac{d\sigma_A}{d\pi_A}(\mu_A)\bigg)\, d\sigma_{A\to B}(\mu_B|\mu_A)\, d\sigma_A(\mu_A)
\end{aligned}$$

and hence, by (39),

$$R(\mathbb{P}^{\widehat{\Psi}} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) = R(\sigma_A \,\|\, \pi_A) + \int_{\mathcal{N}_\infty} R(\sigma_{A\to B}(\cdot|\mu_A) \,\|\, \pi_B)\, d\sigma_A(\mu_A).$$

Thus, in view of the strict convexity of the relative entropy mapping $\mathcal{P}(\mathcal{N}_\infty) \ni \mathcal{Q} \mapsto R(\mathcal{Q} \,\|\, \pi_B)$ on its domain of finiteness (see Lemma 1.4.3(c) in Dupuis and Ellis (1997)) we get, using Jensen's inequality,

$$R(\mathbb{P}^{\widehat{\Psi}} \,\|\, \mathbb{P}^{\Pi_{\lambda_+}}) \ge R(\sigma_A \,\|\, \pi_A) + R(\sigma_B \,\|\, \pi_B)$$

with equality iff

$$\sigma_{A\to B}(\cdot|\mu_A) = \sigma_B$$

for σ_A-almost all μ_A. This completes the proof of (9). ∎

Acknowledgments. I wish to thank the anonymous referee, whose remarks have been helpful in improving this paper. I also gratefully acknowledge the support of the Foundation for Polish Science (FNP).

References

J. A. Bucklew (1990), *Large Deviation Techniques in Decision, Simulation and Estimation*, Wiley, New York.

D. J. Daley and D. Vere-Jones (1988), *An Introduction to the Theory of Point Processes*, Springer, New York.

A. Dembo and O. Zeitouni (1993), *Large Deviations Techniques and Applications*, Jones and Bartlett, London.

J.-D. Deuschel and D. W. Stroock (1989), *Large Deviations*, Academic Press, Boston.

P. Dupuis and R. S. Ellis (1997), *A Weak Convergence Approach to the Theory of Large Deviations*, Wiley, Chichester.

S. Ellis (1985), *Entropy, Large Deviations and Statistical Mechanics*, Springer, New York.

D. Florens and H. Pham (1998), *Large deviation probabilities in estimation of Poisson random measures*, Stochastic Process. Appl. 76, 117–139.

H.-O. Georgii (1988), *Gibbs Measures and Phase Transitions*, Walter de Gruyter, Berlin.

P. Groeneboom, J. Oosterhoff and F. H. Ruymgaart (1979), *Large deviation theorems for empirical probability measures*, Ann. Probab. 7, 553–586.

A. F. Karr (1991), *Point Processes and Their Statistical Inference*, Probab. Pure Appl. 7 Marcel Dekker, New York.

J. Kerstan, K. Matthes and J. Mecke (1982), *Infinitely Divisible Point Processes*, Nauka, Moscow (Russian edition).

C. Léonard (2000), *Large deviations for Poisson random measures and processes with independent increments*, Stochastic Process. Appl. 85, 93–121.

I. N. Sanov (1957), *On the probability of large deviations of random variables*, Mat. Sb. (N.S.) 42 (84), 11–44 (in Russian); English transl. in: Selected Translations in Mathematical Statistics and Probability, Vol. 1, Inst. Math. Statist. and Amer. Math. Soc., Providence, 1961, 213–244.

Faculty of Mathematics and Computer Science
Nicolaus Copernicus University
Chopina 12/18
87-100 Toruń, Poland
E-mail: tomeks@mat.uni.torun.pl

Received 24 September 2001;
revised 15 June 2003 (4110)