

# Estimating large deviation rate functions

Ken R. Duffy\*

Brendan D. Williamson†

21<sup>st</sup> August 2015

## Abstract

Establishing a Large Deviation Principle (LDP) proves to be a powerful result for a vast number of stochastic models in many application areas of probability theory. The key object of an LDP is the large deviations rate function, from which probabilistic estimates of rare events can be determined. In order to make these results empirically applicable, it would be necessary to estimate the rate function from observations. This is the question we address in this article for the best known and most widely used LDP: Cramér's theorem for random walks.

We establish that even when only a narrow LDP holds for Cramér's Theorem, as occurs for heavy-tailed increments, one gets a LDP for estimating the random walk's rate function in the space of convex lower-semicontinuous functions equipped with the Attouch-Wets topology via empirical estimates of the moment generating function. This result may seem surprising as it is saying that for Cramér's theorem, one can quickly form non-parametric estimates of the function that governs the likelihood of rare events.

## 1 Introduction

Large deviation theory [47, 8], the study of the exponential decay in probability of unlikely events, has been used extensively in fields such as statistical mechanics [17, 46], insurance mathematics [2], queueing systems [45, 19], importance sampling [16] and many others. In each of these fields, it is typically the case that a Large Deviation Principle (LDP) is shown to hold based on an assumed underlying stochastic model of the process of interest. To quantify the rate of decay in the probability of events as a function of system size, the LDP rate function, the negative of the statistical mechanical entropy, is identified in terms of the properties of the underlying stochastic process, enabling direct estimates on the probability of rare events.

In many experimental systems, however, an *a priori* parameterization of the underlying stochastic nature of the system is unknown and must be garnered from data. In these situations, to transfer large deviation results from theory to empirical practice, it is necessary to estimate the associated rate function from data as a random function. It is the question of whether, in a non-parametric setting, this is possible and, if so, what is the speed of convergence of the estimates that is the subject considered here.

In the most commonly used LDP, Cramér's theorem for random walks, which underlies many other LDP results, we establish that non-parametric estimation of the rate function is not only possible, but that the probability of mis-estimation is, in appropriate sense, decaying exponentially in the observed sample size.

---

\*Hamilton Institute, National University of Ireland Maynooth, Ireland. E-mail: [ken.duffy@nuim.ie](mailto:ken.duffy@nuim.ie)

†Department of Mathematics, Duke University, Durham, North Carolina, USA. E-mail: [brendan.williamson@duke.edu](mailto:brendan.williamson@duke.edu)

To make matters more precise, let  $\{X_i\}$  be a sequence of real-valued i.i.d. random variables, and define  $S_n = (X_1 + \dots + X_n)/n$  to be the sample mean. If the Moment Generating Function (MGF) of  $X_1$ ,  $M(\theta) = E(\exp(\theta X_1))$  for  $\theta \in \mathbb{R}$ , is finite in a neighbourhood of the origin, then by Cramér's theorem  $\{S_n\}$  satisfies the LDP [47, 8] with a convex rate function  $I$  that has compact level sets,

$$I(x) = \sup_{\theta \in \mathbb{R}} (\theta x - \Lambda(\theta)), \text{ where } \Lambda(\theta) = \log M(\theta).$$

Speaking roughly, for large  $n$  this is suggestive of  $dP(S_n = x) \asymp \exp(-nI(x))dx$ . If, on the other hand, the MGF of  $X_1$  is not finite in a neighbourhood of the origin, as happens with heavy-tailed increments, then  $\{S_n\}$  satisfies a narrow LDP [8], where the rate function does not need to have compact level sets.

If we do not know the distribution of  $X_1$ , but observe a sequence  $X_1, \dots, X_n$ , is it possible to create non-parametric estimates of the rate function  $I$  that are well-behaved in that they converge quickly as a function of the sample size? Given that  $I$  captures the probability of unlikely events, the perhaps surprising answer will prove to be yes.

Drawing parallels with chemical engineers who estimate entropy directly rather than building parametric models, Duffield et al. [9], based on private communication with A. Dembo, proposed using the logarithm of the Maximum Likelihood Estimator (MLE) for the MGF as an estimate of the cumulant generating function of tele-traffic streams. Even though the estimator proved resistant to rigorous determination of its analytic properties, it seemed practically applicable and so was put to use, e.g. [35]. Independently and a little later, a similar approach was developed in statistical mechanics as a means of estimating equilibrium free energy differences, where it is called Jarzynski's estimator [27]. Significant examples of its use in an experimental context can be found in [26, 36, 24, 44], with a recent theoretical study provided in [42].

Given observations  $X_1, \dots, X_n$ , the estimator considered in [9] is the maximum likelihood estimator of the MGF as a random convex function:

$$M_n(\theta) = \frac{1}{n} \sum_{i=1}^n e^{\theta X_i} \text{ for } \theta \in \mathbb{R}. \quad (1)$$

From this, their proposed estimate of the Cumulant Generating Function (CGF) given  $n$  observations is

$$\Lambda_n(\theta) = \log M_n(\theta) \text{ for } \theta \in \mathbb{R},$$

which is also a convex function. Jarzynski's estimator, which could be considered as an empirical estimate of the Effective Bandwidth in teletraffic engineering [28], is given by

$$J_n(\theta) = \frac{1}{\theta} \Lambda_n(\theta) \text{ for } \theta \in \mathbb{R}. \quad (2)$$

Following [12], we use the Legendre-Fenchel transform of the CGF estimator to give the following random function as an estimate of the rate function:

$$I_n(x) = \sup_{\theta \in \mathbb{R}} (\theta x - \Lambda_n(\theta)) \text{ for } x \in \mathbb{R}. \quad (3)$$

It is the large deviation behavior of the random functions  $\{M_n\}$ ,  $\{\Lambda_n\}$ ,  $\{J_n\}$ , and  $\{I_n\}$  that is of interest to us.

Point-estimate properties for a single fixed  $\theta$  have been established for  $\{\Lambda_n\}$ . For example, assuming  $X_1$  is bounded, [25] provides concentration inequalities establishing speed of convergence of the estimate, [20] provides a means for correcting implicit bias in the estimation of effective

bandwidths, while [18, 21] provide for a Bayesian approach. Motivated by Jarzynski's estimator, the recent study [42] considers unbounded random variables, focusing on the argument  $\theta$  where the estimates become unreliable. As both our estimate of  $I$ , and many estimates of interest, depend upon  $\Lambda_n(\theta)$  for all  $\theta$ , however, for many applications it is necessary to consider  $\Lambda_n$  as a random function rather than a random, extended real-valued estimate.

As examples of functions of interest, in the random walk case, it is known, e.g. [23, 10, 32, 2], that if  $E(X_1) < 0$ , then the tail of the supremum of the random walk satisfies

$$\lim_{q \rightarrow \infty} \frac{1}{q} \log P \left( \sup_{k \geq 0} S_k > q \right) = - \inf_{x > 0} x I \left( \frac{1}{x} \right) = - \sup \{ \theta : \Lambda(\theta) \leq 0 \} =: -\delta. \quad (4)$$

This tail asymptote, dubbed Loynes' exponent in [14], has practical significance through its interpretation in terms of the ruin probabilities of an insurance company [2] and in terms of the tail asymptote for the waiting time of a single server queue [1]. A natural estimate [9] of Loynes' exponent is

$$\delta_n = \sup \{ \theta : \Lambda_n(\theta) \leq 0 \}. \quad (5)$$

As an application, results concerning the behavior of the estimates  $\{\delta_n\}$  will be established here.

As a second example of why it is valuable to have results on the stochastic properties of the estimators as random functions rather than single point values, it has recently been proved [15] that the most likely paths to a large integrated random walk with negative drift [37, 13, 30, 7] mimic scaled versions of the function  $-\Lambda(\theta)$  for  $\theta \in [0, \delta]$ , where  $\delta$  is defined in equation (4). In order to empirically estimate these nature of these paths, one must estimate the entire random function  $\Lambda$  directly from observations.

In prior work [12] it was shown that if  $X_1$  is bounded then  $\{I_n\}$ , considered as a sequence of random lower-semicontinuous functions, satisfies the LDP in a suitable topological space. The methods there do not generalize to the unbounded random variable setting, which is often of interest in practice. Motivated by the estimation of Loynes' exponent in equation (4), in [14] it is conjectured that such a generalization is true. Using a significantly distinct approach from that in [12], here we establish that this is the case for any distribution of  $X_1$  on  $\mathbb{R}$ .

## 2 A topological setup suitable for the LDP

Recall that a sequence of random elements  $\{Y_n\}$  taking values in a topological space  $(\mathcal{Y}, \tau)$  satisfies the LDP [47, 8] if there exists a lower-semicontinuous function  $I : \mathcal{Y} \mapsto [0, \infty]$  that has compact level sets such that for all  $G$  open and all  $F$  closed

$$- \inf_{y \in G} I(y) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(Y_n \in G) \text{ and } \limsup_{n \rightarrow \infty} \frac{1}{n} \log P(Y_n \in F) \leq - \inf_{y \in F} I(y). \quad (6)$$

Considering the sequences of estimators  $\{M_n\}$ ,  $\{\Lambda_n\}$  and  $\{I_n\}$  as random lower semi-continuous convex functions, we prove that they satisfy the LDP in suitable spaces equipped with appropriate topologies. In particular, we consider the  $M_n$  first as elements of

$$\mathcal{X}'_M = \{f : \mathbb{R} \mapsto [0, \infty] : f \text{ is a lower-semicontinuous convex function with } f(0) \text{ finite}\},$$

and later as elements of

$$\mathcal{X}_M = \{f : \mathbb{R} \mapsto (0, \infty] : f \in \mathcal{X}'_M, \text{ and there exists a probability measure } \nu \text{ on } \mathbb{R} \text{ such that } f(\theta) < \infty \text{ implies } f(\theta) = E_\nu(\exp(\theta x))\}.$$

The  $\Lambda_n$  will be elements of

$$\mathcal{X}_\Lambda = \{f : \mathbb{R} \mapsto (-\infty, \infty] : f(\theta) = \log g(\theta) \text{ for all } \theta \in \mathbb{R} \text{ and for some } g \in \mathcal{X}_M\}$$

and the  $I_n$  be elements of

$$\mathcal{X}_I = \left\{ f : \mathbb{R} \mapsto [0, \infty] : f(x) = \sup_{\theta \in \mathbb{R}} (\theta x - g(\theta)) \text{ for all } x \in \mathbb{R} \text{ and for some } g \in \mathcal{X}_\Lambda \right\}.$$

The space for Jarzynsky estimators,  $\{J_n\}$ , defined in (2), is a little more complex and we be elements of

$$\mathcal{X}_J = \left\{ f : \mathbb{R} \mapsto [-\infty, \infty] : f(\theta) = \frac{g(\theta)}{\theta} \text{ for all } \theta \neq 0 \text{ and for some } g \in \mathcal{X}_\Lambda, f(0) = \liminf_{\theta \rightarrow 0} f(\theta) \right\}.$$

All of these spaces are equipped with the Attouch-Wets topology [3, 4, 5], denoted  $\tau_{AW}$ , which is also known as the bounded-Hausdorff topology, and its Borel  $\sigma$ -algebra. This topology was first developed to capture a good notion of convergence of optimization problems defined through a sequence of lower-semicontinuous functions. At a fundamental level, its appropriateness for our needs is demonstrated by the functional continuity of the Legendre-Fenchel transform, as used in equation (3), [5][Theorem 7.2.11].

For proper extended real-valued functions defined over the reals, the topology is constructed in the following fashion. Each lower-semicontinuous function  $f : \mathbb{R} \rightarrow [-\infty, \infty]$  is uniquely identified with a closed set in  $\mathbb{R}^2$ , its epigraph  $\text{epi}(f) = \{(\theta, b) : b \geq f(\theta)\}$ . One then defines convergence of the functions based on a notion of convergence of closed sets. In particular, it is based on a projective limits topology using a bounded-Hausdorff idea: a sequence  $\{f_n\}$  of lower-semicontinuous functions converges to  $f$  in  $\tau_{AW}$ , if given any bounded set  $B \in \mathbb{R} \times \mathbb{R}$  and any  $\epsilon > 0$ , there exists  $N_\epsilon$  such that

$$\sup_{x \in B} |d(x, \text{epi}(f_n)) - d(x, \text{epi}(f))| < \epsilon \text{ for all } n > N_\epsilon, \text{ where } d(x, \text{epi}(f)) = \inf_{y \in \text{epi}(f)} d(x, y) \quad (7)$$

and we employ the box metric in  $\mathbb{R}^2$ ,  $d(x, y) = \max(|x_1 - y_1|, |x_2 - y_2|)$ .

Note that not every element of  $\mathcal{X}_J$  is lower semi-continuous. Every  $f \in \mathcal{X}_J$  is continuous on the interior of the interval on which it is finite, and lower semi-continuous on  $(0, \infty)$ , as in both cases every CGF is. Therefore the only point at which  $f$  can fail to be lower semi-continuous is  $a = \inf\{\theta : f(\theta) > -\infty\}$  if  $a > -\infty$ ,  $a < 0$  and  $f(a) > -\infty$ . For such functions  $f$  we will associate  $f$  with the closure of its epigraph in  $\mathbb{R}^2$ , or equivalently, with the epigraph of its lower semi-continuous regularisation,  $\bar{f}(\theta) = \liminf_{\gamma \rightarrow \theta} f(\gamma)$ . The symmetric difference of  $\text{epi}(f)$  and its closure is at most one half-line in  $\mathbb{R}^2$ , and moreover the mapping from functions in  $\mathcal{X}_J$  to the closure of their epigraphs is injective, justifying this approach. For this reason, and for ease of notation, for every  $f \in \mathcal{X}_J$  we will use the notation  $\text{epi}(f)$  to denote the closure of the epigraph of  $f$  without clarification.

As well as the continuity of the Legendre-Fenchel transform, this topology has many properties that are appropriate for our estimation problem and that more commonly used function space topologies do not possess. For example, the following sequence of functions, intended to be indicative of possible estimates of the cumulant generating function when  $X_1$  has an exponential distribution with rate 1,

$$f_n(\theta) = \begin{cases} e^\theta & \text{if } \theta \leq 1, \\ e + n(\theta - 1) & \text{if } \theta > 1, \end{cases}$$

would not be convergent in the topology of uniform convergence on compacts or the Skorohod topology, but in  $\tau_{\text{AW}}$  converge to

$$f(\theta) = \begin{cases} e^\theta & \text{if } \theta \leq 1, \\ +\infty & \text{if } \theta > 1, \end{cases}$$

which is self-evidently desirable. In particular, the topology captures closeness of functions when their effective domains do not coincide.

The LDP for each of these collection of estimators is established in the coming sections. We first prove that the sequence of maximum likelihood estimators for the MGFs,  $\{M_n\}$  defined in equation (1), satisfy the LDP. From this the LDP for estimates of the CGF and the rate function shall be obtained by the contraction principle [8].

### 3 Statement of Main Results

In order to establish the results, a substantial volume of work is necessary, including the characterization of Attouch-Wets limits of MGFs. The statement of the main results are also a little involved as it happens that  $I_M(f) < \infty$  for some functions  $f \in \mathcal{X}'_M$  that are not MGFs; such functions are what motivates the definition of  $\mathcal{X}_M$ . All proofs are deferred to Section 8.

Let the measure on  $\mathbb{R}$  corresponding to  $X_1$  be denoted  $\mu$  and let  $\nu$  be any other measure. We define the relative entropy, e.g. [8], as

$$H(\nu|\mu) = \begin{cases} \int_{\mathbb{R}} \frac{d\nu}{d\mu} \log \left( \frac{d\nu}{d\mu} \right) d\mu & \text{if } \nu \ll \mu \\ +\infty & \text{otherwise,} \end{cases}$$

where  $\nu \ll \mu$  indicates that  $\nu$  is absolutely continuous with respect to  $\mu$  and  $d\nu/d\mu$  is the Radon-Nikodym derivative.

**Definition 3.1.** For any measure  $\nu$  we define the MGF associated to  $\nu$ ,  $f_\nu \in \mathcal{X}'_M$ , by

$$f_\nu(\theta) = \nu(\exp(\theta \cdot)) = \int_{\mathbb{R}} e^{\theta x} d\nu. \quad (8)$$

**Definition 3.2.** For any function  $f \in \mathcal{X}'_M$ , let  $\mathcal{D}_f = \{\theta : f(\theta) < \infty\}$  denote its effective domain and  $\overline{\mathcal{D}_f}$  the closure of  $\mathcal{D}_f$ . For any  $\alpha, \beta$  satisfying  $-\infty \leq \alpha \leq 0 \leq \beta \leq \infty$  define  $\mathcal{D}_{[\alpha, \beta]}^c = \{f \in \mathcal{X}'_M : \overline{\mathcal{D}_f} \subset [\alpha, \beta]\}$ , the set of all functions  $f \in \mathcal{X}'_M$  whose closure of effective domain is a subset of  $[\alpha, \beta]$ . Define  $\mathcal{D}_{[\alpha, \beta]} = \{f \in \mathcal{X}'_M : \overline{\mathcal{D}_f} = [\alpha, \beta]\}$ , the set of all functions  $f$  whose closure of effective domain is  $[\alpha, \beta]$ .

When considering  $[\alpha, \beta] \subset \mathbb{R}$  we identify  $[\alpha$  with  $(\alpha$  when  $\alpha = -\infty$ , and  $\beta]$  with  $\beta)$  when  $\beta = \infty$ .

**Definition 3.3.** For any function  $f \in \mathcal{X}'_M$  that is not a MGF, i.e. for which there exists no probability measure  $\nu$  such that  $f(\theta) = f_\nu(\theta)$  for all  $\theta \in \mathbb{R}$ , but instead satisfies  $f(\theta) = f_\nu(\theta)$  for some  $\nu$  and all  $\theta \in \mathcal{D}_f$ , then we say that  $f$  mimics  $f_\nu$ .

Note that  $0 \in \mathcal{D}_f$  for all  $f \in \mathcal{X}'_M$ , so if  $f$  mimics  $f_\nu$  then  $f(0) = 1$ . Also,  $f \in \mathcal{X}'_M$  satisfies  $f \in \mathcal{X}_M$  if and only if  $f$  is a MGF or mimics one. Armed with one more set of definitions, we can state our main results.

**Definition 3.4.** Define the logarithmic operator  $\mathcal{L} : \mathcal{X}_M \rightarrow \mathcal{X}_\Lambda$  by

$$\mathcal{L}(f)(\theta) = \log(f(\theta)) \text{ for all } \theta \in \mathbb{R} \quad (9)$$

where by convention  $\log(\infty) = \infty$ . Define the Jarzynky operator  $\mathcal{J} : \mathcal{X}_\Lambda \rightarrow \mathcal{X}_J$  by

$$\mathcal{J}(f)(\theta) = \begin{cases} f(\theta)/\theta & \theta \neq 0 \\ \liminf_{\theta \rightarrow 0} f(\theta) & \theta = 0. \end{cases} \quad (10)$$

Finally, define the Legendre-Fenchel transform  $\mathcal{LF} : \mathcal{X}_\Lambda \rightarrow \mathcal{X}_I$  by

$$\mathcal{LF}(f)(x) = \sup_{\theta \in \mathbb{R}} (\theta x - f(\theta)), \quad (11)$$

for  $x \in \mathbb{R}$ .

Notice that these functions are bijective, so that their inverse exists. In fact,  $\mathcal{LF}$  is an involution [41][Theorem 26.5], justifying the presentation of the following statements.

**Theorem 3.1** (LDP for large deviation estimates). *If  $\{X_n\}$  are i.i.d. real valued random variables, then the following hold.*

1. *The sequence of empirical MGF estimators,  $\{M_n\}$  defined in (1), satisfies the LDP in  $\mathcal{X}_M$  equipped with  $\tau_{AW}$  and with the convex rate function  $I_M$  that possesses the following properties.*

(a) *For any  $f_\nu$  finite on a non-empty open interval,*

$$I_M(f_\nu) = H(\nu|\mu).$$

(b) *For  $f$  such that  $f(0) = 1$  and  $f(\theta) = \infty$  otherwise,*

$$I_M(f) = \inf_{\{\nu: f_\nu=f\}} H(\nu|\mu).$$

(c) *For any  $f \in \mathcal{D}_{[\alpha,\beta]}$  that is not a MGF, but mimics a MGF  $f_\nu$ ,*

$$I_M(f) = \begin{cases} I_M(f_\nu) & \text{if } I_M(g) < \infty \text{ for some } g \in \mathcal{D}_{[\alpha,\beta]}, g \text{ a MGF} \\ +\infty & \text{otherwise.} \end{cases}$$

2. *The sequence of empirical CGF estimators,  $\{\Lambda_n\}$ , satisfies the LDP in  $\mathcal{X}_\Lambda$  equipped with  $\tau_{AW}$  and rate function*

$$I_\Lambda(f) = I_M(\mathcal{L}^{-1}(f)) = \{I_M(g) : g(\theta) = \exp(f(\theta)) \text{ for all } \theta \in \mathbb{R}\}.$$

3. *The sequence of empirical Jarzynski estimators,  $\{J_n\}$ , satisfies the LDP in  $\mathcal{X}_J$  equipped with  $\tau_{AW}$  and rate function*

$$I_J(f) = I_M(\mathcal{L}^{-1}(\mathcal{J}^{-1}(f))) = \{I_M(g) : g(\theta) = \exp(\theta f(\theta)) \text{ for all } \theta \neq 0, g(0) = 1\}.$$

4. *The sequence of empirical rate function estimators,  $\{I_n\}$ , satisfies the LDP in  $\mathcal{X}_I$  equipped with  $\tau_{AW}$  and rate function*

$$I_I(f) = I_M(\mathcal{L}^{-1}(\mathcal{LF}^{-1}(f))) = \left\{ I_M(g) : g(\theta) = \exp\left(\sup_{x \in \mathbb{R}} (x\theta - f(x))\right) \text{ for all } \theta \in \mathbb{R} \right\}.$$

The LDPs for the sequences  $\{\Lambda_n\}$ ,  $\{J_n\}$  and  $\{I_n\}$  follow from that for  $\{M_n\}$  by the contraction principle on noting that the functionals mapping  $M_n$  to  $\Lambda_n$ ,  $\Lambda_n$  to  $J_n$  and  $\Lambda_n$  to  $I_n$  are continuous. Thus the primary effort in establishing Theorem 3.1 is to prove the LDP for the MGF estimators,  $\{M_n\}$ , which can be found in Section 4, followed by the characterization of the rate function, which is described in Section 5.

From Theorem 3.1, we prove that the estimators  $\{M_n\}$ ,  $\{\Lambda_n\}$ ,  $\{J_n\}$  and  $\{I_n\}$  converge in probability to  $f_\mu$ ,  $\Lambda_\mu$ ,  $J_\mu$  and  $I_\mu$ , the MGF, CGF, effective bandwidth and rate function of the underlying distribution. This is harder to establish than one might reasonably expect. It is a well-known result of Large Deviation Theory, e.g. [33], that the sequences of probability measures satisfying a LDP with rate function  $I$  are eventually concentrated on the level set  $\{x : I(x) = 0\}$ , which is compact. Proving eventual concentration on smaller sets has also been explored [34]. As  $I_M$  does not have a unique zero in general, convergence of  $\{M_n\}$  to  $f_\mu$  in probability is not immediate. However we do not apply the results of [34] and instead take an alternate approach to prove the following result.

**Corollary 3.1** (Weak laws).  *$\{M_n\}$  converges in probability to  $f_\mu$ , while  $\{\Lambda_n\}$ ,  $\{J_n\}$  and  $\{I_n\}$  converge in probability to  $\Lambda_\mu$ ,  $J_\mu$  and  $I_\mu$ , respectively.*

In Section 7, as an example application of these results, the following LDP is established for estimates of Loynes's exponent along with a discussion of some of the properties of its associated rate function,  $I_\delta$ .

**Theorem 3.2** (LDP for Loynes' exponent estimates). *The sequence of Loynes' estimators,  $\{\delta_n\}$  defined in (5), satisfies a LDP in  $[0, \infty]$  with rate function  $I_\delta : [0, \infty] \rightarrow [0, \infty]$ ,*

$$I_\delta(x) = \inf_{f \in C_x} I_M(f),$$

where

$$\begin{aligned} C_x &= \{f : f(x) = 1, \text{ or } f(x) < 1 \text{ and } f(y) = \infty \text{ for all } y > x\} \text{ for } x \in (0, \infty), \\ C_0 &= \{f : f(x) \geq 1 \text{ for all } x > 0\}, \\ \text{and } C_\infty &= \{f : f(x) \leq 1 \text{ for all } x \geq 0\}. \end{aligned}$$

In the proofs of some of the results in Section 7 other characterisations of  $I_\delta$  are considered. The form in Theorem 3.2 arises most naturally in our proof of the LDP and, as discussed in Section 7, is illustrative of the discontinuity of the Loynes' exponent mapping.

## 4 MGF Estimation

For the MGF estimates, we will first establish a LDP in  $\mathcal{X}'_M$ , and then reduce it to a LDP in  $\mathcal{X}_M$ . Our method of proof will be first to exclude functions that could not possibly be close to estimates and then use the super-additivity methodology pioneered by Ruelle [43] and Lanford [31], and elucidated in [33, 8], to establish the LDP. Namely, for any open  $G \in \tau_{AW}$ , define

$$\underline{m}(G) = \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(M_n \in G). \text{ and } \overline{m}(G) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log P(M_n \in G)$$

and their inf-derivatives

$$\inf_{G \ni f} \underline{m}(G) \in [-\infty, 0], \text{ and } \inf_{G \ni f} \overline{m}(G) \in [-\infty, 0]$$

where the infimum is taken over all open sets  $G$  containing  $f$  or, indeed,  $G$  in any local base of the topology around  $f$ . The inf-derivatives are referred to as the lower and upper deviation functions, respectively, [33]. When they coincide for all  $f \in \mathcal{X}'_M$ , they provide the candidate rate function

$$I_M(f) := - \inf_{G \ni f} \underline{m}(G) = - \inf_{G \ni f} \overline{m}(G) \in [0, \infty], \quad (12)$$

and the sequence  $\{M_n\}$  satisfies the weak [8], or vague [33], large deviation principle with rate function  $I_M$ . That is, with the LDP upper bound, equation (6), only holding for all compact sets rather than all closed sets. The full LDP, including goodness of the rate function, is then proved by establishing that exponential tightness holds; i.e. that there is a sequence of compact sets whose complementary probabilities are decaying at an arbitrarily high rate.

This super-additivity approach does not provide the characterisation of  $I_M$  described in Theorem 3.1 and instead that is developed in Section 5.

#### 4.1 Reduction of the space

We wish to show that equation (12) holds for all  $f \in \mathcal{X}'_M$ . We begin this process by eliminating cases where necessarily  $\inf_{G \ni f} \overline{m}(G) = -\infty$ . To determine which functions we must consider, we need to characterise the closure of the support of  $P(M_n \in \cdot)$ , as any functions,  $f$ , outside this set will have an open neighbourhood  $G$  such that  $P(M_n \in G) = 0$  for all  $n$ , so that  $\inf_{G \ni f} \overline{m}(G) = \inf_{G \ni f} \underline{m}(G) = -\infty$  and (12) holds.

Defining

$$\mathcal{X}_B = \{f_\nu : \nu \text{ is compactly supported in } \mathbb{R}\} \subseteq \mathcal{X}'_M, \quad (13)$$

we have that  $P(M_n \in \mathcal{X}_B) = 1$  for all  $n$ . To see which functions lie in its closure  $\overline{\mathcal{X}_B}$ , whose complement forms part of the set of impossible estimates, we establish the following result.

**Proposition 4.1** (Characterization of possible limits of MGFs in  $\mathcal{X}'_M$ ). *If  $\{f_n\} \subset \mathcal{X}_B$  is a convergent sequence in  $\mathcal{X}'_M$  with limit  $f \in \mathcal{X}'_M$  in  $\tau_{AW}$ , then  $f$  satisfies one of the following:*

1.  $f$  is a MGF;
2.  $f(0) < 1$ ;
3.  $f$  mimics a MGF.

*Proof of Proposition 4.1.* See Section 8. □

For the remainder of this section we refer to these classes of functions as Type 1, 2 and 3 respectively. Establishing the possible existence of these limits can be done by example.

That Type 1 functions exist is self-evident. For Type 2 functions, consider  $X_n$  equal to  $n$  or 0, each with probability  $1/2$ . Then the MGF of  $X_n$  is  $f_n(\theta) = \frac{1}{2} + e^{\theta n}/2$ . This converges in  $\tau_{AW}$  to  $f(\theta) = 1/2$  for  $\theta \leq 0$  and  $f(\theta) = \infty$  otherwise. For Type 3 functions, the function  $f(\theta) = 1$  for  $\theta \leq 0$  and  $f(\theta) = \infty$  otherwise is seen to be the limit of the functions  $f_n = (n-1)/n + e^{\theta n}/n$ , which are the MGFs of the random variables  $X_n$  that are equal to  $n$  with probability  $1/n$  and 0 otherwise. The function  $f$  mimics the MGF of the weak limit of  $\{X_n\}$ , however  $f$  itself is not a MGF. Showing that all limits of  $\{f_n\}$  are in one of these classes is a bigger task for which we adopt a common tactic when considering the limits of MGFs: an application of the Helly Selection Principle, e.g. [39], to the corresponding sequence of distribution functions.

It is worth noting that every MGF is in the closure of  $\mathcal{X}_B$ , defined in equation (13). If we take any distribution  $\nu$ , and let  $\nu_n(dx) = \nu(dx)/\nu([-n, n])$  be  $\nu$  conditioned on  $[-n, n]$ , then  $f_{\nu_n} \in \mathcal{X}_B$  for all  $n$  and  $f_{\nu_n} \rightarrow f_\nu$  in  $\tau_{AW}$ . Choosing to let  $\mathcal{X}_B$  only contain MGFs whose distributions have compact support is done only to ensure every element of  $\mathcal{X}_B$  is finite everywhere, which simplifies the proof of Proposition 4.1.

## 4.2 A local convex base

Let  $B_r = \{x : d(0, x) < r\} \in \mathbb{R}^2$  be the open ball of radius  $r$  in  $\mathbb{R}^2$  and  $\overline{B}_r = \{x : d(0, x) \leq r\} \in \mathbb{R}^2$  be its closure. To use the super-additivity approach to establish (12) we need to construct a local convex base for the topology  $\tau_{AW}$ , but this is not possible in general. The following collection of sets

$$V_k(f) = \left\{ g : \sup_{x \in \overline{B}_k} |d(x, \text{epi}(g)) - d(x, \text{epi}(f))| < \frac{1}{k} \right\}, \text{ for } k \in \mathbb{N}, \quad (14)$$

is known to form a local base for the Attouch-Wets topology [5]. The sets  $V_k(f)$  defined in equation (14) are not, however, typically convex in the sense that if  $g, h \in V_k(f)$  then we cannot deduce that  $l_\alpha$ , defined by  $l_\alpha(\theta) = \alpha g(\theta) + (1 - \alpha)h(\theta)$  for all  $\theta \in \mathbb{R}$  and for every  $\alpha \in (0, 1)$ , is in  $V_k(f)$ . As a counter example, graphically illustrated in Figure 1, consider  $V_2(g)$  where, for any  $\beta > 2$ ,

$$g(\theta) = \begin{cases} 0 & \text{if } \theta = 0, \\ \infty & \text{if } \theta \neq 0, \end{cases} \quad h(\theta) = \begin{cases} 1 - \beta\theta & \text{if } \theta \in [0, 1/\beta], \\ \infty & \text{if } \theta \notin [0, 1/\beta], \end{cases} \quad \text{and thus } l_{1/2}(\theta) = \begin{cases} 1/2 & \text{if } \theta = 0, \\ \infty & \text{if } \theta \neq 0, \end{cases}$$

so that

$$\sup_{x \in \overline{B}_2} |d(x, \text{epi}(g)) - d(x, \text{epi}(h))| = 1/\beta, \text{ but } \sup_{x \in \overline{B}_2} |d(x, \text{epi}(g)) - d(x, \text{epi}(l_{1/2}))| = 1/2,$$

so that  $g, h \in V_2(g)$ , but  $l_{1/2} \notin V_2(g)$ .

Although the base  $\{V_k(f)\}$  is not convex, it will suffice for our initial elimination of impossible MGF estimates. For the core result, we introduce a new base that is convex where it matters; that is, at functions that could appear as limits of MGF estimates.

As convergence in  $\tau_{AW}$  does not imply point-wise convergence [5], even though  $M_n(0) = 1$  for all  $n$  it is possible that  $M_n(0)$  converges to a value less than 1 in  $\tau_{AW}$ . As a direct example, consider the following sequence:

$$f_n(\theta) = \begin{cases} n\theta + 1 & \text{if } |\theta| \leq 1/n, \\ +\infty & \text{if } |\theta| > 1/n. \end{cases}$$

Point-wise we have that  $f_n(0)$  converges to 1, but  $\text{epi}(f_n)$  converges in  $\tau_{AW}$  to the epigraph of the function that is 0 at 0 and  $+\infty$  elsewhere. This is illustrated in Figure 2 and occurs as the topology of point-wise convergence is neither stronger nor weaker than  $\tau_{AW}$  [5].

As a result, we must include in our considerations functions for which  $f(0) < 1$ , but it is not always the case that a local convex base exists for these functions. To see this consider  $f$  satisfying  $f(0) < 1$  and  $f(\theta) = \infty$  for  $\theta \neq 0$ . Assume  $C_k(f)$  forms a local convex base for  $f$ , and consider  $k$  so that  $C_k(f)$  does not contain the function  $g$  satisfying  $g(0) = 1$  and  $g(\theta) = \infty$  for  $\theta \neq 0$ . Then take two functions in  $C_k(f)$ , one infinite on the right half-plane, and one infinite on the left half-plane; any non-trivial convex combination of them will equal  $g$ .

Similarly, when constructing the convex base we must rely on functions that satisfy  $f(0) > 1$ , which is why we included them in  $\mathcal{X}'_M$ . Despite these issues, the following result shows directly that the rate function evaluated at these functions is  $+\infty$ .

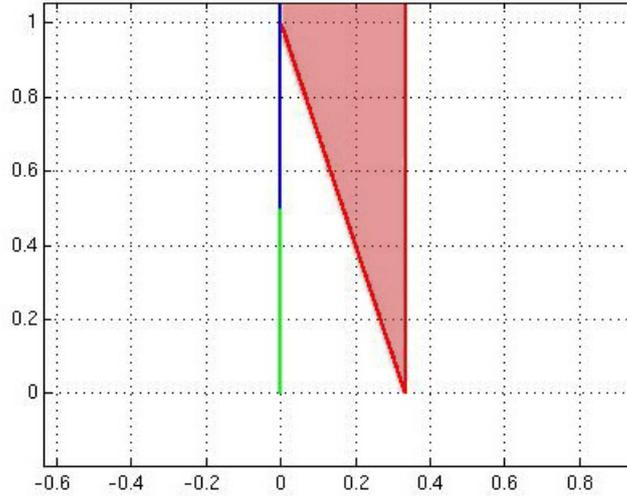


Figure 1: The usual local base for  $\tau_{AW}$  is not convex. For  $\beta = 3$  and with the functions defined in the text, this is illustrated here for  $V_2(g)$ , where  $g$  (green) and  $h$  (red) are in  $V_2(g)$ , but their linear combination  $l_{1/2}$  (blue) is not.

**Proposition 4.2** (Functions with infinite rate). *For any  $f \in \mathcal{X}'_M$  such that  $f(0) \neq 1$ ,*

$$\inf_{G \ni f} \overline{m}(G) = -\infty.$$

*Proof.* See Section 8. □

For the function satisfying  $f(0) = 1$  and  $f(x) = \infty$  for  $x \neq 0$ , we rely on the idea that  $V_k(f) \cap \mathcal{X}_B$  is convex to prove subadditivity. For the general class of functions satisfying  $f(0) = 1$  and finite on a non-empty open interval, we can construct a local convex base  $\{A_k(f)\}$  such that

$$\overline{m}(A_k(f)) = \underline{m}(A_k(f)). \quad (15)$$

This will enable us to deduce the weak [8] (or vague [33]) LDP from which the LDP follows by Proposition 4.5.

The idea for the base  $A_k(f)$  is to consider a small vertical shift of  $f$ , decrease its effective domain slightly and then intersect an element of the non-convex base defined in equation (14) around this new function with all the functions whose epigraph strictly contains the resulting curtailed function's epigraph. We begin this process by defining the shifted and curtailed functions  $\{f_k^\sharp\}$  from which the base will be built.

**Definition 4.1.** *For each  $f \in \mathcal{X}'_M$  such that  $f(0) = 1$  and  $\mathcal{D}_f \neq \{0\}$ , and for each  $k \in \mathbb{N}$ , let*

$$\eta_{f,l,k} = \inf\{\theta : (\theta, f(\theta)) \in B_{2k+2}\} \text{ and } \eta_{f,r,k} = \sup\{\theta : (\theta, f(\theta)) \in B_{2k+2}\}. \quad (16)$$

Let  $0 < \epsilon < \min\{\eta_{f,r,k}, -\eta_{f,l,k}\}$  be such that

$$d((\alpha, f(\alpha)), (\beta, f(\beta))) < \frac{1}{2k} \text{ for all } \alpha, \beta \in [\eta_{f,l,k}, \eta_{f,l,k} + \epsilon] \quad (17)$$

$$\text{and } d((\alpha, f(\alpha)), (\beta, f(\beta))) < \frac{1}{2k} \text{ for all } \alpha, \beta \in [\eta_{f,r,k} - \epsilon, \eta_{f,r,k}]. \quad (18)$$

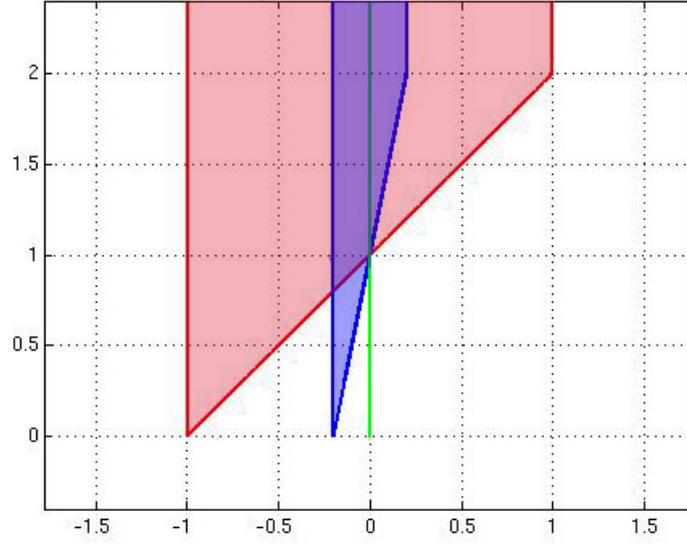


Figure 2:  $f_1$  in red,  $f_5$  in blue, and  $f$  in green.

Then we define  $f_k^\sharp$  by

$$f_k^\sharp(\theta) = \begin{cases} f(\theta) + \frac{1}{2k} & \text{if } \theta \in [\eta_{f,l,k} + \epsilon, \eta_{f,r,k} - \epsilon] \\ +\infty & \text{otherwise.} \end{cases}$$

This curtailing process is illustrated in Figure 3.

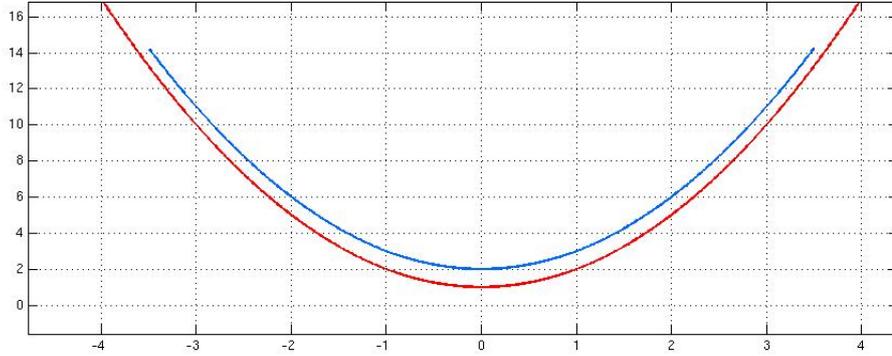


Figure 3: Illustration of the curtailing of  $f(x) = x^2 + 1$  (red) to  $f_7^\sharp$  (blue), whose shrunk domain is  $\eta_{f,r,7} = \sqrt{15} = -\eta_{f,l,7}$ . Both the  $\epsilon$  curtailing and the  $1/2k$  vertical shift of  $f_7^\sharp$  are exaggerated for illustrative purposes.

Using these shifted functions, we construct a collection of sets that we shall prove form a local, convex base at  $f$ . In order to do so, we require the following piece of notation.

**Definition 4.2.** For each  $f \in \mathcal{X}'_M$  such that  $f(0) = 1$  and  $\mathcal{D}_f \neq \{0\}$ , define

$$\theta_{f,l} = \inf \mathcal{D}_f \text{ and } \theta_{f,r} = \sup \mathcal{D}_f.$$

For two functions  $f$  and  $g$ , we write  $g \lll f$  if  $\theta_{g,l} < \theta_{f,l}$ ,  $\theta_{f,r} < \theta_{g,r}$ , and  $g(\theta) < f(\theta)$  for all  $\theta \in [\theta_{f,l}, \theta_{f,r}]$ .

Notice that equivalently we can say that  $g \lll f$  if  $\text{epi}(f) \subset \text{epi}(g)$ ,  $\theta_{f,l}$  and  $\theta_{f,r}$  are finite and

$$\inf_{x \in \partial \text{epi}(f), y \in \partial \text{epi}(g)} d(x, y) > 0,$$

where  $\partial$  denotes the boundary of a set, which is the property that motivates this definition.

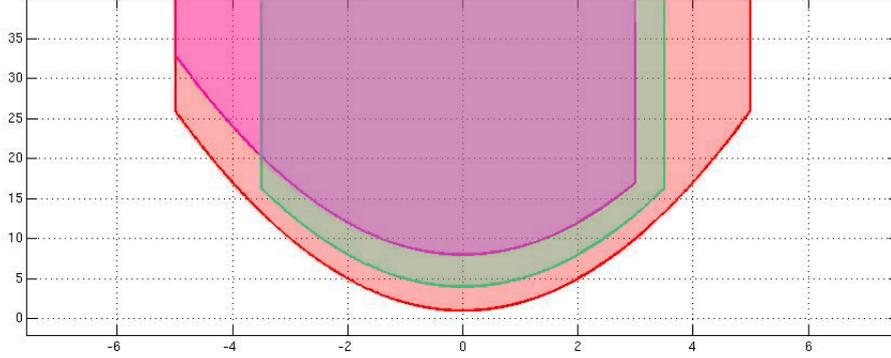


Figure 4:  $f$ ,  $g$  and  $h$  in red, purple and green respectively.  $h \lll f$ , but  $g \lll f$ , as  $\theta_{g,l} = \theta_{f,l}$ .

**Definition 4.3.** For each  $f$  such that  $f(0) = 1$  and  $\mathcal{D}_f \neq \{0\}$  and each  $k \in \mathbb{N}$ , define the set

$$A_k(f) = V_k(f_k^\#) \cap W(f_k^\#),$$

where  $V_k$  is defined in equation (14) and  $W(f_k^\#) = \{g : g \lll f_k^\#\}$ .

**Proposition 4.3** (Local convex base). For each  $f$  such that  $f(0) = 1$  and  $\mathcal{D}_f \neq \{0\}$ ,  $\{A_k(f)\}$  forms a local convex base at  $f$ .

*Proof.* See Section 8. □

### 4.3 Coincidence of the deviation functions, exponential tightness and the LDP

Using the new base, we can prove the following result, following the super-additivity method of Ruelle and Lanford, to establish Cramér's Theorem.

**Proposition 4.4** (Super-additivity). If  $f$  is such that  $f(0) = 1$  and  $\mathcal{D}_f \neq \{0\}$  then for each  $k \in \mathbb{N}$ ,

$$\overline{m}(A_k(f)) = \underline{m}(A_k(f)).$$

That is, equation (15) holds. If  $f(0) = 1$  and  $f(\theta) = +\infty$  for  $\theta \neq 0$ ,

$$\overline{m}(V_k(f)) = \underline{m}(V_k(f)).$$

*Proof.* See Section 8. □

In our setting, exponential tightness for  $\{M_n\}$  will prove to be near automatic due to the following proposition.

**Proposition 4.5** (Compactness). The set  $\{f \in \mathcal{X}'_M : f(0) \leq 1\}$  is compact.

*Proof.* See Section 8. □

As  $P(M_n(0) > 1) = 0$ , exponential tightness is immediate.

A combination of these results leads us to the LDP for  $\{M_n\}$ , albeit without a good characterization of the rate function.

**Theorem 4.1** (LDP for MGF estimators). *The sequence of empirical MGF estimates,  $\{M_n\}$ , satisfies the LDP in  $\mathcal{X}_M = \overline{\mathcal{X}_B} \cap \{f : f(0) = 1\} \subseteq \mathcal{X}'_M$  equipped with  $\tau_{AW}$  and rate function  $I_M : \mathcal{X}_M \mapsto [0, \infty]$ ,*

$$I_M(f) = -\inf_k \overline{m}(A_k(f)) = -\inf_k \underline{m}(A_k(f)).$$

*Proof.* See Section 8. □

Although as yet we do not have a good characterisation of  $I_M$ , from this result proving the LDP for the MLEs of the MGF as random functions in the Attouch-Wets topology, we can establish the LDP for the CGF estimates and the rate function estimates via somewhat involved applications of the contraction principle. We use the contraction principle with the map  $\mathcal{L}$  defined in (9) to prove the LDP for the CGF estimators,  $\{\Lambda_n\}$ .

**Lemma 4.1** (Continuity of  $\mathcal{L}$ ). *The functional  $\mathcal{L}$ , defined in (9) is continuous.*

*Proof.* See Section 8. □

Using this continuity, we are in a position to prove the result for the CGF estimates.

**Theorem 4.2** (LDP for CGF estimators). *The sequence of empirical cumulant generating function estimators,  $\{\Lambda_n\}$ , satisfies the LDP in  $\mathcal{X}_\Lambda$  equipped with  $\tau_{AW}$  and rate function*

$$I_\Lambda(f) = I_M(\mathcal{L}^{-1}(f)) = \{I_M(g) : g(\theta) = \exp(f(\theta)) \text{ for all } \theta \in \mathbb{R}\}.$$

*Proof.* See Section 8. □

Similarly, we can prove the LDP for the Jarzynski estimators,  $\{J_n\}$  by first establishing the continuity of the map  $\mathcal{J}$  defined in (10).

**Lemma 4.2** (Continuity of  $\mathcal{J}$ ). *The functional  $\mathcal{J}$ , defined in (10) is continuous.*

*Proof.* See Section 8. □

**Theorem 4.3** (LDP for Jarzynski estimators). *The sequence of empirical Jarzynski estimators,  $\{J_n\}$ , satisfies the LDP in  $\mathcal{X}_J$  equipped with  $\tau_{AW}$  and rate function*

$$I_J(f) = I_M(\mathcal{L}^{-1}(\mathcal{J}^{-1}(f))) = \{I_M(g) : g(\theta) = \exp(\theta f(\theta)) \text{ for all } \theta \neq 0, g(0) = 1\}.$$

*Proof.* See Section 8. □

Having established the LDP for the CGF estimates in Theorem 4.2, the rate function estimator result follows from another application of the contraction principle in conjunction with the continuity of the Legendre-Fenchel transform defined in equation (11). Considering  $\mathcal{LF} : \mathcal{X}_\Lambda \mapsto \mathcal{X}_I$ , the map  $\mathcal{LF}$  is a homeomorphism [4, 5] and, indeed, this is in part what leads us to this topology in order to establish these results; it correctly captures smoothness in convex conjugation.

**Theorem 4.4** (LDP for rate function estimators). *The sequence of empirical rate function estimators,  $\{I_n\}$ , satisfies the LDP in  $\mathcal{X}_I$  equipped with  $\tau_{AW}$  and rate function*

$$I_I(f) = I_M(\mathcal{L}^{-1}(\mathcal{L}\mathcal{F}^{-1}(f))) = \left\{ I_M(g) : g(\theta) = \exp\left(\sup_{x \in \mathbb{R}} (x\theta - f(x))\right) \text{ for all } \theta \in \mathbb{R} \right\}.$$

*Proof.* See Section 8. □

## 5 Characterizing the Rate Function

The super-additivity approach does not directly provide a useful form for the resulting rate functions governing the LDPs. As  $\mathcal{L}$  and  $\mathcal{L}\mathcal{F}$  are injective, to characterise  $I_I$  and  $I_\Lambda$  it suffices to characterise  $I_M$ . Based on the appearance of relative entropy in the more restricted results in [12], one anticipates it has a role to play here.

The approach we take to create a useful characterization relies heavily on continuity and inverse continuity of mappings from subsets of the set of MGFs to measures on  $\mathbb{R}$ . This reliance on continuity suggests that there may be some version of the contraction principle that could be applied directly to prove the LDP. However, since the mapping from  $\mathcal{X}_M$  to  $\mathcal{M}(\mathbb{R})$  is not well-defined (consider the function finite only at 0), and the inverse mapping is not surjective (there are functions in the effective domain of  $I_M$  that are not mapped to by any measure) this seems unlikely.

Garcia's extension of the contraction principle, [22][Theorem 1.1], almost suffices if we consider Sanov's Theorem for empirical measures in the weak topology [8] and the map taking measures to their MGFs. However, due to the possible unboundedness of the support of those measures, that map is not continuous for any  $x \in \mathcal{M}(\mathbb{R})$  and in any case  $I_M$  will turn out not to coincide with the rate function given in [22]; if  $f$  mimics  $f_\nu$ , we may get  $I_M(f) = \infty$  even though, if it were possible, an application of the contraction principle would give a finite value. Moreover, there are conditions independent of the continuity that characterise  $I_M$ , so it appears that an alternate approach is necessary.

Propositions 5.1, 5.2 and 5.3, which follow, together provide the characterization in Theorem 3.1.

### 5.1 Convexity of $I_M$

In the process of proving that  $I_M$  is convex, we must establish that  $\{(f, g) : (f + g)/2 \in G\} \subset \mathcal{X}_M \times \mathcal{X}_M$  is an open set in the product topology for each open  $G \in \mathcal{X}_M$ . That is, that averaging is a continuous operation in  $\mathcal{X}_M$ . The proof of this result indicates why convex combinations are not a continuous operation in  $\mathcal{X}'_M$ , hence the need to restrict our LDP to  $\mathcal{X}_M$  in Theorem 4.1. As an example, consider  $f_n(\theta) = 1 + \theta n$  and  $g_n(\theta) = 1 - n\theta$  for  $|\theta| \leq 1/n$  and infinite otherwise, converging to  $f$  and  $g$  finite only at 0, where they both equal 0.  $(f_n + g_n)/2$  is equal to 1 on  $[-1/n, 1/n]$  and infinite otherwise, so that it converges to  $(f + g)/2 + 1$ , not  $(f + g)/2$ .

We establish global convexity of  $I_M$  by creating an argument along the lines of [8][Lemma 4.1.21]. The conditions of that Lemma as stated do not hold here as  $\mathcal{X}_M$  is not a topological vector space: we do not have additive inverses, closure under scalar multiplication or addition, and nor is there a zero element. The proof of [8][Lemma 4.1.21], however, relies only on two deductive conclusions of those hypotheses, which we establish directly in Section 8.

**Proposition 5.1** (Convexity of  $I_M$ ). *The MGF rate function,  $I_M$ , is convex on its entire domain.*

*Proof.* See Section 8. □

## 5.2 Characterizing $I_M$ for MGFs

**Proposition 5.2** ( $I_M(f)$  for  $f$  an MGF). *For all moment generating functions  $f_\nu$  finite on a non-empty open interval,*

$$I_M(f_\nu) = H(\nu|\mu).$$

*Moreover, for  $f \in \mathcal{X}_M$  finite only at 0,*

$$I_M(f) \leq \inf_{\nu: f=f_\nu} H(\nu|\mu).$$

*Proof.* See Section 8. □

This result is proved by three smaller results, the first two of which rely on continuity or inverse continuity of the  $\nu \mapsto f_\nu$  operation when restricted to certain subsets of  $\mathcal{M}(\mathbb{R})$  or  $\mathcal{X}_M$ . First, for any moment generating function  $f_\nu$  finite on a non-empty open interval,  $I_M(f_\nu) \geq H(\nu|\mu)$ . Second, for any  $f_\nu \in \mathcal{X}_B$ ,  $I_M(f_\nu) \leq H(\nu|\mu)$ . Finally, for any moment generating function  $f_\nu$ ,  $I_M(f_\nu) \leq \inf_{\nu: f=f_\nu} H(\nu|\mu)$ . For any MGFs  $f$  finite on a non-empty open interval,  $f = f_\nu$  for exactly one measure  $\nu$ , giving us  $I_M(f_\nu) = H(\nu|\mu)$ . For  $f$  finite only at 0, more work is required.

## 5.3 Characterizing $I_M$ for the MGF finite only at 0 and for non-MGFs

In short, we prove the following result, contained in Theorem 3.1.

**Proposition 5.3** ( $I_M(f)$  for  $f$  mimicking a MGF).

(a) *For  $f$  satisfying  $f(0) = 1$  and  $f(\theta) = \infty$  otherwise,*

$$I_M(f) = \inf_{\nu: f=f_\nu} H(\nu|\mu).$$

(b) *For any  $f \in \mathcal{D}_{[\alpha, \beta]}$  that is not a MGF but instead mimics the MGF  $f_\nu$ ,*

$$\begin{aligned} I_M(f) &= I_M(f_\nu) + \inf_{\{g \in \mathcal{D}_{[\alpha, \beta]}, g \text{ a MGF}\}} I_M(g) \\ &= \begin{cases} I_M(f_\nu) & \text{if } I_M(g) < \infty \text{ for some } g \in \mathcal{D}_{[\alpha, \beta]}, g \text{ a MGF} \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

*Proof.* Proposition 5.3 will follow from Lemmas 5.1 and 5.2 below; see Section 8. □

Part (a) is, perhaps, to be expected in light of earlier results. The first equality in part (b) appears less useful than the second, but it gives more insight into the reason that the rate function takes the form that it does. It may seem surprising that functions in  $\mathcal{X}_M$  that are not MGFs lie in the effective domain of  $I_M$ , but this is the case.

We begin with a condition for when a function  $f \in \mathcal{D}_{[\alpha, \beta]}$  that is not moment generating function is in the effective domain of  $I_M$ , i.e. that any moment generating function in  $\mathcal{D}_{[\alpha, \beta]}$  is in the effective domain of  $I_M$ , along with the moment generating function that  $f$  is mimicking.

**Lemma 5.1** (Characterizations for  $f$  not a MGF). *We have the following characterizations.*

1. For  $f$  satisfying  $f(0) = 1$  and  $f(\theta) = \infty$  for  $\theta \neq 0$ ,  $I_M(f) \in \{0, \infty\}$ .
2. If  $I_M(g) < \infty$  for some function  $g \in \mathcal{D}_{[\alpha, \beta]}$ ,  $[\alpha, \beta] \neq [0, 0]$ , then  $I_M(f) \leq I_M(f_\nu)$  for  $f \in \mathcal{D}_{[\alpha, \beta]}$  mimicking  $f_\nu$ .
3. If  $f \in \mathcal{D}_{[\alpha, \beta]}$  mimics  $f_\nu$  for some  $[\alpha, \beta] \neq [0, 0]$  then  $I_M(f) \in \{I_M(f_\nu), \infty\}$ .

*Proof.* See Section 8. □

Combining statements 2 and 3 gives us that if  $f_\nu$ ,  $f$  and  $g$  satisfy  $I_M(f_\nu), I_M(g) < \infty$ ,  $g, f \in \mathcal{D}_{[\alpha, \beta]}$ , and  $f$  mimics  $f_\nu$ , then  $I_M(f) < \infty$ . Establishing conditions for when  $I_M(f) = \infty$  is quite involved and to do so we prove the following.

**Lemma 5.2** (Five equivalences). *The following are equivalent for all pairs  $(\alpha, \beta) \in [-\infty, 0] \times [0, \infty]$ :*

1. The random walk associated with  $e^{\gamma X_1}$ , where the distribution of  $X_1$  corresponds to the measure  $\mu$ , satisfies Cramér's Theorem for some  $\gamma \notin [\alpha, \beta]$ ,
2.  $I_M(f) = \infty$  for all  $f \in \mathcal{D}_{[\alpha, \beta]}^c$ ,
3.  $H(\nu|\mu) = \infty$  for all  $\nu$  with  $f_\nu \in \mathcal{D}_{[\alpha, \beta]}^c$ ,
4.  $H(\nu|\mu) = \infty$  for all  $\nu$  with  $f_\nu \in \mathcal{D}_{[\alpha, \beta]}$ ,
5.  $I_M(f) = \infty$  for all  $f \in \mathcal{D}_{[\alpha, \beta]}$ .

*Proof.* See Section 8. □

Note that in light of Proposition 5.2, when  $[\alpha, \beta] \neq [0, 0]$ , (2) and (4) can be restated as  $I_M(f_\nu) = \infty$  for all  $f \in \mathcal{D}_{[\alpha, \beta]}^c$  and  $f_\nu \in \mathcal{D}_{[\alpha, \beta]}$  respectively. Therefore (4)  $\Rightarrow$  (5) implies that  $I_M(f) = \infty$  for all  $f \in \mathcal{D}_{[\alpha, \beta]}$  if  $I_M(f) = \infty$  for all MGFs in  $\mathcal{D}_{[\alpha, \beta]}$ , giving us a sufficient condition for when  $I_M(f) = \infty$  for  $f \in \mathcal{D}_{[\alpha, \beta]}$ ,  $f$  not a MGF.

Some of these statements may not be of interest in their own right. The condition (1) is included mainly as an aid to establishing the other equivalences. Indeed, in light of Proposition 5.2 we only need (4)  $\Rightarrow$  (5) to prove Proposition 5.3. However, it will be seen in Section 6 that (1)  $\Rightarrow$  (4) is useful for characterising  $I_M$  for specific distributions. The relation (5)  $\Rightarrow$  (1) is also of interest, as it gives us the following corollary. Although it will not be used to prove any of the main results, it is useful in practical characterisations of  $I_M$  for specific distributions  $\mu$ , as seen in Section 6.

**Corollary 5.1** (To lemma 5.2). *Under the assumptions of the Theorem 3.1,*

1. There exist  $\alpha_0 \in [-\infty, 0]$ ,  $\beta_0 \in [0, \infty]$  such that  $f \in \mathcal{D}_{[\alpha, \beta]}$  mimicking  $f_\nu$  with  $I_M(f_\nu) < \infty$  satisfies  $I_M(f) = I_M(f_\nu)$  if  $[\alpha_0, \beta_0] \subset [\alpha, \beta]$ , otherwise  $I_M(f) = \infty$ .
2. For any  $f \in \mathcal{X}_M$ ,  $I_M(f) < \infty \Rightarrow \overline{\mathcal{D}_f} \supset [\alpha_0, \beta_0]$ .
3. For  $f \in \mathcal{X}_M$  finite only at 0,  $I_M(f) = 0$  if and only if  $\alpha_0 = \beta_0 = 0$ .

*Proof.* See Section 8. □

The first statement entirely characterises  $I_M$  for functions  $f$  mimicking  $f_\nu$ , saying that if the closure of the effective domain of  $f$  contains some interval then  $I_M(f) = I_M(f_\nu)$ , otherwise  $I_M(f) = \infty$ . It also tells us that  $I_M$  has a unique zero if and only if  $[\alpha_0, \beta_0] \ni \mathcal{D}_{f_\mu}$ . The second states that this same interval must be in the closure of the effective domain of any function  $g$  such that  $I_M(g) < \infty$ , giving a necessary condition for functions to be in the effective domain of  $I_M$ .

## 5.4 Proving the main results

Equipped with these results we are now able to prove Proposition 5.3. Using Propositions 5.2 and 5.3 and Theorem 4.1, as well as Theorems 4.2, 4.3 and 4.4 we can prove Theorem 3.1, from which Corollary 3.1 follows. The proofs appear subsequently in Section 8. Theorem 3.2 is established in Section 7.

## 6 Examples

As summarised in Theorem 3.1,  $I_M(f_\nu) = H(\nu|\mu)$  for all MGFs,  $f_\nu$ , finite on a non-empty open interval, while  $I_M(f) = \infty$  if  $f \in \mathcal{X}_M$  is neither a MGF nor mimics one. The question remains as to the value of  $I_M(f)$  for  $f$  mimicking a moment generating function,  $f_\nu$ , and for the special function

$$f(\theta) = \begin{cases} 1 & \text{if } \theta = 0 \\ \infty & \text{otherwise.} \end{cases}$$

This issue is reduced to that of finding  $\alpha_0$  and  $\beta_0$  defined in Corollary 5.1.

The following table gives the values of  $\alpha_0$  and  $\beta_0$  in a variety of contexts. The calculations are straight forward and are not shown. They can be evaluated by considering for what values of  $\gamma$  does  $e^{\gamma X_1}$  have a MGF finite in a neighbourhood of the origin, giving upper and lower bounds for both  $\alpha_0$  and  $\beta_0$  respectively, and by showing for what values of  $\alpha$  and  $\beta$  Lemma 5.2 (4) does not hold, implying that Lemma 5.2 (1) does not hold and hence giving lower and upper bounds for  $\alpha_0$  and  $\beta_0$  respectively.

Distribution	$\alpha_0$	$\beta_0$
$\mu$ compactly supported	$-\infty$	$\infty$
$\mu \sim \text{Normal}(\eta, \sigma^2)$	0	0
$\mu(dx) \propto e^{-e^{\lambda x}} dx$ for $x > 0$ , $\lambda > 0$	$-\infty$	$\lambda$
$\mu(dx) \propto e^{-e^{x^\lambda}} dx$ for $x > 0$ , $\lambda > 1$	$-\infty$	$\infty$

For  $\mu$  compactly supported,  $\alpha_0 = -\infty$  and  $\beta_0 = \infty$  so that  $I_M$  is finite only at MGFs, recovering [12][Theorem 1]. As the value of  $\beta_0$  depends only on the right tail, and a heavier right tail in  $\mu$  implies a smaller or equal  $\beta_0$ , and similar for left tails and  $\alpha_0$ , the values of  $\alpha_0, \beta_0$  can be gleaned for many other distributions using the results above. For example, if  $\mu \sim \text{Exp}(\lambda)$  then its left tail behaves like a bounded random variable and so  $\alpha_0 = -\infty$ , and its right tail is heavier than that of a Normal random variable, and so  $\beta_0 = 0$ . The final two distributions serve as examples of unbounded distributions for which  $\beta_0 \in (0, \infty]$ . Similar constructions yield  $\alpha_0 \in [-\infty, 0)$  and so using a combination of these distributions with others, any pair  $(\alpha_0, \beta_0)$  is attained by some distribution  $\mu$ .

## 7 Estimating Loynes' Exponent

### 7.1 A LDP for $\{\delta_n\}$

Consider Loynes' exponent, (4):

$$\delta = \sup\{\theta : \Lambda(\theta) \leq 0\} = \sup\{\theta : M(\theta) \leq 1\} \in [0, \infty].$$

If we decide to estimate Loynes' exponent, we might do so using  $M_n$  in the following way:

$$\delta_n = \sup\{\theta : M_n(\theta) \leq 1\}.$$

Although this mapping from  $\mathcal{X}_M$  to  $[0, \infty]$  is not continuous (consider the sequence of MGFs  $f_n(\theta) = e^{\theta/n}$ ), we can still use the LDP of  $\{M_n\}$  to construct a LDP for  $\{\delta_n\}$  with rate function  $I_\delta$  using Garcia's extension of the contraction principle [22]. Assuming  $P(X_1 \in (-r, r)) = 0$  for some  $r > 0$ , as in [14], makes proof immediate by an application of Puhalskii's extension of the contraction principle [40][Theorem 2.2]. Here we will deal with the general case. Throughout this section that that  $P(X_1 > 0), P(X_1 < 0) \in (0, 1)$ , as otherwise proof of an LDP and characterization of the rate function is trivially equal to that in Theorem 3.2. We also let  $x$  be the argument of the functions in  $\mathcal{X}_M$  so as not to confuse  $\delta$  and  $\delta_n$  with  $\theta$ .

The LDP for Loynes' exponent is proved directly using Garcia's extension of the contraction principle [22], the relevant part of which is recapitulated below.

**Theorem 7.1** (Garcia [22], Theorem 1). *Assume  $\Omega \xrightarrow{X_n} \mathcal{X} \xrightarrow{G} \mathcal{Y}$ ,  $\mathcal{X}, \mathcal{Y}$  are metric spaces, and  $\{X_n\}$  satisfies the large deviation principle with good rate function  $I^\#$ . Define  $G^x = \{y \in \mathcal{Y} | (\exists x_n \rightarrow x) G(x_n) \rightarrow y\}$ . If for every  $x$  with  $I^\#(x) < \infty$ ,  $G^x$  satisfies*

1. *Every sequence converging to  $x$  has a subsequence along which the function  $G$  converges.*
2. *For every  $y \in G^x$  there is a sequence  $\{x_n\}$  converging to  $x$  such that  $G(x_n) \rightarrow y$ ,  $G$  is continuous at  $x_n$  and  $I^\#(x_n) \rightarrow I^\#(x)$ .*

*Then  $\{G(X_n)\}$  satisfies the large deviation principle with good rate function  $I$  defined by*

$$I(y) = \inf\{I^\#(x) : y \in G^x\}.$$

Here,  $G$  is the Loynes' exponent mapping,  $\mathcal{X} = \mathcal{X}_M$  and  $\mathcal{Y} = [0, \infty]$ . With this result we are ready to prove Theorem 3.2; see Section 8.

Note in the statement of Theorem 3.2 that  $C_x$  is closed (and therefore compact in  $\mathcal{X}'_M$ ) for all  $x \in [0, \infty]$ . Indeed, for each  $x \in [0, \infty]$ ,  $C_x$  is the closure of the set  $\{f : \sup\{y : f(y) \leq 1\} = x\}$ , inverse image of  $x$  under the Loynes' exponent mapping. For  $x = \infty$  this set is closed, but in general  $C_x = \{f : \sup\{y : f(y) \leq 1\} = x\} \cup A_{\delta_0}$ , where  $A_{\delta_0} = \{f : f(x) \in \{1, \infty\} \text{ for all } x\}$ , the set containing the MGF of the Dirac measure  $\delta_0$ , and all functions mimicking it. This characterisation of  $C_x$  is useful as it helps us to prove certain properties of  $I_\delta$ .

### 7.2 Properties of $I_\delta$

It should be noted that a weak law can be proven for  $\{\delta_n\}$  without applying Theorem 3.2, by showing that  $P(\delta_n \in \cdot)$  is eventually concentrated on any set of the form  $[0, a)$  for any  $a > \delta_\mu$  or  $(b, \infty]$  for  $b < \delta_\mu$ , and deducing concentration on their intersection. Interestingly

this is a necessary deduction, as the rate function  $I_\delta$  may not have a unique zero. Consider the  $\text{Exp}(\lambda_1, \lambda_2)$  distribution with  $\lambda_1, \lambda_2 > 0$ , defined by the measure

$$\mu(dx) = \begin{cases} \frac{\lambda_1}{2} e^{\lambda_1 x} dx & x < 0 \\ \frac{\lambda_2}{2} e^{-\lambda_2 x} dx & x \geq 0. \end{cases}$$

This distribution has mean  $1/2(1/\lambda_2 - 1/\lambda_1)$ , and so  $\delta_\mu > 0$  if  $\lambda_2 > \lambda_1$ . Also  $H(\nu|\mu) < \infty$  for any  $\nu \sim \text{Exp}(\gamma_1, \gamma_2)$ , and  $f_\nu$  is finite on  $(-\gamma_1, \gamma_2)$ . So by Theorem 3.1(c) any  $f$  that mimics  $f_\mu$  satisfies  $I_M(f) = 0$ , and so  $I_\delta(x) = 0$  for all  $x < \theta_\mu$ . Note however, that we cannot have  $I_\delta(x) = 0$  for  $x > \theta_\mu$ .

**Lemma 7.1** (Positive on  $(\delta_\mu, \infty]$ ).  $I_\delta(x) > 0$  for all  $x > \delta_\mu$ .

*Proof.* See Section 8. □

A necessary and sufficient condition for a unique zero is stated below.

**Proposition 7.1** (Conditions for unique zero).  $I_\delta(x) > 0$  for  $x < \theta_\mu$  if and only if the random walk associated with  $e^{yX_1}$  satisfies the conditions of Cramér's Theorem for some  $y > x$ . Therefore  $I_\delta$  has a unique zero if and only if the random walk associated with  $e^{yX_1}$  obeys Cramér's Theorem for all  $y \in (0, \delta_\mu)$ .

*Proof.* See Section 8. □

Although the condition in Proposition 7.1 may seem strict, it is true for any distribution bounded above.

Without too much analysis, we can prove a number of properties of  $I_\delta$ .

**Theorem 7.2** (Properties of  $I_\delta$ ).

- (a)  $I_\delta$  is increasing on  $[\theta_\mu, \infty]$  and decreasing on  $[0, \theta_\mu]$ ,
- (b)  $I_\delta$  is finite everywhere and therefore bounded,
- (c) For all  $x \in (\theta_\mu, \infty)$ ,  $I_\delta(x) = I_M(f)$  for some  $f$  satisfying  $f(x) = 1$ ,
- (d) If  $y$  is the smallest value for which  $I_\delta(y) = 0$ , and moreover  $y > 0$  then  $I_\delta$  is continuous at all  $x \neq y$ . If  $y = 0$  then  $I_\delta$  is continuous everywhere.

*Proof.* See Section 8. □

After considering the proof of Theorem 7.2(d) it should be clear that the methods involved will not suffice to prove continuity at the smallest value of  $x$  for which  $I_\delta(x) = 0$  if  $x > 0$ , as it is possible that  $I_\delta(x) = I_M(f)$  for some  $f$  satisfying  $f(x) < 1$ . In fact this is true if  $x < \theta_\mu$ . However it is easy to prove continuity for this  $x$  if  $f(x) = 1$  or if there exists some  $g \in \mathcal{D}_{[-\infty, x]}$  with  $g(x) = \infty$  and  $I_M(g) < \infty$ , as we can then use  $g$  instead of  $f_{\mu+}$  in the proof that  $\lim_{\epsilon \downarrow 0} I_\delta(x - \epsilon) \leq I_\delta(x)$ , because although  $f(x - \epsilon) \not\rightarrow 1$ ,  $g(x - \epsilon) \rightarrow \infty$  and so  $a_\epsilon \rightarrow 1$ , as required. Whether or not there exists a  $\mu$  such that  $I_\delta$  is discontinuous at this point is not considered further here.

## 8 Proofs

### 8.1 Section 4

**Proposition 4.1. Characterization of possible limits of MGFs in  $\mathcal{X}'_M$ .** Consider any  $f \in \mathcal{X}'_M$  that is the limit of a sequence of MGFs  $\{f_n\} \subset \mathcal{X}_B$ , with corresponding distribution functions  $\{F_n\}$ . Since  $f_n \in \mathcal{X}_B$ , each  $F_n$  is uniquely determined. An application of the Helly Selection Principle tells us that if we have a sequence of distribution functions  $\{F_n\}$ , then a subsequence of them converge point-wise to a function  $F$ . We can replace  $\{F_n\}$  with this convergent subsequence and so replacing  $\{f_n\}$  with the corresponding subsequence of MGFs, we find they still converge to  $f$ .  $F$  can be seen to be monotonic increasing with codomain  $[0, 1]$ , and therefore can only have countably many discontinuities. Its upper semi-continuous regularisation [33], denoted  $F^\diamond$ , is therefore a distribution function when the sequence of probability measures corresponding to  $F_n$  is tight, i.e. when  $\sup_{x \in \mathbb{R}} \inf_n (F_n(x) - F_n(-x)) = 1$ . Since  $F$  and  $F^\diamond$  only have a countable number of discontinuities they are equal almost everywhere with respect to Lebesgue measure and so this, along with monotonicity, can be used to show that they have the same limit as  $x \rightarrow \pm\infty$ .

Here we deal with three cases regarding the limit of the distribution functions, which are exhaustive and completely characterise the possible limit functions  $F$  and  $f$ . Let  $a = \lim_{x \rightarrow -\infty} F(x)$  and  $b = \lim_{x \rightarrow \infty} F(x)$ . It suffices to prove the following. If

$$a > 0 \text{ and } b < 1,$$

then  $f$  is of Type 1 or 2. If

$$a = 0 \text{ and } b < 1, \text{ or } a > 0 \text{ and } b = 1,$$

then  $f$  is of Type 2. If

$$a = 0 \text{ and } b = 1,$$

then  $f$  is of Type 1 or 3. The proof of this final statement generalises ideas presented in [29], which does not deal with the case of moment generating functions that were not finite in a neighbourhood of the origin.

In the first case, it holds that

$$\text{for all } \epsilon > 0, x \in \mathbb{R} \text{ there exists } N_{\epsilon, x} \text{ such that } F_n(x) < F(x) + \epsilon \leq b + \epsilon \text{ for all } n > N_{\epsilon, x}.$$

Fix  $\epsilon > 0, x \in \mathbb{R}, \theta > 0$ , and see that for  $n > N_{\epsilon, x}$ ,

$$f_n(\theta) = \int_{-\infty}^x e^{\theta y} dF_n(y) + \int_x^{\infty} e^{\theta y} dF_n(y) \geq \int_x^{\infty} e^{\theta y} dF_n(y) \geq e^{\theta x} (1 - F_n(x)) > e^{\theta x} (1 - (b + \epsilon)).$$

This is true for any  $x$ , so in particular we can increase  $x$  without bound and this is still true, although it does change  $N_{\epsilon, x}$ . This gives us  $\lim_{n \rightarrow \infty} f_n(\theta) = \infty$  as we can choose  $\epsilon$  so that  $1 - (b + \epsilon) > 0$ . An analogous argument using  $a > 0$  will give us the same result for  $\theta < 0$ . Thus the  $\tau_{AW}$  limit of  $f_n$  must also satisfy  $f(\theta) = \infty$  for  $\theta \neq 0$ . It can be shown easily that  $f(0) \leq 1$ ; assume that  $f(0) > 1$ . Then  $(0, 1) \notin \text{epi}(f)$ , so that  $d((0, 1), \text{epi}(f)) = \delta > 0$ . Then as  $(0, 1) \in \text{epi}(f_n)$  for all  $n$ , we can show that for  $\epsilon < \delta$  and any set  $B \ni (0, 1)$ , (7) does not hold for any  $n$ , so that we cannot have  $f_n \rightarrow f$ . Therefore  $f(0) \leq 1$ ;  $f$  is of Type 2 if  $f(0) < 1$ , and of Type 1 if  $f(0) = 1$ .

For the second case, we can assume  $b < 1$  and  $a = 0$ , as proving the other case is analogous. In this case we still have  $f(\theta) = \infty$  for  $\theta > 0$  using the arguments from the previous case. Assume

that  $f(\theta) < \infty$  for some  $\theta < 0$ , as otherwise we are back to the aforementioned case of the function infinite everywhere but at 0. We have that

$$\begin{aligned} & \text{for all } \epsilon > 0, \text{ there exists } x_\epsilon \text{ such that } F(x_\epsilon) > b - \epsilon, \\ & \text{for all } \epsilon > 0, \text{ there exists } N_\epsilon \text{ such that } F_n(x_\epsilon) > F(x_\epsilon) - \epsilon \text{ for all } n > N_\epsilon, \\ \Rightarrow & \text{for all } \epsilon > 0, \text{ there exists } x_\epsilon, N_\epsilon \text{ such that } F_n(x_\epsilon) > b - 2\epsilon \text{ for all } n > N_\epsilon. \end{aligned}$$

Similarly as before we have that

$$\text{for all } \epsilon > 0, \text{ there exists } N_\epsilon \text{ such that } F_n(x_\epsilon) < b + \epsilon \text{ for all } n > N_\epsilon.$$

We can assume that  $x_\epsilon \rightarrow \infty$  as  $\epsilon \rightarrow 0$ . Fixing  $\epsilon > 0$ , and  $\theta < 0$  in the interior of  $\mathcal{D}_f$ , see that for  $n > N_\epsilon$  where  $N_\epsilon$  fits both of the above conditions,

$$\begin{aligned} f_n(\theta) &= \int_{-\infty}^{x_\epsilon} e^{\theta x} dF_n(x) + \int_{x_\epsilon}^{\infty} e^{\theta x} dF_n(x) \leq \int_{-\infty}^{x_\epsilon} e^{\theta x} dF_n(x) + e^{\theta x_\epsilon}(1 - F_n(x_\epsilon)) \\ &< \int_{-\infty}^{x_\epsilon} e^{\theta x} dF_n(x) + e^{\theta x_\epsilon}(1 - (b - 2\epsilon)). \end{aligned}$$

For each  $n$ , define  $X_n^*$  by the distribution function  $F_n^*(x) = F_n(x)/F_n(x_\epsilon)$  for  $x \leq x_\epsilon$ . Then  $dF_n(x) = F_n(x_\epsilon)dF_n^*(x)$  and

$$\begin{aligned} f_n(\theta) &< F_n(x_\epsilon)E(e^{\theta X_n^*}) + e^{\theta x_\epsilon}(1 - (b - 2\epsilon)) \\ &< (b + \epsilon)E(e^{\theta X_n^*}) + e^{\theta x_\epsilon}(1 - (b - 2\epsilon)). \end{aligned}$$

It can similarly be shown that

$$f_n(\theta) \geq F_n(x_\epsilon)E(e^{\theta X_n^*}) > (b - 2\epsilon)E(e^{\theta X_n^*}).$$

If we look at the point-wise limits of these functions, this gives us

$$\begin{aligned} (b - 2\epsilon) \limsup_{n \rightarrow \infty} E(e^{\theta X_n^*}) &\leq \limsup_{n \rightarrow \infty} f_n(\theta) \leq (b + \epsilon) \limsup_{n \rightarrow \infty} E(e^{\theta X_n^*}) + e^{\theta x_\epsilon}(1 - (b - 2\epsilon)) \\ \Rightarrow b \limsup_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} E(e^{\theta X_n^*}) &\leq \limsup_{n \rightarrow \infty} f_n(\theta) \leq b \limsup_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} E(e^{\theta X_n^*}) \\ &\Rightarrow \limsup_{n \rightarrow \infty} f_n(\theta) = b \limsup_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} E(e^{\theta X_n^*}). \end{aligned}$$

The  $\tau_{\text{AW}}$  limit  $f$  and  $\{f_n\}$  are convex and continuous on the interior of  $\mathcal{D}_f$  and  $\{\mathcal{D}_{f_n}\}$  respectively, so the point-wise limit of  $f_n(\theta)$  exists and equals  $f(\theta)$  on the interior of  $\mathcal{D}_f$ , which can be shown by a minor modification of [5][Lemma 7.1.2]. Therefore we have

$$f(\theta) = b \limsup_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} E(e^{\theta X_n^*}).$$

As limit superiors of convex functions are convex,  $\limsup_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} E(e^{\theta X_n^*})$  is a non-negative convex function for all  $\theta \in \mathbb{R}$  with value 1 at  $\theta = 0$  and finite on some interval in  $(-\infty, 0)$ . Therefore we know that  $\lim_{\theta \uparrow 0} \limsup_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} E(e^{\theta X_n^*}) \leq 1$ . So as  $f$  is lower semi-continuous and convex,

$$f(0) = \lim_{\theta \uparrow 0} f(\theta) = b \lim_{\theta \uparrow 0} \limsup_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} E(e^{\theta X_n^*}) \leq b,$$

so  $f$  is of Type 2.

In the third case the usc-regularisation of  $F$ ,  $F^\diamond$  is a monotonic right-continuous function, also with the above limits, and so is a distribution function. It is from this point that we follow the

work in [29]. As again we need not consider the case of  $f$  finite only at 0, let  $f$  be finite for some interval on the left of the origin. To be precise, let  $f$  be finite on  $(\alpha, 0]$ , but not on  $(-\infty, \alpha)$  (if  $-\infty < \alpha$ ). We cannot apply [38][Theorem 2] as we cannot assume that  $f$  is a moment generating function. We also cannot apply [29][Theorem 2(a)] because we cannot assume  $f$  is finite in a neighbourhood of 0, but we do not need to. Instead, we show that if  $M(x) = \sup_n F_n(x)$ , that

$$\lim_{x \rightarrow -\infty} M(x)e^{tx} = 0 \text{ for all } t \in (\alpha, 0). \quad (19)$$

Choose  $t$  and  $\gamma$  so that  $\alpha < \gamma < t < 0$ , and see that for  $x < 0$ ,

$$\begin{aligned} F_n(x) &= \int_{-\infty}^x dF_n(u) \leq \int_{-\infty}^x e^{\gamma(u-x)} dF_n(u) \leq e^{-x\gamma} f_n(\gamma), \\ \Rightarrow M(x)e^{tx} &\leq e^{x(t-\gamma)} \sup_n f_n(\gamma). \end{aligned}$$

As  $\gamma$  is in the interior of  $\mathcal{D}_f$ ,  $\{f_n(\gamma)\}$  converges and is finite for all  $n$ , so the supremum over  $n$  is finite, and taking the limit as  $x \rightarrow -\infty$  reduces the above expression to 0, as  $t - \gamma > 0$ . Now let  $\theta$  and  $\gamma$  be such that  $\alpha < \gamma < \theta < 0$ , and let  $M_\gamma = \sup_{x < 0} M(x)e^{\gamma x}$ . By the above we have shown that  $M_\gamma$  is finite. Using integration by parts, and assuming  $-N$  is a continuity point of  $F_n$ , we see that

$$\begin{aligned} \int_{-\infty}^{-N} e^{x\theta} dF_n(x) &= e^{x\theta} F_n(x) \Big|_{-\infty}^{-N} - \theta \int_{-\infty}^{-N} e^{x\theta} F_n(x) dx \\ &\leq M(-N)e^{-N\theta} - 0 - \theta \int_{-\infty}^{-N} e^{x(\theta-\gamma)} e^{\gamma x} M(x) dx \\ &\leq M(-N)e^{-N\theta} - \theta \int_{-\infty}^{-N} e^{x(\theta-\gamma)} M_\gamma dx \\ &= M(-N)e^{-N\theta} - M_\gamma \frac{\theta}{\theta - \gamma} e^{-N(\theta-\gamma)}. \end{aligned}$$

By (19), we can make this expression as small as we want for large enough  $N$ . Also, as  $F$  also satisfies  $F(x) \leq M(x)$ , we can do the same if we replace  $dF_n$  with  $dF$ . Moreover, as  $F$  and  $F^\diamond$  only disagree on at most countably many points, we can choose a large  $N$  (specifically so that  $-N$  is a point of continuity of  $F$  and  $F^\diamond$ ) so that integrating with respect to  $dF$  is equivalent to integrating with respect to  $dF^\diamond$ . Finally, as  $F_n \rightarrow F$  point-wise and therefore to  $F^\diamond$  weakly,  $-N$  is a point of continuity of  $F$  and  $F^\diamond$ , and  $e^{\theta x}$  is bounded on  $x \in [-N, \infty)$  we have

$$\int_{-N}^{\infty} e^{\theta x} dF_n(x) \rightarrow \int_{-N}^{\infty} e^{\theta x} dF(x) = \int_{-N}^{\infty} e^{\theta x} dF^\diamond(x)$$

point-wise for each  $\theta \in (\gamma, 0)$ . Combining this with the above result for the integral from  $-\infty$  to  $-N$  gives us that  $f_n(\theta) \rightarrow f_\nu(\theta)$  point-wise on  $(\gamma, 0)$ , where  $f_\nu$  is the MGF corresponding to  $F^\diamond$ . But we can make  $\gamma$  as close to  $\alpha$  as we want, so we have point-wise convergence on  $(\alpha, 0)$ . As discussed before  $\tau_{\text{AW}}$  convergence implies that  $f_n \rightarrow f$  pointwise on  $(\alpha, 0)$ , so that  $f(\theta) = f_\nu(\theta)$  for  $\theta \in (\alpha, 0)$ . Moreover as  $f$  and  $f_\nu$  are convex and lower-semicontinuous,  $f(0) = \lim_{\theta \uparrow 0} f(\theta) = \lim_{\theta \uparrow 0} f_\nu(\theta) = 1$  and  $f(\alpha) = \lim_{\theta \downarrow \alpha} f(\theta) = \lim_{\theta \downarrow \alpha} f_\nu(\theta) = f_\nu(\alpha)$  if  $\alpha$  is finite, even if  $f_\nu(\alpha) = \infty$ . If  $f$  was finite on  $[0, \beta)$  but not on  $(\beta, \infty)$  we could similarly show that  $f(\theta) = f_\nu(\theta)$  for  $\theta \in [0, \beta]$ , and so in general  $f(\theta) = f_\nu(\theta)$  for all  $\theta \in \mathcal{D}_f$ . If  $\mathcal{D}_f = \mathcal{D}_{f_\nu}$  then  $f = f_\nu$  and  $f$  is a moment generating function, and so is of Type 1. Otherwise  $f$  mimics  $f_\nu$  and so is of Type 3. □

**Proposition 4.2. Functions with infinite rate.** If we define  $\mathcal{X}_{INT}$  to be the space of all moment generating functions finite on some open interval, equipped with the subspace topology, and  $\mathcal{M}(\mathbb{R})$  to be the space of probability measures on  $\mathbb{R}$  equipped with the weak topology, then the mapping from  $\mathcal{X}_{INT}$  into the space of probability measures,  $\Phi : \mathcal{X}_{INT} \rightarrow \mathcal{M}(\mathbb{R})$ , is well defined, as seen in [38][Theorem 2]. Let  $G_m$  be a descending countable base containing  $f$ , and let  $\Phi(G_m \cap \mathcal{X}_B) = A_m$ . Let  $L_n$  be the empirical measure of  $\{X_i\}_{i=1}^n$ , and see that

$$\begin{aligned} \inf_{G \ni f} \overline{m}(G) &= \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P(M_n \in G_m) \\ &= \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P(M_n \in G_m \cap \mathcal{X}_B) \\ &= \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P(L_n \in A_m) \\ &\leq - \lim_{m \rightarrow \infty} \inf_{\nu \in \overline{A_m}} H(\nu | \mu) \\ &= - \inf_{\nu \in C} H(\nu | \mu) \end{aligned}$$

where  $C = \bigcap_{m=1}^{\infty} \overline{A_m}$ . The last line follows from [8] [Lemma 4.1.6 (b)]. Let  $\nu \in C$ . As  $\mathcal{M}(\mathbb{R})$  is metrizable, let  $U_m$  be a descending countable base for  $\nu$ , and for each  $m$  let  $\{\nu_{m,n}\}_{n=1}^{\infty} \subset A_m$  be a sequence satisfying  $\nu_{m,n} \rightarrow \nu$  as  $n \rightarrow \infty$ . If  $N_m$  is such that for  $n \geq N_m$ ,  $\nu_{m,n} \in U_m$ , and if  $\nu_m^* = \nu_{m,N_m}$ ,  $\nu_m^* \in A_m$  and  $\nu_m^* \in U_m$ , so that  $\nu_m^* \rightarrow \nu$ . For each  $\nu_m^*$  there is a corresponding  $f_m^* \in G_m$ , and so  $f_m^* \rightarrow f$ . If  $f(0) < 1$  and  $f_m^* \rightarrow f$ , then by the proof of Proposition 4.1 a subsequence of their corresponding distribution functions tends to a function that is not a distribution function, and so their measures do not converge. By contradiction, it must follow that  $C$  is empty, and so  $\inf_{G \ni f} \overline{m}(G) = -\infty$ .  $\square$

**Proposition 4.3. Local convex base.** In order to establish Proposition 4.3, it suffices to demonstrate:

1. For each  $f$  such that  $f(0) = 1$  and  $\mathcal{D}_f \neq \{0\}$  and each  $k \in \mathbb{N}$ ,  $f \in A_k(f)$ .
2. For each  $f$  such that  $f(0) = 1$  and  $\mathcal{D}_f \neq \{0\}$  and each  $k \in \mathbb{N}$ ,  $A_k(f)$  is open.
3. For each  $f$  such that  $f(0) = 1$  and  $\mathcal{D}_f \neq \{0\}$  and each  $k \in \mathbb{N}$ ,  $A_k(f)$  is convex.
4. For each  $f$  such that  $f(0) = 1$  and  $\mathcal{D}_f \neq \{0\}$  and each  $k \in \mathbb{N}$ ,  $A_{2k}(f) \subset V_k(f)$ .

As our construction of  $W(f_k^\sharp)$  is based on nested epigraphs, the following estimate, readily deducible from [5][Lemma 1.5.1], will prove useful. As  $\text{epi}(f^\sharp) \subset \text{epi}(f)$ ,

$$\begin{aligned} \sup_{x \in \overline{B}_k} |d(x, \text{epi}(f^\sharp)) - d(x, \text{epi}(f))| &= \sup_{x \in \overline{B}_k} d(x, \text{epi}(f^\sharp) \cap \overline{B}_{2k+2}) - d(x, \text{epi}(f) \cap \overline{B}_{2k+2}) \\ &\leq e_d(\text{epi}(f) \cap \overline{B}_{2k+2}, \text{epi}(f^\sharp) \cap \overline{B}_{2k+2}) \\ &= \sup_{x \in \text{epi}(f) \cap \overline{B}_{2k+2}} d(x, \text{epi}(f^\sharp) \cap \overline{B}_{2k+2}), \end{aligned} \tag{20}$$

where  $e_d(A, B)$  is the excess between two sets  $A$  and  $B$ . The first equality comes from the fact that  $x \in \overline{B}_k$  and  $(0, 2) \in \text{epi}(f^\sharp)$ , so that  $d(x, \text{epi}(f^\sharp)) = d(x, (a, b)) \leq d(x, (0, 2))$  for some  $(a, b) \in \text{epi}(f^\sharp)$ , and so  $(a, b) \in \overline{B}_{2k+2}$ . The same is true for  $f$ . Also, as  $\text{epi}(f^\sharp) \subset \text{epi}(f)$ ,  $d(x, \text{epi}(f^\sharp)) \geq d(x, \text{epi}(f))$  for all  $x \in \mathbb{R}^2$ , which justifies the removal of the absolute value.

For item 1, by construction,  $\theta_{f,l} < \theta_{f_k^\sharp,l} < \theta_{f_k^\sharp,r} < \theta_{f,r}$  and  $f(\theta) < f_k^\sharp(\theta)$  for all  $\theta \in [\theta_{f_k^\sharp,l}, \theta_{f_k^\sharp,r}]$  so that  $f \lll f_k^\sharp$  and thus  $f \in W(f_k^\sharp)$  for all  $k$ . It remains to show that  $f$  is also an element of

$V_k(f_k^\sharp)$ . From equation (20) it suffices to prove that

$$e_d(\text{epi}(f) \cap \overline{B}_{2k+2}, \text{epi}(f_k^\sharp) \cap \overline{B}_{2k+2}) < 1/k, \quad (21)$$

which will follow from the construction of  $f_k^\sharp$ . It suffices to consider  $x = (\theta, f(\theta))$  for  $\theta \in [-2k-2, 2k+2]$ . If  $\theta \in [\theta_{f_k^\sharp, l}, \theta_{f_k^\sharp, r}]$ ,  $d(x, \text{epi}(f_k^\sharp)) \leq 1/(2k)$ . If  $\theta \in [\eta_{f, l, k}, \theta_{f_k^\sharp, l}]$  or  $\theta \in [\theta_{f_k^\sharp, r}, \eta_{f, r, k}]$ , defined in equation (16), more care is needed. Consider the former and note that by the triangle inequality

$$d((\theta, f(\theta)), \text{epi}(f_k^\sharp)) \leq d\left((\theta, f(\theta)), \left(\theta_{f_k^\sharp, l}, f(\theta_{f_k^\sharp, l})\right)\right) + d\left(\left(\theta_{f_k^\sharp, l}, f(\theta_{f_k^\sharp, l})\right), \text{epi}(f_k^\sharp)\right).$$

By equation (18), the first term is less than  $1/(2k)$  and, as above, the second term is less than or equal to  $1/(2k)$  so that equation (21) is ensured and  $f \in A_k(f)$ .

For item 2, the set  $V_k(f_k^\sharp)$  is open as it is an element of a local base for  $f_k^\sharp$ , so it suffices to show that  $W(f_k^\sharp)$  is open. For ease of notation, we shall show that  $W(f)$  is open for any  $f$  such that  $-\infty < \theta_{f, l} < \theta_{f, r} < \infty$ . To do this, let  $g \in W(f)$ . We shall construct a set  $V_k(g)$  such that  $V_k(g) \subset W(f)$ .

As  $g \lll f$  and  $g$  is convex,  $g$  is continuous on  $[\theta_{f, l}, \theta_{f, r}]$  so that  $f - g$  is lower semi-continuous on this range and therefore its infimum is attained and positive:

$$\delta = \inf_{\theta \in [\theta_{f, l}, \theta_{f, r}]} (f(\theta) - g(\theta)) > 0.$$

Let

$$0 < \epsilon < \min\left(\theta_{f, l} - \theta_{g, l}, \theta_{g, r} - \theta_{f, r}, \frac{\delta}{2}, 1\right) \text{ be such that}$$

$$\max(g(\theta_{f, l} - \epsilon) - g(\theta_{f, l}), g(\theta_{f, r} + \epsilon) - g(\theta_{f, r})) < \frac{\delta}{2}.$$

By convexity, the second condition ensures that  $g(\theta \pm \epsilon) - g(\theta) < \delta/2$  for any  $\theta \in [\theta_{f, l}, \theta_{f, r}]$ . Now choose  $k \in \mathbb{N}$  so that

$$(\theta, g(\theta)) \in B_k, \text{ for all } \theta \in [\theta_{f, l} - \epsilon, \theta_{f, r} + \epsilon], \text{ and } \frac{1}{k} < \epsilon.$$

Thus for any  $h \in V_k(g)$ ,

$$\sup_{x \in B_k} |d(x, \text{epi}(g)) - d(x, \text{epi}(h))| < \frac{1}{k} \text{ and so } \sup_{\theta \in [\theta_{f, l} - \epsilon, \theta_{f, r} + \epsilon]} d((\theta, g(\theta)), \text{epi}(h)) < \epsilon.$$

So, for any  $\theta \in [\theta_{f, l} - \epsilon, \theta_{f, r} + \epsilon]$  there exists  $(x_\theta, y_\theta) \in \text{epi}(h)$  such that

$$\max(|\theta - x_\theta|, |g(\theta) - y_\theta|) < \epsilon. \quad (22)$$

In particular, consider  $\theta = \theta_{f, l} - \epsilon$ , then  $x_\theta < \theta_{f, l} - \epsilon + \epsilon = \theta_{f, l}$  and  $h(x_\theta) \leq y_\theta < g(\theta_{f, l} - \epsilon) + \epsilon < \infty$  and so  $\theta_{h, l} < \theta_{f, l}$ . It can similarly be shown that  $\theta_{h, r} > \theta_{f, r}$ .

To show that  $h \lll f$ , it remains to be proven that for any  $\theta \in [\theta_{f, l}, \theta_{f, r}]$ , that  $h(\theta) < f(\theta)$ . This will follow from convexity of  $g$  and  $h$ . Consider such a  $\theta$ , then by equation (22) at  $\theta \pm \epsilon$ , since  $x_{\theta - \epsilon} \leq \theta \leq x_{\theta + \epsilon}$  we have that

$$h(\theta) \leq \max(h(x_{\theta - \epsilon}), h(x_{\theta + \epsilon})) < \max(g(\theta - \epsilon), g(\theta + \epsilon)) + \epsilon < \left(g(\theta) + \frac{\delta}{2}\right) + \frac{\delta}{2} \leq f(\theta).$$

So if  $h \in V_k(g)$ , then  $h \in W(f)$  and thus  $W(f)$  is open as required.

For item 3, assume  $g, h \in A_k(f)$ , let  $\alpha \in [0, 1]$  and let  $l_\alpha(\theta) = \alpha g(\theta) + (1 - \alpha)h(\theta)$  for all  $\theta$ . We wish to show that  $l_\alpha \in A_k(f)$ .  $l_\alpha \in W(f_k^\sharp)$  as  $l_\alpha(\theta) \leq \max\{g(\theta), h(\theta)\}$  for all  $\theta$  and therefore  $l_\alpha \lll f_k^\sharp$ .

We must show that  $l_\alpha \in V_k(f_k^\sharp)$ . Note that for any  $x$ ,  $d(x, \text{epi}(l_\alpha)) \geq \min\{d(x, \text{epi}(g)), d(x, \text{epi}(h))\}$  as  $\text{epi}(l_\alpha) \subset \text{epi}(g) \cup \text{epi}(h)$ . Thus as  $l_\alpha, g, h \in W(f_k^\sharp)$ , we have that

$$\begin{aligned} |d(x, \text{epi}(f_k^\sharp)) - d(x, \text{epi}(l_\alpha))| &= d(x, \text{epi}(f_k^\sharp)) - d(x, \text{epi}(l_\alpha)) \\ &\leq d(x, \text{epi}(f_k^\sharp)) - \min\{d(x, \text{epi}(g)), d(x, \text{epi}(h))\} \\ &= \max\{|d(x, \text{epi}(f_k^\sharp)) - d(x, \text{epi}(h))|, |d(x, \text{epi}(f_k^\sharp)) - d(x, \text{epi}(g))|\}. \end{aligned}$$

Taking the supremum over all  $x \in \bar{B}_k$ , as  $g, h \in V_k(f_k^\sharp)$  the right hand side is less  $1/k$  and so  $l_\alpha \in V_k(f_k^\sharp)$ .

Finally, for item 4, we have that  $A_{2k}(f) = V_{2k}(f_{2k}^\sharp) \cap W(f_{2k}^\sharp) \subset V_{2k}(f_{2k}^\sharp)$  so that it suffices to show that  $V_{2k}(f_{2k}^\sharp) \subset V_k(f)$ . Let  $g \in V_{2k}(f_{2k}^\sharp)$  and, using the triangle inequality, consider

$$\begin{aligned} &\sup_{x \in \bar{B}_k} |d(x, \text{epi}(f)) - d(x, \text{epi}(g))| \\ &\leq \sup_{x \in \bar{B}_{2k}} |d(x, \text{epi}(f)) - d(x, \text{epi}(g))| \\ &\leq \sup_{x \in \bar{B}_{2k}} |d(x, \text{epi}(f)) - d(x, \text{epi}(f_{2k}^\sharp))| + \sup_{x \in \bar{B}_{2k}} |d(x, \text{epi}(f_{2k}^\sharp)) - d(x, \text{epi}(g))| \\ &< \frac{1}{2k} + \frac{1}{2k} = \frac{1}{k}, \end{aligned}$$

and so  $g \in V_k(f)$  as required.  $\square$

**Proposition 4.4. Super-additivity.** For each  $m > n$ , define the partial estimates

$$M_{n+1, m}(\theta) = \frac{1}{m - n} \sum_{i=n+1}^m \exp(\theta X_i)$$

and note that, for all  $\theta \in \mathbb{R}$ ,  $M_{1, n+m}$  is a convex combination of  $M_{1, n}$  and  $M_{n+1, n+m}$ :

$$M_{1, n+m}(\theta) = \frac{n}{n+m} M_{1, n}(\theta) + \frac{m}{n+m} M_{n+1, n+m}(\theta).$$

Assume  $f$  is such that  $f(0) = 1$  and  $\mathcal{D}_f \neq \{0\}$ , and consider fixed  $A_k(f)$ . By the convexity of  $A_k(f)$  proved in Proposition 4.3, we have that if  $M_{1, n} \in A_k(f)$  and  $M_{n+1, n+m} \in A_k(f)$ , then  $M_{1, n+m} \in A_k(f)$ , so that, by independence and identical distribution of increments,

$$\begin{aligned} P(M_{1, n+m} \in A_k(f)) &\geq P(M_{1, n} \in A_k(f), M_{n+1, n+m} \in A_k(f)) \\ &= P(M_{1, n} \in A_k(f)) P(M_{n+1, n+m} \in A_k(f)) \\ &= P(M_{1, n} \in A_k(f)) P(M_{1, m} \in A_k(f)). \end{aligned}$$

Thus the sequence  $\{\log P(M_n \in A_k(f))\}$  is super-additive and we have existence of the limit

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(M_n \in A_k(f)) = \sup_n \left( \frac{1}{n} \log P(M_n \in A_k(f)) \right),$$

as required.

If  $f(0) = 1$  and  $f(\theta) = +\infty$ , then we cannot use Proposition 4.3, but for this specific function we can show that  $V_k(f) \cap \mathcal{X}_B$  is convex in much the same way that we showed that  $A_k(f)$

is in Proposition 4.3 so that the result follows as above. See that if  $g, h \in V_k(f) \cap \mathcal{X}_B$  then  $l_\alpha = \alpha g + (1 - \alpha)h \in \mathcal{X}_B$ . Also, as they all take the value 1 at the origin, we have that  $\text{epi}(f) \subset \text{epi}(g), \text{epi}(h), \text{epi}(l_\alpha)$ , so

$$\begin{aligned} |d(x, \text{epi}(f)) - d(x, \text{epi}(l_\alpha))| &= d(x, \text{epi}(f)) - d(x, \text{epi}(l_\alpha)) \\ &\leq d(x, \text{epi}(f)) - \min\{d(x, \text{epi}(g)), d(x, \text{epi}(h))\} \\ &= \max(d(x, \text{epi}(f)) - d(x, \text{epi}(h)), |d(x, \text{epi}(f)) - d(x, \text{epi}(g))|). \end{aligned}$$

Again, by taking the supremum over all  $x \in \overline{B}_k$  we can show that  $l_\alpha \in V_k(f)$ , so that  $V_k(f) \cap \mathcal{X}_B$  is convex. As  $\mathbb{P}[M_n \in V_k(f)] = \mathbb{P}[M_n \in V_k(f) \cap \mathcal{X}_B]$  it follows as for  $A_k(f)$  that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(M_n \in V_k(f)) = \sup_n \left( \frac{1}{n} \log P(M_n \in V_k(f)) \right).$$

□

**Proposition 4.5. Compactness.** The collection of all closed sets in  $\mathbb{R}^2$  equipped with  $\tau_{AW}$  is compact. This can be deduced as a result of [5][Theorem 5.13], which proves that the space is compact when equipped with the Fell topology, and [5][Exercise 10(b), pg 144], which shows the Attouch-Wets and Fell topologies coincide on  $\mathbb{R}^2$  as the closed and bounded subsets of  $\mathbb{R}^2$  are compact. Thus to prove that  $\{f \in \mathcal{X}'_M : f(0) \leq 1\}$  is compact, it suffices to show that it is closed in this larger space.

To establish this, consider a sequence of functions  $\{f_n\} \subset \mathcal{X}'_M$  such that  $f_n(0) \leq 1$  and  $\lim_{n \rightarrow \infty} \text{epi}(f_n) = A$  in the  $\tau_{AW}$  topology. That  $A = \text{epi}(f)$  for some  $f$  can be readily shown by assuming that there exists an  $(x, y) \in A$  and  $z > 0$  such that  $(x, y + z) \notin A$ , and showing it must follow that the same holds for some  $\text{epi}(f_n)$ , which cannot be true. Thus, to show that  $\{f \in \mathcal{X}'_M : f(0) \leq 1\}$  is closed it is sufficient to show that  $f = \lim_{n \rightarrow \infty} f_n$  satisfies  $f \in \mathcal{X}'_M$  and  $f(0) \leq 1$ . Hence we will prove that: (i)  $f$  is lower semi-continuous; (ii)  $f$  is non-negative; (iii)  $(0, 1) \in \text{epi}(f)$ ; and (iv)  $f$  is convex.

(i) This follows as  $(2^{\mathbb{R}^2}, \tau_{AW_d})$  is compact and therefore complete, so that  $\text{epi}(f)$  is closed.

(ii) Let  $(x, y) \in \overline{B}_k$ . As

$$|d((x, y), \text{epi}(f_n)) - d((x, y), \text{epi}(f))| \leq \sup_{z \in \overline{B}_k} |d(z, \text{epi}(f_n)) - d(z, \text{epi}(f))|,$$

it follows that  $d((x, y), \text{epi}(f_n)) \rightarrow d((x, y), \text{epi}(f))$  as  $n \rightarrow \infty$ . This is true for all  $(x, y) \in \mathbb{R}^2$  and is a well-known feature of the Attouch-Wets topology. If  $y < 0$ ,  $d((x, y), \text{epi}(f_n))$  is bounded below by  $|y|$  by the non-negativity of  $f_n$ . Therefore  $d((x, y), \text{epi}(f)) \geq |y|$  and so  $(x, y) \notin \text{epi}(f)$ .

(iii)  $d((0, 1), \text{epi}(f)) = \lim_{n \rightarrow \infty} d((0, 1), \text{epi}(f_n)) = \lim_{n \rightarrow \infty} 0 = 0$ . As  $\text{epi}(f)$  is closed it follows that  $(0, 1) \in \text{epi}(f)$ .

(iv) Fix  $(x_1, y_1), (x_2, y_2) \in \text{epi}(f)$ ,  $\alpha \in [0, 1]$ , and let  $\alpha(x_1, y_1) + (1 - \alpha)(x_2, y_2) = (x_3, y_3)$ . As  $\text{epi}(f_n)$  is closed, there exist some  $\delta_i, \epsilon_i$ ,  $i = 1, 2$  such that

$$d((x_i, y_i), \text{epi}(f_n)) = d((x_i, y_i), (x_i + \delta_i, y_i + \epsilon_i))$$

and  $(x_i + \delta_i, y_i + \epsilon_i) \in \text{epi}(f_n)$ . Thus there exist  $\delta, \epsilon$  satisfying  $\delta = \alpha\delta_1 + (1 - \alpha)\delta_2$ ,  $\epsilon = \alpha\epsilon_1 + (1 - \alpha)\epsilon_2$  such that

$$\alpha(x_1 + \delta_1, y_1 + \epsilon_1) + (1 - \alpha)(x_2 + \delta_2, y_2 + \epsilon_2) = (x_3, y_3) + (\delta, \epsilon) \in \text{epi}(f_n)$$

and so

$$\begin{aligned} d((x_3, y_3), \text{epi}(f_n)) &\leq \max\{|\delta|, |\epsilon|\} \leq \max\{|\delta_1|, |\delta_2|, |\epsilon_1|, |\epsilon_2|\} \\ &= \max\{d((x_1, y_1), \text{epi}(f_n)), d((x_2, y_2), \text{epi}(f_n))\}. \end{aligned}$$

Therefore  $d((x_3, y_3), \text{epi}(f)) = \lim_{n \rightarrow \infty} d((x_3, y_3), \text{epi}(f_n)) = 0$ , and so  $(x_3, y_3) \in \text{epi}(f)$ . □

**Theorem 4.1. LDP for MGF estimators.** First see that  $f \in \mathcal{X}_M$  if and only if it is of Type 1 or 3 as given in the statement of Proposition 4.1, so that  $\mathcal{X}_M = \overline{\mathcal{X}_B} \cap \{f : f(0) = 1\}$ . Proposition 4.5 gives that the measures are exponentially tight, while Propositions 4.2 and 4.4 give coincidence of the upper and lower deviation functions, and therefore by [8][Lemma 1.2.18], we obtain the result for the MGF estimators  $\{M_n\}$ . We can then reduce this to an LDP in  $\mathcal{X}_M$  by [8][Lemma 4.1.5(b)] as the effective domain of  $I_M$  is a subset of  $\mathcal{X}_M$ , which is measurable.  $\square$

**Lemma 4.1. Continuity of  $\mathcal{L}$ .** This proof follows from [11][Proposition 3], although not immediately as the map  $\mathcal{L}$  fails that proposition's criteria. We will by bypass this difficulty by considering a sequence of restrictions of  $\mathcal{L}$  that satisfy the propositions criteria and such that the images of elements of  $\mathcal{X}_M$  under such restrictions are equal to their image under  $\mathcal{L}$  when intersected with an increasingly large neighbourhood of the origin. First, consider the function  $g : \mathbb{R}^2 \mapsto \mathbb{R} \times (0, \infty) : (\theta, \psi) \mapsto (\theta, \exp(\psi))$ . Then  $\text{epi}(\mathcal{L}(f)) = g^*(\text{epi}(f))$ , where  $g^* : \mathcal{P}(\mathbb{R} \times (0, \infty)) \mapsto \mathcal{P}(\mathbb{R}^2)$ , with  $\mathcal{P}$  denoting the power set, and  $g^*(D) = \{(\theta, \psi) : g(\theta, \psi) \in D\}$  is the pull back. Although  $g$  is bijective, continuous and maps bounded sets to bounded sets, its inverse fails to be uniformly continuous on bounded sets, so that the assumptions of [11][Proposition 3] do not hold.

Now consider  $g_n = g|_{\mathbb{R} \times [-n, \infty)}$ , the function  $g$  restricted to the set  $\mathbb{R} \times [-n, \infty)$  with codomain  $\mathbb{R} \times [\exp(-n), \infty)$ , and the corresponding pullback  $g_n^* : \mathcal{P}(\mathbb{R} \times [\exp(-n), \infty)) \mapsto \mathcal{P}(\mathbb{R} \times [-n, \infty))$ . Then  $g_n$  is bijective, continuous and with an inverse uniformly continuous on bounded subsets. Therefore by [11][Proposition 3]  $g_n^*$  is continuous. Now consider any sequence of closed subsets of  $\mathbb{R} \times (0, \infty)$ ,  $\{A_m\}$ , all containing the point  $(0, 1)$  and converging in  $(\mathcal{P}(\mathbb{R} \times (0, \infty)), \tau_{\text{AW}})$  to  $A$ , which, by properties of  $\tau_{\text{AW}}$  convergence, must also contain the point  $(0, 1)$ . Then by [5][Exercise 7.4.1]  $A_m \cap (\mathbb{R} \times [\exp(-n), \infty)) \rightarrow A \cap (\mathbb{R} \times [\exp(-n), \infty))$  in  $(\mathcal{P}(\mathbb{R} \times (0, \infty)), \tau_{\text{AW}})$  as  $m \rightarrow \infty$ . By (7) we can easily show that this implies convergence in  $(\mathcal{P}(\mathbb{R} \times [\exp(-n), \infty)), \tau_{\text{AW}})$ . Therefore  $g_n^*(A_m \cap (\mathbb{R} \times [\exp(-n), \infty))) \rightarrow g_n^*(A \cap (\mathbb{R} \times [\exp(-n), \infty)))$  as  $m \rightarrow \infty$  for any  $n$ . Note that  $g_n^*(A_m \cap (\mathbb{R} \times [\exp(-n), \infty))) = g^*(A_m) \cap (\mathbb{R} \times [-n, \infty))$  when considered as elements of  $\mathcal{P}(\mathbb{R}^2)$ . Therefore (7) holds for  $\{g^*(A_m) \cap (\mathbb{R} \times [-n, \infty))\}_{m=1}^\infty$  for any  $\epsilon > 0$  and any bounded set  $B \subset \mathbb{R} \times [-n, \infty)$ .

Now fix any bounded  $B \subset \mathbb{R}^2$  and any  $\epsilon > 0$ . Then  $B \subset \mathbb{R} \times [-n, \infty)$  for some  $n$ , and for any  $C \subset \mathbb{R}^2$  containing the point  $(0, 0)$ , such as  $g^*(A_m)$  and  $g^*(A)$ ,  $d(x, C) = d(x, C \cap B_{2r})$  for any  $x \in B$  where  $r$  is such that  $B \subset B_r$ , and so for  $n > 2r$  and any  $x \in B$ ,

$$\begin{aligned} d(x, g^*(A_m) \cap (\mathbb{R} \times [-n, \infty))) &= d(x, g^*(A_m) \cap (\mathbb{R} \times [-n, \infty)) \cap B_{2r}) = d(x, g^*(A_m) \cap B_{2r}) \\ &= d(x, g^*(A_m)). \end{aligned}$$

The same is true for  $g^*(A)$  and so (7) also holds for  $\{g^*(A_m)\}$ , any  $\epsilon > 0$  and any  $B \subset \mathbb{R}^2$ , so that  $g^*(A_m) \rightarrow g^*(A)$  in  $(\mathcal{P}(\mathbb{R}^2), \tau_{\text{AW}})$ . This is true for every sequence  $\{A_m\}$  containing the point  $(0, 1)$ , and so  $g^*$  is continuous when its domain is restricted to sets containing the point  $(0, 1)$ , which implies that the functional  $L : \mathcal{X}_M \rightarrow \mathcal{X}_\Lambda$  is continuous.  $\square$

**Theorem 4.2. LDP for CGF estimators.** This result follows via an application of the contraction principle from the LDP for  $\{M_n\}$  in  $\mathcal{X}_M$  proved in Theorem 4.1 through the function defined in Lemma 4.1.  $\square$

**Lemma 4.2. Continuity of  $\mathcal{J}$ .** Although the notation  $\mathcal{D}_f$  has thus far been used for MGFs, here we will use it for elements of  $\mathcal{X}_\Lambda$  to denote their effective domain. Fix some  $f \in \mathcal{X}_\Lambda$  satisfying  $\mathcal{D}_f \neq \{0\}$ ,  $\mathcal{D}_f \subset [0, \infty)$ , and let  $f_n \rightarrow f$  in  $\tau_{\text{AW}}$ . Then  $f_n \rightarrow f$  pointwise on the interior of  $\mathcal{D}_f$  and  $f_n \rightarrow \infty$  pointwise outside  $\overline{\mathcal{D}_f}$ . Therefore  $\mathcal{J}(f_n) \rightarrow \mathcal{J}(f)$  pointwise on the interior of  $\mathcal{D}_f$ ,  $\mathcal{J}(f_n) \rightarrow \infty$  outside  $\mathcal{D}_f \cup [0, \infty)$  and  $\mathcal{J}(f_n) \rightarrow -\infty$  on  $(-\infty, 0)$ . This can be extended to uniform convergence on compact subsets of the interior of  $\mathcal{D}_f$  by the convexity of

$\mathcal{J}(f)$  and  $\mathcal{J}(f_n)$  on  $(0, \infty)$ , and so by a minor modification of [5][Lemma 7.1.2] we have that  $\mathcal{J}(f_n) \rightarrow \mathcal{J}(f)$  in  $(\mathcal{X}_J, \tau_{AW})$ , so that  $\mathcal{J}$  is continuous at  $f$ . Proof is identical if  $\mathcal{D}_f \neq \{0\}$ ,  $\mathcal{D}_f \subset (-\infty, 0]$ , as  $\mathcal{J}(f)$  and  $\mathcal{J}(f_n)$  are concave on  $(-\infty, 0)$ .

Now assume that  $f$  is finite in a neighbourhood of the origin, and let  $f_n \rightarrow f$  in  $\tau_{AW}$ . Let  $f^-$  and  $f^+$  mimic  $f$  with  $\mathcal{D}_{f^-} = \mathcal{D}_f \cap (-\infty, 0]$ ,  $\mathcal{D}_{f^+} = \mathcal{D}_f \cap [0, \infty)$ , and define  $f_n^-, f_n^+$  similarly. It follows from [5][Exercise 7.4.1] that  $f_n^- \rightarrow f^-$ ,  $f_n^+ \rightarrow f^+$  in  $\tau_{AW}$ , and so  $\mathcal{J}(f_n^-) \rightarrow \mathcal{J}(f^-)$ ,  $\mathcal{J}(f_n^+) \rightarrow \mathcal{J}(f^+)$ . Defining  $A_n^+ = \text{epi}(\mathcal{J}(f_n^+))$ ,  $A_n^- = \text{epi}(\mathcal{J}(f_n^-))$ ,  $C_n^+ = A_n^+ \cap \{(x, y) : x \geq 0, \text{ or } x = 0, y \geq \mathcal{J}_n(f)(0)\}$ , and  $A^+$ ,  $A^-$  and  $C^+$  similarly, as  $\{(x, y) : x \geq 0, \text{ or } x = 0, y \geq \mathcal{J}_n(f)(0)\}$  is closed it follows again by [5][Exercise 7.4.1] that  $C_n^+ \rightarrow C^+$  in  $\tau_{AW}$ . As  $\text{epi}(\mathcal{J}(f)) = A^- \cup C^+$  and similarly for  $\text{epi}(f_n)$ , for any fixed closed and bounded set  $B \subset \mathbb{R}^2$ ,  $x \in B$  and  $n \geq 1$ , without loss of generality we can assume  $d(x, A_n^- \cup C_n^+) = d(x, A_n^-)$  and that  $|d(x, \text{epi}(\mathcal{J}(f))) - d(x, \text{epi}(\mathcal{J}(f_n)))| = d(x, \text{epi}(\mathcal{J}(f))) - d(x, \text{epi}(\mathcal{J}(f_n)))$  so that

$$\begin{aligned} |d(x, \text{epi}(\mathcal{J}(f))) - d(x, \text{epi}(\mathcal{J}(f_n)))| &= d(x, \text{epi}(\mathcal{J}(f))) - d(x, \text{epi}(\mathcal{J}(f_n))) \\ &= d(x, A^- \cup C^+) - d(x, A_n^- \cup C_n^+) \\ &= d(x, A^- \cup C^+) - d(x, A_n^-) \\ &\leq d(x, A^-) - d(x, A_n^-) \\ &\leq |d(x, \text{epi}(\mathcal{J}(f^-))) - d(x, \text{epi}(\mathcal{J}(f_n^-)))|. \end{aligned}$$

This is true for all  $n$  and for all  $x \in B$ , so that

$$\begin{aligned} &\sup_{x \in B} |d(x, \text{epi}(\mathcal{J}(f))) - d(x, \text{epi}(\mathcal{J}(f_n)))| \\ &\leq \max\{\sup_{x \in B} |d(x, \text{epi}(\mathcal{J}(f^-))) - d(x, \text{epi}(\mathcal{J}(f_n^-)))|, \sup_{x \in B} |d(x, \text{epi}(\mathcal{J}(f^+))) - d(x, \text{epi}(\mathcal{J}(f_n^+)))|\}, \end{aligned}$$

and so as  $\mathcal{J}(f_n^-) \rightarrow \mathcal{J}(f^-)$  and  $\mathcal{J}(f_n^+) \rightarrow \mathcal{J}(f^+)$ , we can apply (7) to show that  $\mathcal{J}(f_n) \rightarrow \mathcal{J}(f)$  in  $\tau_{AW}$ .

Now assume that  $f \in \mathcal{X}_\Lambda$  satisfies  $\mathcal{D}_f = \{0\}$ . Then for any sequence  $f_n \rightarrow f$ ,  $\mathcal{J}(f_n)(\theta) \rightarrow \infty$  for all  $\theta > 0$  and  $\mathcal{J}(f_n)(\theta) \rightarrow -\infty$  for all  $\theta < 0$ . To prove that  $\mathcal{J}(f_n) \rightarrow \mathcal{J}(f)$  in  $\tau_{AW}$ , we will apply [5][Theorem 3.1.7] that states that for a collection  $\{A_n\}$  and  $A$ , non-empty closed subsets of a metric space  $X$ ,  $A_n \rightarrow A$  in  $\tau_{AW}$  if and only if for every  $k \geq 1$ ,

$$\sup_{x \in A \cap \overline{B}_k(x_0)} d(x, A_n) \rightarrow 0 \text{ and } \sup_{x \in A_n \cap \overline{B}_k(x_0)} d(x, A) \rightarrow 0$$

as  $n \rightarrow \infty$ , where  $x_0$  is any fixed point in  $X$ ,  $\overline{B}_k(x_0)$  is a closed ball of radius  $k$  and centre  $x_0$ , and by convention  $\sup_{x \in \emptyset} d(x, C) = 0$  for any non-empty set  $C$ .

In our case,  $A_n = \text{epi}(\mathcal{J}(f_n))$ ,  $A = \text{epi}(\mathcal{J}(f))$  and  $\overline{B}_k(x_0) = \overline{B}_k$ , the closed ball in  $\mathbb{R}^2$  of radius  $k$  around the origin. First consider  $\sup_{x \in \text{epi}(\mathcal{J}(f)) \cap \overline{B}_k} d(x, \text{epi}(\mathcal{J}(f_n)))$  for some  $k \geq 1$ . As  $\text{epi}(\mathcal{J}(f)) \cap \overline{B}_k$  is closed, for all  $n$  there exists some  $(x_n, y_n) \in \overline{B}_k$  with  $x_n \leq 0$  such that  $d((x_n, y_n), \text{epi}(\mathcal{J}(f_n))) = \sup_{x \in \text{epi}(\mathcal{J}(f)) \cap \overline{B}_k} d(x, \text{epi}(\mathcal{J}(f_n)))$ . As  $x_n \in [-k, 0]$  for all  $n$ , a subsequence of  $x_n$  converges, say to some  $x \in [-k, 0]$ . First assume  $x < 0$  and fix some  $\epsilon > 0$ . Then along this subsequence, when  $n$  is large enough so that  $|x_n - x| < \epsilon$  and  $\mathcal{J}(f_n)(x) < -k$ , then as  $y_n \geq -k$ ,  $(x, y_n) \in \text{epi}(\mathcal{J}(f_n))$  and so  $d((x_n, y_n), \text{epi}(\mathcal{J}(f_n))) \leq d((x_n, y_n), (x, y_n)) < \epsilon$ . This is true for all  $\epsilon > 0$  and so along any subsequence of  $\{f_n\}$  such that  $\{x_n\}$  converges to some  $x < 0$ ,  $\sup_{x \in \text{epi}(\mathcal{J}(f)) \cap \overline{B}_k} d(x, \text{epi}(\mathcal{J}(f_n))) \rightarrow 0$ . Now assume a subsequence of  $\{x_n\}$  converges to 0. Then for any  $\epsilon > 0$  and  $n$  large enough so that  $x_n < -\epsilon$  and  $\mathcal{J}(f_n)(-\epsilon/2) < -k$ , we again have that  $d((x_n, y_n), \text{epi}(\mathcal{J}(f_n))) < \epsilon$ . Therefore  $\sup_{x \in \text{epi}(\mathcal{J}(f)) \cap \overline{B}_k} d(x, \text{epi}(\mathcal{J}(f_n))) \rightarrow 0$  along any convergent subsequence of  $\{x_n\}$ . If  $\sup_{x \in \text{epi}(\mathcal{J}(f)) \cap \overline{B}_k} d(x, \text{epi}(\mathcal{J}(f_n))) \not\rightarrow 0$ , that is  $\sup_{x \in \text{epi}(\mathcal{J}(f)) \cap \overline{B}_k} d(x, \text{epi}(\mathcal{J}(f_n))) > \epsilon$  infinitely often for any  $\epsilon > 0$ , then along this subsequence

we can find a sub-subsequence such that  $\{x_n\}$  converges, which would yield a contradiction. Therefore  $\sup_{x \in \text{epi}(\mathcal{J}(f)) \cap \overline{B}_k} d(x, \text{epi}(\mathcal{J}(f_n))) \rightarrow 0$  along the entire sequence.

Now consider  $\sup_{x \in \text{epi}(\mathcal{J}(f_n)) \cap \overline{B}_k} d(x, \text{epi}(\mathcal{J}(f)))$  for some  $k \geq 1$ . Assume that  $\text{epi}(\mathcal{J}(f_n)) \cap \overline{B}_k \cap ([0, \infty) \times \mathbb{R}) \neq \emptyset$  for any  $n$ , as if this is the case then  $\sup_{x \in \text{epi}(\mathcal{J}(f_n)) \cap \overline{B}_k} d(x, \text{epi}(\mathcal{J}(f))) = 0$ . If this is not true infinitely often then we still need to show convergence along that subsequence, and if it is not true only finitely often then convergence holds immediately. With this assumption, notice that for all  $n$  there exists some  $(x_n, y_n) \in \overline{B}_k$  with  $x_n \geq 0$  such that  $\sup_{x \in \text{epi}(\mathcal{J}(f_n)) \cap \overline{B}_k} d(x, \text{epi}(\mathcal{J}(f))) = d((x_n, y_n), \text{epi}(\mathcal{J}(f))) = x_n$ . It is easy to see that  $x_n = \sup\{\theta \in [0, k] : f_n(\theta) \leq k\}$ , and by lower semi-continuity of  $\mathcal{J}(f_n)$  on  $(0, \infty)$ ,  $\mathcal{J}(f_n)(x_n) \leq k$ . Also,  $\mathcal{J}(f_n)(x_n) \leq k \Rightarrow f_n(x_n) \leq kx_n \leq k^2$ , so that  $\sup_{x \in \text{epi}(f_n) \cap \overline{B}_{k^2}} d(x, \text{epi}(f)) \geq x_n$ . But as  $f_n \rightarrow f$ ,  $\sup_{x \in \text{epi}(f_n) \cap \overline{B}_{k^2}} d(x, \text{epi}(f)) \rightarrow 0$  so that  $\sup_{x \in \text{epi}(\mathcal{J}(f_n)) \cap \overline{B}_k} d(x, \text{epi}(\mathcal{J}(f))) = x_n \rightarrow 0$ .  $\square$

**Theorem 4.3. LDP for Jarzynski estimators.** This result follows from the continuity of  $\mathcal{J}$  and the contraction principle.  $\square$

**Theorem 4.4. LDP for rate function estimators.** This result follows from the continuity of  $\mathcal{LF}$  and the contraction principle.  $\square$

## Section 5

**Proposition 5.1. Convexity of  $I_M$ .** We establish global convexity of  $I_M$  by creating an argument along the lines of [8][Lemma 4.1.21], but the conditions of that Lemma as stated do not hold here as  $\mathcal{X}_M$  is not a topological vector space. The proof of [8][Lemma 4.1.21], however, hinges only on two properties that we instead establish directly.

First, we need to show continuity of averaging in  $\mathcal{X}_M$ . That is, if  $\{f_n\}, \{g_n\} \subset \mathcal{X}_M$  are such that  $f_n \rightarrow f \in \mathcal{X}_M$  and  $g_n \rightarrow g \in \mathcal{X}_M$  in  $\tau_{\text{AW}}$ , then  $(f_n + g_n)/2 \rightarrow (f + g)/2$  in  $\tau_{\text{AW}}$ . If  $f$  or  $g$  is finite and continuous at one point in the domain of the other, then we can apply [5][Theorem 7.4.5] to show that  $\lim_{n \rightarrow \infty} (f_n + g_n) = f + g$  in  $\mathcal{X}'_M$ . To show continuity of multiplication by  $1/2$ , see that  $\{(f_n + g_n)/2\}$  is a sequence in  $\{h : h(0) \leq 1\}$  which is compact by Proposition 4.5, so it has a convergent subsequence. Replacing  $\{(f_n + g_n)/2\}$  with this subsequence with limit  $h$  and applying [5][Theorem 7.4.5] it follows that

$$f + g = \lim_{n \rightarrow \infty} [(f_n + g_n)/2 + (f_n + g_n)/2] = 2h$$

so that  $h = (f + g)/2$ . This is true for every subsequence of  $\{(f_n + g_n)/2\}$ , which lies in a compact set, and so it must follow that  $(f_n + g_n)/2 \rightarrow h = (f + g)/2$  and so multiplication by  $1/2$  is a continuous operation in  $\mathcal{X}'_M$ . It then follows that averaging is continuous in  $\mathcal{X}'_M$  and therefore in  $\mathcal{X}_M$ . If it not the case that  $f$  or  $g$  is finite and continuous at one point in the domain of the other, then it must follow that  $f(\theta) = \infty$  for  $\theta > 0$  and  $g(\theta) = \infty$  for  $\theta < 0$ , or vice versa. In this case  $(f + g)/2$  is finite only at 0 and  $(f_n(\theta) + g_n(\theta))/2 \rightarrow \infty$  point-wise for each  $\theta \neq 0$ , so  $h = \lim_{n \rightarrow \infty} (f_n + g_n)/2$  satisfies  $h(\theta) = \infty$  for  $\theta \neq 0$ , and so  $h = (f + g)/2$  if  $h(0) = 1$ . Consider  $(0, y)$  for  $y < 1$ . As  $f(0) = 1$ ,  $d((0, y), \text{epi}(f_n)) \rightarrow d((0, y), \text{epi}(f)) > 0$ , and similarly for  $d((0, y), \text{epi}(g_n))$ . Furthermore,

$$\begin{aligned} & (f_n(\theta) + g_n(\theta))/2 \geq \min\{f_n(\theta), g_n(\theta)\} \\ & \Rightarrow \text{epi}((f_n + g_n)/2) \subset \text{epi}(f_n) \cup \text{epi}(g_n) \\ & \Rightarrow d((0, y), \text{epi}((f_n + g_n)/2)) \geq \min\{d((0, y), \text{epi}(f_n)), d((0, y), \text{epi}(g_n))\} \\ & \Rightarrow d((0, y), \text{epi}((f_n + g_n)/2)) \not\rightarrow 0 \\ & \Rightarrow d((0, y), \text{epi}(h)) > 0. \end{aligned}$$

Therefore  $h(0) > y$  for any  $y < 1$ , so  $h(0) = 1$  as required.

Second, we need to show the continuity of  $\beta f + (1 - \beta)g$  with respect to  $\beta \in (0, 1)$ . First notice that  $\mathcal{D}_{\beta f + (1 - \beta)g} = \mathcal{D}_f \cap \mathcal{D}_g$  for all  $\beta \in (0, 1)$ . Moreover we have pointwise convergence of  $\beta f + (1 - \beta)g$  on the interior of  $\mathcal{D}_f \cap \mathcal{D}_g$  as  $\beta \rightarrow \beta_0 \in (0, 1)$ . This can be extended to uniform convergence on bounded subsets of the interior of  $\mathcal{D}_f \cap \mathcal{D}_g$  by the convexity of  $\beta f + (1 - \beta)g$ . Convergence in  $\tau_{\text{AW}}$  follows from [5][Lemma 7.1.2], with simple modifications to handle the possibility of open domains and common domains not necessarily equal to  $\mathbb{R}$ . Now see that for any sets  $A_1$  and  $A_2$ , letting  $(A_1 + A_2)/2 = \{g : g = (f_1 + f_2)/2, f_1 \in A_1, f_2 \in A_2\}$  and following the notation of Proposition 4.4,

$$\begin{aligned} P(M_n \in A_1, M_{n+1,2n} \in A_2) &\leq P\left[M_{2n} \in \frac{A_1 + A_2}{2}\right] \\ &\Rightarrow P(M_n \in A_1)P(M_n \in A_2) \leq P\left(M_{2n} \in \frac{A_1 + A_2}{2}\right) \\ &\Rightarrow \frac{1}{2n} \log P(M_n \in A_1) + \frac{1}{2n} \log P(M_n \in A_2) \leq \frac{1}{2n} \log P\left(M_{2n} \in \frac{A_1 + A_2}{2}\right) \\ &\Rightarrow \frac{1}{2} \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(M_n \in A_1) + \frac{1}{2} \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(M_n \in A_2) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log P\left(M_n \in \frac{A_1 + A_2}{2}\right). \end{aligned}$$

As  $(f_1, f_2) \mapsto (f_1 + f_2)/2$  is a continuous operation, as is  $\beta \mapsto \beta f + (1 - \beta)g$  for any  $f, g \in \mathcal{X}_M$  and  $\beta \in (0, 1)$ , the remainder of the proof follows identically to that of [8][Lemma 4.1.21].  $\square$

**Proposition 5.2.  $I_M(f)$  for  $f$  an MGF.** First it is worth noting that  $\nu$  is uniquely defined if  $f$  is finite on an open interval, which need not include 0 [38]. As in Proposition 4.2, if we define  $\mathcal{X}_{\text{INT}}$  to be the space of all moment generating functions finite on some open interval, equipped with the subspace topology, then the mapping from MGFs to measures  $\Phi : \mathcal{X}_{\text{INT}} \rightarrow \mathcal{M}(\mathbb{R})$  is well defined and continuous when  $\mathcal{X}_{\text{INT}}$  is equipped with the subspace topology, as convergence of MGFs that are finite on some open interval to another MGF that is finite on some open interval implies convergence of their distributions, as seen in [38] and in the proof of Proposition 4.1. So we have that for every open  $G \subset \mathcal{M}(\mathbb{R})$ ,  $\Phi^{-1}(G) = G' \cap \mathcal{X}_{\text{INT}}$  for some open  $G' \subset \mathcal{X}_M$ . As  $\Phi^{-1}(\nu) = f_\nu$  for  $f_\nu \in \mathcal{X}_{\text{INT}}$ ,  $G \ni \nu \Rightarrow G' \ni f_\nu$  and so if  $L_n$  is the empirical distribution governed by the same sample as for  $M_n$ ,

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(L_n \in G) &= \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(M_n \in G' \cap \mathcal{X}_{\text{INT}}) \\ &= \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(M_n \in G') = \underline{m}(G') \geq -I_M(f_\nu) \end{aligned}$$

This implies that

$$\inf_{G \ni \nu} \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(L_n \in G) \geq -I_M(f_\nu) \Rightarrow -H(\nu|\mu) \geq -I_M(f_\nu).$$

To prove the second inequality, let  $\mathcal{Y}_N = \{\nu \in \mathcal{M}(\mathbb{R}) : \nu \text{ is supported on } [-N, N]\}$ , and let  $\Gamma$  be the mapping from measures to MGFs. Since convergence of measures in  $\mathcal{Y}_N$  implies convergence of the corresponding moment generating functions, we have that  $\Gamma|_{\mathcal{Y}_N}$  is continuous and so  $G_N \cap \mathcal{Y}_N = \Gamma^{-1}(G' \cap \Gamma(\mathcal{Y}_N))$  is open for every open set  $G' \subset \mathcal{X}_M$ . Note that we do not know if the same  $G \subset \mathcal{M}(\mathbb{R})$  forms the inverse image of  $G' \cap \Gamma(\mathcal{Y}_N)$  for every  $N$ , hence the use of the

subscript, yet we can demand that  $G_N \subset G_{N+1}$ . Note that

$$\begin{aligned} \Gamma^{-1}(G' \cap \Gamma(\mathcal{Y}_N)) &= \Gamma^{-1}((G' \cap \Gamma(\mathcal{Y}_N)) \cap (G' \cap \Gamma(\mathcal{Y}_{N+1}))) \\ &= \Gamma^{-1}(G' \cap \Gamma(\mathcal{Y}_N)) \cap \Gamma^{-1}(G' \cap \Gamma(\mathcal{Y}_{N+1})) \\ &= (G_N \cap \mathcal{Y}_N) \cap (G_{N+1} \cap \mathcal{Y}_{N+1}) \\ &= (G_N \cap G_{N+1}) \cap \mathcal{Y}_N. \end{aligned}$$

So if  $G_N \not\subset G_{N+1}$ , we can replace  $G_N$  with  $G_N \cap G_{N+1}$ . So fixing  $f_\nu \in \mathcal{X}_B$ ,  $f_\nu \in \Gamma(\mathcal{Y}_M)$  for some  $M$ , and so we have that for every  $G' \ni f_\nu$ ,  $\nu \in G_M$  and so

$$\begin{aligned} P(M_n \in G') &= P(M_n \in G' \cap \mathcal{X}_B) = P(M_n \in \cup_{N=M}^{\infty} (G' \cap \Gamma(\mathcal{Y}_N))) = P(L_n \in \cup_{N=M}^{\infty} (G_N \cap \mathcal{Y}_N)) \\ &\geq P(L_n \in \cup_{N=M}^{\infty} (G_M \cap \mathcal{Y}_N)) = P(L_n \in G_M \cap \Gamma^{-1}(\mathcal{X}_B)) = P(L_n \in G_M). \end{aligned}$$

This implies that

$$\underline{m}(G') \geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(L_n \in G_M) \geq \inf_{G' \ni f_\nu} \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(L_n \in G) = -H(\nu|\mu).$$

This is true for every  $G' \ni f_\nu$ , and so

$$\inf_{G' \ni f_\nu} \underline{m}(G) \geq -H(\nu|\mu) \Rightarrow -I_M(f_\nu) \geq -H(\nu|\mu).$$

To prove the third inequality, extending the second inequality to all moment generating functions, fix any moment generating function  $f_\nu \in \mathcal{X}_M$ . Let  $\nu_n(dx) = \nu(dx)/\nu([n, n])$  on  $[-n, n]$ . It can easily be shown that  $H(\nu_n|\mu) \rightarrow H(\nu|\mu)$  as  $n \rightarrow \infty$ . As discussed before,  $f_{\nu_n} \rightarrow f_\nu$  in  $\tau_{AW}$ ,  $I_M(f_{\nu_n}) = H(\nu_n|\mu)$  by the previous two lemmas and so by lower semi-continuity of  $I_M$ ,

$$I_M(f_\nu) \leq \lim_{n \rightarrow \infty} I_M(f_{\nu_n}) = \lim_{n \rightarrow \infty} H(\nu_n|\mu) = H(\nu|\mu).$$

In the case of the moment generating function  $f$  finite only at 0, this statement is true for all  $\nu$  such that  $f = f_\nu$ , and so  $I_M(f) \leq \inf_{\nu: f=f_\nu} H(\nu|\mu)$ .  $\square$

**Lemma 5.1. Characterizations for  $f$  not a MGF.** Point 1. Assume  $I_M(f) < \infty$ . Then for any  $g \in \mathcal{X}_M$  and any  $\gamma \in (0, 1)$ ,  $\gamma f + (1 - \gamma)g = f$ . By the convexity of  $I_M$  proved in Proposition 5.1,  $I_M(f) = I_M(\gamma f + (1 - \gamma)g) \leq \gamma I_M(f) + (1 - \gamma)I_M(g)$ , so that  $I_M(f) \leq I_M(g)$ . This is true for all  $g \in \mathcal{X}_M$ , and so  $I_M(f) = 0$ .

Point 2. See first that  $\gamma g + (1 - \gamma)f_\nu \rightarrow f$  in  $\tau_{AW}$  as  $\gamma \rightarrow 0$ , as they converge uniformly on closed subsets of  $(\alpha, \beta)$  and are infinite outside  $[\alpha, \beta]$ , and so  $\tau_{AW}$  convergence can be proven by a minor modification of [5][Proposition 7.1.3]. Thus

$$I_M(f) \leq \lim_{\gamma \rightarrow 0} I_M(\gamma g + (1 - \gamma)f_\nu) \leq \lim_{\gamma \rightarrow 0} (\gamma I_M(g) + (1 - \gamma)I_M(f_\nu)) = I_M(f_\nu).$$

Point 3 First we show that  $I_M(f) \geq I_M(f_\nu)$ , in a manner very similar to how we showed that  $I_M(g) = \infty$  for  $g(0) < 1$  in Proposition 4.2, except in this case it happens that  $C$  is non-empty. Like before, let  $G_m$  be a descending countable open base of  $f$ , and let  $A_m$  be defined by  $\Phi(G_m \cap \mathcal{X}_B) = A_m$ . See that

$$\begin{aligned} I_M(f) &= - \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P(M_n \in G_m) = - \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P(L_n \in A_m) \\ &\geq \lim_{m \rightarrow \infty} \inf_{\kappa \in A_m} H(\kappa|\mu) = \inf_{\kappa \in C} H(\kappa|\mu), \end{aligned}$$

where  $C = \bigcap_{m=1}^{\infty} \overline{A_m}$ . It can be shown that  $C$  is non-empty, however here it is unnecessary to do so, as if  $C$  is empty then  $I_M(f) = \infty$  and so  $I_M(f) \geq I_M(f_\nu)$  trivially. Let  $\kappa \in C$ . As  $\mathcal{M}(\mathbb{R})$  is metrizable, let  $U_m$  be a descending countable base for  $\kappa$ , and let  $\{\kappa_{m,n}\}_{n=1}^{\infty} \subset A_m$  be a sequence satisfying  $\kappa_{m,n} \rightarrow \kappa$  as  $n \rightarrow \infty$ . Let  $N_m$  be such that for  $n \geq N_m$ ,  $\kappa_{m,n} \in U_m$ . Then if  $\kappa_m^* = \kappa_{m,N_m}$ ,  $\kappa_m^* \in A_m$  and  $\kappa_m^* \in U_m$ , so that  $\kappa_m^* \rightarrow \kappa$ . For each  $\kappa_m^*$  there is a corresponding  $f_m^* \in G_m$ , and so  $f_m^* \rightarrow f$ . If  $f$  mimics  $f_\nu$ , then we have a sequence of moment generating functions converging to a function mimicking a moment generating function  $f_\nu$ , and a corresponding convergent sequence of measures, then it follows that  $\kappa = \nu$  from the proof of Proposition 4.1, and from [38][Theorem 2]. So  $C = \{\nu\}$  and thus  $I_M(f) \geq H(\nu|\mu) = I_M(f_\nu)$ . Now assume that  $I_M(f) < \infty$ , then we can apply statement 2 of Lemma 5.1 with  $f = g$  and so  $I_M(f) = I_M(f_\nu)$ .  $\square$

**Lemma 5.2. Five equivalences.** (1)  $\Rightarrow$  (2). Assume (1), so that the sequence  $1/n \sum_{i=1}^n e^{\gamma X_i}$  satisfies the conditions of Cramér's Theorem. Assume without loss of generality that  $\gamma < \alpha$ , and so  $\alpha \leq 0$  is finite. Fix  $f \in \mathcal{D}_{[\alpha,\beta]}^{\subset}$ . Then for  $k$  large enough so that  $\alpha - 1/k > -k$  and  $\gamma < \alpha - 1/k$ ,

$$M_n \in V_k(f) \Rightarrow M_n(\alpha - 1/k) > k \Rightarrow M_n(\gamma) > k$$

as  $d((\alpha - 1/k, k), \text{epi}(f)) \geq 1/k$ , so we can't have  $(\alpha - 1/k, k) \in \text{epi}(M_n)$ . So

$$\overline{m}(V_k(f)) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log P(M_n(\gamma) > k) \leq - \inf_{x \geq k} I_\gamma^*(x)$$

where  $I_\gamma^*(x)$ , the Cramér's Theorem rate function for  $1/n \sum_{i=1}^n e^{\gamma X_i}$ , has compact level sets. Then

$$\begin{aligned} -I_M(f) &= \lim_{k \rightarrow \infty} \overline{m}(V_k(f)) \leq \lim_{k \rightarrow \infty} - \inf_{x \geq k} I_\gamma^*(x) = -\infty \\ &\Rightarrow I_M(f) = \infty. \end{aligned}$$

This is true for all  $f \in \mathcal{D}_{[\alpha,\beta]}^{\subset}$ , so (1)  $\Rightarrow$  (2).

(2)  $\Rightarrow$  (3) follows immediately from Proposition 5.2.

(3)  $\Rightarrow$  (4) is trivial as  $\mathcal{D}_{[\alpha,\beta]} \subset \mathcal{D}_{[\alpha,\beta]}^{\subset}$ .

(4)  $\Rightarrow$  (1) requires the lengthiest of proof. We first establish the result when  $\mu$  is a discrete distribution. Consider all  $\nu$  such that  $f_\nu \in \mathcal{D}_{[\alpha,\beta]}$ . Then assuming (4),

$$\sum_{k \geq 0} \nu(k) \log \left( \frac{\nu(k)}{\mu(k)} \right) = \infty \text{ for all such } \nu \text{ or } \sum_{k < 0} \nu(k) \log \left( \frac{\nu(k)}{\mu(k)} \right) = \infty \text{ for all such } \nu,$$

as otherwise, if  $\nu$  and  $\kappa$  were distributions with  $f_\nu, f_\kappa \in \mathcal{D}_{[\alpha,\beta]}$  that give a finite sum in the first and second case respectively, then a distribution  $\pi$  satisfying  $\pi(k) \propto \nu(k)$  for  $k \geq 0$  and  $\pi(k) \propto \kappa(k)$  for  $k < 0$  would still have moment generating function in  $\mathcal{D}_{[\alpha,\beta]}$  and would have  $H(\pi|\mu) < \infty$ . Without loss of generality assume that

$$\sum_{k \geq 0} \nu(k) \log \left( \frac{\nu(k)}{\mu(k)} \right) = \infty \text{ for all such } \nu. \quad (23)$$

If  $\beta = \infty$ , then for any  $\nu$  with  $f_\nu \in \mathcal{D}_{[\alpha,\beta]}$ ,  $\kappa$  satisfying  $\kappa(k) = \nu(k)/\nu(-\infty, 0]$  for  $k \leq 0$  and  $\kappa(k) = 0$  otherwise, also has  $f_\kappa \in \mathcal{D}_{[\alpha,\beta]}$ , but (23) does not hold. So it follows that  $\beta < \infty$ . Also

$f_\mu(\epsilon) < \infty$  for some  $\epsilon > 0$ . To see this, assume the contrary, that  $f_\mu \in \mathcal{D}_{[\alpha^*, 0]}$  for some  $\alpha^* \leq 0$ . Then  $\nu(k) \propto e^{-\beta k} \mu(k)$  for  $k \geq 0$  has  $f_\nu \in \mathcal{D}_{[\alpha, \beta]}$  for some suitable behaviour of the left tail. But

$$\sum_{k \geq 0} \nu(k) \log \left( \frac{\nu(k)}{\mu(k)} \right) = \sum_{k \geq 0} e^{-\beta k} (-\beta k) \mu(k) < \infty$$

as  $e^{-\beta k} (-\beta k) \leq 0$  for all  $k$ . As this is impossible, it must follow that  $f_\mu(\epsilon) < \infty$  for some  $\epsilon > 0$ .

We can assume  $X_1$  is not bounded above, as otherwise (1) holds trivially for any  $\gamma > \beta$ . Since the support of  $\mu$  has no upper bound we can find a  $\nu$  supported on a subset of the support of  $\mu$  that satisfies  $\lim_{k \rightarrow \infty} \nu(k) e^{\gamma k} = \infty$  for all  $\gamma > 0$ , where the limit is taken along the support of  $\nu$ . To see this, let  $\{b_n\}$  contained in the support  $\mu$  satisfy  $b_n > n$  for all  $n$ , and let  $\{b_n\}$  be the support of  $\nu$  with  $\nu(b_n) \propto n^{-2}$ . We can write  $\nu$  as  $\nu(k) = C_\nu g(k) \mu(k)$  for some non-negative function  $g$  defined on the support of  $\mu$ .  $C_\nu$  is a constant chosen so that  $\nu$  is a distribution. Without loss of generality we can set  $g(k) \geq 3$  for all  $k \geq 0$ , as  $\sum_{k \geq 0} (g(k) + 3) \mu(k) < \infty$  if and only if  $\sum_{k \geq 0} g(k) \mu(k) < \infty$ . As stated before  $\sum_{k \geq 0} e^{\epsilon k} \mu(k) < \infty$  for small enough positive  $\epsilon$  and so

$$\lim_{k \rightarrow \infty} e^{\epsilon k} (g(k) + 3) \mu(k) = \infty \Leftrightarrow \lim_{k \rightarrow \infty} e^{\epsilon k} g(k) \mu(k) = \infty,$$

if the limit as  $k \rightarrow \infty$  is taken along the support of  $\nu$ . So  $\nu$  still satisfies  $\lim_{k \rightarrow \infty} \nu(k) e^{\gamma k} = \infty$  for all  $\gamma > 0$ . Define  $\kappa(k) = D_\nu \nu(k) / \log g(k)$  for all  $k \geq 0$ , where  $D_\nu$  is chosen so that  $\kappa$  is a distribution. As  $\log g(k) > 1$ ,  $\kappa(k)$  is summable and so  $D_\nu$  exists. The behaviour of  $\kappa(k)$  for  $k < 0$  is unimportant. It follows that

$$\begin{aligned} \sum_{k \geq 0} \kappa(k) \log \left( \frac{\kappa(k)}{\mu(k)} \right) &= \sum_{k \geq 0} D_\nu \frac{\nu(k)}{\log g(k)} \log \left( D_\nu C_\nu \frac{g(k)}{\log g(k)} \right) \\ &\leq \sum_{k \geq 0} D_\nu \frac{\nu(k)}{\log g(k)} \log (D_\nu C_\nu) + \sum_{k \geq 0} D_\nu \frac{\nu(k)}{\log g(k)} \log (g(k)) \\ &\leq \sum_{k \geq 0} D_\nu \nu(k) \log (D_\nu C_\nu) + \sum_{k \geq 0} D_\nu \nu(k) \\ &< \infty. \end{aligned}$$

So  $H(\kappa|\mu) < \infty$  and so  $f_\kappa(\gamma) < \infty$  for some  $\gamma \notin [\alpha, \beta]$ . Since this is true no matter what the behaviour of  $\kappa$  along its left tail it must follow that  $f_\kappa(\gamma) < \infty$  for some  $\gamma > \beta$ , and so

$$\sum_{k \geq 0} e^{(2\theta+\beta)k} \frac{\nu(k)}{\log g(k)} < \infty \Rightarrow e^{(2\theta+\beta)k} \frac{\nu(k)}{\log g(k)} \rightarrow 0$$

as  $k \rightarrow \infty$  for some  $\theta > 0$ . But  $e^{\theta k} \nu(k) \rightarrow \infty$ , so it follows that

$$\frac{e^{(\theta+\beta)k}}{\log g(k)} \rightarrow 0.$$

Thus there exists  $K \geq 0$  such that for all  $k > K$ ,  $\log g(k) > e^{(\theta+\beta)k}$ . Then for any  $\lambda \geq 0$

$$\begin{aligned} E(e^{\lambda e^{(\theta+\beta)X_1}}) &\leq e^\lambda + \sum_{k \geq 0} e^{\lambda e^{(\theta+\beta)k}} \mu(k) \\ &= e^\lambda + \frac{1}{C_\nu} \sum_{k \geq 0} e^{\lambda e^{(\theta+\beta)k}} \frac{\nu(k)}{g(k)} \\ &\leq e^\lambda + \frac{1}{C_\nu} \left( \sum_{k=0}^K e^{\lambda e^{(\theta+\beta)k}} \frac{\nu(k)}{g(k)} + \sum_{k > K} e^{\lambda e^{(\theta+\beta)k}} e^{-e^{(\theta+\beta)k}} \nu(k) \right). \end{aligned}$$

For  $0 \leq \lambda < 1$  the above series converges. As  $E(e^{\lambda e^{(\theta+\beta)X_\mu}}) \leq 1$  for  $\lambda < 0$ ,  $e^{(\beta+\theta)X_1}$  has a moment generating function finite in a neighbourhood of the origin. As finiteness of the moment generating function in a neighbourhood of the origin is a sufficient condition for the random walk associated with a distribution to satisfy the conditions of Cramér's Theorem, we have that (4)  $\Rightarrow$  (1).

In order to extend the result to an arbitrary distribution, we need the following notation. For any distribution  $\nu$ , define  $\nu^*(k) = \nu([k+1])$  as the discretisation of  $\nu$ . Notice that if  $X$  is a random variable with distribution  $\nu$  and  $X^* = \lfloor X \rfloor$  where  $\lfloor \cdot \rfloor$  is the floor function, then  $X^*$  has distribution  $\nu^*$  and  $X-1 \leq X^* \leq X$ , which implies that  $\mathcal{D}_{f_\nu} = \mathcal{D}_{f_{\nu^*}}$ .

Now assume (4) for some arbitrary distribution  $\mu$ , i.e. that  $H(\nu|\mu) = \infty$  for all  $\nu$  with  $f_\nu \in \mathcal{D}_{[\alpha,\beta]}$ . Consider  $\mu^*$ , and any  $\kappa$  supported on a subset of the support of  $\mu^*$  with  $f_\kappa \in \mathcal{D}_{[\alpha,\beta]}$ . Then  $\kappa(k) = g(k)\mu^*(k)$  for some non-negative function  $g$  defined on  $\mathbb{Z}$ . Define  $\nu(dx) = g(\lfloor x \rfloor)\mu(dx)$ , then  $\nu^* = \kappa$ , and the effective domain of the moment generating function of  $\nu$  is the same as that of  $\nu^* = \kappa$ , so  $H(\nu|\mu) = \infty$ . Then

$$\begin{aligned} \infty &= \int_{-\infty}^{\infty} \nu(dx) \log \frac{\nu(dx)}{\mu(dx)} = \int_{-\infty}^{\infty} g(\lfloor x \rfloor)\mu(dx) \log g(\lfloor x \rfloor) \\ &= \sum_{k=-\infty}^{\infty} g(k)\mu(k) \log g(k) = \sum_{k=-\infty}^{\infty} \kappa(k) \log \left( \frac{\kappa(k)}{\mu^*(k)} \right). \end{aligned}$$

So  $H(\kappa|\mu^*) = \infty$ . This is true for any  $\kappa$  supported on a subset of the support of  $\mu^*$  with  $f_\kappa \in \mathcal{D}_{[\alpha,\beta]}$ . As any  $\kappa$  that is not supported on a subset of the support of  $\mu^*$  also satisfies  $H(\kappa|\mu^*) = \infty$ , Lemma 5.2 (4) holds for  $\mu^*$  and so as we already established (4)  $\Rightarrow$  (1) for discrete distributions, if  $X_{\mu^*}$  is a random variable with distribution  $\mu^*$  then the moment generating function of  $e^{\gamma X_{\mu^*}}$  is finite in a neighbourhood of the origin for some  $\gamma \notin [\alpha, \beta]$ . By arguments similar to those showing domain equivalence of MGFs of discretised and non-discretised distributions, it follows that the moment generating function of  $e^{\gamma X_1}$  is also finite in a neighbourhood of the origin for the same  $\gamma \notin [\alpha, \beta]$ . As finiteness of the moment generating function in a neighbourhood of the origin is a sufficient condition for the random walk associated with a distribution to satisfy the conditions of Cramér's Theorem, we have that (4)  $\Rightarrow$  (1).

The proof is now almost complete. (1)  $\Rightarrow$  (2)  $\Rightarrow$  (3)  $\Rightarrow$  (4)  $\Rightarrow$  (1). That (5)  $\Rightarrow$  (4) is readily seen from Proposition 5.2. But then (4)  $\Rightarrow$  (2) by the equivalence of (1)-(4), and (2)  $\Rightarrow$  (5) as  $\mathcal{D}_{[\alpha,\beta]} \subset \mathcal{D}_{[\alpha,\beta]}^c$ , so that (5)  $\Leftrightarrow$  (4) and we have equivalence of all 5 statements.  $\square$

### Corollary 5.1.

1. Let  $\beta_0$  be the supremum over all values of  $\gamma$  such that the random walk associated with  $e^{\gamma X_1}$  satisfies the conditions of Cramér's Theorem, and let  $\alpha_0$  be the infimum. Note that  $\beta_0 \geq 0$ ,  $\alpha_0 \leq 0$ . Let  $\beta \geq \beta_0$ ,  $\alpha \leq \alpha_0$ , and let  $f_\nu$  satisfy  $I_M(f_\nu) < \infty$  and  $[\alpha, \beta] \subset \mathcal{D}_{f_\nu}$ . Then for any  $f \in \mathcal{D}_{[\alpha,\beta]}$  mimicking  $f_\nu$ ,  $I_M(f) = I_M(f_\nu)$  as Lemma 5.2 (1) does not hold for  $\alpha, \beta$ , so that Lemma 5.2 (5) does not hold, and so we can apply statements 2 and 3 of Lemma 5.1. If  $\alpha > \alpha_0$  or  $\beta < \beta_0$ , Lemma 5.2 (1) holds, so that Lemma 5.2 (5) holds and so any  $f \in \mathcal{D}_{[\alpha,\beta]}$  mimicking  $f_\nu$  satisfies  $I_M(f) = \infty \neq I_M(f_\nu)$ .
2. Let  $f$  be such that  $I_M(f) < \infty$ ,  $f \in \mathcal{D}_{[\alpha,\beta]}$ . Then Lemma 5.2 (5) does not hold, so that Lemma 5.2 (1) does not hold, so that  $\overline{\mathcal{D}}_f = [\alpha, \beta] \supset [\alpha_0, \beta_0]$ .
3. If  $\alpha_0 = \beta_0 = 0$ ,  $f_n \in \mathcal{D}_{[-\frac{1}{n}, \frac{1}{n}]}$  mimic  $f_\mu$  and  $f$  is the MGF finite only at 0, then  $f_n \rightarrow f$  and so  $I_M(f) \leq \lim_{n \rightarrow \infty} I_M(f_n) = 0$ . If  $[\alpha_0, \beta_0] \neq \{0\}$  then letting  $\alpha, \beta = 0$  Lemma 5.2 (1) holds for some  $\gamma \neq 0$  so that Lemma 5.2 (5) holds and so  $I_M(f) = \infty$ .

□

**Proposition 5.3.**  $I_M(f)$  for  $f$  mimics.

- (a) Let  $[\alpha, \beta] = [0, 0]$  in the statement of Lemma 5.2. Then we have that for  $f \in \mathcal{X}_M$  finite only at 0

$$I_M(f) = \infty \Leftrightarrow \inf_{\nu: f_\nu=f} H(\nu|\mu) = \infty$$

using (4)  $\Leftrightarrow$  (5). We have already shown that  $I_M(f) \in \{0, \infty\}$  by Lemma 5.1; we will show the same for the other term. See that for any  $\nu$  with  $f_\nu = f$ ,  $\kappa = a\nu + (1-a)\mu$  satisfies  $f_\kappa = f$  for any  $a \in (0, 1]$ , so that if  $H(\nu|\mu) < \infty$  for some such  $\nu$ ,

$$\inf_{\nu: f_\nu=f} H(\nu|\mu) \leq H(\kappa|\mu) \leq aH(\nu|\mu) \rightarrow 0$$

as  $a \rightarrow 0$  by the convexity of  $H(\cdot|\mu)$ , so  $\inf_{\nu: f_\nu=f} H(\nu|\mu) \in \{0, \infty\}$  also and equality holds.

- (b) See first that  $[\alpha, \beta] \neq [0, 0]$  if  $f \in \mathcal{D}_{[\alpha, \beta]}$  is not a MGF but mimics one. Note that for any  $g \in \mathcal{D}_{[\alpha, \beta]}$  and  $\mu_n(dx) = \mu(dx)/\mu([-n, n])$  for any  $n$ ,  $\gamma g + (1-\gamma)f_{\mu_n} \in \mathcal{D}_{[\alpha, \beta]}$  for any  $\gamma \in (0, 1]$ , so if  $I_M(g) < \infty$  for some  $g \in \mathcal{D}_{[\alpha, \beta]}$

$$\inf_{\{g \in \mathcal{D}_{[\alpha, \beta]}, g \text{ a MGF}\}} I_M(g) \leq I_M(\gamma g + (1-\gamma)f_{\mu_n}) \leq \gamma I_M(g) + (1-\gamma)I_M(f_{\mu_n}) \rightarrow I_M(f_{\mu_n})$$

as  $\gamma \rightarrow 0$ . By letting  $n \rightarrow \infty$  we get  $\inf_{\{g \in \mathcal{D}_{[\alpha, \beta]}, g \text{ a MGF}\}} I_M(g) = 0$ , and so in general

$$\inf_{\{g \in \mathcal{D}_{[\alpha, \beta]}, g \text{ a MGF}\}} I_M(g) \in \{0, \infty\}.$$

For the 0 value equality holds in the statement of Proposition 5.3(b) by statements 2 and 3 of Lemma 5.1, and for the  $\infty$  value equality holds by Lemma 5.2 (4)  $\Rightarrow$  (5), as by Proposition 5.2 (4) is equivalent to the expression  $\inf_{\{g \in \mathcal{D}_{[\alpha, \beta]}, g \text{ a MGF}\}} I_M(g)$  taking the value  $\infty$ .

□

## 8.2 Section 3

**Theorem 3.1. LDP for large deviation estimates.** By Theorem 4.1 we have that a LDP holds in  $\mathcal{X}_M$ . Proposition 5.2 contains part (a), while Proposition 5.3 is equivalent to parts part (b) and (c). The subsequent statements about  $\{\Lambda_n\}$ ,  $\{J_n\}$  and  $\{I_n\}$  are proved in Theorems 4.2, 4.3 and 4.4 respectively. □

**Corollary 3.1. Weak laws.** As  $\{M_n\}$  satisfies a LDP in  $(\mathcal{X}_M, \tau_{AW})$ , we know that  $P(M_n \in \cdot)$  is eventually concentrated on the closed set  $A_0 = \{f : I_M(f) = 0\}$  [33], i.e.  $P(M_n \in G) \rightarrow 1$  for every open  $G$  containing  $A_0$ . First note that  $f \in A_0$  implies that  $f(\theta) = f_\mu(\theta)$  for all  $\theta \in \mathcal{D}_f$ , as  $H(\nu|\mu)$  has a unique zero at  $\nu = \mu$ . See that  $C_x^\gamma = \{f : f(\gamma) \leq x\}$  is a closed set for every  $x$  and every  $\gamma$  by a proof identical to that of Proposition 4.5, so fix some  $\gamma$  such that  $f_\mu(\gamma) = E(e^{\gamma X_1}) < \infty$ , and let  $x > f_\mu(\theta)$ . Then by the weak law of large numbers

$$P(M_n \in C_x^\gamma) = P(M_n(\gamma) \leq x) \rightarrow 1$$

and so  $P(M_n \in \cdot)$  is eventually concentrated on  $C_x^\gamma$ . Also note that since  $\mathcal{X}_M$  is a Normal space, disjoint closed sets have disjoint open neighbourhoods and so we can show that  $P(M_n \in \cdot)$

is eventually concentrated on  $C_x^\gamma \cap A_0$ , in the following way. Fix any neighbourhood  $G$  of  $C_x^\gamma \cap A_0$ , and note that  $C_x^\gamma \setminus G$  and  $A_0 \setminus G$  are disjoint closed sets. Then they have disjoint open neighbourhoods  $U$  and  $V$  respectively, and so  $G \cup U$  and  $G \cup V$  are open neighbourhoods of  $C_x^\gamma$  and  $A_0$  respectively, with intersection  $G$ . Thus  $P(M_n \in G \cup U) \rightarrow 1, P(M_n \in G \cup V) \rightarrow 1$ , and so by the inequality

$$1 \geq P(M_n \in (G \cup U) \cup (G \cup V)) = P(M_n \in G \cup U) + P(M_n \in G \cup V) - P(M_n \in G)$$

we have that  $P(M_n \in G) \rightarrow 1$ . Thus  $P(M_n \in \cdot)$  is eventually concentrated on  $C_x^\gamma \cap A_0$ , which consists of functions  $f$  satisfying  $f(\theta) = f_\mu(\theta)$  everywhere  $f$  is finite, which includes  $\gamma$ . We can similarly show that the measures are eventually concentrated on  $C_{x_1}^{\gamma_1} \cap C_{x_2}^{\gamma_2} \cap A_0$  if  $f_\mu(\gamma_1) < x_1, f_\mu(\gamma_2) < x_2$ . Assume now that  $f_\mu$  is finite in a neighbourhood of the origin. Fix some neighbourhood  $V_k(f_\mu)$  of  $f_\mu$ , and let  $\gamma_+ > 0$  and  $\gamma_- < 0$  be large enough so that  $C_{x_+}^{\gamma_+} \cap C_{x_-}^{\gamma_-} \cap A_0 \subset V_k(f_\mu)$  for  $x_+ > f_\mu(\gamma_+), x_- > f_\mu(\gamma_-)$ . Then  $V_k(f_\mu)$  is a neighbourhood of  $C_{x_+}^{\gamma_+} \cap C_{x_-}^{\gamma_-} \cap A_0$  and so  $P(M_n \in V_k(f_\mu)) \rightarrow 1$  as  $n \rightarrow \infty$ . This can be done for all  $k$  and thus  $P(M_n \in \cdot)$  is eventually concentrated on the singleton set  $\{f_\mu\}$  and we have a weak law. If  $f_\mu(\theta) = \infty$  for all  $\theta > 0$  (resp.  $\theta < 0$ ), then in the above construction we need only use  $\gamma^-$  (resp.  $\gamma^+$ ), and if  $f_\mu$  is finite only at 0 then  $A_0$  is already a singleton set.

The results for  $\{\Lambda_n\}, \{J_n\}$  and  $\{I_n\}$  follow from the continuous mapping theorem, e.g. [6][Theorem 2.7].

□

## Section 7

**Theorem 3.2. LDP for Loynes' exponent estimates.** First see that  $\mathcal{X}_M$  is metrizable [5], as is  $[0, \infty]$ . Let  $\delta_0$  be the Dirac measure at 0, let  $G$  be the Loynes' exponent mapping, and let  $f \in \mathcal{X}_M$  neither equal  $f_{\delta_0}$  nor mimic it. It must follow that  $f$  is equal to 1 at at most one non-zero point in its domain. Therefore if  $G(f) \in [0, \infty)$ , for  $x > G(f)$   $d((x, 1), \text{epi}(f)) > 0$  and so for any sequence  $\{f_n\}$  converging to  $f$  in  $\tau_{\text{AW}}$   $f_n(x) > 1$  for large enough  $n$  and so  $G(f_n) \leq x$ . Similarly if  $G(f) \in (0, \infty]$ , for any  $0 < x < G(f)$   $f(x) < 1$  and as  $\tau_{\text{AW}}$  convergence implies pointwise convergence on the interior of  $\mathcal{D}_f$ ,  $f_n(x) < 1$  for large enough  $n$  and so  $G(f_n) \geq x$ . Together this proves that  $G(f_n) \rightarrow G(f)$  and so  $G$  is continuous at  $f$ . Therefore  $G^f$  is a singleton set containing only  $G(f)$ , the first condition of Theorem 7.1 holds, and for the sequence  $f_n = f$  for all  $n$  the second condition holds.

Now consider  $f_{\delta_0}$ . Assume  $I_M(f_{\delta_0}) < \infty$  as otherwise we need not consider it. Note that this implies that  $\mu$  has a point mass at 0. Fix any  $y \in [0, \infty]$  and any MGF  $f_\nu \in \mathcal{X}_M$  finite everywhere satisfying  $G(f_\nu) = y$  and  $I_M(f_\nu) < \infty$ . Letting  $\nu$  equal some convex combination of  $\mu$  conditioned on  $(-\infty, 0)$  and  $\mu$  conditioned on  $(0, M)$  for some  $M$  will show that such an  $f_\nu$  exists for any  $y \in [0, \infty]$ . Let  $l_\alpha = \alpha f + (1 - \alpha)f_{\delta_0}$ . Then  $l_\alpha \rightarrow f_{\delta_0}$  as  $\alpha \rightarrow 0$  and  $G(l_\alpha) = y$  for all  $\alpha \in (0, 1)$ , so that  $G(l_\alpha) \rightarrow y$  as  $\alpha \rightarrow 0$ . This gives us that  $G^{f_{\delta_0}} = [0, \infty]$ , and as  $[0, \infty]$  is compact the first condition of Theorem 7.1 holds trivially. Moreover as  $l_\alpha = f_{\alpha\nu + (1-\alpha)\delta_0}$ ,

$$\begin{aligned}
I_M(l_\alpha) &= H(\alpha\nu + (1-\alpha)\delta_0|\mu) \\
&= \int_{\mathbb{R}} (\alpha\nu(dx) + (1-\alpha)\delta_0(dx)) \log \left( \frac{\alpha\nu(dx) + (1-\alpha)\delta_0(dx)}{\mu(dx)} \right) \\
&= \alpha \int_{\mathbb{R} \setminus \{0\}} \nu(dx) \log \left( \frac{\alpha\nu(dx)}{\mu(dx)} \right) + (\alpha\nu(0) + (1-\alpha)\delta_0(0)) \log \left( \frac{\alpha\nu(0) + 1 - \alpha\delta_0(0)}{\mu(0)} \right) \\
&\rightarrow \delta_0(0) \log \left( \frac{\delta_0(0)}{\mu(0)} \right) \text{ as } \alpha \rightarrow 0 \\
&= H(\delta_0|\mu) \\
&= I_M(f_{\delta_0}).
\end{aligned}$$

Taking the limit has the desired result as  $H(\nu|\mu) < \infty$ . Therefore for every  $y \in G^{f_{\delta_0}}$ , the second condition of Theorem 7.1 holds. If  $g$  mimics  $f_{\delta_0}$  and  $I_M(g) < \infty$ , say with  $\sup \mathcal{D}_g = b$  then the proof is similar. For any  $y \in [0, b]$  the same  $f_\nu$  can be chosen, so that  $G^g$  contains  $[0, b]$ . As any sequence converging to  $g$  must obey  $f_n(\theta) \rightarrow \infty$  for all  $\theta > b$  it follows that  $\limsup_{n \rightarrow \infty} G(f_n) \leq b$  and so  $G^g = [0, b]$  and again the first condition of Theorem 7.1 holds trivially. With  $l_\alpha = \alpha f_\nu + (1-\alpha)g$  we have that  $l_\alpha$  mimics the MGF  $f_{\alpha\nu+(1-\alpha)\delta_0}$ . As  $I_M(g) < \infty$  and  $\mathcal{D}_g = \mathcal{D}_{l_\alpha}$  we can use statements 2 and 3 of Lemma 5.1 to show that  $I_M(l_\alpha) = I_M(f_{\alpha\nu+(1-\alpha)\delta_0}) = H(\alpha\nu+(1-\alpha)\delta_0|\mu)$ . Identical calculations follow as above to show that  $I_M(l_\alpha) \rightarrow I_M(f)$ .

Therefore we have that for all  $f \in \mathcal{X}_M$  satisfying  $I_M(f) < \infty$ , the conditions of Theorem 7.1 hold, and so we have that  $\{\delta_n\}$  satisfies the LDP in  $[0, \infty]$  with rate function

$$I_\delta(x) = \inf_{f: x \in G^f} I_M(f).$$

As seen by the characterisations of  $G^f$  above for all  $f \in \mathcal{X}_M$ , for  $x \in [0, \infty]$   $x \in G^f$  if and only if  $G(f) = x$ ,  $f = f_{\delta_0}$ , or  $f$  mimics  $f_{\delta_0}$  and is finite at  $x$ , i.e., if and only if  $f \in C_x$ . Therefore

$$I_\delta(x) = \inf_{f \in C_x} I_M(f),$$

as required. □

**Lemma 7.1. Positive on  $(\theta_\mu, \infty]$ .** As  $C_x$  is closed,  $I_\delta(x) = I_M(f)$  for some  $f$  satisfying  $f(x) \leq 1$ . As  $f_\mu(x) > 1$ ,  $f$  does not mimic  $f_\mu$  and so  $I_M(f) > 0$ . □

**Proposition 7.1. Conditions for unique zero.** Fix  $x \in [0, \theta_\mu)$ . If  $I_\delta(x) = 0$ , then  $I_\delta(x) = I_M(f)$  for some  $f \in \mathcal{D}_{(-\infty, x]}^c$  mimicking  $f_\mu$  with  $f(x) \leq 1$ , and as  $I_M(f) = 0$ , by the contrapositive of Lemma 5.2(5)  $\Rightarrow$  (1) Cramér's Theorem does not hold for  $e^{yX_1}$  for any  $y > x$ . If  $I_\delta(x) > 0$ , then  $I_M(f) = \infty$  for  $f$  mimicking  $f_\mu$  with  $\mathcal{D}_f = \mathcal{D}_{f_\mu} \cap (-\infty, x]$ , and so  $I_M(g) = \infty$  for all  $g$  satisfying  $\mathcal{D}_g = \mathcal{D}_{f_\mu} \cap (-\infty, x]$  by Theorem 3.1(c). So by Lemma 5.2(5)  $\Rightarrow$  (1) Cramér's Theorem holds for  $e^{yX_1}$  for some  $y \notin \mathcal{D}_{f_\mu} \cap (-\infty, x]$ . As this cannot be true for any  $y \notin \mathcal{D}_{f_\mu}$ , it must hold for some  $y > x$ , proving the first statement. Therefore if  $e^{yX_1}$  satisfies the conditions of Cramér's Theorem for all  $y \in (0, \theta_\mu)$  then  $I_\delta(x) > 0$  for all  $x \in [0, \theta_\mu)$  and so  $I_\delta$  has a unique zero, and if  $e^{xX_1}$  does not satisfy the conditions of Cramér's Theorem for some  $x \in (0, \theta_\mu)$ , then  $e^{yX_1}$  does not satisfy the conditions of Cramér's Theorem for any  $y > x$ , so  $I_\delta(x) = 0$  and the second statement also holds. □

**Theorem 7.2. Properties of  $I_\delta$ .**

- (a) Fix any finite  $x > \theta_\mu$ , and any function  $g$  with  $g(x) \leq 1$ . Then let  $f_{\mu_n}$  be the moment generating function of  $\mu$  conditioned on  $[-n, n]$ . As discussed before  $I_M(f_{\mu_n}) \rightarrow I_M(f_\mu)$ , and as  $f_\mu(x) > 1$ ,  $f_{\mu_n}(x) > 1$  for large enough  $n$ . For such an  $n$ ,

$$ag(x) + (1 - a)f_{\mu_n}(x) = 1$$

for some  $a \in (0, 1]$ , and so

$$\begin{aligned} I_\delta(x) &\leq I_M(ag + (1 - a)f_{\mu_n}) \leq aI_M(g) + (1 - a)I_M(f_{\mu_n}) \leq I_M(g) + I_M(f_{\mu_n}) \\ \Rightarrow I_\delta(x) &\leq \lim_{n \rightarrow \infty} (I_M(g) + I_M(f_{\mu_n})) = I_M(g). \end{aligned}$$

We used  $f_{\mu_n}$  because we didn't know that  $f_\mu$  was finite at  $x$ . This is true for all  $g$  with  $g(x) \leq 1$ , and so it follows from the definition of  $I_\delta$  that

$$I_\delta(x) = \inf_{f: f(x) \leq 1} I_M(f).$$

Then for all  $\theta_\mu < x < y$

$$I_\delta(\theta_\mu) \leq I_\delta(x) = \inf_{f: f(x) \leq 1} I_M(f) \leq \inf_{f: f(y) \leq 1} I_M(f) = I_\delta(y)$$

as  $f(y) \leq 1 \Rightarrow f(x) \leq 1$ . Moreover

$$I_\delta(x) = \inf_{f: f(x) \leq 1} I_M(f) \leq \inf_{f \in C_\infty} I_M(f) = I_\delta(\infty),$$

so  $I_\delta$  is increasing on  $[\theta_\mu, \infty]$ . For  $x < \theta_\mu$  satisfying  $I_\delta(x) > 0$ ,

$$I_\delta(x) = \inf_{f: f(x)=1} I_M(x) \tag{24}$$

as since  $I(x) > 0$  all functions  $f$  infinite for all  $y > x$  satisfies  $I_M(f) = \infty$  by Proposition 7.1 and Lemma 5.2. For any  $g$  with  $g(x) \geq 1$

$$ag(x) + (1 - a)f_\mu(x) = 1$$

for some  $a \in (0, 1]$  and so

$$I_\delta(x) \leq I_M(ag + (1 - a)f_\mu) \leq aI_M(g) + (1 - a)I_M(f_\mu) \leq I_M(g).$$

This is true for all  $g$  with  $g(x) \geq 1$  and so using (24)

$$I_\delta(x) = \inf_{f: f(x) \geq 1} I_M(f).$$

Similar to before we can now show that  $I_\delta(\theta_\mu) \leq I_M(x) \leq I_M(y)$  for any  $y < x$ . This is trivially true for  $y < x < \theta_\mu$  with  $I_\delta(x) = 0$ . Moreover

$$I_\delta(x) = \inf_{f: f(x) \geq 1} I_M(f) \leq \inf_{f \in C_0} I_M(f) = I_\delta(0),$$

and so  $I_\delta$  is decreasing on  $[0, \theta_\mu]$ .

- (b) If  $\mu_+, \mu_-$  are  $\mu$  conditioned on  $[0, \infty), (-\infty, 0]$  respectively then  $I_\delta(0) \leq I_M(f_{\mu_+}) < \infty$  and  $I_\delta(\infty) \leq I_M(f_{\mu_-}) < \infty$ , so together with part (a) this proves that  $I_\delta$  is finite everywhere and bounded.

- (c) Let  $\infty > x > \theta_\mu$ .  $I_\delta(x) = I_M(f) > 0$  for some  $f \in C_x$ . If  $f(x) < 1$  then for  $f_{\mu_n}$  with  $\mu_n$  the distribution  $\mu$  conditioned on  $(-\infty, n]$  we can choose  $n$  large enough so that  $f_{\mu_n}(x) > 1$  and  $I_M(f_{\mu_n}) < I_M(f)$ , as  $I_M(f_{\mu_n}) \rightarrow I_M(f_\mu) = 0$ . Moreover

$$af_{\mu_n}(x) + (1-a)f(x) = 1$$

for some  $a \in (0, 1)$  and so

$$I_\delta(x) \leq aI_M(f_{\mu_n}) + (1-a)I_M(f) < I_M(f).$$

So we cannot have  $f(x) < 1$ , and therefore we must have  $f(x) = 1$ .

- (d) First let  $\infty > x \geq \theta_\mu$ . By monotonicity and lower semi-continuity, we need only show that

$$\lim_{\epsilon \downarrow 0} I_\delta(x + \epsilon) \leq I_\delta(x) \quad (25)$$

in order to show continuity at  $x$ .  $I_\delta(x) = I_M(f)$  for some  $f \in C_x$ . We may assume that  $f$  is a moment generating function, as if  $x > \theta_\mu$  then  $f(x) = 1$  by (c), so if it is mimicking a moment generating function  $f_\nu$  then it also holds that  $f_\nu \in C_x$ . If  $x = \theta_\mu$  then we can have  $f = f_\mu$ . So  $f = f_\nu$  for some distribution  $\nu$ . Assume for now that  $f_\nu(y)$  is finite for some  $y > x$ . Then  $f_\nu(x + \epsilon) > 1$  for all  $\epsilon > 0$ , and for sufficiently small  $\epsilon$

$$a_\epsilon f_\nu(x + \epsilon) + (1 - a_\epsilon) f_{\mu_-}(x + \epsilon) = 1$$

for

$$a_\epsilon = \frac{f_{\mu_-}(x + \epsilon) - 1}{f_{\mu_-}(x + \epsilon) - f_\nu(x + \epsilon)} \rightarrow 1$$

as  $\epsilon \rightarrow 0$ . Therefore

$$I_\delta(x + \epsilon) \leq I_M(a_\epsilon f_\nu + (1 - a_\epsilon) f_{\mu_-}) \leq a_\epsilon I_M(f_\nu) + (1 - a_\epsilon) I_M(f_{\mu_-}) \rightarrow I_M(f_\nu) = I_M(x)$$

as  $\epsilon \downarrow 0$ . Now assume  $f_\nu(y) = \infty$  for all  $y > x$ . Then the support of  $\nu$  is not bounded above. Let  $f_{\nu_n}$  be the moment generating function of  $\nu$  conditioned on  $(-\infty, n]$ . Then  $f_{\nu_n}$  is finite on  $[0, \infty)$ ,  $f_{\nu_n}(x) < f_\nu(x)$ , and  $f_{\nu_n}$  converges point-wise to  $f_\nu$ , or to  $\infty$  where  $f_\nu$  is infinite. Therefore if  $\epsilon_n$  is such that  $f_{\nu_n}(x + \epsilon_n) = 1$  ( $\epsilon_n$  may not exist for small  $n$  if  $\nu([0, n]) = 0$ ), then  $\epsilon_n > 0$  and  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . Moreover

$$\begin{aligned} I_\delta(x + \epsilon_n) &\leq I_M(f_{\nu_n}) \\ \Rightarrow \lim_{n \rightarrow \infty} I_\delta(x + \epsilon_n) &\leq \lim_{n \rightarrow \infty} I_M(f_{\nu_n}) = I_M(f_\nu). \end{aligned}$$

As  $I_\delta$  is increasing on  $[\theta_\mu, \infty)$ , to prove (25) it is sufficient to prove that  $\lim_{n \rightarrow \infty} I_\delta(x + \epsilon_n) \leq I_\delta(x)$  for some positive sequence  $\epsilon_n$  converging to 0. Now assume  $0 < x \leq \theta_\mu$ . This time it is sufficient to show that

$$\lim_{\epsilon \downarrow 0} I_\delta(x - \epsilon) \leq I_\delta(x) \quad (26)$$

in order to prove continuity at  $x$ . If  $I_M(x) = 0$  and moreover  $I_M(y) = 0$  for some  $y < x$  then proof is trivial. So we need only prove continuity for  $x$  satisfying  $I_M(x) > 0$ . Then  $I_\delta(x) = I_M(f) > 0$ , so that  $f(x) = 1$  as shown in part (a). Then for all  $\epsilon > 0$ ,

$$a_\epsilon f(x - \epsilon) + (1 - a_\epsilon) f_{\mu_+}(x - \epsilon) = 1$$

for

$$a_\epsilon = \frac{f_{\mu_+}(x - \epsilon) - 1}{f_{\mu_+}(x - \epsilon) - f(x - \epsilon)} \rightarrow 1$$

as  $\epsilon \rightarrow 0$ . Moreover,

$$\begin{aligned} I_\delta(x - \epsilon) &\leq a_\epsilon I_M(f) + (1 - a_\epsilon) I_M(f_{\mu_+}) \\ \Rightarrow \lim_{\epsilon \downarrow 0} I_\delta(x - \epsilon) &\leq I_M(f) = I_\delta(x), \end{aligned}$$

as required. Continuity at  $\infty$  and at 0 is immediate by monotonicity and lower-semicontinuity, completing the proof.

□

**Acknowledgments:** The authors thank Gerald Beer (California State University) for pointing them in the right direction in proving the compactness in Proposition 4.5. This work was conducted while Brendan Williamson was visiting the National University of Ireland Maynooth from Dublin City University and then Duke University, for which funding is gratefully acknowledged.

## References

- [1] S. Asmussen. *Applied probability and queues*. Springer-Verlag, New York, 2003.
- [2] S. Asmussen and H. Albrecher. *Ruin probabilities*. World Scientific Publishing Co., Hackensack, NJ, 2010.
- [3] H. Attouch and R. J.-B. Wets. A convergence theory for saddle functions. *Trans. Amer. Math. Soc.*, 280(1):1–41, 1983.
- [4] H. Attouch and R. J.-B. Wets. Isometries for the Legendre-Fenchel transform. *Trans. Amer. Math. Soc.*, 296(1):33–60, 1986.
- [5] G. Beer. *Topologies on closed and closed convex sets*. Kluwer Academic, 1993.
- [6] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, Inc., New York, second edition, 1999.
- [7] J. Blanchet, P. Glynn, and S. Meyn. Large deviations for the empirical mean of an M/M/1 queue. *Queueing Systems*, 73(4):425–446, 2013.
- [8] A. Dembo and O. Zeitouni. *Large Deviation Techniques and Applications*. Springer, 1998.
- [9] N. G. Duffield, J. T. Lewis, N. O’Connell, R. Russell, and F. Toomey. Entropy of ATM traffic streams: a tool for estimating QoS parameters. *IEEE Journal of Selected Areas in Communications*, 13:981–990, 1995.
- [10] K. Duffy, J. T. Lewis, and W. G. Sullivan. Logarithmic asymptotics for the supremum of a stochastic process. *Ann. Appl. Probab.*, 13(2):430–445, 2003.
- [11] K. Duffy and A. P. Metcalfe. How to estimate the rate function of a cumulative process. *J. Appl. Probab.*, 42(4):1044–1052, 2005.

- [12] K. Duffy and A. P. Metcalfe. The large deviations of estimating rate functions. *J. Appl. Probab.*, 42(1):267–274, 2005.
- [13] K. R. Duffy and S. P. Meyn. Most likely paths to error when estimating the mean of a reflected random walk. *Perform. Evaluation*, 67(12):1290–1303, 2010.
- [14] K. R. Duffy and S. P. Meyn. Estimating Loynes’ exponent. *Queueing Syst.*, 68(3-4):285–293, 2011.
- [15] K. R. Duffy and S. P. Meyn. Large deviation asymptotics for busy periods. *Stochastic Syst.*, 4(1):300–319, 2014.
- [16] P. Dupuis and H. Wang. Importance sampling, large deviations, and differential games. *Stoch. Stoch. Rep.*, 76(6):481–508, 2004.
- [17] R. S. Ellis. *Entropy, large deviations, and statistical mechanics*. Classics in Mathematics. Springer-Verlag, Berlin, 2006.
- [18] A. Ganesh, P. Green, N. O’Connell, and S. Pitts. Bayesian network management. *Queueing Systems Theory Appl.*, 28(1-3):267–282, 1998.
- [19] A. Ganesh, N. O’Connell, and D. Wischik. *Big queues*. Springer-Verlag, Berlin, 2004.
- [20] A. J. Ganesh. Bias correction in effective bandwidth estimation. *Perform. Evaluation*, 27(8):319–330, 1996.
- [21] A. J. Ganesh and N. O’Connell. A large deviation principle with queueing applications. *Stoch. Stoch. Rep.*, 73(1-2):25–35, 2002.
- [22] J. Garcia. An extension of the contraction principle. *J. Theor. Probab.*, 17(2):403–434, 2004.
- [23] P. Glynn and W. Whitt. Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. *J. Appl. Probab.*, 31A:413–430, 1994.
- [24] A. N. Gupta, A. Vincent, K. Neupane, H. Yu, F. Wang, and M. T. Woodside. Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements. *Nature Phys.*, 7(8):631–634, 2011.
- [25] L. Györfi, A. Rácz, K. Duffy, J. T. Lewis, and F. Toomey. Distribution-free confidence intervals for measurement of effective bandwidths. *J. Appl. Probab.*, 37:1–12, 2000.
- [26] G. Hummer and A. Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. U.S.A.*, 98(7):3658–3661, 2001.
- [27] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690–2693, 1997.
- [28] F. P. Kelly. Notes on effective bandwidths. In F. P. Kelly, S. Zachary, and I. B. Ziedins, editors, *Stochastic Networks: Theory and Applications*, pages 141–168. Oxford University Press, 1996.
- [29] W. Kozakiewicz. On the convergence of sequences of moment generating functions. *Ann. Math. Stat.*, 18(1):61–69, 1947.
- [30] R. Kulik and Z. Palmowski. Tail behaviour of the area under a random process, with applications to queueing systems, insurance and percolations. *Queueing Syst.*, 68(3–4):275–284, 2011.

- [31] O. E. Lanford. Entropy and equilibrium states in classical statistical mechanics. In *Lecture notes in physics*, volume 20, pages 1–113. Springer, Berlin, 1973.
- [32] M. Lelarge. Tail asymptotics for discrete event systems. *Discrete Event Dyn. Syst.*, 18(4):563–584, 2008.
- [33] J. T. Lewis and C. E. Pfister. Thermodynamic probability theory: some aspects of large deviations. *Russ. Math. Surv.*, 50(2):279–317, 1995.
- [34] J. T. Lewis, C.-E. Pfister, and W. G. Sullivan. Entropy, concentration of probability and conditional limit theorems. *Markov Process. Related Fields*, 1:319–386, 1995.
- [35] J. T. Lewis, R. Russell, F. Toomey, B. McGurk, S. Crosby, and I. Leslie. Practical connection admission control for ATM networks based on on-line measurements. *Comput. Commun.*, 21(17):1585–1596, 1998.
- [36] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski’s equality. *Science*, 296(5574):1832–1835, 2002.
- [37] S. P. Meyn. *Control techniques for complex networks*. Cambridge University Press, 2008.
- [38] A. Mukherjea, M. Rao, and S. Suen. A note on moment generating functions. *Statist. Probab. Lett.*, 76(11):1185–1189, 2006.
- [39] I. P. Natanson. *Theory of functions of a real variable*. Frederick Ungar Publishing Co., New York, 1955.
- [40] A. Puhalskii. On functional principle of large deviations. *New trends in probability and statistics*, 1, 1991.
- [41] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- [42] C. M. Rohwer, F. Angeletti, and H. Touchette. Convergence of free energy and large deviation estimators. *Phys. Rev. E.*, 92:052104, 2015.
- [43] D. Ruelle. Correlation functionals. *J. Mathematical Phys.*, 6:201–220, 1965.
- [44] O.-P. Saira, Y. Yoon, T. Tanttu, M. Möttönen, D. V. Averin, and J. P. Pekola. Test of the Jarzynski and Crooks fluctuation relations in an electronic system. *Phys. Rev. Lett.*, 109(18):180601, 2012.
- [45] A. Shwartz and A. Weiss. *Large Deviations for Performance Analysis*. Chapman and Hall, 1995.
- [46] H. Touchette. The large deviation approach to statistical mechanics. *Phys. Rep.*, 478(1-3):1–69, 2009.
- [47] S. R. S. Varadhan. Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, 19:261–286, 1966.