Regret and Jeffreys Integrals in Exp. Families

Peter Grünwald and Peter Harremoës
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
Emails: Peter.Grunwald@cwi.nl and Peter.Harremoes@cwi.nl

I. Preliminaries

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

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

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*. We let $\beta_{\sup} = \sup\{\beta \mid \beta \in \Gamma^{\text{can}}\}$, and $\beta_{\inf}$ likewise.

The elements of the exponential family are also parametrized 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. For any x the maximum likelihood distribution is P_{β_x}. The *Shtarkov integral* S is defined as

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

The variance function V is the function that maps $\mu \in M$ into the variance of P^μ. 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

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

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

As first established by [3], 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

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

It thus becomes quite relevant to investigate whether the same thing still holds if the parameter spaces are *not* restricted to an ineccsi set. Whether or not this is so is discussed at length and posed as an open problem in [2, Chapter 11, Section 11.1].

II. Results

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 $\beta \to P_\beta$ has finite capacity.
7) There exists $\beta_0 \in \Gamma^{\text{can}}$ such that

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

Most of the equivalences between (1)–(6) are quite straightforward. The surprising part is the fact that statements (1)–(6) are also equivalent to (7).

Theorem 2: Let $(\Gamma_0^{\text{can}}, Q)$ represent a left-truncated exponential family. If the Shtarkov integral is infinite, then the Jeffreys integral is infinite.

The converse does not hold in general.

Theorem 3: Let $(\Gamma_0^{\text{can}}, Q)$ represent a left-truncated exponential family such that $\beta_{\sup} = 0$ 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 integral $\int_\beta^0 I(\gamma)^{1/2}\ d\gamma$ is finite.

In most cases finite Shtarkov implies finite Jeffreys.

Theorem 4: Let Q be a measure on the real line with support I. Assume that $\mu_{\inf}$ is the left end point of I. If Q has density $f(x) = (x-\mu_{\inf})^{\gamma-1} g(x)$ in an interval just to the right of a where g is an analytic function and $g(\mu_{\inf}) > 0$ 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.

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

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

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. The exponential family based on Q has redundancy upper bounded by 1 bit but the Jeffreys integral is infinite.

For exponential families in more dimensions the analysis becomes more more involved and one may even have exponential families with finite redundancy and infinite regret.

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] J. Rissanen, "Fisher information and stochastic complexity," *IEEE Trans. Inform. Theory*, vol. 42, no. 1, pp. 40–47, 1996.