Sankhyā : The Indian Journal of Statistics
2013, Volume 75-B, Part 1, pp. 16-28

The power divergence and the density power divergence families: the mathematical connection

Sujayendu Patra, Avijit Maji and
Ayanendranath Basu
Indian Statistical Institute, Kolkata, India
Leandro Pardo
Complutense Universidad, Madrid, Spain

Abstract

The power divergence family of Cressie and Read (1984) is a highly popular family of density-based divergences which is widely used in robust parametric estimation and multinomial goodness-of-fit testing. This family forms a subclass of the family of ϕ-divergences (Csiszár, 1963; Pardo, 2006) or disparities (Lindsay, 1994). The more recently described family of density power divergences (Basu et al., 1998) is also extremely useful in robust parametric estimation. This paper explores the mathematical connection between these two families and establishes some interesting links.

AMS (2000) *subject classification.* Primary 62F10, 62F35; Secondary 62-07.
Keywords and phrases. Density power divergence, power divergence.

1 Introduction

Parametric estimation based on minimum chi-square type methods has been studied by many authors. The general form for the chi-square type distances was proposed, independently, by Ali and Silvey (1966) and Csiszár (1963). This class of divergences has been referred to as ϕ-divergences by several authors, and represents a measure of discrepancy between two probability density functions. Pardo (2006) provides a comprehensive description of the class of ϕ-divergences in discrete models. Lindsay (1994) has referred to such divergences as disparities and explored the geometric structure which makes the corresponding estimators robust under data contamination.

Basu et al. (1998) proposed the family of density power divergences which provides another rich class of density-based divergences leading to robust parameter estimates. Both the power divergence and the density power divergence families have a similar philosophy in downweighting outlying observations with powers of model densities. The two families are essentially distinct with the Kullback-Leibler divergence being the only common element.

In this paper we compare and contrast the mathematical forms of the two classes of divergences and show that there is an interesting link which allows either family to be developed from the other through specific motivations. To set up the standard, we will define a divergence between two probability densities as a measure which is always nonnegative and is equal to zero if and only if the densities are identically equal.

2 The divergence families

2.1. The power divergence family. Consider probability density functions g and f with respect to a common dominating measure (say, Lebesgue measure). The power divergence measure corresponding to the tuning parameter λ is given by

$$\mathrm{PD}_\lambda(g,f) = \frac{1}{\lambda(1+\lambda)} \int \left\{ g \left[\left(\frac{g}{f}\right)^\lambda - 1 \right] \right\} dx, \quad \lambda \in \mathbb{R}. \tag{2.1}$$

We have dropped the dummy variable x from the arguments of the densities to make the presentation simple. This divergence measure can be easily shown to be nonnegative, although the same can not be said about the integrand itself. However, a reorganization of the terms of the above measure gives

$$\begin{aligned} &\int \left\{ \frac{g}{\lambda(1+\lambda)} \left[\left(\frac{g}{f}\right)^\lambda - 1 \right] + \frac{f-g}{1+\lambda} \right\} dx \\ &= \int \left\{ \frac{1}{\lambda(1+\lambda)} \left[\left(\frac{g}{f}\right)^{1+\lambda} - \left(\frac{g}{f}\right) \right] + \frac{1-g/f}{1+\lambda} \right\} f dx, \end{aligned} \tag{2.2}$$

and while the value of the integral does not change, the integrand on the right-hand side of the above equation is easily seen to be nonnegative. The divergences for the cases $\lambda = 0$ and $\lambda = -1$ cannot be directly obtained by replacing these values in equation (2.1), but can be obtained as the limits of the expressions in equation (2.1) as λ tends to 0 and -1 respectively. The distances in these cases are

$$\mathrm{PD}_0(g,f) = \int g \log\left(\frac{g}{f}\right) dx \;\text{ and }\; \mathrm{PD}_1(g,f) = \int f \log\left(\frac{f}{g}\right) dx,$$

and the corresponding versions with nonnegative integrands are given by

$$\mathrm{PD}_0(g,f) = \int \left\{ \frac{g}{f} \log\left(\frac{g}{f}\right) + (1 - g/f) \right\} f dx \quad \text{and} \quad \mathrm{PD}_1(g,f) = \int \left\{ \log\left(\frac{f}{g}\right) + (g/f - 1) \right\} f dx.$$

In minimum divergence estimation based on the power divergence family, the first argument of $\mathrm{PD}_\lambda(\cdot,\cdot)$ represents the data density, while the second component represents the density of the parametric model $\{F_\theta\}$. The minimizer of $\mathrm{PD}_\lambda(g, f_\theta)$ over $\theta \in \Theta$, the parameter space, is the minimum divergence functional at the density g. In particular, the minimizer of $\mathrm{PD}_0(g, f_\theta)$ over $\theta \in \Theta$ is the maximum likelihood functional at g. When the true distribution G is modeled by $\{F_\theta\}$ and the estimation of θ based on an independently and identically distributed sample is of interest, one minimizes $\mathrm{PD}_\lambda(g^*, f_\theta)$ over $\theta \in \Theta$ for a suitable value of the tuning parameter λ; here g^* is a nonparametric density estimate obtained from the data, and f_θ is the model density, The class of minimum divergence estimators within the power divergence family are distinguished by the fact that these estimators are generally first order efficient at the model under suitable regularity conditions (Lindsay, 1994; Pardo, 2006; Basu, Shioya and Park, 2011). Large negative values of λ are preferred from the robustness point of view.

A more general structure for density-based divergences of the χ^2 type between the densities g and f_θ is given by

$$\rho(g, f_\theta) = \int G(\delta(x)) f_\theta(x) dx \tag{2.3}$$

where $\delta = g(x)/f_\theta(x) - 1$ is the Pearson residual at the value x. The function G is strictly convex and satisfies $G(0) = 0$. It is easy to see that all the divergences defined through (2.2) have the form given in (2.3) for specific G functions.

One drawback of doing density-based minimum divergence estimation based on the power divergence family (or performing minimum disparity estimation in general) in the continuous model is that the use of an appropriate density estimation technique (such as kernel density estimation) is unavoidable, and as a result many additional complications creep into the estimation process. The only member within the power divergence family (or more generally within the entire class of disparities) for which such nonparametric smoothing techniques may be avoided is the Kullback-Leibler divergence PD_0, in which case the density g only appears linearly in the essential objective function that has to be optimized for the minimization of $\mathrm{PD}_0(g, f_\theta)$; the objective function can therefore be expressed in terms of the empirical distribution function, and no nonparametric smoothing is necessary.

2.2. The density power divergence family. Basu et al. (1998) defined the density power divergence family as a function of a nonnegative tuning parameter α; for given densities g and f this divergence, corresponding to the parameter α, is defined as

$$d_\alpha(g, f) = \int \left\{ f^{1+\alpha} - \left(1 + \frac{1}{\alpha}\right) g f^\alpha + \frac{1}{\alpha} g^{1+\alpha} \right\} dx. \tag{2.4}$$

For the model family of densities $\{f_\theta\}$, the minimum density power divergence functional at the density g corresponds to the minimizer of $d_\alpha(g, f_\theta)$ over $\theta \in \Theta$. Basu et al. (1998) developed this class of divergences so as to make the associated estimating equation correspond to a weighted version of the ordinary likelihood score equation, where the weighting is provided by a positive power of the model density. If an outlying observation makes a large contribution to the score equation, it is expected that a positive power of the model density function at that point will provide a downweighting effect to offset the influence of the outlier. Remarkably, none of the members of the family of density power divergences require any nonparametric density estimation to estimate the minimum divergence estimator associated with $d_\alpha(g, f_\theta)$. This is due to the fact that the minimization of the density power divergence $d_\alpha(g, f_\theta)$ over θ essentially considers the part

$$\int \left\{ f_\theta^{1+\alpha} - \left(1 + \frac{1}{\alpha}\right) g f_\theta^\alpha \right\}$$

where the density g only appears as a linear term. When one approximates the distribution G with the empirical distribution function G_n, one can substitute dG with dG_n avoiding the use of any nonparametric smoothing technique. Notice that the integrand in equation (2.4) is nonnegative for all $\alpha \geq 0$; the divergence $d_0(g, f)$, obtained by taking the limit of $d_\alpha(g, f)$ as $\alpha \to 0$ coincides with $\mathrm{PD}_0(g, f)$, the Kullback Leibler divergence. This divergence is the only common member between the power divergence and the density power divergence families.

Although Basu et al. (1998) considered only nonnegative tuning parameters α, one can conceptually extend the domain of the tuning parameter and define the density power divergence measures for negative values of α. Consider the scaled divergence

$$(1 + \alpha)^{-1} d_\alpha(g, f). \tag{2.5}$$

This divergence is clearly nonnegative for all values of $\alpha \in \mathbb{R}$. However, from the construction of the divergence it is clear that there is no robustness (or efficiency) benefit for a minimum density power divergence estimator corresponding to negative values of α – compared to maximum likelihood – and there is no particular reason for the statistician to choose such a divergence when constructing a minimum divergence estimator.

3 The connection between the families

Let us suppose that our interest is in developing a class of divergences, starting from the PD_λ measure, which can be used for minimum divergence estimation without taking recourse to nonparametric density estimation. A moments reflection shows that to achieve this one needs to eliminate such terms in the expression of $\mathrm{PD}_\lambda(g, f_\theta)$ which contain a product of a nonlinear function of g with some function of f_θ. A nonlinear term of g by itself (without a factor of f_θ as a multiplier) does not cause any difficulty, since it does not matter in the minimization process. By studying the divergence $\mathrm{PD}_\lambda(g, f_\theta)$ using the general form of PD_λ given on the right-hand side of equation (2.2), it is immediately seen that the term $(g/f_\theta)^{1+\lambda}$ is the one which needs to be adjusted. Since the expression within the parentheses in this equation is nonnegative, the nonnegativity of the divergence is not compromised if we replace the outer density f_θ by a suitable power of f_θ different from 1; the divergence still assumes its smallest value zero when $g = f_\theta$, identically. Thus if we replace the outer f_θ term with $f_\theta^{1+\lambda}$ in the expression of the divergence given in (2.2) and readjust the terms, we get the measure

$$
\begin{aligned}
&\int \left\{ \frac{1}{\lambda(1+\lambda)} \left[g^{1+\lambda} - g f_\theta^\lambda \right] + \frac{f_\theta^{1+\lambda} - g f_\theta^\lambda}{1+\lambda} \right\} dx \\
&= \frac{1}{1+\lambda} \int \left\{ \frac{1}{\lambda} \left[g^{1+\lambda} - g f_\theta^\lambda \right] + f_\theta^{1+\lambda} - g f_\theta^\lambda \right\} dx \\
&= \frac{1}{1+\lambda} \int \left\{ f_\theta^{1+\lambda} - \left(1 + \frac{1}{\lambda} \right) g f_\theta^\lambda + \frac{1}{\lambda} g^{1+\lambda} \right\} dx, \qquad (3.1)
\end{aligned}
$$

which is the same as the density power divergence measure $d_\lambda(g, f_\theta)$ measure divided by the constant term $(1+\lambda)$, or the scaled version of the divergence defined in equation (2.5).

The above operation, starting from a power divergence measure corresponding to a tuning parameter λ, produces a simple multiple of the density power divergence measure with the same value of the tuning parameter - the measure defined in equation (2.5). This is interesting, since this transformation produces divergences with very strong robustness properties starting from divergences which are very weak in that respect (or vice versa). The Pearson's chi-square, for example, corresponds to $\lambda = 1$ and produces minimum divergence estimators which are highly unstable under data contamination. Yet its corresponding member within the density power divergence family obtained through the above operation – the L_2 distance – produces estimators with solid robustness properties. The Hellinger distance, on the other hand, corresponds to $\lambda = -1/2$ and generates an estimator with excellent robustness credentials, but the corresponding member within the family of the extended density power divergence will be so weak from the robustness standpoint that such divergences were not even considered by Basu et al. (1998) within their original density power divergence class. Notice that while a positive power of the model density will downweight the effect of a large outlier, a negative power of the model density will only succeed in magnifying it, and there is little justification of choosing a divergence which treats the score function in this manner.

One could take the reverse route, and generate the power divergence family starting from the density power divergence. The density power divergence measure may be written as

$$d_\alpha(g, f_\theta) = \int \left[1 - \left(1 + \frac{1}{\alpha}\right)\frac{g}{f_\theta} + \frac{1}{\alpha}\left(\frac{g}{f_\theta}\right)^{1+\alpha}\right] f_\theta^{1+\alpha} dx. \tag{3.2}$$

The quantity within the square brackets is expressed only in terms of the ratio g/f, and is therefore related to the Pearson residual utilized by Lindsay (1994) to study the structure of disparities. Unlike disparities, however, the outer coefficient is $f_\theta^{1+\alpha}$, and not just f_θ, so the properties of the divergences cannot be compared by simply comparing the expressions with the bracketed term as functions of the Pearson residual. But since the bracketed term is nonnegative, one could generate a family of legitimate divergences by replacing the outer $f_\theta^{1+\alpha}$ with f_θ in (3.2). This leads to the divergence

$$\begin{aligned}
&\int \left[1 - \left(1 + \frac{1}{\alpha}\right)\frac{g}{f_\theta} + \frac{1}{\alpha}\left(\frac{g}{f_\theta}\right)^{1+\alpha}\right] f_\theta dx \\
&= \int \left\{1 - \frac{g}{f_\theta} + \frac{1}{\alpha}\left[\left(\frac{g}{f_\theta}\right)^{1+\alpha} - \frac{g}{f_\theta}\right]\right\} f_\theta dx \\
&= (1+\alpha)\int \left\{\frac{1}{\alpha(1+\alpha)}\left[\left(\frac{g}{f_\theta}\right)^{1+\alpha} - \frac{g}{f_\theta}\right] + \frac{1 - g/f_\theta}{1+\alpha}\right\} f_\theta dx,
\end{aligned} \tag{3.3}$$

and comparing the last expression with equation (2.2), this is just $(1+\alpha)$ times the density power divergence corresponding to the tuning parameter α. Starting from equation (2.5), therefore, one generates the power divergence family exactly.

Once again notice that a density power divergence corresponding to a positive value of α corresponds to the member of the power divergence family with the same value of α. Thus a density power divergence which is associated with a highly robust estimator gets transformed to a member of the power divergence family which leads to estimators with very weak robustness properties. Notice also that the Kullback-Leibler divergence $PD_0(\cdot,\cdot) = d_0(\cdot,\cdot)$ is the only one which does not require the adjustment of either equation (3.1) or equation (3.3), and automatically satisfies both requirements.

4 The uniqueness of the density power divergence

In this section we consider a general class of divergences and show that within this framework the density power divergence family is only one for which nonparametric smoothing techniques may be avoided for the purpose of estimation. Let us consider the class of divergence measures of the form

$$\rho(g, f_\theta) = \int h(\delta + 1) f_\theta^\beta dx \tag{4.1}$$

where $\beta > 0$, $\delta = g/f_\theta - 1$ is the Pearson's residual and $h(\gamma) = \sum_{t\in T} a_t \gamma^t$ for some finite set T with elements in $\mathbb{R}$ and real coefficients $\{a_t\}$ such that $h(\gamma)$ is nonnegative on $[0,\infty)$ and $h(\gamma) = 0$ only when $\gamma = 1$. Thus

$$\rho(g, f_\theta) = \sum_{t\in T} a_t \int \gamma^t f_\theta^\beta dx$$

where $\gamma = \delta + 1 = g/f_\theta$. For smoothing techniques to be avoided, the density g must appear linearly in those terms of the objective functions where both f_θ and g are present. This implies that $a_t = 0$ for $t \neq 0, 1, \beta$, and we can rewrite the divergence as

$$\rho(g, f_\theta) = c \int \left(1 + c_1\gamma + c_2\gamma^\beta\right) f_\theta^\beta dx$$

for real constants c, c_1, and c_2, with $c > 0$. The value of the constant c does not affect the divergence properties of the measure, and we can take $c = 1$ without loss of generality. In order that the distance is nonnegative and equals zero if and only if $g = f_\theta$, we must have $c_1 + c_2 = -1$ and $1 - (1 + c_2)\gamma + c_2\gamma^\beta \geq 0$ for $\gamma \in [0, \infty)$. Thus

$$c_2 \geq \frac{\gamma - 1}{\gamma^\beta - \gamma} \quad \text{if} \ \ \gamma > 1 \tag{4.2}$$

$$c_2 \leq \frac{1 - \gamma}{\gamma - \gamma^\beta} \quad \text{if} \ \ \gamma < 1, \tag{4.3}$$

and taking $\gamma \downarrow 1$ and $\gamma \uparrow 1$ respectively, we obtain that $c_2 \geq 1/(\beta - 1)$ and $c_2 \leq 1/(\beta - 1)$. So the only admissible value of c_2 is $1/(\beta - 1)$, and writing $\beta = 1 + \alpha$ we get

$$\rho(g, f_\theta) = \int \left\{ f_\theta^{1+\alpha} - \left(1 + \frac{1}{\alpha}\right) g f_\theta^\alpha + \frac{1}{\alpha} g^{1+\alpha} \right\} dx,$$

which is the density power divergence family for $\alpha \geq 0$. When one chooses $\beta \in (0, 1)$, the inequalities in (4.2) and (4.3) are reversed; however taking the limits $\gamma \downarrow 1$ and $\gamma \uparrow 1$, c_2 is again trapped between the same quantities. with $1/(\beta - 1)$ being the only allowable value. Setting $\beta = 1 + \alpha$ shows that the class of divergences in (4.1) represent the density power divergence family for $\alpha > -1$.

The construction can be extended to the scaled divergence (2.5) for all values of $\alpha \in \mathbb{R}$. Considering the scaled version of the measure (4.1) given by

$$\frac{1}{\beta} \int h(g/f_\theta) f_\theta^\beta dx, \beta \in \mathbb{R} \tag{4.4}$$

we see that the modified requirements (i) $h(\gamma)/\beta$ is nonnegetive, (ii) $h(\gamma)/\beta = 0$ if $\gamma = 1$, and (iii) smoothing techniques are to be avoided, lead to the conditions $c_1 + c_2 = -1$ and $1 - (1 + c_2)\gamma + c_2\gamma^\beta \leq 0$ for $\beta \leq 0$. The equivalence of (4.4) and (2.5) is then easily established by arguments similar to the previous ones. Clearly, the required conditions imply that the scaled family of measures in (4.4) is equivalent to the scaled density power divergence family in (2.5) for all $\alpha \in \mathbb{R}$.

Some more discussion on the robustness properties of these families of divergences will be relevant here. We have already noted that in spite of the observed correspondence between the two families in question, the robustness of either family may not transpire to the other along the corresponding divergences, and this aspect needs a little bit of further investigation. In this respect, it may be useful to consider the tuning parameters α and λ as nuisance parameters so that the robustness can be studied in a more general set up. However, this has to be undertaken with care, since the density power divergence measures (or the power divergence measures) vary over different ranges for different values of the tuning parameters α and λ. Empirical evidence shows that the determination of the parameter θ when α (or λ) is treated as the nuisance parameter generally leads to a strongly robust estimator. We consider this aspect of these divergences in the numerical example presented in Section 5.

In this connection it is also pertinent to mention that several authors have attempted to determine optimal choices, encompassing the robustness and efficiency concepts, of the tuning parameter α within the density power divergence family with the aim of finding globally satisfactory results. The most representative result in this context appears to be the one due to Warwick and Jones (2005), which determines the value α by empirically estimating the mean squared error of the minimum density power divergence estimator as a function of α. One then chooses the value of α for which this mean squared error is minimum over an appropriate range of α. We work out this procedure in the example presented in the next section to determine the optimal value of α according to the above criterion.

5 Data example

We now consider a real data example to illustrate the difference in the robustness pattern of the associated members of the power divergence and density power divergence families. These data represent a chemical mutagenicity experiment, and were analyzed previously by Simpson (1987); the details of the experimental protocol are available in Woodruff et al. (1984). In a sex linked study involving Drosophila (fruit flies), the experimenter exposed groups of male flies to different doses of a chemical to be screened. Subsequently each male was mated with unexposed females; the experimenter sampled, roughly, 100 daughter flies from each male and noted the number of daughters carrying a recessive lethal mutation on the X chromosome. For one such experimental run, the observed frequencies of males having 0, 1, 2 $\ldots$, recessive lethal daughters is presented in Table 1. The observed frequencies are presented in the top of the table, while the expected frequencies under the Poisson model for several different estimation methods are presented in the body of the table. The parameter estimates are given in the last column of the table. We have considered the power divergence measures corresponding to $\lambda = 0.5$ and $= -0.5$, with the associated density power divergence measures corresponding to $\alpha = 0.5$ and $\alpha = -0.5$. For comparison we have also included the maximum likelihood fit to the Poisson model for

Table 1: Fits of the Poisson model to the Drosophila data using several estimation methods.

	Recessive lethal count						
	0	1	2	3	4	≥ 5	$\hat{\theta}$
Observed	23	3	0	1	1	0	
$PD_{\lambda=0.5}$	15.58	9.13	2.68	0.52	0.08	0.01	0.5862
$DPD_{\alpha=0.5}$	24.11	3.60	0.27	0.01	–	–	0.1493
$PD_{\lambda=-0.5}$	24.70	3.09	0.19	0.01	–	–	0.1252
$DPD_{\alpha=-0.5}$	13.06	9.96	3.80	0.96	0.18	0.03	0.7624
$PD_{\lambda=-0.8181}$	25.72	2.19	0.09	–	–	–	0.0851
$DPD_{\alpha=1}$	23.78	3.88	0.32	0.02	–	–	0.1633
ML	19.59	7.00	1.25	0.15	0.01	–	0.3571
ML+D	24.95	2.88	0.17	0.01	–	–	0.1154

the full data (in the ML row), as well as the maximum likelihood fit for the cleaned data after deleting the two moderate outliers (in the ML+D row).

Notice that the difference between the maximum likelihood estimator and the outlier deleted maximum likelihood estimator is substantial. The maximum likelihood estimator is clearly inadequate in this situation – a consequence of the presence of the two mild outliers. The two robust estimators ($PD_{\lambda=-0.5}$ and $DPD_{\alpha=0.5}$) successfully withstand the effect of these outliers and produce expected frequencies which clearly provide a satisfactory description of the observed data. The other two estimators ($PD_{\lambda=0.5}$ and $DPD_{\alpha=-0.5}$) are in fact worse than the maximum likelihood estimator in terms of outlier stability, as one would expect.

The lack of robustness of the density power divergence family for negative values of α can also be clearly observed in the influence functions of the corresponding estimators. Figure 1 provides the influence function of several of the estimators within the density power divergence family. The functions are bounded for $\alpha > 0$, increase linearly for $\alpha = 0$ (likelihood), and increase even more sharply in either direction for $\alpha < 0$.

In order to study the overall robustness aspect of the power divergence and density power divergence methods, we have also performed an overall minimization of each of the divergences by considering their tuning parameters to be nuisance parameters varying within a reasonable range. For this purpose we have minimized the power divergence and the density power divergence families jointly with respect to the parameter of interest (over the

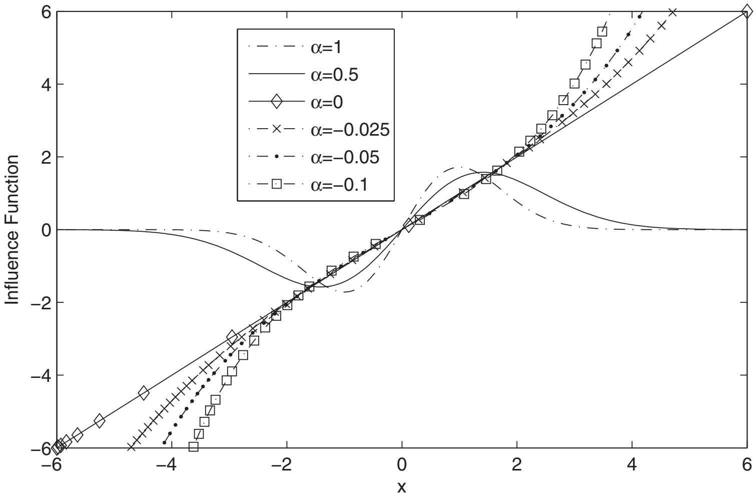

Figure 1: Influence function of different members of the density power divergence family.

parameter space Θ) and the tuning parameter (over an appropriate range of interest). In this example we have allowed α to vary over $[0, 1]$ for the density power divergence family, and have chosen $\lambda \in [-1, 1]$ for the power divergence family. The overall minimization of the power divergence family occurs at $\lambda = -0.8181$ (with corresponding parameter estimate 0.0851), and

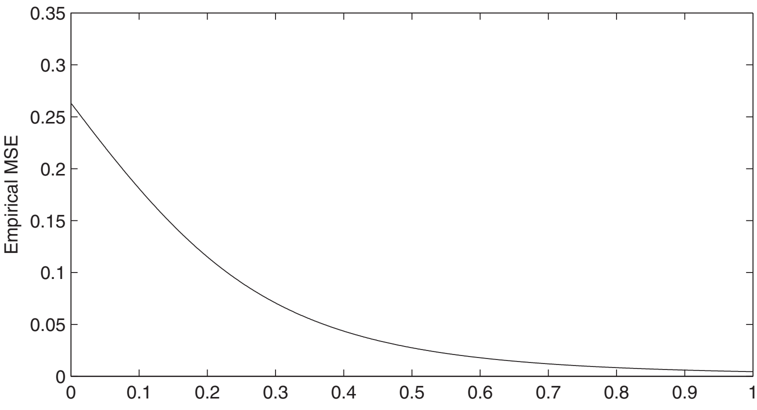

Figure 2: Estimated empirical MSEs for different values of α within the DPD family for the Poisson example.

that for the density power divergence occurs at the boundary $\alpha = 1$ (with a corresponding parameter estimate 0.1633). In either case, the overall minimum corresponds to one of the more robust members of the family.

The application of the Warwick and Jones (2005) method, interestingly, gives us the same value of the tuning parameter $\alpha = 1$. The variation of the empirical mean square error of the estimator is presented in Figure 2 as a function of α, which clearly shows that the empirical mean squared error continuously decreases in the interval $\alpha \in [0, 1]$.

6 Concluding remarks

In this paper we have established some interesting links between two prominent subclasses of the class of density based divergences, each of which has its own advantages in parametric inference. Interestingly, this connection links a robust member of either family with a member of the other family which has weak robustness properties. We expect that further investigations of the structure of these divergences may be useful in developing estimators and statistical procedures which could fare better in terms of combining the desirable properties of both the divergence families.

The following observation is also of interest. The power divergence family is, from a statistical point of view, the most important family of divergences within the class of ϕ-divergences or disparities. Similarly the density power divergence is the most important family within the class of Bregman divergences. The above results show that the most prominent members of the power divergence family and the density power divergence family share an interesting link. This result, however, appears to be specific to these two families, and at least for the moment it is not clear whether such properties are generally shared between the class of ϕ-divergences and the family of Bregman divergences.

The power divergence and the density power divergence are both established tools that are being extensively used in statistics and information theory. As such, researchers have been interested in possible links between them for some time; see for example, Harris and Basu (1997) and Vonta and Karagrigoriou (2010). Our approach here differs from the others in that we are primarily concerned in the modification of the power divergence family which allows us to construct useful minimum divergence procedures that avoid the use of nonparametric smoothing techniques in parametric estimation and hypothesis testing. This endeavor leads directly to the density power divergence family.

References

Ali, S.M. and Silvey, S.D. (1966). A general class of coefficients of divergence of one distribution from another. *J. R. Stat. Soc. Ser. B*, **28**, 131–142.

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.

Basu, A., Shioya, H. and Park, C. (2011). *Statistical inference: the minimum distance approach.* CRC Press, Boca Raton, Florida.

Cressie, N. and Read, T.R.C. (1984). Multinomial goodness-of-fit tests. *J. R. Stat. Soc. Ser. B*, **46**, 440–464.

Csiszár, I. (1963). Eine informations theoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizitat von Markoffschen Ketten. *Publ. Math. Inst. Hungar. Acad. Sci.*, **3**, 85–07.

Harris, I.R. and Basu, A. (1997). A Generalized Divergence Measure. Technical Report. Applied Statistics Unit, Indian Statistical Institute, Calcutta 700 035, India.

Lindsay, B.G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Ann. Statist.*, **22**, 1081–1114.

Pardo, L. (2006). *Statistical inference based on divergences.* CRC/Chapman & Hall, Boca Raton, Florida.

Simpson, D.G. (1987). Minimum Hellinger distance estimation for the analysis of count data. *J. Amer. Statist. Assoc.*, **82**, 802–807.

Vonta, F. and Karagrigoriou, A. (2010). Generalized measures of divergences in survival analysis and reliability. *J. Appl. Probab.*, **47**, 216–234.

Warwick, J. and Jones, M.C. (2005). Choosing a robustness tuning parameter. *J. Stat. Comput. Simul.*, **75**, 581–588.

Woodruff, R.C., Mason, J.M., Valencia, R. and Zimmering, A. (1984). Chemical mutagenesis testing in drosophila I: Comparison of positive and negative control data for sex-linked recessive lethal mutations and reciprocal translocations in three laboratories. *Environ. Mutagen.*, **6**, 189–202.

Sujayendu Patra, Avijit Maji and
Ayanendranath Basu
Indian Statistical Institute
Kolkata 700 108, India

Leandro Pardo
Complutense Universidad
Madrid, Spain

Paper received: 20 December 2011; revised: 29 February 2012.