Positivity   CrossMark

A variational formula on the Cramér function of series of independent random variables

Krzysztof Zajkowski[1]

Abstract   In (Zajkowski, Positivity 19:529–537, 2015) it has been proved some variational formula on the Legendre–Fenchel transform of the cumulant generating function (the Cramér function) of Rademacher series with coefficients in the space ℓ^1. In this paper we show a generalization of this formula to series of a larger class of any independent random variables with coefficients that belong to the space ℓ^2.

1 Introduction

The *Legendre–Fenchel transforms* of *cumulant generating functions* of given random variables are at the core of the *large deviations theory* (see e.g. [3,4]). The Cramér function gives the rate of the exponentially decay of tails of distributions for the empirical means of sequences of i.i.d. random variables. It provides a nice connection between convex analysis and statistics.

Donsker and Varadhan in [5] proved that the Legendre–Fenchel transform ψ_X^* of the cumulant generating function $\psi_X(s) = \ln E e^{sX}$ of a random variable X satisfies the following variational principle

✉ Krzysztof Zajkowski
    kryza@math.uwb.edu.pl

[1]   Institute of Mathematics, University of Bialystok, K. Ciołkowskiego 1M, 15-245 Białystok, Poland

$$\psi_X^*(a) = \inf_{m \ll \mu_X, \int x dm = a} D(m \| \mu_X), \tag{1}$$

where $D(m \| \mu_X) = \int \ln \frac{dm}{d\mu_X} dm$ is the *relative entropy* of a probability distribution m with respect to the distribution μ_X of X.

The aim of this paper is to prove some variational formula for the Cramér functions of series of independent random variables that depends on coefficients and Cramér functions of summands of a given series; see Theorem 2.3.

For series $\sum t_i g_i$ of independent standard normal r.v.s, where $\sum t_i^2 < \infty$ and $g_i \in \mathcal{N}(0, 1)$, it is known their tail estimation of the form

$$Pr\left(\sum t_i g_i > \alpha\right) \le \exp\left(-\frac{\alpha^2}{2 \sum t_i^2}\right);$$

see for instance [10]. The function $\frac{\alpha^2}{2 \sum t_i^2}$ is the Cramér function of the random series $\sum t_i g_i$ (see Example 2.5).

To realize our purposes we will need the general notion of the Legendre–Fenchel transform in topological spaces (see [6] or [1]). Let V be a real locally convex Hausdorff space and V^* its dual space. By $\langle \cdot, \cdot \rangle$ we denote the canonical pairing between V and V^*. Let $f : V \mapsto \mathbb{R} \cup \{\infty\}$ be a function non-identically ∞. By $\mathcal{D}(f)$ we denote the *effective domain* of f, i.e. $\mathcal{D}(f) = \{u \in V : f(u) < \infty\}$. A function $f^* : V^* \mapsto \mathbb{R} \cup \{\infty\}$ defined by

$$f^*(u^*) = \sup_{u \in V} \left\{\langle u, u^* \rangle - f(u)\right\} = \sup_{u \in \mathcal{D}(f)} \left\{\langle u, u^* \rangle - f(u)\right\} \qquad (u^* \in V^*)$$

is called the *Legendre–Fenchel transform* (*convex conjugate*) of f and a function $f^{**} : V \mapsto \mathbb{R} \cup \{\infty\}$ defined by

$$f^{**}(u) = \sup_{u^* \in V^*} \left\{\langle u, u^* \rangle - f^*(u^*)\right\} = \sup_{u^* \in \mathcal{D}(f^*)} \left\{\langle u, u^* \rangle - f^*(u^*)\right\} \qquad (u \in V)$$

is called the *convex biconjugate* of f.

The functions f^* and f^{**} are convex and lower semicontinuous in the weak* and weak topology on V^* and V, respectively. Moreover, the *biconjugate theorem* states that the function $f : V \mapsto \mathbb{R} \cup \{\infty\}$ not identically equal to $+\infty$ is convex and lower semicontinuous if and only if $f = f^{**}$.

Let us mention additional properties of the convex conjugates; see 4.3 Examples in [6]. Let V be a normed space. We denote by $\| \cdot \|$ the norm of V and by $\| \cdot \|_*$ the norm of V^*. For conjugate exponents $p, q \in (1, \infty)$ ($\frac{1}{p} + \frac{1}{q} = 1$), a function $\frac{1}{q} \|u^*\|_*^q$ is the convex conjugate of $\frac{1}{p} \|u\|^p$.

Remark 1.1 Let us emphasize that in Hilbert spaces a function $\frac{1}{2} \|u\|^2$ one can treat as the function invariant with respect to the Legendre–Fenchel transform.

Let us list two properties and the notion of the infimal convolution. The convex-conjugation is order-reversing:

$$\text{if} \quad f \le g \quad \text{then} \quad f^* \ge g^* \tag{2}$$

and

$$\text{if} \quad g(u) = f(au) \quad \text{where} \quad (a \ne 0) \quad \text{then} \quad g^*(u^*) = f^*\left(\frac{u^*}{a}\right). \tag{3}$$

Let functions $f_1, \ldots, f_n$ are convex, lower semicontinuous and not identically equal to $+\infty$. Suppose there is a point in $\bigcap_{i=1}^{n} \mathcal{D}(f_i)$ at which $f_1, \ldots, f_{n-1}$ are continuous. Then the convex conjugate of their sum is given by the so called infimal (in this case even minimal) convolution, i.e.

$$(f_1 + \cdots + f_n)^*(u^*) = \min_{u_1^* + \cdots + u_n^* = u^*} \{f_1^*(u_1^*) + \cdots + f_n^*(u_n^*)\}$$

(see e.g. [7, Th. 1]).

It will turn out (see Remark 2.8) that the variational formula for the Cramér function of series of independent random variables is an example of an application of a generalization of the infimal convolution to the infinite case of summands. Often generalizations of formulas from finite numbers parameters (variables) to the case of infinite ones are not obvious. Other examples of generalizations of the convex conjugates of the logarithm of series of analytic functions, with applications to investigations of the convex conjugates of the spectral radius of the functions of weighted composition operators, one can find in [8,12].

2 Main theorem

The cumulant generating function $\psi_X(s) = \ln E e^{sX}$ of any random variable X is convex and lower semicontinuous on $\mathbb{R}$ (analytic on $int\mathcal{D}(\psi_X)$). It maps $\mathbb{R}$ into $\mathbb{R} \cup \{\infty\}$ and takes value zero at zero but it is possible that $\psi_X(s) = \infty$ when $s \ne 0$. We will assume that it is finite on some neighborhood of zero, i.e. X satisfies condition: $\exists_{\lambda > 0}$ s.t. $E e^{\lambda |X|} < \infty$. Let us emphasize that if $EX = 0$ then $\psi_X \ge 0$ but the Cramér function ψ_X^* is always nonnegative and attains 0 at the value EX.

Let $I \subset \mathbb{N}$ and $(X_i)_{i \in I}$ be a sequence of independent r.v.s. For $\mathbf{t} = (t_i)_{i \in I} \in \ell^2(I) \equiv \ell^2$ consider $X_{\mathbf{t}} = \sum_{i \in I} t_i X_i$ (convergence in L^2 and almost surely). The cumulant generating function of $X_{\mathbf{t}}$ we will denote by $\psi_{\mathbf{t}}$, i.e. $\psi_{\mathbf{t}} = \psi_{X_{\mathbf{t}}}$. Notice that for fixed s we can consider $\psi_{\mathbf{t}}(s)$ as a functional of the variable $\mathbf{t}$ in ℓ^2. We will denote it by ψ^s. Let us emphasize that $\psi^s(\mathbf{t}) = \psi_{\mathbf{t}}(s)$.

Before proving our main theorem, we show forms of the cumulant generating function and the Legendre–Fenchel transform of series of independent random variables.

Proposition 2.1 *Let $(X_i)_{i \in I}$ be a sequence of zero-mean independent random variables with common bounded second moments. Let for each $i \in I$ the cumulant generating function $\psi_i(s) := \psi_{X_i}(s) = \ln E e^{sX_i}$ is finite on some neighborhood*

of zero. Then for each $\mathbf{t} = (t_i)_{i \in I} \in \ell^2$ *the cumulant generating function of a random series* $X_{\mathbf{t}} = \sum_{i \in I} t_i X_i$ *is given by the following equality*

$$\psi_{\mathbf{t}}(s) := \psi_{X_{\mathbf{t}}}(s) = \sum_{i \in I} \psi_i(st_i).$$

Proof Because (X_i) are independent, centered and have common bounded second moments then for every $\mathbf{t} \in \ell^2$ the series $X_{\mathbf{t}} = \sum_{i \in I} t_i X_i$ converges in L^2 and a.s.. Let us emphasize that the convergence of series $X_{\mathbf{t}}$ in L^2 is equivalent to the convergence of sequences $\mathbf{t}$ in ℓ^2. Fixing s we can consider the cumulant generating function $\psi_{\mathbf{t}}(s)$ as a functional of the variable $\mathbf{t}$ in ℓ^2. We will denote it by $\psi^s(\mathbf{t})$, i.e. $\psi^s(\mathbf{t}) = \psi_{\mathbf{t}}(s)$. We show that for every $s \in \mathbb{R}$ the functional ψ^s is convex and lower semicontinuous on ℓ^2.

Convexity one may check by using the Hölder inequality. Let $\mathbf{t}, \mathbf{u} \in \ell^2$ and $\lambda \in (0, 1)$ then

$$\psi^s(\lambda\mathbf{t} + (1 - \lambda)\mathbf{u}) = \ln E e^{s \sum_{i \in I}(\lambda t_i + (1-\lambda)u_i)X_i}$$
$$= \ln E\left[\left(e^{s \sum_{i \in I} t_i X_i}\right)^{\lambda} \left(e^{s \sum_{i \in I} u_i X_i}\right)^{1-\lambda}\right]$$
$$= \ln E\left[\left(e^{sX_{\mathbf{t}}}\right)^{\lambda} \left(e^{sX_{\mathbf{u}}}\right)^{1-\lambda}\right].$$

By the Hölder inequality for exponents $1/\lambda$ and $1/(1 - \lambda)$ we get

$$E\left[\left(e^{sX_{\mathbf{t}}}\right)^{\lambda} \left(e^{sX_{\mathbf{u}}}\right)^{1-\lambda}\right] \leq \left(E e^{sX_{\mathbf{t}}}\right)^{\lambda} \left(E e^{sX_{\mathbf{u}}}\right)^{1-\lambda}$$

and, in consequence,

$$\psi^s(\lambda\mathbf{t} + (1 - \lambda)\mathbf{u}) \leq \lambda \ln E e^{sX_{\mathbf{t}}} + (1 - \lambda) \ln E e^{sX_{\mathbf{u}}}$$
$$= \lambda\psi^s(\mathbf{t}) + (1 - \lambda)\psi^s(\mathbf{u}).$$

Lower semicontinuity follows from Fatou's lemma. Let $\mathbf{t}^n \to \mathbf{t}^0$ in ℓ^2. Note that $X_{\mathbf{t}^n}$ converges a.s. to $X_{\mathbf{t}^0}$. Then

$$\liminf_{n \to \infty} \psi^s(\mathbf{t}^n) = \liminf_{n \to \infty} \ln E e^{sX_{\mathbf{t}^n}} \geq \ln E(\liminf_{n \to \infty} e^{sX_{\mathbf{t}^n}})$$
$$= \ln E(e^{s \lim_{n \to \infty} X_{\mathbf{t}^n}}) = \psi^s(\mathbf{t}^0).$$

It means that ψ^s is lower semicontinuous on ℓ^2.

Let ℓ_0 denote the space of sequences with finite supports. Observe that ℓ_0 is a dense subset of ℓ^2. For $\mathbf{t} \in \ell_0$ we have

$$\psi^s(\mathbf{t}) = \ln E e^{sX_{\mathbf{t}}} = \ln \prod_{i \in I} E e^{st_i X_i}$$
$$= \sum_{i \in I} \psi_i(st_i).$$

For $\mathbf{t} \in \ell^2$ consider a series $\sum_{i \in I} \psi_i(st_i)$. Since $EX_i = 0$, $\psi_i \geq 0$, it follows that $\sum_{i \in I} \psi_i(st_i)$ is convergent or divergent to plus infinity. Since ψ_i are convex, this series defines a convex function on the whole ℓ^2. Let $\mathbf{t}^n \to \mathbf{t}^0$ in ℓ^2. Hence for every $i \in I$ $t_i^n \to t_i^0$. By superadditivity of the limit inferior and, next, by lower semicontinuity of each ψ_i, we get

$$\liminf_{n \to \infty} \psi^s(\mathbf{t}^n) = \liminf_{n \to \infty} \sum_{i \in I} \psi_i(st_i^n) \geq \sum_{i \in I} \liminf_{n \to \infty} \psi_i(st_i^n)$$
$$\geq \sum_{i \in I} \psi_i(st_i^0) = \psi^s(\mathbf{t}^0).$$

Notice that both functions: $\psi^s(\mathbf{t})$ and the series $\sum_{i \in I} \psi_i(st_i)$ are convex and lower semicontinuous on ℓ^2 and moreover coincide on ℓ_0 (a dense subset of ℓ^2). It follows that these functions are equal on whole ℓ^2, i.e.

$$\psi^s(\mathbf{t}) = \sum_{i \in I} \psi_i(st_i)$$

for every $\mathbf{t}$ in ℓ^2. □

Let us observe that for $s = 0$ we have $\psi^0 \equiv 0$ and its convex conjugate $(\psi^0)^*(\mathbf{a}) = 0$ for $\mathbf{a} = \mathbf{0}$ and ∞ otherwise. From now on we assume that $s \neq 0$. A form of $(\psi^s)^*$ for $s \neq 0$ is described in the following:

Proposition 2.2 *Under the assumptions of Proposition* 2.1 *the convex conjugate of* $\psi^s(\mathbf{t}) = \sum_{i \in I} \psi_i(st_i)$ *defined on* ℓ^2 *equals*

$$(\psi^s)^*(\mathbf{a}) = \sum_{i \in I} \psi_i^* \left(\frac{a_i}{s} \right) \quad (s \neq 0)$$

for $\mathbf{a} \in \ell^2$, *where* ψ_i^*'s *are the Cramér functions of* X_i's.

Proof The convex conjugate $(\psi^s)^*$ is convex and lower semicontinuous on $(\ell^2)^* \simeq \ell^2$. Assume first that I is a finite set. By virtue of the form of ψ^s, the convex conjugate of a separated sum (see e.g. [2, Prop. 13.27]) and the property (3), for $\mathbf{a}$ in $\ell^2(I)$, we get

$$(\psi^s)^*(\mathbf{a}) = \sum_{i \in I} \psi_i^* \left(\frac{a_i}{s} \right) \quad (s \neq 0).$$

Define now a functional $\sum_{i \in I} \psi_i^*(\frac{a_i}{s})$ on whole space ℓ^2. Since ψ_i^*'s are convex and lower semicontinuous, this functional is convex and, similarly as in the case of $\sum_{i \in I} \psi_i$, one can show that it is also lower semicontinuous on ℓ^2. Because this functional coincides with $(\psi^s)^*$ on the dense subspace ℓ_0 then both functionals are equal on ℓ^2. □

Let us emphasize that the functions $(\psi^s)^*$ are nonnegative and lower semicontinuous. In large deviation theory such functions are called rate functions (good rate functions when sublevel sets are not only closed but also compact). In the main theorem below we show that the contraction principle applied to the function $(\psi^1)^*$ by using a functional $\langle \mathbf{t}, \cdot \rangle$ over ℓ^2 gives the Cramér function of $X_{\mathbf{t}}$.

Theorem 2.3 *Let a sequence of r.v.'s $(X_i)_{i \in I}$ satisfies the assumptions of Proposition 2.1. Then for every $\mathbf{t} = (t_i)_{i \in I} \in \ell^2$ the Cramér function $\psi^*_{X_{\mathbf{t}}} = \psi^*_{\mathbf{t}}$ of a random series $X_{\mathbf{t}} = \sum_{i \in I} t_i X_i$ is given by the following variational formula*

$$\psi^*_{\mathbf{t}}(\alpha) = \inf_{\mathbf{b} \in \ell^2: \langle \mathbf{t}, \mathbf{b} \rangle = \alpha} \sum_{i \in I} \psi^*_i(b_i),$$

for $\alpha \in int\mathcal{D}(\psi^*_{\mathbf{t}})$, where ψ^*_i's are the Cramér function of X_i's.

Proof The functional ψ^s is convex and lower semicontinuous on ℓ^2. By virtue of the biconjugate theorem we have

$$\psi^s(\mathbf{t}) = \sup_{\mathbf{a} \in \ell^2} \{ \langle \mathbf{t}, \mathbf{a} \rangle - (\psi^s)^*(\mathbf{a}) \},$$

where $(\psi^s)^*(\mathbf{a}) = \sum_{i \in I} (\psi_i)^*(\frac{a_i}{s})$ $(s \neq 0)$. Substituting $\mathbf{a} = s\mathbf{b}$ we get

$$(\psi^s)^*(\mathbf{a}) = \sum_{i \in I} \psi^*_i \left(\frac{a_i}{s} \right) = \sum_{i \in I} \psi^*_i(b_i) = (\psi^1)^*(\mathbf{b})$$

and we can rewrite the above as follows

$$\psi^s(\mathbf{t}) = \sup_{\mathbf{b} \in \ell^2} \{ s \langle \mathbf{t}, \mathbf{b} \rangle - (\psi^1)^*(\mathbf{b}) \}. \tag{4}$$

Let us return to the function $\psi_{\mathbf{t}}$ which is convex and lower semicontinuous on $\mathbb{R}$. By the biconjugate theorem we have

$$\psi_{\mathbf{t}}(s) = \sup_{\alpha \in \mathbb{R}} \{ s\alpha - \psi^*_{\mathbf{t}}(\alpha) \}.$$

Let us recall that $\psi_{\mathbf{t}}(s) = \psi^s(\mathbf{t})$. If we split the supremum of (4) into two parts: over $\mathbb{R}$ and hyperplanes $\{ \mathbf{b} \in \ell^2 : \langle \mathbf{b}, \mathbf{t} \rangle = constant \}$ then we get

$$\psi_{\mathbf{t}}(s) = \psi^s(\mathbf{t}) = \sup_{\alpha \in \mathbb{R}} \sup_{\substack{\mathbf{b} \in \ell^2: \\ \langle \mathbf{t}, \mathbf{b} \rangle = \alpha}} \{ s \langle \mathbf{t}, \mathbf{b} \rangle - (\psi^1)^*(\mathbf{b}) \}$$

$$= \sup_{\alpha \in \mathbb{R}} \{ s\alpha - \inf_{\mathbf{b} \in \ell^2: \langle \mathbf{t}, \mathbf{b} \rangle = \alpha} (\psi^1)^*(\mathbf{b}) \}.$$

Let $\varphi_{\mathbf{t}}(\alpha)$ denote the function $\inf_{\mathbf{b} \in \ell^2: \langle \mathbf{t}, \mathbf{b} \rangle = \alpha} (\psi^1)^*(\mathbf{b})$. The functional $(\psi^1)^*$ is convex on ℓ^2. Convexity is preserved under contraction by linear transformation (see [4,

Th. III.32]). It suffices to state that $\varphi_{\mathbf{t}}$ and $\psi_{\mathbf{t}}^*$ coincide on $int\mathcal{D}(\psi_{\mathbf{t}}^*)$ (both ones take ∞ on a complement of $cl\mathcal{D}(\psi_{\mathbf{t}}^*)$) that is

$$\psi_{\mathbf{t}}^*(\alpha) = \inf_{\mathbf{b}\in\ell^2:\ \langle\mathbf{t},\mathbf{b}\rangle=\alpha} (\psi^1)^*(\mathbf{b}),$$

for $\alpha \in int\mathcal{D}(\psi_{\mathbf{t}}^*)$, where $(\psi^1)^*(\mathbf{b}) = \sum_{i\in I} \psi_i^*(b_i)$.   □

Remark 2.4 We cannot prove in general that $\varphi_{\mathbf{t}}$ is lower semicontinuous. Sometimes it is obvious when for instance the effective domain of $\psi_{\mathbf{t}}^*$ is open subset of $\mathbb{R}$ (or even whole $\mathbb{R}$; see Examples 2.5 and 2.6). Under an assumption that $(\psi^1)^*$ is a good rate function with respect to weak* topology we can state lower-semicontinuity of $\varphi_{\mathbf{t}}$ (see Example 2.7).

Example 2.5 The moment generating function of a standard normal r.v. $g \in \mathcal{N}(0, 1)$ equals $Ee^{sg} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{st-\frac{t^2}{2}} dt = e^{\frac{s^2}{2}}$ and its cumulant generating function to be

$$\psi_g(s) := \ln Ee^{sg} = \frac{s^2}{2}.$$

The function $\frac{s^2}{2}$ is invariant with respect to Legendre transform (see Remark 1.1) that is the Cramér function of g is given by

$$\psi_g^*(\alpha) = \frac{\alpha^2}{2}.$$

By virtue of Proposition 2.1 the cumulant generating function of the series $X_{\mathbf{t}} = \sum_{i\in I} t_i g_i$, where g_i are independent and standard normal distributed, equals

$$\psi_{\mathbf{t}}(s) = \sum_{i\in I} \psi_g(st_i) = \frac{1}{2}s^2 \sum_{i\in I} t_i^2 = \frac{1}{2}s^2\|\mathbf{t}\|^2.$$

By the scaling property (3) we obtain the evident form of the Cramér function

$$\psi_{\mathbf{t}}^*(\alpha) = \frac{\alpha^2}{2\|\mathbf{t}\|^2}.$$

On the other hand by Theorem 2.3 we get

$$\psi_{\mathbf{t}}^*(\alpha) = \frac{1}{2} \inf_{\substack{\mathbf{b}\in\ell^2:\\ \langle\mathbf{t},\mathbf{b}\rangle=\alpha}} \|\mathbf{b}\|^2.$$

Using the Lagrange multipliers technique one can check that this infimum is attained at a sequence $\mathbf{b} = (\frac{\alpha t_i}{\|\mathbf{t}\|^2})_{i\in I}$.

Example 2.6 Let X be r.v. with the Laplace density $\frac{1}{2}e^{-|x|}$. Its moment generating function $Ee^{sX} = \frac{1}{1-s^2}$ for $|s| < 1$ and ∞ otherwise. Let us observe that

$$Ee^{sX} = \frac{1}{1-s^2} = \sum_{n=0}^{\infty} s^{2n} \geq \sum_{n=0}^{\infty} \frac{s^{2n}}{n!2^n} = e^{\frac{s^2}{2}} = Ee^{sg},$$

where g is standard normal distributed. Thus $\psi_X \geq \psi_g$ and, since the Legendre–Fenchel transform is order-reversing, we have

$$\psi_X^*(\alpha) \leq \psi_g^*(\alpha) = \frac{\alpha^2}{2}. \tag{5}$$

Moreover by using the classical Legendre transform we can calculate an evident form of ψ_X^* and get

$$\psi_X^*(\alpha) = \frac{\alpha^2}{\sqrt{1+\alpha^2}+1} + \ln \frac{2}{\sqrt{1+\alpha^2}+1}.$$

Consider a sequence $(X_i)_{i \in I}$ of independent r.v.s with the same Laplace distribution. By (5) we have

$$(\psi^1)^*(\mathbf{b}) = \sum_{i \in I} \psi_X^*(b_i) \leq \frac{1}{2}\|\mathbf{b}\|^2.$$

It means that $(\psi^1)^*$ takes finite values on the whole space ℓ^2 and for every $\alpha \in \mathbb{R}$ there is a finite infimum:

$$\inf_{\mathbf{b} \in \ell^2:\, \langle \mathbf{t}, \mathbf{b} \rangle = \alpha} \sum_{i \in I} \left(\frac{b_i^2}{\sqrt{1+b_i^2}+1} + \ln \frac{2}{\sqrt{1+b_i^2}+1} \right),$$

that is the finite value of $\psi_{\mathbf{t}}^*$ at α.

In the paper [11] one can find an example of the variational formula for the Cramér function of series of weighted symmetric Bernoulli random variables but with coefficients belonging to the space ℓ^1. In the context of our Theorem 2.3, we recall the main result of this paper but now with coefficients in the larger space ℓ^2.

Example 2.7 If X is a symmetric Bernoulli r.v., i.e. $Pr(X = \pm 1) = \frac{1}{2}$, then $Ee^{sX} = \cosh s$. By power series expansions one has $\cosh s \leq \exp(\frac{s^2}{2})$. In this example, conversely as in previous one, we get that $\psi_X(s) \leq \psi_g(s) = \frac{s^2}{2}$ and

$$\psi_X^*(\alpha) \geq \frac{\alpha^2}{2}. \tag{6}$$

One can check that

$$\psi_X^*(\alpha) = \frac{1}{2}[(1 + \alpha) \ln(1 + \alpha) + (1 - \alpha) \ln(1 - \alpha)]$$

for $|\alpha| \leq 1$ and ∞ otherwise; we take $0 \ln 0 = 0$. Note that $\psi_X^*(\pm 1) = \ln 2$.

For a sequence of independent Bernoulli r.v.s, by the above inequality, we have

$$(\psi^1)^*(\mathbf{b}) \geq \frac{1}{2} \|\mathbf{b}\|^2.$$

Since $(\psi^1)^*$ is lower semicontinuous in the weak* topology, sublevel sets $\{\mathbf{b} \in \ell^2 : (\psi^1)^*(\mathbf{b}) \leq c\}$ are weak* closed; for $c < 0$ are empty sets. By the above, for each $c \geq 0$ the sublevel set of $(\psi^1)^*$ is contained in a closed ball $\overline{B}(0; \sqrt{2c}) = \{\mathbf{b} \in \ell^2 : \|\mathbf{b}\| \leq \sqrt{2c}\}$. By virtue of the Banach–Alaoglo theorem (see [9, Th. 3.15]) balls $\overline{B}(0; \sqrt{2c})$ are weak* compact (are polar sets of balls $\overline{B}(0; 1/\sqrt{2c})$). It follows that the sublevel sets $\{\mathbf{b} \in \ell^2 : (\psi^1)^*(\mathbf{b}) \leq c\}$ are weak* compact as closed subsets of compact sets. It means that in this topology $(\psi^1)^*$ is a good rate function and by the contraction principle (see [3, Th. 4.2.1]) $\varphi_{\mathbf{t}}$ is also good a rate function on $\mathbb{R}$, in particular, it is lower semicontinuous.

Let us emphasize that if we know that $(\psi^1)^*$ is a good rate function then we can prove lower-semicontinuity of $\varphi_{\mathbf{t}}$.

Remark 2.8 Let assume now that I is a finite set and $Y_i = t_i X_i$ then $X_{\mathbf{t}} = \sum_{i \in I} Y_i$. Note that

$$\psi_{\mathbf{t}}(s) = \sum_{i \in I} \psi_{Y_i}(s).$$

The functions ψ_{Y_i}, $i \in I$, are convex, lower semicontinuous, not identically equals $+\infty$ and continuous at 0 that is ones satisfy the assumption of Theorem 1 [7]. The convex conjugate of their sum is given by the infimal convolution

$$\psi_{\mathbf{t}}^*(\alpha) = \min_{\sum \alpha_i = \alpha} \sum_{i \in I} \psi_{Y_i}^*(\alpha_i). \tag{7}$$

Since $\psi_{Y_i}(s) = \psi_{t_i X_i}(s) = \psi_{X_i}(t_i s)$, by scaling property (3), one has

$$\psi_{Y_i}^*(\alpha_i) = \psi_{t_i X_i}^*(\alpha_i) = \psi_{X_i}^*\left(\frac{\alpha_i}{t_i}\right).$$

Substituting $\alpha_i = b_i t_i$ into (7) we obtain

$$\psi_{\mathbf{t}}^*(\alpha) = \min_{\sum b_i t_i = \alpha} \sum_{i \in I} \psi_{X_i}^*(b_i).$$

It means that Theorem 2.3 one can treat as the special case of a generalization of the infimal convolution to the situation of infinite numbers of summands.

References

1. Barbu, V., Precupanu, T.: Convexity and Optimization in Banach Spaces, 4th edn. Springer Monographs in Mathematics. Springer, Dordrecht (2012)
2. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer Sciences+Business Media (2011)
3. Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications. Corrected reprint of the second (1998) edition, Stochastic Modeling and Applied Probability, vol. 38. Springer, Berlin (2010)
4. den Hollander, F.: Large Deviations, Fields Institute Monographs, vol. 14. American Mathematical Society, Providence (2000)
5. Donsker, M.D., Varadhan, S.R.S.: Asymptotic evaluation of certain Markov process expectations for large time III. Commun. Pure Appl. Math **29**, 389–461 (1976)
6. Ekeland, I., Témam, R.: Convex Analysis and Variational Problems, Translated from French. Corrected reprint of the 1976 English edition. Classics in Applied Mathematics, vol. 28. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1999)
7. Hiriart-Urruty, J.-B.: A note on the Legendre–Fenchel Transform of Convex Composite Functions, Nonsmooth Mechanics and Analysis, pp. 35–46, Adv. Mech. Math., vol. 12. Springer, New York (2006)
8. Ostaszewska, U., Zajkowski, K.: Legendre–Fenchel transform of the spectral exponent of analytic functions of weighted composition operators. J. Convex Anal. **18**(2), 367–377 (2011)
9. Rudin, W.: Functional Analysis, 2nd edn. International Series in Pure and Applied Mathematics (1991)
10. Talagrand, M.: Majorizing measures: the generic chaining. Ann. Probab. **24**, 1049–1103 (1996)
11. Zajkowski, K.: Cramér transform of Rademacher series. Positivity **19**, 529–537 (2015)
12. Zajkowski, K.: Convex conjugates of analytic functions of logarithmically convex functional. J. Convex Anal. **20**(1), 243–252 (2013)