CRAMÉR'S THEOREM IS ATYPICAL

By Nina Gantert [*] , Steven Soojin Kim [†,§] , and Kavita Ramanan [*,‡,§]

Technische Universität München, Brown University, and Brown University

The empirical mean of n independent and identically distributed (i.i.d.) random variables $(X_1, \ldots, X_n)$ can be viewed as a suitably normalized scalar projection of the n-dimensional random vector $X^{(n)} \doteq (X_1, \ldots, X_n)$ onto the unit vector $n^{-1/2}(1, 1, \ldots, 1) \in \mathbb{S}^{n-1}$. The large deviation principle (LDP) for such projections as $n \to \infty$ is given by the classical Cramér's theorem. We prove an LDP for the sequence of normalized scalar projections of $X^{(n)}$ onto a generic unit vector $\theta^{(n)} \in \mathbb{S}^{n-1}$, as $n \to \infty$. This LDP holds under fairly general conditions on the distribution of X_1, and for "almost every" sequence of directions $(\theta^{(n)})_{n \in \mathbb{N}}$. The associated rate function is "universal" in the sense that it does not depend on the particular sequence of directions. Moreover, under mild additional conditions on the law of X_1, we show that the universal rate function differs from the Cramér rate function, thus showing that the sequence of directions $n^{-1/2}(1, 1, \ldots, 1) \in \mathbb{S}^{n-1}$, $n \in \mathbb{N}$, corresponding to Cramér's theorem is atypical.

1. Introduction. Let $X^{(n)} = (X_1, \ldots, X_n)$ be a sequence of n independent and identically distributed (i.i.d.) $\mathbb{R}$-valued random variables with common distribution $\gamma \in \mathcal{P}(\mathbb{R})$. A fundamental probabilistic question is how the empirical mean of $X^{(n)}$ behaves as the length of the sequence n increases. From a geometric perspective, the empirical mean is a suitably normalized version of (the scalar component of) the projection of the n-dimensional vector $X^{(n)}$ onto the unit vector $\iota^{(n)}$, defined by

$$(1.1) \qquad \iota^{(n)} \doteq \tfrac{1}{\sqrt{n}} (\overbrace{1, 1, \ldots, 1}^{n \text{ times}}) \in \mathbb{S}^{n-1}.$$

In other words, we can write

$$(1.2) \qquad W_\iota^{(n)} \doteq \frac{1}{\sqrt{n}} \langle X^{(n)}, \iota^{(n)} \rangle_n = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

where $\langle \cdot, \cdot \rangle_n$ denotes the Euclidean inner product. With some abuse of terminology, for $x \in \mathbb{R}^n$ and $v \in \mathbb{S}^{n-1}$, we hereby write the "*projection* of x onto v" to refer to the scalar component $\langle x, v \rangle_n \in \mathbb{R}$ (rather than the vector $\langle x, v \rangle_n v \in \mathbb{R}^n$). Then, the expression (1.2) indicates

that questions on the empirical mean for large n can be rephrased in a geometric language as questions on suitably normalized projections of high-dimensional random vectors.

The classical Cramér's theorem characterizes the large deviations behavior of (1.2), the empirical mean of i.i.d. random variables, as $n \to \infty$. In particular, if $X_1 \sim \gamma$ has some finite exponential moments, in the sense that

$$(1.3) \qquad \exists t_0 > 0 \text{ s.t. } \forall |t| < t_0, \quad \Lambda(t) \doteq \log \mathbb{E}[e^{tX_1}] < \infty,$$

then we have the limit

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(W_\iota^{(n)} \geq x) = -\Lambda^*(x),$$

where * denotes the *Legendre transform*,

$$(1.4) \qquad \Lambda^*(x) \doteq \sup_{t \in \mathbb{R}} \{tx - \Lambda(t)\}.$$

We refer to [17, §12] for a review of the Legendre transform (also known as the *convex conjugate*).

Given the geometric view of empirical means given by (1.2), it is natural to investigate analogs of Cramér's theorem for normalized projections onto directions $\theta^{(n)} \in \mathbb{S}^{n-1}$ other than $\iota^{(n)}$. Such projections correspond to *weighted* means,

$$(1.5) \qquad W_\theta^{(n)} \doteq \frac{1}{\sqrt{n}} \langle X^{(n)}, \theta^{(n)} \rangle_n = \frac{1}{n} \sum_{i=1}^n X_i \sqrt{n} \theta_i^{(n)}.$$

Our main result is an LDP for $(W_\theta^{(n)})_{n \in \mathbb{N}}$ for almost every (in a sense that is specified below) sequence of directions $\theta = (\theta^{(1)}, \theta^{(2)}, \dots)$. In particular, we show that the associated rate function does not depend on θ, and that it differs from the Cramér rate function Λ^*. That is, the sequence of directions $(\iota^{(n)})_{n \in \mathbb{N}}$ corresponding to Cramér's theorem is "atypical"!

REMARK 1.1.   While the LDP for (1.5) is novel, the corresponding law of large numbers (LLN) and central limit theorem (CLT) for weighted sums are well known. For example, a weak LLN follows from Chebyshev's inequality, and a CLT follows from the Lindeberg conditions (see, e.g., [11, §VIII.4, Theorem 3]).

The outline of this note is as follows. In Section 2, we state our main results and discuss their relation to prior work. In Section 3, we prove the claimed LDP. In Section 4, we establish that Cramér's theorem is atypical, and also comment on a generalization that is considered in [12].

2. Main results.   We first set some notation. Suppose the random variables $X_1, X_2, \dots$ are all defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\| \cdot \|_n$ denote the Euclidean norm on $\mathbb{R}^n$. Write σ_{n-1} for the unique rotation invariant probability measure on $\mathbb{S}^{n-1}$, the unit sphere in $\mathbb{R}^n$. Let $\mathbb{S} \doteq \prod_{n \in \mathbb{N}} \mathbb{S}^{n-1}$, and let $\pi_n : \mathbb{S} \to \mathbb{S}^{n-1}$ be the coordinate map such

that for $\theta = (\theta^{(1)}, \theta^{(2)}, \dots) \in \mathbb{S}$, we have $\pi_n(\theta) = \theta^{(n)}$. Let σ be a probability measure on (the Borel sets of) $\mathbb{S}$ such that

(H1) $$\sigma \circ \pi_n^{-1} = \sigma_{n-1}, \quad n \in \mathbb{N}.$$

The generic example to keep in mind that satisfies (H1) is the product measure $\sigma = \bigotimes_{n \in \mathbb{N}} \sigma_{n-1}$, in which case the projection directions $\theta^{(n)}$, $n \in \mathbb{N}$, are independent under σ. However, our results allow for more general dependencies; for more discussion on σ and the condition (H1), see Remark 3.3.

For σ-a.e. $\theta \in \mathbb{S}^{n-1}$, we prove a large deviation principle for the sequence $(W_\theta^{(n)})_{n \in \mathbb{N}}$ with a rate function that does not depend on θ. We refer to [9] for general background on large deviations. In particular, recall the following definition:

DEFINITION 2.1. The sequence of probability measures $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}(\mathbb{R})$ is said to satisfy a *large deviation principle (LDP)* with a *rate function* $\mathbb{I} : \mathbb{R} \to [0, \infty]$ if $\mathbb{I}$ is lower semicontinuous, and for all Borel measurable sets $\Gamma \subset \mathbb{R}$,

$$-\inf_{x \in \Gamma^\circ} \mathbb{I}(x) \le \liminf_{n \to \infty} \tfrac{1}{n} \log \mu_n(\Gamma^\circ) \le \limsup_{n \to \infty} \tfrac{1}{n} \log \mu_n(\bar{\Gamma}) \le -\inf_{x \in \bar{\Gamma}} \mathbb{I}(x),$$

where Γ° and $\bar{\Gamma}$ denote the interior and closure of Γ, respectively. Furthermore, $\mathbb{I}$ is said to be a *good rate function* if it has compact level sets.

We say the sequence of $\mathbb{R}$-valued random variables $(\xi_n)_{n \in \mathbb{N}}$ satisfies an LDP if the sequence of laws $(\mu_n)_{n \in \mathbb{N}}$ given by $\mu_n = \mathbb{P} \circ \xi_n^{-1}$ satisfies an LDP.

In particular, for empirical means of i.i.d. random variables, we recall the following classical result, due to [7, 5].

THEOREM 2.2 (Cramér). *Let $(X_n)_{n \in \mathbb{N}}$ be an i.i.d. sequence such that* (1.3) *holds, and let $\iota = (\iota^{(1)}, \iota^{(2)}, \dots)$ be defined as in* (1.1)*. Then the sequence $(W_\iota^{(n)})_{n \in \mathbb{N}}$ of* (1.2) *satisfies an LDP with the good rate function $\mathbb{I}_\iota$, given by*

(2.1) $$\mathbb{I}_\iota(w) \doteq \Lambda^*(w) = \sup_{t \in \mathbb{R}} \{tw - \Lambda(t)\}.$$

Let $\nu \in \mathcal{P}(\mathbb{R})$ denote the standard one-dimensional Gaussian measure. In the sequel, we assume the following condition on Λ, the logarithmic moment generating function (log mgf) of $X_1 \sim \gamma$:

(H2) $$\forall t \in \mathbb{R}, \quad \int_{\mathbb{R}} |\Lambda(tu)|^4 \nu(du) < \infty.$$

Note that (H2) is stronger than even requiring the exponential moment condition in (1.3) to hold with $t_0 = \infty$. For an absolutely continuous γ with density, a sufficient condition for (H2) is that the decay of the tail of the density is strictly faster than exponential, in the following sense:

LEMMA 2.3.   *Suppose that γ has density f, and that there exist $p \in (1, \infty)$ and constants $0 < C_1, C_2, C_3 < \infty$ such that for $|x| > C_1$, we have*

$$f(x) \leq C_2 e^{-C_3 |x|^p}.$$

Then there exists some constant $C < \infty$ such that Λ, the log mgf of γ, satisfies the following upper bound for all $t \in \mathbb{R}$:

$$\Lambda(t) \leq C|t|^{p/(p-1)} + C.$$

Moreover, this implies that Λ satisfies the condition (H2).

PROOF.   By Young's inequality applied to the conjugate exponents p and $\frac{p}{p-1}$, for $\varepsilon > 0$ and $t, y \in \mathbb{R}$,

$$ty \leq \left(\varepsilon^{-1/p}|t| \right) \left(\varepsilon^{1/p}|y| \right) \leq \frac{p-1}{p} \varepsilon^{-1/(p-1)}|t|^{p/(p-1)} + \frac{\varepsilon |y|^p}{p}.$$

In the following, let C absorb all constants, and let $0 < \varepsilon < C_3 p$ to find that for $t \in \mathbb{R}$,

$$\begin{aligned}
\Lambda(t) &= \log \int_{|y| \leq C_1} e^{ty} f(y) dy + \log \int_{|y| > C_1} e^{ty} f(y) dy \\
&\leq C_1 |t| + \log C_2 + \log \int_{\mathbb{R}} e^{ty} e^{-C_3 |y|^p} dy \\
&\leq C_1 |t|^{p/(p-1)} + C_1 + \frac{p-1}{p} \varepsilon^{-1/(p-1)} |t|^{p/(p-1)} - \frac{1}{p} \log(C_3 p - \varepsilon) + 1 \\
&= C|t|^{p/(p-1)} + C.
\end{aligned}$$

From the preceding inequalities, since the Gaussian measure ν has finite moments of every order, it is clear that Λ satisfies the integrability condition (H2). $\square$

We define the following analog of the log mgf in the case of weighted sums,

$$(2.2) \qquad \Psi(t) \doteq \int_{\mathbb{R}} \Lambda(tu) \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du, \quad t \in \mathbb{R}.$$

Our first main result is the following.

THEOREM 2.4 (Weighted LDP).   *Assume* (H1) *and* (H2). *Then, for σ-a.e. $\theta \in \mathbb{S}$, the sequence $(W_\theta^{(n)})_{n \in \mathbb{N}}$ of* (1.5) *satisfies an LDP with the convex good rate function $\mathbb{I}_\sigma$, given by*

$$(2.3) \qquad \mathbb{I}_\sigma(w) \doteq \Psi^*(w) = \sup_{t \in \mathbb{R}} \{ tw - \Psi(t) \}.$$

The proof of Theorem 2.4 is given in Section 3, with intermediate steps established in Section 3.1 and Section 3.2, and the proof completed in Section 3.3.

In principle, the rate function $\mathbb{I}_\sigma$ of Theorem 2.4 could depend on the particular choice of θ, but our result shows that the rate function is the same for σ-a.e. θ. In the case where

σ is the product measure $\sigma = \bigotimes_{n\in\mathbb{N}} \sigma_{n-1}$, this follows immediately from the Kolmogorov zero-one law. That is, let $\mathcal{T}_n$ be the sigma-algebra generated by $(\theta^{(k)})_{k\geq n}$, and let

$$(2.4) \qquad\qquad\qquad\qquad \mathcal{T} \doteq \bigcap_{n=1}^{\infty} \mathcal{T}_n$$

denote the *tail sigma-algebra* induced by $(\theta^{(1)}, \theta^{(2)}, \dots)$. The rate function $\mathbb{I}_\sigma$ is measurable with respect to $\mathcal{T}$, and the Kolmogorov zero-one law states that $\mathcal{T}$ is trivial under the product measure. Hence, $\mathbb{I}_\sigma$ coincides for σ-a.e. $\theta \in \mathbb{S}$. However, our claim holds for general σ satisfying (H1). In particular, Example 3.2(ii) gives an example of σ such that $\theta^{(1)}, \theta^{(2)}, \dots$ are highly dependent, $\mathcal{T}$ is not trivial, and hence, the lack of dependence of the rate function $\mathbb{I}_\sigma$ on θ is not *a priori* obvious.

Given the σ-a.e. statement of Theorem 2.4, it is natural to ask what happens on the set of measure zero in $\mathbb{S}$ where the stated LDP does not hold. In particular, our second main result Theorem 2.5 shows that under certain additional conditions on Λ, the sequence of directions ι associated with Cramér's theorem is exceptional, in the sense that Cramér's rate function $\mathbb{I}_\iota$ differs from the universal rate function $\mathbb{I}_\sigma$. For the following theorem, we assume γ is symmetric, or specifically:

$$(H3) \qquad\qquad\qquad\qquad \forall t \in \mathbb{R}, \quad \Lambda(t) = \Lambda(-t).$$

THEOREM 2.5 (Atypicality).   *Assume Λ satisfies (H3), and let $\mathbb{I}_\iota$ and $\mathbb{I}_\sigma$ be given by (2.1) and (2.3), respectively.*

   a. *If $\Lambda \circ \sqrt{\cdot}$ is concave on $\mathbb{R}_+$, then $\mathbb{I}_\sigma(w) \geq \mathbb{I}_\iota(w)$ for all $w \in \mathbb{R}$.*
   b. *If $\Lambda \circ \sqrt{\cdot}$ is convex on $\mathbb{R}_+$, then $\mathbb{I}_\sigma(w) \leq \mathbb{I}_\iota(w)$ for all $w \in \mathbb{R}$.*
   c. *If $\Lambda \circ \sqrt{\cdot}$ is concave or convex, but not linear, on $\mathbb{R}_+$, then $\mathbb{I}_\sigma(w) = \mathbb{I}_\iota(w) < \infty$ if and only if $w = 0$.*

The proof of Theorem 2.5 is given in Section 4.

We now provide some sufficient conditions (established in [1]) for the convexity or concavity conditions of Theorem 2.5 to hold.

PROPOSITION 2.6.   *Assume the exponential moment condition (1.3) and the symmetry condition (H3).*

   i. *Suppose $\gamma \neq \delta_0$, the Dirac mass at 0. Define $\varphi : \mathbb{N} \to \mathbb{R}$ by*

$$\varphi(k) \doteq (2k+1)\frac{\mathbb{E}[|X_1|^{2k}]}{\mathbb{E}[|X_1|^{2k+2}]}, \quad k \in \mathbb{N}.$$

   If φ is non-decreasing (resp., non-increasing), then $\Lambda \circ \sqrt{\cdot}$ is concave (resp., convex) on $\mathbb{R}_+$.
   ii. *Suppose γ has density f such that $\log f \circ \sqrt{\cdot}$ is concave (resp., convex) on $\mathbb{R}_+$. Then $\Lambda \circ \sqrt{\cdot}$ is concave (resp., convex) on $\mathbb{R}_+$.*

PROOF. Part i. is established in Theorem 7 of [1]. Part ii. follows from applying Theorem 12 of [1] with their f replaced by our $f \circ \sqrt{\cdot}$, and noticing that the integrability of $f \circ \sqrt{\cdot}$ follows from the fact that f has finite first moment, due to the exponential moment condition of (1.3). □

EXAMPLE 2.7. Suppose γ is the *generalized normal distribution* with location 0, scale $\alpha > 0$, and shape $\beta > 1$; that is, $\gamma = \mu_{\alpha,\beta}$, where

$$\mu_{\alpha,\beta}(dx) \doteq \frac{1}{2\alpha\Gamma(1 + \frac{1}{\beta})} e^{-(|x|/\alpha)^{\beta}} dx$$

It follows from Lemma 2.3 that $\mu_{\alpha,\beta}$ satisfies (H2), which implies (1.3). It is also easy to see that $\mu_{\alpha,\beta}$ satisfies (H3). Thus, the conditions of Proposition 2.6 are satisfied. It follows immediately from Proposition 2.6(ii) that for $\beta \geq 2$ (resp., for $\beta \leq 2$), $\Lambda \circ \sqrt{\cdot}$ is concave (resp., convex). In fact, for $\beta \neq 2$, the concavity (resp., convexity) is strict.

The preceding example suggests the particular role of the Gaussian, which corresponds to $\beta = 2$. In particular, $\gamma = \mu_{\alpha,2}$ for some $\alpha > 0$ if and only if $\Lambda \circ \sqrt{\cdot}$ is linear. Thus, we could interpret the conditions of Theorem 2.5 as evaluating whether our distribution of interest is "more" or "less" log-concave than the Gaussian. We also have the following result in the Gaussian case (i.e., when $\gamma = \mu_{\alpha,2}$), which holds for *all* θ as opposed to just for σ-a.e. θ.

PROPOSITION 2.8. *Suppose $\gamma = \mu_{\alpha,2}$ for some $\alpha > 0$. Then, for all $\theta \in \mathbb{S}$, the sequence $(W_{\theta}^{(n)})_{n\in\mathbb{N}}$ satisfies an LDP with the good rate function $\Psi^*(w) = \Lambda^*(w) = (w/\alpha)^2$, where Λ^* is defined in (1.4) with Λ the log mgf of the Gaussian with mean 0 and variance $\alpha^2/2$.*

PROOF. This follows from the fact that for all $n \in \mathbb{N}$, the Gaussian measure on $\mathbb{R}^n$ is spherically symmetric, and hence, for any $\theta^{(n)} \in \mathbb{S}^{n-1}$, the law of $\langle X^{(n)}, \theta^{(n)} \rangle_n$ is the same as the law of $\langle X^{(n)}, \iota^{(n)} \rangle_n$. Thus, the LDP for $(W_{\theta}^{(n)})_{n\in\mathbb{N}}$ follows from the classical Cramér's theorem for empirical means of i.i.d. Gaussians, for which the rate function can be easily computed to be $\Lambda^*(w) = (w/\alpha)^2$. □

REMARK 2.9. It is not clear whether a converse of Proposition 2.8 holds. That is, whether $\mathbb{I}_{\sigma} \equiv \mathbb{I}_{\iota}$ if and only if γ is Gaussian. As one possible approach in this direction, it would be sufficient to show that for any measure γ satisfying both (H2) and (H3) (and possibly some additional natural conditions), the function $\Lambda \circ \sqrt{\cdot}$ must be either concave or convex.

Aside from the sequence of Cramér directions $\iota \in \mathbb{S}$, another natural sequence of directions to consider is the sequence of canonical basis vectors, $e_1 = (e_1^{(1)}, e_1^{(2)}, \ldots) \in \mathbb{S}$, where

$$e_1^{(n)} \doteq (1, \overbrace{0, \ldots, 0}^{n-1 \text{ times}}) \in \mathbb{S}^{n-1}.$$

Then $W_{e_1}^{(n)} = X_1/\sqrt{n}$ for all n. The following result states that under certain tail conditions, such normalized projections yield a trivial LDP, again with a rate function different from $\mathbb{I}_\sigma$.

PROPOSITION 2.10. *Assume the following condition (which is stronger than (H2)):*

(H2′) $\quad\quad\quad \exists C < \infty, r \in [0,2)$ such that $\forall t \in \mathbb{R}, \quad \Lambda(t) \leq C(1+|t|^r).$

Then the sequence $(W_{e_1}^{(n)})_{n\in\mathbb{N}}$ satisfies an LDP with the trivial good rate function χ_0 given by

$$\chi_0(x) \doteq \left\{ \begin{array}{ll} 0, & x = 0; \\ \infty, & x \neq 0. \end{array} \right.$$

PROOF. Consider the limit log mgf associated with the Gärtner-Ellis theorem (recalled for convenience later in Theorem 3.7). For all $t \in \mathbb{R}$,

$$\Lambda_n(t) \doteq \frac{1}{n} \log \mathbb{E}[\exp(tnW_{e_1}^{(n)})] = \frac{1}{n} \log \mathbb{E}[\exp(t\sqrt{n}X_1)] = \frac{1}{n}\Lambda(t\sqrt{n}) \leq \frac{1}{n}(C|t|^r n^{r/2} + C).$$

Since the exponent r of (H2′) satisfies $r < 2$ by assumption, we have $\lim_{n\to\infty} \Lambda_n(t) = 0$ for all $t \in \mathbb{R}$. Thus, by the Gärtner-Ellis theorem, the sequence $(W_{e_1}^{(n)})_{n\in\mathbb{N}}$ satisfies an LDP with good rate function $0^* = \chi_0$. □

2.1. *Relation to prior work.* There is a wealth of literature on large deviations for weighted sums, but our work seems to be the first to emphasize the unique position of Cramér's theorem in the geometric setting. Moreover, it appears that none of the existing literature is readily adaptable to our particular problem. We offer a partial (but inevitably, incomplete) survey of existing results.

In the somewhat classical works of Book, [2] and [3], we can find asymptotics bounds for quantities of the form

$$\mathbb{P}\left(\frac{\sum_{k=1}^n a_{nk}X_k}{\sum_{k=1}^n a_{nk}} > c\right),$$

where $(a_{nk})_{k\leq n, n\in\mathbb{N}}$ is a triangular array of weights such that $\sum_{k=1}^n a_{nk}^2 = 1$ for all n. However, this does not address our setting because if we let $a_{nk} = \theta_k^{(n)}$, we have $\sum_{k=1}^n a_{nk}^2 = 1$, but this only yields tail bounds of the form $\mathbb{P}(W_\theta^{(n)} > cn^{-1/2}\sum_{k=1}^n \theta_k^{(n)})$, as opposed to the desired asymptotics for $\mathbb{P}(W_\theta^{(n)} > c)$. Furthermore, Book does not establish an LDP or identify a rate function.

In a more recent line of work, consider [14], where on their p. 932, their λ and ν correspond to our n and k, respectively. For $Z \sim N(0,1)$, we have the correspondence:

$$a_j(n) = \mathbf{1}_{\{j\leq n-1\}}\frac{1}{n}\sqrt{n}\theta_j^{(n)}, \quad j,n \in \mathbb{N};$$

$$\sum_{j=0}^\infty a_j(n)^k \approx \sum_{j=0}^{n-1} \frac{1}{n^k}z_j^k \overset{(\zeta\text{-a.e.})}{\approx} \frac{\mathbb{E}[|Z|^k]}{n^{k-1}} \doteq \frac{a_k}{n^{k-1}}, \quad k,n \in \mathbb{N};$$

$$\phi(n) = n, \quad n \in \mathbb{N}.$$

Suppose that the sequence $(a_k)_{k\in\mathbb{N}}$ (which depends on the particular choice of weights $a_j(n)$, $j,n\in\mathbb{N}$) satisfies the following condition (from p. 932 of [14]):

$$(2.5) \qquad\qquad \lim_{k\to\infty}|a_k|^{1/k} < \infty.$$

The main result of [14] is that for a sequence of i.i.d. random variables $(X_k)_{k\in\mathbb{N}}$ with cumulants $(c_k)_{k\in\mathbb{N}}$, if condition (2.5) holds, then the sequence of weighted means $\frac{1}{n}\sum_{j=1}^n a_j(n)X_j$, $n\in\mathbb{N}$, satisfies an LDP with rate function χ^*, the Legendre transform of $\chi(t) \doteq \sum_{k=2}^\infty \frac{a_k c_k}{k!}t^k$. However, the finiteness condition (2.5) does not hold in our setting of $a_k = \mathbb{E}[|Z|^k]$, since the following limit is infinite:

$$\lim_{k\to\infty}\mathbb{E}[|Z|^k]^{1/k} = \sqrt{2}\lim_{k\to\infty}\Gamma(\tfrac{k+1}{2})^{1/k} = \infty.$$

Therefore, the weighted mean LDP of [14] does not apply in our setting.

Yet more recently, [16] proves an LDP for weighted empirical means similar to (1.5), except with weights that are uniformly bounded (in n). Our results correspond to *unbounded* weights $\sqrt{n}\theta_i^{(n)}$ which are not covered by their results. Similarly, [6] proves an LDP for empirical means of certain bounded functionals, which again fails to apply to our unbounded weights.

In the context of information theory, [8] states an LDP for sums of the form $\frac{1}{n}\sum_{i=1}^n \rho(x_i,Y_i)$, where $(x_i)_{i\in\mathbb{N}}$ are "weights", $(Y_i)_{i\in\mathbb{N}}$ is a sequence of random variables satisfying certain mixing properties, and $\rho:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}_+$ for Polish spaces $\mathcal{X}$ and $\mathcal{Y}$. The LDP is stated in the form of a *generalized asymptotic equipartition property* for "distortion measures". However, note that ρ is assumed to be nonnegative, so a function like $\rho(x,y)=xy$ (corresponding to projections) does not fit within the setting of [8]. Moreover, their weights $(x_i)_{i\in\mathbb{N}}$ are assumed to be a realization of a stationary ergodic process, which is not the case for our weights $\sqrt{n}\theta^{(n)}$ that are drawn from the scaled sphere $\sqrt{n}\mathbb{S}^{n-1}$. This lends our work a geometric rather than information-theoretic interpretation.

The paper [13], co-authored by the first and third authors of this work, also analyzes weighted sums of i.i.d. random variables, but there the emphasis is on sums of subexponential random variables, rather than the weights themselves.

The most closely related work to our own is the recent work of [4], which gives strong large deviations (i.e., refined asymptotics) for weighted sums of i.i.d. random variables and i.i.d. weights, conditioned on the weights. Our weights $\sqrt{n}\theta_i^{(n)}$ are not i.i.d., but in Section 3.1 and Section 3.2, we prove that Theorem 2.4 can be reduced to an LDP for the sequence $(\widehat{W}_z^{(n)})_{n\in\mathbb{N}}$ defined in (3.2), which is an i.i.d. weighted sum, conditional on given weights. With some additional calculations from this point, the rate function $\mathbb{I}_\sigma$ of Theorem 2.4 could then be deduced from the conditional LDP of [4], stated in their Theorem 1.6 with rate function defined in their equation (1.13). Note that condition (iii) of their Theorem 1.6 has two parts, but our integrability condition (H2) corresponds only to their first part; in fact, it follows from Lemma 3.8 that their second part follows from our condition (3.5), which is weaker than (H2), and thus, need not be assumed separately. Moreover, our research

(completed independently) differs due to our emphasis on a geometric point of view; as a consequence, we can explicitly identify a rate function $\mathbb{I}_\sigma$ and highlight the atypical position occupied by Cramér's theorem.

Lastly, the methods we use are a simplification of those developed in a companion paper [12], where we consider normalized projections of certain *non*-product measures, as well as projections onto random directions.

3. The σ-almost everywhere LDP.

3.1. *The surface measure on $\mathbb{S}^{n-1}$.* In this section, we recall a convenient representation for a random vector distributed according to the surface measure on $\mathbb{S}^{n-1}$, in order to obtain (3.3), which reduces σ-a.e. statements into more tractable statements about Gaussian random variables. Let $\mathbb{A} \doteq \prod_{n \in \mathbb{N}} \mathbb{R}^n$ denote the space of infinite triangular arrays. That is, $z \in \mathbb{A}$ is of the form $z = (z^{(1)}, z^{(2)}, \dots)$ where $z^{(n)} \in \mathbb{R}^n$ for all $n \in \mathbb{N}$. Let $\mathcal{R} : \mathbb{A} \to \mathbb{A}$ be the map such that for $z \in \mathbb{A}$, the n-th row of $\mathcal{R}(z)$ is

$$[\mathcal{R}(z)]^{(n)} \doteq \frac{z^{(n)}}{\|z^{(n)}\|_n}.$$

Let $\bar{\pi}_n : \mathbb{A} \to \mathbb{R}^n$ denote the n-th row map such that $\bar{\pi}_n(z) = z^{(n)}$. Let ν denote the Gaussian measure on $\mathbb{R}$, and let $\nu^{\otimes n}$ denote the standard Gaussian measure on $\mathbb{R}^n$.

LEMMA 3.1. *If $\zeta \in \mathcal{P}(\mathbb{A})$ is such that*

(3.1) $$\zeta \circ \bar{\pi}_n^{-1} = \nu^{\otimes n}, \quad n \in \mathbb{N},$$

then $\sigma \doteq \zeta \circ \mathcal{R}^{-1}$ satisfies (H1)*. Conversely, if $\sigma \in \mathcal{P}(\mathbb{S})$ satisfies* (H1)*, then there exists some $\zeta \in \mathcal{P}(\mathbb{A})$ satisfying* (3.1) *such that $\sigma = \zeta \circ \mathcal{R}^{-1}$.*

PROOF. Both results are merely a restatement of the well known fact that if $Z^{(n)}$ has the n-dimensional standard Gaussian distribution, then $Z^{(n)}/\|Z^{(n)}\|_n$ is uniformly distributed on the unit sphere $\mathbb{S}^{n-1}$, and independent of $\|Z^{(n)}\|_n$.

□

Note that Lemma 3.1 states that for any given $\sigma \in \mathcal{P}(\mathbb{S})$, we can find a corresponding $\zeta \in \mathcal{P}(\mathbb{A})$. Fix such a pair (σ, ζ). Now, for $z \in \mathbb{A}$, define

(3.2) $$\widehat{W}_z^{(n)} \doteq \frac{1}{n} \sum_{i=1}^n X_i z_i^{(n)}.$$

Then, given $W_\theta^{(n)}$ as defined in (1.5), and any good rate function $\mathbb{I} : \mathbb{R} \to [0, \infty]$, Lemma 3.1 implies that

$$\sigma \left(\theta \in \mathbb{S} : (W_\theta^{(n)})_{n \in \mathbb{N}} \text{ satisfies an LDP with good rate function } \mathbb{I} \right)$$

(3.3) $$= \zeta \left(z \in \mathbb{A} : \left(\frac{\sqrt{n}}{\|z^{(n)}\|_n} \widehat{W}_z^{(n)} \right)_{n \in \mathbb{N}} \text{ satisfies an LDP with good rate function } \mathbb{I} \right).$$

In addition, Lemma 3.1 yields a large class of examples of σ satisfying (H1), constructed via ζ satisfying (3.1). We specify two such examples below.

EXAMPLE 3.2.

a. Consider the completely independent case, where the elements $Z_i^{(n)}$, $i = 1, \ldots, n$, $n \in \mathbb{N}$, are all independent; then the law of $\mathcal{R}(Z)$ is the product measure $\sigma = \bigotimes_{n \in \mathbb{N}} \sigma_{n-1}$, where each row $\theta^{(n)}$ of θ is independent under σ. As previously noted, the tail sigma-algebra $\mathcal{T}$ induced by the rows (defined in (2.4)), is trivial in this case due to the Kolmogorov zero-one law.

b. Alternatively, consider the following highly dependent case: let $\zeta \in \mathcal{P}(\mathbb{A})$ satisfy (3.1) such that for ζ-a.e. $z \in \mathbb{A}$, we have $z_i^{(n)} = z_i^{(m)}$ for all $i \in \mathbb{N}$ and $m, n \geq i$ (i.e., constant within columns). Then, let $\sigma = \zeta \circ \mathcal{R}^{-1}$, so that σ satisfies (H1) by Lemma 3.1. In this case, there is strong dependence across rows which precludes a claim regarding triviality of the tail sigma-algebra $\mathcal{T}$ induced by the rows. In fact, consider the event

$$A \doteq \left\{ \theta \in \mathbb{S} : \lim_{n \to \infty} \sqrt{n} \theta_1^{(n)} > 0 \right\}$$

Note that A is measurable with respect to $\mathcal{T}$. However, we also have due to the strong law of large numbers (ζ-a.e., as stated precisely in (3.4)),

$$\sigma(A) = \zeta \left(z \in \mathbb{A} : \lim_{n \to \infty} \sqrt{n} z_1^{(n)} / \|z^{(n)}\|_{n,2} > 0 \right) = \zeta \left(z \in \mathbb{A} : z_1^{(1)} > 0 \right) = \frac{1}{2}.$$

That is, $\mathcal{T}$ is non-trivial, and so $\mathbb{I}_\sigma$ cannot *a priori* be declared as σ-a.e. constant through a simple analysis of the tail sigma-algebra.

REMARK 3.3. We assume the condition (H1) not in an attempt to be as general as possible, but rather to point out that the universality of the rate function is a genuinely interesting phenomenon. Specifically, if we only consider the independent case of Example 3.2a., then the fact that $\mathbb{I}_\sigma$ is "universal" (in that it does not depend on θ) is a consequence of the fact that the tail sigma-algebra $\mathcal{T}$ is trivial. However, Example 3.2b. shows that universality of the rate function is a more general phenomenon that holds even when $\mathcal{T}$ is non-trivial. The condition (H1) only imposes constraints on the "marginal" distribution of the n-th row of the array θ, and imposes no restrictions on the dependence across different rows $\theta^{(n)}$, $n \in \mathbb{N}$. In fact, for $Z \sim \zeta$ satisfying (3.1), the elements of Z need not even be *jointly* Gaussian in order for the law of $\mathcal{R}(Z)$ to satisfy (H1).

3.2. *Exponential equivalence.* As a consequence of Lemma 3.1 and the equality in (3.3), we can replace σ-a.e. statements about $W_\theta^{(n)}$, $n \in \mathbb{N}$ with ζ-a.e. statements about $(\sqrt{n}/\|z^{(n)}\|_n)\widehat{W}_z^{(n)}$, $n \in \mathbb{N}$. In this section, we go further and explain why in the large deviations setting, we can ignore the contribution of the multiplicative factor $\sqrt{n}/\|z^{(n)}\|_n$. That is, we show that such a factor yields an exponentially equivalent sequence, defined as follows.

DEFINITION 3.4. Let $(\xi_n)_{n\in\mathbb{N}}$ and $(\tilde{\xi}_n)_{n\in\mathbb{N}}$ be two sequences of $\mathbb{R}$-valued random variables such that for all $\delta > 0$,

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(|\xi_n - \tilde{\xi}_n| > \delta) = -\infty;$$

then $(\xi_n)_{n\in\mathbb{N}}$ and $(\tilde{\xi}_n)_{n\in\mathbb{N}}$ are said to be *exponentially equivalent*.

PROPOSITION 3.5 ([9]). *If $(\xi_n)_{n\in\mathbb{N}}$ is a sequence of random variables that satisfies an LDP with good rate function $\mathbb{I}$, and $(\tilde{\xi}_n)_{n\in\mathbb{N}}$ is another sequence that is exponentially equivalent to $(\xi_n)_{n\in\mathbb{N}}$, then $(\tilde{\xi}_n)_{n\in\mathbb{N}}$ satisfies an LDP with good rate function $\mathbb{I}$.*

LEMMA 3.6. *Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of random variables that satisfies an LDP with a good rate function $\mathbb{I}$. Let $(a_n)_{n\in\mathbb{N}}$ be a deterministic sequence such that $a_n \to 1$ as $n \to \infty$, and let $(\tilde{\xi}_n)_{n\in\mathbb{N}}$ be another sequence defined by:*

$$\tilde{\xi}_n = a_n\xi_n, \quad n \in \mathbb{N}.$$

If $\mathbb{I}$ is quasiconvex — that is, if the set $\{x \in \mathbb{R} : \mathbb{I}(x) \in (-\infty, c)\}$ is convex for all $c \in \mathbb{R}$ — then $(\xi_n)_{n\in\mathbb{N}}$ and $(\tilde{\xi}_n)_{n\in\mathbb{N}}$ are exponentially equivalent.

PROOF. For $\varepsilon > 0$, let $N_\varepsilon < \infty$ be such that for all $n \geq N_\varepsilon$, we have $|1 - a_n| < \varepsilon$. For $n \geq N_\varepsilon$ and any $\delta > 0$,

$$|\tilde{\xi}_n - \xi_n| \geq \delta \quad \Leftrightarrow \quad |\xi_n| \cdot |1 - a_n| \geq \delta \quad \Rightarrow \quad |\xi_n| \geq \tfrac{\delta}{\varepsilon}.$$

Because $\mathbb{I}$ is lower semicontinuous and has compact level sets, it achieves its global minimum at some (not necessarily unique) $\bar{x} \in \mathbb{R}$. Fix $\delta > 0$ and let $\varepsilon > 0$ be small enough such that $|\bar{x}| < \tfrac{\delta}{\varepsilon}$. Then,

$$\begin{aligned}
\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(|\tilde{\xi}_n - \xi_n| > \delta) &\leq \limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(|\xi_n| \geq \tfrac{\delta}{\varepsilon}) \\
&\leq -\inf_{|x|\geq\delta/\varepsilon}\mathbb{I}(x) \\
&= -\min\left[\mathbb{I}(\tfrac{\delta}{\varepsilon}), \mathbb{I}(-\tfrac{\delta}{\varepsilon})\right].
\end{aligned}$$

The second inequality follows from the LDP for $(\xi_n)_{n\in\mathbb{N}}$. The last equality follows from the fact that if a quasiconvex function has a global minimizer $\bar{x}$, then it is non-increasing for $x < \bar{x}$, and non-decreasing for $x > \bar{x}$ [15, Lemma 1]. Hence, since the rate function $\mathbb{I}$ is quasiconvex and has a global minimizer $\bar{x}$ which satisfies $|\bar{x}| < \delta/\varepsilon$, it follows that if $x \geq \delta/\varepsilon$ (resp., $x \leq -\delta/\varepsilon$), then we have $\mathbb{I}(x) \geq \mathbb{I}(\delta/\varepsilon)$ (resp., $\mathbb{I}(x) \geq \mathbb{I}(-\delta/\varepsilon)$). Lastly, take the limit as $\varepsilon \to 0$, and use the compactness of the level sets of $\mathbb{I}$ to conclude that $\mathbb{I}(\tfrac{\delta}{\varepsilon}) \to +\infty$ and $\mathbb{I}(-\tfrac{\delta}{\varepsilon}) \to +\infty$. This proves the required exponential equivalence. $\square$

Fix ζ satisfying (3.1). Due to the strong law of large numbers, we have that for ζ-a.e. $z \in \mathbb{A}$,

$$(3.4) \qquad \frac{\sqrt{n}}{\|z^{(n)}\|_n} = \left(\frac{1}{n} \sum_{i=1}^{n} (z_i^{(n)})^2 \right)^{-1/2} \xrightarrow{n \to \infty} 1.$$

Thus, we are in a prime position to apply Lemma 3.6, which motivates the analysis of an LDP for $(\widehat{W}_z^{(n)})_{n \in \mathbb{N}}$.

3.3. *Proof of the LDP for* $(W_\theta^{(n)})_{n \in \mathbb{N}}$. We aim to prove an LDP for the sequence $(\widehat{W}_z^{(n)})_{n \in \mathbb{N}}$; that is, an LDP for sums of independent but not identically distributed random variables (where the lack of identical distribution comes from the inhomogeneous weights $z_i^{(n)}$ within the sum). The Gärtner-Ellis theorem (recalled below) is well suited for such an LDP.

THEOREM 3.7 (Gärtner-Ellis). *Let* $(\xi_n)_{n \in \mathbb{N}}$ *be a sequence of* $\mathbb{R}$-*valued random variables. Suppose that the* limit log mgf $\bar{\Lambda} : \mathbb{R} \to [0, \infty)$ *defined by*

$$\bar{\Lambda}(t) \doteq \lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}[e^{tn\xi_n}]$$

is finite and differentiable at all $t \in \mathbb{R}$. *Then* $(\xi_n)_{n \in \mathbb{N}}$ *satisfies an LDP with the convex good rate function* $\bar{\Lambda}^*$, *the Legendre transform of* $\bar{\Lambda}$.

For a proof of Theorem 3.7, we refer to [10, Theorem V.6], which also includes a more general version of the Gärtner-Ellis theorem that applies even if $\bar{\Lambda}$ is finite for only *some* $t \in \mathbb{R}$ (under mild additional conditions).

The following lemma establishes a property of Ψ which will be used in the application of the Gärtner-Ellis theorem.

LEMMA 3.8. *Suppose that*

$$(3.5) \qquad \forall t \in \mathbb{R}, \quad \int_{\mathbb{R}} |\Lambda(tu)| \, \nu(du) < \infty.$$

Then, the function Ψ *of* (2.2) *is differentiable on* $\mathbb{R}$.

PROOF. For each $t \in \mathbb{R}$, differentiability of Ψ at t follows from the differentiability of $t \mapsto \Lambda(tu)$ for all $u \in \mathbb{R}$, and an application of the dominated convergence theorem with the dominating function

$$g_t(u) \doteq |\Lambda'((t-1)u)u| + |\Lambda'((t+1)u)u|, \quad u \in \mathbb{R}.$$

Indeed, fix $t \in \mathbb{R}$ and for each $\delta \in (-1, 1)$ and $u \in \mathbb{R}$, define the difference quotient $R_{t,\delta}(u) \doteq [\Lambda((t+\delta)u) - \Lambda(tu)]/\delta$. Then,

$$|R_{t,\delta}(u)| \leq \sup \left\{ |\Lambda'((t+\alpha)u)u| : \alpha \in [-1, 1] \right\} \leq g(u),$$

where the last inequality uses the fact that $t \mapsto u\Lambda'(tu)$ is monotone. To show that g_t is integrable, first note that the convexity of Λ implies that for $u, s \in \mathbb{R}$,

$$\Lambda(su) - \Lambda(0) \le \Lambda'(su)\,su \le \Lambda(2su) - \Lambda(su),$$

and hence,

$$|\Lambda'(su)su| \le |\Lambda(0)| + |\Lambda(su)| + |\Lambda(2su)|.$$

Since, by the assumption (3.5), for every $s \in \mathbb{R}$, the right-hand side is an integrable function of u, it follows that g_t is also integrable for every $t \in \mathbb{R}$.

$\square$

PROOF OF THEOREM 2.4. Due to Lemma 3.1 (in particular, its consequence, (3.3)), it suffices to prove a ζ-a.e. LDP for the sequence $((\sqrt{n}/\|z\|_n)\widehat{W}_z^{(n)})_{n\in\mathbb{N}}$, where $\widehat{W}_z^{(n)}$ is defined as in (3.2). Due to Lemma 3.6 and the limit (3.4), it suffices to prove a ζ-a.e. LDP for $(\widehat{W}_z^{(n)})_{n\in\mathbb{N}}$. To this end, we consider the Gärtner-Ellis limit log mgf for the sequence $(\widehat{W}_z^{(n)})_{n\in\mathbb{N}}$. For every $n \in \mathbb{N}$ and $t \in \mathbb{R}$, we have due to the independence of X_i, $i = 1, \cdots, n$,

$$(3.6) \quad \Lambda_{n,z}(t) \doteq \frac{1}{n}\log\mathbb{E}\left[\exp\left(tn\widehat{W}_z^{(n)}\right)\right] = \frac{1}{n}\log\prod_{i=1}^{n}\mathbb{E}\left[\exp\left(tX_iz_i^{(n)}\right)\right] = \frac{1}{n}\sum_{i=1}^{n}\Lambda(tz_i^{(n)}).$$

We first claim that for ζ-a.e. $z \in \mathbb{A}$, the Gärtner-Ellis limit log mgf, the limit of (3.6), satisfies, for each $t \in \mathbb{R}$,

$$(3.7) \qquad\qquad \lim_{n\to\infty}\Lambda_{n,z}(t) = \int_{\mathbb{R}}\Lambda(tu)\nu(du) = \Psi(t),$$

with Ψ as defined in (2.2).

We proceed by proving the following modified claim (obtained by interchanging the quantifiers in our original claim): for each $t \in \mathbb{R}$, for ζ-a.e. $z \in \mathbb{A}$, the expression (3.7) holds. Note that if z were an i.i.d. sequence instead of a triangular array, our modified claim would follow from the usual strong law of large numbers. However, the strong LLN does not necessarily extend to empirical means of rows of i.i.d. random variables in a triangular array (see, e.g., [18, Example 5.41]). On the other hand, if the common distribution of the i.i.d. elements (in our case, each of the random variables $\Lambda(tz_i^{(n)})$, $i = 1, \ldots, n$, $n \in \mathbb{N}$) has finite fourth moment, then the strong LLN follows from a standard weak LLN and Borel-Cantelli argument [18, p.113, (i)]. Due to our assumption (H2), it follows that for all $t \in \mathbb{R}$, for ζ-a.e. $z \in \mathbb{A}$, the limit (3.7) holds.

Next, we aim to interchange the quantifiers to establish the original claim. Note that for each $n \in \mathbb{N}$, $\Lambda_{n,z}$ of (3.6) is a convex function (since it is the sum of convex functions). Now, let $T \subset \mathbb{R}$ be countable and dense. Then, it follows from countable additivity that for ζ-a.e. $z \in \mathbb{A}$, the convex functions $\Lambda_{n,z}(t)$ converge pointwise as $n \to \infty$ to $\Psi(t)$, for all t in the dense subset $T \subset \mathbb{R}$. Hence, the convex analytic considerations of [17, Theorem 10.8] imply that the pointwise convergence of $\Lambda_{n,z}(t)$ to $\Psi(t)$ holds for all $t \in \mathbb{R}$. That is, for ζ-a.e. $z \in \mathbb{A}$, for all $t \in \mathbb{R}$, the limit (3.7) holds, proving our original claim.

Since (H2) holds, $\Psi(t) < \infty$ for all $t \in \mathbb{R}$ and, because (3.5) follows trivially from (H2), Lemma 3.8 implies that Ψ is differentiable on $\mathbb{R}$. Therefore, by the Gärtner-Ellis Theorem (Theorem 3.7), for ζ-a.e. $z \in \mathbb{A}$, the sequence $(\widehat{W}_z^{(n)})_{n \in \mathbb{N}}$ satisfies an LDP with good rate function Ψ^*. $\square$

4. Atypicality.   In this section, we compare the rate function $\mathbb{I}_\sigma$ with the Cramér rate function $\mathbb{I}_t$. We first use Jensen's inequality to compare the associated log mgfs Ψ and Λ.

LEMMA 4.1.   *Assume* (H3)*, and let* Λ *and* Ψ *be defined as in* (1.3) *and* (2.2)*, respectively.*

   a. *If* $\Lambda \circ \sqrt{\cdot}$ *is concave on* $\mathbb{R}_+$*, then* $\Psi(t) \leq \Lambda(t)$ *for all* $t \in \mathbb{R}$*.*
   b. *If* $\Lambda \circ \sqrt{\cdot}$ *is convex on* $\mathbb{R}_+$*, then* $\Psi(t) \geq \Lambda(t)$ *for all* $t \in \mathbb{R}$*.*
   c. *If* $\Lambda \circ \sqrt{\cdot}$ *is concave or convex, but not linear, on* $\mathbb{R}_+$*, then* $\Lambda(t) = \Psi(t)$ *if and only if* $t = 0$*.*

PROOF.   We begin with part a. Let ν be the standard Gaussian distribution, and let $Z \sim \nu$ be a standard Gaussian random variable. Then, for all $t \in \mathbb{R}$, we have

$$\Psi(t) = \mathbb{E}[\Lambda(tZ)]$$
$$\text{(symmetry)} \quad = \mathbb{E}\left[\Lambda\left((t^2 Z^2)^{1/2}\right)\right]$$
$$\text{(Jensen)} \quad \leq \Lambda\left(\mathbb{E}[t^2 Z^2]^{1/2}\right)$$
$$= \Lambda(t).$$

Similar calculations can be used to establish part b. As for part c., recall that in Jensen's inequality, equality holds if and only if either: (i) $\Lambda \circ \sqrt{\cdot}$ is linear; or (ii) the underlying random variable is almost surely constant. Note that (i) is not the case by assumption. As for (ii), this holds if and only if $t^2 Z^2$ is almost surely constant, which is the case if and only if $t = 0$. $\square$

Before we prove the theorem, we recall some basic facts about the log mgf of $X_1 \sim \gamma$. Let the *domain* of a function $f : \mathbb{R} \to \mathbb{R}$ be the set $D_f \doteq \{x \in \mathbb{R} : f(x) < \infty\}$. For a set $D \subset \mathbb{R}$, let D° denote the interior of D.

LEMMA 4.2.   *Let* $\Lambda(t) = \log \mathbb{E}[e^{tX_1}]$ *be the log mgf of some random variable* X_1*. Then,*

   1. Λ *is lower semicontinuous;*
   2. Λ *is smooth in* D_Λ°*;*
   3. Λ *is convex.*

Furthermore, if X_1 *is non-degenerate (i.e., not a.s. constant), then*

   4. Λ *is strictly convex in* D_Λ°*;*
   5. Λ^* *is differentiable in* $D_{\Lambda^*}^\circ$*;*

6. *for $x \in D_{\Lambda^*}^\circ$, the maximum in the definition of the Legendre transform is uniquely attained — that is, the following quantity is well defined:*

$$(4.1) \qquad\qquad t_x \doteq \arg\max\{tx - \Lambda(t)\}.$$

PROOF. These are mostly standard, but we provide sketches of the proofs. For 1., lower semicontinuity follows from Fatou's lemma. For 2., smoothness follows from interchanging differentiation and expectation. Convexity in 3. and strict convexity in 4. follow from Hölder's inequality. As for 5., it is classical that if a function is lower semicontinuous and strictly convex in the interior of its domain, then its Legendre transform is differentiable in the interior of its domain (see [17, Theorem 26.3]). Lastly, for 6., it is also classical that for $x \in D_{\Lambda^*}^\circ$, we have $t_x = (\Lambda^*)'(x)$ (see [17, Theorem 26.5]). $\qquad\square$

PROOF OF THEOREM 2.5. Assume without loss of generality that X_1 is non-degenerate. If it were degenerate, then due to the symmetry condition (H3), the law of X_1 must be that $\gamma = \delta_0$, in which case $\Lambda = \Psi = 0$. Therefore, $\mathbb{I}_\sigma$ and $\mathbb{I}_\iota$ are both equal to the characteristic function at 0 (which is equal to 0 at $w = 0$ and $+\infty$ for all other w), and the result is trivial.

Suppose $\Lambda \circ \sqrt{\cdot}$ is concave (the convex case is similar, but with inequalities reversed). Due to Lemma 4.1, we have $\Psi(t) \leq \Lambda(t)$ for all $t \in \mathbb{R}$, which due to the definition of the Legendre transform implies that $\mathbb{I}_\sigma(w) = \Psi^*(w) \geq \Lambda^*(w) = \mathbb{I}_\iota(w)$ for all $w \in \mathbb{R}$, thus proving a. (and b. for the convex case).

Further assume the stronger condition of c., that $\Lambda \circ \sqrt{\cdot}$ is concave but not linear. Then, for $w \in \mathbb{R}$ such that $\Lambda^*(w) < \infty$, let t_w be as in (4.1), which is well defined due to the non-degeneracy condition of Lemma 4.2. Then,

$$\begin{aligned}
\mathbb{I}_\sigma(w) = \Psi^*(w) &\geq t_w w - \Psi(t_w) \\
&\geq t_w w - \Lambda(t_w) \\
&= \Lambda^*(w) = \mathbb{I}_\iota(w).
\end{aligned}$$

Due to Lemma 4.1, the second inequality above is an equality if and only if $t_w = 0$, which occurs if and only if $(\Lambda^*)'(w) = 0$. Note that Λ is symmetric, so Λ^* is also symmetric (by definition of the Legendre transform). Moreover, the smoothness of Λ (see Lemma 4.2), implies the strict convexity of Λ^* within its domain (see [17, Theorem 26.3]). Thus, $(\Lambda^*)'(w) = 0$ if and only if $w = 0$. This yields the claim of part c. $\qquad\square$

REMARK 4.3. In this paper, we address the "atypical" nature of the directions $\iota^{(n)} = (1, 1, \ldots, 1)$ associated with Cramér's theorem for large deviations of product measures. But in fact, the notions of atypicality and universal rate function extend beyond the product case. In particular, the companion paper [12] establishes LDPs for random projections of random vectors distributed according to the uniform measure on ℓ^p balls, again with a rate function that coincides for σ-a.e. sequence of directions, and the sequence of directions $\iota^{(n)} = (1, 1, \ldots, 1)$, $n \in \mathbb{N}$, can be shown to be atypical in that setting as well.

REFERENCES

[1] BARTHE, F. and KOLDOBSKY, A. (2003). Extremal slabs in the cube and the Laplace transform. *Advances in Mathematics* **174** 89–114.

[2] BOOK, S. A. (1972). Large deviation probabilities for weighted sums. *The Annals of Mathematical Statistics* **43** 1221–1234.

[3] BOOK, S. A. (1973). A large deviation theorem for weighted sums. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **26** 43–49.

[4] BOVIER, A. and MAYER, H. (2015). A conditional strong large deviation result and a functional central limit theorem for the rate function. *ALEA, Latin American Journal of Probability And Mathematical Statistics* **12** 533-550.

[5] CHERNOFF, H. (1956). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics* **27** 1-22.

[6] CHI, Z. (2001). Stochastic sub-additivity approach to the conditional large deviation principle. *The Annals of Probability* **29** 1303–1328.

[7] CRAMÉR, H. (1938). Sur un nouveau théoréme–limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles* **736** 5-23.

[8] DEMBO, A. and KONTOYIANNIS, I. (2002). Source coding, large deviations, and approximate pattern matching. *IEEE Transactions on Information Theory* **48** 1590–1615.

[9] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2 ed. Springer.

[10] DEN HOLLANDER, F. (2008). *Large Deviations. Fields Institute Monographs* **14**. American Mathematical Society.

[11] FELLER, W. (1970). *An Introduction to Probability Theory and Its Applications, volume II*. John Wiley & Sons.

[12] GANTERT, N., KIM, S. S. and RAMANAN, K. (2015). Large deviations for random projections of ℓ^p balls. *Preprint*.

[13] GANTERT, N., RAMANAN, K. and REMBART, F. (2014). Large deviations for weighted sums of stretched exponential random variables. *Electronic Communications in Probability* **19** 1–14.

[14] KIESEL, R. and STADTMÜLLER, U. (2000). A large deviation principle for weighted sums of independent identically distributed random variables. *Journal of Mathematical Analysis and Applications* **251** 929–939.

[15] LUENBERGER, D. G. (1968). Quasi-convex programming. *SIAM Journal on Applied Mathematics* **16** 1090–1095.

[16] NAJIM, J. (2002). A Cramér type theorem for weighted random variables. *Electronic Journal of Probability* **7** 1-32.

[17] ROCKAFELLAR, R. T. (1970). *Convex Analysis* **28**. Princeton University Press.

[18] ROMANO, J. P. and SIEGEL, A. F. (1986). *Counterexamples in Probability and Statistics*. CRC Press.

FAKULTÄT FÜR MATHEMATIK,
TECHNISCHE UNIVERSITÄT MÜNCHEN
E-MAIL: gantert@ma.tum.de

DIVISION OF APPLIED MATHEMATICS,
BROWN UNIVERSITY
E-MAIL: steven_kim@brown.edu; kavita_ramanan@brown.edu