ODIF FOR THE EXPECTATION VS. FINITE SAMPLE IF

Sari Peltonen and Pauli Kuosmanen

Signal Processing Laboratory, Digital Media Institute, Tampere University of Technology
P.O. Box 553, FIN-33101 TAMPERE, FINLAND
e-mail: sari@cs.tut.fi and pqo@cs.tut.fi

ABSTRACT

In this paper we consider different methods for assessing the robustness of a finite length filter. These methods are a recently introduced method called output distributional influence function (ODIF) and three finite sample influence functions (IFs), namely empirical IF, sensitivity curve (SC) and a jackknife based method. The finite sample IFs give information of only one realization of the inputs and depend largely on the samples in this realization. In this paper we illustrate, by using the mean and the median filters as examples, that the repeated use of these methods for random realizations and subsequent averaging leads to the ODIF for the expectation.

1. INFLUENCE FUNCTION

Influence function (IF) is a useful heuristic tool of robust statistics introduced by Hampel [2] under the name influence curve (IC) for studying the performance of filters under noisy conditions.

Definition 1. The IF of estimator T at underlying probability distribution F is given by

$$\mathrm{IF}(y) = \lim_{t \to 0^+} \frac{T\left((1-t)F + t\Delta_y\right) - T(F)}{t}$$

for those y where this limit exists.

In this definition Δ_y is the probability measure which puts mass 1 at the point y. The IF gives the effect that an infinitesimal contamination at point y has on the estimator T when divided by the mass of the contamination. So the IF gives asymptotic bias caused by the contamination and thus characterizes properties of the estimator as the number of observations approaches infinity.

We denote by Φ and ϕ the distribution and the density functions of the standard normal distribution. The influence functions for the mean and the median are shown in Figure 1 where the underlying distribution $F = \Phi$. For the mean the gross error sensitivity, i.e., the worst influence which a small amount of contamination of fixed size can have on the value of the estimator, equals infinity and for the median it is finite and equals $\sqrt{\frac{\pi}{2}} \approx 1.253$. So for the mean single outlier can carry the estimate over all bounds but for the median an outlier has a fixed influence.

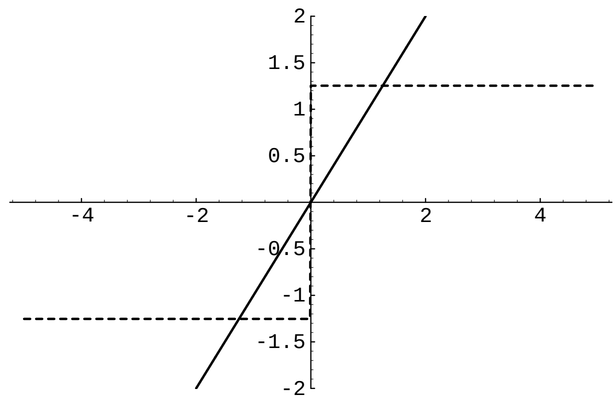

Figure 1: The IFs of the mean (–) and the median (- -) at $F = \Phi$.

2. FINITE SAMPLE INFLUENCE FUNCTION

Since the IF is an asymptotic measure, it describes properties of infinite length filters which may differ from those of finite length filters used in the real world filtering applications. It would be more useful and more interesting to examine properties of these finite length filters rather than the asymptotic properties. For this purpose some finite sample versions of the IF based on an actual sample $(X_1, X_2, \ldots, X_{N-1})$ have been proposed. Stylized versions [1] of them can be obtained by using instead of an actual sample an artificial sample which can be obtained for example by taking the $N-1$ expected order statistics from a random sample of $N-1$ or by setting $X_i = \Phi^{-1}(\frac{i}{N})$.

We present here briefly three well known finite sample versions of the IF: empirical IF, sensitivity curve (SC) and a version using jackknife (see e.g. [3] and the references therein). The first of these, the empirical IF of the estimator T_N, $N \geq 1$, at sample $(X_1, X_2, \ldots, X_{N-1})$ is a plot of

$$T_N(X_1, X_2, \ldots, X_{N-1}, y)$$

as a function of y. The SC is defined as

$$\begin{aligned} \mathrm{SC}_N(y) \quad &= \quad N\big[T_N(X_1, X_2, \ldots, X_{N-1}, y) \\ &\qquad -T_{N-1}(X_1, X_2, \ldots, X_{N-1})\big] \end{aligned}$$

or when when estimator is a functional, i.e.

$$T_N(X_1, X_2, \ldots, X_N) = T(\Phi_N),$$

as

$$\mathrm{SC}_N(y) = \frac{T\left(\left(1-\frac{1}{N}\right)\Phi_{N-1}+\frac{1}{N}\Delta_y\right)-T\left(\Phi_{N-1}\right)}{\frac{1}{N}},$$

where Φ_{N-1} is the empirical distribution function of $(X_1, X_2, \ldots, X_{N-1})$. Since the SC is a translated and rescaled version of the empirical IF, we consider in the following only the SC of these two. Instead of adding one more observation SC could also be defined by replacing an observation. This definition is related to the jackknife, the third version of the finite sample IF. It is based on the i^{th} jackknifed pseudovalue

$$\begin{aligned} T^*_{Ni} &= NT_N(X_1, X_2, \ldots, X_N) \\ &\quad -(N-1)T_{N-1}(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_N). \end{aligned}$$

Now, the finite sample IF using jackknife is at point X_i defined as

$$T^*_{Ni} - T_N(X_1, X_2, \ldots, X_N).$$

3. OUTPUT DISTRIBUTIONAL INFLUENCE FUNCTION

For the finite sample IFs either a real sample $(X_1, X_2, \ldots, X_{N-1})$ or an artificial sample generated from the distribution F of the input samples is needed and this sample itself or the way it is derived from the distribution F affects the result. What we would like to have is a general method which uses the distribution function F of the input samples itself and not any artificial sample derived from F. In the case where the output distribution of a filter can be expressed in a closed form as a function of the distribution functions of the input samples we introduced in [4] output distributional influence function (ODIF) for analyzing the robustness of the finite length filters.

We assume here that the input samples are independent and identically distributed (i.i.d.) random variables. First we need a way to denote the output distribution function of a filter when a fraction ε of the input samples has different distribution than the rest of the samples. We denote by $H_{(1-\varepsilon)F+\varepsilon G_y}(\cdot)$ the output distribution $H_F(\cdot)$ of the filter of length N where every occurrence of the common distribution function F of the input samples is replaced by $(1-\varepsilon)F+\varepsilon G_y$ and G_y is a distribution function with mean y. As usual, we define $h_{(1-\varepsilon)F+\varepsilon G_y}(x) = \frac{d}{dx}H_{(1-\varepsilon)F+\varepsilon G_y}(x)$. We gave the following definition for the ODIF for the distribution function in [4].

Definition 2 . Let the output distribution function of a filter be $H_F(\cdot)$ where $F(\cdot)$ is the common distribution function of the input samples and let $G_y(\cdot)$ be a distribution function having mean y. Then the ODIF for the distribution function $\Omega(\cdot)$ is

$$\Omega(x,y) = \lim_{\varepsilon\to 0^+} \frac{H_{(1-\varepsilon)F+\varepsilon G_y}(x) - H_F(x)}{\varepsilon}$$

for those x and y where this limit exists.

In [4] we defined the ODIF in the same way as for the distribution function in Definition 2 also for the density function and moments. For the example, the ODIF for the expectation was defined as

$$\omega_\mu(y) = \lim_{\varepsilon\to 0^+} \frac{\int_{-\infty}^{\infty} x h_{(1-\varepsilon)F+\varepsilon G_y}(x)dx - \int_{-\infty}^{\infty} x h_F(x)dx}{\varepsilon}.$$

The ODIF for the expectation gives the effect that the infinitesimal contamination in the input has on the expectation of the output of the filter when divided by the mass of this contamination. When the ODIF for the expectation is negative, the contamination G_y has decreased the expectation of the filter, and when it is positive, the expectation has increased. So the ODIF for the expectation actually is similar to the IF but it is defined for the finite length filters and the distribution function of the contamination is not limited to be Δ_y but can be any distribution function G_y having mean y.

The ODIFs for the density function and expectation can also be obtained from the ODIF for the distribution function by using the formulas

$$\omega(x,y) = \frac{d}{dx}\Omega(x,y)$$

and

$$\omega_\mu(y) = \int_{-\infty}^{\infty} x\omega(x,y)dx.$$

4. SENSITIVITY CURVE

When the input samples are independent but not identically distributed we define vector $\mathbf{F} = (F_1, F_2, \ldots, F_N)$ consisting of the distribution functions of the input samples. When large amount of SCs are calculated for random samples of size $N-1$ and the obtained results are averaged, the averaging can be replaced by expectation operation and the result is equal to

$$\begin{aligned} &\int_{-\infty}^{\infty} x\mathrm{SC}_N(y)dx \\ &\quad = N\left(\int_{-\infty}^{\infty} x h_{\mathbf{F}}(x)dx - \int_{-\infty}^{\infty} x h_{F,N-1}(x)dx\right), \quad (1) \end{aligned}$$

where $F_1, F_2, \ldots, F_{N-1} = F$, $F_N = \Delta_y$ and $h_{F,N-1}(x)$ is the output density function of $N-1$ samples. Thus the first $N-1$ samples have common distribution F but the last sample has distribution Δ_y.

We consider in the following the mean and the median filters as examples of how the SC and the ODIF for the expectation are related to each other.

4.1. Mean Filter

The stylized SC is equal to the IF for both ways of creating the artificial sample.

In [4] the ODIF for the expectation of the mean filter was derived to be $\omega_\mu(y) = y - \mu_F$ for any length N.

When we calculate the SCs for many random samples of size $N-1$ following the distribution F and average the obtained SCs, we obtain from equation (1) that

$$\begin{aligned} \int_{-\infty}^{\infty} x\mathrm{SC}_N(y)dx &= N^{N+1}\int_{-\infty}^{\infty} xf(Nx)^{*(N-1)} * \delta_y(Nx)dx \\ &\quad -N(N-1)^{N-1}\int_{-\infty}^{\infty} xf((N-1)x)^{*(N-1)}dx, \end{aligned}$$

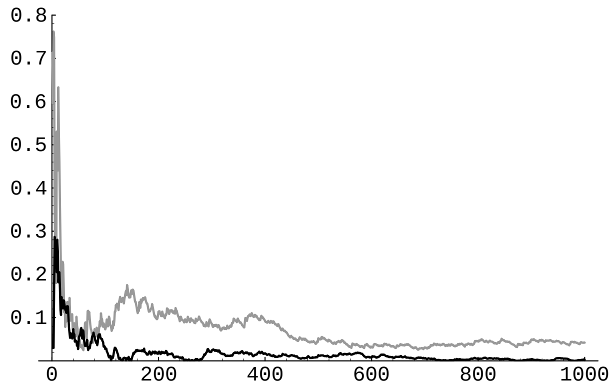

Figure 2: The mean absolute errors (MAEs) between the average of the SCs for random samples of size 4 and the ODIF for the expectation of length 5 for the mean (black) and the median (gray) filters at $F = \Phi$ and $G_y = \Delta_y$. On the x-axis is the number of SCs to be averaged.

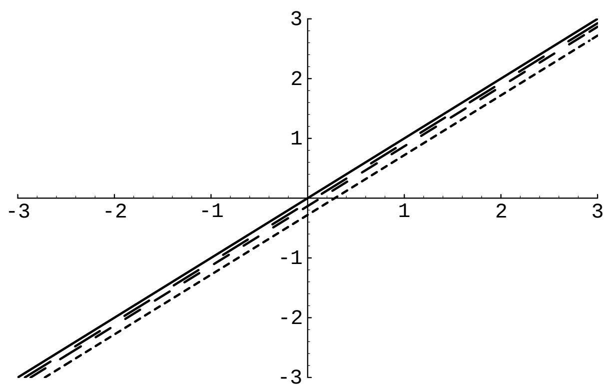

Figure 3: The ODIF for the expectation of 5 samples (solid line) and the average of 10 (small dashes), 20 (medium dashes) and 40 (long dashes) SCs of random samples of size 4 for mean when $F = \Phi$ and $G_y = \Delta_y$.

where $f(x)^{*N} = \overbrace{f(x) * f(x) * \ldots * f(x)}^{N \text{ times}}$. After some derivation we obtain that

$$\int_{-\infty}^{\infty} x\mathrm{SC}_N(y)dx = y - \mu_F.$$

So the averaging of the SCs over large amount of random samples gives the same result as the IF and the ODIF for the expectation for the mean filter.

In Figure 2 is shown (black line) how fast the average of the SCs converges to the same value as the ODIF for the expectation when the amount of the SCs to be averaged is from 1 to 1000. In Figure 3 are shown the averages of three different amounts of the SCs by the dashed lines and for comparison the ODIF for the expectation by the solid line.

From these two figures it can be seen that the result gets closer to the ODIF for the expectation as the number of SCs to be averaged grows. For the mean filter it took about 100 random samples to get a result which differs so slightly from the ODIF for the expectation that if plotted

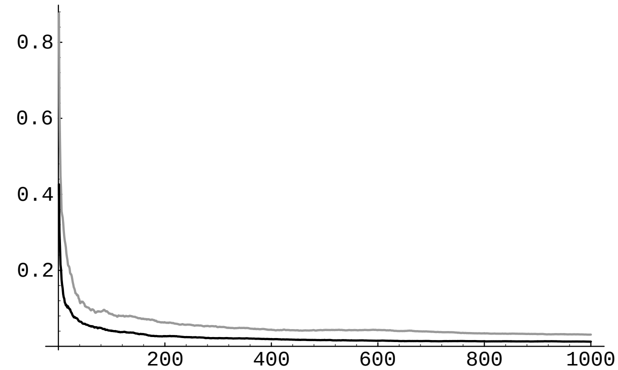

Figure 4: The average of 100 MAEs between the average of the SCs for random samples of size 4 and the ODIF for the expectation of length 5 for the mean (black) and the median (gray) filters at $F = \Phi$ and $G_y = \Delta_y$. On the x-axis is the number of SCs to be averaged.

to Figure 3 the line could not be differentiated from the solid line.

In Figure 4 is the result of averaging 100 MAE graphs like the ones in Figure 2. Averaging makes the graphs smoother and we can observe a clear decreasing pattern as the number of SCs to be averaged grows. Also we can notice that the mean filter (black) converges much faster than the median filter (gray) which we will consider in the next section.

4.2. Median Filter

In Figure 5 are shown by the dashed lines the both stylized versions of the SC for the median filter of length 5. The version where X_i is the i^{th} expected normal order statistic from a sample of 4 (small dashes) gives similar result as the ODIF for the expectation (solid line) but shows worse behavior near the origin and does not fully show how much better the behavior of the finite sample median actually is than what the IF shows. The version using $X_i = \Phi^{-1}(\frac{i}{5})$ (long dashes) approximates better the IF but has the same problem near the origin as the previous version.

When $F = \Phi$ and $G_y = \Delta_y$, an expression was obtained in [4] for the ODIF for the expectation of the median of N samples. This can be further simplified to the form

$$\omega_\mu(y) = \frac{N!}{(n!)^2}\left[\int_{-y}^{y} xn\Phi(x)^{n+1}(1-\Phi(x))^{n-1}\phi(x)dx + y\Phi(y)^n\left(1-\Phi(y)\right)^n\right].$$

When we calculate for the median filter the average of the SCs for many random samples of size $N-1$ following the distribution Φ we obtain from equation (1) that also for the median

$$\int_{-\infty}^{\infty} x\mathrm{SC}_N(y)dx = \omega_\mu(y).$$

In Figure 2 is shown (gray line) the MAE on interval $[-5, 5]$ between the average of the SCs and the ODIF for the expectation when the amount of the SCs to be averaged is

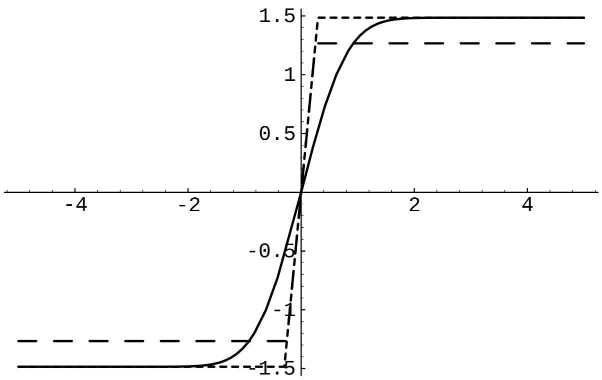

Figure 5: The ODIF for the expectation of 5 samples (solid line) and stylized versions of the SC for the median filter when X_i is the i^{th} expected normal order statistic from a sample of 4 (small dashes) and when $X_i = \Phi^{-1}(\frac{i}{5})$ (long dashes).

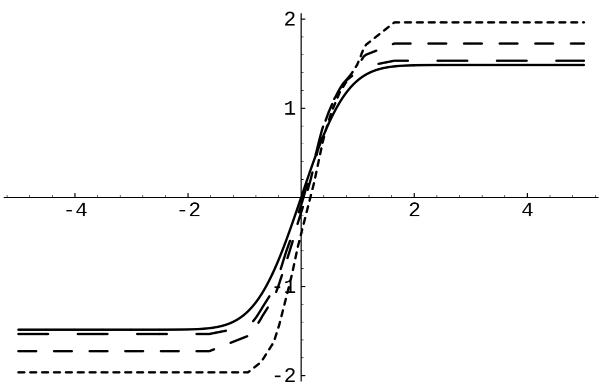

Figure 6: The ODIF for the expectation of 5 samples (solid line) and the average of 10 (small dashes), 20 (medium dashes) and 40 (long dashes) SCs of random samples of size 4 for median when $F = \Phi$ and $G_y = \Delta_y$.

from 1 to 1000. In Figure 6 are shown the averages of three different amounts of the SCs by the dashed lines and the ODIF for the expectation by the solid line for comparison. The convergence to the ODIF for the expectation is clearly not as fast for the median as it was for the mean for this case. Even after 1000 averages of the SCs the error is quite large.

5. JACKKNIFE

It is easy to derive that the expectation of the jackknife based finite sample IF gives the same result as the one obtained for the sensitivity curve but scaled by $\frac{N-1}{N}$.

When we calculate the finite sample IF by using the jackknife we obtain only N values of the graph corresponding to the points $X_1, X_2, \ldots, X_N$. So to obtain enough points to plot a continuous graph we have to calculate it many times and average the results in the points where the same random number appears more than once.

In Figure 7 are shown the results for the mean and the median when 500000 random samples are used. Even after so many random samples there are quite large deviations

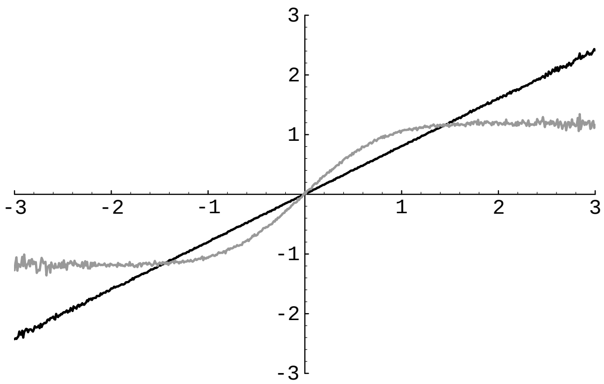

Figure 7: The average of jackknife based finite sample IF of 500000 random samples of size 5 following the distribution Φ for the mean (black) and the median (gray) filters.

in the graphs and the absolute differences between these graphs scaled by $\frac{5}{4}$ and the ODIFs for the expectation (from Figures 3 and 6) are also quite large especially further from the origin. Most of the random numbers give points near the origin and those parts of the graphs are smooth and and very close to the ODIF for the expectation but values further from the origin appear seldom and the number of the random samples should be increased a lot to obtain smooth graphs also in these areas. This however is very time consuming and not useful since the same result can be obtained much faster by simply calculating the ODIF for the expectation.

However, if the output distribution function is unknown the SC and the jackknife based methods are useful in obtaining the ODIF for the expectation.

6. CONCLUSION

In this paper we showed that when large amount of SCs or finite sample IFs based on jackknife multiplied by $\frac{N}{N-1}$ are averaged we obtain the ODIF for the expectation for the mean and the median filters. This further confirms the validity of using the ODIF for the expectation as a measure of the robustness of finite length filters.

7. REFERENCES

[1] Andrews D.F., Bickel P.J., Hampel F.R., Huber P.J., Rogers W.H. and Tukey J.W. *Robust Estimates of Location: Survey and Advances.* Princeton University Press, Princeton, New Jersey, 1972.

[2] Hampel F.R. "The Influence Curve and Its Role in Robust Estimation". *Journal of the American Statistical Association*, 69(346):383–393, June 1974.

[3] Hampel F.R., Rousseeuw P.J., Ronchetti E.M. and Stahel W.A. *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York, 1986.

[4] Peltonen S., Kuosmanen P. and Astola J. "Output Distributional Influence Function". *Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, Antalya, Turkey, June 20-23, 1999, pages 33–37.