DOI: 10.2478/bile-2022-0006

Biometrical Letters
Vol. 59 (2022), No. 1, 65-75

Considering distance measures in Statistics

Christos P. Kitsos[1], Constantinos - Symeon Nisiotis[2]

[1]Department of Informatics and Computer Engineering, University of West Attica, Agiou Spiridonos 28, 12243, Aigaleo, Greece, e-mail: xkitsos@uniwa.gr
[2]Department of Public and Community Health, University of West Attica, Agiou Spiridonos 28, 12243, Aigaleo, Greece, e-mail: csnisiotis@uniwa.gr

Summary

The target of this paper is to offer a compact review of the so called distance methods in Statistics, which cover all the known estimation methods. Based on this fact we propose a new step, to adopt from Information Theory, the divergence measures, as distance methods, to compare two distributions, and not only to investigate if the means or the variances of the distributions are equal. Some useful results towards this line of thought are presented, adopting a compact form for all known divergence measures, and are appropriately analyzed for Biometrical, and not only, applications.

Key words: Euclidean distance, information theory, bioassays

1. Introduction

The sense of distance is also important in the main aspects of Statistics, such as Estimation. Among the pioneering work one could consider the work of Blyth (1970), Wolfowitz (1957), were it was pointed out that the Maximum Likelihood Estimators (MLE), the Least Square Estimators (LSE), the Chi-Square test, as well as the Kolmogorov-Smirnov test obey the minimum distance principle.

In principle, in Statistics and therefore in Information Theory, the term “distance” does not fulfill the “triangular inequality”, as in Linear Algebra or in Analysis. Moreover, the “symmetry” it is not also available in most cases. But we keep on using the term “distance” for these information distance measures! Moreover, there are “information distance” measures, which are discussed and can be proved useful in Geographical Data Analysis, as well in other Data Analysis fields, see also Kitsos and Iliopoulou (2021, 2022).

It has been defined, that the minimum distance estimators, are those which minimize the distance between the empirical distribution F_n and the actual distribution, F_θ, in terms that:

$$\inf d(F_\theta, F_n) = d(F_{\hat{\theta}}, F_n) \tag{1}$$

where $\theta, \hat{\theta}$ being the parameter space. In (1) the "minimum distance estimator" is defined, and $\hat{\theta}$ is the evaluated estimate. Notice that by the Law of "Large Numbers" the empirical distribution converges a.s to the distribution function F, while the Glivenkov-Cantelli theorem insures that the convergence in fact happens uniformly in R, in terms that:

$$||F_n - F||_\infty = sup\{|F_n(x) - F(x)|, x \in R\} \tag{2}$$

Moreover by the Central Limit Theorem (CLT) the empirical distribution has asymptotically normal distribution, with $\sqrt{n}$ rate of convergence, mean zero and variance $F(x)(1-F(x))$. In principle the discrepancy measure is reflected to $d(F_\theta, F_n)$, for the estimation case under consideration. Typical example can be the Kolmogorov - Smirnov (K-S) test Smirnov (1948), Darling (1957) were:

$$d(F_{\hat{\theta}}, F_n) = \sup_{x \in R} |F_\theta(x) - F_n(x)| \tag{3}$$

So, it is clear that the K-S statistic quantifies a distance measure between the empirical distribution function, of the collected sample, of size n and the cumulative distribution function of the reference distribution. In the same way of thinking, Least Squares have been considered through the Euclidean distance in R^n. There is a Geometrical inside in most Statistical methods, see, Saville and Wood (1991).That is why, we believe, Maximum Likelihood estimation, introduced in by Fisher (1922), it is an impressed evolution: although no typical distance measure is used, at the end, it is a distance measure. Adopting the distance methods in Bioassays Kitsos and Sotiropoulos (2009), it was clear that the affine transformation preserves all the properties, for a class of Binary response models. In applications not only the K-S test is adopted, but also the Shapiro-Wilk test as a test of normality in frequentist Statistics, Shapiro and Wilks (1965).

So moving from the Euclidean distance between two points in R^n, we moved to Minkowski's distance for given points, of order m is:

$$d_m(x, y) = \left[\sum_{i=1}^{n} |x_i - y_i|^m\right]^{1/m} \tag{4}$$

then the distance of a given point from a Normal distribution. The Mahalanobis distance, d_M, the distance of the measurements $x \in R^n$ form the Normal distribution $N(\mu, \Sigma)$ with mean $\mu \in R^n$ and covariance matrix $\Sigma \in R^{n \times n}$, Mahalanobis (1936):

$$d_M = [(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})]^{1/2} \tag{5}$$

The next step is the distance between the empirical distribution and reference distribution,

$$||F_n - F||_\infty = sup\,\{|F_n(x) - F(x)|, x \in R\} \tag{6}$$

Consider two probability measures P and Q say, on a sigma-algebra $\boldsymbol{A}$ of subsets of the sample space Ω. Then, the total variation distance, V day, between P and Q is defined :

$$V(P, Q) = 2sup\,[|P(A) - Q(A)| : A \in \boldsymbol{A}] \tag{7}$$

It also holds that:

$$V(P, Q) = ||P - Q||_1 = \sum |P(\omega) - Q(\omega)|, \omega \in \Omega \tag{8}$$

So, the total variation V is related to L_1 norm. Moreover, from the Radon-Nikodym theorem there exist functions f, g, respectively to P and Q integrable such that, with μ the appropriate measure:

$$f = \frac{dP}{d\mu}, \quad g = \frac{dQ}{d\mu} \tag{9}$$

Then it holds, Kitsos and Iliopoulou (2022), among others :

$$d_{TV}(P, Q) = \int_A |f - g| d\mu \tag{10}$$

This compact review supports the line of thought introduced in section 2. Adopt the information theory approach, were the distance between distributions is adopted, to provide Data Analysis and it will be referred with capital D.

2. The information theory point of view of distance

The information distance measures (idm) succeed to qualify the information provided by the random variable (r.v) X, coming from the distribution P, say, in relation with the random variable Y, coming from the given distribution Q. In such case, the most well-known information measure, is that of Kullback-Leibler one, Kulback - Leibler (1951), widely known as K-L divergence, also known as relative entropy Cover and Thomas (1991). The K-L divergence it is neither symmetric nor triangular, but offers a measure to quantify how one probability distribution is different from a second, considering as reference probability distribution. This is why it is known as "divergence" and not as "distance" measure. It varies from zero (when P and Q perfectly much) to infinity (the K-L divergence between Cauchy and Normal distribution, Schervish (1995), the perfect disagreement) and is considering as an information-based measure of disparity among probability distributions, defined for the continuous or discrete case as:

$$\begin{aligned} D_{KL}(X,Y) \equiv D_{KL}(P,Q) \quad &:= \int p(x) \log \tfrac{p(x)}{q(x)} dx && \text{Continuous case} \\ &:= \textstyle\sum_{i=1}^{n} p_i \log \tfrac{p_i}{q_i} && \text{Discrete case} \end{aligned}$$

Usually p(x) is considered the "true" distribution and "q(x)" the approximate, so it is eventually the expectation of the log difference between the probability of data in the original distribution with the approximating distribution. Moreover K-L information needs no many regularity conditions, as Fisher's information needs, but still is related to it, Schervish (1995).

The K-L diverge usually works with the Normal distribution, while there is a broader class of "normal" distributions: The Generalized Normal Distribution (GND) was defined in Kitsos and Tavoularis (2009), through the Logarithm Sobolev Inequalities (LSI) for the r.v $X \equiv N_{\gamma}^{p}(\mu, \Sigma)$ with three parameters: for position $\mu \in R^p$, variance $\Sigma \in R^{p \times p}$ and the shape parameter γ. In GND the introduced new parameter, the shape parameter γ, offers the possibility for creating a family of "normal distributions with fat tails". Moreover, the "universal constant", the quantity:

$$\left(\frac{1}{\gamma - 1}\right)^{1/(\gamma-1)} or \left(\frac{\gamma}{\gamma - 1}\right)^{\gamma/(\gamma-1)} \tag{11}$$

and as it appears in GND, Kitsos and Tavoularis (2009), Kitsos and Toulias (2010, 2017) see also (12) in $q_i(x)$, $i = 1, 2$ and in Applied Analysis Takasi and Takashi (2011), is γ associated with the order GND theory. Where the normality does not hold, this can be a solution, while in this paper the normality of the observations is the main assumption. For the Generalized Normal distributions $KT^p_\gamma(\mu_1, \sigma_1^2|p)$ and $KT^p_\gamma(\mu_0, \sigma_0^2|p)$ the (information) distance KLI^p_γ of Kullback-Leibler, for given γ is equal to Kitsos and Toulias (2017):

$$KLI^p_\gamma = \frac{C(p,\gamma)}{\sigma_1^p}\left[\frac{p}{2}(ln\frac{\sigma_0^2}{\sigma_1^2})\int_{R^p} e^{-q_1(x)}dx - \int_{R^p} e^{-q_1(x)}q_1(x)dx + \int_{R^p} e^{-q_1(x)}q_0(x)dx\right], \tag{12}$$

where $q_i(x) = \frac{\gamma-1}{\gamma}(\sigma_i^{-1}x - \mu_i)^{\frac{\gamma}{(\gamma-1)}}, x \in R^p, \quad i = 0, 1$

Moreover, the evaluated KLI^p_2, between two classical Normal distributions, is a linear expression:

$$KLI^p_2 = \xi p + \frac{|\mu_1 - \mu_0|^2}{2\sigma_0^2}, \qquad \xi = \frac{1}{2}\left[\left(ln\frac{\sigma_0^2}{\sigma_1^2}\right) - 1 + \frac{\sigma_1^2}{\sigma_0^2}\right] \geqslant 0 \tag{13}$$

Considering (13) as a linear function of p the slope is positive, as well as the constant term.

A general form of the information distance measure (idm) was proposed in Kitsos and Toulias (2017). For given "smooth" functions g, h and f_X, f_Y the pdf of given rv's X and Y the idm is defined to be, D_{KT}, as:

$$D_{KT}(X,Y) = g\left(\int_{R^p} h(f_X, f_Y)\right) \tag{14}$$

Under (14) the following cases can be considered for the function g:
Case 1: For $g(t) = \sqrt{t}$
Then with:

$$h(f_X, f_Y) = \frac{1}{2}\left(\sqrt{f_X} - \sqrt{f_Y}\right)^2 \tag{15}$$

the Hellinger's distance $D_H(X,Y)$ is obtained.
Case 2: For $g(t) = id(t) = t$
With different values of $h(\cdot,\cdot)$ different distance measures are obtained:
2i. With $h(f_X, f_Y) = f_X \log\frac{f_X}{f_Y}$ the D_{KL} is obtained, the K-L distance.
2ii. With $h(f_X, f_Y) = f_X\left[\log\frac{f_X}{f_Y}\right]^2$ the exponential distribution, D_e, ids obtained.

2iii. With $h(f_X, f_Y) = f_X[1 - \frac{f_Y}{f_X}]^\alpha, \alpha \geqslant 1$ the Vajda, D_v, distance is emerged.

2iv. With $h(f_X, f_Y) = \frac{1}{2} f_X \left[1 - \frac{f_Y}{f_X}\right]^2$ the Kagan (or X^2), D_{X^2}, distance is obtained.

2v. With $h(f_X, f_Y) = f_Y \phi(\frac{f_Y}{f_X})$, ϕ convex, the Csiszar, D_c, distance defined.

2vi. With $h(f_X, f_Y) = f_X^\alpha f_Y^{1-\alpha}, \alpha \in (0, 1)$ the $D_{ch}^{(\alpha)}$, Chernoff's distance is obtained.

2vii. With $h(f_X, f_Y) = |f_X - f_Y|$ we obtain the very popular, total variation distance, D_{TV}.

Case 3: For $g(t) = log(t)$

With $h(f_X, f_Y) = (f_X \cdot f_Y)^{1/2}$ also denoted as BC (Bhattacharyya Coefficient), one comes across to the Bhattacharyya's , D_B, distance.

Notice that in most cases the ratio $\left(\frac{f_X}{f_Y}\right)$ is used and this can offer an easy explanation that there is, in principle, no symmetry for the idm. It is needed a special consideration for the cases where symmetry can be present. Moreover, under certain restriction the idm are related between them, as one can be a function of another one - but the technical insight is different. One example is that Hellinger's D_H distance, which obeys to the triangle inequality, is related to Bhattacharyya's, D_B, which is symmetric, but not under triangular inequality, see Corollary 1.

The cases 2i and 2ii are related, as well as the cases 2iii and 2iv. The 2v case offers a generalization for the ratio $\frac{f_Y}{f_X}$, with first term the function f_Y. The cases 2vi and 2vii are based on completely different line of thought. In this paper the Hellinger's, D_H, distance is adopted, which obeys the triangular relation, and offers possibilities, discussed in this section, for symmetry. The K-L divergence needs a special consideration, see section 2, as it is useful is a number of applications, as all the idm, Kamps (1989), that can be extended.

In principle the following Proposition 1 and the Corollary, for the multivariate and univariate case Kitsos and Toulias (2017) offer the way that these measures can be evaluated.

Proposition 1. Let us consider the p-dimensional normal distribution $N_p(\mu, \Sigma)$, $\mu \in R^p$, $\Sigma \in R^{p\times p}$. Then for the given rv's X and with

$$X \sim N_p(\mu_X, \Sigma_X), Y \sim N_p(\mu_Y, \Sigma_Y), \mu_X \neq \mu_Y, \Sigma_X \neq \Sigma_Y \tag{16}$$

The corresponding information distance measures (idm) are, see also

Nielsen (2011):

(1) The square of idm of Hellinger's D_H is:

$$D_H^2(X,Y) = 1 - \frac{|\Sigma_X|^{1/4}|\Sigma_Y|^{1/4}}{|\Sigma|}\exp[W]$$

with $\Sigma = \frac{1}{2}(\Sigma_X + \Sigma_Y)$ and $W = -\frac{1}{8}\left[(\mu_X - \mu_Y)^T\Sigma^{-1}(\mu_X - \mu_Y)\right]$

(2) The α-idm of Chernoff is:

$$\begin{aligned} D_{ch}^{(\alpha)}(X,Y) =& \frac{1}{2}\log\frac{|\alpha\Sigma_X + (1-\alpha)\Sigma_Y|}{|\Sigma_X|^{\alpha}|\Sigma_Y|^{1-\alpha}} + \\ & \frac{\alpha(1-\alpha)}{2}(\mu_X - \mu_Y)^T(\alpha\Sigma_X + (1-\alpha)\Sigma_Y)(\mu_X - \mu_Y), \\ & \alpha \in (0,1) \end{aligned} \tag{17}$$

(3) The Bhattacharyya's idm is:

$$D_B = \frac{1}{2}\log\frac{|\Sigma|}{|\Sigma_X\Sigma_Y|^{1/2}} - W, \quad \text{with } W \text{ and } \Sigma \text{ as in (21).} \tag{18}$$

Restriction: In this study, and for the problem we discuss, it is essential to impose the restriction that:

$$\frac{|\alpha\Sigma_X + (1-\alpha)\Sigma_Y|}{|\Sigma_X|^{\alpha}|\Sigma|^{1-\alpha}} > 1 \tag{19}$$

So that to obtain positive values for the corresponding distance.

Hellinger's distance is related to the Bhattacharyya coefficient BC, Bhattacharyya (1943), see also Case 3 above, as $D_H = (1 - BC)^{1/2}$ and as the Bhattacharyya's distance is originally defined to be $D_B = -ln(BC)$ it is easy to see that (19), as above, holds. Moreover for the Hellinger's distance:

Theorem 1. $D_H(P,Q) = [1 - BC(P,Q)]$, where $BC(P,Q) = \int\sqrt{fg}d\mu$ is known as the affinity of P and Q or the Bhattacharyya coefficient, related to Bhattacharyya's distance D_B.

Theorem 2. $D_H^2(P,Q) \leqslant d_{TV}(P,Q)$

Proof:

$$D_H^2(P,Q) = \int(\sqrt{f}-\sqrt{g})^2 d\mu \leqslant \int|\sqrt{f}-\sqrt{g}||\sqrt{f}+\sqrt{g}|d\mu = \int|f-g|d\mu = d_{TV}$$

Moreover from (23) the existence of Bhattacharyya's distance D_B depends on the value of $det(\Sigma)$, it is necessary to be positive one, as well as $det(\Sigma_X \Sigma_Y)$. It can be proved that the square of Hellinger's distance is smaller or equal the total variation one, while is related to Entropy Kamps (1989), among others.

It is clear that Chernoff's idm in (22) it is not symmetric. But when $\alpha = 1/2$ it is symmetric, while Hellinger's distance in (21) can be considered symmetric, while obeys, in principle, to triangular inequality. Moreover in univariate case is clear, symmetry holds for idm of Hellinger, and also when $\alpha = 1/2$ for the Chernoff idm, due to the following Corollary. Indeed:

Corollary 1. For the one-dimension rv's $X \sim N(\mu_X, \sigma_X^2), Y \sim (\mu_Y, \sigma_Y^2)$ holds:

$$D_H^2(X,Y) = 1 - \left(\frac{2\sigma_X \sigma_Y}{\sigma_X^2 + \sigma_Y^2} \right)^{1/2} \exp(w_0) \quad \text{with} \quad w_0 = -\frac{(\mu_X - \mu_Y)^2}{4(\sigma_X^2 + \sigma_Y^2)} \tag{20}$$

$$D_{ch}^{(\alpha)}(X,Y) = \frac{1}{2} \log \frac{\alpha\sigma_X^2 + (1-\alpha)\sigma_Y^2}{\left[\sigma_X^{\alpha}\sigma_Y^{(1-\alpha)}\right]^2} \tag{21}$$

$$D_B = -\log(1 - D_H^2) \tag{22}$$

Notice that $(\mu_X - \mu_Y)^2$ is a Euclidean distance of the means of the Normal distributions that the data set follow. Therefore, the involved parameter w_0 is a weighted distance measure and so does any continuous function, therefore $d^* = exp(w_0)$ is also a distance of the means. Hence:

Corollary 2. A distance between the mean values of the Normal univariate variables X and Y under consideration is expressed by $d^* = \exp(w_0)$. Therefore, Hellinger's idm between two Normal distributions is a function of the distance d^* between their mean values, $D_H^2(X,Y) = D_H^2(X,Y;d^*)$.

Moreover, it easy to see that:

Corollary 3. For the $^1/_2$-Chernoff information distance measure holds:

$$D_{Ch}^{1/2}(X,Y) = \frac{1}{2}\log\left[\frac{1}{2}\frac{\sigma_X^2 + \sigma_Y^2}{\sigma_X\sigma_Y}\right] = D_{Ch}^{1/2}(Y,X) \tag{23}$$

Notice that it is symmetric.

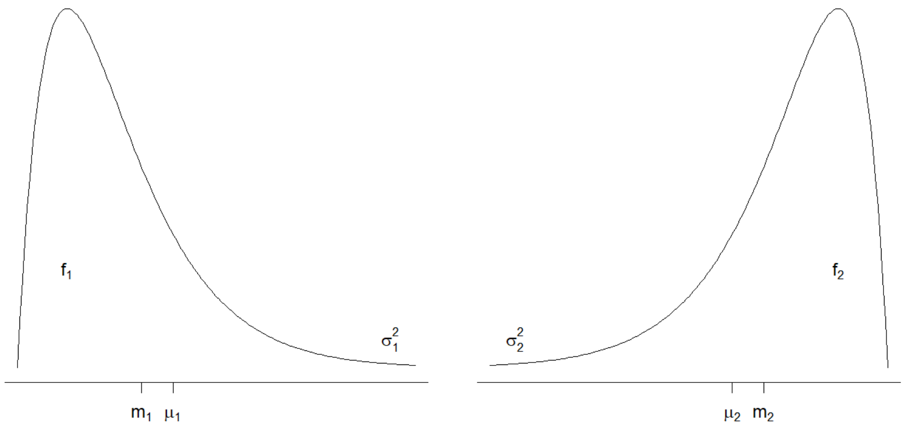

Figure 1. Compare f_1 and f_2 not means μ_1 and μ_2 or modes m_1 and m_2 or variances σ_1^2 and σ_2^2. Evaluate the distance between f_1 and f_2.

The $^1/_2$-Chernoff distance is related to Hellinger's square distance, see (22) and also (26). The distance, obeying to symmetry, but not to triangular inequality, Bhattacharyya's, D_B, is related to Hellinger's distance, see (18) is also adopting to this paper. Notice that Mahalanobis' distance is a special case of Bhattacharyya's distance Bhattacharyya (1983). That explains why Hellinger's distance plays a dominant role - most of the other distances related to it.

3. Discussion

We believe that distance measures can be proved useful in Bioassays and Medical Statistics, comparing the "whole population" and not only the means or variances, see Figure 1, Lu et. al. (2014), Shanon (1948), Wilson and Martinez (1997), and not only for the Information Theory, Cover and Thomas (1991). It was not that easy a number of statistical tests to be considered as distance measures Comber, et. al. (2020), while some application are using non-Euclidean distances, or are improving distance functions. It might be useful to be adopted on Biometrical methods, Kitsos and Sotiropoulos (2009), and see "how far" are the distribution function describing different populations, under the same disease under investigation. As an index of disagreement between the distributions, one could use the distance between the Normal distribution and the Cauchy distribution. Certainly more work is needed and we try to follow this line of though - Data Analysis under the Information Theory line of thought, Geoghegan (2008), adopting the appropriate distances. As far as the symmetry concerns, it might be essential to social-economical problems, but not necessarily for the Environmental problems - interest is focused to measure the information from the source of the pollution, say place X to the place, say Y, in a distance and not the opposite. Same line of thought might be for some Biometrically data - from the "source of the Ca" to other organs and not vice versa. Certainly more experience is needed in practical situations, and we are investigating such cases.

Acknowledgements

We would like to thank, the unknown referee for the very helpful comments, which improved this paper.

References

Bhattacharyya A. (1943): On a measure of divergence between two statistical populations defined by their probability distributions, Bull. Calcutta Math. Soc. 35, 99–110.

Blyth R.C. (1970): On the Inference and Decision Models of Statistics. The Annals of Mathematical Statistics 41(3), 1034–1058.

Comber A., Chi K., Huy Q.M., Nguyen Q., Lu B., Phe H.H., Harris P. (2020): Distance metric choice can both reduce and induce collinearity in geographically weighted regression, Environ. Plan. B Urban Anal. City Sci, 47, 489–507.

Cover T.M., Thomas J.A. (1991): Elements of Information Theory, Wiley, New York.

Darling D.A. (1957): The Kolmogorov - Smirnov, Gramer - Von Miss tests, Ann Math Stat. 28, 283–838.

Fisher R.A. (1922): On the mathematical foundation of theoretical statistics, Phil. Trans. Roy. Soc. London, series A 22, 165–177.

Geoghegan D.B. (2008): The Histographic conceptualization of information: A Critical Survey, IEEE Ann. Of History of Computing 30, 66–81.

Iliopoulou P., Kitsos C. (2019): Statistical applications in Geography: Spatial Analysis, in Contemporary Essays in Social Sciences & Geography: Theory and Policies, Korres, G., Kourliouros, E., and Kokkinou, A. eds., University of the Aegean, 212–224 (in Greek).

Kamps U. (1989): Hellinger distance and α-entropy in one parameter class of density functions, Stat. Papers, 30, 263–269.

Kitsos C., Iliopoulou P. (2021): Distance measures in spatial statistics, in Social Sciences & Geography: Theory, Methods and Spatial Analysis techniques, Kalabokidis, K., Korres, G., Soulakellis, N., and Feidas, H. eds. University of the Aegean, 96–108 (in Greek).

Kitsos C., Iliopoulou P. (2022): Adopting Information Distance Measures for Geographical Data Analysis. J. of Regional and Socio-Economic Issues, Vol 12(1), 6–27.

Kitsos C., Toulias L.T. (2010): New information measures for the generalized normal distribution, Information 1, 13–27.

Kitsos C., Toulias L.T. (2017): Hellinger distance between generalized normal distributions, British Journal of Mathematics and Computer Science 21, 1–16.

Kitsos C., Tavoularis K.N. (2009): Logarithmic Sobolev inequalities for information measures, IEEE Transactions on Information Theory 55 (2009), 2554–2561.

Kitsos C., Sotiropoulos M. (2009). Distance methods for bioassays, Biometrie und Medizinische Informatik Greifswalder Seminarberichte 15, 55–74.

Kolmogorov A. (1933): Sulla determinazione empirica di una legge di distribuzione, G. Ist. Ital. Attuari 4, 83–91.

Kullback S., Leibler A.R. (1951): On Information and sufficiency, Ann. Math. Statist. 22 (1951), 79–86.

Lu B., Charlton M., Harris P., Fotheringham A.S. (2014): Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data, Int J Geogr. Inf. Sci. 28, 660–681.

Mahalanobis C.P. (1936): On the generalized distance in Statistics, Proc. Nat. Inst. of Sciences of India, 2.

Nielsen F. (2011): Chernoff information of exponential families. IEEE Signal Processing Letters 20.3, 269 – 272

Saville J.D., Wood R.G. (1991): Statistical Methods : The Geometric Approach. Springer-Verlag.

Schervish J.M. (1995): Theory of Statistics, Springer.

Shanon C.E. (1948): A mathematical theory of communication, Bell Syst. Tech. J. 27, 379–423, 623–656.

Shapiro S.S., Wilks M.B. (1965): An analysis of variance test for normality (complete samples), Biometrika 65, 591–611.

Smirnov N. (1948): Table for estimating the goodness of fit of empirical distributions, Annals of Math. Stat. 19, 279–281.

Takasi S., Takashi S. (2011): Applied Analysis: Mathematical Methods in Natural Science, 2d Edition, Imperial College Press.

Wilson R., Martinez T. (1997): Improved heterogeneous distance functions, J. of Artif. Intell. Res. 6, 1–37.

Wolfowitz J. (1957): The minimum distance method, Ann. Math. Stat 28, 75–87.