

Influence function analysis of the restricted minimum divergence estimators: A general form

Abhik Ghosh

*Indian Statistical Institute
203 B. T. Road
Kolkata 700 018, India
e-mail: abhianik@gmail.com*

Abstract: The minimum divergence estimators have proved to be useful tools in the area of robust inference. The robustness of such estimators are often measured using the classical Influence Function analysis. However, in many complex situations like that of testing a composite null hypothesis require the estimators to be restricted over some proper subspace of the parameter space. The robustness of these restricted minimum divergence estimators are very crucial in order to have overall robust inference. In this paper we provide a comprehensive description of the robustness of such restricted estimators in terms of their influence function for a general class of density based divergences along with their unrestricted versions. In particular, the robustness of some popular minimum divergence estimators are also demonstrated under certain usual restrictions through examples. Thus the paper provides a general framework for the influence function analysis of a large class of minimum divergence estimators with or without restrictions on the parameters and provides theoretical solutions for measuring the impact of the parameter restrictions on the robustness of the corresponding estimators.

AMS 2000 subject classifications: Primary 62F30, 62F35.

Keywords and phrases: Minimum divergence estimator, robustness, influence function, parameter restriction.

Received September 2014.

Contents

1	Introduction	1018
2	Density-based minimum divergence estimators (MDEs) and their influence functions: A general form	1019
3	The influence function of the restricted MDE: A general form	1023
4	Applications: Some particular divergences	1029
4.1	Disparity measures	1029
4.2	Density power divergences	1031
4.3	S -divergence family	1033
5	Impact of restrictions on the robustness: Some qualitative illustrations	1034
6	Conclusion	1036
	Acknowledgment	1037

Appendix: Proof of the results	1037
A.1 Proof of Theorem 2.1	1037
A.2 Proof of Theorem 3.1	1038
A.3 Detailed calculation for Example 3.3	1038
References	1039

1. Introduction

The minimum divergence approach has proved to be a very useful one in the context of parametric statistical inference. The idea behind this approach is to quantify the discrepancy between the sample data and the parametric model through an appropriate divergence and minimize this discrepancy measure over the parameter space. There are two ways of such quantification in the literature – either through distribution functions or through probability density functions (with respect to some suitable dominating measure). Most of the density based minimum divergence methods are seen to be particularly useful due to their strong robustness properties along with high efficiencies.

However, in many complex statistical problems we need to estimate the parameter of interest under some pre-specified restrictions on the parameter space. For example, when testing a composite null hypothesis, we need to estimate the parameter under the restrictions imposed by the null hypothesis. For such cases we need to minimize the divergence measures only over a restricted subspace of the parameter space. Simpson [20], Lindsay [17] and Basu et al. [4] used such restricted minimum divergence estimators in the context of testing statistical hypotheses and derived their asymptotic properties. But they did not consider the robustness of these restricted estimators separately, although it is also very important in order to obtain robust solutions for the overall inference problem.

One can provide several examples of such a theoretical void in the literature in understanding the robustness of estimators under parameter restrictions. To motivate the issue further, let us consider the problem of testing the equality of means for two independent normal populations, $N(\theta_1, \sigma^2)$ and $N(\theta_2, \sigma^2)$ with common variance σ^2 . For testing this hypothesis, we need to estimate the common mean under the null hypothesis $\theta_1 = \theta_2$. It is well-known that the maximum likelihood estimators of the mean and the variance under the normal model are highly affected in the presence of outliers in the data, which can be examined by the unbounded nature of the corresponding influence functions. However, it is often empirically observed that the conclusion of the testing problem of the hypothesis $\theta_1 = \theta_2$ based on the maximum likelihood estimator of the common mean is not affected significantly in the presence of outliers having a similar pattern in both the samples, although the corresponding estimate of σ^2 is affected significantly. Further, if outliers are present in only one sample, then the end result of the testing problem may get severely affected. This is quite intuitive, but there is no theoretical result in terms of influence functions to support this fact. Similar problems exist even in the case of one normal sample, where a similar question about the robustness of the estimators of the mean or the

variance parameter under the restriction on the other parameter has no answer from the existing theory of robustness. Yet it is very natural to estimate the mean under a known variance or estimate the variance under a known mean. The problem becomes more difficult beyond the case of normal models and in presence of more complicated restrictions. Indeed, the robustness aspect of the restricted minimum divergence estimators are not well studied in the literature at all. In this paper we will consider this very important issue and describe the robustness of the general minimum divergence estimators in terms of the Influence Function Analysis.

The rest of this paper is organized as follows. Section 2 describes the concept of the minimum divergence estimators and presents a general form for their influence function analysis. In Section 3, we will derive a general form of the influence function of the restricted minimum divergence estimators. In Section 4, we will apply the general results in case of some popular divergences – disparities, density power divergences and S-divergences – under some standard restrictions on the parameter of interest. The qualitative impact of the usual restrictions on the robustness of the corresponding minimum divergence estimators is discussed in Section 5, where we have tried to answer all the above motivating questions along with some more examples and discussions. Finally we end the paper with some concluding remark in Section 6. For a better flow of presentation, we have move all the proofs and detailed calculations to the [Appendix](#).

2. Density-based minimum divergence estimators (MDEs) and their influence functions: A general form

Let us begin our discussion with a general parametric estimation problem. We have n independent and identically distributed observations X_1, \dots, X_n from a distribution G . We want to model it by a parametric family of distributions

$$\mathcal{F}_\theta = \{F_\theta : \theta \in \Theta \subseteq \mathbb{R}^p\}.$$

We assume the support of G and the parametric model \mathcal{F}_θ be the same. Also let both G and \mathcal{F}_θ belong to \mathcal{G} , the class of all distributions having densities with respect to an appropriate σ -finite measure μ on the relevant σ -field and f_θ, g be the density functions of F_θ, G respectively with respect to μ . We want to estimate the parameter θ based on the available sample data.

In case of density-based minimum divergence estimation, this is done by choosing the model element that provides the closest match to the data where the separation between the model and data is quantified by a nonnegative function $\rho(\cdot, \cdot)$ from $\mathcal{G} \times \mathcal{G}$ to $[0, \infty)$ that equals zero if and only if its arguments are identically equal. Such a function $\rho(\cdot, \cdot)$ is termed as a Statistical Divergence and the estimator $\hat{\theta}$ of θ obtained by minimizing $\rho(\hat{g}, f_\theta)$ with respect to $\theta \in \Theta$, where \hat{g} is some nonparametric estimator of g based on the sample data, is called the Minimum Divergence Estimator (MDE) of θ .

Definition 2.1. A Statistical Divergence measure is a nonnegative function

$$\rho(\cdot, \cdot) : \mathcal{G} \times \mathcal{G} \mapsto [0, \infty),$$

which satisfies

$$\rho(g, f) = 0 \quad \Leftrightarrow \quad g = f \text{ a.e. } [\mu], \quad \text{for all } f, g \in \mathcal{G}.$$

Definition 2.2. The Minimum Divergence Estimator (MDE) $\widehat{\theta}_\rho$ of the parameter θ corresponding to the divergence $\rho(\cdot, \cdot)$ is defined as

$$\widehat{\theta}_\rho = \arg \min_{\theta \in \Theta} \rho(\widehat{g}, f_\theta).$$

In terms of statistical functionals, the Minimum Divergence Functional $T_\rho(G)$ corresponding to the divergence $\rho(\cdot, \cdot)$ is defined by the relation

$$\rho(g, f_{T_\rho(G)}) = \min_{\theta \in \Theta} \rho(g, f_\theta),$$

provided such a minimum exists.

There are several popular examples of statistical divergences that generate highly robust and efficient estimators including the disparity family [17], Cressie-Read power divergences family [5], density power divergence family [2] etc. See Csiszár [6, 7, 8], Ali and Silvey [1], Vajda [21], Pardo [18] and Basu et al. [3] for further examples and the details of the minimum divergence estimators including their asymptotic and robustness properties. Throughout the present paper, we will restrict our attention to a special class of divergences as given in the following definition which contains most of the popular divergences used for statistical inference.

Definition 2.3. Define a class of divergence measures given by

$$\rho(g, f) = \int D(g(x), f(x)) d\mu(x) \quad (2.1)$$

for some suitable function $D(\cdot, \cdot) : \mathbb{R} \times \mathbb{R} \mapsto [0, \infty)$ which satisfies $D(a, b) = 0$ whenever $a = b$.

The class of divergences defined in Definition 2.3 contains most of the popular statistical divergence measures for particular choices of the $D(\cdot, \cdot)$ function. For example, the choice

$$D(a, b) = \phi\left(\frac{a}{b} - 1\right) b \quad (2.2)$$

for a non-negative thrice differentiable strictly convex function ϕ on $[-1, \infty)$ with $\phi(0) = 0 = \phi'(0)$ generates the class of disparities [17]. For any $\alpha \geq 0$, the choice

$$D(a, b) = \begin{cases} b^{1+\alpha} - \left(\frac{1+\alpha}{\alpha}\right) b^\alpha a + \frac{1}{\alpha} a^{1+\alpha}, & \text{if } \alpha > 0, \\ a \log(a/b), & \text{if } \alpha = 0, \end{cases} \quad (2.3)$$

generates the density power divergence family of Basu et al. [2]. We will consider these two classes of divergences in detail again in Section 4.

Now, let us consider the estimating equation of the MDE based on the general divergences considered in Definition 2.3 which turns out to be

$$\nabla \rho(g, f_\theta) = \int \nabla D(g(x), f_\theta(x)) d\mu(x) = 0. \quad (2.4)$$

Here ∇ represents the derivative with respect to θ . Note that, this estimating equation does not necessarily give us an M-estimator; it does so only when the terms in $\nabla D(g, f_\theta)$ containing f_θ include only linear functions of g or some constant independent of g . However, the number of divergences satisfying this condition is limited [See, e.g. 19] so that we can not always apply the theory of M-estimators to describe the properties of the MDEs. However, all the MDEs obtained as a solution to (2.4) will be Fisher consistent by definition of the divergence $\rho(\cdot, \cdot)$.

The popularity of the MDEs is largely due to their strong robustness properties and in this context a useful tool is the Influence Function [14, 15] which is an indicator of their classical first-order robustness, as well as of their asymptotic efficiency. To obtain the influence function of the minimum divergence estimators based on the divergence $\rho(\cdot, \cdot)$, we consider the ϵ contaminated version of the true density g given by $g_\epsilon(x) = (1 - \epsilon)g(x) + \epsilon\chi_y(x)$. The corresponding contaminated distribution function is given by $G_\epsilon(x) = (1 - \epsilon)G(x) + \epsilon\wedge_y(x)$; here $\chi_y(x)$ and $\wedge_y(x)$ are respectively density and distribution functions of the degenerate distribution at y . Let $\theta^g = T_\rho(G)$ and $\theta_\epsilon = T_\rho(G_\epsilon)$ be the functional obtained via the minimization of $\rho(g, f_\theta)$ and $\rho(g_\epsilon, f_\theta)$ respectively. Then the influence function of the minimum divergence functional $T_\rho(\cdot)$ is defined as

$$IF(y, T_\rho, G) = \left. \frac{\partial \theta_\epsilon}{\partial \epsilon} \right|_{\epsilon=0}.$$

Suppose, $D^{(i)}(\cdot, \cdot)$ denotes the first order partial derivative of $D(\cdot, \cdot)$ with respect to its i^{th} argument, $D^{(i,j)}(\cdot, \cdot)$ denotes its second order partial derivative with respect to i^{th} and j^{th} arguments ($i, j = 1, 2$). Further, assume that the standard regularity conditions hold for the model densities so that all above derivatives exist and can be interchanged with the integrals with respect to μ . Then the following theorem provides a general form of the influence function of the MDEs corresponding to the particular divergence given in Equation (2.1); for brevity in presentation the proof is given in Appendix A.1.

Theorem 2.1. *Under standard regularity conditions on model densities, the influence function of the minimum divergence functional T_ρ corresponding the particular divergences given by Equation (2.1) has the form*

$$IF(y, T_\rho, G) = N(\theta^g)^{-1}[\xi(\theta^g) - M(y; \theta^g)], \quad (2.5)$$

where $\theta^g = T_\rho(G)$ and

$$\begin{aligned} N(\theta) &= \int \left[D^{(2)}(g(x), f_\theta(x)) \nabla_2 f_\theta(x) \right. \\ &\quad \left. + D^{(2,2)}(g(x), f_\theta(x)) \{ \nabla f_\theta(x) \} \{ \nabla f_\theta(x) \}^T \right] d\mu(x), \\ M(y; \theta) &= D^{(1,2)}(g(y), f_\theta(y)) [\nabla f_\theta(y)], \end{aligned}$$

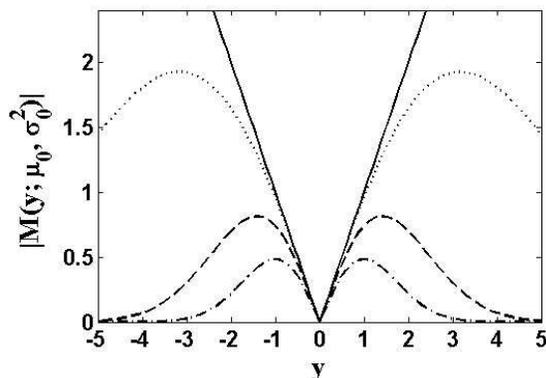


FIG 1. Plot of the function $|M(y; \mu_0, \sigma_0^2)|$ in Example 2.1 with $\mu_0 = 0$, $\sigma_0 = 1$ over the contamination point y for different α [solid line: $\alpha = 0$, dotted line: $\alpha = 0.1$, dashed line: $\alpha = 0.5$ and dot-dashed line: $\alpha = 1$].

$$\xi(\theta) = \int D^{(1,2)}(g(x), f_\theta(x))[\nabla f_\theta(x)]g(x)d\mu(x) = E_g[M(X; \theta)].$$

In particular, when the true distribution G belongs to the parametric model with $g(x) = f_{\theta_0}(x)$ for some $\theta_0 \in \Theta$, we get $\theta^g = \theta_0$ and the above influence function becomes

$$IF(y, T_\rho, F_{\theta_0}) = N(\theta_0)^{-1}[\xi(\theta_0) - M(y; \theta_0)],$$

where each of $N(\theta_0)$, $\xi(\theta_0)$ and $M(y; \theta_0)$ are evaluated at $g = f_{\theta_0}$. Therefore, the influence functions of the MDEs will be bounded at the model for all those divergences for which the function $|M(y; \theta)|$ is bounded in y for all θ . This will further imply the robustness of the corresponding MDEs.

Example 2.1. Consider the estimation of normal mean θ based on an i.i.d. sample X_1, \dots, X_n from the $N(\theta, \sigma^2)$ distribution with known σ^2 . Assume that $\theta \in \mathbb{R}$ and the true data generating density belongs to the model with true parameter value θ_0 . Then, for the minimum density power divergence estimator [2] with tuning parameter α defined by Equation (2.3), we have

$$M(y; \theta_0) = -(1 + \alpha) \frac{(y - \theta_0)}{(2\pi)^{\alpha/2} \sigma_0^{\frac{\alpha}{2} + 2}} e^{-\frac{\alpha(y - \theta_0)^2}{2\sigma_0^2}}, \quad y \in \mathbb{R}, \quad \alpha \geq 0;$$

see Section 4 for the derivation in case of general model families. Note that, the function $|M(y; \theta_0)|$ is bounded in y whenever $\alpha > 0$ but unbounded at $\alpha = 0$ (see Figure 1). So from the general theory developed above, the minimum density power divergence estimators with $\alpha > 0$ are robust with respect to the outliers and that corresponding to $\alpha = 0$ is non-robust. This fact exactly coincides with the corresponding results derived independently in Basu et al. [2]. \square

Further, as expected from the interpretation of the influence function by Hampel et al. [16], we have

$$\begin{aligned} \int IF(y, T_\rho, G)d(G(y)) &= \int IF(y, T_\rho, G)g(y)d\mu(y) \\ &= N(\theta_0)^{-1} \int [\xi(\theta_0) - M(y; \theta_0)]g(y)d\mu(y) \\ &= N(\theta_0)^{-1} \left[\xi(\theta_0) - \int M(y; \theta_0)g(y)d\mu(y) \right] \\ &= 0. \end{aligned}$$

Then the asymptotic distribution of the MDE follows from Hampel et al. [16], as noted in the following remark.

Remark 2.1. If the MDE $\hat{\theta}_\rho = T_\rho(\hat{G})$ is an \sqrt{n} -consistent estimator of $T_\rho(G)$, then under standard regularity conditions the asymptotic distribution of $\sqrt{n}(\hat{\theta}_\rho - T_\rho(G))$ is asymptotically normal with mean zero and variance

$$\begin{aligned} V(T_\rho, G) &= \int IF(y, T_\rho, G)IF(y, T_\rho, G)^T d(G(y)) \\ &= N(\theta^g)^{-1}Var_g[M(X; \theta^g)]N(\theta^g)^{-1}, \end{aligned}$$

where $Var_g(\cdot)$ denotes the variance under the distribution of g .

3. The influence function of the restricted MDE: A general form

We will now consider the case of restricted minimum divergence estimators and derive a general expression for its influence function extending the concepts of the previous section. Consider the set-up of Section 2, but now we want to estimate the parameter θ under a restricted (proper) subspace Θ_0 of the entire parameter space Θ . In case of composite null hypotheses, the subspace Θ_0 is given by the null parameter space. Generally, we can define the subspace Θ_0 by a set of $r(< p)$ restrictions of the form

$$h(\theta) = 0 \quad \text{on } \Theta_0, \tag{3.1}$$

for some function $h : \mathbb{R}^p \mapsto \mathbb{R}^r$ satisfying the property that the $p \times r$ matrix

$$H(\theta) = \frac{\partial h(\theta)}{\partial \theta}$$

exists with rank r and is continuous in θ . Thus, under Θ_0 , the parameter θ essentially contains $p - r$ independent components.

We can solve the above estimation problem by minimizing $\rho(\hat{g}, f_\theta)$ with respect to $\theta \in \Theta_0$ and the estimator obtained from this minimization exercise will be called the Restricted Minimum Divergence Estimator (RMDE).

Definition 3.1. The Restricted Minimum Divergence Estimator (RMDE) $\tilde{\theta}_\rho$ of the parameter θ corresponding to the divergence $\rho(\cdot, \cdot)$ and restriction (3.1)

is defined as

$$\tilde{\theta}_\rho = \arg \min_{\theta \in \Theta_0} \rho(\hat{g}, f_\theta) = \arg \min_{\theta; h(\theta)=0} \rho(\hat{g}, f_\theta).$$

In terms of statistical functional, the Restricted Minimum Divergence Functional $\tilde{T}_\rho(G)$ is then given by the relation

$$\rho(g, f_{\tilde{T}_\rho(G)}) = \min_{\theta \in \Theta_0} \rho(g, f_\theta) = \min_{h(\theta)=0} \rho(g, f_\theta),$$

provided such a minimum exists.

We can easily solve this minimization problem using the Lagrange multiplier method. Lindsay [17] and Basu et al. [4] derived the asymptotic distribution of such RMDE for the disparities and the density power divergences respectively. Here we will derive the influence function of the Restricted Minimum divergence functional.

As in Section 2, let us consider the ϵ -contaminated density g_ϵ , contaminated distribution G_ϵ and define $\tilde{\theta}^g = \tilde{T}_\rho(G)$ and $\tilde{\theta}_\epsilon = \tilde{T}_\rho(G_\epsilon)$. Note that $\tilde{\theta}_\epsilon$ is the minimizer of $\rho(g_\epsilon, f_\theta)$ subject to (3.1). We will consider only the restrictions which can be substituted explicitly in the expression of $\rho(g_\epsilon, f_\theta)$ before taking its derivatives with respect to θ . Then, the resulting derivative will be zero at $\theta = \tilde{\theta}_\epsilon$, which can be used to derive the required influence function. The following theorem presents the general expression for the influence function of the restricted minimum divergence functional; see Appendix A.2 for a proof.

Theorem 3.1. *Consider the notations developed so far and assume that rank of $H(\tilde{\theta}^g)$ is r . Then the influence function of the restricted minimum divergence estimator corresponding to the divergences (2.1) is given by*

$$\begin{aligned} & IF(y, \tilde{T}_\rho, G) \\ &= \left[N_0(\tilde{\theta}^g)^T N_0(\tilde{\theta}^g) + H(\tilde{\theta}^g) H(\tilde{\theta}^g)^T \right]^{-1} N_0(\tilde{\theta}^g)^T \left[\xi_0(\tilde{\theta}^g) - M_0(y; \tilde{\theta}^g) \right], \end{aligned} \quad (3.2)$$

where $\tilde{\theta}^g = \tilde{T}_\rho(G)$ and $N_0(\theta)$, $\xi_0(\theta)$, $M_0(y; \theta)$ are the same as $N(\theta)$, $\xi(\theta)$, $M(y; \theta)$ respectively but with an additional prior restriction of $h(\theta) = 0$.

In particular, if the true density belongs to the model family, i.e., $g = f_{\theta_0}$ for some θ_0 satisfying $h(\theta_0) = 0$, then we put $\tilde{\theta}^g = \theta_0$ in (3.2) to obtain the corresponding influence function. Therefore, the influence functions of the RMDEs will be bounded at the model for all those divergences for which the function $|M_0(y; \theta)|$ is bounded in y for all θ . In particular, whenever the influence function of the MDE at the model is bounded, that of the corresponding RMDE will also be bounded at the model for any given set of restrictions; but the converse is not true as shown in the following example.

Example 3.1 (Continuation of Example 2.1). Consider again the problem of estimating normal mean θ as in Example 2.1. But, now let us assume that the mean θ can take only values in a proper subset Θ_0 of the real line. Then, for this

restricted case, the function $M_0(y; \theta_0)$ corresponding to the minimum density power divergence estimator can be seen to have the form

$$M_0(y; \theta_0) = \begin{cases} M(y; \theta_0), & \text{if } y \in \Theta_0, \\ 0, & \text{if } y \notin \Theta_0, \end{cases} \quad \alpha \geq 0,$$

where $M(y; \theta_0)$ is as given in Example 2.1.

Then the function $M_0(y; \theta_0)$ is bounded in y whenever $\alpha > 0$ for any choice of Θ_0 ; but for $\alpha = 0$, the function $M_0(y; \theta_0)$ is bounded only if Θ_0 is a bounded subset and unbounded otherwise. Therefore, the minimum density power divergence estimators of the normal mean θ corresponding to $\alpha > 0$, which are originally robust without any restriction on the parameter space (Example 2.1), remain robust also under any restrictions imposed on the parameter.

However, the corresponding estimator with $\alpha = 0$ (which is in fact the maximum likelihood estimator as shown in Basu et al. [2] and non-robust in presence of no restriction on the parameter space) becomes robust if we impose suitable restrictions on the set Θ of possible values of the parameter to make it bounded (e.g., the mean θ belongs to a finite interval $[a, b]$ or it is a fixed given real number specified by the simple null hypothesis). But, for the restrictions where the set of possible parameter values is unbounded (e.g., set of all positive reals) this restricted estimator continues to be non-robust with respect to outliers. \square

Further, as in the unrestricted case, one can verify that

$$\int IF(y, \tilde{T}_\rho, G)d(G(y)) = \int IF(y, \tilde{T}_\rho, G)g(y)d\mu(y) = 0.$$

Then, the asymptotic distribution of the RMDE follows by an argument similar to that presented in Hampel et al. [16] for unrestricted case.

Remark 3.1. If the RMDE $\tilde{\theta}_\rho = \tilde{T}_\rho(\hat{G})$ is an \sqrt{n} -consistent estimator of $\tilde{T}_\rho(G)$, then the asymptotic distribution of $\sqrt{n}(\tilde{\theta}_\rho - \tilde{T}_\rho(G))$ is asymptotically normal with mean zero and variance

$$\begin{aligned} V(\tilde{T}, G) &= \int IF(y, \tilde{T}_\rho, G)IF(y, \tilde{T}_\rho, G)^T d(G(y)) \\ &= \left[N_0(\tilde{\theta}^g)^T N_0(\tilde{\theta}^g) + H(\tilde{\theta}^g)H(\tilde{\theta}^g)^T \right]^{-1} N_0(\tilde{\theta}^g)^T Var_g[M_0(y; \tilde{\theta}^g)] \\ &\quad \times N_0(\tilde{\theta}^g) \left[N_0(\tilde{\theta}^g)^T N_0(\tilde{\theta}^g) + H(\tilde{\theta}^g)H(\tilde{\theta}^g)^T \right]^{-1}. \end{aligned}$$

At the model $g = f_{\theta_0}$ for some θ_0 satisfying $h(\theta_0) = 0$, the above expression of asymptotic variance can be further simplified by substituting $\tilde{\theta}^g = \theta_0$ and using the structure of corresponding model density.

We will now explore a couple of particular cases of restrictions that are commonly used in parametric inference.

Example 3.2. First we consider a simple and perhaps most popular case of restrictions where few components of the parameter θ is pre-specified. Precisely,

let $\theta = (\theta_1 \ \theta_2)^T$ where θ_1 is an r -vector and its value is specified as $\theta_{1,0}$. Thus we consider the RMDE of θ under the restriction $\theta_1 = \theta_{1,0}$. Note that, in this case, the RMDE of θ_1 will be fixed at $\theta_{1,0}$ having no influence of the outliers and the influence function analysis of the RMDE of θ_2 should be the same as that of the unrestricted MDE considering θ_2 as the only parameter of interest. We will now apply the general formulas derived above to this simple case and verify if the general results are in-line with this common intuition.

Now, the influence function of the MDE of θ in the unrestricted case is given by

$$IF(y, T_\rho, G) = N(\theta^g)^{-1}[\xi(\theta^g) - M(y; \theta^g)].$$

Let us partition this result in terms of θ_1 and θ_2 to get

$$M(y; \theta) = (M_1(y; \theta) \ M_2(y; \theta))^T, \quad \xi(\theta) = (\xi_1(\theta) \ \xi_2(\theta))^T$$

and

$$N(\theta) = \begin{pmatrix} N_{11}(\theta) & N_{12}(\theta) \\ N_{12}(\theta)^T & N_{22}(\theta) \end{pmatrix},$$

where M_1 and ξ_1 are r -vectors and N_{11} is the matrix of order $r \times r$. Using simple matrix algebra, we get the influence functions of the MDE of θ_1 and θ_2 separately, although they are not independent for general parametric models. However, one can verify that these two (unrestricted) influence functions of θ_1 and θ_2 will be independent whenever $N_{12}(\theta^g) = O$, the null matrix of appropriate order.

Next, consider the restricted case where we have $h(\theta) = \theta_1 - \theta_{1,0}$ and

$$H(\theta) = \begin{bmatrix} I_r \\ O_{(p-r) \times r} \end{bmatrix}.$$

Here I_r represents the identity matrix of order r and $O_{p \times q}$ represents the null matrix of order $p \times q$ (we will drop the subscript $p \times q$ whenever the order is clear from the context and will denote the square null matrix of order $p \times p$ by just O_p). Also, note that $\tilde{\theta}^g = (\theta_{1,0}, \theta_2^g)^T$ and hence

$$N_0(\tilde{\theta}^g) = \begin{bmatrix} O_r & O \\ O & N_{22}((\theta_{1,0}, \theta_2^g)^T) \end{bmatrix},$$

$M_0(y; \tilde{\theta}^g) = [0_r^T \ M_2(y; (\theta_{1,0}, \theta_2^g)^T)]^T$ and $\xi_0(\tilde{\theta}^g) = [0_r^T \ \xi_2((\theta_{1,0}, \theta_2^g)^T)]^T$. Therefore, using Theorem 3.1, the influence function of the RMDE of θ becomes

$$IF(y, \tilde{T}_\rho, G) = \begin{pmatrix} 0_r \\ N_{22}((\theta_{1,0}, \theta_2^g)^T)^{-1} [\xi_2((\theta_{1,0}, \theta_2^g)^T) - M_2(y; (\theta_{1,0}, \theta_2^g)^T)] \end{pmatrix}.$$

Thus the influence function of the RMDE corresponding to θ_1 turns out to be identically zero and that corresponding to θ_2 is the same as that obtained in the unrestricted case considering θ_2 only, as expected. \square

Remark 3.2. Note that Theorem 3.1 can only be applied provided the restrictions are such that $rank(H(\tilde{\theta}^g)) = r$. But in many practical situations we need to consider restrictions for which the rank is strictly less than r and we can not apply the above Theorem 3.1 directly to obtain the influence function of the corresponding RMDEs. However, the arguments presented to derive the theorem (in Appendix A.2) can still be applied with some small modifications as required. One such common restriction is considered in Example 3.3 below.

Example 3.3. Consider a slightly complicated case of restrictions where the first r components of θ depend among themselves through only one unknown parameter, say β . Suppose $\theta = (\theta_1, \theta_2)^T$ where θ_1 is an r -vector satisfying $\theta_1 = \phi(\beta)$ for a known function $\phi : \mathbb{R} \mapsto \mathbb{R}^r$. We assume that $\phi(\beta) = (\phi_1(\beta), \dots, \phi_r(\beta))^T$ and each ϕ_i is a twice differentiable real function with a non-zero derivative. Again, let us consider the partitions of the matrices $N(\theta)$, $\xi(\theta)$ and $M(y; \theta)$ in terms of θ_1 and θ_2 as in Example 3.2.

To derive the influence function of the RMDEs in this case, note that $h(\theta) = \theta_1 - \phi(\beta)$ so that

$$H(\theta) = \begin{pmatrix} I_r - B \\ O_{(p-r) \times r} \end{pmatrix},$$

where the $r \times r$ matrix B is defined as $B = \frac{\partial \phi(\beta)}{\partial \theta_1}$. Note that, the $(i, j)^{th}$ element of the matrix B is given by $b_{ij} = \frac{\phi'_j(\beta)}{\phi'_i(\beta)}$ for each $i, j = 1, \dots, r$, where $\phi'_i(\beta)$ denote the first derivative with respect to β . Clearly $rank(H(\tilde{\theta}^g)) = r - 1$ and so Theorem 3.1 cannot be applied directly to obtain the influence function of the RMDE here. However, we can restart with the set of Equations (A.2) and (A.3) in the proof of Theorem 3.1 with $\theta = \tilde{\theta}^g = (\phi(\tilde{\beta}^g), \tilde{\theta}_2^g)^T$ and solve them for the expression of the influence function.

Let us partition the influence function $IF(y, \tilde{T}_\rho, G)$ of \tilde{T}_ρ in terms of that of the functionals $\tilde{T}_{\rho,1}$ and $\tilde{T}_{\rho,2}$ corresponding to θ_1 and θ_2 respectively as

$$IF(y, \tilde{T}_\rho, G) = \begin{bmatrix} IF(y, \tilde{T}_{\rho,1}, G) \\ IF(y, \tilde{T}_{\rho,2}, G) \end{bmatrix}.$$

Then, starting from Equations (A.2) and (A.3), we can show that (details are presented in Appendix A.3) the first component $IF(y, \tilde{T}_{\rho,1}, G)$ is given by the solution of

$$\begin{aligned} & \left\{ B \left[N_{11}(\tilde{\theta}^g) - N_{12}N_{22}(\tilde{\theta}^g)^{-1}N_{21}(\tilde{\theta}^g) \right] \right. \\ & \left. + B^{(1)} \int \left[\frac{\partial f_\theta(x)}{\partial \theta_1} \otimes I_r \right] D^{(2)}(g(x), f_\theta(x)) d\mu(x) \right\} IF(y, \tilde{T}_{\rho,1}, G) \\ & = B \left\{ \left[\xi_1(\tilde{\theta}^g) - M_1(y; \tilde{\theta}^g) \right] - N_{12}(\tilde{\theta}^g)N_{22}(\tilde{\theta}^g)^{-1} \left[\xi_2(\tilde{\theta}^g) - M_2(y; \tilde{\theta}^g) \right] \right\}, \end{aligned} \tag{3.3}$$

subject to the condition

$$B^T IF(y, \tilde{T}_{\rho,1}, G) = IF(y, \tilde{T}_{\rho,1}, G), \tag{3.4}$$

where \otimes denote the Kronecker product of two matrices and the $r \times r^2$ matrix $B^{(1)}$ is defined as $B^{(1)} = \frac{\partial^2 \phi(\beta)}{\partial \theta_1^2}$. The remaining second component $IF(y, \tilde{T}_{\rho,2}, G)$ of the influence function is then given by

$$IF(y, \tilde{T}_{\rho,2}, G) = N_{22}(\tilde{\theta}^g)^{-1} \left[\xi_2(\tilde{\theta}^g) - M_2(y; \tilde{\theta}^g) \right] - N_{22}(\tilde{\theta}^g)^{-1} N_{21}(\tilde{\theta}^g) IF(y, \tilde{T}_{\rho,1}, G). \quad (3.5)$$

In particular, if $N_{12}(\theta) = O$, then the estimators $\tilde{\theta}_1^g$ and $\tilde{\theta}_2^g$ becomes asymptotically independent and their influence functions also become independent of each other. Then, the influence function of $\tilde{\theta}_2^g$ simplifies to

$$IF(y, \tilde{T}_{\rho,2}, G) = N_{22}(\tilde{\theta}^g)^{-1} \left[\xi_2(\tilde{\theta}^g) - M_2(y; \tilde{\theta}^g) \right].$$

It is easy to see that, this is indeed of the same form as the corresponding influence function under the unrestricted case. And the influence function of $\tilde{\theta}_1^g$ in this case is given by the solution of

$$\left\{ BN_{11}(\tilde{\theta}^g) + B^{(1)} \int \left[\frac{\partial f_{\theta}(x)}{\partial \theta_1} \otimes I_r \right] D^{(2)}(g(x), f_{\theta}(x)) d\mu(x) \right\} IF(y, \tilde{T}_{\rho,1}, G) = B \left\{ \left[\xi_1(\tilde{\theta}^g) - M_1(y; \tilde{\theta}^g) \right] \right\},$$

subject to the restriction (3.4). \square

Example 3.4. Let us consider a general form of our motivating example on testing the equality of two sample means. The multivariate generalization of this problem involves the testing for homogeneity of mean among the p normal populations with unknown equal variances. Let us represent this problem by the set-up of multivariate normal model with mean $\mu = (\mu_1, \dots, \mu_p)^T$ and variance $\sigma^2 I_p$ where we are interested in testing the restrictions $\mu_1 = \dots = \mu_p$. Equivalently this restriction can be represented in terms of one unknown parameter, say β , (instead of p -components of μ) and is given by $\mu = \beta(1, \dots, 1)^T$.

Now let us derive the influence function for this motivating example, but with a slightly general restriction

$$\mu = \beta \mu_0$$

for some known p -vector μ_0 (instead of the the vector of ones). Interestingly, this case is a particular situation of the Example 3.3 with $\phi_i(\beta) = \beta \mu_0^i$ for all $i = 1, \dots, p$, where the superscript denote the corresponding component of μ_0 . Hence, we can use the result of Example 3.3 to derive the influence function of the RMDE $(\widetilde{\mu}^g, \widetilde{\sigma}^{2g})$ in this case

Note that, in terms of the notations of Example 3.3, we have $b_{ij} = \text{constant}$ for all i, j and so $B^{(1)} = O$, the null matrix of appropriate order. Further, considering $\theta_1 = \mu$ and $\theta_2 = \sigma^2$, we have $N_{12}(\theta) = 0_p$ for the present case of normal models. Thus, from the last part of Example 3.3, the influence function of $\widetilde{\sigma}^{2g}$ is given by

$$IF(y, \widetilde{\sigma}^{2g}, G) = N_{22}(\tilde{\theta}^g)^{-1} \left[\xi_2(\tilde{\theta}^g) - M_2(y; \tilde{\theta}^g) \right].$$

and that of $\widetilde{\mu}^g$ is given by the solution of

$$BN_{11}(\widetilde{\theta}^g)IF(y, \widetilde{\mu}^g, G) = B \left[\xi_1(\widetilde{\theta}^g) - M_1(y; \widetilde{\theta}^g) \right], \tag{3.6}$$

subject to the restriction

$$B^T IF(y, \widetilde{\mu}^g, G) = IF(y, \widetilde{\mu}^g, G).$$

Thus we will get a non-zero influence function of $\widetilde{\mu}^g$ if the matrix B^T has one of its eigenvalue as 1 and in that case the required influence function is given by the particular eigenvector of B^T corresponding to the eigenvalue 1 which satisfies (3.6). After another simplification, this influence function must be of the form

$$IF(y, \widetilde{\mu}^g, G) = B^T N_{11}(\widetilde{\theta}^g)^{-1} \left[\xi_1(\widetilde{\theta}^g) - M_1(y; \widetilde{\theta}^g) + v \right],$$

where v is a vector in the null-space of the matrix B .

For the special choice $\mu_0 = (1, \dots, 1)^T$ in case of testing homogeneity of means, we have $b_{ij} = 1$ for all i, j so that the matrix B does not have eigenvalue 1 and hence

$$IF(y, \widetilde{\mu}^g, G) = 0. \tag{□}$$

4. Applications: Some particular divergences

Based on the general results obtained in the previous sections, one can describe the influence function analysis and the asymptotic distributions of any MDE or RMDE provided one can prove only their \sqrt{n} -consistency. In this section, we will apply those results for some common divergence measures. Throughout this section, we will assume some common notations from the likelihood theory; $L(\theta; \Theta) = \ln f_\theta(x)$ for all $\theta \in \Theta$ is the likelihood function, $u_\theta(x) = \nabla L(\theta; \Theta)$ is the the likelihood score function, $I(\theta) = E_{f_\theta}[u_\theta(X)u_\theta(X)^T]$ is the Fisher information matrix. Also, we will define similar quantities under a proper subspace $\Theta_0 \subset \Theta$ (defined by the restrictions $h(\theta) = 0$) as $L(\theta; \Theta_0)$ being the restriction of $L(\theta; \Theta)$ onto the subspace Θ_0 , $u_\theta^0(x) = \nabla L(\theta; \Theta_0)$ and $I^0(\theta) = E_{f_\theta}[u_\theta^0(X)u_\theta^0(X)^T]$.

4.1. Disparity measures

One of the most popular family of divergences is the disparity family [17] that yields fully efficient and robust estimators upon minimization. It is defined in terms of a non-negative thrice differentiable strictly convex function ϕ on $[-1, \infty)$ with $\phi(0) = 0$ and $\phi'(0) = 0$, called the disparity generating function, as

$$\rho(g, f) = \int \phi(\delta)f,$$

with $\delta = g/f - 1$. It is of the form of the general divergences defined in Equation (2.1) with the function $D(\cdot, \cdot)$ given by (2.2) so that we can apply all the results

derived in previous sections. Using the same notations, we have,

$$M(y; \theta) = -A'(\delta)u_\theta(y),$$

and

$$N(\theta) = \int [A'(\delta)u_\theta u_\theta^T g - A(\delta)\nabla_2 f_\theta],$$

where the function $A(\delta)$, defined as

$$A(\delta) = C'(\delta)(\delta + 1) - C(\delta),$$

is known as the Residual Adjustment Function in the context of minimum disparity estimation and plays a crucial role in its robustness [17]. Then we get the following result from Equation (2.5).

Result 4.1. The influence function of the minimum disparity estimators is given by

$$IF(y, T_\rho, G) = N(\theta^g)^{-1}[A'(\delta)u_\theta(y) - E_g[A'(\delta)u_\theta(y)]],$$

which is the same as obtained by Lindsay [17] independently. In particular, at the model $g = f_{\theta_0}$, this influence function simplifies to $I(\theta_0)^{-1}u_{\theta_0}(y)$ which is independent of the disparity generating function $\phi(\cdot, \cdot)$ and so is same as that of the MLE. This is unbounded function for most of the common model families.

Next, let us consider the restricted minimum disparity estimation under the restriction $h(\theta) = 0$. Using the notations of Section 3, it is easy to see that

$$M_0(y; \theta) = -A'(\delta)u_\theta^0(y),$$

and

$$N_0(\theta) = \int [A'(\delta)u_\theta^0(u_\theta^0)^T g - A(\delta)\nabla_2 [f_\theta]|_{\Theta_0}].$$

Then, we can derive the influence function of the restricted minimum disparity estimators from Theorem 3.1 and above simplified expressions. However, the interesting case is when the true density belongs to the model family, i.e., $g = f_{\theta_0}$. In that case, we will have $M_0(y; \theta_0) = -u_\theta^0(y)$, $\xi_0(\theta_0) = 0$ and $N_0(\theta_0) = I^0(\theta_0)$. Thus, we have the following new result in the context of the minimum disparity estimation.

Result 4.2. The influence function of the restricted minimum disparity estimator \tilde{T}_ϕ corresponding to the disparity generated by $\phi(\cdot, \cdot)$ at the model is given by

$$IF(y, \tilde{T}_\phi, F_{\theta_0}) = [I^0(\theta_0)^2 + H(\theta_0)H(\theta_0)^T]^{-1} I^0(\theta_0)u_\theta^0(y). \quad (4.1)$$

Note that the above expression is independent of the choice of the disparity generating function and hence it also gives the influence function of the Restricted Maximum Likelihood Estimator.

Further, the above result also helps us to derive asymptotic distribution of the restricted minimum disparity estimators $\tilde{\theta}_\phi$ including that of the restricted maximum likelihood estimators. Following the argument of Lindsay [17], one can easily prove the \sqrt{n} -consistency of $\tilde{\theta}_\phi$. Then Remark 3.1 gives its asymptotic distribution as follows.

Remark 4.1. Under standard regularity conditions, whenever the restricted minimum disparity estimator $\tilde{\theta}_\phi$ is consistent, the asymptotic distribution of $\sqrt{n}(\tilde{\theta}_\phi - \theta_0)$, at the model $g = f_{\theta_0}$, is normal with mean zero and variance given by

$$[I^0(\theta_0)^2 + H(\theta_0)H(\theta_0)^T]^{-1} [I^0(\theta_0)]^3 [I^0(\theta_0)^2 + H(\theta_0)H(\theta_0)^T]^{-1}.$$

This result coincides with asymptotic distribution of the restricted maximum likelihood estimators obtained independently from the likelihood theory. Hence it provides a justification of our general results obtained in this paper.

4.2. Density power divergences

In the recent decades, arguably the most popular divergence measure in the context of the robust minimum divergence estimation is the Density Power Divergence [2]. The increasing popularity of this divergence is mainly due to the fact that corresponding minimum divergence estimation does not require any kernel smoothing in case of the continuous models; which is a major drawback of the disparity measures. The density power divergence is defined in terms of a non-negative tuning parameter α as

$$\rho_\alpha(g, f) = \int f^{1+\alpha} - \frac{1+\alpha}{\alpha} \int f^\alpha g + \frac{1}{\alpha} \int g^{1+\alpha} \quad \text{for } \alpha > 0,$$

and

$$\rho_0(g, f) = \lim_{\alpha \rightarrow 0} \rho_\alpha(g, f) = \int g \log(g/f).$$

Note that the case of $\alpha = 0$ gives the likelihood disparity and so the influence function of the corresponding minimum divergence estimator is already discussed in previous subsection. Let us now consider the case $\alpha > 0$.

Interestingly, for any given fixed $\alpha > 0$, this divergence also belongs to the general family of divergence defined in Equation (2.1) with the function $D(\cdot, \cdot)$ given by (2.3). Now, we can apply all the results derived above for the density power divergences, where one can show that

$$M(y; \theta) = -(1 + \alpha)u_\theta(y)f_\theta^\alpha(y),$$

and

$$\begin{aligned} N(\theta) &= (1 + \alpha) \int [u_\theta u_\theta^T f_\theta^{1+\alpha} + (i_\theta - \alpha u_\theta u_\theta^T)(g - f_\theta)f_\theta^\alpha] \\ &= (1 + \alpha)J_\alpha(\theta), \quad \text{say,} \end{aligned}$$

with $i_\theta = -\nabla u_\theta$. Then we get the corresponding influence function from (2.5).

Result 4.3. The influence function of the minimum density power divergence estimator T_α is given by

$$IF(y, T_\alpha, G) = J_\alpha(\theta^g)^{-1} [u_\theta(y) f_\theta^\alpha(y) - E_g[u_\theta(X) f_\theta^\alpha(X)]]. \quad (4.2)$$

In particular, if we assume $g = f_{\theta_0}$, then the influence function simplifies to

$$IF(y, T_\alpha, F_{\theta_0}) = \left(\int u_{\theta_0} u_{\theta_0}^T f_{\theta_0}^{1+\alpha} \right)^{-1} \left[u_{\theta_0}(y) f_{\theta_0}^\alpha(y) - \int u_{\theta_0} f_{\theta_0}^{1+\alpha} \right]. \quad (4.3)$$

The above influence function is bounded for all $\alpha > 0$ and most of the common model families. The expressions of the influence functions of the minimum density power divergence estimators obtained from our general theory is in fact exactly the same as those derived in Basu et al. [2] independently. This again proves the power and correctness of the general theory developed in this paper.

Next, we will consider the restricted minimum density power divergence estimation under the restrictions $h(\theta) = 0$. Again, we use the notations of Section 3 and get

$$M_0(y; \theta) = -(1 + \alpha) u_\theta^0(y) f_\theta^\alpha(y),$$

and

$$\begin{aligned} N_0(\theta) &= (1 + \alpha) \int [u_\theta^0(u_\theta^0)^T f_\theta^{1+\alpha} + (i_\theta^0 - \alpha u_\theta^0(u_\theta^0)^T)(g - f_\theta) f_\theta^\alpha] \\ &= (1 + \alpha) J_\alpha^0(\theta), \end{aligned}$$

with $i_\theta^0 = -\nabla u_\theta^0$. Then, Theorem 3.1 gives us the expression of the influence function of the restricted density power divergence estimators. The particular case of $g = f_{\theta_0}$ is presented in the following result.

Result 4.4. The influence function of the restricted minimum disparity estimator \tilde{T}_α at the model $g = f_{\theta_0}$ has the simple form

$$IF(y, \tilde{T}_\alpha, F_{\theta_0}) = \Psi(\theta_0)^{-1} \left(\int u_{\theta_0}^0 (u_{\theta_0}^0)^T f_{\theta_0}^{1+\alpha} \right) u_{\theta_0}^0(y) f_{\theta_0}^\alpha(y), \quad (4.4)$$

where

$$\Psi(\theta) = \left[\left(\int u_\theta^0 (u_\theta^0)^T f_\theta^{1+\alpha} \right)^2 + \frac{1}{(1 + \alpha)^2} H(\theta) H(\theta)^T \right].$$

Once again this influence function is generally bounded for all $\alpha > 0$ for most parametric models.

Finally, we can derive the asymptotic distribution of the restricted minimum density power divergence estimators $\tilde{\theta}_\alpha$ from Remark 3.1. The \sqrt{n} -consistency of the restricted minimum density power divergence estimators follows from a modification of the argument of Basu et al. [2] used to prove the same for minimum density power divergence estimators. So, if we have $g = f_{\theta_0}$, then

asymptotic distribution of $\sqrt{n}(\tilde{\theta}_\alpha - \theta_0)$ is normal with mean zero and asymptotic variance

$$\Psi(\theta_0)^{-1} \left(\int u_{\theta_0}^0 (u_{\theta_0}^0)^T f_{\theta_0}^{1+\alpha} \right) \text{Var}_{f_{\theta_0}} [u_{\theta_0}^0(Y) f_{\theta_0}^\alpha(Y)] \times \left(\int u_{\theta_0}^0 (u_{\theta_0}^0)^T f_{\theta_0}^{1+\alpha} \right) \Psi(\theta_0)^{-1}.$$

4.3. S-divergence family

We will end this section by considering a recent family of divergences, namely the *S*-Divergence family, developed by Ghosh et al. [13]. This is a general super-family containing both the density power divergence [2] and the Cressie-Read family of power divergences [5] and along with many other useful divergences. It is defined in terms of two parameters $\lambda \in \mathbb{R}$ and $\alpha \geq 0$ as

$$\rho(g, f) = \frac{1}{A} \int f^{1+\alpha} - \frac{1+\alpha}{AB} \int f^B g^A + \frac{1}{B} \int g^{1+\alpha}, \quad \alpha \in [0, 1], \quad \lambda \in \mathbb{R},$$

where $A = 1 + \lambda(1 - \alpha)$ and $B = \alpha - \lambda(1 - \alpha)$. For either $A = 0$ or $B = 0$, it is defined by the corresponding continuous limit of the divergences; see Ghosh et al. [13] for details. Further applications of this new divergence family in robust statistical inferences can be found in Ghosh [9], Ghosh and Basu [10, 11] and Ghosh et al. [12].

Again, this large family of divergence can be written in the form of equation (2.1) with

$$D(a, b) = \frac{1}{A} b^{1+\alpha} - \left(\frac{1+\alpha}{\alpha} \right) b^B a^A + \frac{1}{B} a^{1+\alpha}.$$

Then, we have

$$M_0(y; \theta) = -(1 + \alpha) u_\theta(y) f_\theta^B(y) g^A(y),$$

and

$$\begin{aligned} N(\theta) &= \frac{(1 + \alpha)}{A} \int [A u_\theta u_\theta^T f_\theta^{1+\alpha} + (i_\theta^0 - B u_\theta u_\theta^T)(g^A - f_\theta^A) f_\theta^B] \\ &= (1 + \alpha) J_{(\alpha, \lambda)}. \end{aligned}$$

So we get the influence function of the minimum *S*-divergence estimator from Equation (2.5).

Result 4.5. The influence function of the minimum *S*-divergence functional $T_{(\alpha, \lambda)}$ is given by

$$\begin{aligned} IF(y, T_{(\alpha, \lambda)}, G) &= J_{(\alpha, \lambda)}(\theta^g)^{-1} [u_\theta(y) f_\theta^B(y) g^A(y) - E_g[u_\theta(X) f_\theta^B(X) g^A(X)]], \end{aligned} \quad (4.5)$$

which is again the exactly same as obtained in Ghosh et al. [13]. For the special case $g = f_{\theta_0}$, this influence function coincides with that of the density power divergence given by equation (4.3).

Finally, the influence function of the restricted minimum S -divergence estimators under the restrictions $h(\theta) = 0$ can be derived from Theorem 3.1. It is then easy to see that, at the model $g = f_{\theta_0}$, the influence function of the restricted minimum S -divergence estimators coincides with that of the restricted density power divergence estimators derived in Equation (4.4).

5. Impact of restrictions on the robustness: Some qualitative illustrations

In the previous sections, we have developed the influence function of a general class of minimum divergence estimators under parameter restrictions and compared them with the general form for the corresponding unrestricted estimators. This work will help us to understand an important question about robustness under parameter restrictions – do we gain more robustness by imposing restrictions on the parameters? In this section, we present some qualitative discussion as an indicative answer to this question through several interesting examples along with the motivating problems mentioned in Section 1.

We have already noted that, for most parametric models, if the unrestricted minimum divergence estimator has bounded influence function (and hence is robust) then the corresponding restricted minimum divergence estimators under most usual restrictions will also be robust and have bounded influence functions. For example, as seen in Example 3.1, the minimum density power divergence estimator of the normal mean will always be robust for $\alpha > 0$ whether we consider the parameter space to be the whole real line or any restricted subspace of it. However, the converse of the above is not true in general; the maximum likelihood estimator of the normal mean is non-robust and have unbounded influence function for the unrestricted parameter space or even for the restricted parameter space if it is an unbounded subset of real line like the set of positive reals etc. But if we restrict the parameter space for the normal mean to a bounded interval of reals, then even the maximum likelihood estimator become robust with respect to the outliers outside that bounded interval. This fact, although intuitively clear, had no rigorous theoretical proof in the existing literature, which is now made available by our work in this paper.

Another interesting case of restriction is considered in Example 3.4 following one of the motivating problems – test for homogeneity of p normal populations with equal variances. In the motivation for this work, we have noted a similar problem with $p = 2$; in case of outliers in both the samples, solution of this testing problem with the non-robust maximum likelihood estimator becomes robust under the null hypothesis imposing the restriction of the equality of two means, which had no theoretical justification. Our work in the present paper justifies this situation by a rigorous proof as seen in Example 3.4. From the last paragraph of the example, we have obtained that the influence function of any restricted minimum divergence estimator (and hence that of the restricted maximum likelihood estimator also) for the common mean of the p normal samples with equal variance is identically zero under the restriction of equality

of the p means. This in turn implies the robustness of those restricted minimum divergence estimators for any divergence of the form (2.1), which includes the maximum likelihood estimator. However, without the restriction of equality of means, there are several minimum divergence estimators of the p means like their maximum likelihood estimators which get affected by the presence of outliers in the samples.

Thus we have seen that, by restricting on the parameter space we can sometimes make a non-robust minimum divergence estimator to be robust so that it provides us correct inference even in the presence of outliers in the data. However, if an unrestricted minimum divergence estimator is already robust with respect to outliers, then the imposition of any restrictions does not generally affect its robustness properties.

Next, we will present one more such implication that will illustrate the effect of restriction in one parameter (or, a set of parameters) of a model on the robustness of its other parameter (or, the set of remaining parameters); this will cover one of our motivating problems about the robustness of the estimators of normal mean when variance is known or vice versa.

Consider a parametric model family f_θ with two sets of parameters θ_1 and θ_2 ; here $\theta = (\theta_1, \theta_2)$. Examples of such model includes the normal family with unknown mean and variance, exponential family with location and shape parameter etc. Suppose we impose some restriction on the first parameter θ_1 ; a natural question is to check whether the robustness of the estimators of the second component θ_2 gets affected by such restrictions. We have already discussed the theoretical results about two such restrictions in Example 3.2 and 3.3, – one specifies the value of θ_1 completely under the restriction (case of simple hypothesis) and the second specifies θ_1 only partially (like the case of composite hypothesis). We have derived the influence function of the RMDE in both the cases following the general theory developed in this paper; the implication of the results can be summarized as follows: If the MDEs of θ_1 and θ_2 are asymptotically independent then, for both the cases, there will not be any effect of the restrictions imposed on θ_1 on the properties (including robustness) of the estimator of θ_2 ; we have seen that the influence function of the RMDE of θ_2 coincides with that of the corresponding MDE of θ_2 as expected from common intuition. See Example 5.1 below for a numerical illustration of this case. However, if the MDEs of θ_1 and θ_2 are not asymptotically independent then the above implication does not necessarily hold in general. But when θ_1 is fully specified (as in Example 3.2) then, even under asymptotic dependency, the influence function of the RMDE of θ_2 under the restriction on θ_1 becomes equivalent to that of the corresponding (unrestricted) MDE implying no effect of restriction on the robustness of the RMDE of θ_2 . Otherwise, in general, the robustness of the RMDE of θ_2 may depends on the (partially specified) restrictions on θ_1 and need to be examined for each particular case separately.

Example 5.1. Consider the problem of testing for the normal variance with known mean. In this case, we fix the mean of the assumed normal model to a given value under the null and estimate the variance parameter σ^2 . Suppose we

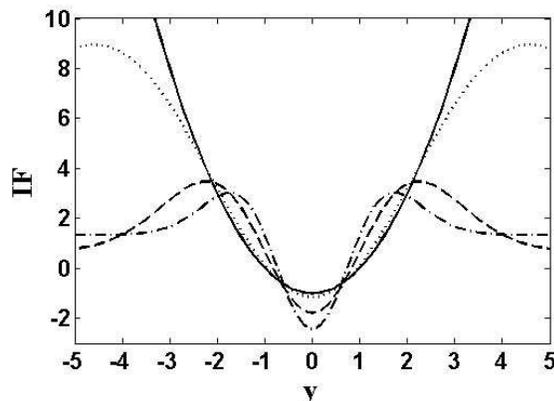


FIG 2. Influence function of the restricted minimum density power divergence estimators of normal variance under the restriction of fixed mean at the $N(0, 1)$, model for different α [solid line: $\alpha = 0$, dotted line: $\alpha = 0.1$, dashed line: $\alpha = 0.5$ and dot-dashed line: $\alpha = 1$].

use the density power divergence measure for this purpose. Then, the influence function of the restricted minimum density power divergence estimators $\widetilde{\sigma}^2$ of σ^2 at the model, under the restriction of given mean, turns out to be

$$IF(y, \widetilde{\sigma}^2, N(\mu_0, \sigma_0^2)) = \frac{2(1+\alpha)^{5/2}}{(\alpha^2+2)} \left\{ \frac{[(x-\mu_0)^2 - \sigma_0^2]}{\sigma_0^2} e^{-\frac{\alpha(x-\mu_0)^2}{2\sigma_0^2}} + \frac{\alpha}{(1+\alpha)^{3/2}} \right\} \sigma_0^2,$$

which coincides with that of the corresponding unrestricted estimators. These influence functions for some particular choices of the tuning parameter α are presented in Figure 2; clearly the RMDEs of σ^2 with $\alpha > 0$ are robust having bounded influence function, but that corresponding to $\alpha = 0$ has unbounded influence function and hence is non-robust with respect to the outliers. \square

These are only a few of the several possible important implications of the theory developed in the present paper, which could help us to understand the robustness properties of different minimum divergence estimators under restrictions in case of different composite hypotheses testing problems. We hope that these would give the reader a fair enough idea about the impact of parameter restriction on the robustness of the corresponding minimum divergence estimators and the usefulness of the general theory developed here in this context.

6. Conclusion

This work presents the derivation of the influence function of the restricted and unrestricted minimum divergence estimators for a general class of density based divergences. It will help researchers to derive the robustness properties of any

minimum divergence estimators under several restrictions on the parameters. For illustrations, we have examined the same for some popular minimum divergence estimators, namely the disparity, density power divergence and S -divergence family; we have also presented examples with a set of linearly dependent restrictions for general model family. Further, this paper gives us several directions for future works including the influence function of more general class of divergences that are possibly based on the distribution functions; author want to solve the related problems in subsequent researches.

Acknowledgment

This work is a part of the author's Ph.D. dissertation under the supervision of Prof. Ayanendranath Basu at the Indian Statistical Institute. The author wants to thank Prof. Basu for his sincere guidance and comments regarding this paper. The author also wishes to express his gratitude to the editor, an AE and two anonymous referees for their valuable comments that improved the content and presentation of this paper.

Appendix: Proof of the results

A.1. Proof of Theorem 2.1

From the definition of θ_ϵ , it must satisfy the estimating equation (2.4). So, substituting g_ϵ and θ_ϵ in place of g and θ respectively in (2.4) and differentiating with respect to ϵ at $\epsilon = 0$ we get,

$$\int \frac{\partial[\nabla D(g(x), f_{\theta^g}(x))]}{\partial g(x)} [-g(x) + \chi_y(x)] d\mu(x) + \int \frac{\partial[\nabla D(g(x), f_{\theta^g}(x))]}{\partial f_{\theta^g}(x)} [\nabla f_{\theta^g}(x)]^T IF(y, T_\rho, G) d\mu(x) = 0. \quad (\text{A.1})$$

But

$$\nabla D(g(x), f_\theta(x)) = D^{(2)}(g(x), f_\theta(x)) \nabla f_\theta(x);$$

hence

$$\frac{\partial[\nabla D(g(x), f_\theta(x))]}{\partial g(x)} = D^{(1,2)}(g(x), f_\theta(x)) \nabla f_\theta(x),$$

and

$$\frac{\partial[\nabla D(g(x), f_\theta(x))]}{\partial f_\theta(x)} = D^{(2)}(g(x), f_\theta(x)) \frac{\partial[\nabla f_\theta(x)]}{\partial f_\theta(x)} + D^{(2,2)}(g(x), f_\theta(x)) \nabla f_\theta(x).$$

Substituting these expressions in (A.1) and simplifying, we get

$$[-\xi(\theta^g) + M(y; \theta^g)] + N(\theta^g) IF(y, T_\rho, G) = 0.$$

Then the theorem follows immediately.

A.2. Proof of Theorem 3.1

Note that $\tilde{\theta}_\epsilon = \tilde{T}_\rho(G_\epsilon)$. By definition, the derivative of $\rho(g_\epsilon, f_\theta)$ with respect to θ will be zero at $\tilde{\theta}_\epsilon$, provided the derivative is taken after the substitution of the restrictions (3.1) in the expression of $\rho(g_\epsilon, f_\theta)$. Then, proceeding as in the proof of Theorem 2.1, we get

$$N_0(\tilde{\theta}^g)IF(y, \tilde{T}_\rho, G) - \xi_0(\tilde{\theta}^g) + M_0(y; \tilde{\theta}^g) = 0, \quad (\text{A.2})$$

where $N_0(\theta)$, $\xi_0(\theta)$, $M_0(y; \theta)$ are as defined in the statement of the theorem. Also, since $\tilde{\theta}_\epsilon$ must satisfy (3.1), a differentiation with respect to ϵ at $\epsilon = 0$ yields

$$H(\tilde{\theta}^g)^T IF(y, \tilde{T}_\rho, G) = 0, \quad (\text{A.3})$$

We need to solve the two equations (A.2) and (A.3) to get a general expression for the influence function $IF(y, \tilde{T}_\rho, G)$. Combining them, we get

$$\begin{pmatrix} N_0(\tilde{\theta}^g) \\ H(\tilde{\theta}^g)^T \end{pmatrix} IF(y, \tilde{T}_\rho, G) = \begin{pmatrix} \xi_0(\tilde{\theta}^g) - M_0(y; \tilde{\theta}^g) \\ 0_r \end{pmatrix}, \quad (\text{A.4})$$

where 0_r represents the zero-vector (column) of length r . After simplification, we get the required expression for the influence function $IF(y, \tilde{T}_\rho, G)$, which completes the proof.

A.3. Detailed calculation for Example 3.3

As noted in Example 3.3, we start with the set of equations (A.2) and (A.3) and solve them to derive the influence function of the RMDE of $\theta = (\theta_1, \theta_2)$. The form of the restrictions and the matrix $H(\theta)$ are already given in the example. Now, simple differentiation gives

$$\nabla f_{(\phi(\beta), \theta_2)^T}(x) = \begin{pmatrix} B \frac{\partial f_\theta(x)}{\partial \theta_1} \\ \frac{\partial f_\theta(x)}{\partial \theta_2} \end{pmatrix} = B^* \nabla f_\theta(x),$$

where B^* is a $p \times p$ matrix defined as

$$B^* = \begin{pmatrix} B & O \\ O & I_{p-r} \end{pmatrix},$$

and

$$\nabla^2 f_{(\phi(\beta), \theta_2)^T}(x) = \begin{bmatrix} B \frac{\partial^2 f_\theta(x)}{\partial \theta_1^2} B^T + B^{(1)} \left[\frac{\partial f_\theta(x)}{\partial \theta_1} \otimes I_r \right] & B \frac{\partial^2 f_\theta(x)}{\partial \theta_2 \partial \theta_1} \\ \frac{\partial f_\theta(x)}{\partial \theta_1 \partial \theta_2} B^T & \frac{\partial^2 f_\theta(x)}{\partial \theta_2^2} \end{bmatrix}.$$

Then we have

$$M_0(y; \theta) = B^* M(y; \theta), \quad \xi_0(\theta) = B^* \xi(\theta)$$

and

$$N_0(\theta) = B^* M(\theta) (B^*)^T + \begin{bmatrix} B^{(1)} \int \left[\frac{\partial f_\theta(x)}{\partial \theta_1} \otimes I_r \right] D^{(2)}(g(x), f_\theta(x)) d\mu(x) & O \\ O & O_{p-r} \end{bmatrix},$$

Now using the special form of the matrices for this case, Equations (A.2) and (A.3) simplify to

$$\begin{aligned} & [BN_{11}(\tilde{\theta}^g)B^T]IF(y, \tilde{T}_{\rho,1}, G) + [BN_{12}(\tilde{\theta}^g)]IF(y, \tilde{T}_{\rho,2}, G) \\ & + B^{(1)} \left\{ \int \left[\frac{\partial f_\theta(x)}{\partial \theta_1} \otimes I_r \right] D^{(2)}(g(x), f_\theta(x)) d\mu(x) \right\} IF(y, \tilde{T}_{\rho,1}, G) \\ & = B \left[\xi_1(\tilde{\theta}^g) - M_1(y; \tilde{\theta}^g) \right], \end{aligned} \quad (\text{A.5})$$

$$[N_{21}(\tilde{\theta}^g)B^T]IF(y, \tilde{T}_{\rho,1}, G) + N_{22}(\tilde{\theta}^g)IF(y, \tilde{T}_{\rho,2}, G) = \left[\xi_2(\tilde{\theta}^g) - M_2(y; \tilde{\theta}^g) \right], \quad (\text{A.6})$$

$$[I_r - B^T]IF(y, \tilde{T}_{\rho,1}, G) = 0. \quad (\text{A.7})$$

Now from Equation (A.6) and (A.6), we get Equation (3.5) and (3.4) respectively. Substituting (3.5) in (A.5), we get a simplified form of Equation (A.5) as given in (3.3). So, we need to solve Equation (3.3) for the first component $IF(y, \tilde{T}_{\rho,1}, G)$ of the partition of the influence function subject to (3.4) and then use Equation (3.5) to get the remaining second component $IF(y, \tilde{T}_{\rho,2}, G)$ of the influence function.

References

- [1] ALI, S. M. and SILVEY, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society B* 28, 131–142. [MR0196777](#)
- [2] BASU, A., HARRIS, I. R., HJORT, N. L., and JONES, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika* 85, 549–559. [MR1665873](#)
- [3] BASU, A., SHIOYA, H., and PARK, C. (2011). *Statistical Inference: The Minimum Distance Approach*. Chapman & Hall/CRC. [MR2830561](#)
- [4] BASU, A., MANDAL, A., MARTIN, N., and PARDO, L. (2013). Density Power Divergence Tests for Composite Null Hypotheses. *ArXiv pre-print*, arXiv:1403.0330 [[stat.ME](#)].
- [5] CRESSIE, N. and READ, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society B* 46, 440–464. [MR0790631](#)
- [6] CSISZÁR, I. (1963). Eine informations theoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publ. Math. Inst. Hungar. Acad. Sci.* 3, 85–107.

- [7] CSISZÁR, I. (1967a). Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica* 2, 299–318. [MR0219345](#)
- [8] CSISZÁR, I. (1967b). On topological properties of f -divergences. *Studia Scientiarum Mathematicarum Hungarica* 2, 329–339.
- [9] GHOSH, A. (2014). Asymptotic properties of minimum S -divergence estimator discrete models. *Sankhya A – The Indian Journal of Statistics*, doi:[10.1007/s13171-014-0063-2](#).
- [10] GHOSH, A. and BASU, A. (2014). The Minimum S -Divergence Estimator under Continuous Models: The Basu-Lindsay Approach. *ArXiv pre-print*, arXiv:[1408.1239](#) [[math.ST](#)].
- [11] GHOSH, A. and BASU, A. (2015). Testing Composite Null Hypothesis Based on S -Divergences. *ArXiv pre-print*, arXiv:[1504.04100](#) [[math.ST](#)].
- [12] GHOSH, A., BASU, A., and PARDO, L. (2015). On the robustness of a divergence based test of simple statistical hypotheses. *Journal of Statistical Planning and Inference* 161, 91–108. [MR3316553](#)
- [13] GHOSH, A., HARRIS, I. R., MAJI, A., BASU, A., and PARDO, L. (2013). A Generalized Divergence for Statistical Inference. *Technical Report, BIRU/2013/3*, Interdisciplinary Statistical Research Unit, Indian Statistical Institute, Kolkata, India.
- [14] HAMPEL, F. R. (1968). *Contributions to the theory of robust estimation*. Ph. D. thesis, University of California, Berkeley, USA. [MR2617979](#)
- [15] HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *Journal of American Statistical Association* 69, 383–393. [MR0362657](#)
- [16] HAMPEL, F. R., RONCHETTI, E., ROUSSEEUW, P. J., and STAHEL, W. (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York, USA: John Wiley & Sons. [MR0829458](#)
- [17] LINDSAY, B. G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Annals of Statistics* 22, 1081–1114. [MR1292557](#)
- [18] PARDO, L. (2006). *Statistical Inference Based on Divergences*. CRC/Chapman-Hall. [MR2183173](#)
- [19] PATRA, S., MAJI, A., BASU, A., and PARDO, L. (2013). The power divergence and the density power divergence families: The mathematical connection. *Sankhya B* 75, 16–28. [MR3082808](#)
- [20] SIMPSON, D. G. (1989). Hellinger deviance test: Efficiency, breakdown points, and examples. *Journal of the American Statistical Association* 84, 107–113. [MR0999667](#)
- [21] VAJDA, I. (1972). On the f -divergence and singularity of probability measures. *Periodica Math. Hungarica* 2, 223–234. [MR0335163](#)