Divergencies and statistical decisions

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

How can we relate an optimal statistical decision to statistical information theory?

By attempting to answer this challenging question, we will illustrate the importance of understanding the structure of information divergences. This can be understood, in particular, through deconvolutions, which lead to optimal statistical inference.

We consider a statistical model with N independent observations $y_1, \ldots, y_N$ that are distributed according to gamma densities.

$$f(y_i|\gamma_i) = \begin{cases} \gamma_i^{v_i} \dfrac{y_i^{v_i-1}}{\Gamma(v_i)} \exp(-\gamma_i y_i), & \text{for } y_i > 0, \\ 0, & \text{for } y_i \leq 0. \end{cases} \tag{1}$$

Here $\gamma := (\gamma_1, \ldots, \gamma_N)$ is a vector of unknown scale parameters, which are the parameters of interest.

This model is motivated, for example, by a situation where we observe time intervals between $(N+1)$ successive random events in a Poisson process.

Model (1) is a regular exponential family, where the sufficient statistics for the canonical parameter $\gamma \in \Gamma$ have the form $t(y) = -y$ and $\Gamma = \{(\gamma_1, \ldots, \gamma_N), \gamma_i > 0; i = 1, \ldots, N\}$. The "covering" property

$$\{t(y) : y \in Y\} \subseteq \{E_\gamma[t(y)] : \gamma \in \Gamma\}$$

allows us to define the I-divergence of the observed vector y in the sense of [2]:

$$I_N(y, \gamma) := I(\hat{\gamma}_y, \gamma).$$

Here, $I(\gamma^\star, \gamma)$ is the Kullback-Leibler divergence between the parameters $\gamma^\star$ and γ. The I-divergence has nice statistical consequences. Let us consider the likelihood ratio (LR) λ_1 of the test of the hypothesis (2) and the LR λ_2 of the test of the homogeneity hypothesis $H_0 : \gamma_1 = \ldots = \gamma_N$ in the family (1). Then we have the following relation for every vector of canonical parameters $(\gamma_0, \ldots, \gamma_0) \in \Gamma^N$:

$$I_N(y, (\gamma_0, \ldots, \gamma_0)) = -\ln \lambda_1 + (-\ln \lambda_2 | \gamma_1 = \ldots = \gamma_N). \tag{2}$$

Here, the variables $-\ln \lambda_1$ and $-\ln \lambda_2 | \gamma_1 = \ldots = \gamma_N$, i.e., the $-\ln \lambda_2$ under the condition $H_0 : \gamma_1 = \ldots = \gamma_N$, are independent. The deconvolution (2) of I_N is a consequence of Theorem 4 in [3].

We will define three tests related to this deconvolution and explore their statistical properties. We will also illustrate the usefulness of I-divergence in applications in reliability engineering (see [4] and [7]), water quality control [6], linguistics [5], or portfolio analysis based on entropy mean-variance frontier [1].

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

References

[1] Hooda, D. S. Portfolio analysis of investments in risk management. *Open Stat. Prob. J. 3*, 1 (Aug. 2011), 21–26.

[2] Pázman, A. *Nonlinear Statistical Models*. Springer Netherlands, Dordrecht, 1993.

[3] Stehlík, M. Distributions of exact tests in the exponential family. *Metrika 57*, 2 (May 2003), 145–164.

[4] Stehlík, M. Homogeneity and scale testing of generalized gamma distribution. *Reliab. Eng. Syst. Saf. 93*, 12 (Dec. 2008), 1809–1813.

[5] Stehlík, M., Pari Ruiz, M., Stehlíková, S., and Lu, Y. On equidistant designs, symmetries and their violations in multivariate models. In *Contemporary Experimental Design, Multivariate Analysis and Data Mining*. Springer International Publishing, Cham, 2020, pp. 217–225.

[6] Stehlík, M., Sabolová, R., Sečkárová, V., Núñez Soza, L., and Kiseľák, J. I-divergence based statistical inference for heteroscedasticity and compounds of arsenic contamination in chile. *Chemometr. Intell. Lab. Syst. 226*, 104579 (July 2022), 104579.

[7] Stehlík, M., Somorčík, J., Střelec, L., and Antoch, J. Approximation of information divergences for statistical learning with applications. *Math. Slovaca 68*, 5 (Oct. 2018), 1149–1172.