Focused model selection and inference using robust estimators

Sam-Erik Walker, Nils Lid Hjort
Dept. of Mathematics, University of Oslo, Norway

BACKGROUND AND MOTIVATION
Selection of parametric models based on general information criteria like Akaike's information criterion (AIC), Bayesian information criterion (BIC), and similar, is a well-established practice within the statistical science. Over the years, more specific model selection criteria have also been developed, like the so-called focused information criterion, or FIC, where models are selected based on specifying a focus parameter - a certain parameter or function of parameters deemed most important in a given statistical setting. Up to now FIC has been mostly based on maximum likelihood estimator (MLE) methodology (Claeskens and Hjort, 2008; Jullum and Hjort, 2017). Such procedures are efficient under model conditions, but not robust.

In this work we extend the theory and application of the FIC to the use of robust estimators, such as the density power divergence (DPD) estimator (Basu et al., 1998), and to a newly developed maximum weighted likelihood (MWL) estimator (Hjort and Walker, 2018). By comparing with a robust nonparametric alternative, we may perform focused model selection and inference also in situations where the model might be misspecified, or where the data might be contaminated with atypical values or outliers.

BASEIC FIC APPROACH
The basic FIC approach is defined as follows:

\[ \hat{\epsilon} = \text{argmin}_{\epsilon} \left( \frac{1}{2} \log \sum \frac{1}{\epsilon^2} + \frac{1}{2} \sum \log \left( 1 + \frac{x_i}{\epsilon} \right) - \sum \log \phi (x_i) \right) \]

where \( \phi \) is the density function of the base model.

It is a two-step extremum estimator (Amemiya, 1985; Newey and McFadden, 1994), where in the first step we use MAD (median of absolute deviations from the median) of the ratios as our focus parameter, which downweights data values which has low probability density under the model, but unlike DPD, weights with square root of FIC on the x-axis and estimated MAD values on the y-axis. For DPD, its tuning parameter is in the range 0 (triangle) to 1.5 (square) in steps of 0.1 (points). For MWL, its tuning parameter is in the range 0 (triangle) to 0.1 (square) in steps of 0.01 (points), except for the Weibull model where \( a \) ranges from 0 to 0.3 in steps of 0.03.

MAXIMUM WEIGHTED LIKELIHOOD ESTIMATOR - MWL
The MWL estimator (Hjort and Walker, 2018) is defined as follows:

\[ \hat{\epsilon} = \text{argmin}_{\epsilon} \left( \frac{1}{2} \log \sum \frac{1}{\epsilon^2} + \frac{1}{2} \sum \log \left( 1 + \frac{x_i}{\epsilon} \right) - \sum \log \phi (x_i) \right) \]

where \( \phi \) is an estimated weight function, with weight parameters \( a, b \), estimated from the same data \( x_i \).

It is a step extremum estimator (Amemiya, 1985; Newey and McFadden, 1994), where in the first step the parameters of the weight function is estimated using e.g. DPD, and where the parameters of the model density is estimated in the second step applying the estimated weight function. MWL is an M-estimator if the weight function is predefined and independent of the data. If the weight function is constant and equal to one, MWL is identical to MLE.

Here we use MWL with a density threshold weight function defined as follows:

\[ \phi(x) = \begin{cases} 1, & x < x_{\text{thres}} \\ \frac{1}{1 + e^{-\beta (x - x_{\text{thres}})}}, & x \geq x_{\text{thres}} \end{cases} \]

where \( \beta \) is a threshold function, which has low probability density under the model, but unlike DPD, weights with downweighting as compared with DPD. MWL will be robust with a bounded influence function (B-robust) if DPD with a > 0 is used as a weight function estimator in Step 1.

DENSITY POWER DIVERGENCE ESTIMATOR - DPD
The DPD estimator (Basu et al., 1998) is defined as follows:

\[ \hat{\epsilon} = \text{argmin}_{\epsilon} \left( \left( \frac{1}{2} \log \sum \frac{1}{\epsilon^2} + \frac{1}{2} \sum \log \left( 1 + \frac{x_i}{\epsilon} \right) - \sum \log \phi (x_i) \right) - \lambda \sum \left( \frac{x_i}{\epsilon} \right)^{2a} \right) \]

where \( \phi \) is the density function of the base model.

It is an M-estimator which is robust with a bounded influence function (B-robust) if the robustness vs. efficiency tuning parameter a > 0. Increasing a leads to increased robustness but typically less efficiency as compared with MLE. DPD approaches MLE in the limit when a \( \rightarrow 0 \).

REAL CASE EXAMPLE: MAMMALS RATIO OF BODY VS. BRAIN WEIGHTS
We consider the “msleep” dataset, publicly available in the R package “ggplot2” (Wickham, 2009: R core team, 2018), which is an updated and expanded version of the “mammals sleep” dataset (Savage and West, 2007). It contains data on average brain weight (kg) and body weight (kg) of n = 56 species of mammals. Here we consider the ratio of body weight divided by brain weight for each of the n species as our data.

We use MAD (median of absolute deviations from the median) of the ratios as our focus parameter, which represents a robust scale estimator for these data. We perform a robust FIC analysis using four parametric candidate models: Exponential; Gamma; Weibull and Lognormal, combined with DPD and MWL estimators for a range of their respective tuning parameters, and compare FIC scores obtained with these with the FIC score obtained from the nonparametric alternative based on estimating MAD directly from the data.

CONCLUDING REMARKS
In this work we extended the theory and application of the FIC to the use of robust estimators, such as the density power divergence estimator (Basu et al., 1998), and to a newly developed maximum weighted likelihood estimator (Hjort and Walker, 2018).

By comparing with nonparametric alternatives, using FIC, we may perform model selection and inference in situations where the model might be misspecified, or where data might be contaminated with atypical values or outliers which could have a large impact on the estimated focus parameter.

ACKNOWLEDGEMENTS
This work was done as part of the FocuStat project at the Dept. of Mathematics, Univ. of Oslo http://www.mn.uio.no/math/english/research/projects/focustat/ which is a 5 year project (2014 - 2018) funded by the Norwegian Research Council.

REFERENCES