Confluentes Mathematici, Vol. 1, No. 2 (2009) 181–196

DIFFERENTIABILITY-FREE CONDITIONS ON THE FREE-ENERGY FUNCTION IMPLYING LARGE DEVIATIONS

HENRI COMMAN

Institute of Mathematics
Pontifical Catholic University of Valparaiso, Chile
henri.comman@ucv.cl

Received 31 July 2009
Revised 31 July 2009

Let (μ_α) be a net of Radon sub-probability measures on $\mathbb{R}$, and (t_α) be a net in $]0,1]$ converging to 0. Assuming that the generalized log-moment generating function $L(\lambda)$ exists for all λ in a nonempty open interval G, we give conditions on the left or right derivatives of $L_{|G}$, implying a vague (and thus narrow when $0 \in G$) large deviation principle. The rate function (which can be nonconvex) is obtained as an abstract Legendre–Fenchel transform. This allows us to strengthen the Gärtner–Ellis theorem by weakening the essential smoothness assumption. A related question of R. S. Ellis is solved.

Keywords: Radon sub-probability measures; generalized log-moment generating function; vague large deviations; Gärtner–Ellis theorem.

AMS Subject Classification: 60F10

1. Introduction

Let (μ_α) be a net of Radon sub-probability measures on a Hausdorff topological space X, and (t_α) be a net in $]0,1]$ converging to 0. Let $\mathcal{B}(X)$ (resp. $\mathcal{C}(X)$) denote the set of $[-\infty,+\infty[$-valued Borel measurable (resp. continuous) functions on X. For each $h \in \mathcal{B}(X)$, we define

$$\underline{\Lambda}(h) = \log \liminf \mu_\alpha^{t_\alpha}(e^{h/t_\alpha})$$

and

$$\overline{\Lambda}(h) = \log \limsup \mu_\alpha^{t_\alpha}(e^{h/t_\alpha}),$$

where $\mu_\alpha^{t_\alpha}(e^{h/t_\alpha})$ stands for $(\int_X e^{h(x)/t_\alpha}\mu_\alpha(dx))^{t_\alpha}$, and write $\Lambda(h)$ when both expressions are equal. When $X = \mathbb{R}$, for each pair of reals (λ,ν), let $h_{\lambda,\nu}$ be the function defined on X by $h_{\lambda,\nu}(x) = \lambda x$ if $x \leq 0$ and $h_{\lambda,\nu}(x) = \nu x$ if $x \geq 0$ (we write simply h_λ in place of $h_{\lambda,\lambda}$). For each real λ, we put $L(\lambda) = \Lambda(h_\lambda)$ when $\Lambda(h_\lambda)$ exists.

A well-known problem of large deviations in $\mathbb{R}$ (usually stated for sequences of probability measures) is the following: assuming that $L(\lambda)$ exists and is finite for all λ in an open interval G containing 0, and that the map $L_{|G}$ is not differentiable on G, what conditions on $L_{|G}$ imply large deviations, and with which rate function?

In relation with this problem, R. S. Ellis posed the following question ([4]): assuming that $\Lambda(h_{\lambda,\nu})$ exists and is finite for all $(\lambda,\nu)\in\mathbb{R}^2$, what conditions on the functional $\Lambda_{|\{h_{\lambda,\nu}:(\lambda,\nu)\in\mathbb{R}^2\}}$ imply large deviations with rate function $J(x) = \sup_{(\lambda,\nu)\in\mathbb{R}^2}\{h_{\lambda,\nu}(x)-\Lambda(h_{\lambda,\nu})\}$ for all $x\in X$?

In this paper, we solve the above problem by giving conditions on $L_{|G}$ involving only its left and right derivatives; the rate function is obtained as an abstract Legendre–Fenchel transform $\Lambda^*_{|\mathcal{S}}$, where $\mathcal{S}$ can be any set in $\mathcal{C}(X)$ containing $\{h_\lambda : \lambda\in G\}$ (Theorem 3). When $\mathcal{S}=\{h_\lambda:\lambda\in G\}$, we get a strengthening of the Gärtner–Ellis theorem by removing the essential smoothness assumption (Corollary 1). Taking $\mathcal{S}=\{h_{\lambda,\nu}:(\lambda,\nu)\in\mathbb{R}^2\}$ gives an answer to the Ellis question (Corollary 2).

The techniques used are refinements of those developed in previous author's works ([1, 2]), where variational forms for $\underline{\Lambda}(h)$ and $\overline{\Lambda}(h)$ are obtained with $h\in\mathcal{B}(X)$ satisfying the usual Varadhan's tail condition (X a general space). We consider here the set $\mathcal{C}_{\mathcal{K}}(X)$ of elements h in $\mathcal{C}(X)$ for which $\{y\in X : e^{h(x)}-\varepsilon\le e^{h(y)}\le e^{h(x)}+\varepsilon\}$ is compact for all $x\in X$ and $\varepsilon>0$ with $e^{h(x)}>\varepsilon$. The first step is Theorem 2, which establishes that for any $\mathcal{T}\subset\mathcal{C}_{\mathcal{K}}(X)$, and under suitable conditions (weaker than vague large deviations), there exist some reals m, M such that

$$\Lambda(h)=\sup_{x\in\{m\le h\le M\}}\{h(x)-l_1(x)\}\quad\text{for all } h\in\mathcal{T},$$

where $l_1(x) = -\log\inf\{\liminf\mu_\alpha^{t_\alpha}(G) : x\in G\subset X, G \text{ open}\}$ for all $x\in X$; in particular, $\Lambda(h)$ exists and has the same form as when large deviations hold. Note that when $X=\mathbb{R}$ and $\mathcal{T}=\{h_\lambda:\lambda\in G\}$ with $0\notin G$, then the sup in the above expression can be taken on a compact set (if $0\in G$, this follows from the exponential tightness). It turns out that any subnet of $(\mu_\alpha^{t_\alpha})$ has a subnet $(\mu_\gamma^{t_\gamma})$ satisfying the above conditions. The second step consists in applying Theorem 2 with $X=\mathbb{R}$, $\mathcal{T}=\{h_\lambda:\lambda\in G\}$ and all these subnets. More precisely, we show that if x is the left or right derivative of L at some point $\lambda_x\in G$, then $l_1^{(\mu_\gamma^{t_\gamma})}(x)\le\lambda_x x-L(\lambda_x)$, whence

$$l_1^{(\mu_\gamma^{t_\gamma})}(x)\le L^*_{|G}(x) \tag{1}$$

(Proposition 1). Let $\mathcal{S}$ be any set in $\mathcal{C}(X)$ containing $\{h_\lambda:\lambda\in G\}$, and assume that $\Lambda(h)$ exists for all $h\in\mathcal{S}$. It is easy to see that

$$L^*_{|G}\le\Lambda^*_{|\mathcal{S}}\le l_0^{(\mu_\gamma^{t_\gamma})}\le l_1^{(\mu_\gamma^{t_\gamma})}, \tag{2}$$

where $l_0^{(\mu_\gamma^{t_\gamma})}(x) = -\log\inf\{\limsup \mu_\gamma^{t_\gamma}(G) : x \in G \subset X, G \text{ open}\}$ for all $x \in X$. Putting together (1) and (2) give

$$L^*_{|G}(x) = \Lambda^*_{|\mathcal{S}}(x) = l_0^{(\mu_\gamma^{t_\gamma})}(x) = l_1^{(\mu_\gamma^{t_\gamma})}(x) \tag{3}$$

for all x in the image of the left (resp. right) derivative of $L_{|G}$; consequently, if the set of these images contains $\{\Lambda^*_{|\mathcal{S}} < +\infty\}$, then $(\mu_\gamma^{t_\gamma})$ satisfies a vague (narrow if $0 \in G$) large deviation principle with powers (t_γ) and rate function $\Lambda^*_{|\mathcal{S}}$, which moreover coincides with $L^*_{|G}$ on its effective domain. Using capacity theory, and in particular a compactness argument, we conclude that the same result holds for the net $(\mu_\alpha^{t_\alpha})$. Furthermore, $\{\Lambda^*_{|\mathcal{S}} < +\infty\}$ can be replaced by its interior, when $\Lambda^*_{|\mathcal{S}}$ is proper convex and lower semi-continuous, which is the case when $\mathcal{S} = \{h_\lambda : \lambda \in G\}$; this allows us to improve a strong version of Gärtner–Ellis theorem given by O'Brien.

Various generalizations are given in order to get large deviations with a rate function coinciding with $\Lambda^*_{|\mathcal{S}}$ and $L^*_{|G}$ only on its effective domain. Note that all our results hold for general nets of sub-probability measures and powers.

The paper is organized as follows. Section 2 fixes the notations and recall some results on large deviations and convexity; Sec. 3 deals with the variational forms of the functionals Λ; Sec. 4 treats the case $X = \mathbb{R}$.

2. Preliminaries

Without explicit mention, X denotes a Hausdorff topological space, (μ_α) a net of Radon sub-probability measures on X, and (t_α) a net in $]0,1]$ converging to 0. Throughout the paper, the notations $\underline{\Lambda}$, $\overline{\Lambda}$, Λ, l_0, l_1 (introduced in Sec. 1) refer to the net $(\mu_\alpha^{t_\alpha})$. We shall write $l_1^{(\mu_\beta^{t_\beta})}$ when in the definition of l_1, $(\mu_\alpha^{t_\alpha})$ is replaced by the subnet $(\mu_\beta^{t_\beta})$. We do not make such distinction for the map Λ, since it does not depend on the subnet along which the limit is taken. We recall that l_0 and l_1 are lower semi-continuous functions.

Definition 1. (a) (μ_α) satisfies a (*narrow*) *large deviation principle* with powers (t_α) if there exists a $[0,+\infty]$-valued lower semi-continuous function J on X such that

$$\limsup \mu_\alpha^{t_\alpha}(F) \le \sup_{x\in F} e^{-J(x)} \quad \text{for all closed } F \subset X \tag{4}$$

and

$$\sup_{x\in G} e^{-J(x)} \le \liminf \mu_\alpha^{t_\alpha}(G) \quad \text{for all open } G \subset X;$$

J is a *rate function* for $(\mu_\alpha^{t_\alpha})$, which is said to be tight when it has compact level sets. When "closed" is replaced by "compact" in (4), we say that a *vague large deviation principle* holds.
(b) (μ_α) is *exponentially tight with respect to* (t_α) if for each $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset X$ such that $\limsup \mu_\alpha^{t_\alpha}(X\backslash K_\varepsilon) < \varepsilon$.

The following results are well known for a net $(\mu_\varepsilon^\varepsilon)_{\varepsilon>0}$, with μ_ε a Radon probability measure ([3]); it is easy to see that the proofs work also for general nets of sub-probability measures and powers.

Lemma 1. (a) *Let X be locally compact Hausdorff. Then, (μ_α) satisfies a vague large deviation principle with powers (t_α) if and only if $l_0 = l_1$. In this case, l_0 is the rate function.*
(b) *If (μ_α) satisfies a vague large deviation principle with powers (t_α), and (μ_α) is exponentially tight with respect to (t_α), then (μ_α) satisfies a large deviation principle with same powers and same rate function.*

A capacity on X is a map c from the powerset of X to $[0, +\infty]$ such that:

(i) $c(\emptyset) = 0$.
(ii) $c(Y) = \sup\{c(K) : K \subset Y, K \text{ compact}\}$ for all $Y \subset X$.
(iii) $c(K) = \inf\{c(G) : K \subset G \subset X, G \text{ open}\}$ for all compact $K \subset X$.

The vague topology on the set of capacities is the coarsest topology for which the maps $c \to c(Y)$ are upper (resp. lower) semi-continuous for all compact (resp. open) $Y \subset X$. Let $\Gamma(X, [0,1])$ denote the set of $[0,1]$-valued capacities on X provided with the vague topology. It is clear that any Radon sub-probability measure and any power of such a measure by a positive number less than 1 belong to $\Gamma(X, [0,1])$, so that $(\mu_\alpha^{t_\alpha})$ is a net in $\Gamma(X, [0,1])$. For each $[0, +\infty]$-valued lower semi-continuous function l on X, we associate the element c_l in $\Gamma(X, [0,1])$ defined by $c_l(Y) = \sup_{x\in Y} e^{-l(x)}$ for all $Y \subset X$. We refer to [9] for the first assertion in the following lemma; the second one is the mere transcription of the definition of a vague large deviation principle in terms of capacities.

Lemma 2. (a) *If X is locally compact Hausdorff, then $\Gamma(X, [0,1])$ is a compact Hausdorff space.*
(b) *(μ_α) satisfies a vague large deviation principle with powers (t_α) and rate function J if and only if $(\mu_\alpha^{t_\alpha})$ converges to c_J in $\Gamma(X, [0,1])$.*

For any $[-\infty, +\infty]$-valued (not necessary convex) function f defined on some topological space, we put $\mathcal{D}\text{om}(f) = \{f < +\infty\}$ (the so-called effective domain), and denote by $\text{int}\mathcal{D}\text{om}(f)$ (resp. $\text{bd}\mathcal{D}\text{om}(f)$) the interior (resp. boundary) of $\mathcal{D}\text{om}(f)$. The range of f is denoted by $\text{ran}f$.

A $[-\infty, +\infty]$-valued convex function f on $\mathbb{R}$ is said to be proper if f is $]-\infty, +\infty]$-valued and takes a finite value on at least one point. The Legendre–Fenchel transform f^* of f is defined by $f^*(x) = \sup_{\lambda\in\mathbb{R}}\{\lambda x - f(\lambda)\}$ for all $x \in \mathbb{R}$; note that f^* is convex lower semi-continuous, and proper when f is proper. Let $I \subset \mathbb{R}$ be a nonempty interval, and $f_{|I}$ be a $]-\infty, +\infty]$-valued convex function on I. We denote by $\widehat{f_{|I}}$ the convex function on $\mathbb{R}$ which coincides with $f_{|I}$ on I, and takes the value $+\infty$ out of I; in this case we write simply $f_{|I}^*$ in place of $\widehat{f_{|I}}^*$. The left and right derivatives of $f_{|I}$ at some point $x \in \mathcal{D}\text{om}(f_{|I})$ are denoted by $f'_{|I_-}(x)$ and $f'_{|I_+}(x)$

respectively. A proper convex function f on $\mathbb{R}$ is said to be essentially smooth if $\mathrm{int}\mathcal{D}\mathrm{om}(f) \neq \emptyset$, f is differentiable on $\mathrm{int}\mathcal{D}\mathrm{om}(f)$, and $\lim |f'(x_n)| = +\infty$ for all sequences (x_n) in $\mathrm{int}\mathcal{D}\mathrm{om}(f)$ converging to some $x \in \mathrm{bd}\mathcal{D}\mathrm{om}(f)$ ([10]).

If $L(\lambda)$ exists and is finite for all λ in a nonempty open interval G, then $L_{|G}$ is convex; if moreover $0 \in G$, then (μ_α) is exponentially tight with respect to (t_α). If $L(\lambda)$ exists for all reals λ, then L is a $[-\infty, +\infty]$-valued convex function on $\mathbb{R}$; if moreover $0 \in \mathrm{int}\mathcal{D}\mathrm{om}(L)$, then L is proper (the proof of these facts is obtained by modifying suitably the one of Lemma 2.3.9 in [3]).

Lemma 3. *Let f be a proper convex lower semi-continuous function on $\mathbb{R}$. Then,*

$$\inf_{y \in G} f(y) = \inf_{y \in G \cap \mathrm{int}\mathcal{D}\mathrm{om}(f)} f(y)$$

for all open sets $G \subset \mathbb{R}$.

Proof. Let G be an open subset of $\mathbb{R}$. If $G \cap \mathcal{D}\mathrm{om}(f) = \emptyset$, then the conclusion holds trivially ($\inf \emptyset = +\infty$ by convention). Assume that $G \cap \mathcal{D}\mathrm{om}(f) \neq \emptyset$. By Corollary 6.3.2 of [10], $G \cap \mathrm{int}\mathcal{D}\mathrm{om}(f) \neq \emptyset$. By Theorem VI.3.2 of [5], for each $x \in \mathcal{D}\mathrm{om}(f)$ we can find a sequence (x_n) in $\mathrm{int}\mathcal{D}\mathrm{om}(f)$ converging to x and such that $\lim f(x_n) = f(x)$, which implies $\inf_{G \cap \mathcal{D}\mathrm{om}(f)} f = \inf_{G \cap \mathrm{int}\mathcal{D}\mathrm{om}(f)} f$, and the lemma is proved since $\inf_{G \cap \mathcal{D}\mathrm{om}(f)} f = \inf_G f$. □

3. Variational Forms for Λ on $\mathcal{C}_{\mathcal{K}}(X)$

We begin by defining a notion, which will appear as a key condition in the sequel; it is nothing else but a uniform version of the tail condition in Varadhan's theorem.

Definition 2. We say that a set $\mathcal{T} \subset \mathcal{B}(X)$ *satisfies the tail condition* for $(\mu_\alpha^{t_\alpha})$ if for each $\varepsilon > 0$, there exists a real M such that

$$\limsup \mu_\alpha^{t_\alpha}(e^{h/t_\alpha} 1_{\{h > M\}}) < \varepsilon \quad \text{for all } h \in \mathcal{T}.$$

For each $h \in \mathcal{B}(X)$, each $x \in X$ and each $\varepsilon > 0$, we put $F_{e^{h(x)},\varepsilon} = \{y \in X\colon e^{h(x)} - \varepsilon \leq e^{h(y)} \leq e^{h(x)} + \varepsilon\}$ and $G_{e^{h(x)},\varepsilon} = \{y \in X : e^{h(x)} - \varepsilon < e^{h(y)} < e^{h(x)} + \varepsilon\}$. The following expressions are known when (μ_α) is a net of probability measures, and when $\mathcal{T}$ has only one element, say h (see [1] and [2] for the first and the second assertions, respectively). The proofs reveal that the constant M comes from the above tail condition (assumed to be satisfied by h), so that the uniform versions for a general $\mathcal{T}$ follow immediately; they moreover work as well for the sub-probability case.

Theorem 1. *Let $\mathcal{T} \subset \mathcal{B}(X)$ satisfying the tail condition for $(\mu_\alpha^{t_\alpha})$. There is a real M such that for each $h \in \mathcal{T}$,*

$$e^{\underline{\Lambda}(h)} = \liminf \sup_{x \in X, \varepsilon > 0} \{(e^{h(x)} - \varepsilon)\mu_\alpha^{t_\alpha}(G_{e^{h(x)},\varepsilon})\}$$

$$= \lim_{\varepsilon \to 0} \liminf \sup_{x \in \{h \leq M\}} \{e^{h(x)}\mu_\alpha^{t_\alpha}(G_{e^{h(x)},\varepsilon})\}$$

and

$$e^{\overline{\Lambda}(h)} = \sup_{x\in X,\varepsilon>0}\{(e^{h(x)}-\varepsilon)\limsup \mu_\alpha^{t_\alpha}(G_{e^{h(x)},\varepsilon})\}$$
$$= \sup_{x\in\{h\leq M\},\varepsilon>0}\{(e^{h(x)}-\varepsilon)\limsup \mu_\alpha^{t_\alpha}(G_{e^{h(x)},\varepsilon})\}.$$

In the above expressions, $G_{e^{h(x)},\varepsilon}$ *can be replaced by* $F_{e^{h(x)},\varepsilon}$.

Part (a) of the following theorem shows that under conditions strictly weaker than large deviations, $\Lambda(h)$ exists and has the same form as when large deviations hold, since in this case the rate function coincides with l_1 (Lemma 1); it can be seen as a vague version of Varadhan's theorem. Note that the hypothesis $h \in \mathcal{C}_{\mathcal{K}}(X)$ cannot be dropped: consider a vague large deviation principle for a net of probability measures with rate function $J \equiv +\infty$, take $h \equiv 0$ and get $\Lambda(h) = 0$ and $\sup_X\{h(x) - J(x)\} = -\infty$. Note also that the condition (ii) holds in particular when $(\mu_\alpha^{t_\alpha})$ converges in $\Gamma(X, [0, 1])$.

Theorem 2. *Let* $\mathcal{T} \subset \mathcal{C}(X)$ *with* X *locally compact Hausdorff, and assume that the following conditions are fulfilled*:

(i) $\mathcal{T}$ *satisfies the tail condition for* $(\mu_\alpha^{t_\alpha})$.
(ii) $\limsup \mu_\alpha^{t_\alpha}(K) \leq \liminf \mu_\alpha^{t_\alpha}(G)$ *for each compact* $K \subset X$ *and each open* $G \subset X$ *with* $K \subset G$.
(iii) $\inf_{h\in\mathcal{T}} \overline{\Lambda}(h) > m$ *for some real* m.

The following conclusions hold.

(a) *If* $\mathcal{T} \subset \mathcal{C}_{\mathcal{K}}(X)$, *then* $\Lambda(h)$ *exists for all* $h \in \mathcal{T}$, *and there is a real* M *such that*

$$\Lambda(h) = \sup_{x\in\{m\leq h\leq M\}}\{h(x) - l_1(x)\} = \sup_{x\in X}\{h(x) - l_1(x)\} \quad \textit{for all } h \in \mathcal{T}. \qquad (5)$$

(b) *If* (μ_α) *is exponentially tight with respect to* (t_α), *then* $\Lambda(h)$ *exists for all* $h \in \mathcal{T}$, *and there is a real* M *and a compact* $K \subset X$ *such that*

$$\Lambda(h) = \sup_{x\in K\cap\{m\leq h\leq M\}}\{h(x)-l_1(x)\} = \sup_{x\in X}\{h(x)-l_1(x)\} \quad \textit{for all } h \in \mathcal{T}. \qquad (6)$$

Proof. Assume $\mathcal{T} \subset \mathcal{C}_{\mathcal{K}}(X)$. By (i) and Theorem 1, there is a real M' such that for each $h \in \mathcal{T}$,

$$\sup_{x\in\{h\leq M'+\log 2\}} e^{h(x)}e^{-l_1(x)} \leq \sup_{x\in X} e^{h(x)}e^{-l_1(x)} \leq e^{\underline{\Lambda}(h)}$$
$$\leq e^{\overline{\Lambda}(h)} = \sup_{x\in\{h\leq M'\},\varepsilon>0}\{(e^{h(x)}-\varepsilon)\limsup \mu_\alpha^{t_\alpha}(F_{e^{h(x)},\varepsilon})\}. \qquad (7)$$

Put $M = \log 2 + M'$, and suppose that

$$\sup_{x\in\{h\leq M\}} e^{h(x)}e^{-l_1(x)} + \nu < \sup_{x\in\{h\leq M'\},\varepsilon>0}\{(e^{h(x)}-\varepsilon)\limsup \mu_\alpha^{t_\alpha}(F_{e^{h(x)},\varepsilon})\}$$

for some $h \in \mathcal{T}$ and some $\nu > 0$. Then there exist $x_0 \in \{h \leq M'\}$ and $\varepsilon_0 > 0$ with $e^{h(x_0)} > \varepsilon_0$ such that

$$\sup_{x \in \{h \leq M\}} e^{h(x)} e^{-l_1(x)} < (e^{h(x_0)} - \varepsilon_0 - \nu) \limsup \mu_\alpha^{t_\alpha}(F_{e^{h(x_0)}, \varepsilon_0}). \tag{8}$$

By continuity and local compactness, for each $x \in F_{e^{h(x_0)}, \varepsilon_0}$, there exist some open sets V_x and V'_x satisfying $x \in V_x \subset \overline{V_x} \subset V'_x$ with $\overline{V_x}$ compact, and such that $e^{h(y)} > e^{h(x_0)} - \varepsilon_0 - \nu$ for all $y \in V'_x$. Note that $h(x) \leq M$ for each $x \in F_{e^{h(x_0)}, \varepsilon_0}$, since $e^{h(x_0)} + \varepsilon_0 < 2e^{M'}$. By (8), for each $x \in F_{e^{h(x_0)}, \varepsilon_0}$, there exist some open sets W_x and W'_x satisfying $x \in W_x \subset \overline{W_x} \subset W'_x$ with $\overline{W_x}$ compact, and such that

$$e^{h(x)} \liminf \mu_\alpha^{t_\alpha}(W'_x) < (e^{h(x_0)} - \varepsilon_0 - \nu) \limsup \mu_\alpha^{t_\alpha}(F_{e^{h(x_0)}, \varepsilon_0}). \tag{9}$$

Put $G_x = W_x \cap V_x$ for all $x \in F_{e^{h(x_0)}, \varepsilon_0}$. Since $F_{e^{h(x_0)}, \varepsilon_0}$ is compact, there is a finite set $A \subset F_{e^{h(x_0)}, \varepsilon_0}$ such that $F_{e^{h(x_0)}, \varepsilon_0} \subset \bigcup_{x \in A} G_x$; thus, for some $x \in A$ we have

$$\begin{aligned}(e^{h(x_0)} - \varepsilon_0 - \nu) \limsup \mu_\alpha^{t_\alpha}(F_{e^{h(x_0)}, \varepsilon_0}) &\leq e^{h(x)} \limsup \mu_\alpha^{t_\alpha}(G_x) \\ &\leq e^{h(x)} \limsup \mu_\alpha^{t_\alpha}(\overline{W_x}) \leq e^{h(x)} \liminf \mu_\alpha^{t_\alpha}(W'_x)\end{aligned}$$

(where the third inequality follows from (ii)), which contradicts (9). Therefore, all inequalities in (7) are equalities, that is for each $h \in \mathcal{T}$, $\Lambda(h)$ exists and

$$\Lambda(h) = \sup_{x \in \{h \leq M\}} \{h(x) - l_1(x)\} = \sup_{x \in X} \{h(x) - l_1(x)\} = \sup_{x \in \{m \leq h \leq M\}} \{h(x) - l_1(x)\},$$

(where the third equality follows from (iii)), which proves (a). For (b), the above proof works verbatim replacing $\{h \leq M\}$ and $F_{e^{h(x_0)}, \varepsilon_0}$ by $\{h \leq M\} \cap K$ and $F_{e^{h(x_0)}, \varepsilon_0} \cap K$ respectively, where K is some compact set given by the exponential tightness. $\square$

The following definition extends the usual notion of Legendre–Fenchel transform (when X is a real topological vector space and $\mathcal{S}$ its topological dual) and its generalization proposed in [4] (with $X = \mathbb{R}$ and $\mathcal{S} = \{h_{\lambda,\nu} : (\lambda, \nu) \in \mathbb{R}^2\}$); it coincides with our preceding notations since for $\mathcal{S} = \{h_\lambda : \lambda \in G\}$ with G a nonempty open interval, we have

$$\begin{aligned}L^*_{|G}(x) &= \sup_{\lambda \in \mathbb{R}} \{\lambda x - \widehat{L_{|G}}(\lambda)\} = \sup_{\lambda \in G} \{\lambda x - L(\lambda)\} \\ &= \sup_{\{h_\lambda : \lambda \in G\}} \{h_\lambda(x) - \Lambda(h_\lambda)\} = \Lambda^*_{|S}(x).\end{aligned}$$

In [1] (Corollary 2), we proved that for X completely regular (not necessary Hausdorff), a rate function has always the form $\Lambda^*_{|\mathcal{S}}$, where $\mathcal{S}$ is any set in $\mathcal{C}(X)$ stable by translation, separating suitably points and closed sets, and such that each $h \in \mathcal{S}$ satisfies the tail condition for $(\mu_\alpha^{t_\alpha})$; this is proved in [2] for X normal Hausdorff and $\mathcal{S}$ the set of all bounded continuous functions on X (this case was known under exponential tightness hypothesis as a part of the conclusion of Bryc's theorem). We will identify in the next section other sets $\mathcal{S}$ for which the rate function is given by $\Lambda^*_{|\mathcal{S}}$. Note that $\Lambda^*_{|\mathcal{S}}$ is lower semi-continuous when $\mathcal{S} \subset \mathcal{C}(X)$.

Definition 3. Let $\mathcal{S} \subset \mathcal{B}(X)$ such that $\Lambda(h)$ exists for all $h \in \mathcal{S}$. The map $\Lambda^*_{|\mathcal{S}}$ defined by

$$\Lambda^*_{|\mathcal{S}}(x) = \sup_{h \in \mathcal{S}}\{h(x) - \Lambda(h)\} \quad \text{for all } x \in X,$$

is the abstract Legendre–Fenchel transform of $\Lambda_{|\mathcal{S}}$.

4. The Case $X = \mathbb{R}$

In this section, we take $X = \mathbb{R}$ and apply Theorem 2 with $\mathcal{T} = \{h_\lambda : \lambda \in G\}$ where G is a nonempty open interval. This allows us to compare the values of $l_1^{(\mu_\gamma^{t_\gamma})}$ and those of $L^*_{|G}$ on $\mathrm{ran}L'_{|G_-} \cup \mathrm{ran}L'_{|G_+}$, where $(\mu_\gamma^{t_\gamma})$ is a suitable subnet of $(\mu_\alpha^{t_\alpha})$ (Proposition 1). By means of a compactness argument, we then derive sufficient conditions for large deviations, involving only the left and right derivatives of $L_{|G}$; the rate function is given by an abstract Legendre–Fenchel transform $\Lambda^*_{|\mathcal{S}}$ (Theorem 3). The strengthening of Gärtner–Ellis theorem (Corollary 1) and the solution to the Ellis question (Corollary 2) are obtained by taking suitable $\mathcal{S}$.

Proposition 1. *Let $\lambda_0 \in \mathbb{R}$, and assume that $L(\lambda)$ exists and is finite for all λ in an open interval G containing λ_0. Then, $(\mu_\alpha^{t_\alpha})$ has a subnet $(\mu_\gamma^{t_\gamma})$ such that*

$$l_1^{(\mu_\gamma^{t_\gamma})}(L'_{|G_-}(\lambda_0)) \le \lambda_0 L'_{|G_-}(\lambda_0) - L(\lambda_0)$$

and

$$l_1^{(\mu_\gamma^{t_\gamma})}(L'_{|G_+}(\lambda_0)) \le \lambda_0 (L'_{|G_+}(\lambda_0)) - L(\lambda_0).$$

In particular,

$$l_1^{(\mu_\gamma^{t_\gamma})}(x) \le L^*_{|G}(x) \quad \textit{for all } x \in \mathrm{ran}L'_{|G_-} \cup \mathrm{ran}L'_{|G_+}.$$

Proof. Let G_0 be an open interval such that $\lambda_0 \in G_0 \subset \overline{G_0} \subset G$. Let λ_1 and λ_2 in $G\backslash\{0\}$ such that $\lambda_1 < \lambda < \lambda_2$ for all $\lambda \in G_0$. There exists $\gamma > 1$ such that $\{\gamma\lambda_1, \gamma\lambda_2\} \subset \mathcal{D}\mathrm{om}(L)$ so that h_{λ_1} and h_{λ_2} satisfy (individually) the tail condition by Lemma 4.3.8 of [3] (the proof given there for probability measures works as well for the sub-probability case). Therefore, for each $\varepsilon > 0$ and for each $i \in \{1, 2\}$ there exists $M_{i,\varepsilon}$ such that

$$\limsup \mu_\alpha^{t_\alpha}(e^{h_{\lambda_i}/t_\alpha} 1_{\{h_{\lambda_i} > M_{i,\varepsilon}\}}) < \varepsilon.$$

Put $M_\varepsilon = M_{1,\varepsilon} \vee M_{2,\varepsilon}$, and obtain for each $\lambda \in G_0$,

$$\begin{aligned}
&\int_{\{x:\lambda x > M_\varepsilon\}} e^{\lambda x/t_\alpha} \mu_\alpha(dx) \\
&\qquad = \int_{\{x:\lambda x > M_\varepsilon\}\cap\mathbb{R}_-} e^{\lambda x/t_\alpha} \mu_\alpha(dx) + \int_{\{x:\lambda x > M_\varepsilon\}\cap\mathbb{R}_+} e^{\lambda x/t_\alpha} \mu_\alpha(dx) \\
&\qquad \le \int_{\{x:\lambda_1 x > M_{1,\varepsilon}\}\cap\mathbb{R}_-} e^{\lambda_1 x/t_\alpha} \mu_\alpha(dx) + \int_{\{x:\lambda_2 x > M_{2,\varepsilon}\}\cap\mathbb{R}_+} e^{\lambda_2 x/t_\alpha} \mu_\alpha(dx),
\end{aligned}$$

hence

$$\begin{aligned}
&\forall\, \lambda \in G_0, \quad \limsup \mu_\alpha^{t_\alpha}(e^{h_\lambda/t_\alpha} 1_{\{h_\lambda > M_\varepsilon\}}) \\
&\qquad \le \limsup \mu_\alpha^{t_\alpha}(e^{h_{\lambda_1}/t_\alpha} 1_{\{h_{\lambda_1} > M_{1,\varepsilon}\}}) \vee \limsup \mu_\alpha^{t_\alpha}(e^{h_{\lambda_2}/t_\alpha} 1_{\{h_{\lambda_2} > M_{2,\varepsilon}\}}) < \varepsilon.
\end{aligned}$$

It follows that $\{h_\lambda : \lambda \in G_0\}$ satisfies the tail condition for $(\mu_\alpha^{t_\alpha})$. Since $L_{|G}$ is continuous and $\overline{G_0}$ compact, $L_{|G_0}$ is bounded and in particular $\inf_{\lambda\in G_0} L(\lambda) > m$ for some real m. Let $(\mu_\gamma^{t_\gamma})$ be a subnet of $(\mu_\alpha^{t_\alpha})$ converging in $\Gamma(X,[0,1])$ (given by Lemma 2), put $\mathcal{T} = \{h_\lambda : \lambda \in G_0\}$, and note that all the hypotheses of Theorem 2 hold for $\mathcal{T}$ and $(\mu_\gamma^{t_\gamma})$, with moreover $\mathcal{T} \subset \mathcal{C}_{\mathcal{K}}(X)$. If $\lambda_0 \neq 0$ (say $\lambda_0 > 0$), then λ_1 and λ_2 can be chosen such that $0 < \lambda_1 < \lambda < \lambda_2$ for all $\lambda \in G_0$. Since for each real $M \ge m$, there is a compact K_M such that $\bigcup_{\lambda\in G_0}\{m \le h_\lambda \le M\} \subset K_M$, by Theorem 2(a) we get a compact K such that

$$L(\lambda) = \sup_{x\in K}\{\lambda x - l_1^{(\mu_\gamma^{t_\gamma})}(x)\} \quad \text{for all } \lambda \in G_0. \tag{10}$$

If $\lambda_0 = 0$, then (μ_α) (resp. (μ_γ)) is exponentially tight with respect to (t_α) (resp. (t_γ)), and we apply Theorem 2(b) to obtain (10). Therefore, for each $\lambda \in G_0$ there exists $x_\lambda \in K$ such that $L(\lambda) = \lambda x_\lambda - l_1^{(\mu_\gamma^{t_\gamma})}(x_\lambda)$. Put $x = L'_{|G_+}(\lambda_0)$, and let $(x_{\lambda'+\lambda_0})$ be a subnet of $(x_{\lambda+\lambda_0})_{\lambda+\lambda_0\in G_0,\lambda>0}$. Since $x_{\lambda+\lambda_0} \in K$ for all $\lambda + \lambda_0 \in G_0$, $(x_{\lambda'+\lambda_0})$ has a subnet $(x_{\lambda''+\lambda_0})$ converging to some point $x'' \in K$ when $\lambda'' \to 0^+$, so that

$$\begin{aligned}
x &= \lim_{\lambda''\to 0^+} \frac{L(\lambda''+\lambda_0) - L(\lambda_0)}{\lambda''} \\
&= \lim_{\lambda''\to 0^+} \frac{(\lambda''+\lambda_0)x_{\lambda''+\lambda_0} - l_1^{(\mu_\gamma^{t_\gamma})}(x_{\lambda''+\lambda_0}) - L(\lambda_0)}{\lambda''} \\
&= x'' + \lim_{\lambda''\to 0^+} \frac{\lambda_0 x_{\lambda''+\lambda_0} - l_1^{(\mu_\gamma^{t_\gamma})}(x_{\lambda''+\lambda_0}) - L(\lambda_0)}{\lambda''},
\end{aligned}$$

which implies $x'' = x$ and

$$0 = \lim_{\lambda''\to 0^+} \lambda_0 x_{\lambda''+\lambda_0} - l_1^{(\mu_\gamma^{t_\gamma})}(x_{\lambda''+\lambda_0}) - L(\lambda_0) \le \lambda_0 x - l_1^{(\mu_\gamma^{t_\gamma})}(x) - L(\lambda_0),$$

which proves the assertion concerning $L'_{|G_+}(\lambda_0)$. A similar proof works for $L'_{|G_-}(\lambda_0)$. □

Theorem 3. *Let $\mathcal{S} \subset \mathcal{C}(X)$ and $G \subset X$ be a nonempty open interval such that $\mathcal{S} \supset \{h_\lambda : \lambda \in G\}$, and assume that $\Lambda(h)$ exists for all $h \in \mathcal{S}$ with $L(\lambda)$ finite for all $\lambda \in G$.*

(a) *If*

$$\mathrm{ran}L'_{|G_-} \cup \mathrm{ran}L'_{|G_+} \supset \mathcal{D}\mathrm{om}(l_0) \cap \{l_1 > -\overline{\Lambda}(0)\}, \tag{11}$$

then (μ_α) satisfies a vague large deviation principle with powers (t_α) and rate function J satisfying

$$J(x) = L^*_{|G}(x) = \Lambda^*_{|\mathcal{S}}(x) \quad \text{for all } x \in \mathcal{D}\mathrm{om}(J) \cap \{J > -\overline{\Lambda}(0)\}. \tag{12}$$

If moreover $0 \in G$, then the principle is narrow and

$$J(x) = L^*_{|G}(x) = \Lambda^*_{|\mathcal{S}}(x) \quad \text{for all } x \in \mathcal{D}\mathrm{om}(J). \tag{13}$$

(b) *If*

$$\mathrm{ran}L'_{|G_-} \cup \mathrm{ran}L'_{|G_+} \supset \mathcal{D}\mathrm{om}(l_0), \tag{14}$$

then (μ_α) satisfies a vague large deviation principle with powers (t_α) and rate function J satisfying

$$J(x) = L^*_{|G}(x) = \Lambda^*_{|\mathcal{S}}(x) \quad \text{for all } x \in \mathcal{D}\mathrm{om}(J). \tag{15}$$

If moreover $0 \in G$, then the principle is narrow.

(c) *If*

$$\mathrm{ran}L'_{|G_-} \cup \mathrm{ran}L'_{|G_+} \supset \mathcal{D}\mathrm{om}(\Lambda^*_{|\mathcal{S}}) \cap \{l_1 > -\overline{\Lambda}(0)\}, \tag{16}$$

then (μ_α) satisfies a vague large deviation principle with powers (t_α) and rate function J satisfying

$$J(x) = \Lambda^*_{|\mathcal{S}}(x) \quad \text{for all } x \in \{J > -\overline{\Lambda}(0)\}, \tag{17}$$

and

$$J(x) = L^*_{|G}(x) \quad \text{for all } x \in \mathcal{D}\mathrm{om}(\Lambda^*_{|\mathcal{S}}) \cap \{J > -\overline{\Lambda}(0)\}. \tag{18}$$

If moreover $0 \in G$, then the principle is narrow with $J = \Lambda^*_{|\mathcal{S}}$ satisfying

$$J(x) = L^*_{|G}(x) \quad \text{for all } x \in \mathcal{D}\mathrm{om}(J). \tag{19}$$

(d) *If*

$$\mathrm{ran}L'_{|G_-} \cup \mathrm{ran}L'_{|G_+} \supset \mathcal{D}\mathrm{om}(\Lambda^*_{|\mathcal{S}}), \tag{20}$$

then (μ_α) satisfies a vague large deviation principle with powers (t_α) and rate function $J = \Lambda^*_{|\mathcal{S}}$ satisfying

$$J(x) = L^*_{|G}(x) \quad \text{for all } x \in \mathcal{D}\mathrm{om}(J). \tag{21}$$

If moreover $0 \in G$, then the principle is narrow.

(e) *If* l_0 *is proper convex, then* (*a*) (*resp.* (*b*)) *holds verbatim replacing the symbol* $\mathcal{D}$om *by* int$\mathcal{D}$om *in* (11)–(13) (*resp.* (14) *and* (15)).
(f) *If* $\Lambda^*_{|\mathcal{S}}$ *is proper convex, then* (*c*) (*resp.* (*d*)) *holds verbatim replacing the symbol* $\mathcal{D}$om *by* int$\mathcal{D}$om *in* (16), (18), (19)) (*resp.* (20) *and* (21)).

Proof. For all $h \in \mathcal{S}$ and all $x \in X$ we have by Theorem 1 (since $\Lambda(h) \geq \overline{\Lambda}(h1_{\{h\leq M\}} + (-\infty)1_{\{h>M\}})$ for all reals M),

$$\begin{aligned}\Lambda(h) - h(x) &\geq \sup_{M\in\mathbb{R}} \sup_{\{h\leq M\}} \{h(y) - l_0(y)\} - h(x)\\ &\geq \sup_{y\in X}\{h(y) - l_0(y)\} - h(x) \geq -l_0(x),\end{aligned}$$

so that

$$L^*_{|G}(x) \leq \Lambda^*_{|\mathcal{S}}(x) \leq l_0(x) \quad \text{for all } x \in X. \tag{22}$$

Assume that (11) holds, and let $(\mu_\beta^{t_\beta})$ be a subnet of $(\mu_\alpha^{t_\alpha})$. By Proposition 1 applied to $(\mu_\beta^{t_\beta})$ in place of $(\mu_\alpha^{t_\alpha})$, $(\mu_\beta^{t_\beta})$ has a subnet $(\mu_\gamma^{t_\gamma})$ such that

$$l_1^{(\mu_\gamma^{t_\gamma})}(x) \leq L^*_{|G}(x) \quad \text{for all } x \in \mathcal{D}\text{om}(l_0) \cap \{l_1 > -\overline{\Lambda}(0)\}. \tag{23}$$

Since

$$l_0 \leq l_0^{(\mu_\gamma^{t_\gamma})} \leq l_1^{(\mu_\gamma^{t_\gamma})} \leq l_1, \tag{24}$$

(22) and (23) imply for each $x \in \mathcal{D}\text{om}(l_0) \cap \{l_1 > -\overline{\Lambda}(0)\}$,

$$l_0^{(\mu_\gamma^{t_\gamma})}(x) = l_1^{(\mu_\gamma^{t_\gamma})}(x) = L^*_{|G}(x) = \Lambda^*_{|\mathcal{S}}(x) = l_0(x). \tag{25}$$

If $x \notin \mathcal{D}\text{om}(l_0)$, then $l_0^{(\mu_\gamma^{t_\gamma})}(x) = l_1^{(\mu_\gamma^{t_\gamma})}(x) = +\infty$ by (24). If $l_1(x) \leq -\overline{\Lambda}(0)$, then

$$l_0^{(\mu_\gamma^{t_\gamma})}(x) = l_1^{(\mu_\gamma^{t_\gamma})}(x) = l_0(x) = l_1(x) = -\overline{\Lambda}(0).$$

Therefore, $l_0^{(\mu_\gamma^{t_\gamma})}(x) = l_1^{(\mu_\gamma^{t_\gamma})}(x)$ for all $x \in X$. By Lemma 1 applied to $(\mu_\gamma^{t_\gamma})$, (μ_γ) satisfies a vague large deviation principle with powers (t_γ) and rate function

$$J(x) = \begin{cases} \Lambda^*_{|\mathcal{S}} & \text{if } x \in \mathcal{D}\text{om}(l_0) \cap \{l_1 > -\overline{\Lambda}(0)\}\\ -\overline{\Lambda}(0) & \text{if } l_1(x) \leq -\overline{\Lambda}(0)\\ +\infty & \text{if } x \notin \mathcal{D}\text{om}(l_0). \end{cases} \tag{26}$$

By Lemma 2(b), $(\mu_\gamma^{t_\gamma})$ converges to c_J in $\Gamma(X, [0,1])$. Since $(\mu_\beta^{t_\beta})$ is arbitrary, we have proved that any subnet of $(\mu_\alpha^{t_\alpha})$ has a subnet converging vaguely to c_J. By Lemma 2(a), it follows that $(\mu_\alpha^{t_\alpha})$ converges vaguely to c_J, which proves the first assertion of (a) ((12) follows from (25) and (26), since $J = l_0 = l_1$). If $0 \in G$, then (13) follows from (22) and (26) since $-L(0) \leq L^*_{|G}$, and the principle is narrow by exponential tightness. The proofs of (b)–(d) are similar. Assume that l_0 is proper convex, and

$$\text{ran}L'_{|G_-} \cup \text{ran}L'_{|G_+} \supset \text{int}\mathcal{D}\text{om}(l_0) \cap \{l_1 > -\overline{\Lambda}(0)\}.$$

In the same way as above we get for each $x \in \mathrm{int}\mathcal{D}\mathrm{om}(l_0) \cap \{l_1 > -\overline{\Lambda}(0)\}$,

$$l_0^{(\mu_\gamma^{t_\gamma})}(x) = l_1^{(\mu_\gamma^{t_\gamma})}(x) = L^*_{|G}(x) = \Lambda^*_{|\mathcal{S}}(x) = l_0(x). \tag{27}$$

Suppose that $l_1^{(\mu_\gamma^{t_\gamma})}(x) > l_0(x)$ for some $x \in \{l_1 > -\overline{\Lambda}(0)\}$. Since l_1 and $l_1^{(\mu_\gamma^{t_\gamma})}$ are lower semi-continuous, there is an open set G_0 containing x such that

$$\inf_{G_0 \cap \{l_1 > -\overline{\Lambda}(0)\}} l_1^{(\mu_\gamma^{t_\gamma})} > \inf_{G_0 \cap \{l_1 > -\overline{\Lambda}(0)\}} l_0 = \inf_{G_0 \cap \{l_1 > -\overline{\Lambda}(0)\} \cap \mathrm{int}\mathcal{D}\mathrm{om}(l_0)} l_0,$$

where the equality follows from Lemma 3 applied to l_0 and $G_0 \cap \{l_1 > -\overline{\Lambda}(0)\}$. Then, there exists $y \in G_0 \cap \{l_1 > -\overline{\Lambda}(0)\} \cap \mathrm{int}\mathcal{D}\mathrm{om}(l_0)$ such that $l_1^{(\mu_\gamma^{t_\gamma})}(y) > l_0(y)$, which contradicts (27). We then have $l_1^{(\mu_\gamma^{t_\gamma})}(x) \leq l_0(x)$ for all $x \in \{l_1 > -\overline{\Lambda}(0)\}$, and by (24),

$$l_0^{(\mu_\gamma^{t_\gamma})}(x) = l_1^{(\mu_\gamma^{t_\gamma})}(x) = l_0(x) \quad \text{for all } x \in \{l_1 > -\overline{\Lambda}(0)\}.$$

Since

$$l_0^{(\mu_\gamma^{t_\gamma})}(x) = l_1^{(\mu_\gamma^{t_\gamma})}(x) = l_0(x) = l_1(x) = -\overline{\Lambda}(0) \quad \text{for all } x \in \{l_1 \leq -\overline{\Lambda}(0)\},$$

it follows as above that $(\mu_\alpha^{t_\alpha})$ converges vaguely to c_J, with $J = l_0 = l_1$ satisfying by (27),

$$J(x) = L^*_{|G}(x) = \Lambda^*_{|\mathcal{S}}(x) \quad \text{for all } x \in \mathrm{int}\mathcal{D}\mathrm{om}(J) \cap \{J > -\overline{\Lambda}(0)\}. \tag{28}$$

If $0 \in G$, then $-L(0) \leq L^*_{|G}$, and by (22) and (28) we get

$$J(x) = L^*_{|G}(x) = \Lambda^*_{|\mathcal{S}}(x) \quad \text{for all } x \in \mathrm{int}\mathcal{D}\mathrm{om}(J).$$

This proves the assertion of (e) concerning (a); the one concerning (b) is proved similarly. Assume that $\Lambda^*_{|\mathcal{S}}$ is proper convex, and

$$\mathrm{ran}L'_{|G_-} \cup \mathrm{ran}L'_{|G_+} \supset \mathrm{int}\mathcal{D}\mathrm{om}(\Lambda^*_{|\mathcal{S}}) \cap \{l_1 > -\overline{\Lambda}(0)\}.$$

As above we get

$$l_0^{(\mu_\gamma^{t_\gamma})}(x) = l_1^{(\mu_\gamma^{t_\gamma})}(x) = L^*_{|G}(x) = \Lambda^*_{|\mathcal{S}}(x) \tag{29}$$

for all $x \in \mathrm{int}\mathcal{D}\mathrm{om}(\Lambda^*_{|\mathcal{S}}) \cap \{l_1 > -\overline{\Lambda}(0)\}$. The same reasoning as in the proof of (e) (with $\Lambda^*_{|\mathcal{S}}$ in place of l_0) gives $l_1^{(\mu_\gamma^{t_\gamma})}(x) \leq \Lambda^*_{|\mathcal{S}}(x)$ for all $x \in \{l_1 > -\overline{\Lambda}(0)\}$, and by (22),

$$l_0^{(\mu_\gamma^{t_\gamma})}(x) = l_1^{(\mu_\gamma^{t_\gamma})}(x) = \Lambda^*_{|\mathcal{S}}(x) = l_0(x) \quad \text{for all } x \in \{l_1 > -\overline{\Lambda}(0)\}. \tag{30}$$

Since

$$l_0^{(\mu_\gamma^{t_\gamma})}(x) = l_1^{(\mu_\gamma^{t_\gamma})}(x) = -\overline{\Lambda}(0) \quad \text{for all } x \in \{l_1 \leq -\overline{\Lambda}(0)\},$$

it follows as above that $(\mu_\alpha^{t_\alpha})$ converges vaguely to c_J, with J satisfying (17). Since $J = l_1$, (29) gives

$$J(x) = L^*_{|G}(x) \quad \text{for all } x \in \mathrm{int}\mathcal{D}\mathrm{om}(\Lambda^*_{|\mathcal{S}}) \cap \{J > -\overline{\Lambda}(0)\}. \tag{31}$$

Since $0 \in G$ implies $-L(0) \leq L^*_{|G}$, by (22), (30), (31), we obtain $J = \Lambda^*_{|\mathcal{S}}$ and

$$J(x) = L^*_{|G}(x) \quad \text{for all } x \in \mathrm{int}\mathcal{D}\mathrm{om}(\Lambda^*_{|\mathcal{S}}).$$

This proves the assertion of (f) concerning (c); the one concerning (d) is proved similarly. □

The standard Gärtner–Ellis theorem deals with the case where (μ_α) is a sequence of Borel probability measures; it states that if $L(\lambda)$ exists for all reals λ, L is lower semi-continuous essentially smooth and $0 \in \mathrm{int}\mathcal{D}\mathrm{om}(L)$, then (μ_α) satisfies a large deviation principle with powers (t_α) and rate function L^* ([3], Theorem 2.3.6; [7,6]). A stronger version has been given by O'Brien ([8], Theorem 5.1): if $L(\lambda)$ exists and is finite for all λ in a nonempty open interval G and if $\widehat{L_{|G}}$ is essentially smooth, then (μ_α) satisfies a vague large deviation principle with powers (t_α) and rate function $L^*_{|G}$; if moreover $0 \in G$, then the principle is narrow. The former version is recovered by taking $G = \mathrm{int}\mathcal{D}\mathrm{om}(L)$ (the hypotheses implying $L^* = L^*_{|G}$ with $\widehat{L_{|G}}$ essentially smooth). The improvements consist in the obtention of the vague large deviations, and in the fact that L in not assumed to exist out G (even when L exists on X, it is not assumed to be lower semi-continuous).

The following corollary summarizes the case where $\mathcal{S} = \{h_\lambda\colon \lambda \in G\}$ in Theorem 3, and where large deviations hold with rate function $L^*_{|G}$ $(= \Lambda^*_{|\mathcal{S}})$. It strengthens the O'Brien's version of Gärtner–Ellis theorem by obtaining the same conclusions, with the essential smoothness hypothesis replaced by the weaker condition (32) (or (33) when $0 \in G$); in particular, there is no differentiability assumption. Furthermore, it works for general nets of Radon sub-probability measures.

Corollary 1. *We assume that $L(\lambda)$ exists and is finite for all λ in a nonempty open interval $G \subset X$.*

(a) *If*

$$\mathrm{ran}L'_{|G_-} \cup \mathrm{ran}L'_{|G_+} \supset \mathrm{int}\mathcal{D}\mathrm{om}(L^*_{|G}), \tag{32}$$

then (μ_α) satisfies a vague large deviation principle with powers (t_α) and rate function $L^*_{|G}$. The condition (32) is satisfied in particular when $\widehat{L_{|G}}$ is essentially smooth.

(b) *If $0 \in G$ and*

$$\mathrm{ran}L'_{|G_-} \cup \mathrm{ran}L'_{|G_+} \supset \mathrm{int}\mathcal{D}\mathrm{om}(L^*_{|G}) \cap \{l_1 > -L(0)\}, \tag{33}$$

then (μ_α) satisfies a large deviation principle with powers (t_α) and rate function $L^*_{|G}$.

Proof. (b) and the first assertion of (a) follow from Theorem 3(f) with $\mathcal{S} = \{h_\lambda\colon \lambda \in G\}$. Assume that $\widehat{L_{|G}}$ is essentially smooth. Extend $L_{|G}$ by continuity to a convex function $L_{|\overline{G}}$ on $\overline{G}$, so that $\widehat{L_{|\overline{G}}}$ is a proper convex lower semi-continuous function on X with $G = \mathrm{int}\mathcal{D}\mathrm{om}(\widehat{L_{|\overline{G}}})$; moreover, $\widehat{L_{|\overline{G}}}$ is essentially smooth. By Theorem 26.1 and Corollary 26.4.1 of [10], we have

$$\mathrm{ran}L'_{|\overline{G}} \supset \mathrm{int}\mathcal{D}\mathrm{om}(L^*_{|\overline{G}}), \tag{34}$$

which gives (32) since $\mathrm{ran}L'_{|\overline{G}} = \mathrm{ran}L'_{|G}$ and $L^*_{|\overline{G}} = L^*_{|G}$. □

The solution to the Ellis question (with in fact weaker hypotheses) is a direct consequence of Theorem 3(c), by taking $\mathcal{S} = \{h_{\lambda,\nu}\colon (\lambda,\nu) \in \mathbb{R}^2\}$.

Corollary 2. *Put $\mathcal{S} = \{h_{\lambda,\nu}\colon (\lambda,\nu) \in \mathbb{R}^2\}$, and assume that $\Lambda(h_{\lambda,\nu})$ exists for all $(\lambda,\nu) \in \mathbb{R}^2$ and is finite for all pairs (λ,λ) with λ in some open interval G containing 0. If $\mathrm{ran}L'_{|G_-} \cup \mathrm{ran}L'_{|G_+} \supset \mathcal{D}\mathrm{om}(\Lambda^*_{|\mathcal{S}}) \cap \{l_1 > -L(0)\}$, then (μ_α) satisfies a large deviation principle with powers (t_α) and rate function $J = \Lambda^*_{|\mathcal{S}}$. Moreover,*

$$J(x) = L^*_{|G}(x) \quad \textit{for all } x \in \mathcal{D}\mathrm{om}(J).$$

The following example is often cited as a typical case not covered by the Gärtner–Ellis theorem ([4, 8]).

Example 1. Consider the sequence $(\mu_n^{1/n})$ where $\mu_n\{-1\} = \mu_n\{1\} = \frac{1}{2}$ for all $n \in \mathbb{N}$. Then $L(\lambda) = |\lambda|$ for all reals λ. Take $\mathcal{S} = \{h_{\lambda,\nu}\colon (\lambda,\nu) \in \mathbb{R}^2\}$ and compute

$$\Lambda(h_{\lambda,\nu}) = -\lambda \vee \nu \quad \text{for all } (\lambda,\nu) \in \mathbb{R}^2,$$

hence

$$\Lambda^*_{|\mathcal{S}}(x) = \begin{cases} 0 & \text{if } |x| = 1, \\ +\infty & \text{if } |x| \neq 1. \end{cases}$$

Then, $\mathrm{ran}L'_- \cup \mathrm{ran}L'_+ = \{-1,1\} \supset \mathcal{D}\mathrm{om}(\Lambda^*_{|\mathcal{S}})$, and by Corollary 2, (μ_n) satisfies a large deviation principle with powers $(1/n)$ and rate function $J = \Lambda^*_{|\mathcal{S}}$. Since

$$L^*(x) = \begin{cases} 0 & \text{if } |x| \leq 1, \\ +\infty & \text{if } |x| > 1, \end{cases}$$

we have $J(x) = L^*(x)$ for all $x \in \{-1,1\} = \mathcal{D}\mathrm{om}(J)$. Note that for any nonempty open set $G \subset]-1,1[$,

$$\mathrm{ran}L'_{|G_-} \cup \mathrm{ran}L'_{|G_+} \not\supset \mathrm{int}\mathcal{D}\mathrm{om}(L^*_{|G}) \cap \{J > 0\} \supset]-1,1[,$$

and the condition (33) of Corollary 1 does not hold.

The following example exhibits a situation with convex rate function, where both above corollaries do not work; we then apply Theorem 3 with another set $\mathcal{S}$.

Example 2. Consider the net $(\mu_\varepsilon^\varepsilon)_{\varepsilon>0}$, where μ_ε is the probability measure on X defined by $\mu_\varepsilon(0) = 1 - 2p_\varepsilon$, $\mu_\varepsilon(-\varepsilon\log p_\varepsilon) = \mu_\varepsilon(\varepsilon\log p_\varepsilon) = p_\varepsilon$, and assume that $\lim \varepsilon\log p_\varepsilon = -\infty$. Put $Q_n(x) = n|x|e^{-|x|} - x$ for all $n \in \mathbb{N}$ and all $x \in X$, and take $\mathcal{S} = \{Q_n\colon n \in \mathbb{N}\} \cup \{h_\lambda\colon \lambda \in \,]-1,1[\}$. Easy calculations give $\Lambda(Q_n) = 0$ for all $n \in \mathbb{N}$, and

$$L(\lambda) = \begin{cases} 0 & \text{if } |\lambda| \leq 1, \\ +\infty & \text{if } |\lambda| > 1, \end{cases}$$

so that

$$L^*_{|]-1,1[}(x) = L^*(x) = |x| \quad \text{for all } x \in X,$$

and

$$\Lambda^*_{|\mathcal{S}}(x) = \sup_{n\in\mathbb{N}}\{Q_n(x) - \Lambda(Q_n)\} \vee L^*_{|]-1,1[}(x) = \begin{cases} 0 & \text{if } x = 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Then, $\mathrm{ran}L'_{|]-1,1[} = \{0\} \supset \mathcal{D}\mathrm{om}(\Lambda^*_{|\mathcal{S}})$, and by Theorem 3(d), (μ_ε) satisfies a large deviation principle with powers $(\varepsilon)_{\varepsilon>0}$ and rate function $J = \Lambda^*_{|\mathcal{S}}$. Note that J is convex but $J \neq L^*$ (however, J coincides with L^* on $\mathcal{D}\mathrm{om}(J)$); in particular, L is not essentially smooth and the Gärtner-Ellis theorem does not work. Furthermore, for any nonempty open set $G \subset \,]-1,1[$,

$$\{0\} = \mathrm{ran}L'_{|G} \not\supset \mathrm{int}\mathcal{D}\mathrm{om}(L^*_{|G}) \cap \{J > 0\} \supset X\backslash\{0\}$$

and the condition (33) of Corollary 1 does not hold either. We observe also that Corollary 2 does not apply; indeed, the set $\{h_{\lambda,\nu}\colon (\lambda,\nu) \in \mathbb{R}^2\}$ is not suitable since

$$\Lambda(h_{\lambda,\nu}) = \begin{cases} 0 & \text{if } \lambda \geq -1 \text{ and } \nu \leq 1, \\ +\infty & \text{otherwise} \end{cases}$$

gives $\Lambda^*_{|\{h_{\lambda,\nu}:(\lambda,\nu)\in\mathbb{R}^2\}}(x) = L^*(x)$ for all $x \in X$.

References

1. H. Comman, Functional approach of large deviations in general spaces, *J. Theor. Probab.* **18** (2005) 187–207.
2. H. Comman, Criteria for large deviations, *Trans. Amer. Math. Soc.* **355** (2003) 2905–2923.
3. A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd edn. (Springer-Verlag, 1998).
4. R. S. Ellis, An overview of the theory of large deviations and applications to statistical mechanics, *Scand. Actuarial J.* **1** (1995) 97–142.
5. R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics* (Springer-Verlag, 1985).
6. R. S. Ellis, Large deviations for a general class of random vectors, *Ann. Probab.* **12** (1984) 1–12.

7. J. Gärtner, On large deviations from the invariant measure, *Th. Probab. Appl.* **22** (1977) 24–39.
8. G. L. O'Brien, Compactness in the theory of large deviations, *Stoch. Process. Appl.* **57** (1995) 1–10.
9. G. L. O'Brien and W. Vervaat, Capacities, large deviations and loglog laws, in *Stable Processes and Related Topics*, eds. S. Cambanis, G. Samorodniski and M. S. Taqqu (Birkhäuser, 1991), pp. 43–83.
10. R. T. Rockafeller, *Convex Analysis* (Princeton Univ. Press, 1970).