Abstract

In this contribution we study Rényi pseudodistance estimators which are based on minimization of information-theoretic divergences between empirical and hypothetical probability distribution. These distances are more robust (than e.g. MLE estimators) against outliers and other measurement errors potentially present in the data sets. Robustness of these estimators is described by influence function. In [1] and [4] authors found explicit formulas for enumeration of Rényi distances in normal families and for their influence functions. We focus on finding explicit formulas for other families (Weibull, Cauchy, Exponential) and finding influence functions for these estimators. We perform computer simulations for pseudorandom contaminated and uncontaminated data sets, different sample sizes and different Rényi distance parameters.

1 Introduction and basic definitions

Let $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}^m\}$ be set of probability measures on measurable space $(\mathcal{X}, \mathcal{A})$. We will apply the estimators in the statistical model with i.i.d. observations $X_1, \dots, X_n$ governed by distribution P_0. Because we will be interested in robustness, we allow the case $P_0 \notin \mathcal{P}$ and therefore we define another set $\mathcal{P}^+ = \mathcal{P} \cup \{P_r\}$.

Definition 1.1. We say that mapping $\mathfrak{D} : \mathcal{P} \times \mathcal{P}^+ \to \mathbb{R}$ is *pseudodistance* between probability measures $P \in \mathcal{P}$ and $Q \in \mathcal{P}^+$ if it holds

$$\mathfrak{D}(P_\theta, Q) \geq 0 \quad \forall \theta \in \Theta \quad \text{and} \quad Q \in \mathcal{P}^+ \tag{1}$$

and

$$\mathfrak{D}(P_\theta, P_{\tilde{\theta}}) = 0 \iff \theta = \tilde{\theta}. \tag{2}$$

This pseudodistance is *decomposable* if there exist functionals so that $\mathfrak{D}^0 : \mathcal{P} \to \mathbb{R}$, $\mathfrak{D}^1 : \mathcal{P}^+ \to \mathbb{R}$ and measurable mapping $\rho_\theta : \mathcal{X} \to \mathbb{R}$, $\theta \in \Theta$, so that $\forall \theta \in \Theta$ and $\forall Q \in \mathcal{P}^+$ the expectation $\int \rho_\theta \mathrm{d}Q$ exists and

$$\mathfrak{D}(P_\theta, Q) = \mathfrak{D}^0(P_\theta) + \mathfrak{D}^1(Q) + \int \rho_\theta \mathrm{d}Q. \tag{3}$$

Definition 1.2. We say that a functional $T_\mathfrak{D} : \mathcal{Q} \to \Theta$, for $\mathcal{Q} = \mathcal{P}^+ \cup \mathcal{P}_{\text{emp}}$ defines minimum pseudodistance estimator (min $\mathfrak{D}$-estimator) if $\mathfrak{D}(P_\theta, Q)$ is a decomposable pseudodistance on $\mathcal{P} \times \mathcal{P}^+$ and parameters $T_\mathfrak{D}(Q) \in \Theta$ minimize $\mathfrak{D}^0 + \int \rho_\theta \mathrm{d}Q$, that means

$$T_\mathfrak{D}(Q) = \arg\min_{\theta \in \Theta} \left[\mathfrak{D}^0(P_\theta) + \int \rho_\theta \mathrm{d}Q \right] \quad \forall Q \in \mathcal{Q}. \tag{4}$$

In particular, for $Q = P_n = \frac{1}{n}\sum_{i-1}^{n} \delta_{X_i} \in \mathcal{P}_{emp}$

$$\hat{\theta}_{\mathfrak{D},n} = T_{\mathfrak{D}}(P_n) = \arg\min_{\theta\in\Theta} \left[\mathfrak{D}^0(P_\theta) + \frac{1}{n}\sum_{i-1}^{n} \rho_\theta(X_i) \right]. \tag{5}$$

Every min $\mathfrak{D}$-estimator is Fisher consistent in the sense that

$$T_{\mathfrak{D}}(P_{\theta_0}) = \arg\min_{\theta\in\Theta} \mathfrak{D}(P_\theta, P_{\theta_0}) = \theta_0, \quad \forall\theta_0 \in \Theta. \tag{6}$$

Theorem 1.1. *Let for some $\beta > 0$*

$$p^\beta, q^\beta, \ln p \in \mathrm{L}_1(Q), \quad \forall P \in \mathcal{P}, Q \in \mathcal{P}^+.$$

holds. Then $\forall\alpha$, $0 < \alpha \le \beta$, and for $P \in \mathcal{P}$, $Q \in \mathcal{P}^+$ the expression

$$\mathcal{R}_\alpha(P,Q) = \frac{1}{1+\alpha}\ln\left(\int p^\alpha \mathrm{d}P\right) + \frac{1}{\alpha(1+\alpha)}\ln\left(\int q^\alpha \mathrm{d}Q\right) - \frac{1}{\alpha}\ln\left(\int p^\alpha \mathrm{d}Q\right) \tag{7}$$

represents the family of pseudodistances decomposable in the sense

$$\mathcal{R}_\alpha(P,Q) = \mathcal{R}_\alpha^0(P) + \mathcal{R}_\alpha^1(Q) - \frac{1}{\alpha}\ln\left(\int p^\alpha \mathrm{d}Q\right),$$

where

$$\mathcal{R}_\alpha^0(P) = \frac{1}{1+\alpha}\ln\left(\int p^\alpha \mathrm{d}P\right), \quad \mathcal{R}_\alpha^1(Q) = \frac{1}{\alpha(1+\alpha)}\ln\left(\int q^\alpha \mathrm{d}Q\right).$$

Moreover for $\alpha \searrow 0$ it holds

$$\mathcal{R}_0(P,Q) = \lim_{\alpha\to 0} \mathcal{R}_\alpha(P,Q) = \int (\ln q - \ln p)\, \mathrm{d}Q$$

References

[1] Michel Broniatowski, Igor Vajda. *Several Applications of Divergence Criteria in Continuous Families.* Research report No 2257 September 2009, UTIA AV CR, Prague, 2009.

[2] Igor Vajda. *Information - Theoretic Methods in Statistics.* Research report No 1834 October 1995, UTIA AV CR, Prague, 1995.

[3] Michel Broniatowski, Aida Toma, Igor Vajda. *Decomposable Pseudodistances and Applications in Statistical Estimation.* arXiv:1104.1541v1, 2011.

[4] Radim Demut *Robust properties of minimum divergence density estimators.* Diploma Thesis, ČVUT, Prague 2010.

[5] By Ayanendranath Basu, Ian R. Harris, Nils L. Hjort, M. C. Jones *Robust and eficient estimation by minimising a density power divergence.* In Biometrika, 85, 549-559, 1998.

[6] Friedrich Liese, Igor Vajda. *On Divergences and Informations in Statistics and Information Theory.* IEEE Transactions on Information Theory, Vol. 52, No. 10,4394-4412, 2006.