Fundamental limits for weighted empirical approximations of exponentially tilted distributions

Anonymous Author
Anonymous Institution

Abstract

Generating samples from exponentially tilting a given distribution of random vectors when samples from the given distribution are available finds applications in fields such as finance and climate science and in the broad area of rare event simulation. In this article, we discuss the asymptotic efficiency of an estimator obtained by exponentially tilting the empirical distribution. We provide a sharp characterization of how much one can accurately tilt distributions given a certain number of samples. Our findings reveal a surprising dichotomy: While twisting unbounded distributions is a fundamentally hard task, for bounded distributions, one can accurately tilt by a large amount using much fewer samples.

1 INTRODUCTION

Exponential tilting of random variables or vectors is a technique in the field of rare event sampling and Monte Carlo simulation, that relies upon changing the underlying measure of a probability space, making rare events easier to sample from. It has wide applications ranging from finance to climate science (Ragone et al. (2018), McLeish & Men (2015)). An excellent review of exponential tilting may be found in (Alvo, 2022, Chapter 1). For a historical account of exponential tilting in the context of rare-event estimation, the reader is referred to Juneja & Shahabuddin (2006) and (Asmussen & Glynn, 2007, Chapter 6).

Exponentially tilted distributions are also central to large deviations theory when the underlying random variables are light-tailed. Conditioned on sums of a large number of independent identically distributed random variables taking a large deviation, the limiting distribution of each individual component (as the number of variables tends to infinity) is the exponentially tilted version of the original distribution (See Dembo & Zeitouni (2010)).

Our aim is to answer the question : *when is empirical tilting of random vectors hard*? We believe that the answer lies in the following two heuristics : a large tilt cannot be performed with a small number of samples, and tilting unbounded random vectors is fundamentally harder than tilting bounded ones. We shall develop theoretical versions of the aforementioned heuristics and demonstrate their accuracy through simulations.

The central objects in this article are the distributions of the tilted random vectors, and their empirical approximations. If X is a random vector on $\mathbb{R}^d, d \geq 1$, $g : \mathbb{R}^d \to \mathbb{R}^d$ is a suitable site-specific tilt, and $\theta \in \mathbb{R}^d$ is a vector such that $\mathbb{E}[e^{\theta^T g(X)}] < \infty$, then we denote by X_θ the tilted random vector, whose distribution is given by

$$\mathbb{P}[X_\theta \in A] = \frac{\mathbb{E}[e^{\theta^T g(X)} \mathbf{1}_{X \in A}]}{\mathbb{E}[e^{\theta^T g(X)}]}. \tag{1}$$

Given independent and identically distributed samples $X_1, X_2, \ldots, X_n$ of X, the reweighed empirical estimator $R_{n,\theta}$ is given by

$$\begin{aligned}\mathbb{P}[R_{n,\theta} \in A] &= \frac{\sum_{i=1}^n e^{\theta^T g(X_i)} \mathbf{1}_{X_i \in A}}{\sum_{i=1}^n e^{\theta^T g(X_i)}} \\ &= \sum_{i=1}^n \frac{e^{\theta^T g(X_i)}}{\sum_{i=1}^n e^{\theta^T g(X_i)}} 1_{X_i \in A}. \end{aligned} \tag{2}$$

The quantity central to our investigation is given by

$$M_\theta = \frac{\mathbb{E}[e^{2\theta^T g(X)}]}{\mathbb{E}[e^{\theta^T g(X)}]^2}. \tag{3}$$

Observe that $M_\theta = 1 + \mathsf{cv}(e^{\theta^T g(X)})^2$, where

$$\mathsf{cv}(Y) = \frac{\sqrt{\mathbb{V}\mathrm{ar}\,(Y)}}{\mathbb{E}[Y]}$$

Preliminary work. Under review by AISTATS 2026. Do not distribute.

is the coefficient of variation of a random variable Y, which is often used to quantify the accuracy of empirical estimation of Y. This should not be surprising, in light of the observation that $R_{n,\theta}$ is obtained by replacing $\mathbb{E}[e^{\theta^T g(X)}]$ by its empirical mean.

In this article, our objective is to delineate asymptotic regimes under which the reweighed empirical estimator $R_{n,\theta}$ accurately estimates X_θ as (n,θ) jointly vary. In particular, if KS denotes the Kolmogorov-Smirnov distance, then we achieve each of the following aims.

(a) In one dimension, fixing $\theta \in \mathbb{R}$, we provide a rate of convergence of $\mathsf{KS}(R_{n,\theta}, X_\theta)$ to zero, and demonstrate that the scaled limit's expectation and fluctuations are controlled by M_θ.

(b) If $\theta_n, n \geq 1$ converges to infinity, we create a regime in terms of M_{θ_n} within which $\mathsf{KS}(R_{n,\theta_n}, X_{\theta_n})$ converges to zero with an *a priori* rate of convergence.

(c) If X is a bounded random variable, we make tail assumptions on X under which the above regime is maximally large, and comment on the asymptotic behavior of R_{n,θ_n} if $\mathsf{KS}(X_{\theta_n}, R_{n,\theta_n}) \not\to 0$. In particular, we prove that an accurate tilt by θ_n can be performed with a number of samples which is only *polynomial* in θ_n.

(d) We repeat the above exercise for bounded random vectors in multiple dimensions and show that the number of samples required for accurate tilting is polynomial in θ_n.

(e) For unbounded random vectors, we prove that accurate tilting requires *exponentially* many samples in θ_n, in contrast to the bounded case.

We address each of these points in separate sections. Namely, in Section 2 we address tilting by fixed θ. In Section 3 we address tilted sampling as $\theta_n \to \infty$. In Section 4 we establish tight regimes on the accuracy of tilting in 1 dimension, followed by multiple dimensions in Section 5. Finally, unbounded random vectors are discussed in Section 6.

In Appendix A we prove the results from Section 2. In Appendix B we prove the results from Section 3. In Appendix C we prove a multitude of results that apply to all the regimes considered in Sections 4, 5 and 6. Following this, in Appendices D, E and F we prove the results in Sections 4, 5 and 6 respectively. Finally, Appendix G is reserved for experiments.

Notation : Whenever we speak of random vectors, we shall also include the one-dimensional case in our purview, and refer to a random vector as a random variable if it is one-dimensional. We always denote by X the random vector which is being subject to tilting. θ denotes the vector direction/scalar along which the tilt is being affected.

Throughout this article, F_Y denotes the cumulative distribution function (CDF) of any untilted random vector Y. The CDF of X_θ will be denoted by F_θ and that of the empirical distribution $R_{n,\theta}$ will be denoted by $F_{n,\theta}$. Convergence in distribution will be denoted by $\xrightarrow{d}$, while convergence in any other notion will be denoted by $\to$ and mentioned explicitly.

Measures will be indicated by μ, ν etc. and integration with respect to them will be indicated by $d\mu, d\nu$ (e.g. $\int_0^\infty f(x) d\nu(x)$). Sometimes the domain of integration may be omitted if it is the entire support of the measure.

The indicator function of a set A will be denoted by $\mathbf{1}_A$, and its complement will be denoted by A^c. The topological boundary ∂A of a Borel set $A \subset \mathbb{R}^d$ is the intersection of A^c with the closure of A.

The nomenclature $C_1, C_2, \ldots$ is reserved for constants which are independent of the parameters in the result (typically X, θ), while parameter-dependent quantities will be indicated using subscripts e.g. M_θ.

2 TILTING FOR FIXED θ

Recall the true and empirical estimators given by equation 1 and equation 2 respectively, for a random vector X and vector θ. In this section, we assume that $d = 1$ and that X is a *continuous* random variable. Furthermore, let $n \geq 1$ be a positive integer, $\theta \in \mathbb{R}$ be a *fixed* positive real number and $g : \mathbb{R} \to \mathbb{R}$ be a strictly continuous increasing function on $\mathbb{R}$. We will let n tend to infinity in what follows, and assume that $\mathbb{E}[e^{\theta g(X)}] < \infty$.

For the sake of clarity, we redefine the one-dimensional versions of equation 1 and equation 2. That is,

$$\mathbb{P}(X_\theta \in A) = \frac{\mathbb{E}[e^{\theta g(X)} \mathbf{1}_{X \in A}]}{\mathbb{E}[e^{\theta g(X)}]}, \tag{4}$$

and if $X_1, X_2, \ldots, X_n$ are iid with distribution X, then

$$\mathbb{P}(R_{n,\theta} \in A) = \frac{\sum_{i=1}^n e^{\theta g(X_i)} \mathbf{1}_{X_i \in A}}{\sum_{i=1}^n e^{\theta g(X_i)}}. \tag{5}$$

Note that $F_{n,\theta}(x)$, the CDF of $R_{n,\theta}$ is a random CDF since $R_{n,\theta}$ depends upon the random samples X_i. Let F_θ be the CDF of X_θ. Define the Kolmogorov-Smirnov (KS) distance between two random variables U, V with CDFs F_U, F_V by

$$\mathsf{KS}(U, V) = \sup_{x \in \mathbb{R}} |F_U(x) - F_V(x)|. \tag{6}$$

Our first main result reads as follows. Recall that a Gaussian field $\mathcal{G}$ on $\mathbb{R}$ is an $\mathbb{R}$-indexed stochastic process $\mathcal{G} = \{\mathcal{G}(x) : x \in \mathbb{R}\}$ such that for every $x_1, x_2, \ldots, x_m \in \mathbb{R}, m \geq 1$, the random vector $\{\mathcal{G}(x_1), \ldots, \mathcal{G}(x_m)\}$ is Gaussian.

Theorem 1. *There exists a Gaussian random field $\mathcal{G}_\theta$ on $\mathbb{R}$ such that*

$$\sqrt{n}\mathsf{KS}(F_{n,\theta}, F_\theta) \xrightarrow{d} Z,$$

where $Z = \sup_x |\mathcal{G}_\theta(x)|$. Furthermore, there exist constants $C_1, C_2 > 0$ independent of X and θ such that

$$C_1\sqrt{M_\theta} \leq \mathbb{E}[Z] \leq C_2\sqrt{M_\theta},$$

where M_θ is as in equation 3. Furthermore,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| > u) \leq e^{-u^2/M_\theta^2} \tag{7}$$

for all $u > 0$.

This theorem has multiple layers to it, which we uncover now. First, note that the appropriate rate at which the KS distance converges is $\sqrt{n}$, which is independent of θ. However, as we shall observe (see Proposition 2 for the precise result), the covariance functional of $\mathcal{G}_\theta$ depends on θ.

Studying the supremum Z of $|G_\theta|$ is essential for demonstrating asymptotic sample complexity bounds, and the theorem states that Z is strongly concentrated about its expectation whenever M_θ is small, demonstrating the cruciality of this quantity to establishing the accuracy of sampling. Furthermore, that the expectation of Z itself is of the order of M_θ implies that the average maximal fluctuation in the limit is $\Theta(M_\theta)$ with high probability.

Next, We shall provide a brief technical overview of the proof of this theorem. Since $F_{n,\theta}$ is obtained by replacing $\mathbb{E}[e^{\theta^T g(X)}]$ with its empirical estimator, it follows by the law of large numbers and the continuous mapping theorem that $\mathsf{KS}(F_{n,\theta}, F_\theta) \to 0$ almost surely. To understand the fluctuations in this convergence we use the delta-method, which combines the continuous mapping theorem with the Donsker Central-Limit-Theorem. Since the random elements $F_{n,\theta}, F_\theta$ are infinite-dimensional, we require fairly sophisicated machinery to establish these results, which can be found in (van der Vaart & Wellner, 1996, Section 3.9). This helps us establish the existence of a Gaussian process in the scaling limit. Finally, another application of the continuous mapping theorem brings us the limit Z in the theorem above.

In order to establish bounds on $\mathbb{E}[Z]$, we make use of well-established upper and lower bounds on the expected suprema of Gaussian processes, in particular the Sudakov-Fernique theorem ((van der Vaart & Wellner, 1996, Theorem A.2.5)) for the upper bound and Sudakov's inequality ((van der Vaart & Wellner, 1996, Corollary 2.2.8)) for the lower bound. However, these results require estimation of the covering numbers of a metric space associated to G_θ. In particular, on $\mathbb{R}$ define

$$d(s,t) = \sqrt{\mathbb{E}[|G_\theta(s) - G_\theta(t)|^2]} = \|\mathcal{G}_\theta(s) - \mathcal{G}_\theta(t)\|_2$$

where $\|\cdot\|_2$ denotes the L^2 distance between random variables. For any $\epsilon > 0$, the covering number $N(\epsilon)$ is the smallest size of an ϵ-cover i.e. a set such that every real number is at most ϵ away from some point in this set, in the distance d.

We develop estimates for $N(\epsilon)$ in $(\mathbb{R}, d)$ in terms of its diameter, and then show that the diameter is controlled by $\sqrt{M_\theta}$. This, along with the Sudakov-Fernique theorem and Sudakov inequality, completes the proof of the expectation bounds.

The concentration bound is a standard consequence of the Borell-TIS inequality (Theorem 16). For the detailed proof of Theorem 1, the reader is referred to Appendix A.

3 ACCURATE TILTING IN ONE DIMENSION

In this section, let X be any random variable. Unlike the previous section, we allow X to contain atoms. Let $g : \mathbb{R} \to \mathbb{R}$ be a measurable site-specific tilt function. Assume that $\mathbb{E}[e^{\eta g(X)}] < \infty$ for all $\eta > 0$, which is certainly true if $g(X)$ is bounded.

Let n be a positive integer and θ be a positive real number. In this section, θ and n will be made to tend to infinity jointly. Recall equation 4 and equation 5, the distributions of X_θ and $R_{n,\theta}$ respectively, and their associated CDFs F_θ and $F_{n,\theta}$.

Our main result for this section reads as follows.

Theorem 2. *Let $\theta_n, n \geq 1$ be a sequence. If*

$$\lim_{n\to\infty} \frac{M_{\theta_n}}{n} = 0,$$

then $s_n\mathsf{KS}(F_{n,\theta_n}(x), F_{\theta_n}(x)) \xrightarrow{d} 0$, where s_n is any sequence such that

$$\lim_{n\to\infty} s_n^2 \frac{M_{\theta_n}}{n} = 0. \tag{8}$$

That is, provided that θ_n doesn't grow too fast, it is possible to asymptotically accurately sample from X_θ with an explicit bound on the KS accuracy. Note that this includes the previous case where θ is fixed and $g(x) = x$, which implies that $s_n = o(\sqrt{n})$ works. However, no scaling limit is provided, hence this result is weaker than Theorem 1 for this specific case.

An illustrative example of this result is $X = Exp(1)$, with $\theta_n = \frac{1}{2} - \frac{1}{\sqrt{n}}$ and site-specific tilt $g(x) = x$. One computes that

$$M_{\theta_n} = \frac{(\sqrt{n}+2)^2}{8\sqrt{n}} = \Theta(\sqrt{n}).$$

Therefore, $s_n^2 \frac{M_{\theta_n}}{n} \to 0$ provided that $s_n = o\left(n^{\frac{1}{4}}\right)$. This is particularly illustrative in light of the fact that $\theta_n \to +\infty$ is not necessary.

Another application of the lemma is when X is a discrete random variable. Let X be uniform on the set $\{1, 2, \dots, 6\}$ i.e. X is the result from the roll of a fair dice. Let $\theta_n = C \log n$ for some $C < \frac{1}{6}$, $g(x) = x$, and note that

$$\frac{M_{\theta_n}}{n} = \frac{\mathbb{E}[e^{\theta_n X}]}{n} = \frac{1}{6}\sum_{i=1}^{6} e^{\theta_n i - \log n} = \frac{1}{6}\sum_{i=1}^{6} n^{Ci-1},$$

which implies that

$$\frac{M_{\theta_n}}{n} \to 0, \quad \frac{\frac{M_{\theta_n}}{n}}{n^{6C-1}} = \frac{1}{6}.$$

In particular, if $s_n = o(n^{\frac{1-6C}{2}})$ then $s_n^2 \frac{M_{\theta_n}}{n} \to 0$. We remark that this particular application cannot be replicated by any of the other results, whose tail or continuity assumptions would debar X.

We shall now provide the following rough idea of the positive bound. Note that the numerator of equation 4 is similar to the numerator of equation 5, and likewise the denominators are expected to be close for large n. Thus, the two CDFs should also be close to each other, once we combine these two observations. More rigorously, we will prove the following result.

Proposition 1. *For any $t > 0$, $n \geq 1$ and $\theta_n \in \mathbb{R}$, let $T_n = \frac{t}{2}\mathbb{E}[e^{\theta_n g(X)}]$. Then, we have*

$$\begin{aligned}
&\mathbb{P}\left[\sup_x |F_{n,\theta_n}(x) - F_{\theta_n}(x)| \geq t\right] \\
\leq &\mathbb{P}\left[\sup_x \left|\frac{1}{n}\sum_{i=1}^{n} e^{\theta_n g(X_i)}\mathbf{1}_{X_i \leq x} - \mathbb{E}[e^{\theta_n g(X)}\mathbf{1}_{X \leq x}]\right| \geq T_n\right] \\
+ &\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} e^{\theta_n g(X_i)} - \mathbb{E}[e^{\theta_n g(X)}]\right| \geq T_n\right].
\end{aligned}$$

While the second term above is rather easily bounded, the first term requires a specific concentration result from empirical process theory, which we proceed to apply in our result. We will prove the following result using (LEDERER & GEER, 2014, Corollary 3.1).

Theorem 3. *Let $Y_1, \dots, Y_n$ be iid random variables with distribution Y such that $\mathbb{E}[Y^2] < \infty$. Let*

$$Z_n = \sup_x \left|\sum_{i=1}^{n} Y_i 1_{Y_i \leq x} - \mathbb{E}[Y 1_{Y \leq x}]\right|.$$

Then, for every $\epsilon, x > 0$,

$$\mathbb{P}\left(Z_n \geq (1+\epsilon)\mathbb{E}[Z_n] + x\right) \leq \frac{c_\epsilon \sqrt{M}}{x\sqrt{n}}$$

where $M = \mathbb{E}[Y^2]$.

The proof of Theorem 2 follows by treating the second term on the right hand side of Proposition 1 using Chebyshev's inequality, and the first term using Theorem 3. We relegate its proof to Appendix B.

4 TILTING IN THE 1D WEIBULL REGIME

While the previous section creates a regime in which tilting can be performed accurately, it does not contain any negative results. It turns out that tail assumptions are necessary to obtain negative results. We demonstrate that accuracy of one-dimensional tilting can be completely disseminated, with the results looking strikingly simple in comparison to corresponding results in multiple dimensions. Furthermore, our assumptions in one dimension motivate analogous assumptions in high dimensions.

Therefore, throughout this section, let X be an *upper bounded* random variable, and

$$\mathcal{M} = \sup\{x : \mathbb{P}(X \leq x) < 1\}$$

be the supremum of the support of X. As usual, let $\theta_n, n \geq 1$ be any sequence, and recall M_{θ_n} from equation 3. We shall, under tail assumptions on X, establish results on the behavior of $\mathsf{KS}(R_{n,\theta_n}, X_{\theta_n})$ as $n \to \infty$, in the following three regimes :

(a) $\frac{M_{\theta_n}}{n} \to 0$.

(b) $\frac{M_{\theta_n}}{n} \to c$ for some $c \in (0, \infty)$.

(c) $\frac{M_{\theta_n}}{n} \to \infty$.

Our method of establishing such behavior relies on the idea of scaling limits. In particular, the following result associates scaling limits to asymptotic KS behavior.

Lemma 1. *Let $X_{1,n}, X_{2,n}$ be two sequences of random variables such that for some sequences a_n and $b_n > 0$, we have*

$$\frac{X_{i,n} - a_n}{b_n} \xrightarrow{d} Z_i, \tag{9}$$

for some random variables Z_1, Z_2.

(a) If $Z_1 \overset{d}{=} Z_2$, then $\mathsf{KS}(X_{1,n}, X_{2,n}) \to 0$ if Z_1 is continuous.

(b) If $Z_1(x) \overset{d}{\neq} Z_2(x)$ then $\mathsf{KS}(X_{1,n}, X_{2,n}) \not\to 0$ (even if one of Z_1, Z_2 has atoms).

We will prove the lemma in Appendix A. as a result, we can focus our attention on scaling limits for X_{θ_n} and R_{n,θ_n}. Note that as $\theta_n \to \infty$, samples which are closer to $\mathcal{M}$ receive more weight in the empirical estimator. This observation contributes crucially to our final results.

The famous Fisher-Tippett-Gnedenko theorem states that the maximum of $X_1, X_2, ..., X_n$ possesses a limit only in three well-known scenarios in one dimension. These are known as the Weibull, Gumbel and Frechet limits, and we restrict our attention to the Weibull regime, since it pertains specifically to (upper) bounded random variables.

Before we define the Weibull regime, we require the definition of a regularly varying function (see (Bingham et al., 1987, Page 18, Section 2)). A function $f : \mathbb{R} \to \mathbb{R}_+$ is regularly varying at 0 with index $\alpha \in \mathbb{R}$ if for all $u > 0$,

$$\lim_{t \to 0} \frac{f(tu)}{f(t)} = u^\alpha. \tag{10}$$

An analogous definition holds for f being regularly varying at (positive) infinity. As a canonical example, x^α is regularly varying of order α at 0 and infinity, for any $\alpha \in \mathbb{R}$.

Assumption 1. *We assume that the random variable X with CDF F_X falls in the* Weibull regime *with parameter α for some $\alpha > 0$. That is, the function*

$$x \mapsto \mathbb{P}(X > \mathcal{M} - x) = 1 - F_X(\mathcal{M} - x)$$

is regularly varying at 0 of order α.

Note that $\alpha = 0$ is not allowed : indeed, some bounded random variables which satisfy the above assumption but with $\alpha = 0$ do not fall into any of the extreme value regimes. For example, if $\mathbb{P}(X = \mathcal{M}) > 0$ then X does not lie in the Weibull regime, since for any $u > 0$,

$$\lim_{t \to 0} \frac{1 - F_X(\mathcal{M} - tu)}{1 - F_X(\mathcal{M} - u)} = \frac{\mathbb{P}(X = \mathcal{M})}{\mathbb{P}(X = \mathcal{M})} = 1 = u^0.$$

Since $\alpha = 0$ is debarred, X does not satisfy our assumption. Note that we were nevertheless able to obtain results on tilting such a random variable (fair dice roll) in the previous section using Theorem 2.

Some examples are as follows. If $X = U[0,1]$ then $\mathcal{M} = 1$ and $\mathbb{P}(X > 1 - u) = u$ is regularly varying of order 1 at 0. Hence, a uniform random variable is in the Weibull regime with parameter $\alpha = 1$. We include some more examples below, whose justifications will be provided in Appendix D.

Examples of Weibull Random Variables: Each of the following random variables lies in the Weibull regime.

(a) *Beta*(a, b), for any $a, b > 0$ are in the Weibull regime with parameter $\alpha = b$.

(b) Truncated normal and truncated exponential random variables are in the Weibull regime with parameter $\alpha = 1$.

(c) The sum of two (possibly dependent) random variables X, Y, where X and Y themselves lie in the Weibull regime. In fact if X, Y have parameters α, β respectively then $X + Y$ has parameter $\alpha + \beta$.

Note that it is not necessary that X needs to be a continuous random variable, even in a neighborhood of $\mathcal{M}$, since regularly varying functions do not necessarily need to be continuous. However, X must be continuous at $\mathcal{M}$, as we saw earlier.

As remarked before, the Weibull regime allows for the sample maximum to have a scaling limit (cf. (Resnick, 1987, Proposition 0.3)).

Lemma 2 (Fischer-Tippett-Gnedenko)**.** *Suppose that X is in the Weibull regime, and $X_1, X_2, ..., X_n$ are iid with distribution X. Then,*

$$\frac{\mathcal{M} - \max\{X_1, X_2, ..., X_n\}}{\mathcal{M} - F_X^{-1}(1 - \frac{1}{n})} \to -W_\alpha$$

in distribution, where W_α is a Weibull random variable having the distribution

$$G_\alpha(x) = \exp(-(1 + \alpha x)^{-1/\alpha}) \text{ for } x < 0.$$

We will also deal with general tilts in this section. Recall that $g : \mathbb{R} \to \mathbb{R}$ is a site-specific tilt function. Typical examples of g include $g(x) = x$ (exponential tilting), $g(x) = x^\alpha$ for some $\alpha > 0$, and $g(x) = C \log x$ for some $C > 0$ (polynomial tilting). However, the latter form of tilt is rather mild, and is therefore applied most often while tilting heavy-tailed random variables, which we do not consider.

Thus, we can assume that g has a regularly varying tail at $\mathcal{M}$ of positive index. However, observe that all the above functions are also strictly increasing and continuous. This motivates the assumption :

Assumption 2. *$g : \mathbb{R} \to \mathbb{R}$ is strictly increasing and continuous. Furthermore, $g(\mathcal{M}) - g(\mathcal{M} - x)$ is regularly varying at 0 of index $\beta > 0$.*

The following result shows that if X is in the Weibull regime, then $Y = g(X)$ is in the Weibull regime as well.

Lemma 3. *If X is in the Weibull regime with index $\alpha > 0$ and g satisfies Assumption 2 with parameter $\beta > 0$, then $g(X)$ is in the Weibull regime with index $\frac{\alpha}{\beta} > 0$.*

We are now ready to state our main theorems of this section. Throughout the rest of this section, let X be in the Weibull regime with parameter $\alpha > 0$, maximum value $\mathcal{M}$ and CDF F_X. Keeping Lemma 1 in mind, we know that scaling limits for X_{θ_n} and R_{n,θ_n} under the three regimes defined at the beginning of this section imply corresponding statements on the asymptotic KS accuracy of the empirical estimator. Since all the regimes depend upon the asymptotic behavior of M_{θ_n}, our first result establishes *polynomial growth* of M_{θ_n}.

Theorem 4. *We have*

$$\left(1 - F_X\left(\mathcal{M} - \frac{1}{\theta_n}\right)\right) M_{\theta_n} \to 2^{-\alpha}.$$

By Assumption 1 and a combination of (Bingham et al., 1987, Theorem 1.4.1 and Proposition 1.5.1), it is clear that M_{θ_n} grows at most polynomially. This theorem is a simple corollary of the following lemma, which follows from Karamata's Tauberian theorem for regularly varying functions.

Lemma 4. *The sequence $\mathbb{E}[e^{\theta_n X}]$, as $\theta_n \to \infty$ satisfies*

$$\frac{\mathbb{E}[e^{\theta_n X}]}{e^{\theta_n \mathcal{M}}\left(1 - F_X\left(\mathcal{M} - \frac{1}{\theta_n}\right)\right)} \to \Gamma(1+\alpha).$$

The appearance of the Γ function above hints at its appearance in the scaling limit for X_θ as well. While the analysis is subtler than the proof of Lemma 4, our next result explicitly establishes the scaling limit for X_θ in this regime as a Gamma random variable.

Theorem 5. *Let X be in the Weibull regime with parameter $\alpha > 0$ and maximum value $\mathcal{M}$. As $\theta_n \to \infty$, we have*

$$\theta_n(\mathcal{M} - X_{\theta_n}) \xrightarrow{d} \Gamma(\alpha, 1),$$

where $\Gamma(a, b)$ is the positive random variable with density

$$f_{\Gamma(a,b)}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \quad x > 0.$$

Finally, under the three regimes mentioned at the start of this section, we state the scaling limits of R_{n,θ_n}. However, a brief explanation of the below results follow for clarity.

Suppose that θ_n diverges very slowly. In that case, using the same idea as in the proof of Theorem 2, the numerators of equation 4 and equation 5 become comparable as $n \to \infty$, and likewise the denominators. Thus, we expect that in such a regime the scaling limits are likely to be the same, and given by Theorem 5.

On the other hand, if θ_n diverges too fast to infinity, then by equation 5, too much weight is placed on the maximal value of the sample $X_1, X_2, \dots, X_n$ in comparison to the rest. Therefore, the empirical estimator essentially behaves like the maximal value of the sample, whose scaling limit is known from the Weibull regime (see Lemma 2).

Our analysis, however, reveals the answer in the intermediate "critical" regime as well, which turns out to be a functional of a limiting Poisson random measure (PRM) associated to the extremal process of the samples $X_i, i \geq 1$. We defer the discussion of this interesting phenomena to Appendix C, but refer the reader to (Resnick, 1987, Chapter 3) for more details on point processes.

Theorem 6. *For a sequence $\theta_n, n \geq 1$ such that $\theta_n \to \infty$,*

(a) If $\frac{M_{\theta_n}}{n} \to 0$, then

$$\theta_n(\mathcal{M} - R_{n,\theta_n}) \to \Gamma(\alpha, 1). \tag{11}$$

(b) If $\frac{M_{\theta_n}}{n} \to c \in (0, \infty)$, then

$$\theta_n(\mathcal{M} - R_{n,\theta_n}) \to Z_{c,PRM} \tag{12}$$

for a random variable $Z_{c,PRM} \neq \Gamma(\alpha, 1)$ which depends on c and a limiting PRM.

(c) If $\frac{M_{\theta_n}}{n} \to \infty$, then

$$\frac{(\mathcal{M} - R_{n,\theta_n})}{\mathcal{M} - F_X^{-1}(1 - \frac{1}{n})} \to -W_\alpha, \tag{13}$$

where W_α is a Weibull random variable of index α.

Thus, combining Lemma 1 and Theorems 4, 5 and 6, we assert that for random variable in the Weibull domain, one can perform exponential tilting accurately with polynomially many samples in the amount of tilt.

We illustrate these results with an example. Let $X = U[0, 1]$ and $g(x) = x$, for which we know that $\alpha = 1$ and $\mathcal{M} = 1$. Furthermore,

$$1 - F_X\left(1 - \frac{1}{\theta_n}\right) = \frac{1}{\theta_n}$$

for any $\theta_n > 1$. By Theorem 4, if $\theta_n \to \infty$ we have

$$\frac{M_{\theta_n}}{\theta_n} \to \frac{1}{2}.$$

By Theorem 5,

$$\theta_n(1 - X_{\theta_n}) \xrightarrow{d} \Gamma(1,1) = Exp(1),$$

where $Exp(1)$ is the exponential random variable with parameter 1 and density λe^{-x} for $x > 0$. Note that $\frac{M_{\theta_n}}{n} \to 0$ if and only if $\frac{\theta_n}{n} \to 0$. In this case, by Theorem 6(a) it follows that

$$\theta_n(1 - R_{n,\theta_n}) \xrightarrow{d} \Gamma(1,1) = Exp(1).$$

Thus, by Lemma 1,

$$\mathsf{KS}(X_{\theta_n}, R_{n,\theta_n}) \to 0$$

provided that $n = \Omega(\theta_n)$. This shows that a substantial twist can be achieved with very few samples. If, on the other hand, $\frac{\theta_n}{n} \to c$ for some $c \in (0,\infty)$, then $\frac{M_{\theta_n}}{n} \to \frac{c}{2}$, and we have by Theorem 6(b) that

$$\theta_n(1 - R_{n,\theta_n}) \xrightarrow{d} Z_{\frac{c}{2},PRM}.$$

This random variable will not be equal to $Exp(1)$, demonstrating that $n = \Theta(\theta_n)$ samples are not enough. Furthermore, if very few samples are considered and $\frac{\theta_n}{n} \to +\infty$, then $\frac{M_{\theta_n}}{n} \to +\infty$ as well. Note that

$$1 - F_X^{-1}\left(1 - \frac{1}{n}\right) = 1 - \left(1 - \frac{1}{n}\right) = \frac{1}{n}.$$

Therefore,

$$n(1 - R_{n,\theta_n}) \to -W_1,$$

which implies that the scaling limits are totally different for $X = U[0,1]$.

All the results in this section will be proved in Appendix D.

5 TILTING IN HIGH DIMENSIONS

In this section, let X be a random vector with bounded support on $\mathbb{R}^d, d \geq 1$. Let $\theta \in \mathbb{R}^d$ be a non-zero vector, $g : \mathbb{R}^d \to \mathbb{R}^d$ be a continuous site-specific tilt weight, and recall the tilted distributions X_θ and $R_{n,\theta}$ from equation 1 and equation 2 respectively. Define the support of $X, \mathsf{supp}(X)$ as the largest closed set outside which X takes values with probability 0.

As in the one-dimensional setup, we wish to study accuracy of empirical approximation through scaling limits. However, the KS distance is fundamentally one-dimensional. The problem with considering an extension as in Jacobs et al. (2025), is that scaling limits are not respected under this extension.

Instead, we consider a weaker notion of convergence, and prove its compatibility with scaling limits.

Lemma 5. *Let $X_{1,n}, X_{2,n}$ be two sequences of random vectors such that for some sequences $a_n \in \mathbb{R}^d$ and $b_n > 0$, we have*

$$\frac{X_{i,n} - a_n}{b_n} \xrightarrow{d} Z_i, \tag{14}$$

for some random vectors Z_1, Z_2.

- *a If $Z_1 \overset{d}{=} Z_2$ and Z_1 is continuous, then $\sup_{x \in \mathcal{R}} |F_{X_1}(c) - F_{X_2}(c)| \to 0$ for every compact $\mathcal{R} \subset \mathbb{R}^d$.*
- *b If $Z_1(x) \overset{d}{\neq} Z_2(x)$ then for some compact $\mathcal{R} \subset \mathbb{R}^d$ we have $\sup_{x \in \mathcal{R}} |F_{X_1}(c) - F_{X_2}(c)| \not\to 0$.*

To begin the discussion of exponential tilting in multiple dimensions, note that we must define a notion of convergence of θ to infinity in order to discuss asymptotic accuracy. A rather natural choice would be to fix a unit vector $\theta \in \mathbb{R}^d, \|\theta\|_2 = 1$, and consider a large parameter $c > 0$. We can thus look at scaling limits of $X_{c\theta}$ and $R_{n,c\theta}$ as c converges to infinity. This would correspond to tilting towards samples $x \in \mathsf{supp}(X)$ such that $\theta^T g(x)$ is relatively large i.e. samples which, after a transformation, point in a direction close to θ.

Note that if $c \to \infty$, then we expect $X_{c\theta}$ to concentrate in regions where $\theta^T g(x)$ is maximized. Suppose there are multiple such points : then the analysis for the limit can get rather complicated and interesting. We impose the following assumption which ensures that a scaling limit is easier to obtain.

Assumption 3. *The following assumptions hold.*

(a) The functional $x \to \theta^T g(x)$ has a unique maximizer x_θ in $\mathsf{supp}(X)$.

(b) For every $\epsilon > 0$, $\mathbb{P}[\theta^T(g(X) - g(x_\theta)) < \epsilon] > 0$.

The assumption ensures that as $c \to \infty$, $X_{c\theta}$ concentrates around x_θ, and not a multitude of different values. Part (b) of the assumption ensures that x_θ is not an isolated point in $\mathsf{supp}(X)$.

This assumption is naturally satisfied in many multidimensional settings with $g(x) = x$. For instance, any random vector whose support is a convex polytope satisfies the above assumption for all directions θ. These include Dirichlet random vectors and uniform random vectors on any such shape, for instance.

Our first result establishes a "law of large numbers" for exponential tilting.

Theorem 7. *Let X be bounded and satisfy Assumption 3 in some direction θ. Then,*

$$X_{c\theta} \xrightarrow{d} x_\theta$$

as $c \to \infty$.

Proof. It is enough to prove that if A is a subset of the support of X such that the closure of A doesn't contain x_θ, then $\mathbb{P}(X_{c\theta} \in A) \to 0$. Without loss of generality, we assume that A is closed (or we can replace it by its closure below).

Since A is closed and doesn't contain x_θ, by the continuity of the function $\theta^T g$, there exists $\delta > 0$ such that $\theta^T g(y) < \theta^T g(x_\theta) - \delta$ for all $x \in A$. We write

$$\begin{aligned}\mathbb{P}(X_{c\theta} \in A) =& \frac{\mathbb{E}[e^{c\theta^T g(X)}\mathbf{1}_{X\in A}]}{\mathbb{E}[e^{c\theta^T g(X)}]} \\ =& \frac{\mathbb{E}[e^{c\theta^T(-g(x_\theta)+g(X))+c\delta/2}\mathbf{1}_{X\in A}]}{\mathbb{E}[e^{c\theta^T(-g(x_\theta)+g(X))+c\delta/2}]}. \end{aligned} \quad (15)$$

Then,

$$\mathbb{E}[e^{c\theta^T(-g(x_\theta)+g(X))+c\delta/2}\mathbf{1}_{X\in A}] \le e^{-c\delta/2}\mathbb{E}[\mathbf{1}_A] \to 0$$

as c converges to infinity. It now remains to confirm that the denominator of equation 15 is non-zero in the limit. Let

$$B = \{x \in \mathsf{supp}(X) : \theta^T(g(x) - g(x_\theta)) < \delta/4\}.$$

We know that $\mathbb{P}(X \in B) > 0$. However,

$$\mathbb{E}[e^{c\theta^T(-g(x_\theta)+g(X))+c\delta/2}] \ge \mathbb{E}[e^{c\delta/4}\mathbf{1}_B] > 0$$

regardless of c. Thus, it follows that $\mathbb{P}(X_{c\theta} \in A) \to 0$. This completes the proof. □

We shall now proceed to the scaling limit. For this, we require a few additional assumptions on g and the "tail" of X at x_θ. This, we address through a multivariate regularly-varying framework(see (Resnick, 2007, Section 2.3) and (Resnick, 1987, Chapter 5)).

Assumption 4. *We assume that X is regularly varying at x_θ i.e. there exists a non-degenerate Radon measure ν on $\{y \in \mathbb{R}^d : \theta^T y > 0\}$ such that*

$$\frac{1}{U(t)}\int f d\mathbb{P}\left(\frac{x_\theta - X}{t}\right) \to \int f d\nu$$

for all compactly supported f on $\{\theta^T y > 0\}$ which are ν-a.s. continuous, and some function $U(t)$ which is regularly varying at 0 with index $\alpha > 0$.

Furthermore, we assume that

$$\nu(\{y : 0 < \theta^T y \le 1\}) \in (0, \infty).$$

This is equivalent to saying that with non-trivial probability, empirical samples will approach x_θ along the asymptotic direction θ from the support of X.

Note that vague convergence is equivalent to

$$\frac{1}{U(t)}\mathbb{P}\left(\frac{x_\theta - X}{t} \in A\right) \to \nu(A)$$

for all compact A such that $\nu(\partial A) = 0$ where ∂A is the topological boundary of A. Using approximation arguments, it is sufficient that the above convergence holds for all A belonging to a determining class, such as all compact hypercubes or balls in $\mathbb{R}^d$.

The following facts about the limit measure ν are useful. While the first and second establish scaling properties, the third establishes a key integrability property that will be used to determine scaling limits.

Lemma 6. *Let X satisfy Assumption 4, and suppose θ, ν and α are as in the assumption. Then,*

(a) ν satisfies $\nu(cA) = c^\alpha \nu(A)$ for all compact sets A such that $\nu(\partial A) = 0$, and $c > 0$.

(b) ν is a product measure in the following sense : there exists a finite measure μ on $\mathcal{S} = \{\theta^T y = 1\}$ such that for every $a, b > 0$ and Borel $B \subset S$,

$$\nu\left(\left\{y : \theta^T y \in [a, b], \frac{y}{\theta^T y} \in B\right\}\right) = (b^\alpha - a^\alpha)\mu(B).$$

(c) We have

$$\int_{\{y:\theta^T y>0\}} e^{-\theta^T x} d\nu(x) < \infty.$$

Assumption 4 is rather abstract. This necessitates the need for examples, which are provided by the following lemma.

Examples: The following random vectors all satisfy Assumption 4.

(a) If X is a random *variable* in the Weibull regime with parameter $\alpha > 0$, CDF F_X and maximum value $\mathcal{M}$, then it satisfies Assumption 4 with $\nu([0, y]) = y^\alpha$ and $U(t) = 1 - F_X(\mathcal{M} - t)$.

(b) Bounded random vectors having independent components $X_i, 1 \le i \le d$ with maximum values $\mathcal{M}_i, 1 \le i \le d$ that lie in the one-dimensional Weibull regime (see Assumption 1) are regularly varying at $(\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_d)$.

(c) For any fixed $r > 0$, a truncated normal random vector $X = N(0, I)\mathbf{1}_{N(0,I)\le r}$ is regularly varying at any point $y \in \mathbb{R}^d$ with $\|y\| = r$.

(d) A uniform random variable on any polytope is regularly varying at any extreme point of the polytope.

We remark that the above assumption does not imply continuity of X near x_θ.

The assumption that we need to make on the site-specific tilt g motivates itself as a high-dimensional version of the corresponding one-dimensional Assumption 2. In order to keep our presentation simple, we restrict ourselves to exponential twists.

Assumption 5. *We assume that g is of the form*

$$g(x_1,\ldots,x_d) = (x_1^{\alpha_1},\ldots,x_d^{\alpha_d}),$$

for some $\alpha_1,\ldots,\alpha_d > 0$.

Exactly as in the one-dimensional case, general tilting preserves the tail assumption.

Lemma 7. *If X satisfies Assumption 4 at x_θ then $g(X)$ satisfies Assumption 4 at the point $g(x_\theta)$.*

The proof of this result is in Appendix E. From this point we assume that X satisfies Assumptions 3 and 4, and that $U(t)$ is regularly varying of order $\alpha > 0$.

As in the previous section, we begin by considering the quantity $M_{c\theta}$ and its asymptotics as $c \to \infty$, followed by the scaling limit of $X_{c\theta}$. We remark that Assumption 4 makes the proofs rather easy.

Theorem 8. *We have*

$$U\left(\frac{1}{c}\right) M_{c\theta} \to 2^{-\alpha}.$$

In particular, $M_{c\theta}$ grows at most polynomially with c. Next, we state a scaling limit for $X_{c\theta}$ as $c \to \infty$.

Theorem 9. *As $c \to \infty$,*

$$c(x_\theta - X_{c\theta}) \to Z$$

where Z is a random variable whose density is proportional to $e^{-\theta^T y}\nu(dy)$.

We remark that by the scaling equation 64, one can prove that Z above has Γ marginals, which shows that it extends the one-dimensional Theorem 4. However, these marginals may be dependent on each other based on the nature of ν.

Finally, we can obtain asymptotics for $R_{n,c\theta}$ in the same three regimes as in the previous section. The first and second regimes intuitively follow as in the previous section, (with the second regime depending upon a PRM).

However, for rapidly growing θ_n, a glance at equation 2 reveals that $R_{n,c\theta}$ should behave asymptotically like the sample maximum of $\theta^T y$ i.e. $\arg\max_{y=X_1,\ldots,X_n} \theta^T y$. This is indeed the case, but unlike the one-dimensional case, due to angularity constraints the asymptotic distribution of the latter quantity is fairly complex.

Theorem 10. *For a sequence $c_n, n \geq 1$ converging to infinity,*

(a) If $\frac{M_{c_n\theta}}{n} \to 0$, then

$$c_n(x_\theta - R_{n,c_n\theta}) \to Z, \tag{16}$$

where Z is as in Theorem 9.

(b) If $\frac{M_{c_n\theta}}{n} \to c \in (0,\infty)$, then

$$c_n(x_\theta - R_{n,c_n\theta}) \to Z_{PRM} \tag{17}$$

for a random variable $Z_{PRM} \neq Z$.

(c) If $\frac{M_{c_n\theta}}{n} \to \infty$, then

$$\frac{(x_\theta - R_{n,c_n\theta})}{U^{-1}(\frac{1}{n})} \to V, \tag{18}$$

where V is a random vector depending on ν and the limiting PRM from the previous result.

We note that the final random variable V has Weibull marginals and therefore extends the one-dimensional result. However, a combination of Lemma 5 and Theorems 8, 9 and 10 tells us that for multidimensional random variables with regularly varying tails, polynomially many samples are sufficient for tilting.

As an example, let $X = (X_1, X_2)$ where $X_1 \sim U[0,1]$ and $X_2 \sim U[0,1]^2$ are independent. Let $\theta = (1,1)$. In this case, $x_\theta = (1,1)$ is the maximizer of $\theta^T x$ for $x \in \mathsf{supp}(X) = [0,1]^2$.

We claim that X is multivariate regularly varying at x_θ. To prove this, observe that if $A = [0,x_1] \times [0,x_2]$ for some $x_1, x_2 > 0$ then for any $t > 0$,

$$\begin{aligned}\mathbb{P}\left(\frac{x_\theta - X}{t} \in A\right) =&\mathbb{P}(X \in (1,1) - tA)\\ =&\mathbb{P}(X_1 \in [1-tx_1, 1])\mathbb{P}(X_2 \in [1-tx_2, 1])\\ =&tx_1(1-\sqrt{1-tx_2})\end{aligned}$$

since X_1, X_2 are independent. Now, we have

$$\frac{1}{t^2}\mathbb{P}\left(\frac{x_\theta - X}{t} \in A\right) = x_1\frac{(1-\sqrt{1-tx_2})}{t}.$$

as $t \to 0$, we get $\frac{x_1 x_2}{2}$ in the limit. Therefore, the limit measure ν is uniquely defined by its values on rectangles, which is given by $\nu([0,x_1]\times[0,x_2]) = \frac{x_1 x_2}{2}$, and $U(t) = t^2$, which gives $\alpha = 2$.

We can now apply the theorems. In particular, Theorem 8 implies that as $c \to \infty$,

$$\frac{M_{c\theta}}{c^2} \to \frac{1}{2}.$$

In particular, $M_{c\theta}$ grows at the same rate as c^2. By Theorem 9 we have

$$c((1,1) - X_{c\theta}) \to Z$$

where Z has density proportional to $e^{-\theta^T y}\nu(dy)$, which in particular implies that

$$\begin{aligned}
&\mathbb{P}(Z \in [0,x_1] \times [0,x_2]) \\
\propto &\int_{[0,x_1]\times[0,x_2]} e^{-\theta^T y} d\nu(y) \\
\propto &\frac{1}{2}\int_0^{x_1} e^{-y_1} dy_1 \int_0^{x_2} e^{-y_2} dy_2 \\
\propto &\mathbb{P}(\Gamma(1,1) \in [0,x_1])\mathbb{P}(\Gamma(1,1) \in [0,x_2])
\end{aligned}$$

In particular, Z is a product of two independent $\Gamma(1,1) = Exp(1)$ random variables in this case. By Theorem 10(a), if $\frac{M_{c\theta}}{n} \to 0$ then $n = \Omega(c^2)$, and

$$c((1,1) - R_{n,c\theta}) \to Z$$

where Z has independent $Exp(1)$ random variables as its components, as before. On the other hand, if $\frac{c^2}{n} \to C \in (0,\infty)$ then $\frac{M_{c\theta}}{n} \to \frac{C^2}{2}$, in which case Theorem 10(b) implies that

$$c((1,1) - R_{n,c\theta}) \to Z_{\frac{C}{2},PRM}$$

for some random variable $Z_{\frac{C}{2},PRM}$ which is not equal to Z in distribution. Finally, if $n = o(c^2)$ then $\frac{M_\theta}{n} \to +\infty$ and

$$\sqrt{n}((1,1) - R_{n,c\theta}) \to V$$

where V is the scaling limit of $\arg\max_i (X_i)_1 + (X_i)_2$. It is possible to prove that V has $-W_1$-marginals, however these marginals will be correlated.

The proofs of all results in this section will be included in Appendix E.

6 TILTING IN THE UNBOUNDED SETTING

In this section, we will provide some heuristic arguments as to why tilting unbounded random variables is likely to be hard. Our discussion focuses on the role of the quantity M_θ defined by equation 3, and why, unlike the previous two sections, it is either the wrong quantity to focus on, or indicative of requiring too many samples for a small twist.

6.1 Tilting Exponential Random Variables

Consider the exponential random variable $X = Exp(\lambda)$ for some $\lambda > 0$, and recall X_θ from equation 1. It is easy to see that X_θ exists only till $\theta < \lambda$, and that $X_\theta \overset{d}{=} Exp(\lambda - \theta)$.

If $\frac{\lambda}{2} < \theta < \lambda$, then note that $M_\theta = +\infty$. Therefore, unlike theorems in the previous sections, this quantity cannot dictate the rate at which the empirical estimator $R_{n,\theta}$ given by equation 5 converges to the true distribution X_θ, in the regime $\frac{\lambda}{2} < \theta_n < \lambda$. Contrast this with the discussion following Theorem 2. Thus, unbounded random variables whose moment generating function have finite radii require separate treatment for high tilts : such a question does not arise in the bounded case.

6.2 Tilting Normal-like Random Variables

In the bounded setting, the polynomial growth rate of M_{θ_n} as $\theta \to \infty$ (see Theorems 4 and 8) allowed us to tilt bounded distributions accurately with fewer samples (see Theorems 6(a) and 10). We will demonstrate a similar trichotomy and prove that for unbounded light tailed random variables, the rate of growth of M_θ is exponential in θ. This verifies our assertion that twisting unbounded random variables is much harder!

To simplify our assumptions as much as possible, let X be a continuous random vector with full support on $\mathbb{R}^d$ and density f. Suppose that there exist $\alpha, K, L > 0$ such that

$$\lim_{x\to\infty} \frac{f(x)}{e^{-K\|x\|^\alpha}} \to L, \tag{19}$$

in that for every $\epsilon > 0$ there is an $R > 0$ such that if $\|x\| > R$ then $\frac{f(x)}{e^{-K\|x\|^\alpha}} \in (L-\epsilon, L+\epsilon)$. This includes standard Gaussian random vectors, with $\alpha = 2$ and some $K, L > 0$, for instance.

As in the previous section, fix a direction $\theta \in \mathbb{R}^d$ with $\|\theta\| = 1$, and let $c > 0$ be arbitrary. In this section, for ease of presentation we only deal with exponential tilting i.e. $g(x) = x$.

Our results, like the previous sections, are divided into three parts : the asymptotics of $M_{c\theta}$, a scaling limit for $X_{c\theta}$, and scaling limits for $R_{n,c\theta}$ in three regimes depending upon $\frac{M_{c\theta}}{n}$. We remark, however, that the results are significantly harder to prove than the bounded case, even under our extremely natural and simple assumptions.

We introduce some notation that will be useful in stating and proving our results. Let

$$\Phi_c(x) = c\theta^T x - K\|x\|^\alpha.$$

The following properties of Φ_c are important.

Lemma 8. *For each $c > 0$, the function $\Phi_c(x)$*

(a) attains its unique maximum at a point

$$m_c = \left(\frac{c}{\alpha K}\right)^{\frac{1}{\alpha-1}} \theta,$$

and

$$\Phi_c(m_c) = (\alpha-1)\alpha^{-\alpha/(\alpha-1)} c^{\alpha/(\alpha-1)} K^{-1/(\alpha-1)}.$$

(b) is thrice continuously differentiable in $\mathbb{R}^d \setminus \{0\}$, and

$$\nabla^2 \Phi_c(m_c) = -K\alpha m_c^{\alpha-2}(I + (\alpha-2)\theta\theta^T).$$

(Note : in the second term, the product of vectors is the outer product).

We are now ready to state our results.

Theorem 11. *There exists constants $p_{\alpha,K}, q_{\alpha,K}$ depending only upon α and K such that as $c_n \to \infty$,*

$$\lim_{c\to\infty} \frac{M_{c_n\theta}}{e^{p_{\alpha,K} c_n^{\alpha/(\alpha-1)}} c_n^{-d(\alpha-2)/(\alpha-1)}} = q_{\alpha,K}.$$

In particular, observe that $M_{c\theta}$ grows *exponentially* with c, since $\alpha/(\alpha-1) > 0$. The following scaling limit holds for $X_{c\theta}$.

Theorem 12. *As $c_n \to \infty$,*

$$(-\nabla^2 \Phi_c(m_c))^{-\frac{1}{2}}(X_{c_n\theta} - m_{c_n}) \xrightarrow{d} N(0, I).$$

The scaling limits for $R_{n,c\theta}$ can also be specified. Since the expressions are lengthy, we informally state this result below, and remark that they are motivated by the discussions preceding Theorems 6 and 10.

Theorem 13. *If $\frac{M_{c_n}}{n} \to 0$ then $X_n, R_{n,c_n\theta}$ have the same scaling limit. If $\frac{M_{c_n}}{n} \to c \in (0,\infty)$ then $R_{n,c_n\theta}$ converges to a PRM-based functional. Finally, if $\frac{M_{c_n}}{n} \to \infty$ then $R_{n,c_n\theta}$ has the same scaling limit as the sample maximizer of $\theta^T y$.*

These results will be proved in Appendix F.

7 CONCLUSION

In this work, we discussed the asymptotic efficiency of a distribution estimator obtained by exponentially tilting the empirical distribution in different regimes. We provided scaling limits and characterized the asymptotic sample complexity needed to accurately twist distributions in the Weibull regime. We also showed the fundamental hardness of exponentially tilting unbounded distributions.

The main limitation of this work is that we did not address unbounded random variables and vectors, particularly heavy tailed ones which lie in the Frechet regime. However, we believe this is a genuinely hard problem since such distributions can only be subject to polynomial tilting.

References

Robert J. Adler. *The Geometry of Random Fields*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia, USA, 2010. doi: 10.1137/1.9780898718980. URL https://doi.org/10.1137/1.9780898718980.

Mayer Alvo. *Exponential Tilting and Its Applications*, pp. 171–193. Springer International Publishing, Cham, 2022. ISBN 978-3-031-06784-6. doi: 10.1007/978-3-031-06784-6_6. URL https://doi.org/10.1007/978-3-031-06784-6_6.

Søren Asmussen and Peter W. Glynn. *Stochastic Simulation: Algorithms and Analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2007. ISBN 978-0-387-30679-7. doi: 10.1007/978-0-387-69033-9.

Patrick Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. John Wiley & Sons, New York, 2 edition, 1999. ISBN 978-0-471-19745-4.

N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1987. ISBN 0521307872. doi: 10.1017/CBO9780511721434.

Amir Dembo and Ofer Zeitouni. *Applications-The Finite Dimensional Case*, pp. 71–114. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 978-3-642-03311-7. doi: 10.1007/978-3-642-03311-7_3. URL https://doi.org/10.1007/978-3-642-03311-7_3.

William Feller. *An Introduction to Probability Theory and Its Applications, Volume II*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, 2nd edition, 1971. ISBN 9780471257097. doi: 10.2307/3613733.

Peter Matthew Jacobs, Foad Namjoo, and Jeff M. Phillips. Efficient and stable multi-dimensional kolmogorov-smirnov distance, 2025. URL https://arxiv.org/abs/2504.11299.

S. Juneja and P. Shahabuddin. Chapter 11 rare-event simulation techniques: An introduction and recent advances. In Shane G. Henderson and Barry L. Nelson (eds.), *Simulation*, volume 13 of *Handbooks in Operations Research* *and Management Science*, pp. 291–350. Elsevier, 2006. doi: https://doi.org/10.1016/S0927-0507(06)13011-X. URL `https://www.sciencedirect.com/science/article/pii/S092705070613011X`.

Olav Kallenberg. *Foundations of Modern Probability*. Probability and Its Applications. Springer, New York, 2 edition, 2002. ISBN 978-0-387-95313-7.

JOHANNES LEDERER and SARA VAN DE GEER. New concentration inequalities for suprema of empirical processes. *Bernoulli*, 20(4):2020–2038, 2014. ISSN 13507265. URL `http://www.jstor.org/stable/43590431`.

Donald L. McLeish and Zhongxian Men. Extreme value importance sampling for rare event risk measurement. In *Lecture Notes in Computer Science*, volume 8976, pp. 319–331. Springer, 2015. doi: 10.1007/978-3-319-09114-3_18.

Francesco Ragone, Jeroen Wouters, and Freddy Bouchet. Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences*, 115(1):24–29, 2018. doi: 10.1073/pnas.1712645115. URL `https://www.pnas.org/doi/10.1073/pnas.1712645115`.

Sidney Resnick. Multivariate regular variation on cones: Application to extreme values, hidden regular variation and conditioned limit laws. *Stochastics An International Journal of Probability and Stochastic Processes*, 80, 12 2007. doi: 10.1080/17442500701830423.

Sidney I. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 1987. ISBN 9780387964812. doi: 10.1007/978-0-387-75953-1.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, New York, 1996. ISBN 0387946403.

Supplementary Materials: Proofs and Experiments

A Proofs for Section 2

In this section, we will prove Theorem 1 with the assistance of techniques from empirical process theory and inequalities for the suprema of Gaussian processes.

We begin by recalling notation from Section 2. Let X be a continuous random variable, $n \geq 1$ be a fixed positive integer, $\theta \in \mathbb{R}$ a fixed positive real number and $g : \mathbb{R} \to \mathbb{R}$ a strictly increasing, continuous function, which in particular ensures that $e^{\theta g(x)}$ is strictly increasing in x.

Throughout this section, we set the following notation. Let $X_1, \ldots, X_n$ be independent and identically distributed as X. Let $Y = e^{\theta g(X)}$ and $Y_i = e^{\theta g(X_i)}$.

In the first part of this section, we will prove the first part of Theorem 1 i.e. the existence of the limiting Gaussian process $\mathcal{G}_\theta$ with some specified covariance. In the second part, we will prove the bounds on the limiting random variable.

A.1 Proof of Gaussian limit $\mathcal{G}_\theta$ in Theorem 1

Recall that the true and empirical distributions X_θ and $R_{n,\theta}$ are given by equation 4 and equation 5 respectively, and the KS distance is given by equation 6. Let $F_{n,\theta}(x)$ be the CDF of $R_{n,\theta}$, and for convenience define $G_n(x), x \in (-\infty, +\infty]$ by

$$G_n(x) = \frac{1}{n} \sum_{i=1}^{n} Y_i \mathbf{1}_{X_i \leq x}. \tag{20}$$

Note that $G_n(+\infty) = \frac{1}{n} \sum_{i=1}^{n} Y_i$. By the law of large numbers, the random vector $(G_n(x), G_n(+\infty))$ converges a.s. to $(\mathbb{E}[Y \mathbf{1}_{X \leq x}], \mathbb{E}[Y])$ for each fixed $x \in \mathbb{R}$. By the continuous mapping theorem, we have

$$F_{n,\theta}(x) = \frac{G_n(x)}{G_n(+\infty)} \to \frac{\mathbb{E}[Y \mathbf{1}_{X \leq x}]}{\mathbb{E}[Y]} = F_\theta(x)$$

a.s. for each $x \in \mathbb{R}$. However, the quantity we are studying is

$$\begin{aligned} \mathsf{KS}(F_{n,\theta}, F_\theta) &= \sup_x |F_{n,\theta}(x) - F_\theta(x)| \\ &= \sup_x \left| \frac{G_n(x)}{G_n(+\infty)} - \frac{\mathbb{E}[Y \mathbf{1}_{Y \leq e^{\theta g(x)}}]}{\mathbb{E}[Y]} \right|. \end{aligned} \tag{21}$$

Therefore, we require uniform control in the convergence rates of each of these random variables; in particular, we must look at the entire stochastic process $(G_n(x))_{x \in (-\infty, +\infty]}$ and show that it converges somewhere in some uniform fashion. However, we have the machinery to demonstrate not just uniform convergence, but the Gaussian limit as well. To set up the appropriate notation, let

$$\mathcal{F} = \left\{ f_x : f_x(y) = y \mathbf{1}_{y \leq e^{\theta g(x)}}, y \in \mathbb{R}, x \in (-\infty, +\infty] \right\}. \tag{22}$$

Here, $f_\infty(y) = y$. By equation 20,

$$G_n(x) = \frac{1}{n} \sum_{i=1}^{n} f_x(Y_i) \quad , \quad \mathbb{E}[G_n(x)] = \mathbb{E}[f_x(Y_i)] = \mathbb{E}[Y \mathbf{1}_{X \leq x}].$$

In order to apply results from empirical process theory, we first quickly verify the equation above (van der Vaart & Wellner, 1996, equation (2.1.1)). Observe that for any $y \in \mathbb{R}$,

$$\sup_{f_x \in \mathcal{F}} |f_x(y) - \mathbb{E}[f_x(Y)]| \le \sup_{f_x \in \mathcal{F}} |f_x(y)| + \sup_{f_x \in \mathcal{F}} \mathbb{E}[f_x(Y)] \le y + \mathbb{E}[Y].$$

The above inequality establishes that if for a class of functions $\mathcal{F}'$ on a normed space, we define

$$l^\infty(\mathcal{F}') = \left\{ h : \mathcal{F}' \to \mathbb{R}, \sup_{f \in \mathcal{F}'} |h(f)| < \infty \right\}, \quad \|h\|_{\mathcal{F}'} = \sup_{f \in \mathcal{F}'} |h(f)|, \tag{23}$$

then G_n can be seen as a random element of $l^\infty(\mathcal{F})$ by letting f_x be the input instead of x i.e. $G_n(f_x) = \frac{1}{n}\sum_{i=1}^n f_x(Y_i)$. However, we abuse notation and let $G_n(x)$ denote the random element itself.

The following lemma shows that the fluctuations $G_n - \mathbb{E}[G_n]$ converge weakly to a Gaussian process on $l^\infty(\mathcal{F})$. For the definition of weak convergence (which we denote by $\Rightarrow$), we refer the reader to (van der Vaart & Wellner, 1996, Section 1.3).

Lemma 9. *We have*

$$\sqrt{n}(G_n - \mathbb{E}[G_n]) \Rightarrow \mathcal{G}'_\theta$$

for a centered Gaussian process $\mathcal{G}'_\theta$ on $(-\infty, +\infty]$ with covariance functional given by

$$\mathbb{C}\text{ov}\left(\mathcal{G}'_\theta(x), \mathcal{G}'_\theta(y)\right) = \mathbb{E}[f_x(Y) f_y(Y)] - \mathbb{E}[f_x(Y)]\mathbb{E}[f_y(Y)] = \mathbb{C}\text{ov}\left(Y 1_{X \le x}, Y \mathbf{1}_{X \le y}\right).$$

Here, the weak convergence takes place on the space $l^\infty(\mathcal{F})$.

Before we begin the proof of the lemma, some definitions are in order. Given two functions $u, l : (-\infty, +\infty] \to \mathbb{R}$, define the "bracket" class of functions $[u, l]$ by

$$[u, l] = \{f : (-\infty, +\infty] \to \mathbb{R} : u \le f \le l\}. \tag{24}$$

Given $\epsilon > 0$, an ϵ-bracket is a class of functions $[u, l]$ such that $\sqrt{\mathbb{E}[(l(Y) - u(Y))^2]} < \epsilon$. We call a class $\mathcal{S}$ of ϵ-brackets as an ϵ-bracket cover of $\mathcal{F}$, if the union of all ϵ-brackets in $\mathcal{S}$ contains $\mathcal{F}$. The smallest size of such a cover $\mathcal{S}$ will be denoted by $N_{[]}(\epsilon, \mathcal{F}, L^2(Y))$, the *$\epsilon$-bracket covering number* of $\mathcal{F}$.

We now begin the proof of Lemma 9.

Proof of Lemma 9. By the discussion at the start of (van der Vaart & Wellner, 1996, Section 2.5.2), it suffices to prove that

$$\int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L^2(Y))} d\epsilon < \infty. \tag{25}$$

For this, we need an upper bound on $N_{[]}(\epsilon, \mathcal{F}, L^2(Y))$. This is obtained by finding at least one ϵ-bracket cover for each $\epsilon > 0$, and finding its size.

Instead, we propose a sequence of classes containing brackets, which are naturally inspired by the structure of $\mathcal{F}$. For each $n \ge 1$, let $x_1 < x_2 < \ldots < x_n$ be real numbers chosen such that if $x_{n+1} = +\infty$ and $x_0 = -\infty$ then

$$\mathbb{E}[(f_{x_i}(Y) - f_{x_{i-1}}(Y))^2] = \mathbb{E}[Y^2 1_{X \in (x_{i-1}, x_i)}] = \frac{\mathbb{E}[Y^2]}{n+1}. \tag{26}$$

for all $i = 1, 2, \ldots, n+1$. Note that such a choice requires the continuity of Y and X (and hence X alone).

Let

$$\mathcal{S}_n = \{[f_{x_i}, f_{x_{i+1}}] : i = 1, \ldots, n\} \cup \{[0, f_{x_1}]\}. \tag{27}$$

Then, $\mathcal{S}_n$ covers $\mathcal{F}$ by equation 24 and equation 22. Furthermore, by equation 26, $\mathcal{S}_n$ is a $\sqrt{\frac{\mathbb{E}[Y^2]}{n+1}}$-bracket cover of $\mathcal{F}$. Finally, $|\mathcal{S}_n| = n + 1$ by equation 27.

Thus, let $\epsilon > 0$ be arbitrary. Observe that if $\epsilon > \sqrt{\frac{\mathbb{E}[Y^2]}{n+1}}$, then $\mathcal{S}_n$ is also an ϵ-bracket cover of $\mathcal{F}$. By the definition of $N_{[]}(\epsilon, \mathcal{F}, L^2(Y))$,

$$\sqrt{\frac{\mathbb{E}[Y^2]}{n+1}} < \epsilon < \sqrt{\frac{\mathbb{E}[Y^2]}{n}} \implies \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L^2(Y))} \leq \sqrt{\log(n+1)}. \tag{28}$$

Finally, if $\epsilon > \sqrt{\mathbb{E}[Y^2]}$, then $\mathcal{S} = \{[0, f_{+\infty}]\}$ is an ϵ-bracket cover of $\mathcal{F}$. Thus, $\log N_{[]}(\epsilon, \mathcal{F}, L^2(Y)) = 0$. We can now upper bound the left hand side of equation 25 using equation 28 and this observation.

$$\begin{aligned}
&\int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L^2(Y))} d\epsilon \\
=& \int_0^{\sqrt{\mathbb{E}[Y^2]}} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L^2(Y))} d\epsilon \\
\leq& \sum_{n=1}^{\infty} \left(\sqrt{\frac{\mathbb{E}[Y^2]}{n}} - \sqrt{\frac{\mathbb{E}[Y^2]}{n+1}} \right) \sqrt{\log(n+1)} \\
\leq& \sqrt{\mathbb{E}[Y^2]} \sum_{n=1}^{\infty} \left(\frac{1}{\sqrt{n}} - \frac{1}{\sqrt{n+1}} \right) \sqrt{\log(n+1)} < C\sqrt{\mathbb{E}[Y^2]} < \infty,
\end{aligned}$$

where

$$C = \sum_{n=1}^{\infty} \left(\frac{1}{\sqrt{n}} - \frac{1}{\sqrt{n+1}} \right) \sqrt{\log(n+1)} < \infty.$$

Thus equation 25 has been proved, and consequently we have

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} f_x(Y_i) - \mathbb{E}[f_x(Y_i)] \right) \Rightarrow \mathcal{G}'_\theta \tag{29}$$

for some Gaussian process $\mathcal{G}'_\theta$ on $(-\infty, +\infty]$. The covariance functional in the lemma follows from (van der Vaart & Wellner, 1996, equation (2.1.2)). $\square$

An application of the δ-method should help us interpolate between the process $(G_n(x))_{x \in (-\infty, +\infty]}$ and $(F_{n,\theta}(x))_{x \in (-\infty, +\infty)}$.

Proposition 2. *We have*

$$\sqrt{n}(F_{n,\theta}(x) - F_\theta(x)) \Rightarrow \mathcal{G}_\theta$$

where $\mathcal{G}_\theta$ is a centered Gaussian process on $\mathbb{R}$ with covariance functional $Cov(\mathcal{G}_\theta(x_1), \mathcal{G}_\theta(x_2))$ equal to

$$\frac{1}{y_3^2} \left(v_{12} - v_{23} \frac{y_1}{y_3} - v_{13} \frac{y_2}{y_3} + v_{33} \frac{y_1 y_2}{y_3^2} \right).$$

Here, $x_3 = +\infty$, $v_{ij} = \mathbb{C}\mathrm{ov}\left(Y \mathbf{1}_{X \leq x_i}, Y \mathbf{1}_{X \leq x_j}\right)$ and $y_i = \mathbb{E}[Y \mathbf{1}_{X \leq x_i}]$.

(Note : This weak convergence takes place on the space $l^\infty(\mathcal{F} \setminus \{f_\infty\})$, which we recall from equation 23.)

Furthermore, recalling the KS distance from equation 6,

$$\sqrt{n} \mathsf{KS}(F_{n,\theta}, F_\theta) \Rightarrow \sup_x |\mathcal{G}_\theta(x)|$$

in distribution (in the usual sense i.e. as real valued random variables).

Proof. Define a map $\phi : l^\infty(\mathcal{F}) \to l^\infty(\mathcal{F} \setminus \{f_\infty\})$ by

$$[\phi(h)](f_x) = \frac{h(f_x)}{h(f_\infty)}.$$

Observe that $\phi(G_n) = F_{n,\theta}$ by equation 20 and equation 5, and similarly $\phi(\mathbb{E}[G_n]) = F_\theta$, where $\mathbb{E}[G_n]$ is just a constant functional on $\mathcal{F}$. We wish to apply (van der Vaart & Wellner, 1996, Theorem 3.9.4), for which it is sufficient to check that ϕ is Hadamard differentiable (see page 372, (van der Vaart & Wellner, 1996, Section 3.9.1) for the definition of Hadamard differentiability) on $h = \mathbb{E}[G_n]$, where we note that $\mathbb{E}[G_n(+\infty)] \neq 0$.

This is a rather standard exercise which we include for the sake of completeness. Let j_n be a sequence converging to j in $l^\infty(\mathcal{F})$, with $t_n \to 0$. For $f \in \mathcal{F}$ we must assess the limit of the quantity

$$\begin{aligned}
\left[\frac{\phi(h+t_n j_n) - \phi(h)}{t_n}\right](f) =& \frac{1}{t_n}\left[\frac{(h+t_n j_n)(f)}{(h+t_n j_n)(f_\infty)} - \frac{h(f)}{h(f_\infty)}\right] \\
=& \frac{1}{t_n}\left[(h+t_n j_n)(f)\left(\frac{1}{(h+t_n j_n)(f_\infty)} - \frac{1}{h(f_\infty)}\right) + \frac{t_n j_n(f)}{h(f_\infty)}\right] \\
=& -\frac{(h+t_n j_n)(f) j_n(f_\infty)}{(h+t_n j_n)(f_\infty) h(f_\infty)} + \frac{j_n(f)}{h(f_\infty)} \\
\to& -\frac{h(f) j(f_\infty)}{h(f_\infty)^2} + \frac{j(f)}{h(f_\infty)} = [[\phi'(h)](j)](f).
\end{aligned} \tag{30}$$

Thus, ϕ is Hadamard differentiable at h with the above derivative. By (van der Vaart & Wellner, 1996, Theorem 3.9.4),

$$\sqrt{n}(F_{n,\theta} - F_\theta) \Rightarrow [\phi'[h]](\mathcal{G}'_\theta) = \mathcal{G}_\theta. \tag{31}$$

Next, we unravel the notation on the right hand side to ascertain the true limit. Indeed, since $\phi'[h]$ is a linear map, $\mathcal{G}_\theta$ is also a centered Gaussian process. We have, by definition, that $h(f_x) = \mathbb{E}[Y 1_{X \le x}]$ and $h(f_\infty) = \mathbb{E}[Y]$. Therefore, by equation 30,

$$\mathcal{G}_\theta(f_x) = \frac{\mathcal{G}'_\theta(f_x)}{\mathbb{E}[Y]} - \frac{\mathbb{E}[Y 1_{X\le x}]\mathcal{G}'_\theta(f_\infty)}{\mathbb{E}[Y]^2}.$$

All we must do now is find the covariance using the above expression and Lemma 9. For any $x, y \in \mathbb{R}$, if we treat $\mathcal{G}_\theta(f_x) \equiv \mathcal{G}_\theta(x)$ as a stochastic process on $\mathbb{R}$ instead of a random element of $\mathcal{F} \setminus \{f_\infty\}$, we have

$$\begin{aligned}
\mathbb{C}\mathrm{ov}\left(\mathcal{G}_\theta(x), \mathcal{G}_\theta(y)\right) =& \frac{\mathbb{E}[Y 1_{X\le x}]\mathbb{E}[Y \mathbf{1}_{X \le y}]}{\mathbb{E}[Y]^4}\mathbb{V}\mathrm{ar}\left(\mathcal{G}'_\theta(\infty)\right) - \mathbb{C}\mathrm{ov}\left(\mathcal{G}'_\theta(x), \mathcal{G}'_\theta(\infty)\right)\frac{\mathbb{E}[Y\mathbf{1}_{X\le y}]}{\mathbb{E}[Y]^3} \\
&- \mathbb{C}\mathrm{ov}\left(\mathcal{G}'_\theta(y), \mathcal{G}'_\theta(\infty)\right)\frac{\mathbb{E}[Y\mathbf{1}_{X\le x}]}{\mathbb{E}[Y]^3} + \frac{\mathbb{C}\mathrm{ov}\left(\mathcal{G}'_\theta(x), \mathcal{G}'_\theta(y)\right)}{\mathbb{E}[Y]^2} \\
=& \frac{\mathbb{E}[Y 1_{X\le x}]\mathbb{E}[Y \mathbf{1}_{X \le y}]\mathbb{V}\mathrm{ar}\left(Y\right)}{\mathbb{E}[Y]^4} - \frac{\mathbb{E}[Y\mathbf{1}_{X\le y}]\mathbb{C}\mathrm{ov}\left(Y, Y\mathbf{1}_{X\le x}\right)}{\mathbb{E}[Y]^3} \\
&- \frac{\mathbb{E}[Y\mathbf{1}_{X\le x}]\mathbb{C}\mathrm{ov}\left(Y, Y\mathbf{1}_{X\le y}\right)}{\mathbb{E}[Y]^3} + \frac{\mathbb{C}\mathrm{ov}\left(Y\mathbf{1}_{Y\le x}, Y\mathbf{1}_{Y\le y}\right)}{\mathbb{E}[Y]^2}
\end{aligned}$$

which, after observation, is of the form given in the proposition, as desired.

The convergence of $\sqrt{n}\mathsf{KS}(F_{n,\theta}, F_\theta)$ follows, since

$$\mathsf{KS}(F_{n,\theta}, F_\theta) = \sup_x |F_{n,\theta} - F_\theta| = \|F_{n,\theta} - F_\theta\|_{l^\infty(\mathcal{F})}$$

by equation 23. The norm is a continuous functional, therefore by the continuous mapping theorem (van der Vaart & Wellner, 1996, Theorem 1.3.6), it follows that

$$\sqrt{n}\mathsf{KS}(F_{n,\theta}, F_\theta) \xrightarrow{d} \sup_x |\mathcal{G}_\theta(x)|,$$

as desired. $\square$

Note that Proposition 2 has completed a part of the proof of Theorem 1. The bounds on the expectation of the limiting random variable will be shown in the next subsection.

A.2 Proof of bounds in Theorem 1

We shall first state some notions required to prove the required bounds. Indeed, both the upper and lower bound can be given in terms of covering numbers; however, the notion of distance and cover will be simpler than that considered in the previous subsection.

Let $\mathcal{G}_\theta$ be as in Proposition 2 and define a distance on $\mathbb{R}$ by

$$d(s,t) = \sqrt{\mathbb{E}[|\mathcal{G}_\theta(s) - \mathcal{G}_\theta(t)|^2]}. \tag{32}$$

For any $x \in \mathbb{R}, r > 0$ let $B_d(x,r) = \{y \in \mathbb{R} : d(x,y) \le r\}$. We can now define, for any $\epsilon > 0$, the covering number

$$N(\epsilon) = \inf\{N : \exists x_1, x_2, \ldots, x_n \in \mathbb{R}, \mathbb{R} = \cup_{i=1}^n B_d(x_i, r)\}. \tag{33}$$

That is, $N(\epsilon)$ is the smallest size of a set satisfying the property that every real number is at most ϵ away (in distance d) from some point in this set.

We will also define some constants and set some notation that will make analysis easier in the upcoming sections. Define

$$K = \sup_{s,t \in \mathbb{R}} d(s,t). \tag{34}$$

K is the diameter of the metric space $(\mathbb{R}, d)$. We need estimates on K, which are difficult to obtain without having a simpler formula for d. The following lemma obtains such a formula.

Lemma 10. *For any $x_1 < x_2$ (and symmetrically for $x_2 \le x_1$), if $A = \{X \in [x_1, x_2]\}$, then*

$$d^2(x_1, x_2) = \frac{1}{\mathbb{E}[Y]^4}\left(\mathbb{E}[Y^2 \mathbf{1}_A]\mathbb{E}[Y\mathbf{1}_{A^c}]^2 + \mathbb{E}[Y^2\mathbf{1}_{A^c}]\mathbb{E}[Y\mathbf{1}_A]^2\right).$$

Assuming the above lemma, we can obtain bounds on K. Recall M_θ from equation 3.

Proposition 3. *We have*

$$\sqrt{M_\theta} \ge K \ge \frac{1}{2}\sqrt{M_\theta}.$$

Finally, the following lemma will be used to control $d(x,y)$ when x, y are (in the usual metric) close to each other.

Lemma 11. *Let $A = \{X \in [s,t]\}$. If $\frac{\mathbb{E}[Y\mathbf{1}_A]}{\mathbb{E}[Y]} = L \le \frac{1}{2}$, then*

$$d^2(s,t) \le L^2 \frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2} = L^2 M_\theta.$$

The three results above will all be proved in the order they appear in the last subsection. We shall assume these and now proceed to the proof of the lower bound.

A.2.1 Lower bound

The following theorem will be used as a lower bound for $\mathbb{E}\sup_x |\mathcal{G}_\theta(x)|$. It is (van der Vaart & Wellner, 1996, Theorem A.2.5).

Theorem 14 (Sudakov-Fernique)**.** *For any $\epsilon > 0$,*

$$\mathbb{E}\left[\sup_x |\mathcal{G}_\theta(x)|\right] \ge \frac{1}{3}\epsilon\sqrt{\log N(\epsilon)}.$$

The lower bound in Theorem 1 is now an easy consequence of the above theorem.

Proof of lower bound in Theorem 1. Let $\epsilon = K/3$. We claim that $N(\epsilon) \ge 2$. Indeed, if $N(\epsilon) = 1$, then there is a point $y \in \mathbb{R}$ such that $d(y,x) < \frac{K}{3}$ for all $x \in \mathbb{R}$. However, by the triangle inequality, for any $x, z \in \mathbb{R}$,

$$d(x,z) \le d(x,y) + d(y,z) \le \frac{2K}{3},$$

contradicting the fact that K is the diameter of the space. By Theorem 14 and Proposition 3, it follows that

$$\mathbb{E}\left[\sup_x |\mathcal{G}_\theta(x)|\right] \geq \frac{K}{9}\sqrt{\log 2} \geq C\sqrt{M_\theta}.$$

for some constant $C = \frac{\sqrt{\log 2}}{18}$ independent of $\mathcal{G}_\theta$. □

A.2.2 Upper bound

We shall now derive an upper bound on $\mathbb{E}[\sup_x |\mathcal{G}(x)|]$. For this, we use the following inequality, which is (van der Vaart & Wellner, 1996, Corollary 2.2.8), augmented with the relation between covering and packing numbers under (van der Vaart & Wellner, 1996, Definition 2.2.3, page 98).

Theorem 15. *Recall the covering number $N(\epsilon)$ from equation 33. There is a universal constant κ such that for any $t_0 \in \mathbb{R}$, we have*

$$\mathbb{E}\left[\sup_x |\mathcal{G}_\theta(x)|\right] \leq \mathbb{E}|\mathcal{G}_\theta(t_0)| + \kappa \int_0^\infty \sqrt{\log N\left(\frac{1}{2}\epsilon\right)}d\epsilon.$$

We will now prove the upper bound in Theorem 1. The proof is very similar to the proof of Lemma 9 but they differ mildly in the manner in which the cover is constructed.

Proof of upper bound in Theorem 1. We will look to bound the right hand side of Theorem 15. In order to derive an upper bound for $N(\epsilon)$ for any $\epsilon > 0$, we must find at least one set of points $S \subset \mathbb{R}$ such that for every $x \in \mathbb{R}$, $B_d(x, \epsilon) \cap S \neq \emptyset$. We can then say that $|S| \geq N(\epsilon)$.

Suppose $n \in \mathbb{N}$. By continuity of Y, we can find points $x_1, x_2, \ldots, x_n$ such that if $x_0 = -\infty, x_{n+1} = +\infty$ and $A_n = \{X \in [x_n, x_{n+1}]\}$, then

$$\frac{\mathbb{E}[Y\mathbf{1}_{A_n}]}{\mathbb{E}[Y]} = \frac{1}{n+1}. \tag{35}$$

We claim that if $\epsilon > \frac{\sqrt{M_\theta}}{n+1}$ then $\cup_{i=1}^n B_d(x_i, \epsilon) = \mathbb{R}$. Indeed, suppose that $y \in \mathbb{R}$. Then, $y \in A_i$ for some i. Let x_k be any one of the endpoints of A_i (or the only real endpoint, if the other endpoint is infinite). The interval $[y, x_k]$ (or $[x_k, y]$ depending upon the order between the two) is contained in A_i. Therefore, if $L = \frac{\mathbb{E}[Y\mathbf{1}_{X\in[y,x_k]}]}{\mathbb{E}[Y]}$, then $L \leq \frac{1}{n+1}$. By Lemma 11 and equation 35,

$$d(y, x_k) \leq L\sqrt{M_\theta} \leq \frac{\sqrt{M_\theta}}{n+1} < \epsilon.$$

This completes the proof of the claim since $y \in \mathbb{R}$ was arbitrary.

Having obtained this, we are now ready to estimate the integral on the right hand side in Theorem 15. Observe that $N\left(\frac{1}{2}\epsilon\right) \geq 2$ if $K \geq \frac{\epsilon}{2} > \frac{\sqrt{M_\theta}}{3}$, where K is defined in equation 34. On the other hand, we have shown that for any $n \geq 3 \in \mathbb{N}$ that

$$\frac{\sqrt{M_\theta}}{(n+1)} < \frac{\epsilon}{2} \leq \frac{\sqrt{M_\theta}}{n} \implies N\left(\frac{\epsilon}{2}\right) \leq n.$$

Therefore,

$$\int_0^K \sqrt{\log N\left(\frac{\epsilon}{2}\right)}d\epsilon \leq \left(2K - \frac{2\sqrt{M_\theta}}{3}\right)\sqrt{\log 2} + \sum_{n=3}^\infty \sqrt{2\log(n)}\left(\frac{\sqrt{M_\theta}}{n} - \frac{\sqrt{M_\theta}}{n+1}\right) \tag{36}$$

It only remains to show that the right hand side is bounded by a constant multiple of $\sqrt{M_\theta}$, completing the proof. However, $K \leq \sqrt{M_\theta}$ by Proposition 3. Therefore,

$$\left(K - \frac{\sqrt{M_\theta}}{3}\right)\sqrt{\log 2} \leq \left(\frac{2}{3}\sqrt{\log 2}\right)\sqrt{M_\theta} = C\sqrt{M_\theta}.$$

for some C independent of M. Applying this, the right hand side of equation 36 is bounded by

$$\sqrt{M_\theta}\left(C + \sum_{n=3}^\infty \sqrt{\log(n)}\left(\frac{1}{n} - \frac{1}{n+1}\right)\right) \leq C'\sqrt{M_\theta}$$

for some constant $C' > 0$, since the series within the bracket converges.

Combining Theorem 15, equation 36 and the above estimate, we obtain for any $t_0 \in \mathbb{R}$ that

$$\mathbb{E}\left[\sup_x |\mathcal{G}_\theta(x)|\right] \le \mathbb{E}[|\mathcal{G}_\theta(t_0)|] + C_4\sqrt{M_\theta}. \tag{37}$$

Now, $\mathcal{G}_\theta$ is a mean zero Gaussian random variable with variance specified by Proposition 2. As $t_0 \to -\infty$, observe that $Y1_{X\le t_0} \to 0$ in L^2 by the dominated convergence theorem. Therefore, $\mathbb{V}\mathrm{ar}\left(\mathcal{G}_\theta(t_0)\right) \to 0$ as $t_0 \to -\infty$. However, noting that

$$\mathbb{E}[|\mathcal{G}_\theta(t_0)|] = \sqrt{\mathbb{V}\mathrm{ar}\left(\mathcal{G}_\theta(t_0)\right)}\sqrt{2/\pi}$$

it follows that this particular term can be made as small as necessary. In particular, let t_0 be chosen so that $\mathbb{E}[|\mathcal{G}_\theta(t_0)|] \le \frac{1}{2}$. Then, by equation 37,

$$\mathbb{E}\left[\sup_x |\mathcal{G}_\theta(x)|\right] \le \frac{1}{2} + C_4\sqrt{M_\theta} \le \left(\frac{1}{2} + C_4\right)\sqrt{M_\theta}$$

since $M_\theta \ge 1$. This completes the proof. ◻

A.2.3 Proof of Lemma 10, Proposition 3 and Lemma 11

Having completed the proof of the bounds in Theorem 1, we now present the proofs of the lemmas and theorems which we used to prove them, starting with Lemma 10.

Proof of Lemma 10. Recall the notation used to define the covariance in Proposition 2. We will use this notation in our proof. That is, let $x_1, x_2 \in \mathbb{R}$, and $x_3 = +\infty$. We reuse the notation $v_{ij} = \mathbb{C}\mathrm{ov}\left(Y\mathbf{1}_{X\le x_i}, Y\mathbf{1}_{X\le x_j}\right)$ and $y_i = \mathbb{E}[Y\mathbf{1}_{X\le x_i}]$. With this, if $Z_i = \mathcal{G}_\theta(x_i)$, then

$$\begin{aligned}
\mathbb{C}\mathrm{ov}\left(Z_1, Z_1\right) &= \frac{1}{y_3^2}\left(v_{11} - 2v_{13}\frac{y_1}{y_3} + v_{33}\frac{y_1^2}{y_3^2}\right).\\
\mathbb{C}\mathrm{ov}\left(Z_2, Z_2\right) &= \frac{1}{y_3^2}\left(v_{22} - 2v_{23}\frac{y_2}{y_3} + v_{33}\frac{y_2^2}{y_3^2}\right)\\
-2\mathbb{C}\mathrm{ov}\left(Z_1, Z_2\right) &= \frac{-2}{y_3^2}\left(v_{12} - v_{23}\frac{y_1}{y_3} - v_{13}\frac{y_2}{y_3} + v_{33}\frac{y_1 y_2}{y_3^2}\right).
\end{aligned}$$

Adding up the following and multiplying by the common factor y_3^2, we obtain

$$\begin{aligned}
y_3^2 d^2(x_1, x_2) = (v_{11} + v_{22} - 2v_{12})&\\
- 2(v_{23} - v_{13})&\left(\frac{y_2 - y_1}{y_3}\right)\\
+ v_{33}&\left(\frac{(y_2 - y_1)^2}{y_3^2}\right)
\end{aligned}$$

Now, $v_{11} + v_{22} - 2v_{12} = \mathbb{V}\mathrm{ar}\left(Y\mathbf{1}_{X\in[x_1,x_2]}\right)$. Then,

$$(v_{23} - v_{13})\left(\frac{y_2 - y_1}{y_3}\right) = \mathbb{C}\mathrm{ov}\left(Y\mathbf{1}_{X\in[x_1,x_2]}, Y\left(\frac{y_2 - y_1}{y_3}\right)\right).$$

Finally,

$$v_{33}\left(\frac{(y_2 - y_1)^2}{y_3^2}\right) = \mathbb{V}\mathrm{ar}\left(Y\left(\frac{y_2 - y_1}{y_3}\right)\right).$$

Combining these three terms and adjusting the y_3^2 term (and using the identity $\mathbb{V}\mathrm{ar}\left(V + W\right) = \mathbb{V}\mathrm{ar}\left(V\right) + \mathbb{V}\mathrm{ar}\left(W\right) - 2\mathbb{C}\mathrm{ov}\left(V, W\right)$ for any two random variables V, W), we obtain

$$\mathbb{E}[Y]^2 d^2(x_1, x_2) = \mathbb{V}\mathrm{ar}\left(Y\left(\mathbf{1}_{X\in[x_1,x_2]} - \frac{\mathbb{E}[Y\mathbf{1}_{X\in[x_1,x_2]}]}{\mathbb{E}[Y]}\right)\right).$$

Let $A = \{X \in [x_1, x_2]\}$. Since the random variable on the right has expectation zero, its variance is equal to the square of its expectation i.e.

$$\begin{aligned}
\mathbb{E}[Y]^2 d^2(x_1, x_2) &= \mathbb{E}\left[Y^2 \left(\mathbf{1}_A - \frac{\mathbb{E}[Y\mathbf{1}_A]}{\mathbb{E}[Y]}\right)^2\right] \\
&= \mathbb{E}\left[Y^2 \mathbf{1}_A\right] + \frac{\mathbb{E}\left[Y^2\right]\mathbb{E}[Y\mathbf{1}_A]^2}{\mathbb{E}[Y]^2} - 2\mathbb{E}[Y^2\mathbf{1}_A]\frac{\mathbb{E}[Y\mathbf{1}_A]}{\mathbb{E}[Y]} \\
&= \frac{1}{\mathbb{E}[Y]^2}\left(\mathbb{E}[Y^2\mathbf{1}_A]\mathbb{E}[Y]^2 + \mathbb{E}[Y^2]\mathbb{E}[Y\mathbf{1}_A]^2 - 2\mathbb{E}[Y^2\mathbf{1}_A]\mathbb{E}[Y\mathbf{1}_A]\mathbb{E}[Y]\right). \qquad (38)
\end{aligned}$$

Note that by writing $\mathbb{E}[Y] = \mathbb{E}[Y\mathbf{1}_A] + \mathbb{E}[Y\mathbf{1}_{A^c}]$ and $\mathbb{E}[Y^2] = \mathbb{E}[Y^2\mathbf{1}_A] + \mathbb{E}[Y^2\mathbf{1}_{A^c}]$, we may break each of the terms into ones involving A and A^c. Doing this for the first term,

$$\mathbb{E}[Y^2\mathbf{1}_A]\mathbb{E}[Y]^2 = \mathbb{E}[Y^2\mathbf{1}_A](\mathbb{E}[Y\mathbf{1}_A]^2 + \mathbb{E}[Y\mathbf{1}_{A^c}]^2 + 2\mathbb{E}[Y\mathbf{1}_A]\mathbb{E}[Y\mathbf{1}_{A^c}]). \qquad (39)$$

For the second term, we have

$$\mathbb{E}[Y^2]\mathbb{E}[Y\mathbf{1}_A]^2 = (\mathbb{E}[Y^2\mathbf{1}_A] + \mathbb{E}[Y^2\mathbf{1}_{A^c}])\mathbb{E}[Y\mathbf{1}_A]^2. \qquad (40)$$

Finally, for the third equation we have

$$2\mathbb{E}[Y^2\mathbf{1}_A]\mathbb{E}[Y\mathbf{1}_A]\mathbb{E}[Y] = 2\mathbb{E}[Y^2\mathbf{1}_A]\mathbb{E}[Y\mathbf{1}_A]^2 + 2\mathbb{E}[Y^2\mathbf{1}_A]\mathbb{E}[Y\mathbf{1}_A]\mathbb{E}[Y\mathbf{1}_{A^c}]. \qquad (41)$$

The right hand side of equation 38 equals $\frac{1}{\mathbb{E}[Y]^2} \times$ (equation 39 +equation 40−equation 41). After multiple cancellations in the expanded expressions, we obtain

$$\mathbb{E}[Y]^2 d^2(x_1, x_2) = \frac{1}{\mathbb{E}[Y]^2}\left(\mathbb{E}[Y^2\mathbf{1}_A]\mathbb{E}[Y\mathbf{1}_{A^c}]^2 + \mathbb{E}[Y^2\mathbf{1}_{A^c}]\mathbb{E}[Y\mathbf{1}_A]^2\right).$$

Dividing by $\mathbb{E}[Y]^2$ gives the conclusion. □

This is followed by the proof of Proposition 3.

Proof of Proposition 3. By equation 34 and Lemma 10,

$$K^2 = \sup_{x_1, x_2 \in \mathbb{R}} \frac{1}{\mathbb{E}[Y]^4}\left(\mathbb{E}[Y^2\mathbf{1}_A]\mathbb{E}[Y\mathbf{1}_{A^c}]^2 + \mathbb{E}[Y^2\mathbf{1}_{A^c}]\mathbb{E}[Y\mathbf{1}_A]^2\right), \qquad (42)$$

We will first prove the upper bound. Note that for any set A, we have $\mathbb{E}[Y\mathbf{1}_A] \leq \mathbb{E}[Y]$. Directly applying this bound in the right hand side of equation 42,

$$K^2 \leq \sup_{x_1, x_2 \in \mathbb{R}} \frac{1}{\mathbb{E}[Y]^2}\left(\mathbb{E}[Y^2\mathbf{1}_A] + \mathbb{E}[Y^2\mathbf{1}_{A^c}]\right) = M_\theta.$$

We now obtain a lower bound. Since Y is a continuous random variable, by the intermediate value theorem there exist points x_1, x_2 such that $\mathbb{E}(Y\mathbf{1}_A) = \frac{1}{2}\mathbb{E}[Y]$, where we recall that $A = \{X \in [x_1, x_2]\}$. Then, $\mathbb{E}(Y\mathbf{1}_{A^c}) = \frac{1}{2}\mathbb{E}[Y]$. This implies that

$$d^2(x_1, x_2) = \frac{1}{4\mathbb{E}[Y]^2}\left(\mathbb{E}[Y^2\mathbf{1}_A] + \mathbb{E}[Y^2\mathbf{1}_{A^c}]\right) = \frac{M_\theta}{4},$$

and by equation 34,

$$K \geq d(x_1, x_2) = \frac{1}{2}\sqrt{M_\theta}.$$

□

Finally, we conclude this section with the proof of Lemma 11.

Proof of Lemma 11. Suppose that $\frac{\mathbb{E}[Y\mathbf{1}_A]}{\mathbb{E}[Y]} = L \leq \frac{1}{2}$. Then, $L \leq 1 - L$. The proof now follows from Lemma 10 :

$$\begin{aligned}
d^2(x_1, x_2) &= \frac{1}{\mathbb{E}[Y]^4} \left(\mathbb{E}[Y^2\mathbf{1}_A]\mathbb{E}[Y\mathbf{1}_{A^c}]^2 + \mathbb{E}[Y^2\mathbf{1}_{A^c}]\mathbb{E}[Y\mathbf{1}_A]^2 \right) \\
&= \frac{\mathbb{E}[Y^2\mathbf{1}_A]}{\mathbb{E}[Y]^2}(1-L)^2 + \frac{\mathbb{E}[Y^2\mathbf{1}_{A^c}]}{\mathbb{E}[Y]^2}L^2 \\
&\leq L^2 \left(\frac{\mathbb{E}[Y^2\mathbf{1}_A]}{\mathbb{E}[Y]^2} + \frac{\mathbb{E}[Y^2\mathbf{1}_{A^c}]}{\mathbb{E}[Y]^2} \right) \\
&\leq L^2 M_\theta.
\end{aligned}$$

□

A.3 Proof of the variance bound equation 7

The key inequality which will be used in the proof of the concentration bound equation 7 is the Borell-Tsirelson-Ibragimov-Sudakov (Borell-TIS) inequality (cf. (Adler, 2010, Theorem 2.1.1)).

Theorem 16 (Borell-TIS)**.** *Let $\mathscr{G}_t$ be an R^d-valued centered Gaussian process with separable index set T which is a.s. bounded. Let $\|f\| = \sup_t |\mathscr{G}_t|$. Then, for all $u > 0$,*

$$\mathbb{P}(\|f\| - \mathbb{E}\|f\| > u) \leq e^{-u^2/2\sigma^2},$$

where

$$\sigma^2 = \sup_{t\in T} \mathbb{E}[|\mathscr{G}_t|^2] = \sup_{t\in T} \mathbb{V}\mathrm{ar}\,(\mathscr{G}_t).$$

We are now ready to complete the proof of the concentration bound.

Proof of concentration bound in Theorem 1. We shall now apply this theorem to the process $\mathcal{G}_\theta(t)$, whose covariance functional is as in Proposition 2. Recall that $\sigma^2 = \sup_{t\in\mathbb{R}} \mathbb{V}\mathrm{ar}\,(\mathcal{G}_\theta(t))$. However, observe that if d is the distance defined in equation 32 then we have

$$\mathbb{V}\mathrm{ar}\,(\mathcal{G}_\theta(t)) = \lim_{x\to\infty} d(t, x).$$

This is a direct consequence of the dominated convergence theorem and the formula for the covariance functional. By Lemma 10 it follows that $\sigma^2 \leq M_\theta$.

Therefore, by Theorem 16, if $Z = \sup_t |\mathcal{G}_\theta(t)|$, then for any $u > 0$,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| > u) \leq e^{-u^2/\sigma^2} \leq e^{-u^2/M_\theta^2}.$$

This is the statement we wanted to prove. □

B Proofs for Section 3

In this section, we shall prove Theorem 2. Recall that X is any random variable and $g : \mathbb{R} \to \mathbb{R}$ is a measurable function such that $\mathbb{E}[e^{\eta g(X)}] < \infty$ for all $\eta > 0$. For a sequence θ_n, $n \geq 1$ let F_{θ_n} be the distribution of X_{θ_n} given by equation 4. Recall R_{n,θ_n} from equation 5 and its CDF F_{θ_n}.

We will begin with the proof of Proposition 1, which is a standard exercise in unwrapping the definitions of F_{θ_n} and F_{n,θ_n}.

Proof of Proposition 1. Note that for any $q \geq p > 0$ and $a, b > 0$ we have the inequality

$$\left| \frac{p}{q} - \frac{a}{b} \right| \leq \frac{|p - q| + |a - b|}{b}.$$

This follows from routine algebra :

$$\left|\frac{p}{q}-\frac{a}{b}\right| = \left|\frac{p\mathcal{M}-aq}{qb}\right| = \left|\frac{p\frac{b}{q}-a}{b}\right| = \left|\frac{p\frac{b}{q}-p+p-a}{b}\right|$$
$$= \left|\frac{\frac{p}{q}(\mathcal{M}-q)+p-a}{b}\right| \le \frac{p}{q}\frac{|\mathcal{M}-q|}{b} + \frac{|p-a|}{b} \le \frac{|\mathcal{M}-q|+|p-a|}{b}.$$

For any sequence $\theta_n, n \ge 1$, $t > 0$ and $x \in \mathbb{R}$, applying this to $|F_{n,\theta_n}(x) - F_{\theta_n}(x)|$ gives

$$\left\{\sup_x |F_{n,\theta_n}(x) - F_{\theta_n}(x)| \ge t\right\}$$
$$= \left\{\sup_x \left|\frac{\frac{1}{n}\sum_{i=1}^n e^{\theta_n g(X_i)}\mathbf{1}_{X_i\le x}}{\frac{1}{n}\sum_{i=1}^n e^{\theta_n g(X_i)}} - \frac{\mathbb{E}[e^{\theta_n g(X)}\mathbf{1}_{X\le x}]}{\mathbb{E}[e^{\theta_n g(X)}]}\right| \ge t\right\}$$
$$\subset \left\{\sup_x \frac{\left|\frac{1}{n}\sum_{i=1}^n e^{\theta_n g(X_i)}\mathbf{1}_{X_i\le x} - \mathbb{E}[e^{\theta_n g(X)}\mathbf{1}_{X\le x}]\right| + \left|\frac{1}{n}\sum_{i=1}^n e^{\theta_n g(X_i)} - \mathbb{E}[e^{\theta_n g(X)}]\right|}{\mathbb{E}[e^{\theta_n g(X)}]} \ge t\right\}$$

We must now separate the above set into two parts, one of which involves only the numerator comparison, and the other the denominator comparison. To do this, we note that if $a+b \ge t$ then either $a \ge t/2$ or $b \ge t/2$. Using this fact,

$$\left\{\sup_x \frac{\left|\frac{1}{n}\sum_{i=1}^n e^{\theta_n g(X_i)}\mathbf{1}_{X_i\le x} - \mathbb{E}[e^{\theta_n g(X)}\mathbf{1}_{X\le x}]\right| + \left|\frac{1}{n}\sum_{i=1}^n e^{\theta_n g(X_i)} - \mathbb{E}[e^{\theta_n g(X)}]\right|}{\mathbb{E}[e^{\theta_n g(X)}]} \ge t\right\}$$
$$\subset \left\{\sup_x \frac{\left|\frac{1}{n}\sum_{i=1}^n e^{\theta_n g(X_i)}\mathbf{1}_{X_i\le x} - \mathbb{E}[e^{\theta_n g(X)}\mathbf{1}_{X\le x}]\right|}{\mathbb{E}[e^{\theta_n g(X)}]} \ge t/2\right\} \cup \left\{\frac{\left|\frac{1}{n}\sum_{i=1}^n e^{\theta_n g(X_i)} - \mathbb{E}[e^{\theta_n g(X)}]\right|}{\mathbb{E}[e^{\theta_n g(X)}]} \ge t/2\right\}$$
$$\subset \left\{\sup_x \left|\frac{1}{n}\sum_{i=1}^n e^{\theta_n g(X_i)}\mathbf{1}_{X_i\le x} - \mathbb{E}[e^{\theta_n g(X)}\mathbf{1}_{X\le x}]\right| \ge \frac{t}{2}\mathbb{E}[e^{\theta_n g(X)}]\right\} \cup \left\{\left|\frac{1}{n}\sum_{i=1}^n e^{\theta_n g(X_i)} - \mathbb{E}[e^{\theta_n g(X)}]\right| \ge \frac{t}{2}\mathbb{E}[e^{\theta_n g(X)}]\right\}$$

Combining the above two set containments and using the union bound,

$$\mathbb{P}\left[\sup_x |F_{n,\theta_n}(x) - F_{\theta_n}(x)| \ge t\right] \le \mathbb{P}\left[\sup_x \left|\frac{1}{n}\sum_{i=1}^n e^{\theta_n g(X_i)}\mathbf{1}_{X_i\le x} - \mathbb{E}[e^{\theta_n g(X)}\mathbf{1}_{X\le x}]\right| \ge \frac{t}{2}\mathbb{E}[e^{\theta_n g(X)}]\right]$$
$$+\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^n e^{\theta_n g(X_i)} - \mathbb{E}[e^{\theta_n g(X)}]\right| \ge \frac{t}{2}\mathbb{E}[e^{\theta_n g(X)}]\right]$$

This is precisely the statement of the proposition with $T_n = \frac{t}{2}\mathbb{E}[e^{\theta_n g(X)}]$. □

Next, we prove Theorem 3.

Proof of Theorem 3. By the remark straddled between pages 3 and 4 of LEDERER & GEER (2014), it is sufficient to prove that if $Y_1, \dots, Y_n$ are iid random variables with distribution Y such that $\mathbb{E}[Y^2] < \infty$, $x_1, x_2, \dots, x_m \in \mathbb{R}$ are arbitrary points, and

$$Z_{m,n} = \sup_{j=1,\dots,m} \left|\frac{1}{n}\sum_{i=1}^n Y_i 1_{Y_i \le x_j} - \mathbb{E}[Y 1_{Y\le x_j}]\right|, \tag{43}$$

then for every $\epsilon, x > 0$,

$$\mathbb{P}\left(Z_{m,n} \ge (1+\epsilon)\mathbb{E}[Z_{m,n}] + x\right) \le \frac{c_\epsilon\sqrt{M}}{x\sqrt{n}} \tag{44}$$

where $M = \mathbb{E}[Y^2]$.

Our attempt, for the rest of this proof, is to match our notation to LEDERER & GEER (2014) and use (LEDERER & GEER, 2014, Corollary 3.1). Let

$$Z(j) = (Z_i(j))_{i=1,\dots,n} = \left(Y_i \mathbf{1}_{Y_i \le x_j} - \mathbb{E}[Y \mathbf{1}_{Y \le x_j}]\right)_{i=1,\dots,n} \tag{45}$$

be an n-dimensional vector depending upon j. Let

$$\sigma = \max_{1 \le j \le m} \sqrt{\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[Z_i(j)^2]} = \max_{1 \le j \le m} \sqrt{\mathbb{V}\mathrm{ar}\,(Y \mathbf{1}_{Y \le x_j})}. \tag{46}$$

Note that σ is independent of n. Furthermore, by equation 43 and equation 45,

$$Z_{m,n} = \max_{1 \le j \le m} \left| \frac{1}{n} \sum_{i=1}^{n} Z_i(j) \right| \tag{47}$$

is in line with (LEDERER & GEER, 2014, (c)). Define $\mathcal{E}_i = Y_i + \mathbb{E}[Y]$ for $1 \le i \le n$, and observe that by equation 45 and equation 47,

$$|Z_i(j)| \le \mathcal{E}_i, \quad \mathbb{E}[\mathcal{E}_i^2] \le \mathbb{E}[Y^2] + 3\mathbb{E}[Y]^2 \le M_0^2,$$

where $M_0 = 2\sqrt{\mathbb{E}[Y^2]}$. Thus, $\mathcal{E}_i$ and M_0 together satisfy (LEDERER & GEER, 2014, (4)) with $p = 2$.

We have everything in place to apply (LEDERER & GEER, 2014, Corollary 3.1) with $l = 1$ and $p = 2$ on the right hand side. Doing so, we have for every $\epsilon, x > 0$ that

$$\mathbb{P}\left(Z_{m,n} \ge (1+\epsilon)\mathbb{E}[Z_{m,n}] + x\right) \le \frac{\left(\frac{64}{\epsilon} + 7 + \epsilon\right)\frac{M_0}{\sqrt{n}} + \frac{4}{n\sigma}}{x}.$$

Note that as $n \to \infty$, $\frac{M_0/\sqrt{n}}{4/(n\sigma)} \to +\infty$. Therefore, there exists a constant $C > 0$ such that $\frac{M_0}{\sqrt{n}} \ge C\frac{4}{n\sigma}$ for all $n \ge 1$, which implies along with $M_0 = 2\sqrt{\mathbb{E}[Y^2]}$ that

$$\mathbb{P}\left(Z_{m,n} \ge (1+\epsilon)\mathbb{E}[Z_{m,n}] + x\right) \le c_\epsilon \frac{\sqrt{\mathbb{E}[Y^2]}}{x\sqrt{n}},$$

which matches equation 44 that was to be proved. ∎

Finally, we conclude the proof of Theorem 2.

Proof of Theorem 2. By definition of convergence in probability and the KS distance equation 6, we must prove that for every $t > 0$,

$$s_n \mathbb{P}\left[\sup_x |F_{n,\theta_n}(x) - F_\theta(x)| \ge t\right] \to 0 \tag{48}$$

as $n \to \infty$, if $s_n^2 \frac{M_{\theta_n}}{n} \to 0$. Ignoring the s_n for now, we directly apply Proposition 1 to the rest of the left hand side :

$$\begin{aligned}\mathbb{P}\left[\sup_x |F_{n,\theta_n}(x) - F_{\theta_n}(x)| \ge t\right] \le &\mathbb{P}\left[\sup_x \left|\frac{1}{n}\sum_{i=1}^{n} e^{\theta_n g(X_i)} \mathbf{1}_{X_i \le x} - \mathbb{E}[e^{\theta_n g(X)} \mathbf{1}_{X \le x}]\right| \ge \frac{t}{2}\mathbb{E}[e^{\theta_n g(X)}]\right] \\ &+ \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} e^{\theta_n g(X_i)} - \mathbb{E}[e^{\theta_n g(X)}]\right| \ge \frac{t}{2}\mathbb{E}[e^{\theta_n g(X)}]\right]\end{aligned} \tag{49}$$

At this point we just need to estimate the two terms on the right hand side above. The second one is much easier and can be bounded by Chebyshev's inequality :

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} e^{\theta_n g(X_i)} - \mathbb{E}[e^{\theta_n g(X)}]\right| \ge \frac{t}{2}\mathbb{E}[e^{\theta_n g(X)}]\right] \le \frac{\mathbb{V}\mathrm{ar}\left(\frac{1}{n}\sum_{i=1}^{n} e^{\theta_n g(X_i)}\right)}{\frac{t^2}{4}\mathbb{E}[e^{\theta_n g(X)}]^2} \le \frac{4M_{\theta_n}}{nt^2}.$$

For the first term on the right hand side of Proposition 1, we apply Theorem 3 with $Y = e^{\theta_n g(X)}$, $\epsilon = 1$ and $x = \frac{t}{2}\mathbb{E}[Y] - 2\mathbb{E}[Z_n]$, where Z_n is as in the theorem. This gives

$$\mathbb{P}\left(Z_n \geq \frac{t}{2}\mathbb{E}[Y]\right) \leq \frac{c}{\left(\frac{t}{2}\mathbb{E}[Y] - 2\mathbb{E}[Z_n]\right)\sqrt{n}}\sqrt{\mathbb{E}[Y^2]} = \frac{c}{\sqrt{n}}\sqrt{\frac{\mathbb{E}[Y^2]}{\left(\frac{t}{2}\mathbb{E}[Y] - 2\mathbb{E}[Z_n]\right)^2}}$$

for some constant $c > 0$ independent of X, θ and n.

Now, let s_n be a sequence satisfying equation 8. Combining the above inequality with equation 50 and equation 49, and multiplying by s_n on both sides,

$$s_n\mathbb{P}\left[\sup_x |F_n(x) - F_{\theta_n}(x)| \geq t\right] \leq s_n\frac{4M_{\theta_n}}{nt^2} + s_n\frac{c}{\sqrt{n}}\sqrt{\frac{\mathbb{E}[Y^2]}{\left(\frac{t}{2}\mathbb{E}[Y] - 2\mathbb{E}[Z_n]\right)^2}}.$$

The first term above converges to 0 by equation 8, since $\frac{M_{\theta_n}}{n} \to 0$. For the second term, it is clear that $\mathbb{E}[Z_n] \to 0$ since $Z_n \xrightarrow{d} 0$ by the Glivenko Cantelli lemma. Consequently,

$$\sqrt{\frac{\mathbb{E}[Y^2]}{\left(\frac{t}{2}\mathbb{E}[Y] - 2\mathbb{E}[Z_n]\right)^2}} \times \frac{1}{\frac{2}{t}\sqrt{M_{\theta_n}}} \to 1 \implies s_n\frac{c}{\sqrt{n}}\sqrt{\frac{\mathbb{E}[Y^2]}{\left(\frac{t}{2}\mathbb{E}[Y] - 2\mathbb{E}[Z_n]\right)^2}} \times \frac{\sqrt{n}}{\frac{2}{t}s_n c\sqrt{M_{\theta_n}}} \to 1$$

as $n \to \infty$. Now, since s_n satisfies equation 8,

$$\frac{\sqrt{n}}{\frac{2}{t}s_n c\sqrt{M_{\theta_n}}} \to +\infty,$$

which implies that the other term must converge to 0 if the product is to converge to 1.

Thus, equation 48 has been proved for arbitrary $t > 0$, completing the proof. □

C Auxiliary material for the proofs in Section 4, 5 and 6

In this section, we will collect results that apply to all the upcoming sections. We divide these into three parts. The first part will be used to prove Lemma 1 and Lemma 5. The second will be used to prove the positive results in each regime i.e. when $\frac{M_{\theta_n}}{n} \to 0$. In the final part we will define Poisson random measures and prove some key lemmas which will allow us to isolate the common aspects of the proofs in the regimes where $\frac{M_{\theta_n}}{n} \not\to 0$.

Throughout this section, let X be a random vector in $\mathbb{R}^d, d \geq 1$ (note : we include the case $d = 1$ where X is just a random variable).

C.1 Strengthening convergence of random vectors

We begin with the following lemma which is key to establishing the proofs of Lemma 1 and Lemma 5. Recall the KS distance from equation 6 and the notion of uniform convergence on compacts from Lemma 5. The first lemma strengthens convergence in distribution to convergence in these distances, under the assumption that the limiting distribution is continuous. The second asserts that two random variables are different if and only if their CDFs differ at a common point of continuity.

Lemma 12. *(a) Suppose that Y_n is a sequence of random vectors which converge, in distribution, to a continuous random vector Z. Then, if Z is continuous, we have $\mathsf{KS}(Y_n, Z) \to 0$ (if $d = 1$), and uniform convergence over compacts of F_{Y_n} to F_Z (if $d > 1$).*

(b) If Z_1, Z_2 are random vectors and if $F_{Z_1}(c) = F_{Z_2}(c)$ at every $c \in \mathbb{R}^d$ at which both F_{Z_1}, F_{Z_2} are continuous, then $Z_1 \stackrel{d}{=} Z_2$.

Proof. We begin with the proof of part (a). For the KS distance, the argument is as in (Resnick, 1987, page 3). We can adapt this argument to uniform convergence on compacts as well.

Let $\mathcal{R} \subset \mathbb{R}^d$ be any axis-aligned compact hypercube. Since Z is continuous, we have $F_{X_n}(c) \to F_Z(c)$ for every $c \in \mathcal{R}$. We claim that, in fact, the convergence above is uniform over all $c \in \mathcal{R}$.

To prove this, note that $\mathcal{R}$ is compact, hence F_Z is uniformly continuous over $\mathcal{R}$. Given $\epsilon > 0$, we can divide $\mathcal{R}$ into finitely many axis-aligned hypercubes $\mathcal{R}_i, 1 \le i \le M$ such that if $x, y \in \mathcal{R}_i$ then $|F_Z(x) - F_Z(y)| < \epsilon$. Let $x_i, y_i \in \mathcal{R}_i$ be the lower left and upper right endpoints of $\mathcal{R}_i$.

Let N be chosen large enough so that if $n > N$, then $|F_{X_n}(z) - F_Z(z)| < \epsilon$ for all $z = x_i, y_i, 1 \le i \le M$. Now, if $t \in \mathcal{R}$, then $t \in \mathcal{R}_i$ for some i. If $n > N$ then

$$\|F_{X_n}(y_i) - F_{X_n}(x_i)\| \le \|F_{X_n}(y_i) - F_Z(y_i)\| + \|F_Z(y_i) - F_Z(x_i)\| + \|F_Z(x_i) - F_{X_n}(x_i)\| < 3\epsilon.$$

Observe that $F_{X_n}(t_i) \in [F_{X_n}(x_i), F_{X_n}(y_i)]$ and $F_Z(t_i) \in [F_Z(x_i), F_Z(y_i)]$. Therefore,

$$\|F_{X_n}(t_i) - F_{X_n}(x_i)\| \le \|F_{X_n}(y_i) - F_{X_n}(x_i)\| < 3\epsilon, \quad \|F_Z(t_i) - F_Z(x_i)\| \le \|F_Z(y_i) - F_Z(x_i)\| < \epsilon$$

Therefore we have

$$\|F_{X_n}(t_i) - F_Z(t_i)\| \le \|F_{X_n}(t_i) - F_{X_n}(x_i)\| + \|F_{X_n}(x_i) - F_Z(x_i)\| + \|F_Z(x_i) - F_Z(t_i)\| < 5\epsilon$$

whenever $n > N$. Thus, $F_{X_n} \to F_Z$ uniformly over $\mathcal{R}$. Since any compact set is contained in a large enough compact axis-aligned hypercube, the proof is complete.

Part (b) directly follows from a particular kind of right-continuity that multivariate CDFs possess. Indeed, if $x_n \in \mathbb{R}^d$ converges to x and x_n is coordinatewise bigger than x for all n, then $F(x_n) \to F(x)$ for any multivariate CDF F (see (Kallenberg, 2002, Theorem 4.25)). It follows from right-continuity and monotonicity of CDFs that their discontinuities form a set of measure zero. That is, given any two random vectors Z_1, Z_2, their CDFs F_{Z_1}, F_{Z_2} are simultaneously continuous outside a set of measure zero.

Thus, for any $c \in \mathbb{R}^d$ we can find points $c_n \to c$ such that F_{Z_1}, F_{Z_2} are both continuous at c_n for all n, and c_n is coordinatewise greater than c. If we assume that $F_{Z_1}(c) = F_{Z_2}(c)$ at all points of continuity, then we have $F_{Z_1}(c_n) = F_{Z_2}(c_n)$ and hence by right continuity we get $F_{Z_1}(c) = F_{Z_2}(c)$, as desired. □

C.2 A key lemma in establishing accuracy in the regime $\frac{M_{\theta_n}}{n} \to 0$

In this section we will outline the key proposition which will be used to prove part(a) of Theorems 6, 10 and 13. As explained prior to the statement of Theorem 6, the idea is that for θ_n small, the numerators of equation 1 and equation 2 are asymptotically equivalent, likewise the denominators.

We make this precise in the following proposition. Recall M_θ from equation 3. We assume that $g(x) = x$ throughout this section, since Lemmas 3 and 7 allow us to replace X with $g(X)$ in all arguments that follow.

Proposition 4. *Let X be a random vector, and suppose θ_n be a sequence of vectors such that $\frac{M_{\theta_n}}{n} \to 0$. Let B_n be a sequence of measurable sets such that*

$$\limsup_{n\to\infty} \frac{\mathbb{E}[e^{\theta_n^T X}]}{\mathbb{E}[e^{\theta_n^T X}\mathbf{1}_{X\in B_n}]} < \infty.$$

Then,

$$\frac{\mathbb{P}(R_{n,\theta_n} \in B_n)}{\mathbb{P}(X_{\theta_n} \in B_n)} \to 1$$

in probability, as $n \to \infty$.

Proof. Observe that by equation 1 and equation 2,

$$\frac{\mathbb{P}(R_{n,\theta_n} \in B_n)}{\mathbb{P}(X_{\theta_n} \in B_n)} = \frac{\frac{1}{n}\sum_{i=1}^n e^{c\theta^T X_i}\mathbf{1}_{X_i\in B_n}}{\mathbb{E}[e^{c\theta^T X}\mathbf{1}_{X\in B_n}]} \frac{\mathbb{E}[e^{c\theta^T X}]}{\frac{1}{n}\sum_{i=1}^n e^{c\theta^T X_i}}. \tag{50}$$

We will prove that each ratio converges to 1 in probability, following which their product will also do so. Fix $\epsilon > 0$ and $n \ge 1$. In the following argument, let $A_n = B_n$ for all $n \ge 1$ or $A_n = \mathbb{R}^d$ for all $n \ge 1$. By Chebyshev's inequality ,

$$\begin{aligned}
&\mathbb{P}\left[\left|\frac{\frac{1}{n}\sum_{i=1}^n e^{\theta_n^T X_i}\mathbf{1}_{X_i\in A_n}}{\mathbb{E}[e^{\theta_n^T X}\mathbf{1}_{X\in A_n}]}-1\right|>\epsilon\right]\\
=&\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^n e^{\theta_n^T X_i}\mathbf{1}_{X_i\in A_n}-\mathbb{E}[e^{\theta_n^T X}\mathbf{1}_{X\in A_n}]\right|>\epsilon\mathbb{E}[e^{\theta_n^T X}\mathbf{1}_{X\in A_n}]\right]\\
\leq&\frac{\mathbb{V}\mathrm{ar}\left(e^{\theta_n^T X}\mathbf{1}_{X\in A_n}\right)}{n\epsilon^2\mathbb{E}\left[e^{\theta_n^T X}\mathbf{1}_{X\in A_n}\right]^2}\\
\leq&\frac{\mathbb{V}\mathrm{ar}\left(e^{\theta_n^T X}\mathbf{1}_{X\in A_n}\right)}{n\epsilon^2\mathbb{E}\left[e^{\theta_n^T X}\mathbf{1}_{X\in A_n}\right]^2}\\
\leq&\frac{\mathbb{E}\left[e^{2\theta_n^T X}\right]}{n\epsilon^2\mathbb{E}\left[e^{\theta_n^T X}\mathbf{1}_{X\in A_n}\right]^2}\\
\leq&\frac{1}{\epsilon^2}\frac{\mathbb{E}[e^{2\theta_n^T X}]}{n\mathbb{E}[e^{\theta_n^T X}]^2}\frac{\mathbb{E}[e^{\theta_n^T X}]^2}{\mathbb{E}[e^{\theta_n^T X}\mathbf{1}_{X\in A_n}]^2}=\frac{1}{\epsilon^2}\frac{M_{\theta_n}}{n}\frac{\mathbb{E}[e^{\theta_n^T X}]^2}{\mathbb{E}[e^{\theta_n^T X}\mathbf{1}_{X\in A_n}]^2}.
\end{aligned}$$

By our assumptions, it follows that this quantity converges to 0 as $n\to\infty$ for all $\epsilon>0$, which implies that

$$\frac{\frac{1}{n}\sum_{i=1}^n e^{\theta_n^T X_i}\mathbf{1}_{X_i\in A_n}}{\mathbb{E}[e^{\theta_n^T X}\mathbf{1}_{X\in A_n}]}\to 1$$

in probability. Applying this result for $A_n=B_n$ and then $A_n=\mathbb{R}^d$, and multiplying the resulting statements together proves that the terms in equation 50 converge to 1 in probability, as desired. $\square$

C.3 Point Processes, Poisson random measures and the regimes $\frac{M_{\theta_n}}{n}\not\to 0$

Next, we introduce point processes and the Poisson random measure, which will be used to explicitly find the limiting random variables in part (b) in each of Theorems 6, 10 and 13. It turns out that they will be essential to the proofs in part (c) of these theorems as well. We will establish some key lemmas which will be used to prove these results together, but the proofs of the theorems themselves will be relegated to future sections. We take our material from (Resnick, 1987, Chapter 3).

Let δ_x denote the Dirac measure $\delta_x(A)=\mathbf{1}_{x\in A}$. A point measure M on $E\subset\mathbb{R}^d, d\geq 1$ is a measure of the form $M=\sum_i\delta_{x_i}$ where $\{x_i\}\subset E$ is an at-most countable collection of points (which can also be finite e.g. the Dirac measure, and can also contain the same point multiple times e.g. twice the Dirac measure). Let $\mathcal{M}_p(E)$ denote the set of all point measures on E. It is equipped with the coarsest topology such that for all Borel sets $A\subset E$, the evaluation map $e_A:\mathcal{M}_p(E)\to[0,\infty]$ given by $e_A(M)=M(A)$ is measurable. It turns out that this topology coincides with vague convergence of measures, restricted to $\mathcal{M}_p(E)$, which we recall from the discussion before Assumption 4.

Any random element taking values in $\mathcal{M}_p(E)$ is called a point process on E. It can thus be thought of as a random collection of points on E.

Definition 1. *Let $E\subset\mathbb{R}^d$ and ν be a Radon measure on E. A point process M taking values on E is called a Poisson random measure, or PRM with intensity measure μ if*

(a) For all $A\subset E$, $M(A)$ is distributed as a Poisson random variable with parameter $\nu(A)$. (Note : If $\nu(A)=0$ then $M(A)=0$ and if $\nu(A)=\infty$ then $M(A)=\infty$ a.s.)

(b) For $A_1,A_2,\ldots,A_n\subset E$ mutually disjoint, $\{M(A_i)\}_{i=1,2,\ldots,n}$ forms an independent collection of random variables.

By (Resnick, 1987, Proposition 3.6), a PRM on any Borel subset of $\mathbb{R}^d$ with any intensity measure ν that is Radon, always exists and is unique up to distribution. We denote a PRM with intensity measure ν by $PRM(\nu)$, and integration with respect to $PRM(\nu)$ will be denoted by $dPRM(\nu)$ with the integration variable being omitted (e.g. $\int f(x)dPRM(\nu)$).

The convergence of a sequence of point processes $M_n \in \mathcal{M}_p(E)$ to a point process $M \in \mathcal{M}_p(E)$ will be considered in the weak sense here, see (Resnick, 1987, Section 3.5) for more details. We use $\Rightarrow$ to denote this convergence. Note that this is not an abuse of notation; the same notion of convergence was used for Gaussian process limits as well earlier.

The following results will now be used in the proofs of various point-process related convergences in our work. We begin with the following vague convergence result.

Lemma 13. *Let X_i be iid random vectors taking values in a Borel $E \subset \mathbb{R}^d, d \geq 1$ with the same distribution X. Suppose that for some sequence of vectors a_n and scalars b_n we have*

$$n\mathbb{P}\left(\frac{X - a_n}{b_n} \in \cdot\right) \to \nu\left(\cdot\right)$$

vaguely (see Assumption 4 for the definition of vague convergence) for some Radon measure ν on E. Then,

$$\sum_{k=1}^{n} \delta_{(X_k - a_n)/b_n} \Rightarrow PRM(\nu)$$

on $\mathcal{M}_p(E)$.

Proof. This is a simple application of the "warm-up" exercise in the proof of (Resnick, 1987, Proposition 3.21) with $X_{n,j} = \frac{X_j - a_n}{b_n}, n, j \geq 1$. □

We will now setup a common lemma which converts point process convergence to convergence in distribution. Since this may require case-by-case analysis, we set the tone.

Lemma 14. *Let X be a random vector in $\mathbb{R}^d, d \geq 1$ and $\theta \in \mathbb{R}^d$ be a unit vector. Let c_n be a sequence of real numbers such that if $\theta_n = c_n\theta$, then $X_{\theta_n} \to x$ in probability as $n \to \infty$, for some $x \in \mathbb{R}^d$. Suppose that, for some real sequence a_n, bounded Borel set $D \in \mathbb{R}^d$ and Radon measure ν on $\mathbb{R}^d$, all the following conditions hold.*

(a) $c_n a_n \to C_1 \in (0, \infty)$, and $c_n \to \infty$. (Hence, $a_n\theta_n \to C_1\theta$)

(b) $n\mathbb{P}\left(\frac{(x-X)}{a_n} \in \cdot\right) \to \nu(\cdot)$ vaguely.

(c) $\frac{x - A_n}{a_n} \to D$ as sets, as $n \to \infty$.

Then,

$$\left(\sum_{i=1}^{n} e^{-\theta_n^T(x - X_i)}\mathbf{1}_{X_i \in A_n}, \sum_{i=1}^{n} e^{-\theta_n^T(x - X_i)}\right) \xrightarrow{d} \left(\int e^{-C\theta^T y}\mathbf{1}_{y \in D} dPRM(\nu), \int e^{-C\theta^T y}\mathbf{1}_{y \in D} dPRM(\nu)\right).$$

Proof. We will show that this convergence holds using the Laplace transform (see (Kallenberg, 2002, Theorem 6.3)). In order to do this, let $s_1, s_2 < 0$ be arbitrary and consider the quantity

$$\mathbb{E}\left[\exp\left(s_1\sum_{i=1}^{n} e^{-\theta_n^T(x - X_i)}1_{A_n} + s_2\sum_{i=1}^{n} e^{-\theta_n^T(x - X_i)}\right)\right]$$

Begin from the left hand side :

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(s_1\sum_{i=1}^n e^{-\theta_n^T(x-X_i)}1_{A_n}+s_2\sum_{i=1}^n e^{-\theta_n^T(x-X_i)}\right)\right]\\
=&\mathbb{E}\left[\exp\left(\sum_{i=1}^n\left(s_1e^{-\theta_n^T(x-X_i)}1_{A_n}+s_2e^{-\theta_n^T(x-X_i)}\right)\right)\right]\\
=&\mathbb{E}\left[\exp\left(s_1e^{-\theta_n^T(x-X_1)}1_{A_n}+s_2e^{-\theta_n^T(x-X_1)}\right)\right]^n\\
=&\left(1-\mathbb{E}\left[1-\exp\left(s_1e^{-\theta_n^T(x-X_1)}1_{A_n}+s_2e^{-\theta_n^T(x-X_1)}\right)\right]\right)^n.
\end{aligned}$$

The term inside the expectation goes to zero as $n\to\infty$, since $\|\theta_n\|=c_n\to\infty$ by assumption (a). Therefore,

$$\lim_{n\to\infty}\frac{\left(1-\mathbb{E}\left[1-\exp\left(s_1e^{-\theta_n^T(x-X_1)}1_{A_n}+s_2e^{-\theta_n^T(x-X_1)}\right)\right]\right)^n}{\exp\left(-n\mathbb{E}\left[1-\exp\left(s_1e^{-\theta_n^T(x-X_1)}1_{A_n}+s_2e^{-\theta_n^T(x-X_1)}\right)\right]\right)}=1. \tag{51}$$

It suffices, therefore, to focus on the limit of

$$n\mathbb{E}\left[1-\exp\left(s_1e^{-\theta_n^T(x-X_1)}1_{A_n}+s_2e^{-\theta_n^T(x-X_1)}\right)\right].$$

Let $R>0$ be arbitrary, and

$$Y_n=\frac{(x-X_1)}{a_n}$$

for notational convenience. By assumption (b), we have $n\mathbb{P}(Y_n\in\cdot)\to\nu(\cdot)$ vaguely. Now,

$$\begin{aligned}
&n\mathbb{E}\left[1-\exp\left(s_1e^{-\theta_n^T(x-X_1)}1_{A_n}+s_2e^{-\theta_n^T(x-X_1)}\right)\right]\\
=&n\mathbb{E}\left[1-\exp\left(s_1e^{-a_n\theta_n^Ty}1_{y\in\frac{x-A_n}{a_n}}+s_2e^{-a_n\theta_n^Ty}\right)\right]
\end{aligned}\tag{52}$$

Let $R>0$ be an arbitrary real number. Note that

$$\begin{aligned}
&n\mathbb{E}\left[1-\exp\left(s_1e^{-a_n\theta_n^Ty}1_{y\in\frac{x-A_n}{a_n}}+s_2e^{-a_n\theta_n^Ty}\right)\right]\\
\geq&n\int_{C_1\theta^Ty\leq R,\|y\|\leq R}\left[1-\exp\left(s_1e^{-a_n\theta_n^Ty}1_{y\in\frac{x-A_n}{a_n}}+s_2e^{-a_n\theta_n^Ty}\right)\right]d\mathbb{P}_{Y_n}(y)
\end{aligned}\tag{53}$$

For R fixed, this term converges due to vague convergence of Y_n, and the fact that we are restricted to a compact set (a small approximation argument is required, whose basic idea is derived from the end of the proof of (Resnick, 1987, Proposition 3.21).)

$$\begin{aligned}
&n\int_{C_1\theta^Ty\leq R}\left[1-\exp\left(s_1e^{-a_n\theta_n^Ty}1_{y\in\frac{x-A_n}{a_n}}+s_2e^{-a_n\theta_n^Ty}\right)\right]d\mathbb{P}_{Y_n}(y)\\
&\qquad\to\int_{C_1\theta^Ty\leq R,\|y\|\leq R}\left[1-\exp\left(s_1e^{-C_1\theta^Ty}1_{y\in D}+s_2e^{-C_1\theta^Ty}\right)\right]d\nu(y).
\end{aligned}\tag{54}$$

Combining this statement and equation 53 we have as $R\to\infty$ that

$$\liminf_{n\to\infty}n\mathbb{E}\left[1-\exp\left(s_1e^{-a_n\theta_n^Ty}1_{y\in\frac{x-A_n}{a_n}}+s_2e^{-a_n\theta_n^Ty}\right)\right]\geq\int\left[1-\exp\left(s_1e^{-C_1\theta^Ty}1_{y\in D}+s_2e^{-C_1\theta^Ty}\right)\right]d\nu(y). \tag{55}$$

On the other hand, suppose that R_n is any sequence of real numbers increasing to $+\infty$. We have

$$\begin{aligned}
&n\mathbb{E}\left[1-\exp\left(s_1e^{-a_n\theta_n^Ty}1_{y\in\frac{x-A_n}{a_n}}+s_2e^{-a_n\theta_n^Ty}\right)\right]\\
=&\int_{\theta_n^Ty\leq R_n}\left[1-\exp\left(s_1e^{-a_n\theta_n^Ty}1_{y\in\frac{x-A_n}{a_n}}+s_2e^{-a_n\theta_n^Ty}\right)\right]d\nu(y)+\int_{\theta_n^Ty>R_n}\left[1-\exp\left(s_1e^{-a_n\theta_n^Ty}1_{y\in\frac{x-A_n}{a_n}}+s_2e^{-a_n\theta_n^Ty}\right)\right]d\nu(y)\\
\leq&\int\left[1-\exp\left(s_1e^{-a_n\theta_n^Ty}1_{y\in\frac{x-A_n}{a_n}}+s_2e^{-a_n\theta_n^Ty}\right)\right]d\nu(y)+n\mathbb{P}(\theta_n^Ty>R_n)(-s_1-s_2)e^{-a_nR_n},
\end{aligned}$$

where we used the inequality

$$1-\exp\left(s_1 e^{-a_n\theta_n^T y}1_{y\in\frac{x-A_n}{a_n}}+s_2 e^{-a_n\theta_n^T y}\right)\leq -\left(s_1 e^{-a_n\theta_n^T y}1_{y\in\frac{x-A_n}{a_n}}+s_2 e^{-a_n\theta_n^T y}\right)\leq -(s_1+s_2)e^{-a_n\theta_n^T y}$$

Let R_n be a sequence chosen so large that $n\mathbb{P}(\theta_n^T Y>R_n)\to 0$. Then, it follows by letting $n\to\infty$ above that

$$\limsup_{n\to\infty} n\mathbb{E}\left[1-\exp\left(s_1 e^{-a_n\theta_n^T y}1_{y\in\frac{x-A_n}{a_n}}+s_2 e^{-a_n\theta_n^T y}\right)\right]\geq \int\left[1-\exp\left(s_1 e^{-C_1\theta^T y}1_{y\in D}+s_2 e^{-C_1\theta^T y}\right)\right]d\nu(y).$$

Combining this with the lower bound equation 55 we have

$$\lim_{n\to\infty} n\mathbb{E}\left[1-\exp\left(s_1 e^{-a_n\theta_n^T y}1_{y\in\frac{x-A_n}{a_n}}+s_2 e^{-a_n\theta_n^T y}\right)\right]= \int\left[1-\exp\left(s_1 e^{-C_1\theta^T y}1_{y\in D}+s_2 e^{-C_1\theta^T y}\right)\right]d\nu(y). \tag{56}$$

It now suffices to look at the Laplace transform at $(-s_1,-s_2)$ of the left hand side of the lemma. That is,

$$\mathbb{E}\left[\exp\left(s_1\int e^{-C_1\theta^T y}1_{y\in D}dPRM(\nu)+s_2\int e^{-C_1\theta^T y}dPRM(\nu)\right)\right]=\mathbb{E}\left[\exp\left(\left(\int(s_1 e^{-C_1\theta^T y}1_{y\in D}+s_2 e^{-C_1\theta^T y})dPRM(\nu)\right)\right]\right].$$

However, this is precisely the Laplace functional of $PRM(\nu)$ evaluated at the function $f(s)=-(s_1e^{-C_1 y}1_{y\in D}+s_2e^{-C_1 y})$ (see (Resnick, 1987, Section 3.2) for the definition). This admits an explicit formula for $PRM(\nu)$ by (Resnick, 1987, Proposition 3.6(ii)). Applying that formula,

$$\mathbb{E}\left[\exp\left(-\int f(s)dPRM(\nu)\right)\right]=\exp\left(-\int_0^\infty(1-e^{-f(s)})d\nu(s)\right)$$

which matches exactly with the right hand side of equation 56, once we use equation 51. Thus, the proof is complete. □

Having proved this important lemma, we shall now prove a result that will help us with part (c) of Theorems 6 and 10. We proceed to motivate the lemma, state it and then prove it. Suppose for ease of clarity that out of the iid samples $X_1,X_2,\ldots,X_n$ of X, that X_1 is the sample maximizer of the quantity $\theta^T x$, for some fixed unit vector θ. Consider the quantity

$$\frac{\sum_{i=1}^n e^{\theta_n^T X_i}}{e^{\theta_n^T X_1}}=\sum_{i=1}^n e^{\theta_n^T(X_i-X_1)}$$

Suppose that a particular functional of X converges to a PRM, for instance let $x\in R^d$ and a_n be a sequence such that $\sum_{i=1}^n\delta_{\frac{x-X_i}{a_n}}\Rightarrow PRM(\nu)$ for some Radon measure ν, and θ_n be a sequence of vectors such that $a_n\theta_n\to\theta$ for some vector θ. Observe that

$$\sum_{i=1}^n e^{\theta_n^T(X_i-X_1)}=\sum_{i=1}^n e^{a_n\theta_n^T\frac{(x-X_1)}{a_n}-a_n\theta_n^T\frac{(x-X_i)}{a_n}}$$

The term on the exponential is a random measure we have already studied, minus its very first term. Since $a_n\theta_n\to\theta$, we expect that the rest will converge to a PRM minus its first term, which is a phenomena general enough to warrant its own lemma. This is the lemma that we will now proceed to prove. A simple point process is one which, when evaluated at a singleton $\{a\}$ for any $a\in\mathbb{R}$, results in either 0 or 1.

Lemma 15. *Let η_n be a sequence of simple point processes on $\mathbb{R}^d$ such that η_n has n points. Suppose $\eta_n\Rightarrow\eta$. For any simple point process μ on $\mathbb{R}^d$ with finitely many points, let $s_\mu=\arg\min\theta^T y$ be the sample minimizer of $\theta^T y$ in μ. Suppose s_μ is unique and isolated a.s. in μ. then, $\eta_n-s_{\eta_n}\Rightarrow\eta-s_\eta$.*

Proof. We will show that the map $\mu\to\mu-s_\mu$ is continuous on the subspace of simple point processes. Then, this result follows directly from the continuous mapping theorem. In order to do this, we must first prove that $\mu\to s_\mu$ is a continuous real-valued function on the space of all μ with a unique minimizer of $\theta^T y$.

Let μ_n converge to μ vaguely. We will prove that s_{μ_n} converges to s_μ. Let $R_1 = \theta^T s_\mu$ and $R_2 = \|s_\mu\|$. Note that for any $r_1, r_2 > 0$, we have

$$\mu_n(\{\theta^T y < r_1, \|y\| \le r_2\}) \to \mu(\{\theta^T y < r_1, \|y\| \le r_2\})$$

Let $\epsilon > 0$ be chosen small enough so that $\theta^T s_\mu < \inf_{\mu \setminus s_\mu} \theta^T y - \epsilon$. Then, for any $\epsilon' > 0$,

$$\mu(\{\theta^T y < R - \epsilon, \|y\| \le R_2 - \epsilon'\}) = 0, \mu(\{\theta^T y < R + \epsilon, \|y\| \le R_2 + \epsilon'\}) = 1.$$

Thus, for large enough n we have by vague convergence that

$$\mu_n(\{\theta^T y < R - \epsilon, \|y\| \le R_2\}) = 0, \mu_n(\{\theta^T y < R + \epsilon, \|y\| \le R_2\}) \to 1,$$

which implies that $\theta^T s_{\mu_n} \in (R_1 - \epsilon, R_1 + \epsilon)$ and $\|s_{\mu_n}\| \in (R_2 - \epsilon, R_2 + \epsilon)$. By the choice of ϵ, it follows that $s_{\mu_n} = s_\mu$ for n large enough.

Now, suppose that A is a set such that $\partial A = 0$. We must prove that $(\mu_n - s_{\mu_n})(A) \to (\mu - s_\mu)(A)$. This is equivalent to

$$\mu_n(A + s_{\mu_n}) \to \mu(A + s_\mu).$$

For any $\epsilon > 0$, consider the open set

$$A^{-\epsilon} = \{x \in A + s_\mu : \|x - y\| < \epsilon \text{ for all } y \in \partial(A + s_\mu)\}$$

and the compact set

$$A^{+\epsilon} = \{x \in A + s_\mu : \|x - y\| \le \epsilon \text{ for some } y \in A + s_\mu\}$$

Then, $A^{-\epsilon} \subset A + s_\mu \subset A^{+\epsilon}$, and since $s_{\mu_n} \to s_\mu$, we have that $\mathbf{1}_{A+s_{\mu_n}} \to \mathbf{1}_{A+s_\mu}$ pointwise as functions on $\mathbb{R}^d$. In particular, this implies that for some $N \in \mathbb{N}$, $A^{-\epsilon} \subset A + s_{\mu_n} \subset A^{+\epsilon}$ for all $n > N$. Thus,

$$\mu_n(A^{-\epsilon}) \le \mu_n(A + s_{\mu_n}) \le \mu_n(A^{+\epsilon})$$

for all $n > N$.

However, by Proposition 3.12, Resnick, $\limsup_n \mu_n(A^{+\epsilon}) \le \mu(A^{+\epsilon})$ and $\liminf_n \mu_n(A^{-\epsilon}) \ge \mu(A^{-\epsilon})$. Thus, we have

$$\mu(A^{-\epsilon}) \le \liminf_n \mu_n(A^{-\epsilon}) \le \liminf_n \mu_n(A + s_{\mu_n}) \le \limsup_n \mu_n(A + s_{\mu_n}) \le \limsup_n \mu_n(A^{+\epsilon}) \le \mu(A^{+\epsilon}).$$

This is true for any $\epsilon > 0$. The result now follows by letting $\epsilon \to 0$.

Having shown that $\mu \to \mu - s_\mu$ is continuous on the space of simple point processes, we conclude by the continuous mapping theorem (see (Billingsley, 1999, Theorem 5.1)) since η is a simple point process a.s. $\square$

PRM HAS UNIQUE SMU PROPERTY : NEEDS PROOF.

We shall now proceed to prove the theorems in each of the remaining sections in their order of appearance. It will be apparent that the results in the forthcoming sections will utilize the preceding results heavily.

D Proofs for Section 4

In this section, we will prove all the results in Section 4. Let X be a random variable in the Weibull regime (see Assumption 1), with maximum value $\mathcal{M} > 0$ and tail index $\alpha > 0$.

D.1 Proof of Lemma 1 and Weibull regime examples

We begin by proving Lemma 1 and the examples of random variables in the Weibull regime.

Proof of Lemma 1. Recalling that $b_n > 0$, note that for any $n \geq 1$,

$$\begin{aligned}\sup_{x\in\mathbb{R}} |\mathbb{P}(X_{1,n} \leq x) - \mathbb{P}(X_{2,n} \leq x)| &= \sup_{x\in\mathbb{R}} \left|\mathbb{P}\left(\frac{X_{1,n}-a_n}{b_n} \leq \frac{x-a_n}{b_n}\right) - \mathbb{P}\left(\frac{X_{2,n}-a_n}{b_n} \leq \frac{x-a_n}{b_n}\right)\right| \\ &= \sup_{y\in\mathbb{R}} \left|\mathbb{P}\left(\frac{X_{1,n}-a_n}{b_n} \leq y\right) - \mathbb{P}\left(\frac{X_{2,n}-a_n}{b_n} \leq y\right)\right| \end{aligned} \tag{57}$$

since the map $x \mapsto y = \frac{x-a_n}{b_n}$ is a bijective mapping on $\mathbb{R}$. Now recalling the definition equation 6 of KS we have

$$\mathsf{KS}(X_{1,n}, X_{2,n}) = \sup_x \left|\mathbb{P}\left(\frac{X_{1,n}-a_n}{b_n} \leq x\right) - \mathbb{P}\left(\frac{X_{2,n}-a_n}{b_n} \leq x\right)\right| = \mathsf{KS}\left(\frac{X_{1,n}-a_n}{b_n}, \frac{X_{2,n}-a_n}{b_n}\right)$$

For part (a), since $Z_1 = Z_2 = Z$ is continuous, by Lemma 12 we have $\mathsf{KS}\left(\frac{X_{i,n}-a_n}{b_n}, Z\right) \to 0$. The result now follows from the above equality and the triangle inequality for the KS distance.

For part (b), since $Z_1 \neq Z_2$ in distribution, by Lemma 12 there is a point c at which both the CDF of Z_1 and the CDF of Z_2 are continuous, such that $F_{Z_1}(c) \neq F_{Z_2}(c)$. But $\mathbb{P}\left(\frac{X_{i,n}-a_n}{b_n} \leq c\right) \to F_{Z_i}(c)$ by convergence in distribution. That the KS distance between the scaled random variables doesn't go to zero, now follows directly from this observation and equation 57. ☐

Next, we will sketch the correctness of the three examples provided for the Weibull regime assumption. (See 4 for details on these examples)

Proof of Weibull examples. Let $X = Beta(a,b)$ for some $a.b > 0$. Then, X has the density $cx^{a-1}(1-x)^{\mathcal{M}-1}$ for $x \in [0,1]$ and a normalizing constant c. Clearly its maximum value is $\mathcal{M} = 1$, and by the chain rule, the function $1 - F_X(1-x)$ has the derivative $c(1-x)^{a-1}x^{\mathcal{M}-1}$.

Note that $c(1-x)^{a-1}$ is a function which is continuous and non-zero at 0, hence regularly varying with index 0 at 0. On the other hand $x^{\mathcal{M}-1}$ is regularly varying of order $\mathcal{M}-1$. By the product rule (Bingham et al., 1987, Proposition 1.5.7(iv)) it is readily seen that $c(1-x)^{a-1}x^{\mathcal{M}-1}$ is regularly varying of order $\mathcal{M}-1$ at 0.

By (Bingham et al., 1987, Proposition 1.5.8), its integral i.e. $1 - F_X(1-x)$ is regularly varying at 0 of index $b > 0$. Hence, it lies in the Weibull regime with parameter b, completing the proof of the first example.

Coming to the second example, suppose that X is either a normal or exponential random variable, and let $M > 0$. It is easily checked that if f is the density of the truncated random variable $X_M = X\mathbf{1}_{X\leq M}$, then $\lim_{x\to M^-} f(x) = f(M)$ is a positive number. Therefore, it is clear that f is regularly varying at M or order 0. By an argument similar to the one made for $Beta(a,b)$, it follows that X is in the Weibull regime with $1 - F_X(M-x)$ regularly varying at 0 with index 1.

We shall now address the third example in detail. Let X, Y be in the Weibull regime with parameters α, β respectively and maximum values $\mathcal{M}_X, \mathcal{M}_Y$. Then, $X + Y$ has maximum value $\mathcal{M}_X + \mathcal{M}_Y$. Let F_X be the CDF of X and F_Y the CDF of Y. Note that for any $u > 0$,

$$\begin{aligned}\mathbb{P}(X+Y > (\mathcal{M}_X + \mathcal{M}_Y - u)) &= \int_{\mathcal{M}_X-u}^{\mathcal{M}_X} (1 - F_Y(\mathcal{M}_X + \mathcal{M}_Y - u - x))dF_X(x) \\ &= \int_0^u (1 - F_Y(\mathcal{M}_Y - u + x))d(1 - F_X(\mathcal{M}_X - x))) \\ &= \int_0^u (1 - F_X(\mathcal{M}_X - x))d(1 - F_Y(\mathcal{M}_Y - u + x)) \end{aligned} \tag{58}$$

where we used integration by parts and noted that $\lim_{x\to 0^+} 1 - F_X(\mathcal{M}_X - x) = 0$ and $\lim_{x\to u^-} 1 - F_Y(\mathcal{M}_Y - u + x) = 0$.

Similarly, for any $t, u > 0$,

$$\begin{aligned}\mathbb{P}(X+Y > (\mathcal{M}_X + \mathcal{M}_Y - tu)) &= \int_0^{tu} (1 - F_Y(\mathcal{M}_Y - tu + x)) d(1 - F_X(\mathcal{M}_X - x))) \\ &= t \int_0^{u} (1 - F_Y(\mathcal{M}_Y - tu + tx)) d(1 - F_X(\mathcal{M}_X - tx))) \\ &= t \int_0^{u} (1 - F_X(\mathcal{M} - tx)) d(1 - F_Y(\mathcal{M}_Y - tu + tx)),\end{aligned} \tag{59}$$

where we used integration by parts and noted that $\lim_{x \to 0^+} 1 - F_X(\mathcal{M}_X - tx) = 0$ and $\lim_{x \to u^-} 1 - F_X(\mathcal{M}_X - tu + tx) = 0$.

The representations so far clearly hint at the appearance of regularly varying terms in the integrand and integrator of equation 58 and equation 59. Thus, we expect the ratio of these two terms to behave regularly, which is what will be proved now. We shall now fix $t > 1$ and study the ratio of the left hand sides of equation 58 and equation 59 as $u \to 0$. An analogous argument will apply for $t < 1$.

Let $A > 1, \delta > 0$ be arbitrary. By the definition in equation 1 of the Weibull regime and Potter's theorem (Bingham et al., 1987, Theorem 1.5.6(iii)), there exists $X > 0$ such that for $p, q \leq X$,

$$A^{-1} q^{\alpha - \delta} \leq \frac{(1 - F_Y(\mathcal{M}_Y - qp))}{(1 - F_Y(\mathcal{M}_Y - p))} \leq A q^{\alpha + \delta} \tag{60}$$

$$A^{-1} q^{\beta - \delta} \leq \frac{(1 - F_X(\mathcal{M}_X - qp))}{(1 - F_X(\mathcal{M}_X - p))} \leq A q^{\beta + \delta} \tag{61}$$

Now, applying equation 60 to every $u < X/t$ and integrating these inequalities from 0 to u with respect to $1 - F_X(\mathcal{M}_X - x)$,

$$A^{-1} t^{\alpha - \delta} \leq \frac{\int_0^u (1 - F_Y(\mathcal{M}_Y - tu + tx)) d(1 - F_X(\mathcal{M} - x))}{\int_0^u (1 - F_Y(\mathcal{M}_Y - u + x)) d(1 - F_X(\mathcal{M} - x))} \leq A t^{\alpha + \delta} \tag{62}$$

Similarly, applying equation 61 to every $u < X/t$ and integrating these inequalities from 0 to u with respect to $1 - F_Y(\mathcal{M}_Y - tu + tx)$,

$$A^{-1} t^{\beta - \delta} \leq \frac{\int_0^u (1 - F_X(\mathcal{M} - tx)) d(1 - F_Y(\mathcal{M}_Y - tu + tx))}{\int_0^u (1 - F_X(\mathcal{M} - x)) d(1 - F_Y(\mathcal{M}_Y - tu + tx))} \leq A t^{\beta + \delta} \tag{63}$$

Multiplying equation 62 and equation 63, while also noting the equalities in equation 58 and equation 59,

$$A^{-2} t^{\beta + \alpha - 2\delta} \leq \frac{\mathbb{P}(X+Y > (\mathcal{M}_X + \mathcal{M}_Y - tu))}{\mathbb{P}(X+Y > (\mathcal{M}_X + \mathcal{M}_Y - u))} \leq A^2 t^{\beta + \alpha + 2\delta}$$

for all $u, t > 0$ such that $tu < X$. Now, fixing t and letting $u \to 0$ above gives that

$$A^{-2} t^{\beta + \alpha - 2\delta} \leq \liminf_{u \to 0} \frac{\mathbb{P}(X+Y > (\mathcal{M}_X + \mathcal{M}_Y - tu))}{\mathbb{P}(X+Y > (\mathcal{M}_X + \mathcal{M}_Y - u))} \leq \limsup_{u \to 0} \frac{\mathbb{P}(X+Y > (\mathcal{M}_X + \mathcal{M}_Y - tu))}{\mathbb{P}(X+Y > (\mathcal{M}_X + \mathcal{M}_Y - u))} \leq A^2 t^{\beta + \alpha + 2\delta}.$$

Since this is true for all $\delta > 0$ and $A > 1$, letting $A \downarrow 1$ and $\delta \downarrow 0$ gives

$$\lim_{u \to 0} \frac{\mathbb{P}(X+Y > (\mathcal{M}_X + \mathcal{M}_Y - tu))}{\mathbb{P}(X+Y > (\mathcal{M}_X + \mathcal{M}_Y - u))} = t^{\alpha + \beta}.$$

The same argument as here can be repeated with $t < 1$, showing that the above statement holds for any $t > 0$ (for $t = 1$ it is immediate). Consequently, the result has been proved. □

We will next prove Lemma 3 i.e. that the site-specific tilt g retains the Weibull assumption.

Proof of Lemma 3. Since g is strictly increasing, the maximum value of $g(X)$ is $g(\mathcal{M})$. Let F_X be the CDF of X, and $F_{g(X)}$ be the CDF of $g(X)$. Consider the quantity

$$\begin{aligned}1 - F_{g(X)}(g(\mathcal{M}) - x) &= \mathbb{P}(g(X) > g(\mathcal{M}) - x)\\ &= \mathbb{P}(X > g^{-1}(g(\mathcal{M}) - x))\\ &= 1 - F_X(g^{-1}(g(\mathcal{M}) - x)).\end{aligned}$$

Here, g^{-1} is the set theoretic inverse of g. Note that since g is strictly increasing, $g^{-1}(y) < x$ if and only if $y < g(x)$.

We must prove that $1 - F_{g(X)}(g(\mathcal{M}) - x)$ is regularly varying at 0. However, we have written this function as a composition of two functions

$$1 - F_{g(X)}(g(\mathcal{M}) - x) = h(j(x))$$

where $h(z) = 1 - F_X(\mathcal{M} - z)$ and $j(y) = \mathcal{M} - g^{-1}(g(\mathcal{M}) - y)$.

Observe that $j(y) \to 0$ as $y \to 0$ by the continuity of f. Furthermore, h is regularly varying at 0 with index α. If we can show that j is regularly varying at 0 with index $\frac{1}{\beta}$, then by (Bingham et al., 1987, Proposition 1.5.7), $x \mapsto 1 - F_{g(X)}(g(\mathcal{M}) - x)$ is regularly varying at 0 of order $\frac{\alpha}{\beta}$, completing the proof.

However, observe that the definition of $j(y)$ implies that $g(\mathcal{M}) - g(\mathcal{M} - j(y)) = y$. In other words, j is the set theoretic inverse of the function $x \mapsto g(\mathcal{M}) - g(\mathcal{M} - x)$. By (Bingham et al., 1987, Theorem 1.5.12), j is regularly varying at 0 of order $\frac{1}{\beta}$. $\square$

Thus, general twists are absorbed into our setup. We will therefore continue to assume that X is in the Weibull regime with index $\alpha > 0$ and maximum value B. Let $\theta_n, n \geq 1$ be a sequence, and recall M_{θ_n} from equation 3. Also recall X_{θ_n} from equation 4.

We will now prove Lemma 6.

Proof of Lemma 6. For part (a), suppose that $A \subset \mathbb{R}^d$ is compact and satisfies $\nu(A) = 0$. Then, by Assumption 4,

$$\begin{aligned}\nu(cA) &= \lim_{t\to 0} \frac{1}{U(t)} \mathbb{P}\left(\frac{x_\theta - X}{t} \in cA\right)\\ &= \lim_{t\to 0} \frac{1}{U(t)} \mathbb{P}\left(\frac{x_\theta - X}{ct} \in A\right)\\ &= \lim_{t\to 0} \frac{c^\alpha}{U(ct)} \mathbb{P}\left(\frac{x_\theta - X}{ct} \in A\right)\\ &= c^\alpha \nu(A).\end{aligned} \tag{64}$$

To prove part (b), let $\nu\{0 \leq \theta^T y \leq 1\} = C \in (0, \infty)$, and define the measure μ on $\mathcal{S} = \{\theta^T y = 1\}$ by

$$\mu(B) = \nu\left\{0 < \theta^T y < 1, \frac{y}{\theta^T y} \in B\right\}.$$

Observe that $\mu(\mathcal{S}) = C$. Now, observe that for any $c, x > 0$ and $B \subset \mathcal{S}$, by part (a) we have

$$\nu\left(\left\{\theta^T y \in (0, c], \frac{y}{\theta^T y} \in B\right\}\right) = c^\alpha\left(\left\{\theta^T y \in (0, 1], \frac{y}{\theta^T y} \in B\right\}\right) = c^\alpha \mu(B). \tag{65}$$

Since sets of the form

$$\left\{\theta^T y \in (0, c], \frac{y}{\theta^T y} \in B\right\}$$

generate $\{\theta^T y > 0\}$, it follows that ν is uniquely determined by the scaling property, and if $a, b > 0$ and $B \subset \mathcal{S}$ are arbitrary then

$$\nu\left(\left\{\theta^T y \in (a, b], \frac{y}{\theta^T y} \in B\right\}\right) = \nu\left(\left\{\theta^T y \in (0, b], \frac{y}{\theta^T y} \in B\right\}\right) - \nu\left(\left\{\theta^T y \in (0, c], \frac{y}{\theta^T y} \in B\right\}\right) = (b^\alpha - a^\alpha)\mu(B)$$

by equation 65.

To prove part(c), observe that

$$\int_{\theta^T y>0} e^{-\theta^T y} d\nu(y) < \sum_{n=1}^{\infty} e^{-n} \nu\left(\{n-1<\theta^T y \leq n\}\right) \leq C \sum_{n=1}^{\infty} e^{-n}(n^{\alpha}-(n-1)^{\alpha}) < \infty,$$

where we used part (b). □

We will now dedicate a section to the asymptotic behavior of M_{θ_n} and the proof of Theorem 4.

D.2 Proof of Theorem 4

We first prove Lemma 4 about the growth of $\mathbb{E}[e^{\theta_n X}]$ as $\theta_n \to \infty$. Throughout this section let F_X denote the CDF of X.

Proof of Lemma 4. Write

$$\begin{aligned}
\mathbb{E}[e^{\theta_n X}] &= \int_{-\infty}^{\mathcal{M}} e^{\theta_n y} dF_X(y) \\
&= e^{\theta_n \mathcal{M}} \int_{-\infty}^{\mathcal{M}} e^{\theta_n (y-\mathcal{M})} dF_X(y) \\
&= -e^{\theta_n \mathcal{M}} \int_{0}^{\infty} e^{-\theta_n u} dF_X(\mathcal{M}-u) \\
&= e^{-\theta_n \mathcal{M}} \int_{0}^{\infty} e^{-\theta_n u} dG_X(u),
\end{aligned} \tag{66}$$

where $G_X(u) = 1 - F_X(\mathcal{M}-u)$. Now, $G_X(u)$ is regularly varying at 0 of order $\alpha > 0$. By the Karamata Tauberian theorem (see (Feller, 1971, Theorem 3, Section XIII.5)) it follows that

$$\frac{\int_0^{\infty} e^{-\theta_n u} dG_X(u)}{\Gamma(\alpha+1)\left(1-F_X\left(\mathcal{M}-\frac{1}{\theta_n}\right)\right)} \to 1.$$

Our result directly follows from the above equation, equation 66 and algebraic rearrangements. □

The proof of Theorem 4 is a corollary of this.

Proof of Theorem 4. If $\mathcal{M}$ is the maximum value of X, then we have by the previous lemma that

$$\frac{\mathbb{E}[e^{2\theta_n X}]}{e^{2\theta_n \mathcal{M}}\left(1-F_X\left(\mathcal{M}-\frac{1}{2\theta_n}\right)\right)} \to \Gamma(1+\alpha)$$

$$\frac{e^{2\theta_n \mathcal{M}}\left(1-F_X\left(\mathcal{M}-\frac{1}{\theta_n}\right)\right)^2}{\mathbb{E}[e^{\theta_n X}]^2} \to \frac{1}{\Gamma(1+\alpha)}$$

Multiplying these together and noting that $\frac{\left(1-F_X\left(\mathcal{M}-\frac{1}{\theta_n}\right)\right)}{\left(1-F_X\left(\mathcal{M}-\frac{1}{2\theta_n}\right)\right)} \to 2^{-\alpha}$ by Assumption 1, we obtain the result. □

In the next section, we will prove the scaling limit of X_θ i.e. Theorem 5.

D.3 Proof of Theorem 5

The following lemma covers the asymptotics of rare event probabilities under large twists. We will use it to prove Theorem 5.

Lemma 16. *Let θ_n be any sequence converging to infinity, and r_n be any sequence increasing to a real number $\mathcal{M}$. Suppose $\theta_n(\mathcal{M} - r_n) \to C \in (0, \infty)$. Then, for any random variable X belonging to the Weibull regime with parameter $\alpha > 0$ and maximum value $\mathcal{M}$, we have*

$$\frac{\mathbb{E}[e^{\theta_n X}\mathbf{1}_{X>r_n}]}{\mathbb{E}[e^{\theta_n X}]} \to \frac{\int_0^C t^{\alpha-1}e^{-t}dt}{\Gamma(\alpha)} \in (0,1).$$

Proof. As usual, let F_X be the CDF of X. We begin by noting that

$$\mathbb{E}[e^{\theta_n X}\mathbf{1}_{X>r_n}] = \int_{r_n}^{\mathcal{M}} e^{\theta_n y} dF_X(y).$$

We perform the following manipulations, taking $G_X(u) = 1 - F_X(\mathcal{M} - u)$. Since this is regularly varying at 0 by Assumption 1, by (Bingham et al., 1987, Theorem 1.4.1) we have $G_X(u) = L(u)u^\alpha$ for some slowly varying function L (i.e. regularly varying with index 0) at zero.

$$\begin{aligned}
\int_{r_n}^{\mathcal{M}} e^{\theta_n y} dF_X(y) =& e^{\theta_n \mathcal{M}} \int_{r_n}^{\mathcal{M}} e^{\theta_n(y-\mathcal{M})} dF_X(y) \\
=& - e^{\theta_n \mathcal{M}} \int_0^{\mathcal{M}-r_n} e^{-\theta_n u} dF_X(\mathcal{M}-u) \\
=& e^{\theta_n \mathcal{M}} \int_0^{\mathcal{M}-r_n} e^{-\theta_n u} dG_X(u) \\
\overset{IBP}{=}& e^{\theta_n \mathcal{M}} \left[[e^{-\theta_n u} G_X(u)]_0^{\mathcal{M}-r_n} + \int_0^{\mathcal{M}-r_n} G_X(u)\theta_n e^{-\theta_n u} du \right] \\
=& e^{\theta_n \mathcal{M}} \left[e^{-\theta_n(\mathcal{M}-r_n)} G_X(\mathcal{M}-r_n) + \int_0^{\theta_n(\mathcal{M}-r_n)} L\left(\frac{t}{\theta_n}\right)\left(\frac{t}{\theta_n}\right)^\alpha e^{-t} dt \right] \\
=& e^{\theta_n \mathcal{M}} \theta_n^{-\alpha} L\left(\frac{1}{\theta_n}\right) \left[e^{-\theta_n(\mathcal{M}-r_n)} L(\mathcal{M}-r_n)(\mathcal{M}-r_n)^\alpha L^{-1}\left(\frac{1}{\theta_n}\right)\theta_n^\alpha + \int_0^{\theta_n(\mathcal{M}-r_n)} \frac{L(\frac{t}{\theta_n})}{L\left(\frac{1}{\theta_n}\right)} t^\alpha e^{-t} dt \right]
\end{aligned} \tag{67}$$

We used the fact that $G_X(\mathcal{M}) = 0$ above. We will now consider the limit of both quantities inside the bracket on the last line.

Clearly, the first term is easy to handle :

$$e^{-\theta_n(\mathcal{M}-r_n)} L(\mathcal{M}-r_n)(\mathcal{M}-r_n)^\alpha L^{-1}\left(\frac{1}{\theta_n}\right)\theta_n^\alpha \to e^{-C}C^\alpha, \tag{68}$$

since $\theta_n(\mathcal{M} - r_n) \to C$ and L is slowly varying.

For the second term in equation 67, note that $\theta_n(\mathcal{M} - r_n) \to C$. Thus, it is bounded, and we assume that $\theta_n(\mathcal{M} - r_n) < M$ for all $N > 0$. Pointwise, on $[0, M]$ we have $\frac{L(\frac{t}{\theta_n})}{L(\frac{1}{\theta_n})} \to 1$. Thus, we have

$$\frac{L(\frac{t}{\theta_n})}{L\left(\frac{1}{\theta_n}\right)} t^\alpha e^{-t} \mathbf{1}_{[0,\theta_n(\mathcal{M}-r_n)]} \to t^\alpha e^{-t} \mathbf{1}_{[0,C]}$$

as n converges to infinity, pointwise on $[0, M]$. To apply the dominated convergence theorem, we note that L is slowly varying at 0, and therefore $L(s)s^{\alpha/4} \to 0$ as $s \to 0$, and $L(s)s^{-\alpha/4} \to +\infty$ as $s \to 0$ ((Bingham et al., 1987, Proposition 1.5.1)). Therefore, there is a constant K such that $Ks^{\alpha/4} \leq L(s) \leq Ks^{-\alpha/4}$ for $s \in [0, M]$, which implies that

$$\frac{L(\frac{t}{\theta_n})}{L\left(\frac{1}{\theta_n}\right)} t^{\alpha} e^{-t} \mathbf{1}_{[0,\theta_n(\mathcal{M}-r_n)]} \leq K^2 t^{\alpha/2} e^{-t} \mathbf{1}_{[0,M]}$$

which is integrable. By the dominated convergence theorem,

$$\int_0^{\theta_n(\mathcal{M}-r_n)} \frac{L(\frac{t}{\theta_n})}{L\left(\frac{1}{\theta_n}\right)} t^{\alpha} e^{-t} dt \to \int_0^C t^{\alpha} e^{-t} dt.$$

Combining the above with equation 68, equation 67 and Lemma 4,

$$\frac{\mathbb{E}[e^{\theta_n Y} \mathbf{1}_{Y > r_n}]}{\mathbb{E}[e^{\theta_n Y}]} \to \frac{e^{-C} C^{\alpha} + \int_0^C t^{\alpha} e^{-t} dt}{\Gamma(\alpha + 1)} = \frac{\int_0^C t^{\alpha-1} e^{-t} dt}{\Gamma(\alpha)},$$

where the numerator simplifies by integration-by-parts, and the denominator satisfies $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$. This completes the proof. □

With this, one can complete the proof of Theorem 5.

Proof of Theorem 5. Let $C > 0$ be arbitrary. For a sequence θ_n increasing to infinity, let $s_n = \frac{C}{\theta_n}$. We have following some algebraic rearrangement,

$$\{\theta_n(\mathcal{M} - X_{\theta_n}) \leq C\} = \left\{X_{\theta_n} \geq \mathcal{M} - \frac{C}{\theta_n}\right\} = \{X_{\theta_n} \geq \mathcal{M} - s_n\}.$$

Now, by the definition of exponential tilting,

$$\mathbb{P}(X_{\theta_n} \geq \mathcal{M} - s_n) = \frac{\mathbb{E}[e^{\theta_n X} \mathbf{1}_{X \geq \mathcal{M} - s_n}]}{\mathbb{E}[e^{\theta_n X}]} = \frac{\mathbb{E}[e^{\theta_n X} \mathbf{1}_{X \geq \mathcal{M} - s_n}]}{\mathbb{E}[e^{\theta_n X}]}.$$

By Lemma 16 applied with X and $r_n = \mathcal{M} - s_n$,

$$\mathbb{P}(X_{\theta_n} \geq \mathcal{M} - s_n) \to \frac{\int_0^C t^{\alpha-1} e^{-t} dt}{\Gamma(\alpha)} = \mathbb{P}(\Gamma(\alpha, 1) \leq C).$$

It follows that

$$\mathbb{P}(\theta_n(\mathcal{M} - X_{\theta_n}) \leq C) \to \mathbb{P}(\Gamma(\alpha, 1) \leq C)$$

as $n \to \infty$. This is sufficient to show the theorem, since we only need to consider $C \in (0, \infty)$ for non-negative random variables to converge in distribution. □

Finally, recall the estimator random variable $R_{n,\theta}$ defined by

$$\mathbb{P}(R_{n,\theta} \in A) = \frac{\sum_{i=1}^n e^{\theta f(X_i)} \mathbf{1}_{X_i \in A}}{\sum_{i=1}^n e^{\theta f(X_i)}}.$$

We are now ready to start proving Theorem 6. We will prove its parts in the order in which they were stated.

D.4 Proof of Theorem 6

We begin with the proof of part (a). Note that the rough idea was already discussed above the statement of the theorem.

Proof of Theorem 6(a). We use Proposition 4 with θ_n as in the statement of the theorem. Let $C > 0$ be arbitrary, and let $r_n = \mathcal{M} - \frac{C}{\theta_n}$, with $B_n = (r_n, \infty)$. Then, by Lemma 16 we have that

$$\lim_{n\to\infty} \frac{\mathbb{E}[e^{\theta_n X}]}{\mathbb{E}[e^{\theta_n X}\mathbf{1}_{X\in B_n}]} < \infty.$$

By assumption, $\frac{M_{\theta_n}}{n} \to 0$. Thus, by Proposition 4,

$$\frac{\mathbb{P}(R_{n,\theta_n} \geq r_n)}{\mathbb{P}(X_{\theta_n} \geq r_n)} = \frac{\mathbb{P}(R_{n,\theta_n} \geq r_n)}{\frac{\mathbb{E}[e^{\theta_n(X_i-B)}\mathbf{1}_{X_i\geq r_n}]}{\mathbb{E}[e^{\theta_n(X_i-B)}]}} \to 1.$$

in probability. (Note that $\mathbb{P}(R_{n,\theta_n} \geq r_n)$ is a random variable, since it depends on the X_i). The asymptotics of the denominator follow directly from Lemma 16, and we obtain

$$\mathbb{P}(R_{n,\theta_n} \geq r_n) \to \mathbb{P}(\Gamma(\alpha, 1) \leq C).$$

Since this is true for any $C \in (0, \infty)$, the result follows exactly as in the proof of Theorem 5. $\square$

Next, we prove Theorem 6(b).

In this section, we explicitly find the limiting random variable in Theorem 6(b) using Lemma 14.

Proof of Theorem 6. (b) Suppose that $\frac{M_{\theta_n}}{n} \to C \in (0, \infty)$. Then, by Theorem 4 we know that $\theta_n(B - F^{-1}(1 - \frac{1}{n})) \to c_1$ for some $c_1 > 0$, and hence $\theta_n \to \infty$. Let $c_2 \in (0, \infty)$ and $A_n = [\mathcal{M} - c_2(\mathcal{M} - F^{-1}(1 - \frac{1}{n})), B]$.

Consider $\mathbb{P}(R_{n,\theta} \in A_n)$, and divide the top and bottom by $e^{\theta B}$ to obtain

$$\begin{aligned}
\mathbb{P}(R_{n,\theta} \in A_n) =& \frac{\sum_{i=1}^n e^{\theta X_i}\mathbf{1}_{X_i\in A_n}}{\sum_{i=1}^n e^{\theta X_i}} \\
=& \frac{\sum_{i=1}^n e^{\theta(X_i-B)}\mathbf{1}_{X_i\in A_n}}{\sum_{i=1}^n e^{\theta(X_i-B)}} \\
=& \Phi_n\left(\sum_{i=1}^n e^{\theta_n(\mathcal{M}-X_i)} 1_{A_n}, \sum_{i=1}^n e^{\theta_n(\mathcal{M}-X_i)}\right),
\end{aligned}$$

where

$$\Phi(y, z) = \frac{y}{z}.$$

At this point, we verify the hypotheses of Lemma 14. We take $\theta = 1$, $c_n = \theta_n$. Then, $X_{\theta_n} \to B$ in probability. Let

$$a_n = B - F^{-1}(1 - \frac{1}{n}), C_1 = c_1, \nu(dy) = \alpha y^{\alpha-1}dy \text{ and } D = [0, c_2].$$

Hypotheses (a) and (c) of the lemma are easily verified, while hypothesis (b) follows from the Weibull regime being an instance of multivariate regular variation (see the examples in Section 5).

Since Φ is a continuous mapping, and $\int e^{-c_1 y} dPRM(\nu) \neq 0$ with probability 1 by the definition of a PRM, by Lemma 14 it would follow that

$$\mathbb{P}(R_{n,\theta} \in A_n) = \Phi_n\left(\sum_{i=1}^n e^{\theta_n(\mathcal{M}-X_i)} 1_{A_n}, \sum_{i=1}^n e^{\theta_n(\mathcal{M}-X_i)}\right) \xrightarrow{d} \frac{\int e^{-c_1 y} 1_{y\leq c_2} dPRM(\nu)}{\int e^{-c_1 y} dPRM(\nu)}.$$

Finally, note that

$$\mathbb{P}(R_{n,\theta} \in A_n) = \mathbb{P}\left(\theta_n(\mathcal{M} - R_{n,\theta}) \leq c_2\theta_n\left(\mathcal{M} - F^{-1}\left(1 - \frac{1}{n}\right)\right)\right) \approx \mathbb{P}(\theta_n(\mathcal{M} - R_{n,\theta}) \leq c_1 c_2)$$

if n is large enough. Combining the two statements above, if Z is a random variable such that

$$\mathbb{P}(Z \leq c_2) = \frac{\int e^{-c_1 y} 1_{y \leq c_2} dPRM(\nu)}{\int e^{-c_1 y} dPRM(\nu)},$$

then

$$\mathbb{P}(\theta_n(\mathcal{M} - R_{n,\theta}) \leq c_1 c_2) \to \mathbb{P}(Z \leq c_2) = \mathbb{P}(c_1 Z \leq c_1 c_2).$$

Thus, the random variable $Z_{C,PRM} = c_1 Z$ is the desired limit. Note that this random variable depends upon the Poisson random measure, while the other scaling limit $\Gamma(\alpha, 1)$ does not. It follows that these two random variables are not the same, which concludes the proof.

We remark that the limiting random variable Z_{PRM} is continuous, but do not prove this here. □

Now we will prove the last part i.e. Theorem 6(c).

Proof of Theorem 6(c). We claim that in probability,

$$\frac{\sum_{i=1}^n e^{\theta_n X_i}}{e^{\theta_n \max_i X_i}} \to 1$$

if $\theta_n(B - F^{-1}(1 - \frac{1}{n})) \to +\infty$ (which, by Theorem 4 is implied by $\frac{M_{\theta_n}}{n} \to +\infty$). In particular, this implies that

$$\frac{\sum_{i=1}^n e^{\theta_n X_i} \mathbf{1}_{X_i \neq \max_j X_j}}{e^{\theta_n \max_i X_i}} \to 0. \tag{69}$$

By Lemma 13, $\sum_{i=1}^n \delta_{\frac{\mathcal{M} - X_i}{\mathcal{M} - F^{-1}(1 - \frac{1}{n})}} \Rightarrow PRM(\nu)$ where $d\nu(y) = \alpha y^{\alpha - 1}$. Now, note that the smallest point of the former point process is attained when $X_i = \max_j X_j$, which without loss of generality we assume is attained at X_1. Therefore, we get by Lemma 15 (with $\theta = 1$) that

$$\sum_{i=1}^n \delta_{\frac{X_1 - X_i}{\mathcal{M} - F^{-1}(1 - \frac{1}{n})}} \Rightarrow PRM(\nu) - s_{PRM(\nu)}.$$

Let $M > 0$ be an arbitrary but fixed parameter. Let $f(x) = e^{-Mx} \mathbf{1}_{x < M}$. Note that f is continuous and compactly supported on $[0, \infty)$. Thus, by the continuous mapping theorem (see (Resnick, 1987, Section 3.5)) we have

$$\sum_{i=1}^n e^{-M \frac{X_1 - X_i}{\mathcal{M} - F^{-1}(1 - \frac{1}{n})}} \xrightarrow{d} \int e^{-Mx} \mathbf{1}_{x < M} \left[PRM(\nu) - s_{PRM(\nu)} \right]. \tag{70}$$

Now, it is clear that for any $\epsilon > 0$,

$$\lim_{M \to \infty} \mathbb{P}\left[\sum_{i=1}^n e^{-\theta_n (X_1 - X_i)} \mathbf{1}_{X_1 - X_i < (\mathcal{M} - F^{-1}(1 - \frac{1}{n}) M)} > 1 + \epsilon \right] = \mathbb{P}\left[\sum_{i=1}^n e^{-\theta_n (X_i - X_1)} > 1 + \epsilon \right].$$

For any $\epsilon > 0$ and $M > 0$, employing equation 70 gives

$$\begin{aligned}
&\limsup_n \mathbb{P}\left[\sum_{i=1}^n e^{-\theta_n \left(\mathcal{M} - F^{-1}(1 - \frac{1}{n})\right)\left(\frac{X_1 - X_i}{B - F^{-1}(1 - \frac{1}{n})}\right)} \mathbf{1}_{X_1 - X_i < (\mathcal{M} - F^{-1}(1 - \frac{1}{n}) M)} > 1 + \epsilon \right] \\
\leq &\limsup_{n \to \infty} \mathbb{P}\left[\sum_{i=1}^n e^{-M \left(\frac{X_1 - X_i}{B - F^{-1}(1 - \frac{1}{n})}\right)} \mathbf{1}_{X_1 - X_i < (\mathcal{M} - F^{-1}(1 - \frac{1}{n}) M)} > 1 + \epsilon \right] \\
= &\mathbb{P}\left[\int e^{-Mx} \mathbf{1}_{x < M} (PRM(\nu) - s_{PRM(\nu)}) > 1 + \epsilon \right].
\end{aligned} \tag{71}$$

Note that $e^{-Mx}\mathbf{1}_{x<M}$ is dominated by e^{-x} for $M > 1$. Furthermore, as $M \to \infty$, we have $e^{-Mx}\mathbf{1}_{x<M} \to 1_{\{0\}}$ pointwise. In order to use the dominated convergence theorem, we need to prove that e^{-x} is integrable under $PRM(\nu) - s_{PRM(\nu)}$ a.s. : we will prove the stronger statement that e^{-x} is integrable under $PRM(\nu)$ a.s.

In order to prove this, observe that $(1 - e^{-x}) \leq x$ for all $x > 0$. Thus,

$$\int_0^\infty (1 - e^{-e^{-x}}) d\nu(x) \leq \int_0^\infty e^{-x} d\nu(x) = \alpha \int_0^\infty e^{-x} x^{\alpha-1} dx < \infty.$$

However, by (Resnick, 1987, Proposition 3.6(ii)), the left hand side is equal to $-\ln \mathbb{E}[e^{-\int_0^\infty e^{-x} dPRM(\nu)}]$. It follows that $\int_0^\infty e^{-x} dPRM(\nu)$ is finite a.s.

Therefore, by the dominated convergence theorem, $\int e^{-Mx}\mathbf{1}_{x<M}(PRM(\nu) - s_{PRM(\nu)}) \to 1$ a.s., which implies that the probability in equation 71 converges to 0. In particular, it follows that

$$\limsup_n \mathbb{P}\left[\sum_{i=1}^n e^{-\theta_n\left(\mathcal{M} - F^{-1}(1-\frac{1}{n})\right)\left(\frac{X_1 - X_i}{B - F^{-1}(1-\frac{1}{n})}\right)} > 1 + \epsilon\right] = 0,$$

which instantly proves the claim, since the sum must always exceed 1 (by taking the term $i = 1$), therefore it was sufficient to check only exceedance by ϵ to confirm convergence in probability.

Having proved the claim, recall that

$$\begin{aligned}\mathbb{P}(R_{n,\theta_n} \in A_n) &= \frac{\sum_{i=1}^n e^{\theta_n X_i}\mathbf{1}_{X_i \in A_n}}{\sum_{i=1}^n e^{\theta_n X_i}} \\ &= \frac{\sum_{i=1}^n e^{\theta_n X_i}\mathbf{1}_{X_i \in A_n}}{e^{\theta_n \max_i X_i}} \frac{e^{\theta_n \max_i X_i}}{\sum_{i=1}^n e^{\theta_n X_i}}\end{aligned}$$

The second term converges in probability to 1 by our claim, while the first term can be written as

$$\frac{\sum_{i=1}^n e^{\theta_n X_i}\mathbf{1}_{X_i \in A_n}}{e^{\theta_n \max_i X_i}} = \mathbf{1}_{\max_i X_i \in A_n} + \sum_{i=1}^n e^{\theta_n (X_i - \max_i X_i)}\mathbf{1}_{X_i \in A_n, X_i \neq \max_j X_j}.$$

Again, by the claim, the second term converges to 0 in probability. It now follows that

$$\mathbb{P}(R_{n,\theta_n} \in A_n) - \mathbb{P}(\max_i X_i \in A_n) \to 0$$

in probability. In particular, for any fixed positive C take $A_n = B - C(\mathcal{M} - F^{-1}(1 - \frac{1}{n}))$. Then, clearly

$$\mathbb{P}(\max_i X_i \in A_n) = \mathbb{P}\left(\frac{\mathcal{M} - \max_i X_i}{\mathcal{M} - F^{-1}(1 - \frac{1}{n})} \geq C\right) \to \mathbb{P}(-W_\alpha \geq C).$$

Therefore, we have $\mathbb{P}(R_{n,\theta_n} \in A_n) \to \mathbb{P}(-W_\alpha \geq C)$, which rearranges itself to

$$P\left(\frac{\mathcal{M} - R_{n,\theta_n}}{\mathcal{M} - F^{-1}(1 - \frac{1}{n})} \geq C\right) \to \mathbb{P}(-W_\alpha \geq C)$$

as desired. This completes the proof. □

E Proofs for Section 5

In this section, we prove the results from Section 5. Recall that X is now a vector, $\theta \in \mathbb{R}^d$ a fixed unit vector, $c > 0$ a scalar tending to infinity, and the distributions of $X_{c\theta}$ and $R_{n,c\theta}$ are given by equation 1 and equation 2.

We begin with the proof of Lemma 5, which is essentially the same as the proof of Lemma 1.

Proof. Let a_n be a sequence of vectors and $b_n > 0$. Then, for any axis-aligned hypercube R, observe that $\{\frac{r-a_n}{b_n} : r \in R\}$ is still an axis-aligned rectangle. Furthermore, this map is invertible, and thus is a $1-1$ mapping on $\mathcal{R}$, the set of all axis-aligned hypercubes on $\mathcal{R}$. Therefore, for any $n \geq 1$,

$$\sup_R |\mathbb{P}(X_{1,n} \in R) - \mathbb{P}(X_{2,n} \in R)| = \sup_R \left| \mathbb{P}\left(\frac{X_{1,n} - a_n}{b_n} \in R\right) - \mathbb{P}\left(\frac{X_{2,n} - a_n}{b_n} \in R\right) \right|. \tag{72}$$

For part(a), if $Z_1 = Z_2 = Z$ is continuous, then by Lemma 12 we know that the distribution of $\frac{X_{i,n}-a_n}{b_n}$ converges uniformly on compacts to F_Z. Given this, part(a) is clear by equation 72.

On the other hand, for part (b) we note that if $Z_1 \neq Z_2$ then one can find a continuity point c of Z_1, Z_2 at which their CDFs differ, by Lemma 12. However, if $\mathcal{R}_c = \{x : x_i \leq c_i, 1 \leq i \leq d\}$ then the difference between the CDFs at c equals $|\mathbb{P}(Z_1 \in \mathcal{R}_c) - \mathbb{P}(Z_2 \in \mathcal{R}_c)| = \epsilon > 0$, which is non-negative. Let c' be such that $F_{Z_1}(c'), F_{Z_2}(c') < \epsilon/2$. Then, it follows that

$$|\mathbb{P}(Z_1 \in \mathcal{R}_c \setminus \mathcal{R}_{c'}) - \mathbb{P}(Z_2 \in \mathcal{R}_c \setminus \mathcal{R}_{c'})| > \epsilon - \epsilon/2 - \epsilon/2 > 0.$$

Since $X_{i,n} \to Z_i$ in distribution, it follows that their CDFs cannot uniformly converge to the same limit on $\mathcal{R}_c \setminus \mathcal{R}_{c'}$, completing the proof. ☐

Next, we sketch proofs for the examples of random vectors satisfying Assumption 4.

Proof of Examples of vectors satisfying Assumption 4. We begin with the one-dimensional example. Let X be a random variable with maximum value B which is in the Weibull regime with parameter $\alpha > 0$. Let F_X be the CDF of X.

We will take $x_\theta = \mathcal{M}$. Let $A = [0, c]$ for some arbitrary $c > 0$. Let $U(t) = 1 - F_X(\mathcal{M} - t)$, which is regularly varying with index $\alpha > 0$ by Assumption 1. Then,

$$\frac{1}{U(t)} \mathbb{P}\left(\frac{\mathcal{M} - X}{t} \in A\right) = \frac{1}{U(t)} \mathbb{P}(X \geq \mathcal{M} - tc) = \frac{1 - F_X(\mathcal{M} - tc)}{1 - F_X(\mathcal{M} - t)} = c^\alpha.$$

Thus, it follows that $\nu([0, c]) = c^\alpha$ for all $c > 0$, which uniquely determines ν. It is sufficient to consider intervals to prove vague convergence, hence the result follows.

For the second example, let $(X_1, X_2, \ldots, X_d)$ be a bounded random vector with independent components, and let $x = (\mathcal{M}_1, \ldots, \mathcal{M}_d)$ be the vector of maximal values of the X_i. Suppose X_i is in the Weibull regime with index ρ_i. By an approximation argument, it is sufficient to find a function $U(t)$ and a measure ν such that

$$\frac{1}{U(t)} \mathbb{P}(x - X \in tA) \to \nu(A)$$

for all hypercubes A of the form $A = \times_{i=1}^n [a_i, b_i]$ where $a_i \leq b_i$ for $1 \leq i \leq d$. However, in this case,

$$x \in x - tA \text{ if and only if } X_i \in [\mathcal{M}_i - ta_i, \mathcal{M}_i - tb_i] \text{ for all } 1 \leq i \leq d.$$

Therefore,

$$\mathbb{P}(x - X \in tA) = \prod_{i=1}^d \mathbb{P}(X_i \in [\mathcal{M}_i - ta_i, \mathcal{M}_i - tb_i]).$$

Let $U_i(t) = \frac{1}{1 - F_{X_i}(\mathcal{M} - t)}$. Then,

$$\frac{1}{U_i(t)} \mathbb{P}\left(X_i \in [\mathcal{M}_i - ta_i, \mathcal{M}_i - tb_i]\right) = \frac{1}{U_i(t)} \mathbb{P}\left(\frac{\mathcal{M} - X_i}{t} \in [a_i, b_i]\right) \to \nu_i([a_i, b_i])$$

by Assumption 1, where $\nu_i([0, y]) = y^{\rho_i}$. Thus, taking the product of the above convergence over all $i \in 1, \ldots, d$,

$$\frac{1}{\prod_{i=1}^d U_i(t)} \mathbb{P}(x - X \in tA) \to \nu(A)$$

where $\nu = \otimes_{i=1}^n \nu_i$ is the product measure. Since the product of regularly varying functions is still regularly varying, the result follows.

For the third example, suppose that X is a truncated normal $X = N(0,I)\mathbf{1}_{N(0,I)\le r}$ for some $r > 0$. Fix $y \in R^d, \|y\| = r$. Note that $N(0,I)$ possesses a density f which is continuous at y, and $f(y) > 0$. So, for any compact set A we have

$$\frac{\mathbb{P}(X \in y - tA)}{f(y)|(y-tA) \cap \{\|x\| \le r\}|} \to 1.$$

Now, it is sufficient to see that $U(t) = |(y-tA) \cap \{\|x\| \le r\}|$ is regularly varying at 0 of order d. This is in particular easy to see when A is a sphere, and then easy to extend to a disjoint union of spheres, followed by general compact sets using a common approximation argument. Thus, with the vague limit $\nu = 1$, the proof is complete.

Finally, if X is a uniform random vector on any polytope P, then the density is the same everywhere, and therefore if y is an extreme point then like the previous example, A is any compact set and c denotes the common density at all points,

$$\frac{\mathbb{P}(X \in y - tA)}{c|(y-tA) \cap P|} \to 1.$$

Depending upon the geometry of y at P, we can find a truncated cone of some dimension $k \le d$ centered at y which lies entirely inside P and whose boundary coincides with the boundary of P until some positive radius. For instance, if P is a k-hypercube in $\mathbb{R}^d$ then the cone would be of dimension k (e.g. if P is a line segment in the plane then $k = 1$ and $d = 2$). It is now clear that $c|(y-tA) \cap P|$ is regularly varying of order k at zero if A is a hypercube chosen so that its axes align correctly with the faces of P, concluding the proof with vague limit $\nu = 1$ again. $\square$

We now prove Lemma 7.

Proof of Lemma 7. Let X satisfy Assumption 4 at x_θ with function $U(t)$ and vague limit ν. For the sake of clarity, we assume $(x_\theta)_i \neq 0$ for all $1 \le i \le d$: the proof is easily modified if one of these doesn't hold. This implies that for every bounded rectangle A we have

$$\frac{1}{U(t)}\mathbb{P}\left(\frac{x_\theta - X}{t} \in A\right) \to \nu(A).$$

Let $A = \times_{i=1}^d [a_i, b_i]$ with $a_i \le b_i$ and suppose $g(x) = (x_i^{\alpha_i})_{1\le i\le d}$. Now, we have

$$\begin{aligned}\left\{\frac{g(x_\theta) - g(X)}{t} \in A\right\} = \{g(X) \in g(x_\theta) - tA\} =& \{X \in g^{-1}(g(x_\theta) - tA)\} \\ =& \left\{\frac{x - X}{t} \in \frac{x_\theta - g^{-1}(g(x_\theta) - tA)}{t}\right\}\end{aligned}$$

Expanding out the definition of g and g^{-1},

$$\frac{x_\theta - g^{-1}(g(x_\theta) - tA)}{t} = \times_{i=1}^d \left[\frac{x_{\theta_i} - (((x_\theta)_i)^{\alpha_i} - ta_i)^{\frac{1}{\alpha_i}}}{t}, \frac{x_{\theta_i} - (((x_\theta)_i)^{\alpha_i} - tb_i)^{\frac{1}{\alpha_i}}}{t}\right]$$

Observe that as $t \to 0$, the above set converges :

$$\times_{i=1}^d \left[\frac{x_{\theta_i} - (((x_\theta)_i)^{\alpha_i} - ta_i)^{\frac{1}{\alpha_i}}}{t}, \frac{x_{\theta_i} - (((x_\theta)_i)^{\alpha_i} - tb_i)^{\frac{1}{\alpha_i}}}{t}\right] \to \times_{i=1}^d \frac{[a_i, b_i]}{\alpha_i x_{\theta_i}^{\alpha_i - 1}}.$$

Thus, it follows that

$$\frac{1}{U(t)}\mathbb{P}\left(\frac{g(x_\theta) - g(X)}{t} \in A\right) \to \mu(B)$$

where if $B = \times_{i=1}^d [a_i, b_i]$ with $a_i \le b_i$ then $\mu(B) = \nu\left(\times_{i=1}^d \frac{[a_i,b_i]}{\alpha_i x_{\theta_i}^{\alpha_i - 1}}\right)$. That is, μ is just a rescaling of ν. This completes the proof. $\square$

Unlike the previous appendix section, we prove Theorem 8 and Theorem 9 in the same subsection, since their proofs are extremely similar. The three parts of Theorem 10 will then be covered later.

E.0.1 Proof of Theorem 8 and Theorem 9

The following lemma will be extremely helpful in this subsection. Let X satisfy Assumption 4 at x_θ with limiting measure ν and regularly varying function $U(t)$.

Lemma 17. *Let $B \subset \mathbb{R}^d$ be such that $\nu(\partial B) = 0$, where ∂B is the topological boundary of B. We have*

$$\frac{\mathbb{E}[e^{c\theta^T X}\mathbf{1}_{X \in x_\theta - \frac{1}{c}B}]}{e^{c\theta^T x_\theta} U(\frac{1}{c})} \to \int_B e^{-\theta^T y} d\nu(y)$$

as $c \to \infty$.

Proof. We have

$$\begin{aligned}
&\mathbb{E}[e^{c\theta^T X}\mathbf{1}_{X \in x_\theta - \frac{1}{c}B}] \\
=& e^{c\theta^T x_\theta} \int e^{-c\theta^T(x_\theta - X)} \mathbf{1}_{X \in x_\theta - \frac{1}{c}B} d\mathbb{P}(X) \\
=& e^{c\theta^T x_\theta} \int e^{-\theta^T y} \mathbf{1}_{y \in B} d\mathbb{P}\left(c(x_\theta - X)\right) \\
=& e^{c\theta^T x_\theta} U\left(\frac{1}{c}\right) \frac{\int e^{-\theta^T y} \mathbf{1}_{y \in B} d\mathbb{P}\left(c(x_\theta - X)\right)}{U(\frac{1}{c})}.
\end{aligned}$$

Observe that $e^{-\theta^T y}\mathbf{1}_{y \in B}$ is discontinuous only on ∂B, but $\nu(\partial B) = 0$. Thus, as $c \to \infty$, the latter term converges to $\int_B e^{-\theta^T y} d\nu(y)$ by Assumption 4, completing the proof. □

The proof of the theorem follows from the above lemma rather easily.

Proof of Theorem 8. By Lemma 17,

$$\begin{aligned}
\frac{\mathbb{E}[e^{2c\theta^T X}]}{U(1/2c) e^{2c\theta^T x_\theta}} &\to \int e^{-\theta^T y} d\nu(y) \\
\frac{U^2(1/c) e^{2c\theta^T x_\theta}}{\mathbb{E}[e^{c\theta^T X}]^2} &\to \frac{1}{\int e^{-\theta^T y} d\nu(y)}
\end{aligned}$$

Multiplying these together,

$$\frac{\frac{\mathbb{E}[e^{2c\theta^T X}]}{\mathbb{E}[e^{c\theta^T X}]^2}}{U^{-2}(1/c)U(1/2c)} \to 1.$$

Note that

$$\frac{U(1/2c)}{U(1/c)} \to 2^{-\alpha}$$

as $c \to \infty$. Hence,

$$U(1/c) \frac{\mathbb{E}[e^{2c\theta^T X}]}{\mathbb{E}[e^{c\theta^T X}]^2} \to 2^{-\alpha},$$

which was to be proved. □

The above lemma is also enough to provide a simple proof for the scaling limit of $X_{c\theta}$.

Proof of Theorem 9. It suffices to prove that for every set $B \subset \mathbb{R}^d$ such that $\nu(\partial B) = 0$, we have

$$\mathbb{P}\left(c(x_\theta - X_{c\theta}) \in B\right) \to \mathbb{P}\left(Z \in B\right).$$

However, observe that

$$\begin{aligned}
&\mathbb{P}\left(c(x_\theta - X_{c\theta}) \in B\right) \\
=&\mathbb{P}\left(X_{c\theta} \in x_\theta - \frac{1}{c}B\right) \\
=&\frac{\mathbb{E}[e^{c\theta^T X}\mathbf{1}_{X\in x_\theta - \frac{1}{c}B}]}{\mathbb{E}[e^{c\theta^T X}\mathbf{1}_{X\in\mathbb{R}^d}]} \\
\to&\frac{\int_B e^{-\theta^T y} d\nu(y)}{\int e^{-\theta^T y} d\nu(y)},
\end{aligned}$$

where we applied Lemma 17 once for the numerator and once for the denominator with B replaced by $\mathbb{R}^d$. This completes the proof. ☐

In the next subsection, we will prove Theorem 10.

E.1 Proof of Theorem 10

The below proof of Theorem 10(a) is extremely similar to the proof of Theorem 6(a).

Proof of Theorem 10(a). We will use Proposition 4. Observe that $\frac{M_{\theta_n}}{n} \to 0$. Let B be any Borel set such that $\nu(B) > 0$ and $\nu(\partial B) = 0$, and let $B_n = x_\theta - \frac{B}{c_n}$. By Theorem 9 it is clear that $\lim_{n\to\infty} \frac{\mathbb{E}[e^{\theta_n^T X}]}{\mathbb{E}[e^{\theta_n^T X}\mathbf{1}_{X\in B_n}]} < \infty$. Therefore, applying Proposition 4,

$$\frac{\mathbb{P}(R_{n,c\theta} \in x_\theta - \frac{1}{c}B)}{\mathbb{P}(X_{c\theta} \in x_\theta - \frac{1}{c}B)} \to 1$$

in probability. Now, the denominator converges to a known quantity $\mathbb{P}(Z \in B)$ by Theorem 9, where Z has density $e^{-\theta^T y} d\nu(y)$. Hence the numerator must also converge to the same quantity, implying that

$$\mathbb{P}\left(R_{n,c\theta} \in x_\theta - \frac{1}{c}B\right) \to \mathbb{P}(Z \in B)$$

which completes the proof following a simple rewrite. ☐

Next, we will prove Theorem 10(b).

Proof of Theorem 10(b). Suppose that $\frac{M_{c\theta}}{n} \to C \in (0, \infty)$. Then, by Theorem 8 we know that $cU^{-1}(\frac{1}{n}) \to C_1$ for some $c_1 > 0$, and hence $\theta_n \to \infty$. Let D be a Borel set, and let $A_n = x_\theta - U^{-1}(\frac{1}{n})D$.

Consider $\mathbb{P}(R_{c,\theta} \in A_n)$, and divide the top and bottom by $e^{\theta^T x_\theta}$ to obtain

$$\begin{aligned}
\mathbb{P}(R_{c,\theta} \in A_n) =&\frac{\sum_{i=1}^n e^{c\theta^T X_i}\mathbf{1}_{X_i\in A_n}}{\sum_{i=1}^n e^{\theta^T X_i}} \\
=&\frac{\sum_{i=1}^n e^{\theta^T (X_i - x_\theta)}\mathbf{1}_{X_i\in A_n}}{\sum_{i=1}^n e^{\theta^T (X_i - x_\theta)}} \\
=&\Phi_n\left(\sum_{i=1}^n e^{\theta_n^T(x_\theta - X_i)} 1_{A_n}, \sum_{i=1}^n e^{\theta_n^T(x_\theta - X_i)}\right),
\end{aligned}$$

where

$$\Phi(y, z) = \frac{y}{z}.$$

At this point, we verify the hypotheses of Lemma 14. We take $c_n = \|\theta_n\|$ and θ as in the lemma itself. Then, $X_{\theta_n} \to x_\theta$ in probability. Let $a_n = U^{-1}(\frac{1}{n})$, C_1 be as above, ν be as in equation 4 and D as chosen. All the hypotheses of Lemma 14 are easily verified.

Since Φ is a continuous mapping, and $\int e^{-C_1\theta^T y} dPRM(\nu) \neq 0$ with probability 1 by the definition of a PRM, by Lemma 14 it follows that

$$\mathbb{P}(R_{n,\theta} \in A_n) = \Phi_n \left(\sum_{i=1}^{n} e^{c\theta^T(x_\theta - X_i)} 1_{A_n}, \sum_{i=1}^{n} e^{c\theta^T(x_\theta - X_i)} \right) \xrightarrow{d} \frac{\int e^{-C_1 y} 1_{y\in D} dPRM(\nu)}{\int e^{-C_1 y} dPRM(\nu)}.$$

Finally, note that

$$\mathbb{P}(R_{c,\theta} \in A_n) = \mathbb{P}(c\theta(x_\theta - R_{c,\theta}) \in c(x_\theta - A_n)) \approx \mathbb{P}(c\theta(x_\theta - R_{c,\theta}) \in C_1 D)$$

if n is large enough. Combining the two statements above, if Z is a random vector such that

$$\mathbb{P}(Z \in D) = \frac{\int e^{-C_1 y} 1_{y\in D} dPRM(\nu)}{\int e^{-C_1 y} dPRM(\nu)},$$

then

$$\mathbb{P}(c(x_\theta - R_{c,\theta}) \in D) \to \mathbb{P}(C_1 Z \in D).$$

Thus, the random variable $Z_{C,PRM} = C_1 Z$ is the desired limit. Note that this random variable depends upon the Poisson random measure, while the scaling limit Z from part(a) of the theorem does not. It follows that these two random variables are not the same, which concludes the proof.

We remark that, as in the one-dimensional case, the limiting random variable $Z_{C,PRM}$ is continuous, but do not prove this here. $\square$

The only non-trivial part of the proof of Theorem 10(c) is finding the scaling limit for the sample maximizer of $\theta^T X_{(i)}$. Once this is done, our proof will follow exactly as in the one-dimensional case, and we can omit the proofs of these parts. The next lemma explains the origin of the random variable V in Theorem 10(c) as ths aforementioned scaling limit.

Lemma 18. *Let $X_1, X_2, ..., X_n$ be iid X and let $X_{(n)} = \arg\max_i \theta^T X_i$ be the sample maximizer of $\theta^T X_i$. Then,*

$$\frac{x - X_{(n)}}{U^{-1}(\frac{1}{n})} \to V$$

for some random variable V depending upon ν.

Proof. Let $\mathcal{S} = \{w : \theta^T w = 1\}$, and define $T : \{y : \theta^T y > 0\} \to (0,\infty) \times \mathcal{S}$ by

$$T(y) = \left(\theta^T y, \frac{y}{\theta^T y} \right).$$

Clearly T is continuous on this space, and the support of $\frac{x-X}{t}$ is contained in the domain of T for all $t > 0$.

We claim, furthermore, that $T^{-1}(K)$ is compact for every $K \subset (0,\infty) \times \mathcal{S}$ which is compact. This is easy to see : indeed, if K is compact then K is contained in a set of the form $[a,b] \times J$ where $J \subset \mathcal{S}$ is compact. However, it's easy to see that $T^{-1}([a,b] \times J)$ is a closed and bounded rectangle, hence compact (this is particularly easy to see if θ is a multiple of one of the standard basis vectors, for instance). Hence, $T^{-1}(K)$ is a closed subset of a compact set, hence compact.

By (Resnick, 1987, Proposition 3.18), we have that if $\frac{1}{U(t)}\mathbb{P}(\frac{x-X}{t} \in \cdot) \to \nu(\cdot)$ vaguely, then

$$\frac{1}{U(t)} \mathbb{P}\left(\left(\frac{\theta^T(x-X)}{t}, \frac{x-X}{\theta^T(x-X)} \right) \in \cdot \right) \to \nu \circ T^{-1}(\cdot)$$

vaguely. But by Lemma 6(b), $\nu \circ T^{-1}$ is a product measure, of the form

$$\nu \circ T^{-1} = dr \times ds.$$

Now, by (Resnick, 1987, Proposition 3.21) and a restriction argument similar to the proofs in (Resnick, 1987, Section 3.3.2) we obtain

$$\sum_{i=1}^{n} \delta_{(\theta^T(x-X_i), \frac{x-X_i}{\theta^T(x-X_i)})} \Rightarrow PRM(\nu \circ T^{-1}) = PRM(dr \times ds),$$

where we use $dr \times ds$ to denote the spherical and angular parts of $d\nu \circ T^{-1}$. By the continuous mapping theorem and L, it follows that $\frac{x-X_{(n)}}{U^{-1}(1/n)}$ converges to the atom V of the PRM on which $\theta^T y$ is minimized.

It is not difficult to see the distribution of V. Note that the point V is located at a point $(a, s) \in [0, \infty) \times \mathcal{S}$, if and only if the PRM has no point in the set $\{w : 0 < \theta^T w < a\}$, which occurs with probability $e^{-\nu \circ T^{-1}((0,a] \times \mathcal{S})}$. Thus, it follows that

$$\mathbb{P}(V \in B) = \int_B e^{-\nu \circ T^{-1}((0,r] \times \mathcal{S})} PRM(dr \times ds).$$

□

We now complete the proof of Theorem 10(c).

Proof of Theorem 10(c). Suppose, without loss of generality that X_1 is the sample maximizer of $\theta^T x$. We claim that in probability,

$$\frac{\sum_{i=1}^n e^{c\theta^T X_i}}{e^{c\theta^T X_1}} \to 1,$$

if $cU^{-1}(\frac{1}{n}) \to +\infty$, which is implied by Theorem 8 and $\frac{M_{\theta_n}}{n} \to \infty$. In particular, this implies that

$$\frac{\sum_{i=2}^n e^{c\theta^T X_i}}{e^{c\theta^T X_1}} \to 0. \tag{73}$$

To prove this claim, we have by Lemma 13 that $\sum_{i=1}^n \delta_{\frac{x_\theta - X_i}{U^{-1}(\frac{1}{n})}} \Rightarrow PRM(\nu)$ where ν is as in Assumption 4. Now, note that the smallest point of the former point process is attained when $X_i = \max_j X_j$, which without loss of generality we assume is attained at X_1. Therefore, we get by Lemma 15 that

$$\sum_{i=1}^{n} \delta_{\frac{X_1 - X_i}{U^{-1}(\frac{1}{n})}} \Rightarrow PRM(\nu) - s_{PRM(\nu)}.$$

Let $M > 0$ be an arbitrary but fixed parameter. Let $f(x) = e^{-Mx}\mathbf{1}_{\|x\|<M}$. Note that f is continuous and compactly supported. Thus, by the continuous mapping theorem (see (Resnick, 1987, Section 3.5)) we have

$$\sum_{i=1}^{n} e^{-M\frac{\theta^T(X_1-X_i)}{U^{-1}(\frac{1}{n})}} \mathbf{1}_{\|X_1-X_i\|<U^{-1}(\frac{1}{n})M} \xrightarrow{d} \int e^{-M\theta^T x}\mathbf{1}_{\|x\|<M}\left[PRM(\nu) - s_{PRM(\nu)}\right]. \tag{74}$$

Now, it is clear that for any $\epsilon > 0$,

$$\lim_{M\to\infty} \mathbb{P}\left[\sum_{i=1}^{n} e^{-c\theta^T(X_1-X_i)}\mathbf{1}_{\|X_1-X_i\|<MU^{-1}(\frac{1}{n})} > 1+\epsilon\right] = \mathbb{P}\left[\sum_{i=1}^{n} e^{-c\theta^T(X_i-X_1)} > 1+\epsilon\right].$$

For any $\epsilon > 0$ and $M > 0$, employing equation 74 gives

$$\begin{aligned}
&\limsup_{n} \mathbb{P}\left[\sum_{i=1}^{n} e^{-cU^{-1}(\frac{1}{n})\theta^T\left(\frac{X_1-X_i}{U^{-1}(\frac{1}{n})}\right)} \mathbf{1}_{\|X_1-X_i\|<MU^{-1}(\frac{1}{n})} > 1+\epsilon\right] \\
\leq &\limsup_{n\to\infty} \mathbb{P}\left[\sum_{i=1}^{n} e^{-M\left(\frac{\theta^T(X_1-X_i)}{U^{-1}(\frac{1}{n})}\right)} \mathbf{1}_{\|X_1-X_i\|<MU^{-1}(\frac{1}{n})} > 1+\epsilon\right] \\
= &\mathbb{P}\left[\int e^{-M\theta^T x} \mathbf{1}_{\|x\|<M} (PRM(\nu) - s_{PRM(\nu)}) > 1+\epsilon\right]. \qquad (75)
\end{aligned}$$

Note that $e^{-M\theta^T x}\mathbf{1}_{\|x\|<M}$ is dominated by $e^{-\theta^T x}$ for $M > 1$, since the support of ν is contained in $\{\theta^T y > 0\}$. Furthermore, as $M \to \infty$, we have $e^{-M\theta^T x}\mathbf{1}_{\|x\|<M} \to \mathbf{1}_{\theta^T y=0}$ pointwise. In order to use the dominated convergence theorem, we need to prove that $e^{-\theta^T x}$ is integrable under $PRM(\nu) - s_{PRM(\nu)}$ a.s. : we will prove the stronger statement that $e^{-\theta^T x}$ is integrable under $PRM(\nu)$ a.s.

In order to prove this, observe that $(1 - e^{-\theta^T x}) \leq \theta^T x$ whenever $\theta^T x > 0$. Thus,

$$\int_{\theta^T x>0} (1 - e^{-e^{-\theta^T x}}) d\nu(x) \leq \int_{\theta^T x>0} e^{-\theta^T x} d\nu(x) < \infty,$$

by Lemma 6(c).

However, by (Resnick, 1987, Proposition 3.6(ii)), the left hand side is equal to $-\ln \mathbb{E}[e^{-\int_{\theta^T y>0} e^{-\theta^T x} dPRM(\nu)}]$. It follows that $\int_{\theta^T y>0} e^{-\theta^T x} dPRM(\nu)$ is finite a.s.

By the dominated convergence theorem, the probability in equation 75 converges to 0. In particular, it follows that

$$\limsup_{n} \mathbb{P}\left[\sum_{i=1}^{n} e^{-cU^{-1}(\frac{1}{n})\theta^T\left(\frac{X_1-X_i}{U^{-1}(\frac{1}{n})}\right)} > 1+\epsilon\right] = 0,$$

which instantly proves the claim, since the sum must always exceed 1 (by taking the term $i = 1$), therefore it was sufficient to check only exceedance by ϵ to confirm convergence in probability.

Having proved the claim, recall that

$$\begin{aligned}
\mathbb{P}(R_{n,\theta_n} \in A_n) &= \frac{\sum_{i=1}^n e^{c\theta^T X_i} \mathbf{1}_{X_i \in A_n}}{\sum_{i=1}^n e^{c\theta^T X_i}} \\
&= \frac{\sum_{i=1}^n e^{c\theta^T X_i} \mathbf{1}_{X_i \in A_n}}{e^{c\theta^T X_1}} \frac{e^{c\theta^T X_1}}{\sum_{i=1}^n e^{c\theta^T X_i}}
\end{aligned}$$

The second term converges in probability to 1 by our claim, while the first term can be written as

$$\frac{\sum_{i=1}^n e^{c\theta^T X_i} \mathbf{1}_{X_i \in A_n}}{e^{c\theta^T X_1}} = \mathbf{1}_{c\theta^T X_1 \in A_n} + \sum_{i=2}^{n} e^{c\theta^T(X_i - X_1)} \mathbf{1}_{X_i \in A_n}.$$

Again, by the claim, the second term converges to 0 in probability. It now follows that

$$\mathbb{P}(R_{c,\theta} \in A_n) - \mathbb{P}(X_1 \in A_n) \to 0$$

in probability. In particular, for any Borel set B take $A_n = x_\theta - BU^{-1}(\frac{1}{n})$. Then, clearly

$$\mathbb{P}(X_1 \in A_n) = \mathbb{P}\left(\frac{x_\theta - X_1}{U^{-1}(\frac{1}{n})} \in B\right) \to \mathbb{P}(V \in B).$$

Therefore, we have $\mathbb{P}(R_{c,\theta} \in A_n) \to \mathbb{P}(V \in B)$, which rearranges itself to

$$P\left(\frac{x_\theta - R_{c,\theta}}{U^{-1}(\frac{1}{n})} \in B\right) \to \mathbb{P}(V \in B)$$

as desired. This completes the proof. ☐

F Proofs for Section 6

In this section we will prove results from Section 6. Recall, in this section, that X is a continuous random vector with full support on $\mathbb{R}^d$, and density f. Also recall that there are $\alpha > 1, K, L > 0$ such that equation 19 holds. Fix $\theta \in \mathbb{R}^d, \|\theta\| = 1$ and $c > 0$ which will tend to infinity.

For completeness, we restate Lemma 8. However, we also include a part(c) in this statement which was omitted in the original, since it is important for the proof but not for the statements of the theorems in this section.

Lemma 19. *For each $c > 0$, the function $\Phi_c(x)$*

(a) attains its unique maximum at a point $m_c = \left(\frac{c}{\alpha K}\right)^{\frac{1}{\alpha-1}}\theta$, and

$$\Phi_c(m_c) = (\alpha-1)\alpha^{-\alpha/(\alpha-1)}c^{\alpha/(\alpha-1)}K^{-1/(\alpha-1)}.$$

(b) is thrice continuously differentiable in $\mathbb{R}^d \setminus \{0\}$, and

$$\nabla^2\Phi_c(m_c) = -K\alpha m_c^{\alpha-2}(I + (\alpha-2)\theta\theta^T).$$

(Note : in the second term, the product of vectors is the outer product, not the inner product. Hence it leads to a rank-one matrix).

(c) is uniformly second-order Taylor approximable for large c i.e. for every $\epsilon > 0$, there exists c_0 such that $c > c_0$ and $\|x - m_c\| \leq \frac{\|m_c\|}{2}$ implies

$$\left|\Phi_c(x) - \Phi_c(m_c) - \frac{1}{2}(x-m_c)^T(\nabla^2\Phi_c(m_c))(x-m_c)\right| \leq \frac{\epsilon}{2}(x-m_c)^T(\nabla^2\Phi_c(m_c))(x-m_c).$$

Proof. Note that as $\|x\| \to \infty$, since $\theta^T x \leq \|\theta\|\|x\|$ by the Cauchy Schwarz inequality, it follows that $\lim_{\|x\|\to\infty}\Phi_c(x) = -\infty$. Furthermore, Φ_c is differentiable everywhere except possibly at 0, where $\Phi_c(0) = 0$. Therefore, the maximum of Φ_c must be attained at one of its critical points provided we show that the value at this point is positive. Write $\|x\|^\alpha = \left(\|x\|^2\right)^{\alpha/2}$, which is easier to differentiate since the derivative of $\|x\|^2$ equals $2x$.

Differentiating,

$$[\nabla\Phi_c(x)](h) = c\theta^T h - \frac{K\alpha}{2}\|x\|^{\alpha-2} \times 2x^T h. \tag{76}$$

Since we are searching for critical points, this must be set to 0 for all h, whence it follows that

$$c\theta = K\alpha\|x\|^{\alpha-2}x.$$

This tells us that x is a multiple of θ, say $x = L\theta$. But then

$$c = K\alpha L^{\alpha-1} \implies L = \left(\frac{c}{\alpha K}\right)^{\frac{1}{\alpha-1}}.$$

It follows that Φ_c possesses a unique critical point at $m_c = \left(\frac{c}{\alpha K}\right)^{\frac{1}{\alpha-1}}\theta$. A quick computation now reveals that

$$\Phi_c(m_c) = cL - KL^\alpha = (\alpha-1)\alpha^{-\alpha/(\alpha-1)}c^{\alpha/(\alpha-1)}K^{-1/(\alpha-1)}$$

which is a positive constant, completing the proof of part (a).

To compute the Hessian $\nabla^2\Phi_c$ it suffices to differentiate equation 76 once more, leading to

$$\nabla^2\Phi_c(x) = -K\alpha\left((\alpha-2)\|x\|^{\alpha-4}xx^T + \|x\|^{\alpha-2}I\right).$$

Substituting $x = m_c$ gives

$$\nabla^2\Phi_c(m_c) = -K\alpha\|m_c\|^{\alpha-2}(I + (\alpha-2)\theta\theta^T), \tag{77}$$

completing the proof of part(b).

The third derivative of Φ_c can be obtained by differentiating the above expression once more. We do not include the calculations, but it suffices to see that for some constant $D > 0$, $\lim_{\|x\|\to\infty} \frac{\nabla^3\Phi_c(x)}{D\|x\|^{\alpha-3}} = 1$. (This is rather easy to observe in the one-dimensional case, for instance).

Therefore, by the Taylor Theorem with remainder applied to $\Phi_c(x)$ in the ball $B_c = \{x : \|x - m_c\| \le \|m_c\|/2\}$,

$$\Phi_c(x) = \Phi_c(m_c) + \frac{1}{2}(x - m_c)^T \nabla^2\Phi_c(m_c)(x - m_c) + R_3(x), \tag{78}$$

where

$$\|R_3(x)\| \le \sup_{\|x-m_c\|\le\|m_c\|/2} \|\nabla^3\Phi_c(x)\| \le M\|m_c\|^{\alpha-3}$$

for some constant M and all c large enough. In particular, this along with equation 77 implies that

$$\lim_{c\to\infty} \sup_{x\in B_c} \frac{\|R_3(x)\|}{\frac{1}{2}\|(x - m_c)^T \nabla^2\Phi_c(m_c)(x - m_c)\|} = 0.$$

For any $\epsilon > 0$, we may use the above limit to see that for $c > 0$ large enough,

$$\|R_3(x)\| \le \frac{\epsilon}{2}\|(x - m_c)^T \nabla^2\Phi_c(m_c)(x - m_c)\|$$

for all $x \in B_c$. Part (c) follows directly from equation 78 and the triangle inequality. □

We will now prove Theorem 11 and Theorem 12.

F.1 Proof of Theorem 11 and Theorem 12

The following lemma, which we proceed to state, does all the heavy lifting and begets the scaling limit of $X_{c\theta}$ and the growth of $M_{c\theta}$ as corollaries.

Lemma 20. *Let $B \subset \mathbb{R}^d$ be any Borel set. Define*

$$B_c = m_c + (-\nabla^2\Phi_c(m_c))^{-\frac{1}{2}}B,$$

where m_c, Φ_c are as in Lemma 8. Then,

$$\lim_{c\to\infty} \frac{\mathbb{E}[e^{c\theta^T X} 1_{X\in B_c}]}{\frac{e^{\Phi_c(m_c)}}{\sqrt{det(-\nabla^2\Phi_c(m_c))}}} = L\mathbb{P}(N(0, I) \in B).$$

Proof. For notational convenience, let $b_c = \sqrt{det(-\nabla^2\Phi_c(m_c))}$. The proof will be furnished in two parts. We will prove that

$$\liminf_{c\to\infty} \frac{\mathbb{E}[e^{c\theta^T X} 1_{X\in B_c}]}{\frac{e^{\Phi_c(m_c)}}{b_c}} \ge L\mathbb{P}(N(0, I) \in B), \tag{79}$$

and subsequently we will prove that

$$\limsup_{c\to\infty} \frac{\mathbb{E}[e^{c\theta^T X} 1_{X\in B_c}]}{\frac{e^{\Phi_c(m_c)}}{b_c}} \le L\mathbb{P}(N(0, I) \in B), \tag{80}$$

completing the proof.

To start the proof of equation 79 we let $\epsilon, \epsilon' > 0$ be arbitrary. By equation 19 there exists $R_{\epsilon'} > 0$ such that

$$f(x) \ge (L - \epsilon')e^{-K\|x\|^\alpha} \tag{81}$$

for all $\|x\| > R_{\epsilon'}$. By definition of the expectation and equation 81 we have

$$\begin{aligned}
&\mathbb{E}[e^{c\theta^T X} 1_{X \in B_c}] \\
=&\int_{B_c} e^{c\theta^T x} f(x) dx \\
\geq&(L-\epsilon') \int_{B_c \cap \{\|x\| \geq R_{\epsilon'}\}} e^{\Phi_c(x)} dx. \\
\geq&(L-\epsilon') \int_{B_c \cap \{\|x\| \geq R_{\epsilon'}\} \cap \{\|x-m_c\| \leq \frac{\|m_c\|}{2}\}} e^{\Phi_c(m_c) + \frac{1-\epsilon}{2}(x-m_c)^T \nabla^2 \Phi_c(m_c)(x-m_c)} dx,
\end{aligned} \tag{82}$$

where in the last step we used Lemma 8(c). We shall now perform a two-step substitution merely for the sake of clarity : this calculation will not be repeated in the lower bound. To begin with, we perform the substitution $y = x - m_c$. This gives

$$\begin{aligned}
&(L-\epsilon') \int_{B_c \cap \{\|x\| \geq R_{\epsilon'}\} \cap \{\|x-m_c\| \leq \frac{\|m_c\|}{2}\}} e^{\Phi_c(m_c) + \frac{1-\epsilon}{2}(x-m_c)^T \nabla^2 \Phi_c(m_c)(x-m_c)} dx \\
&\qquad = (L-\epsilon') \int_{(B_c - m_c) \cap \{\|y+m_c\| \geq R_{\epsilon'}\} \cap \{\|y\| \leq \frac{\|m_c\|}{2}\}} e^{\Phi_c(m_c) + \frac{1-\epsilon}{2} y^T \nabla^2 \Phi_c(m_c) y} dy.
\end{aligned} \tag{83}$$

Subsequently, we let $z = (-\nabla^2 \Phi_c(m_c))^{-\frac{1}{2}} y$. Effecting this change,

$$\begin{aligned}
&(L-\epsilon') \int_{(B_c - m_c) \cap \{\|y+m_c\| \geq R_{\epsilon'}\} \cap \{\|y\| \leq \frac{\|m_c\|}{2}\}} e^{\Phi_c(m_c) + \frac{1-\epsilon}{2} y^T \nabla^2 \Phi_c(m_c) y} dy \\
&\qquad = (L-\epsilon') \frac{e^{\Phi_c(m_c)}}{b_c} \int_{J_c} e^{-\frac{1-\epsilon}{2}\|z\|^2} dz,
\end{aligned} \tag{84}$$

where

$$J_c = B \cap \{\|(-\nabla^2 \Phi_c(m_c))^{-\frac{1}{2}} z + m_c\| \geq R_{\epsilon'}\} \cap \left\{\|(-\nabla^2 \Phi_c(m_c))^{-\frac{1}{2}} z\| \leq \frac{\|m_c\|}{2}\right\}. \tag{85}$$

Combining equation 81, equation 82, equation 83 and equation 84,

$$\frac{\mathbb{E}[e^{c\theta^T X} 1_{X \in B_c}]}{\frac{e^{\Phi_c(m_c)}}{b_c}} \geq (L-\epsilon') \int_{J_c} e^{-\frac{1-\epsilon}{2}\|z\|^2} dz. \tag{86}$$

The behavior of J_c as $c \to \infty$ will now be analyzed. As a consequence of Lemma 8(b), note that the matrix $M_c = (-\nabla^2 \Phi_c(m_c))^{-\frac{1}{2}}$ exhibits three kinds of behaviours as $c \to \infty$: if $1 < \alpha < 2$ then $M_c \to 0$. If $\alpha = 2$ then $M_c \to K\alpha I$. Finally, if $\alpha > 2$ then $\|M_c z\| \to \infty$ for every $z \neq 0$. Since $\|m_c\| \to \infty$ by Lemma 8(a), it follows that as $c \to \infty$, $\{\|(-\nabla^2 \Phi_c(m_c))^{-\frac{1}{2}} z + m_c\| \geq R_{\epsilon'}\}$ eventually converges to either $\mathbb{R}^d$ or $\mathbb{R}^d \setminus \{0\}$.

Similarly, by Lemma 8(b) it follows that the biggest eigenvalue of $\nabla^2 \Phi_c(m_c)$ grows at the rate of $\|m_c\|^{2-\alpha/2} < \frac{\|m_c\|}{2}$ for any $\alpha > 1$ and large enough c. Thus, $\left\{\|(\nabla^2 \Phi_c(m_c))^{-\frac{1}{2}} z\| \leq \frac{\|m_c\|}{2}\right\}$ converges to $\mathbb{R}^d$ as $c \to \infty$. It follows from these observations and the definition equation 85 of J_c, that either $J_c \to B$ or $J_c \to B \setminus \{0\}$ as $c \to \infty$.

Finally, letting $c \to \infty$ in equation 86,

$$\liminf_{c \to \infty} \frac{\mathbb{E}[e^{c\theta^T X} 1_{X \in B_c}]}{\frac{e^{\Phi_c(m_c)}}{b_c}} \geq (L-\epsilon') \int_B e^{-\frac{1-\epsilon}{2}\|z\|^2} dz.$$

This equation holds for arbitrary $\epsilon', \epsilon > 0$. Thus, letting these parameters tend to zero, the lower bound equation 79 follows.

We will now prove the upper bound equation 80. Once again, let $\epsilon, \epsilon' > 0$ be arbitrary constants. By equation 19 there exists $R_{\epsilon'} > 0$ such that

$$f(x) \leq (L+\epsilon') e^{-K\|x\|^\alpha} \tag{87}$$

for all $\|x\| > R_{\epsilon'}$. Thus,

$$\begin{aligned}
&\mathbb{E}[e^{c\theta^T X} 1_{X \in B_c}] \\
&= \int_{B_c} e^{c\theta^T X} \\
&\leq \int_{B_c \cap \{\|x\| \leq R_{\epsilon'}\}} e^{c\theta^T X} f(x) dx + (L + \epsilon') \int_{B_c \cap \{\|x\| > R_{\epsilon'}\}} e^{\Phi_c(x)} dx. \qquad (88)
\end{aligned}$$

By a discussion similar to that following equation 86 it follows that $B_c \cap \{\|x\| \leq R_{\epsilon'}\}$ converges to either the empty set or $\{0\}$ as $c \to \infty$, which is a set of measure zero. Thus, this term in equation 88 doesn't contribute on the right hand side i.e.

$$\lim_{c\to\infty} \frac{\int_{B_c \cap \{\|x\| \leq R_{\epsilon'}\}} e^{c\theta^T X} f(x) dx + (L + \epsilon') \int_{B_c \cap \{\|x\| > R_{\epsilon'}\}} e^{\Phi_c(x)} dx}{(L + \epsilon') \int_{B_c \cap \{\|x\| > R_{\epsilon'}\}} e^{\Phi_c(x)} dx} = 1. \qquad (89)$$

Owing to this realization, we now switch our focus to the second term of equation 88, and for $c > 0$ large enough apply Lemma 8(b) to get

$$\begin{aligned}
&(L + \epsilon') \int_{B_c \cap \{\|x\| > R_{\epsilon'}\}} e^{\Phi_c(x)} dx \\
&\qquad \leq (L + \epsilon') \int_{B_c \cap \{\|x\| > R_{\epsilon'}\} \cap \{\|x - m_c\| \leq \frac{\|m_c\|}{2}\}} e^{\Phi_c(m_c) + \frac{1+\epsilon}{2}(x - m_c)^T \nabla^2 \Phi_c(m_c)(x - m_c)} dx. \qquad (90)
\end{aligned}$$

Just as in the proof of equation 79, all we must do now is effect the change of variables and ensure consistency in the limit. $x = m_c + (-\nabla^2 \Phi_c(m_c))^{\frac{1}{2}} z$. Then,

$$\begin{aligned}
&(L + \epsilon') \int_{B_c \cap \{\|x\| > R_{\epsilon'}\} \cap \{\|x - m_c\| \leq \frac{\|m_c\|}{2}\}} e^{\Phi_c(m_c) + \frac{1+\epsilon}{2}(x - m_c)^T \nabla^2 \Phi_c(m_c)(x - m_c)} dx \\
&\qquad = (L + \epsilon') \frac{e^{\Phi_c(m_c)}}{b_c} \int_{J_c} e^{-\frac{1+\epsilon}{2}\|z\|^2} dz, \qquad (91)
\end{aligned}$$

where J_c is as in equation 85. As in the discussion there, it follows that $J_c \to \mathbb{R}^d$ or $\mathbb{R}^d \setminus \{0\}$ as $c \to \infty$. From here, taking the limit superior on both sides of equation 91 and noting the arbitrariness of ϵ, ϵ', the proof is immediate. □

The proofs of the two main theorems are now immediate.

Proof of Theorem 11. We proceed exactly as in the proof of Theorem 4. Taking $B = \mathbb{R}^d$ in Lemma 20, we obtain the following results.

$$\lim_{c\to\infty} \frac{\mathbb{E}[e^{2c\theta^T X}]}{\frac{e^{\Phi_{2c}(m_{2c})}}{\sqrt{det(-\nabla^2 \Phi_{2c}(m_{2c}))}}} = L$$

$$\lim_{c\to\infty} \frac{\frac{e^{2\Phi_c(m_c)}}{\det(-\nabla^2 \Phi_c(m_c))}}{\mathbb{E}[e^{c\theta^T X}]^2} = L^2.$$

Multiplying the two of them,

$$\lim_{c\to\infty} M_{c\theta} \times \frac{\frac{e^{2\Phi_c(m_c)}}{\det(-\nabla^2 \Phi_c(m_c))}}{\frac{e^{\Phi_{2c}(m_{2c})}}{\sqrt{det(-\nabla^2 \Phi_{2c}(m_{2c}))}}} = L.$$

EXPAND THIS OUT AND GET THE EXPLICIT VALUES □

Proof of Theorem 12. Let $B \subset \mathbb{R}^d$ be an arbitrary Borel set. Then, if we apply Lemma 20 with B and $\mathbb{R}^d$ separately,

$$\mathbb{P}(X_{c\theta} \in B_c) = \frac{\mathbb{E}[e^{c\theta^T X} \mathbf{1}_{X \in B_c}]}{\mathbb{E}[e^{c\theta^T X}]} \to \mathbb{P}(N(0, I) \in B) \tag{92}$$

as $c \to \infty$, where $B_c = m_c + (-\nabla^2 \Phi_c(m_c))^{-\frac{1}{2}} B$. However,

$$\mathbb{P}(X_{c\theta} \in B_c) = \mathbb{P}((-\nabla^2 \Phi_c(m_c))^{\frac{1}{2}} (X_{c\theta} - m_c) \in B).$$

Therefore, equation 92 is equivalent to saying that $(-\nabla^2 \Phi_c(m_c))^{\frac{1}{2}} (X_{c\theta} - m_c) \xrightarrow{d} N(0, I)$, as desired. $\square$

Now, we will prove Theorem 13.

F.2 Proof of Theorem 13

We begin with the proof of Theorem 13(a). Unsurprisingly, this is along the lines of the proof of Theorem 6(a) and Theorem 10(a).

Proof of Theorem 13(a). We will use Proposition 4 again. Let B be any Borel set of non-zero measure and let $B_c = m_c + (-\nabla^2 \Phi_c(m_c))^{-\frac{1}{2}} B$.

By Lemma 20 it is clear that $\lim_{n \to \infty} \frac{\mathbb{E}[e^{c\theta^T X}]}{\mathbb{E}[e^{c\theta^T X} \mathbf{1}_{X \in B_c}]} < \infty$. Furthermore, $\frac{M_{c\theta}}{n} \to 0$ by assumption. Therefore, applying Proposition 4,

$$\frac{\mathbb{P}(R_{n,c\theta} \in B_c)}{\mathbb{P}(X_{c\theta} \in B_c)} \to 1$$

in probability. Now, the denominator converges to a known quantity $\mathbb{P}(N(0, I) \in B)$ by Theorem 12. Hence the numerator must also converge to the same quantity. The proof is immediate. $\square$

C VS N IS BAD, WE SHOULD KEEP BOTH AS SEQUENCES, OTHERWISE ITS CONFUSING

For the proof of Theorem 13(b), we require the following vague convergence result. Following this, the PRM machinery makes everything go through as usual.

Theorem 17. *Let $a_c = m_c\theta$ and $b_c = \alpha^{-1} m_c^{1-\alpha}$. Then*

$$L^{-1} e^{m_c^\alpha} \alpha^d m_c^{d(\alpha-1)} \mathbb{P}\left(\frac{X - a_c}{b_c} \in \cdot\right) \to \nu(\cdot)$$

vaguely(where we recall vague convergence from Assumption 4), where ν has density $e^{-\theta^T y}$ over $\mathbb{R}^d$.

Proof. Let A be any compact set, and consider $\mathbb{P}\left(\frac{X-a_c}{b_c} \in A\right)$. Suppose that $\epsilon, \epsilon' > 0$ are arbitrary. Let $R > 0$ be large enough such that if $\|x\| > R$ then

$$(1 - \epsilon) L e^{-K\|x\|^\alpha} \leq f(x) \leq (1 + \epsilon) L e^{-K\|x\|^\alpha}. \tag{93}$$

Since A is compact and $\|m_c\| \to \infty$, we have that $b_c \to 0$, and therefore $a_c + b_c A \subset \{\|x\| > R\}$ for large enough $c > 0$.

By the remainder version of Taylor's theorem applied to $f(y) = \|y\|^\alpha$ at the point a_c, for any vector $h \in A$ we have

$$\|a_c + b_c h\|^\alpha = \|a_c\|^\alpha + \alpha b_c \|a_c\|^{\alpha-2} a_c^T h + R_2 \tag{94}$$

where

$$R_2 = \frac{1}{2} h^T \left(b_c^2 \alpha \|a_c + \lambda b_c h\|^{\alpha-2} I + b_c^2 \alpha(\alpha - 2) \|a_c + \lambda b_c h\|^{\alpha-4} (a_c + \lambda b_c h)(a_c + \lambda b_c h)^T\right) h \tag{95}$$

for some $\lambda \in (0,1)$. We will now show that this remainder term goes to 0 uniformly over $h \in A$ and $\lambda \in [0,1]$ as $c \to \infty$.

Since A is compact, $\|\lambda h\|$ is bounded for all $\lambda \in [0,1], h \in A$. Since $\|a_c\| \to \infty$ and $b_c \to 0$, it follows that $\frac{\|a_c+\lambda b_c h\|}{\|a_c\|} \to 1$ uniformly over $h \in A$ and $\lambda \in [0,1]$. Furthermore, by the definition of b_c and a_c it is clear that

$$b_c^2\|a_c\|^{\alpha-2} = \|a_c\|^{-\alpha} \to 0, \quad b_c^4\|a_c\|^{\alpha-4} = \|a_c\|^{-3\alpha} \to 0.$$

Therefore, both terms inside the bracket in equation 95 goes to zero as $c \to \infty$, implying that R_2 in equation 95 goes to 0.

So, by equation 94 we can pick c' large enough so that

$$\left|\|a_c + b_c h\|^\alpha - \|a_c\|^\alpha - \alpha b_c\|a_c\|^{\alpha-2}a_c^T h\right| \leq \epsilon'. \tag{96}$$

for all $c > c'$ and $h \in A$. We can now begin manipulating the original expression for c large enough, first demonstrating a lower bound, followed by an upper one.

$$\begin{aligned}
\mathbb{P}\left(\frac{X-a_c}{b_c} \in A\right) =& \mathbb{P}\left(X \in a_c + b_c A\right) \\
\overset{equation\ 93}{\geq}& (1-\epsilon)L\int_{a_c+b_cA} e^{-K\|x\|^\alpha}dx \\
=& b_c^d\int_A e^{-K\|a_c+b_c y\|^\alpha}dy \\
\overset{equation\ 96}{\geq}& (1-\epsilon)e^{-\epsilon'}Le^{-\|a_c\|^\alpha}b_c^d\int_A e^{-\alpha b_c\|a_c\|^{\alpha-2}a_c^T y}dy.
\end{aligned} \tag{97}$$

Therefore,

$$L^{-1}e^{\|a_c\|^\alpha}b_c^{-d}\mathbb{P}\left(\frac{X-a_c}{b_c} \in A\right) \geq (1-\epsilon)e^{-\epsilon'}\int_A e^{-\alpha b_c\|a_c\|^{\alpha-2}a_c^T y}dy. \tag{98}$$

By taking the limit inferior on both sides as $c \to \infty$, and recalling that $\alpha b_c\|a_c\|^{\alpha-2}a_c^T y \to \theta^T y$ from the definitions of b_c and a_c, we obtain

$$\liminf_{c\to\infty} L^{-1}e^{\|a_c\|^\alpha}b_c^{-d}\mathbb{P}\left(\frac{X-a_c}{b_c} \in A\right) \geq (1-\epsilon)e^{-\epsilon'}\int_A e^{-\theta^T y}dy.$$

Since this is true for all $\epsilon, \epsilon' > 0$ it follows that

$$\liminf_{c\to\infty} L^{-1}e^{\|a_c\|^\alpha}b_c^{-d}\mathbb{P}\left(\frac{X-a_c}{b_c} \in A\right) \geq \int_A e^{-\theta^T y}dy. \tag{99}$$

The other side of this equation i.e.

$$\limsup_{c\to\infty} L^{-1}e^{\|a_c\|^\alpha}b_c^{-d}\mathbb{P}\left(\frac{X-a_c}{b_c} \in A\right) \leq \int_A e^{-\theta^T y}dy. \tag{100}$$

is derived exactly in the same way as equation 99, using the other sides of the inequalities equation 93 and equation 96. This is exactly the definition of vague convergence, whence the result follows. ☐

Once this is true, if $f(c) = L^{-1}e^{m_c^\alpha}\alpha^d m_c^{d(\alpha-1)}$, then

$$n\mathbb{P}\left(\frac{X-a_{f^{-1}(n)}}{b_{f^{-1}(n)}} \in \cdot\right) \to \nu(\cdot)$$

where ν has density $e^{-\theta^T y}$.

From here, the proof of Theorem 13(b) follows exactly as the proof of the previous two theorems of this kind.

THIS PROOF, TO BE WRITTEN, REQUIRES US TO APPROPRIATELY ADJUST THE PROOF OF THAT PART B RESULT

Proof of Theorem 13(b). Suppose that $\frac{M_{c\theta}}{n} \to C \in (0, \infty)$. Then, by Theorem 11 we see that $\frac{c}{f^{-1}(n)} \to C_1$ for some $C_1 > 0$. Let D be a Borel set, and $A_c = m_c - b_c D$.

Consider $\mathbb{P}(R_{c\theta} \in A_n)$, and divide the top and bottom by $e^{\theta^T x_\theta}$ to obtain

$$\begin{aligned}\mathbb{P}(R_{c,\theta} \in A_n) =& \frac{\sum_{i=1}^n e^{c\theta^T X_i} \mathbf{1}_{X_i \in A_n}}{\sum_{i=1}^n e^{\theta^T X_i}} \\ =& \frac{\sum_{i=1}^n e^{\theta^T (X_i - x_\theta)} \mathbf{1}_{X_i \in A_n}}{\sum_{i=1}^n e^{\theta^T (X_i - x_\theta)}} \\ =& \Phi_n \left(\sum_{i=1}^n e^{\theta_n^T (x_\theta - X_i)} 1_{A_n}, \sum_{i=1}^n e^{\theta_n^T (x_\theta - X_i)} \right),\end{aligned}$$

where

$$\Phi(y, z) = \frac{y}{z}.$$

At this point, we verify the hypotheses of Lemma 14. We take $c_n = \|\theta_n\|$ and θ as in the lemma itself. Then, $X_{\theta_n} \to x_\theta$ in probability. Let $a_n = U^{-1}(\frac{1}{n})$, C_1 be as above, ν be as in equation 4 and D as chosen. All the hypotheses of Lemma 14 are easily verified.

Since Φ is a continuous mapping, and $\int e^{-C_1 \theta^T y} dPRM(\nu) \neq 0$ with probability 1 by the definition of a PRM, by Lemma 14 it follows that

$$\mathbb{P}(R_{n,\theta} \in A_n) = \Phi_n \left(\sum_{i=1}^n e^{c\theta^T (x_\theta - X_i)} 1_{A_n}, \sum_{i=1}^n e^{c\theta^T (x_\theta - X_i)} \right) \xrightarrow{d} \frac{\int e^{-C_1 y} 1_{y \in D} dPRM(\nu)}{\int e^{-C_1 y} dPRM(\nu)}.$$

Finally, note that

$$\mathbb{P}(R_{c,\theta} \in A_n) = \mathbb{P}(c\theta(x_\theta - R_{c,\theta}) \in c(x_\theta - A_n)) \approx \mathbb{P}(c\theta(x_\theta - R_{c,\theta}) \in C_1 D)$$

if n is large enough. Combining the two statements above, if Z is a random vector such that

$$\mathbb{P}(Z \in D) = \frac{\int e^{-C_1 y} 1_{y \in D} dPRM(\nu)}{\int e^{-C_1 y} dPRM(\nu)},$$

then

$$\mathbb{P}(c(x_\theta - R_{c,\theta}) \in D) \to \mathbb{P}(C_1 Z \in D).$$

Thus, the random variable $Z_{C,PRM} = C_1 Z$ is the desired limit. Note that this random variable depends upon the Poisson random measure, while the scaling limit Z from part (a) of the theorem does not. It follows that these two random variables are not the same, which concludes the proof.

We remark that, as in the one-dimensional case, the limiting random variable $Z_{C,PRM}$ is continuous, but do not prove this here. $\square$

THIS REQUIRES A PROPER PROOF, NOT JUST A PARAGRAPH. EXPLAIN IT MORE.

Note that this part of the theorem, unlike the previous two parts, is slightly different in the following way. Let $X_{(n)}$ denote the sample maximizer of $\theta^T y$ as in the previous sections. It turns out that while $\theta^T X_{(n)}$ scales to a Gumbel random variable, the component $X_{(n)} - \theta^T X_{(n)}$ orthogonal to θ requires no scaling at all, to converge to a standard normal random vector. We omit the details since they are superfluous, but remark that the estimator R_{n,θ_n}, therefore, doesn't admit a scaling limit in the traditional sense.

All we can show, is that

$$R_{n,\theta_n} - X_{(n)} \xrightarrow{d} 0,$$

which follows exactly as in the proof of Theorem 10(c).

G Experiments

In this section, we present a series of experiments. For simplicity, we get the tilting function to be the identity function, i.e., $g(x) = x$. For the sake of notation, θ will denote the tilting parameter and n will denote the number of samples, unless otherwise specified.[1]

G.1 Unbounded Random Variables

In this section we look at unbounded random variables and how accurately they can be twisted. We start by considering an exponential random variable.

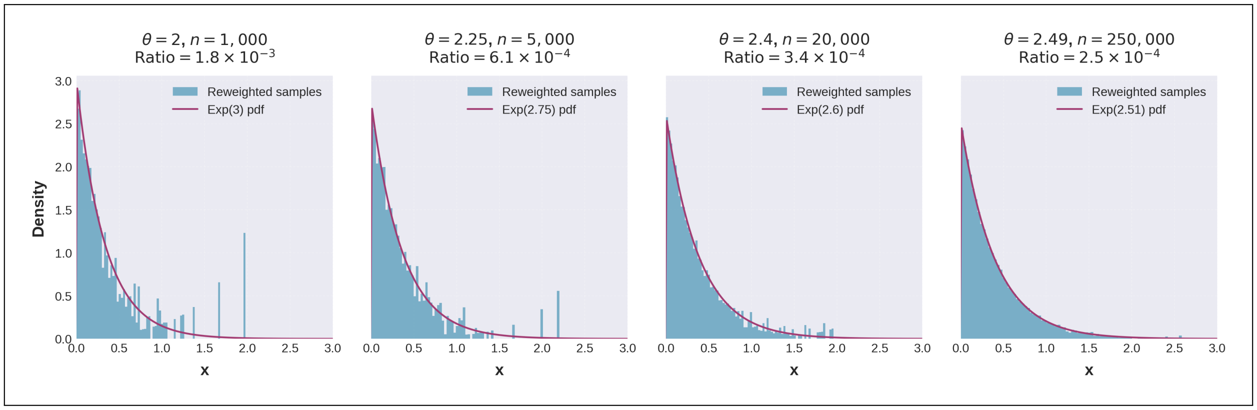

Figure 1: Exponential tilting of Exp(5) distribution with a sequence of (θ_i, n_i) s.t. $M_\theta/n \to 0$.

As Figure 1 shows, if $\frac{M_{\theta_n}}{n} \to 0$ then exponential tilting is asymptotically accurate, reflecting Theorem 2.

G.2 Bounded Random Variables

The next figures represent twists of bounded random variables in the Weibull regime.

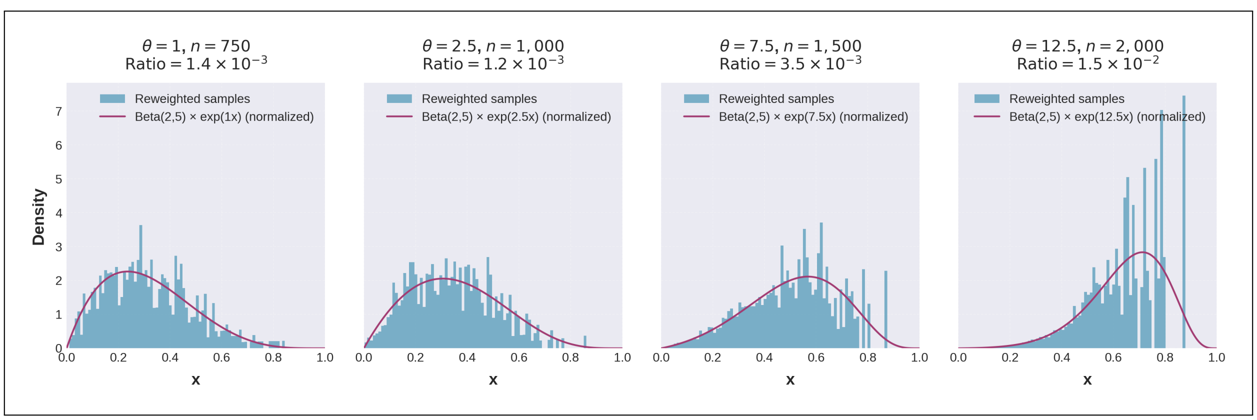

Figure 2: Exponential tilting of Beta(2, 5) distribution with a sequence of (θ_i, n_i) s.t. $M_\theta/n \not\to 0$.

As Figure 2 shows, if $\frac{M_{\theta_n}}{n} \to \infty$ then the sample maximizer receives higher empirical weight than the other samples, demonstrating Theorem 6(c).

[1]The code is available here

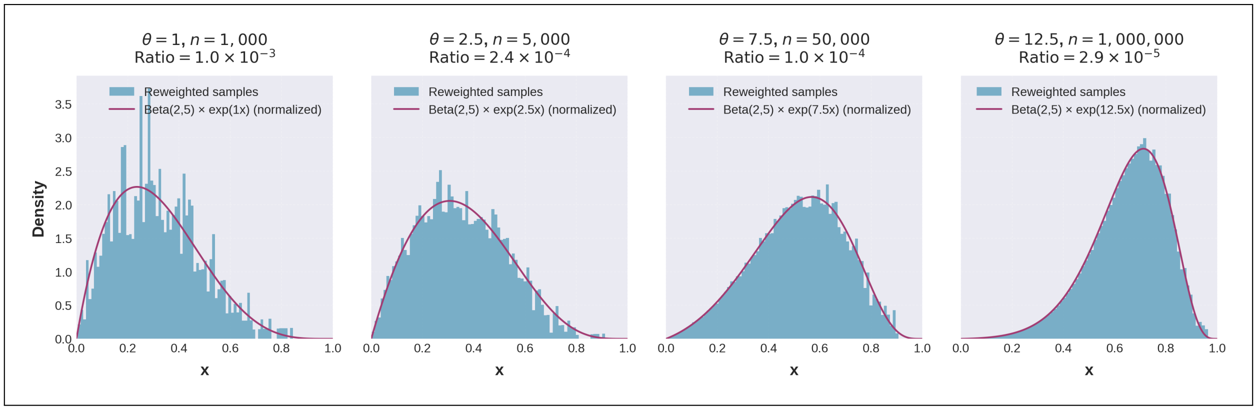

Figure 3: Exponential tilting of Beta(2, 5) distribution with a sequence of (θ_i, n_i) s.t. $M_\theta/n \to 0$.

Conversely, if $\frac{M_{\theta_n}}{n} \to 0$, then twisting does take place accurately in the Weibull regime in Fig 3, demonstrating Theorem 6(a). The next two figures, Fig 4 and Fig 5 demonstrate the same phenomena for exponentially tilting a uniform random variable.

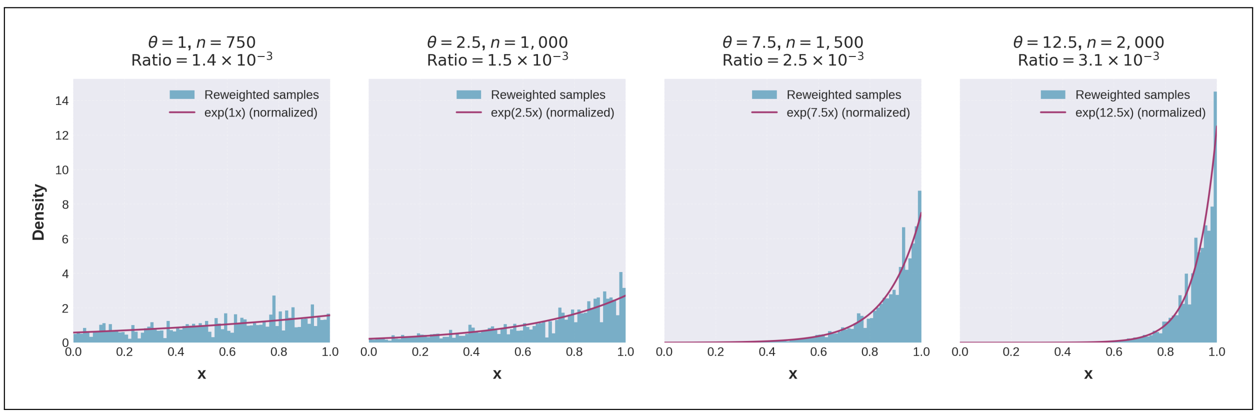

Figure 4: Exponential tilting of Uniform(0, 1) distribution with a sequence of (θ_i, n_i) s.t. $M_\theta/n \not\to 0$.

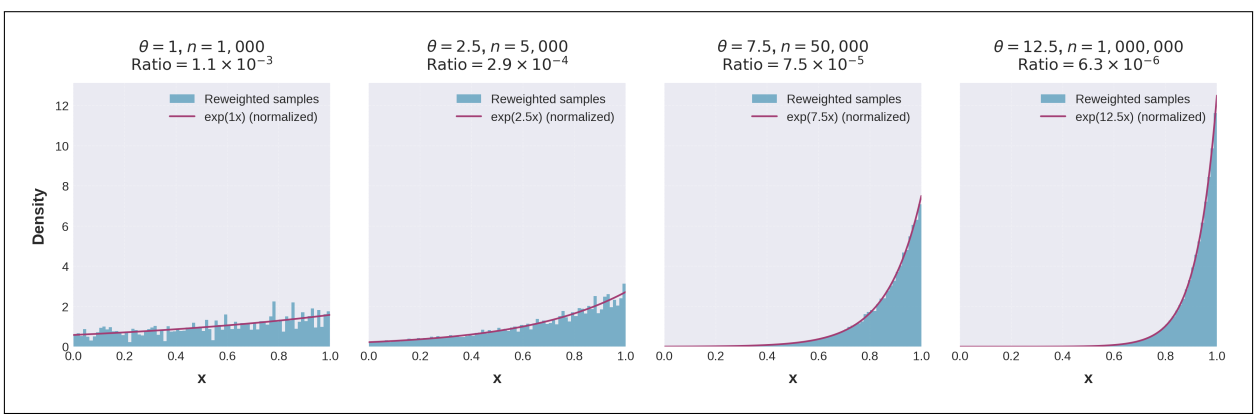

Figure 5: Exponential tilting of Uniform(0, 1) distribution with a sequence of (θ_i, n_i) s.t. $M_\theta/n \to 0$.

Finally, we demonstrate the Gamma scaling limit of the empirical random variable in Fig 6, thereby adding credence to Theorem 6(a).

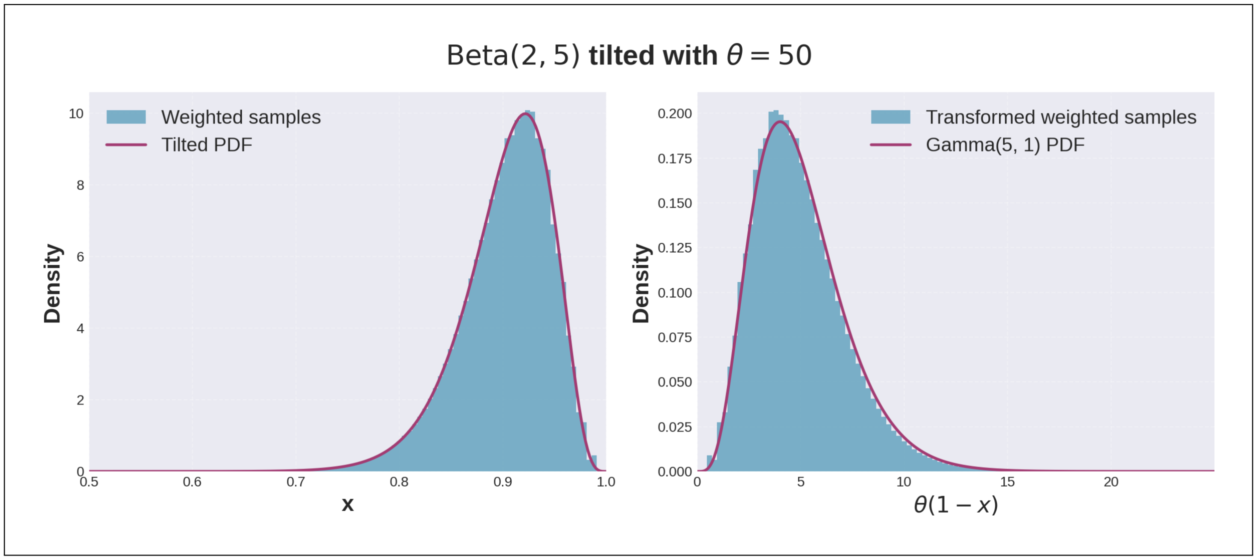

Figure 6: Exponential tilting of Beta(2, 5) distribution with $\theta = 50$. The PDF of the samples and true distribution is on the left. The transformed samples, given in theorem 5, with the PDF of $\Gamma(5, 1)$, the corresponding scaling limit.

References

Robert J. Adler. *The Geometry of Random Fields*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia, USA, 2010. doi: 10.1137/1.9780898718980. URL `https://doi.org/10.1137/1.9780898718980`.

Mayer Alvo. *Exponential Tilting and Its Applications*, pp. 171–193. Springer International Publishing, Cham, 2022. ISBN 978-3-031-06784-6. doi: 10.1007/978-3-031-06784-6_6. URL `https://doi.org/10.1007/978-3-031-06784-6_6`.

Søren Asmussen and Peter W. Glynn. *Stochastic Simulation: Algorithms and Analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2007. ISBN 978-0-387-30679-7. doi: 10.1007/978-0-387-69033-9.

Patrick Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. John Wiley & Sons, New York, 2 edition, 1999. ISBN 978-0-471-19745-4.

N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1987. ISBN 0521307872. doi: 10.1017/CBO9780511721434.

Amir Dembo and Ofer Zeitouni. *Applications-The Finite Dimensional Case*, pp. 71–114. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 978-3-642-03311-7. doi: 10.1007/978-3-642-03311-7_3. URL `https://doi.org/10.1007/978-3-642-03311-7_3`.

William Feller. *An Introduction to Probability Theory and Its Applications, Volume II*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, 2nd edition, 1971. ISBN 9780471257097. doi: 10.2307/3613733.

Peter Matthew Jacobs, Foad Namjoo, and Jeff M. Phillips. Efficient and stable multi-dimensional kolmogorov-smirnov distance, 2025. URL `https://arxiv.org/abs/2504.11299`.

S. Juneja and P. Shahabuddin. Chapter 11 rare-event simulation techniques: An introduction and recent advances. In Shane G. Henderson and Barry L. Nelson (eds.), *Simulation*, volume 13 of *Handbooks in Operations Research and Management Science*, pp. 291–350. Elsevier, 2006. doi: https://doi.org/10.1016/S0927-0507(06)13011-X. URL `https://www.sciencedirect.com/science/article/pii/S092705070613011X`.

Olav Kallenberg. *Foundations of Modern Probability*. Probability and Its Applications. Springer, New York, 2 edition, 2002. ISBN 978-0-387-95313-7.

JOHANNES LEDERER and SARA VAN DE GEER. New concentration inequalities for suprema of empirical processes. *Bernoulli*, 20(4):2020–2038, 2014. ISSN 13507265. URL http://www.jstor.org/stable/43590431.

Donald L. McLeish and Zhongxian Men. Extreme value importance sampling for rare event risk measurement. In *Lecture Notes in Computer Science*, volume 8976, pp. 319–331. Springer, 2015. doi: 10.1007/978-3-319-09114-3_18.

Francesco Ragone, Jeroen Wouters, and Freddy Bouchet. Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences*, 115(1):24–29, 2018. doi: 10.1073/pnas.1712645115. URL https://www.pnas.org/doi/10.1073/pnas.1712645115.

Sidney Resnick. Multivariate regular variation on cones: Application to extreme values, hidden regular variation and conditioned limit laws. *Stochastics An International Journal of Probability and Stochastic Processes*, 80, 12 2007. doi: 10.1080/17442500701830423.

Sidney I. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 1987. ISBN 9780387964812. doi: 10.1007/978-0-387-75953-1.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, New York, 1996. ISBN 0387946403.