Ill-posed relations between two types of information and relations between the ϕ-divergences and statistical information.

Milan Stehlík*,
Institute of Statistics, University of Valparaíso, Chile

An ill-posed relation between two types of information, in a mathematical context, means that small changes in the input data can lead to large, unpredictable changes in the output, making the problem difficult or impossible to solve without additional constraints or assumptions. During this talk, we will concentrate on two Fredholm type I expressed relations. Namely, in the context of Fisher informations for estimation and predictions in parametetric models [8], and, secondly, in relation between f-divergences $D_f(P,Q)$ and statistical informations $\mathcal{I}_\pi(P,Q) \equiv \mathcal{I}_\pi(P,Q)$ (differences $B_\pi - B_\pi(P,Q)$ between the prior and posterior Bayes risks). This relationship has been established by [5]).

The first relationship that can be well motivated from the perspective of designing optimal models for correlated errors is challenging due to the complexity involved. Observations of stochastic processes with parameterized mean and covariance are analyzed over a compact set using information functionals. Efficient designs for correlated process parameters, particularly equidistant designs, are explored, and this is far from an easy analytical part, which brings into play metaheuristic searches, see e.g. [7]. Equidistant designs have been shown to be optimal for trend parameters of certain processes, as shown in [4]. The talk will introduce compound criteria for optimal design and explore their potential for correlated process parameter optimization. These compound criteria form the kernels of Fredholm type I integrals.

In the second relationship between statistical divergence and statistical information, we will formulate some open problems [6]. We will discuss generalization of such relationship to the *alternative ϕ-divergences* $\mathcal{D}_\phi(P_1, P_2, ..., P_n)$ and general *statistical informations* $\mathcal{I}_{\pi_1,\pi_2,...,\pi_{n-1}}(P_1, P_2, ..., P_n)$ of [1, 2].

Here the *alternative ϕ-divergence* $\mathcal{D}_\phi(P_1, P_2, ..., P_n)$ means the integral

$$\int_{\mathcal{X}} \phi(p_1, p_2, ..., p_n) \mathrm{d}\mu \quad for\ p_i = \frac{\mathrm{d}P_i}{\mathrm{d}\mu},\ \mu \gg \{P_1, P_2, ..., P_n\} \tag{1}$$

where $\phi : [0,\infty)^n \longrightarrow (-\infty, \infty]$ is convex, continuous and homogeneous in the sense

$$\phi(\alpha t_1, \alpha t_2, ..., \alpha t_n) = \alpha \phi(t_1, t_2, ..., t_n) \forall \alpha \geq 0.$$

Integral (1) is well defined (but possibly infinite) which follows from the inequality between $\phi(t_1, t_2, ..., t_n)$ and its support plane at the point $(t_1, t_2, ..., t_n) = (1, 1, ..., 1)$.

These ϕ-divergences were introduced by [3]. Igor Vajda extend the definition of ϕ-divergences by:

* *Correspondence Address:* Email: Milan.Stehlik@uv.cl

(i) admitting in (1) convex functions $\phi : [0,\infty)^n \longrightarrow (-\infty,\infty]$ which are finite on $(0,\infty)^n$ and possibly infinite at the boundary,

(ii) replacing the continuity by the lower semicontinuity, and

(iii) assuming strict convexity at $(t_1,t_2,...,t_n) = (1,1,...,1)$ with $\phi(1,...,1) = 0$.

The last assumption guaratnees that $\mathcal{D}_\phi(P_1,P_2,...,P_n)$ is nonnegative, equal zero if and only if all probability measures $P_1,P_2,...,P_n$ coincide.

The *statistical information* $\mathcal{I}_{\pi_1,\pi_2,...,\pi_{n-1}}(P_1,P_2,...,P_n)$ is the difference between the classical prior Bayes risk $B_{\pi_1,\pi_2,...,\pi_{n-1}}$ and the posterior Bayes risk $B_{\pi_1,\pi_2,...,\pi_{n-1}}(P_1,P_2,...,P_n)$ in the statistical decision model with conditional probability measures $P_1,P_2,...,P_n$ on an observation space $\mathcal{X}$ which is equipped with a σ-algebra and a dominating σ-finite measure μ leading to the densities considered in (1). These probability measures are assumed to govern observations with prior probabilities $\pi_1,\pi_2,...,\pi_n$ where $\pi_1,\pi_2,...,\pi_{n-1}$ are from the open simplex

$$S_{n-1} = \left\{\pi_i > 0, \sum_{i=1}^{n-1} \pi_i < 1\right\} \subset \mathbb{R}^{n-1} \quad \textit{and}\ \pi_n = 1 - \sum_{i=1}^{n-1} \pi_i.$$

References

[1] DeGroot, M. H. Uncertainty, information, and sequential experiments. *Ann. Math. Stat. 33*, 2 (June 1962), 404–419.

[2] DeGroot, M. H. *Optimal Statistical Decisions.* McGraw-Hill, New York, 1970.

[3] Györfi, L., and Nemetz, T. f-dissimilarity: A generalization of the affinity of several distributions. *Ann. Inst. Stat. Math. 30*, 1 (Dec. 1978), 105–113.

[4] Kiseľák, J., and Stehlík, M. Equidistant and d-optimal designs for parameters of Ornstein–Uhlenbeck process. *Stat. Probab. Lett. 78*, 12 (Sept. 2008), 1388–1396.

[5] Liese, F., and Vajda, I. On divergences and informations in statistics and information theory. *IEEE Trans. Inf. Theory 52*, 10 (Oct. 2006), 4394–4412.

[6] Stehlik, M. Decompositions of information divergences: Recent development, open problems and applications. AIP.

[7] Stehlík, M., Chen, P.-Y., Wong, W. K., and Kiseľák, J. A double exponential particle swarm optimization with non-uniform variates as stochastic tuning and guaranteed convergence to a global optimum with sample applications to finding optimal exact designs in biostatistics. *Appl. Soft Comput. 163*, 111913 (Sept. 2024), 111913.

[8] Stehlík, M., Kisel'ák, J., Bukina, E., Lu, Y., and Baran, S. Fredholm integral relation between compound estimation and prediction (FIRCEP). *Stoch. Anal. Appl. 38*, 3 (May 2020), 427–459.