Experimental Results (2022), 3, e16, 1–7
doi:10.1017/exp.2022.14

COMPUTER SCIENCE
SUPPLEMENTARY-RESULT

Application of offset estimator of differential entropy and mutual information with multivariate data

Iván Marín-Franch[1,2,*], Martín Sanz-Sabater[3] and David H. Foster[4]

[1]Computational Optometry, Atarfe, Spain, [2]Southwest Eye Institute, Tavistock, United Kingdom, [3]Optics Department, Universitat de València, Valencia, Spain, and [4]Department of Electrical and Electronic Engineering, University of Manchester, Manchester, United Kingdom
*Corresponding author: Email: imarinfr@optocom.es

(Received 17 February 2022; Revised 25 August 2022; Accepted 25 August 2022)

Abstract
Numerical estimators of differential entropy and mutual information can be slow to converge as sample size increases. The offset Kozachenko–Leonenko (KLo) method described here implements an offset version of the Kozachenko–Leonenko estimator that can markedly improve convergence. Its use is illustrated in applications to the comparison of trivariate data from successive scene color images and the comparison of univariate data from stereophonic music tracks. Publicly available code for KLo estimation of both differential entropy and mutual information is provided for R, Python, and MATLAB computing environments at https://github.com/imarinfr/klo.

Key words: information theory; Kozachenko–Leonenko estimator; mutual information; nonparametric statistics; R; Python; and MATLAB

Introduction

Shannon's theory of communication (Shannon, 1948a; 1948b) demonstrated that the information transmitted between systems is a well-defined measurable quantity with fundamental limits. The two key elements of what became known as information theory are entropy and mutual information. The entropy of a system quantifies the uncertainty of the result of making an observation on a signal, and the mutual information quantifies how much of that uncertainty can be reduced by a related signal. From a statistical viewpoint, mutual information is a measure of the probabilistic dependence between univariate or multivariate random variables that is more general than, for example, Pearson's correlation, which measures the linear association between two univariate random variables.

Given the wide applicability of information-theoretic quantities in physics, engineering, and the life sciences, it is important to have available accurate numerical estimators. Unfortunately, the underlying distributions are generally unknown, and nonparametric estimators are usually required, although they may be subject to error, especially with continuous multivariate random variables where the entropy becomes the differential entropy. For some existing estimators, their slow convergence with increasing sample size can be a serious challenge. But their accuracy may be improved by exploiting a simple decomposition of differential entropy. The method was introduced in Marín-Franch and Foster (2013) in an application to artificial image transformations.

The objective here is to illustrate an extension of the method to two real-world datasets and to describe software packages for estimating both differential entropy and mutual information in several computing environments. The two datasets consist of trivariate data from successive scene color images and univariate data from a stereophonic music recording. The first application is more detailed and contains illustrative code for the R computing environment (R Core Team, 2021).

Methods

The nonparametric estimator described here is an offset version of a nearest-neighbor class of estimators for differential entropy (Berrett *et al.*, 2019; Charzyńska & Gambin, 2015; Goria *et al.*, 2005; Holmes & Nemenman, 2019; Kozachenko & Leonenko, 1987; Kraskov *et al.*, 2004), namely the Kozachenko–Leonenko (KL) estimator (Goria *et al.*, 2005; Kozachenko & Leonenko, 1987). Although the offset method entails a decomposition into Gaussian and non-Gaussian components, the distribution of X is not itself assumed to be Gaussian or approximately Gaussian. Estimating the differential entropy $h(X)$ of a d-dimensional multivariate continuous random variable X proceeds as follows.

1. Estimate the differential entropy $\hat{h}_{\mathrm{G}}(X)$ of a Gaussian distribution with the same covariance matrix, c say, as X.
2. Linearly transform X to form a new variable X^* whose differential entropy $h(X^*)$ is such that $h(X) = h(X^*) + \hat{h}_{\mathrm{G}}(X)$. From the scaling property of differential entropy, the transformation, A say, required for the equality to hold is given by $A = (2\pi e)^{-1/2} c^{-1/2}$, so that $X^* = AX$.
3. Apply the KL estimator to X^* and call the result $\hat{h}_{\mathrm{KL}}(X^*)$.
4. Define the offset Kozachenko–Leonenko (KLo) estimator by $\hat{h}_{\mathrm{KLo}}(X) = \hat{h}_{\mathrm{KL}}(X^*) + \hat{h}_{\mathrm{G}}(X)$.

Since the KL estimator is asymptotically unbiased (Goria *et al.*, 2005; Kozachenko & Leonenko, 1987), it follows that the offset version KLo is also asymptotically unbiased. For a proof, see Appendix A in the Supplementary Material, available on the Cambridge Core website. For a description of the software implementation, see Appendix B in the Supplementary Material.

Results

The first example illustrates the sample-size dependence of the KLo and KL estimators for a trivariate dataset, along with results for the popular Kraskov-Stögbauer-Grassberger (KSG) estimator of mutual information (Kraskov, Stögbauer, & Grassberger, 2004). Progressively larger samples, ranging from 2^{10} to 2^{19} points, were drawn randomly and identically from two trivariate images of a scene recorded at successive instants, about 1 min apart, shown in the thumbnail color images in Figure 1. The data were taken from a larger study (Foster, 2021) where image values were expressed not as conventional RGB triplets but as LMS triplets, corresponding to activities in the long-, medium-, and short-wavelength-sensitive cone photoreceptors of the eye. The difference between RGB and LMS representations is immaterial for this illustration. Each image was stored as a $1{,}024 \times 1{,}344 \times 3$ array, where the first two dimensions index pixel coordinates and the third dimension indexes LMS values, each obtained by integrating 12-bit spectral radiance data weighted by photoreceptor sensitivities. The frequency distributions of the LMS values were bimodal.

The main panel in Figure 1 shows the KLo, KL, and KSG estimates of the mutual information plotted against the sample size. Each curve is an average of over 100 repeated random samples. The KLo estimate rapidly asymptotes with increasing sample size, unlike the KL and KSG estimates, which continue to increase even as sample size approaches the maximum available determined by image size. The Gaussian component of the KLo estimator was about 8.0 bits.

The second example is described in Appendix C in the Supplementary Material, available on the Cambridge Core website. It illustrates the similarity of the KLo and KL estimates and the failure of the KSG estimate with a univariate dataset.

https://doi.org/10.1017/exp.2022.14 Published online by Cambridge University Press

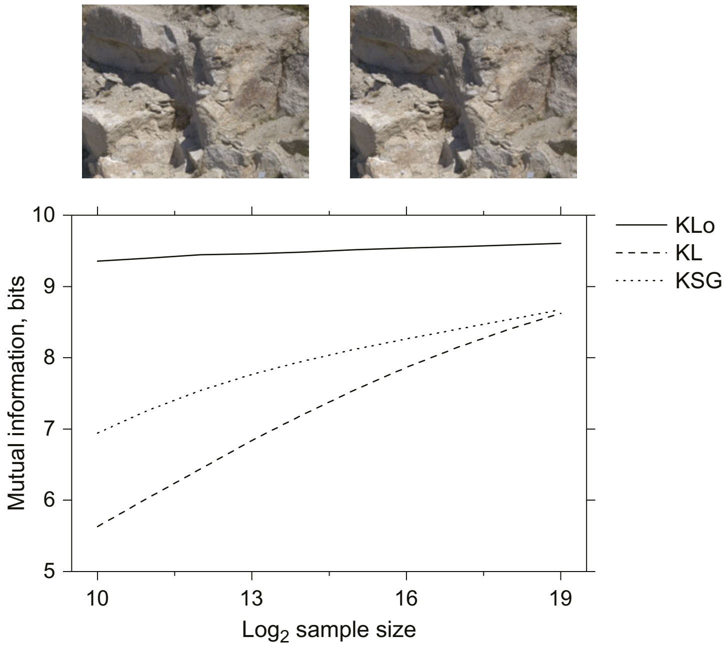

Figure 1. Estimates of mutual information between two color images. The thumbnail images are sRGB renderings (IEC, 1998) of the source data. The plots show mutual information estimates for the offset Kozachenko–Leonenko (KLo), Kozachenko–Leonenko (KL), and Kraskov-Stögbauer-Grassberger (KSG) estimators as a function of sample size. Standard deviations for the KLo and KL estimates ranged from about 0.1 with the smallest sample sizes to 0.006 with the largest sample sizes. Standard deviations for the KSG estimates were a little smaller.

Discussion

The slow convergence of mutual information estimators with increasing sample size is not inevitable. The offset method can clearly improve the convergence of estimates derived with the Kozachenko–Leonenko estimator with some real-world datasets. But the extent of the improvement does depend on the size of the Gaussian component of the underlying differential entropies. For distributions very far from Gaussian, there is no guarantee that the offset method will converge faster than applying the KL estimator directly. The offset method does, though, have the advantage of automatically adjusting itself to the properties of the distributions. It is, moreover, neutral with respect to the choice of differential entropy estimator, so that any other estimator can instead be plugged in.

The present approach may be open to generalization. One possibility is to replace the particular linear transformation used to decompose differential entropy into Gaussian and non-Gaussian components by other transformations. Another possibility is to extend the offset method to estimating related information-theoretic quantities such as Kullback–Liebler divergence and cross-entropy.

Acknowledgment. We are grateful to P. A. Gaydecki for providing the soundtrack samples and to S. M. C. Nascimento and K. Amano for collaborating in producing the hyperspectral images for the color data.

Supplementary Materials. To view supplementary material for this article, please visit http://doi.org/10.1017/exp.2022.14.

Data availability statement. The code is copyrighted by I.M.-F. and D.H.F. and distributed under Apache License v2.0. Both the code and data used in this manuscript are available for R, Python, and MATLAB computing environments at https://github.com/imarinfr/klo.

Funding statement. This work was supported by Computational Optometry (Atarfe, Spain; https://www.optocom.es/). The code was developed as part of two EPSRC grants (EP/B000257/1 and EP/E056512/1).

Conflicts of interest. The authors declare no conflicts of interest.

Authorship contributions. Conceptualization: I.M.-F. and D.H.F.; Methodology: I.M.-F.; Software: I.M.-F. and M.S.-S.; Validation: all authors; Visualization: D.H.F.; Writing—original draft: I.M.-F. and D.H.F.; Writing—review and editing: I. M.-F. and D.H.F.

References

Berrett, T. B., Samworth, R. J., & Yuan, M. (2019). Efficient multivariate entropy estimation via *k*-nearest neighbour distances. *The Annals of Statistics*, **47**, 288–318. https://doi.org/10.1214/18-AOS1688

Charzyńska, A., & Gambin, A. (2015). Improvement of the *k*-NN entropy estimator with applications in systems biology. *Entropy*, **18**,13. https://doi.org/10.3390/e18010013

Foster, D. H. (2021). Fluctuating environmental light limits number of surfaces visually recognizable by colour. *Scientific Reports*, **11**, 2102. https://doi.org/10.1038/s41598-020-80591-9

Goria, M. N., Leonenko, N. N., Mergel, V. V., & Novi Inverardi, P. L. (2005). A new class of random vector entropy estimators and its applications in testing statistical hypotheses. *Journal of Nonparametic Statistics*, **17**, 277–297. https://doi.org/10.1080/104852504200026815

Holmes, C. M., & Nemenman, I. (2019). Estimation of mutual information for real-valued data with error bars and controlled bias. *Physical Review E*, **100**,022404. https://doi.org/10.1103/PhysRevE.100.022404

IEC. (1998). *IEC 61966-2-1. Colour management in multimedia systems—Part 2: Colour Management, Part 2.1: Default RGB colour space—sRGB. Technical report, Technical Committee No. 100.*

Kozachenko, L. F., & Leonenko, N. N. (1987). Sample estimate of the entropy of a random vector [Translated from *Problemy Peredachi Informatsii*, 23(2): 9–16, 1987]. *Problems of Information Transmission*, **23**, 95–101.

Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. *Physical Review E*, **69**, 066138. https://doi.org/10.1103/PhysRevE.69.066138

Marín-Franch, I., & Foster, D. H. (2013). Estimating information from image colors: An application to digital cameras and natural scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**, 78–91. https://doi.org/10.1109/TPAMI.2012.78

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Shannon, C. E. (1948a). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shannon, C. E. (1948b). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

Cite this article: Marín-Franch I, Sanz-Sabater M, Foster DH (2022). Application of offset estimator of differential entropy and mutual information with multivariate data. *Experimental Results,* 3, e16, 1–7. https://doi.org/10.1017/exp.2022.14

Peer Reviews

Reviewing editor: Prof. Emanuele Frontoni
University of Macerata, Information Engineerging Department - DII, Macerata, Italy, 62100

Minor revisions requested.

doi:10.1017/exp.2022.14.pr1

Review 1: Offset estimator of differential entropy and mutual information with multivariate data

Reviewer: Dr. Javier E. Contreras-Reyes

Date of review: 19 March 2022

Conflict of interest statement. None.

Comment

Comments to the Author: Review of "Offset estimator of differential entropy and mutual information with multivariate data"

Authors considered an estimators already published in Marin-French & Foster (2013), which itself is based on Kozachenko-Leonenko (1987) estimator. From 2013 until today, I think that the estimator has been proved in several experiments. Thus, what is the real contribution of the paper? it is the public programming codes (Matlab-R-Python)? or the application?

Decomposition on a Gaussian and non-Gaussian component is an important step of the proposed method, which has widely considered in the literature for differential entropy and mutual information. From the references added in the manuscript, there exist(s) some(s) of them that considered this issue?

About the results, in Appendix A (of supplementary material) is obtained the limits for KL and KLo, where for a large sample size, they converge to differential entropy H. However, in Fig. 1, why not occurs the same for KL and KLo in log_2(n) ~ 19? Also, why is the intention of authors in including these both images? what is the difference among the images?

Score Card

Presentation

4.4 /5

Is the article written in clear and proper English? (30%)	5/5
Is the data presented in the most useful manner? (40%)	5/5
Does the paper cite relevant and related articles appropriately? (30%)	3/5

Context

4.8 /5

Does the title suitably represent the article? (25%)	5/5
Does the abstract correctly embody the content of the article? (25%)	5/5
Does the introduction give appropriate context? (25%)	4/5
Is the objective of the experiment clearly defined? (25%)	5/5

Analysis

4.4 /5

Does the discussion adequately interpret the results presented? (40%)	5/5
Is the conclusion consistent with the results and discussion? (40%)	4/5
Are the limitations of the experiment as well as the contributions of the experiment clearly outlined? (20%)	4/5

https://doi.org/10.1017/exp.2022.14 Published online by Cambridge University Press doi:10.1017/exp.2022.14.pr2

Review 2: Offset estimator of differential entropy and mutual information with multivariate data

Reviewer: Dr. Gbadebo Oladeji-Atanda MSc

Botswana International University of Science and Technology, Computer Science and Information Systems, P Bag 16, Palapye, Botswana

Date of review: 19 August 2022

Conflict of interest statement. Reviewer declares none

Comment

Comments to the Author: Line 40 Correct the word 'asymptically'

Score Card

Presentation

5.0 /5

Is the article written in clear and proper English? (30%)	5/5
Is the data presented in the most useful manner? (40%)	5/5
Does the paper cite relevant and related articles appropriately? (30%)	5/5

Context

4.5 /5

Does the title suitably represent the article? (25%)	4/5
Does the abstract correctly embody the content of the article? (25%)	5/5
Does the introduction give appropriate context? (25%)	5/5
Is the objective of the experiment clearly defined? (25%)	4/5

Analysis

5.0 /5

Does the discussion adequately interpret the results presented? (40%)	5/5
Is the conclusion consistent with the results and discussion? (40%)	5/5
Are the limitations of the experiment as well as the contributions of the experiment clearly outlined? (20%)	5/5