L_1-*Statistical Procedures and Related Topics*
IMS Lecture Notes – Monograph Series (1997) Volume 31

Target estimation and implications to robustness

Luisa Turrin Fernholz

Temple University, Princeton, USA

Abstract: This paper considers target functionals $\widetilde{T}$, which are bias-reduced functionals that can be obtained from a functional T in a parametric setting. It is shown that the L_1-error of the corresponding target estimator decreases and the asymptotic normality is obtained using von Mises expansions with the Hadamard derivative. It is also shown that targeting can improve robustness since the gross-error sensitivity decreases under certain conditions. Applications to M-estimates of location, the sample median, and simultaneous M-estimates of location and scale are given.

Key words: Bias, influence function, L_1-error, parametric family, statistical functionals, target estimates, von Mises expansions.

AMS subject classification: Primary 62G99; secondary 62E20.

1 Introduction

Target estimation was introduced by Cabrera and Fernholz (1996) as a procedure to reduce the bias and the variance of an estimator. In that paper, the von Mises expansion of a statistical functional was used to obtain conditions for this reduction in bias and variance. Moreover, it was also shown that the bias of the target estimator can be expressed in terms of the influence function of the original statistic. It is natural to ask what the relationship is between the von Mises expansion of the original functional and that of the target functional. This issue is addressed in the present paper where the asymptotic distribution of the target functional will follow from the corresponding von Mises expansion. The influnce function of the target functional is also obtained and a condition is given for reducing the gross-error sensitivity.

The paper is organized as follows: In Section 2 we give a review of the basic ideas behind target functionals and we consider the L_1 error and the mean square error of these functionals. In Section 3 we look at the asymptotics and robustness of target functionals through the von Mises expansion and the influence function. Applications to M-estimates of location, the sample median, and simultaneous M-estimates of location and scale are given in Section 4.

Throughout this paper we shall assume that T is a statistical functional and the statistic $T(F_n)$ estimates the parameter $T(F_\theta)$, where F_n is the empirical d.f. corresponding to the sample $X_1, \ldots, X_n$ of i.i.d. random variables with common d.f. F_θ, with $\theta \in \Theta$, for Θ an open subset of the real numbers. When a statistical functional T satisfies $T(F_\theta) = \theta$, the functional is said to be Fisher consistent.

We shall also assume that the expectation of $T(F_n)$, $g_n(\theta) = E_\theta(T(F_n))$, exists for all $\theta \in \Theta$, where E_θ indicates the expectation with respect to F_θ. Moreover the function g_n will be assumed to be one-to-one and differentiable.

2 Target functionals

Definition 1 *Let $g_n(\theta) = E_\theta(T(F_n))$ be a one-to-one function. The functional $\widetilde{T}_n$ induced by T from the relation*

$$g_n^{-1}(T) = \widetilde{T}_n \tag{1}$$

will be called the target functional of T. The statistic $\widetilde{T}_n(F_n)$ will be called the target estimator .

The above definition was introduced by Cabrera and Fernholz (1996) where the goal was the reduction of the bias and the variance of an estimator. Note that when $g_n(\theta) = a\theta + b$ for $a \neq 0$, then the corresponding target functional is $\widetilde{T}_n = (T - b)/a$, which will always be unbiased and its variance will be reduced if and only if $a^2 > 1$. The variance of $\widetilde{T}_n$ will remain unchanged when $|a| = 1$.
It should be noted that Rousseeuw and Ronchetti (1981) used the function g_n to generate functionals $g_n^{-1}(T)$ in an entirely different context to study the influence curve of statistics used in testing hypotheses.
The following two theorems refer to the reduction in the bias $B_{\widetilde{T}}(\theta)$ and mean square error of target functionals.

Theorem 1 *If for a statistical functional T the function $g_n(\theta) = E_\theta(T(F_n))$ satisfies*

$$\begin{array}{rcll} (i) & \theta & < & g_n(\theta) \\ (ii) & 1 & < & g_n'(\theta) \leq b \\ (iii) & 0 & \leq & g_n''(\theta) \end{array}$$

then

$$a)\ E_\theta(\widetilde{T}_n) \leq \theta \leq E_\theta(T) \qquad and \qquad b)\ |(B_{\widetilde{T}_n}(\theta))| < |(B_T(\theta)|.$$

Theorem 2 *If T is a statistical functional with variance V_T and $g_n(\theta) = E_\theta(T(F_n))$ is differentiable with $|g_n'(\theta)| > 1$ for all $\theta \in \Theta$, then the mean square error of $\widetilde{T}_n$ satisfies*

$$MSE_{\widetilde{T}_n} < V_T.$$

The proofs of the above theorems can be found in Cabrera and Fernholz (1996). The other theorems in the current paper are new.
The following result refers to the L_1-error of the target estimator $\widetilde{T}_n$.

Theorem 3 *If T is a statistical functional and $g_n(\theta)$ is differentiable with $|g_n'(\theta)| > 1$ for all $\theta \in \Theta$, then*

$$E_\theta|\widetilde{T}_n - \theta| < E_\theta|T - E_\theta(T)|.$$

Proof: Using the mean value theorem for g_n we have

$$g_n(\widetilde{T}_n) = g_n(\theta) + (\widetilde{T}_n - \theta)g_n'(\xi)$$

for some ξ between θ and $\widetilde{T}_n(F_n)$. Since by definition $g_n(\widetilde{T}_n) = T$, we have

$$\widetilde{T}_n - \theta = \frac{T - g_n(\theta)}{g_n'(\xi)}.$$

By taking absolute values and their expectation of this last equation, we obtain

$$\begin{array}{rcl} E|\widetilde{T}_n - \theta| & = & E|1/g_n'(\xi)||T - g_n(\theta)| \\ & < & E|T - E(T)|, \end{array}$$

and the theorem is proved since $|g_n'(\theta)| > 1$. $\square$

An immediate consequence of this theorem is the following corollary that relates the L_1-errors of T and $\widetilde{T}$ with the median, $M_\theta(T)$, of the distribution of T.

Corollary 4 *Under the conditions of Theorem 3 we have*

$$E|\widetilde{T}_n - \theta| < E|T - \theta| + |M_\theta(T) - E_\theta(T)|$$

Proof: It follows directly from Theorem 3 since

$$\begin{aligned} E|\widetilde{T}_n - \theta| &< E|T - M_\theta(T)| + |M_\theta(T) - E_\theta(T)| \\ &\leq E|T - \theta| + |M_\theta(T) - E_\theta(T)| \end{aligned}$$

by the L_1-minimizing property of the median. □

This corollary shows that when T has a symmetric sampling distribution, the target functional $\widetilde{T}$ has smaller L_1-error.

3 Asymptotics and robustness

The results presented in Section 2 refer to the statistic $\widetilde{T}_n(F_n)$. But for each target functional $\widetilde{T}_n$ we can consider the statistic $\widetilde{T}_n(F_m)$ where n and m do not necessarily coincide. For this purpose, let F_θ be a parametric family and T be a Fisher consistent statistical functional. For each n and $g_n(\theta)$ as above, we obtain the sequence of target functionals $\widetilde{T}_n = g_n^{-1}(T)$, each one of them generating statistics $\widetilde{T}_n(F_m)$ for a sample $X_1, \ldots, X_m$.
Let T be a Fisher consistent functional with influence function ϕ_1 (see Hampel, 1974). For a sample $X_1, \ldots, X_m$ from F_θ, the first order von Mises expansion of $T(F_m)$ is

$$T(F_m) = \theta + \frac{1}{m}\sum_{i=1}^{m}\phi_1(X_i) + Rem_m. \tag{2}$$

With Hadamard or Frechet differentiability under certain regularity conditions, we have $Rem_m = o_P(\, m^{-\frac{1}{2}})$ and T is asymptotically normal with variance $\sigma^2 = E(\phi_1(X))^2$ (see Fernholz, 1983).
The following result gives the von Mises expansion of $\widetilde{T}_n$

Lemma 5 *Let T be a statistical functional and for a fixed n let $\widetilde{T}_n$ be the corresponding target functional. If T is Hadamard differentiable at F_θ with von Mises expansion as in (2), then $\widetilde{T}_n$ is also Hadamard differentiable at F_θ with von Mises expansion*

$$\widetilde{T}_n(F_m) = g_n^{-1}(\theta) + \frac{1}{m}\sum_{i=1}^{m}(1/g_n'(\theta))\phi_1(X_i) + \widetilde{Rem}_m. \tag{3}$$

Proof: First note that $\widetilde{T}_n(F_\theta) = g_n^{-1}(T(F_\theta)) = g_n^{-1}(\theta)$. Since T is Hadamard differentiable, the composition $g_n^{-1}(T) = \widetilde{T}_n$ will be Hadamard differentiable for a differentiable real function g_n by the chain rule (see Fernholz, 1983). Therefore the influence function of $\widetilde{T}_n$ is $(1/g_n'(\theta))\phi_1(x)$ and the lemma is proved. □

In general we would be interested in using target estimates when T is such that $g_n(\theta) \neq \theta$. This clearly implies that when T is Fisher consistent then $\widetilde{T}_n$ will not be. However, $\widetilde{T}_n$ has less bias than T when g_n satisfies certain conditions.
When $n = m$, the expansion in (3) provides a linear approximation of the statistic $\widetilde{T}_n(F_n)$ with the influence function given by $(1/g_n'(\theta))\phi_1(x)$. If we now compare the influence functions of $\widetilde{T}_n$ and T we can conclude immediately that, when $|g_n'(\theta)| > 1$, $\widetilde{T}_n$ is more robust than T in terms of gross-error sensitivity (see Hampel et al., 1986) since we have

$$\sup_x |(1/g_n'(\theta))\phi_1(x)| < \sup_x |\phi_1(x)|.$$

When the function $g(\theta) = a\theta + b$ is linear, $\widetilde{T} = (T - b)/a$ and the corresponding influence function satisfies $\widetilde{\phi}_1(x) = \phi_1(x)/a$. The gross-error sensitivity of $\widetilde{T}$ will be smaller than that of T when $|a| > 1$.
The expansion in (3) above is useful to derive the asymptotic normality and efficiency of $\widetilde{T}_n$ as we see in

Theorem 6 *Let T and $\widetilde{T}_n$ be as in Lemma 4. If T is Hadamard differentiable at $F = F_\theta$ with von Mises expansion as in (2) then, for fixed n and $m \to \infty$ we have :*
a) $\widetilde{T}_n$ is asymptotically normal with

$$\sqrt{m}\,(\widetilde{T}_n(F_m) - \widetilde{T}_n(F)) \xrightarrow{\mathcal{D}} N(0, \sigma_n^2)$$

where the asymptotic variance of $\widetilde{T}_n$ is $\sigma_n^2 = (1/g_n'(\theta))^2\sigma^2$;
b) If $|g_n'(\theta)| > 1$, $\widetilde{T}_n$ is asymptotically more efficient than T.

Proof: Part a) is an immediate consequence of Lemma 5 since the Hadamard differentiability of $\widetilde{T}_n$ implies the asymptotic normality of $\widetilde{T}_n$ (see Fernholz, 1983).
For part b), note that when $|g_n'(\theta)| > 1$, the asymptotic variance of $\widetilde{T}_n$, satisfies

$$\sigma_n^2 = (1/g_n'(\theta))^2\sigma^2 < \sigma^2. \ \square$$

Theorem 6 refers to the asymptotic normality of $\widetilde{T}_n(F_m)$ for fixed n when m tends to infinity. When $n = m$ is large, we have

Theorem 7 *Let T and $\widetilde{T}_n$ be as in Theorem 6. If $m = n$ tends to infinity and for all θ $\sqrt{n}(g_n(\theta) - \theta) \to 0$, and $\sqrt{n}(1/g_n'(\theta) - 1) \to 0$, then for all θ it holds that*

$$\sqrt{n}(\widetilde{T}_n(F_n) - \theta) \xrightarrow{\mathcal{D}} N(0, \sigma^2).$$

Proof: From the expansions (2) and (3) above we obtain

$$\begin{aligned} \sqrt{n}(\widetilde{T}_n(F_n) - \theta) &= \sqrt{n}(\widetilde{T}_n(F_n) - T(F_n) + T(F_n) - \theta) \\ &= \sqrt{n}(\widetilde{T}_n(F_n) - T(F_n)) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_1(X_i) + \sqrt{n} Rem_n . (4) \end{aligned}$$

Since T is Hadamard differentiable, we have

$$\sqrt{n} Rem_n \xrightarrow{P} 0$$

and

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_1(X_i) \xrightarrow{\mathcal{D}} \mathrm{N}(0, \sigma^2).$$

Hence, it suffices to show that the first term in (4) converges to zero in probability. But in

$$\sqrt{n}(\widetilde{T}_n(F_n) - T(F_n)) = \sqrt{n}(g_n^{-1}(\theta) - \theta) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1/g_n'(\theta) - 1)\phi_1(X_i)$$

$$+ \sqrt{n}(\widetilde{Rem}_n - Rem_n) \qquad (5)$$

the first term tends to zero by hypothesis and the third term converges to zero by the Hadamard differentiability of T and $\widetilde{T}_n$. For the second term in (5), Markov's inequality implies that for any $\epsilon > 0$

$$P\Big\{ \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1/g_n'(\theta) - 1)\phi_1(X_i) \Big| > \epsilon \Big\} \geq \sqrt{n} |1/g_n'(\theta) - 1| \, E|\phi_1(X)|/\epsilon,$$

which tends to zero when n tends to infinity since by hypothesis $\sqrt{n}(1/g_n'(\theta) - 1) \to 0$ and the theorem is proved. □

From Theorems 6 and 7 we can see that, when $n = m$ is large, there is little gain for the target estimators in terms of gross-error sensitivity. However these theorems insure that there will be no loss in gross-error sensitivity when we use targeting to reduce bias.

4 Examples

4.1 M-estimates of location

Since in this section the sample size n will always be fixed, we shall simplify the notation by using $g(\theta)$ instead of $g_n(\theta)$ and $\widetilde{T}$ for the target functionals. For a parametric family F_θ, let the functional $T(F_\theta) = \widehat{\theta}$ be defined implicitly as a solution of

$$\int \psi(x - \widehat{\theta}) dF_\theta(x) = 0.$$

The corresponding statistic $T(F_n)$ is the well known M-estimate of location. When the parametric family satisfies $F_\theta(x) = F(x-\theta)$ for all θ and some d.f. F, the corresponding M-estimate is location equivariant, that is $T(F_\theta) = \theta + T(F)$. It was shown in Cabrera and Fernholz (1996) that the bias of an M-estimate of location is constant. So $g(\theta) = \theta + B$, and the influence functions of T and $\widetilde{T}$ coincide.

4.2 The sample median

The functional $T(F_\theta) = F_\theta^{-1}(1/2) = \widehat{\theta}$ corresponds to the sample median $T(F_n) = F_n^{-1}(1/2)$. When F_θ is not symmetric about θ the sample median is biased. The second order von Mises expansion of T (see Fernholz, 1996; Fankhauser, 1996) gives

$$\begin{aligned} g(\theta) &= E_\theta(F_\theta^{-1}(1/2)) \\ &= \theta - \frac{f'(\theta)}{8n\, f^3(\theta)} + o(1/n). \end{aligned}$$

Its derivative is

$$g'(\theta) = 1 + \frac{(f'(\theta))^2 - f(\theta) f''(\theta)}{2n\, f^4(\theta)} + o(1/n),$$

hence for θ such that $f''(\theta) < 0$, we have $g'(\theta) > 1$. In this case T satisfies the hypotheses of most of the theorems presented above. The target estimator $g^{-1}(F_n^{-1}(1/2))$ will have smaller bias, MSE, and gross-error sensitivity than $F_n^{-1}(1/2)$.

4.3 Simultaneous M-estimates of location and scale.

Consider a family of d.f.'s F_θ with $\theta = (\mu, \sigma)$ and the two-dimensional functional $T(F_\theta) = (T_1(F_\theta), T_2(F_\theta))$ defined implicitly by

$$\int \psi\Big(\frac{x - T_1(F_\theta)}{T_2(F_\theta)}\Big) dF_\theta(x) = 0,$$

where $\psi = (\psi_1, \psi_2)$. The corresponding statistic $T(F_n) = (T_1(F_n), T_2(F_n))$ satisfies a system of equations and is called an M-estimate of location and scale. See Huber (1981) and Hampel et al (1986).
When the family of distributions is such that $F_\theta(x) = F(\frac{x-\mu}{\sigma})$ for some fixed d.f. F, the functional T satisfies $T(F_{\mu,\sigma}) = (\mu + \sigma T_1(F), \sigma T_2(F))$ and is said to be location-scale equivariant (see Hampel et al , 1986). In this case, it can be shown that

$$g(\theta) = g(\mu, \sigma) = (\mu + \sigma C_1, \sigma C_2)$$

where C_1 and C_2 are constants independent of μ and σ. Now, the target functional is $\widetilde{T} = (\widetilde{T}_1, \widetilde{T}_2)$ with

$$\begin{aligned}\widetilde{T}_1 &= T_1 - (C_1/C_2)T_2 \\ \widetilde{T}_2 &= (1/C_2)\,T_2,\end{aligned}$$

and so the corresponding two-dimensional influence function is $\widetilde{\phi} = (\widetilde{\phi}_1, \widetilde{\phi}_2)$ where

$$\begin{aligned}\widetilde{\phi}_1(x) &= \widetilde{\phi}_1(x) - (C_1/C_2)\,\phi_2(x) \\ \widetilde{\phi}_2(x) &= (1/C_2)\,\phi_2(x).\end{aligned}$$

If we let $\| \ \|$ denote the Euclidean norm, then for each x

$$\begin{aligned}\|\widetilde{\phi}(x)\|^2 &= \left|\widetilde{\phi_1}(x)\right|^2 + \left|\widetilde{\phi_2}(x)\right|^2 \\ &= |\phi_1(x) - (C_1/C_2)\,\phi_2(x)|^2 + |(1/C_2)\,\phi_2(x)|^2 \\ &\le |\phi_1(x)|^2 + |(C_1/C_2)\,\phi_2(x)|^2 + |(1/C_2)\,\phi_2(x)|^2 \\ &= |\phi_1(x)|^2 + ((C_1/C_2)^2 + (1/C_2)^2)\,|\phi_2(x)|^2 \\ &< |\phi_1(x)|^2 + |\phi_2(x)|^2 \\ &= \|\phi(x)\|^2\end{aligned}$$

when $C_1^2 + 1 < C_2^2$. Therefore the gross-error sensitivity of $\widetilde{T}$ will be smaller than that of T when $C_1^2 + 1 < C_2^2$.

5 Closing remarks.

We should note that the regularity conditions of the theorems in Sections 2 and 3 are not too stringent as shown by the examples and applications of target estimation presented in Section 4 as well as in Cabrera and Fernholz (1996). Moreover, the hypotheses of Theorem 7 are not as restrictive as they may seem. The condition $(g_n(\theta) - \theta) = o(1/\sqrt{n})$ is equivalent to

$E_\theta(Rem_n) = o(1/\sqrt{n})$ which is quite plausible since $Rem_n = o_P(1/\sqrt{n})$ for Hadamard differentiable functionals. The condition referring to the derivative of g_n is also satisfied by the most reasonable statistics. The examples presented in Section 4 all satisfy these conditions.
Target estimation is a computer intensive procedure that has proved to be very effective in reducing the bias and the variance of estimators. Target estimation can also be performed when the function $g_n(\theta)$ is the median of the statistic, as was first presented in Cabrera and Watson (1997). The examples given in Cabrera and Fernholz (1996) and Cabrera and Watson (1997), as well as the applications to practical problems in computer vision of Cabrera and Meer (1996), reveal that target estimation is an effective method of bias and variance reduction in many situations. This paper shows the additional gain in robustness due to targeting.
Although the problems of bias reduction and robustness of an estimator seem to be two independent issues, we showed in this paper that the conditions for bias and variance reduction will assure a smaller bound for the influence function of the bias-reduced functional, which means smaller gross-error sensitivity for the target estimator. The von Mises expansions proved to be a powerful tool to approach the theoretical issues of target estimation. Simulations and practical applications need to be performed now to evaluate more precisely the gain in robustness when estimators are targeted. This is a topic of ongoing research.

References

[1] Cabrera, J. and Fernholz, L.T. (1996). Target estimation for bias and varance reduction. Submitted for publication.

[2] Cabrera, J. and Meer , P. (1996). *Unbiased Estimation of Ellipses by Bootstrapping.* In Press.

[3] Cabrera, J. and Watson, G.S.(1997). Simulation methods for mean and median bias reduction in parametric estimation. *J. Statist. Plann. Infer.* **57** 143-152

[4] Fankhauser, E. (1996). Etude du développement de von Mises au deuxième ordre. Ph.D. dissertation. Ecole Polytechnique Federale, Lausanne, Switzerland.

[5] Fernholz, L. T. (1983). *Von Mises Calculus for Statistical Functionals.* Lecture Notes in Statistics Vol. 19. New York: Springer Verlag.

[6] Fernholz, L.T. (1996). On higher order von Mises expansions. Technical Report 96-04, Temple University, Dept. Statistics.

[7] Hampel, F. (1974). The influence curve and its role in robust estima-

tion. *J. Am. Statis. Assoc.* **69** 383-393.

[8] Hampel, F., Ronchetti, E., Rousseeuw, P., Stahel, W. (1986). *Robust Statistics. The Approach Based on the Influence Function.* New York: Wiley.

[9] Huber, P. (1981). *Robust Statistics.* New York: Wiley.

[10] Rousseeuw, P. and Ronchetti, E. (1981). Influence curves of general statistics. *Journal of Computational and Applied Mathematics* **7** 161-166.