

Finiteness of Redundancy, Regret, Shtarkov Sums, and Jeffreys Integrals in Exponential Families

Peter Grünwald and Peter Harremoës

Centrum Wiskunde & Informatica

Amsterdam, The Netherlands

Emails: Peter.Grunwald@cwi.nl and Peter.Harremoes@cwi.nl

Abstract—The normalized maximum likelihood (NML) distribution plays a fundamental role in the MDL approach to statistical inference. It is only defined for statistical families with a finite Shtarkov sum. Here we characterize, for exponential families, when the Shtarkov sum is finite. This turns out to be the case if and only if the minimax redundancy is finite, thus extending the reach of our results beyond the individual-sequence setting. In practice, the NML/Shtarkov distribution is often approximated by the Bayesian marginal distribution based on Jeffreys’ prior. One serious problem is that in many cases Jeffreys’ prior cannot be normalized. It has been conjectured that Jeffreys’ prior cannot be normalized in exactly the cases where the Shtarkov sum is infinite, i.e. when the minimax redundancy and regret are infinite. We show that the conjecture is true for a large class of exponential families but that there exist examples where the conjecture is violated.

I. PRELIMINARIES

A. Main Definitions

We let $\{P_\beta \mid \beta \in \Gamma^{\text{can}}\}$ be a 1-dimensional exponential family given in a canonical parameterization,

$$\frac{dP_\beta}{dQ} = \frac{1}{Z(\beta)} e^{\beta x}, \quad (1)$$

where Z is the partition function $Z(\beta) = \int \exp(\beta x) dQx$, and $\Gamma^{\text{can}} := \{\beta \mid Z(\beta) < \infty\}$ is the *canonical parameter space*. Note that we allow the measure Q to have both discrete and continuous components.

We let $\beta_{\text{sup}} = \sup\{\beta \mid \beta \in \Gamma^{\text{can}}\}$, and β_{inf} likewise. The trivial case where Γ^{can} has no interior points is excluded from the analysis. See [1] for more details on exponential families.

The elements of the exponential family are also parameterized by their mean value μ . We write μ_β for the mean value corresponding to the canonical parameter β and β_μ for the canonical parameter corresponding to the mean value μ . Note that we allow infinite values of the mean. The element in the exponential family with mean μ is denoted P^μ . The range of $\beta \rightarrow \mu_\beta$ is denoted M . We write $\mu_{\text{sup}} = \sup M$, and $\mu_{\text{inf}} = \inf M$. If Q has a point mass at $\mu_{\text{inf}} > -\infty$ and the support of Q is a subset of $[\mu_{\text{inf}}, \infty[$, then the exponential family is extended by the element $P_{-\infty} = P^{\mu_{\text{inf}}} = \delta_{\mu_{\text{inf}}}$, and likewise the exponential family is extended if Q has a point mass in $\mu_{\text{sup}} < \infty$ and the support of Q is a subset of $]-\infty, \mu_{\text{sup}}]$. For any x the maximum likelihood distribution is P_{β_x} . The *Shtarkov integral* S is defined as

$$S = \int \frac{1}{Z(\beta_x)} e^{\beta_x x} dQx. \quad (2)$$

If Q is discrete the integral reduces to a sum. For this reason, it is often called the *Shtarkov sum*.

The variance function V is the function that maps $\mu \in M$ into the variance of P^μ . If M has interior points then the exponential family is uniquely determined by its variance function. The Fisher information of an exponential family in its canonical parametrization is $I_\beta = V(\mu_\beta)$ and the Fisher information of the exponential family in its mean value parametrization is $I^\mu = (V(\mu))^{-1}$. The *Jeffreys integral* J is defined as

$$J = \int_{\Gamma^{\text{can}}} I_\beta^{1/2} d\beta = \int_M (I^\mu)^{1/2} d\mu. \quad (3)$$

More on Fisher information can be found in [2].

In general, the finiteness of the Shtarkov integral and of the Jeffreys integral relative to the full exponential family defined by Γ^{can} is determined by local behavior both at β_{sup} and β_{inf} , either of which may or may not be (minus) infinity. Thus, there are two relevant limit points. For ease of exposition we will restrict ourselves to families where only one limit point is relevant, by picking an arbitrary $\beta^\circ > \beta_{\text{inf}}$ in the interior of Γ^{can} and considering the subfamily with parameters $\{\beta \in \Gamma^{\text{can}} : \beta^\circ \leq \beta \leq \beta_{\text{sup}}\}$. We call an exponential family of this form *left-truncated*. This ensures that the contribution to the Jeffreys and Shtarkov integrals on the left part of the interval is finite. For such families, the Shtarkov integral becomes $S = \int Z^{-1}(\beta(x)) \exp(\beta(x)) dQx$, with β the ML estimator, $\hat{\beta}(x) = \beta_x$ if $x \geq \mu_{\beta^\circ}$, $\hat{\beta}(x) = \beta^\circ$ if $x < \mu_{\beta^\circ}$.

For our problem it is natural to work with *extended exponential families* as defined in [3]. For a probability distribution Q on \mathbb{R}^d the convex support $cs(Q)$ is the intersection of all convex closed sets that have Q -probability 1. The convex core $cc(Q)$ is the intersection of all convex measurable sets with Q -probability 1, [4]. We have $cc(Q) \subseteq cs(Q)$. An extreme point x in $cs(Q)$ belongs to $cs(Q)$ if and only if $Q(x) > 0$. In its mean value parametrization the exponential family based on a measure with bounded support has a natural extension to $cc(Q)$. In particular δ_x belongs to the extended exponential family if Q has a point mass in x and x is an extreme point of $cs(Q)$.

B. Relevance of the problem

We consider the information channel from Γ^{can} to \mathbb{R} . The capacity of this channel is given by the minimax (expected)

redundancy that equals the maximin redundancy. The capacity tells us how much information the data in \mathbb{R} can at most reveal about the parameter β . The redundancy also tells us how many extra bits we need on average to code data if the parameter β is unknown and the average is calculated with respect to P_β compared with the situation where the parameter β is known. If we are interested in the additional code length for individual sequences rather than on average, then the relevant notion becomes the (individual sequence) regret. The *Shtarkov* or *NML distribution* given by

$$\frac{dQ_{\text{Shtarkov}}}{dQ}(x) = \frac{e^{\beta_x \cdot x}}{Z(\beta_x) \cdot S}.$$

It is well-defined iff S is finite, and in that case, it achieves the minimax regret. In fact, the minimax individual sequence regret is always equal to $\log S$ and hence finite iff S is. Henceforth, as is customary, “redundancy” refers to “expected redundancy” and “regret” refers to individual sequence regret. See, for example, [2] for precise definitions.

In statistics we often consider independent data all modelled by the same parameter β so that P_β is replaced by P_β^n . Exponential families have the nice property that if (Γ^{can}, Q) is an exponential family then $(\Gamma^{\text{can}}, Q^n)$ is an exponential family. Further (Γ^{can}, Q) has finite minimax redundancy if and only if $(\Gamma^{\text{can}}, Q^n)$ has finite minimax redundancy, and similarly for regret. If the sample space is infinite, the minimax regret is infinite in many cases, but hitherto it was unknown exactly what these cases are; only some examples were known, such as the Gaussian, Poisson and geometric families. Because, if it is finite, then the Shtarkov distribution is well-defined and can serve as the basis of MDL model selection, it is of obvious interest to determine which exponential families have finite minimax regret and which have finite minimax redundancy. We address this question in Section II. As we will see, the minimax regret is finite iff the minimax redundancy is finite, so that finiteness of the Shtarkov integral is also relevant if one is interested in average-case rather than individual sequence-coding.

As first established by [5], if the parameter space is restricted to a compact subset of the interior of the parameter space with non-empty interior (called an *ineccsi* set in [2]), then the minimax regret is finite and equal to the logarithm of the Shtarkov integral, which in turn is equal to

$$\frac{k}{2} \log \frac{n}{2\pi} + \log J + o(1). \quad (4)$$

It thus becomes quite relevant to investigate whether the same thing still holds if the parameter spaces are *not* restricted to an *ineccsi* set. Moreover, the same asymptotic regret (4) is achieved by the Bayesian marginal distribution equipped with Jeffreys prior with density $w(\theta) = I_\theta^{1/2}/J$. Instead of the Shtarkov distribution (often hard to calculate) we can also essentially achieve minimax regret by using the more convenient Bayesian marginal distribution. It thus becomes, again, quite relevant to investigate whether the same thing still holds if the parameter spaces are *not* restricted to an *ineccsi*

set. To answer this question, one first needs to know when the Jeffreys integral is finite. This is the second question we address. Hitherto, very little was known about this situation for infinite sample spaces. For all the examples previously considered in the literature, either Jeffreys’ integral and the minimax redundancy and the Shtarkov integral are all finite or all infinite. This leads one to conjecture that this holds in general. Whether or not this is so is discussed at length and posed as an open problem in [2, Chapter 11, Section 11.1]. Here we resolve the issue for one-dimensional exponential families. As we will see, the conjecture is “almost”, but not quite, true: there are intriguing counterexamples.

II. WHEN IS SHARKOV FINITE?

Theorem 1: For a 1-dimensional left-truncated exponential family, the following statements are all equivalent:

- 1) The Shtarkov integral is finite.
- 2) The minimax individual-sequence regret is finite.
- 3) The minimax expected redundancy is finite.
- 4) The exponential family has a dominating distribution Q_{dom} in terms of information divergence, i.e.,

$$\sup_{\beta \in \Gamma^{\text{can}}} D(P_\beta \| Q_{\text{dom}}) < \infty.$$

- 5) There is distribution P_β with $\beta \in \Gamma^{\text{can}}$ that dominates the exponential family in terms of information divergence.
- 6) The information channel $\beta \rightarrow P_\beta$ has finite capacity.
- 7) There exists $\beta_0 \in \Gamma^{\text{can}}$ such that

$$\lim_{\beta \uparrow \beta_{\text{sup}}} D(\beta_0 \| \beta) < \infty \text{ or } \lim_{\beta \uparrow \beta_{\text{sup}}} D(\beta \| \beta_0) < \infty.$$

Example 2: The family of Poisson distributions restricted to have mean $\mu \geq \mu^\circ$ for some $\mu^\circ > 0$, the family of geometric distributions, the family of exponential distributions and the normal location family under the same restriction all have infinite Shtarkov integral. Indeed, an easy calculation shows that for all these families, both information divergences in item (7) tend to infinity for all $\beta_0 \in \Gamma^{\text{can}}$. The Bernoulli family has finite Shtarkov integral, and indeed, taking β_0 such that $P_{\beta_0}(X = 1) = P_{\beta_0}(X = 0) = 1/2$ shows that the second condition in (7) holds. More surprisingly, we will see below that (6) also holds for some practically meaningful exponential families with unbounded support, such as the inverse Gaussian family.

Most of the equivalences between (1)–(6) are quite straightforward, although it may be mildly surprising that, if a dominating distribution Q_{dom} exists at all, then it must reside in the exponential family (Γ^{can}, Q) . The surprising part is the fact that statements (1)–(6) are also equivalent to (7). We will now give a more concrete characterization of the two cases mentioned in (7), that allows us to check, for each given exponential family, whether or not it satisfies (7). We call the first case the β -case and the second the μ -case, because, to some extent, the β - and μ -parameters play a dual role in them.

A. β -Case

The family (Γ^{can}, Q) is of β -type if there exists $\beta_0 \in \Gamma^{\text{can}}$ such that $\lim_{\beta \uparrow \beta_{\text{sup}}} D(\beta_0 \| \beta) < \infty$.

Theorem 3 (β -case): 1) Suppose (Γ^{can}, Q) is of β -type.

Then $\beta_{\text{sup}} < \infty$, $\beta_{\text{sup}} \in \Gamma^{\text{can}}$, and $P_{\beta_{\text{sup}}}$ dominates (Γ^{can}, Q) .

2) Conversely, suppose that $\beta_{\text{sup}} < \infty$ and $\beta_{\text{sup}} \in \Gamma^{\text{can}}$. Then (Γ^{can}, Q) is of β -type.

Part (1) of the theorem implies that $P_{\beta_{\text{sup}}}$ actually dominates the family in the sense of Theorem 1. Concerning part (2), note that, assuming $\beta_{\text{sup}} < \infty$, we may change the parametrization (Γ^{can}, Q) by redefining Q as $Q' = P_{\beta_{\text{sup}}}$. Denoting parameters of the new parameterization as β' , we have $\beta'_{\text{sup}} = 0$ and the underlying measure Q' is now a probability measure.

If the family (Γ^{can}, Q) is such that Q is a probability measure with unbounded support from above (i.e. $\sup\{x : x \in \text{supp}(Q)\} = \infty$, and Q has a sufficiently fat tail, then $Z(0) = 1$ and $Z(\beta) = \infty$ for all $\beta > 0$. For example, if Q has density q relative to Lebesgue measure, this is the case if for all $\beta > 0$, $\exp(\beta x)q(x) \rightarrow \infty$, so for example, if $q(x)$ decreases polynomially or as $\exp(-x^{1/2})$. Part (2) thus implies:

Corollary 4: If the family (Γ^{can}, Q) can be represented by a Q' that is a probability measure with unbounded support from above and fat tails in the sense above, then the family is of β -type and the Shtarkov integral is finite.

Example 5 (The inverse Gaussian distribution): If X is a standard Gaussian random variable then the distribution of $1/X^2$ has density (relative to Lebesgue measure)

$$\frac{\exp\left(-\frac{1}{2x}\right)}{(2\pi x^3)^{1/2}}$$

for $x \in [0; \infty[$. The elements in the corresponding exponential family have densities

$$\frac{\exp\left(-\frac{1}{2x}\left(\frac{x}{\mu} - 1\right)^2\right)}{(2\pi x^3)^{1/2}}$$

where $\mu \in]0; \infty[$ is the mean value. We shall restrict our attention to the subfamily with $\mu \in]1; \infty[$. By setting $\beta = -\mu^{-2}/2$, we see that this is a left-truncated exponential family with underlying measure with density $q(x) = e^{-1/2x}(2\pi x^3)^{-1/2}$. Since q has a polynomial tail, we can apply Corollary 4 to conclude that the Shtarkov integral is finite. Indeed, in this case we can also calculate the integral directly, which gives

$$\int_1^\infty \frac{\exp\left(-\frac{1}{2x}\left(\frac{x}{\mu} - 1\right)^2\right)}{(2\pi x^3)^{1/2}} dx = \int_1^\infty \frac{1}{(2\pi x^3)^{1/2}} dx < \infty.$$

B. μ -Case

A left-truncated family (Γ^{can}, Q) is of μ -type if there exists $\beta_0 \in \Gamma^{\text{can}}$ such that $\lim_{\beta \uparrow \beta_{\text{sup}}} D(\beta_0 \| \beta) < \infty$. It will be convenient to work with right-truncated instead of left-truncated families, i.e. restricted to parameters $\{\beta \in \Gamma^{\text{can}} : \beta_{\text{inf}} \leq \beta \leq$

$\beta^\circ\}$ for some $\beta^\circ < \beta_{\text{sup}}$. This can always be achieved by considering the random variable $-X$ rather than X . We say that a right-truncated family (Γ^{can}, Q) is of μ -type if there exists $\beta_0 \in \Gamma^{\text{can}}$ such that $\lim_{\beta \downarrow \beta_{\text{inf}}} D(\beta_0 \| \beta) < \infty$, or equivalently, there exists $\mu_0 \in M$ such that $\lim_{\mu \downarrow \mu_{\text{inf}}} D(\mu_0 \| \mu) < \infty$. The definition of β -type is adjusted accordingly.

Theorem 6 (μ -case): 1) Suppose the right-truncated family (Γ^{can}, Q) is of μ -type but not of β -type. Then $\mu_{\text{inf}} > -\infty$, $Q(X = \mu_{\text{inf}} > 0)$, the extension of family (Γ^{can}, Q) contains element $\beta_{\text{inf}} = -\infty$ which puts all its mass on $X = \mu_{\text{inf}}$, and the support of Q is bounded from below by μ_{inf} , and every element in Γ^{can} (but not its extension) dominates (Γ^{can}, Q) .

2) Suppose that $\mu_{\text{inf}} > -\infty$ and $Q(X = \mu_{\text{inf}} > 0)$. Then (Γ^{can}, Q) is of μ -type.

Concerning part (1), note that in the β -case, there is a single element $P_{\beta_{\text{sup}}} \in \partial(\Gamma^{\text{can}})$ which dominates (Γ^{can}, Q) ; in general, the other elements of Γ^{can} do not dominate (Γ^{can}, Q) . In the μ -case, the situation is reversed: all elements of Γ^{can} dominate (Γ^{can}, Q) , but in general, the limiting distribution $P_{\beta_{\text{inf}}}$ does not.

Concerning part (2), note that it is a partial converse of part (1). Together with part (1), it implies a useful corollary:

Corollary 7: Suppose that the underlying measure Q has support bounded from below, and let $x_{\text{min}} \in \mathbb{R}$ be the smallest x in the support. Then the family is of μ -type (and the Shtarkov integral is finite) if and only if $Q(X = x_{\text{min}}) > 0$.

Example 8 (The uniform distribution and a modification): Let Q denote the uniform distribution on $[-1; 1]$. Then the partition function of the corresponding exponential family is

$$Z(\beta) = \int_{-1}^1 \frac{1}{2} \exp(\beta x) dx = \frac{\sinh(\beta)}{\beta}.$$

In the full paper we give a straightforward but extensive calculation which shows that the Shtarkov integral is infinite.

Now let Q be a $1/2-1/2$ mixture of the uniform distribution on $[-1; 1]$ and the distribution with point masses $1/2$ in -1 and 1 . Then, by Corollary 7, the Shtarkov integral is finite.

III. WHEN IS JEFFREYS FINITE?

Theorem 9: Let $(\Gamma_0^{\text{can}}, Q)$ represent a left-truncated exponential family. If the Shtarkov integral is infinite, then the Jeffreys integral is infinite.

We note that the condition of left-truncatedness is really necessary in this theorem. We show in the full paper that if, for example, we allow the parameter set to consist of a countably infinite number of disjoint intervals, we can construct a family for which Jeffreys is finite and Shtarkov is infinite.

We might expect that, if we impose the left-truncatedness condition, the converse of Theorem 9 would hold as well, making finiteness of Jeffreys and Shtarkov equivalent. But in fact, the converse does not hold in general. To study this further, let us assume that the Shtarkov integral is finite. Then, by Theorem 1, we are either in the β -case or in the μ -case.

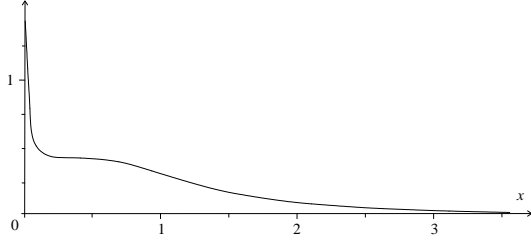


Fig. 1. Density of the exponentiated Cauchy distribution.

A. μ -case

In most cases finite Shtarkov implies finite Jeffreys.

Theorem 10: Let Q be a measure on the real line with support I . Assume that μ_{inf} is the left end point of I . If Q has density $f(x) = (x - \mu_{\text{inf}})^{\gamma-1} g(x)$ in an interval just to the right of a where g is an analytic function and $g(\mu_{\text{inf}}) > 0$ then the left end of the interval I gives a finite contribution to Jeffrey's integral if and only if Q has a point mass in a .

Proof: The variance function can be approximated by $V(\mu) \approx c_0 (\mu - \mu_{\text{inf}})^p$, where $p = 2$ if there is no point mass in μ and $p < 2$ if there is a point mass in μ_{inf} [6, Thm. 4.4]. Therefore the integrand in the Jeffreys integral can be approximated by $c_0^{-p/2} (x - \mu_{\text{inf}})^{-1/2}$ near μ_{inf} so the left endpoint gives a finite contribution to the Jeffreys integral if and only if $p < 2$. ■

B. β -case

Proposition 11: Let $(\Gamma_0^{\text{can}}, Q)$ represent a left-truncated exponential family such that $\beta_{\text{sup}} = 0$ and Q admits a density q either with respect to Lebesgue measure or counting measure. If $q(x) = O(1/x^{1+\alpha})$ for some $\alpha > 0$, then the Jeffreys integrand satisfies $I(\beta)^{1/2} = O((Z''(\beta))^{1/2}) = (-\beta)^{-1+\alpha/2}$, so that the Jeffreys integral $\int_{\beta}^0 I(\gamma)^{1/2} d\gamma$ is finite.

Example 12 (The inverse Gaussian, Cont.): The family of inverse Gaussian distribution has variance function $V(\mu) = \mu^3$. Thus the Jeffrey integral is finite, as expected:

$$\int_1^{\infty} \left(\frac{1}{\mu^3}\right)^{1/2} d\mu = \int_1^{\infty} \mu^{-3/2} d\mu = 2.$$

C. Counterexample

If Y is a Cauchy distributed random variable then $X = \exp(Y)$ has density

$$\frac{1}{\pi} \frac{1}{x(1 + \log^2(x))}.$$

A probability measure Q is defined as a 1/2 and 1/2 mixture of a point mass in 0 and an exponentiated Cauchy distribution. We consider the exponential family based on Q . The partition function is

$$Z(\beta) = \frac{1}{2} + \frac{1}{2\pi} \int_0^{\infty} \frac{\exp(\beta x)}{x(1 + \log^2(x))} dx, \quad \beta \leq 0.$$

We note that $1/2 \leq Z(\beta) \leq 1$ for all $\beta \leq 0$. Then

$$D(Q_{\beta} \| Q) \leq D(Q_{-\infty} \| Q) = 1 \text{ bit}.$$

Therefore the minimax redundancy is at most 1 bit.

The mean value μ as a function of β is

$$\mu \leq 2Z'(\beta) = \frac{1}{\pi} \int_0^{\infty} \frac{\exp(\beta x)}{(1 + \log^2(x))} dx \leq \frac{1}{|\beta|}.$$

The variance as a function of β can be lower bounded as follows:

$$\begin{aligned} I_{\beta} &= \frac{\frac{1}{2}\mu^2 + \frac{1}{2\pi} \int_0^{\infty} (x - \mu)^2 \frac{\exp(\beta x)}{x(1 + \log^2(x))} dx}{Z(\beta)} \\ &\geq \frac{1}{Z(\beta) 2\pi} \int_{2/|\beta|}^{3/|\beta|} \left(x - \frac{1}{|\beta|}\right)^2 \frac{\exp(\beta x)}{x(1 + \log^2(x))} dx \\ &\geq \frac{1}{6\pi e^3} \frac{1}{\beta^2 \left(1 + \log^2\left(\frac{|\beta|}{3}\right)\right)}. \end{aligned}$$

We have $\frac{(1 + \log^2(\frac{|\beta|}{3}))}{\log^2|\beta|} \rightarrow 1$ for $\beta \rightarrow -\infty$ or for $\beta \rightarrow 0$.

Therefore there exists a constant $c > 0$ such that $I_{\beta}^{1/2} \geq c \cdot \frac{1}{|\beta| \log|\beta|}$ for $|\beta|$ sufficient large or for $|\beta|$ sufficiently small. Hence both the left- and the right-truncated exponential families have finite redundancy but infinite Jeffreys integral.

IV. TWEEDIE DISTRIBUTIONS

An interesting class of distributions are the so-called Tweedie distributions. A Tweedie distribution is given by the parameters p, μ and σ^2 where μ is the mean. For fixed p and σ^2 we get an exponential family with variance function

$$V(\mu) = \sigma^2 \mu^p.$$

Some p have the property that if we restrict the exponential family to a compact set then, Jeffreys and Shtarkov are, in a sense, proportional to each other. Important examples are the Gaussian location family ($p = 0$), the gamma distributions including the exponential distribution ($p = 2$), and the inverse Gaussian distribution ($p = 3$). The convex core is \mathbb{R} for the Gaussian distribution. The convex core is $[0; \infty[$ for the Poisson distribution and other Tweedie distributions with $p \in [1; 2[$. The convex core is $]0; \infty[$ for the exponential distributions, the gamma distributions and the inverse Gaussian distributions and other distributions with $p \in [2; \infty[$. In each of these cases Jeffrey's prior is proportional to $x^{-p/2}$. We see that Jeffrey's prior cannot be normalized on the whole parameter space. For $p > 2$ Jeffrey's prior can be normalized if we restrict to a closed subset of the parameter space. For $p \leq 2$ Jeffrey's prior can be normalized if we restrict to a compact subset of the parameter space.

For the Tweedie distributions with $p = 0$ or $p = 2$ or $p = 3$ the integrand in the Shtarkov integral is proportional to $x^{-p/2}$. For the Poisson family ($p = 1$) the Jeffreys integrand and the Shtarkov summand have different domains, but they behave essentially in the same way near infinity.

TABLE I

FINITENESS RESULTS FOR 1-DIMENSIONAL EXPONENTIAL FAMILIES

Examples	Support	Feature	$\mu_{\text{sup}} < \infty$	$\beta_{\text{sup}} < \infty$	cl.?	Shtar.	Jef.	Case
uniform	bounded	no mass at x_{max} (NS)	$< \infty$	$< \infty$	cl.?	$< \infty$	$< \infty$	$< \infty$
Exp. Cauchy + point mass	bounded	mass at x_{max} (NS)	$< \infty$	$< \infty$	cl.?	$< \infty$	$< \infty$	$< \infty$
Bernoulli; Ex. 8	bounded	mass at x_{max} (NS)	$< \infty$	$< \infty$	cl.?	$< \infty$	$< \infty$	$< \infty$
Poisson, geometric	unbounded	$\forall \beta > 0 : \exp(\beta x)q(x) \rightarrow 0$ (NS)	$< \infty$	$< \infty$	cl.?	$< \infty$	$< \infty$	$< \infty$
exponential distributions	unbounded	$\int q(x)dx = \infty$ (NS)	$< \infty$	$< \infty$	cl.?	$< \infty$	$< \infty$	$< \infty$
Not in this note.	unbounded	$\int q(x)dx < \infty, \forall \alpha > 0 : q(x)/x^\alpha \rightarrow 0$ (NS)	$< \infty$	$< \infty$	cl.?	$< \infty$	$< \infty$	$< \infty$
inverse Gaussian	unbounded	$\int q(x)dx < \infty, \exists \alpha > 0 : q(x)/2^\alpha > 0$ (NS)	$< \infty$	$< \infty$	cl.?	$< \infty$	$< \infty$	$< \infty$
nonsteep families	unbounded	$\int xq(x)dx < \infty$ (NS)	$< \infty$	$< \infty$	cl.?	$< \infty$	$< \infty$	$< \infty$

Any left-truncated 1-dimensional exponential family belong to exactly one row in the table. (Un)bounded means “un(bounded) from above”. “cl.” means “closable”, i.e. that $p_{\beta_{\text{sup}}}$ is a member of the (potentially extended) family. x_{max} is the supremum of the support of Q . The example in the third row is in fact for the (essentially equivalent) right-truncated case, with all sup replaced by inf and v.v. The third column lists necessary and sufficient conditions on Q in the special case where Q has a density q with respect to the Lebesgue measure. “S” indicates that the condition is sufficient, i.e. if it holds for a simple family, then the family must be in the corresponding row. “N” indicates that the condition is necessary, i.e. it must hold for all simple families in the row. The “=” and “<” indicates that the (in)equality holds if a parameterization is chosen for which $\beta_{\text{sup}} = 0$.

V. BOUNDED SUPPORT IN MORE DIMENSIONS

A *polytope* is the convex hull of finitely many points. A point \bar{x} in the convex set F is *exposed* if there exists an affine function $f : F \rightarrow \mathbb{R}$ such that $f(\bar{x})$ is maximal and such that $f(\bar{x}) = f(\bar{y}), \bar{y} \in F$ implies $\bar{y} = \bar{x}$. All exposed points are extreme points. For any exposed point \bar{x} of $cs(Q)$ the distribution $\delta_{\bar{x}}$ is in the weak closure of the exponential family generated by Q . Here we consider exponential families with sufficient statistic $\bar{x} \in \mathbb{R}^d$ and canonical parameter $\bar{\beta}$.

Lemma 13: If $P_n \rightarrow \delta_{\bar{x}}$ in the weak topology and $n \rightarrow D(P_n \| Q_{\text{dom}})$ is bounded then $\liminf P_n(\bar{x}) > 0$.

Theorem 14: If the minimax redundancy of the exponential family based on Q is finite then $cs(Q)$ has finitely many extreme points each with a positive point probability.

Proof: Let \bar{x} be an exposed point in $cs(Q)$. Let β_n denote a sequence such that $P_{\beta_n} \rightarrow \delta_{\bar{x}}$ in the weak topology. If the redundancy is finite there exists a distribution Q_{dom} and a constant K such that $D(P_{\beta} \| Q_{\text{dom}}) \leq K$ for all P_{β} in the exponential family. Then P_{β_n} eventually has a point mass in \bar{x} and therefore Q also has a point mass in \bar{x} . Thus, for any exposed points \bar{x}_0 in $cs(Q)$ we have $Q(\bar{x}_0) > 0$. Further we have $D(\delta_{\bar{x}_0} \| Q_{\text{dom}}) = -\log(Q_{\text{dom}}(\bar{x}_0)) \leq K$. Assume that there exist an infinite sequence $\bar{x}_1, \bar{x}_2, \dots$ of exposed points in $cs(Q)$. Then $\sum_{i=1}^{\infty} Q_{\text{dom}}(\bar{x}_i) \leq 1$ which implies that $Q_{\text{dom}}(\bar{x}_i) \rightarrow 0$ for $i \rightarrow \infty$. Hence $-\log(Q_{\text{dom}}(\bar{x}_i)) \rightarrow \infty$ for $i \rightarrow \infty$ and we have a contradiction. Therefore $cs(Q)$ has finitely many exposed points. The set of exposed points is dense in the set of extreme points [7, p. 81, Remark 7 (b)] so the the extreme points are exposed. ■

Theorem 15: If an extended exponential family is restricted to a subset of a polytope in the mean value parametrization, then the Shtarkov integral and the minimax and maximin redundancies are finite.

Proof: The function $\bar{x} \rightarrow \frac{\exp(\hat{\beta} \cdot \bar{x})}{Z(\hat{\beta})}$ is convex and convex functions are bounded on polytopes. Thus the Shtarkov

integral is finite. The Shtarkov integral upper bounds the minimax and maximin redundancies. ■

By using that the regret is lower bounded by the mean redundancy we now get the following theorem. ∞ μ

Theorem 16: Let \mathcal{Q} denote a probability measure μ with bounded support. Then the following conditions are equivalent. 1) The convex core $cc(\mathcal{Q})$ is a polytope. 2) The minimax regret of the exponential family is finite. 3) The minimax redundancy of the exponential family is finite.

4) The information channel $\beta \rightarrow P_{\beta}$ has finite capacity.

VI. DISCUSSION

In this short note we have entirely focused on the question of whether the quantities in question are finite or infinite. We also had to exclude most of the proofs, examples and a more detailed discussion of how restricting the parameter space will effect the results (only for one-dimensional exponential families this question is trivial). Here we shall just mention some further results on the asymptotic behavior of left-truncated families. If the minimax regret is finite at all, then it is upper bounded by $\log n + O(1)$. If moreover Jeffreys prior exists, then under some further regularity conditions, we obtain the same asymptotics $(1/2) \log n + O(1)$ as for parameters restricted to inecsi sets. However, if the Jeffreys integral is infinite but the minimax regret is finite, then these asymptotics do not hold and the minimax regret must behave as $(1/2) \log n + f(n)$ for a function $f(n) \rightarrow \infty$.

ACKNOWLEDGEMENT

The authors thanks Flemming Topsøe, Shaul Bar-Lev, Bent Jørgensen, and Tim van Erven for discussions and comments. This work has been supported by the European Pascal Project.

REFERENCES

- [1] O. Barndorff-Nielsen, *Information and Exponential Families in Statistical Theory*. New York: John Wiley, 1978.
- [2] P. Grünwald, *the Minimum Description Length principle*. MIT Press, 2007.
- [3] I. Csiszár and F. Matúš, “Information projections revisited,” *IEEE Trans. Inform. Theory*, vol. 49, pp. 1474–1490, June 2003.
- [4] I. Csiszár and F. Matúš, “Convex cores of measures on \mathbb{R}^d ,” *Studia Sci. Math. Hungar.*, vol. 38, pp. 177–190, 2001.
- [5] J. Rissanen, “Fisher information and stochastic complexity,” *IEEE Trans. Inform. Theory*, vol. 42, no. 1, pp. 40–47, 1996.
- [6] B. Jørgensen, *The Theory of Dispersion Models*. Chapman & Hall, 1997.
- [7] M. M. Day, *Normed Linear Spaces*. Berlin: Springer Verlag, 1958.