Japanese Journal of Statistics and Data Science (2019) 2:437–440
https://doi.org/10.1007/s42081-019-00059-7

PREFACE

Special feature: information theory and statistics

Takafumi Kanamori[1,2]

Published online: 4 September 2019

In commemoration of the foundation of the Japanese Journal of Statistics and Data Science (JJSD), our third special feature focuses on relationships and collaborations between information theory and statistics. Development and expansion in research areas of information theory and statistics are drawing a great deal of attention. This special feature comprises ten contributions from mainly three research topics: divergence-based statistical inference, high-dimensional sparse learning, and combinatorial design.

Divergence measure, an extension of distances over the set of probability distributions, is a transverse tool in mathematical sciences. In particular, the Kullback-Leibler divergence, or in other words, the relative entropy, is closely related to the maximum likelihood estimator in statistics and the code length in information theory (Kullback 1959; Cover and Thomas 2006). The information criterion (Akaike 1974) and Bayes coding (Clarke and Barron 2006) are interdisciplinary research topics based on the concept of divergences. Today, important classes of divergence measures such as Bregman divergences have been widely applied in data analysis and information sciences (Bregman 1967; Basu et al. 1998; Fujisawa and Eguchi 2008). The following five articles focus mainly on theoretical analysis and practical applications of divergence measures.

– Machida and Takenouchi (2019) are concerned with non-negative matrix factorization (NMF), which is a typical means of feature extraction in the framework of unsupervised learning (Lee and Seung 2001). It is well known that the standard NMF algorithm is not robust against outlier noise. The authors propose robust NMF algorithms by combining statistical modeling of reconstruction and the γ -divergence.

✉ Takafumi Kanamori
kanamori@c.titech.ac.jp; takafumi.kanamori@riken.jp

[1] Department of Mathematical and Computing Science, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan

[2] RIKEN Center for Advanced Intelligence Project (AIP), Nihonbashi 1-chome Mitsui Building, 15th floor,1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan

- Kawashima and Fujisawa (2019) study an extension of robust and sparse linear regression. They propose robust and sparse generalized linear models (McCullagh and Nelder 1989 based on the γ-divergence.
- Ihara (2019) developed a new mathematical tool to prove the optimality of the coding scheme for the feedback capacity of discrete-time additive Gaussian noise. The author also shows that the minimum decoding error probability decreases with an exponential order that increases linearly with block length.
- Sainudiin and Teng (2019) present a data-adaptive multivariate histogram estimator of an unknown density based on independent samples. The authors prove the universal performance guarantee under the L_1 distance.
- Abe and Fujisawa (2019) study a multivariate skew distribution using the transformation of scale. Also, they present additional properties of the distribution such as random number generation, non-degenerated Fisher information, and entropy maximization distribution.

In the past two decades, the notion of *sparsity* has attracted massive attention in the research realm of compressed sensing, high-dimensional statistical inference, and related matters (Tibshirani 1996). Today, highly developed detection technology enables us to observe extremely high-dimensional complex data. Statisticians are thus required to develop statistical methods to deal with such high-dimensional data. Sparsity is an assumption that only a small number of elements in high-dimensional data are significant. The regularization techniques that induce a sparsity pattern are useful for finding such significant elements. The L1-regularization for linear regression models has been intensively studied both in theory and practice (Hastie et al. 2015). In this special feature, the following three articles have to do with the sparse structure of high-dimensional data.

- Komatsu et al. (2019) study group lasso, in which high-dimensional covariates are assumed to be clustered in groups. The authors propose the information criterion for the group Lasso under the framework of generalized linear models. They illustrate that their criterion is almost the same as or better than cross-validation.
- Post-selection inference (Lee et al. 2013) is a statistical technique for determining salient variables after model or variable selection. Umezu and Takeuchi (2019) develop a selective inference framework for binary classification problems. They also conduct several simulation studies to confirm the statistical power of the test.
- The sparse superposition code is known to achieve channel capacity (Joseph and Barron 2012). Takeishi and Takeuchi (2019) show an improved upper bound on its block error probability with least squares decoding, which is a fairly simplified and tighter bound than in previous results.

Combinatorial structure appears in both statistics and information theory. In statistics, the experimental design is a classical research topic in which combinatorics has been widely applied (Fisher 1940; Rao 1947). Also in coding theory, the combinatorial concept is important to design computationally efficient coding. Two articles

provide an interesting connection between combinatorial structure and information processing.

– Hirao and Sawa (2019) present a characterization theorem of a combinatorial structure called almost tight Euclidean design. Furthermore, the article includes a short review of a relationship between Euclidean designs for rotationally symmetric integrals and kernel approximation in machine learning.
– Lu and Jimbo (2019) present a review of the brief history, basic problems, and significant results for constructing the arrays for combinatorial interaction testing. They also propose explicit construction for covering arrays involving information-theoretic methods.

We are grateful to all reviewers for their help in the process of refereeing the contributions and for sharing their time and knowledge. We also want to thank again all authors who have contributed interesting works to this special feature.

References

Abe, T., & Fujisawa, H. (2019). Multivariate skew distributions with mode-invariance through transformation of scale. *Japanese Journal of Statistics and Data Science*. https://doi.org/10.1007/s42081-019-00047-x.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control AC.*, *19*(6), 716–723.

Basu, A., Harris, I. R., Hjort, N. L., & Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, *85*(3), 549–559.

Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, *7*, 200–217.

Clarke, B. S., & Barron, A. R. (2006). Information-theoretic asymptotics of bayes methods. *IEEE Trans Inf Theor*, *36*(3), 453–471. https://doi.org/10.1109/18.54897.

Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. New York: Wiley-Interscience.

Fisher, R. A. (1940). An examination of the different possible solutions of a problem in incomplete blocks. *Annals of Eugenics*, *10*, 52–75.

Fujisawa, H., & Eguchi, S. (2008). Robust parameter estimation with a small bias against heavy contamination. *J Multivar Anal*, *99*(9), 2053–2081.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. London: Chapman & Hall/CRC.

Hirao, M., & Sawa, M. (2019). On almost tight euclidean designs for rotationally symmetric integrals. *Japanese Journal of Statistics and Data Science*. https://doi.org/10.1007/s42081-019-00048-w.

Ihara, S. (2019). On the feedback capacity of the first-order moving average gaussian channel. *Japanese Journal of Statistics and Data Science*. https://doi.org/10.1007/s42081-019-00045-z.

Joseph, A., & Barron, A. R. (2012). Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity. *IEEE Trans Information Theory*, *58*(5), 2541–2557. https://doi.org/10.1109/TIT.2012.2184847.

Kawashima, T., & Fujisawa, H. (2019). Robust and sparse regression in generalized linear model by stochastic optimization. *Japanese Journal of Statistics and Data Science*. https://doi.org/10.1007/s42081-019-00049-9.

Komatsu, S., Yamashita, Y., & Ninomiya, Y. (2019). AIC for the group lasso in generalized linear models. *Japanese Journal of Statistics and Data Science*. https://doi.org/10.1007/s42081-019-00052-0.

Kullback, S. (1959). *Information Theory and Statistics*. New York: Wiley.
Lee, D. D., & Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In: Leen TK, Dietterich TG, Tresp V (eds) *Advances in Neural Information Processing Systems 13*, MIT Press, pp 556–562. http://papers.nips.cc/paper/1861-algorithms-for-non-negative-matrix-factorization.pdf
Lee, J. D., Sun, D. L., Sun, Y., & Taylor, J. E. (2013). Exact post-selection inference, with application to the lasso. https://doi.org/10.1214/15-AOS1371, http://arxiv.org/abs/1311.6238
Lu, X. N., & Jimbo, M. (2019). Arrays for combinatorial interaction testing: A review on constructive approaches. *Japanese Journal of Statistics and Data Science*. https://doi.org/10.1007/s42081-019-00056-w.
Machida, K., & Takenouchi, T. (2019). Statistical modeling of robust non-negative matrix factorization based on γ-divergence and its applications. *Japanese Journal of Statistics and Data Science*. https://doi.org/10.1007/s42081-019-00041-3.
McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman & Hall / CRC.
Rao, R. C. (1947). Factorial experiments derivable from combinatorial arrangements of arrays. *Journal of the Royal Statistical Society (Supplement)*, *9*(1), 128–139.
Sainudiin, R., & Teng, G. (2019). Minimum distance histograms with universal performance guarantees. *Japanese Journal of Statistics and Data Science*. https://doi.org/10.1007/s42081-019-00054-y.
Takeishi, Y., & Takeuchi, J. (2019). An improved analysis of least squares superposition codes with bernoulli dictionary. *Japanese Journal of Statistics and Data Science*. https://doi.org/10.1007/s42081-019-00057-9.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, *58*(1), 267–288.
Umezu, Y., & Takeuchi, I. (2019). Selective inference via marginal screening for high dimensional classification. *Japanese Journal of Statistics and Data Science*. https://doi.org/10.1007/s42081-019-00058-8.