The Super Robustness of Maximum Likelihood Location Estimator of Exponential Power Distribution, When $p<1$

Qinghuai Gao

Abstract We proof that statistically, the maximum likelihood location estimator of exponential power distribution is strict super robust, when $p<1$.

1. Parameter Estimation Problem

A system that transforms an input I to an output O with a transformation T is defined mathematically as below:

$$O = T(I) \tag{1}$$

Estimating the transformation T based on a group of input and output pairs of a system is a central and challenging problem in many pattern matching and computer vision systems. Typical examples are medical image registration, fingerprint matching, and camera model estimation.

The following concepts will be used in this paper:

Estimator An estimation approach to generate the system parameters based on groups of observations.

Experiment A group of observations that is used to generate an estimated transformation.

Strict robustness The capability of an estimator that gives the perfect estimation even when the good observations are perfect but the noise observations have any possible distribution.

Super robustness The characteristics that an estimator generates an estimated transformation whose error to the perfect estimation is bounded even when the noise observations are majority, and they move to infinite.

Suppose $I_1, I_2 \cdots\cdots, I_N$ are N inputs of a system defined in the formulae (1), where I_i is a point in an Euclidean space, and $O_1, O_2 \cdots\cdots, O_N$ are the corresponding outputs, where O_i is a point in an Euclidean space that may have different dimension than the input space. For a transformation T, we define the difference of O_i and $T(I_i)$ as

$$d(O_i, T(I_i)) \tag{2}$$

Thus, the overall difference between the observed outputs and the estimated outputs based on T is

$$\sum_{i=1}^{N} d(O_i, T(I_i)) \tag{3}$$

The problem to estimate T becomes that find a T_b, which satisfies:

$$D_{T_b} = \min_{T} \sum_{i=1}^{N} d(O_i, T(I_i)) \tag{4}$$

The minimum takes on any possible transformation T in a predefined transformation group. We will only discuss translation in this paper. When d is Euclidean distance, it is the least square estimation.

In this paper, we use L^p to define the difference of O_i and $T(I_i)$:

$$|O_i - T(I_i)|^p \tag{5}$$

This leads to the maximum likelihood estimator of the exponential power distribution. We simply call it L^p estimator.

Now, the estimation problem is converted to that find a T_b that satisfies:

$$\sum_{i=1}^{N} |O_i - T_b(I_i)|^p = \min_{T} \sum_{i=1}^{N} |O_i - T(I_i)|^p \tag{6}$$

To observe the robustness characteristics of L^p estimator, we divide the observations into two groups: all observations in the first group are perfect observations: for $i = 1, 2, \cdots\cdots, n$, $O_i = T_b(I_i)$, where T_b is the ideal transformation; all observations in the second group are noisy, that is, $O_i \neq T_b(I_i)$, where $i = n+1, n+2, \cdots\cdots, N$. In this paper, we denote the number of noise observations as M : $M = N - n$.

We will investigate that how much percent of the ideal output out of the total observations still results an ideal estimation T_b or a reliable estimation of T when L^p estimator is used.

To pursue strict robustness, we expect that n satisfies

$$D_T \geq D_{T_b} \tag{7}$$

for any possible transformation T in a transformation group. Or

$$\sum_{i=1}^{n} |O_i - T(I_i)|^p + \sum_{i=1}^{M} |O_{i+n} - T(I_{i+n})|^p \geq \sum_{i=1}^{M} |O_{i+n} - T_b(I_{i+n})|^p \tag{8}$$

since $O_i = T_b(I_i)$ for $i = 1, 2, \cdots\cdots, n$.

2. **Strict Robustness on Translation**

A translation is define as:

$$T(I) = I + a_T \tag{9}$$

For the ideal transformation T_b, we define it as:

$$T_b(I) = I + a_{T_b} \tag{10}$$

For the *i*-th observation in the first group, the difference between the observation and the output of the system with a transform *T* is:

$$O_i - T(I_i) = T_b(I_i) - T(I_i) = a_{T_b} - a_T \tag{11}$$

By denoting that $d_T = \left| a_{T_b} - a_T \right|$, we have:

$$nd_T^p + \sum_{i=1}^{M} \left| O_{i+n} - T(I_{i+n}) \right|^p \geq \sum_{i=1}^{M} \left| O_{i+n} - T_b(I_{i+n}) \right|^p \tag{12}$$

Denote that $d_i = \left| O_{i+n} - T_b(I_{i+n}) \right|$. The right side of the above inequality is

$$\sum_{i=1}^{M} d_i^p \tag{13}$$

The left side is no less than:

$$nd_T^p + \sum_{i=1}^{M} \left| d_i - d_T \right|^p \tag{14}$$

Thus, if

$$nd_T^p + \sum_{i=1}^{M} \left| d_i - d_T \right|^p \geq \sum_{i=1}^{M} d_i^p \quad \textit{or} \quad nd_T^p \geq \sum_{i=1}^{M} d_i^p - \sum_{i=1}^{M} \left| d_i - d_T \right|^p \tag{15}$$

then $D_T \geq D_{T_b}$.

When $d_T \leq d_i$, we have

$$d_i^p - \left| d_i - d_T \right|^p \leq d_T^p \tag{16}$$

This is because $(a+b)^p < a^p + b^p$ when $p < 1$, $a > 0$ and $b > 0$.

When $d_T > d_i$,

$$d_i^p - \left| d_T - d_i \right|^p < d_T^p \tag{17}$$

Thus, when $n \geq M$, we always have $D_T \geq D_{T_b}$ for any translation.

Theorem 1. $L^p(p < 1)$ location estimator is strict robust.

3. **Strict Super Robustness on Translation**

To make estimation simpler, without loss generality, we assume that:

$$d_1 < d_2 < \cdots < d_M \tag{18}$$

When $d_T \in [d_k, d_{k+1})$, we divide the items into two groups:

$$\sum_{i=1}^{k} \left[d_i^p - \left(d_T - d_i \right)^p \right] + \sum_{i=k+1}^{M} \left[d_i^p - \left(d_i - d_T \right)^p \right] \tag{19}$$

The first group with that d_i is no larger than d_T is named as TFG. The second group with $d_i > d_T$ is named as TSG. We will separately estimate the upper bounds of them.

3.1 d_i has a uniform distribution

Before sorted, d_i has a uniform distribution. After sorted, the normalized d_i has a beta distribution $B(i, M+1-i)$ [3], which has a distribution function of $C_{M+1}^{i} x^{i} (1-x)^{M-i}$.

The normalized $d_{M/2}$ has a distribution function $C_{M+1}^{M/2} x^{M/2} (1-x)^{M/2}$. Based on Hoeffding's inequality [4] with $n = M+1$, $p = 1/2 + M^{a-1}$, and $k = M/2$, we have $I_{1/2-M^{a-1}}(M/2+1, M/2+1) \le e^{-2M^{2a-1}}$ so

$$P\left(\left|d_{M/2} - 1/2\right| < M^{a-1}\right) > 1 - e^{-2M^{2a-1}} \tag{20}$$

Similarly, the formulae above should be valid for all d_i. Then we have:

$$P\left(\bigcap_{i=1}^{M} \left|d_i - i/M\right| < M^{a-1}\right) > \left(1 - e^{-2M^{2a-1}}\right)^{M} \tag{21}$$

When $a > 1/2$, the right side goes to 1 when M goes to infinite. Since $2e^{-2M^{2a-1}}$ rapidly reduces to 0, for relatively large M, we find an a such that the right side is no less than any given percentage. The minimum a-*s* for the right side is no less than 99.9% for M from 100 to 1000 are listed in the table below:

M	a
100	0.696
200	0.676
300	0.666
400	0.660
500	0.655
600	0.652
700	0.649
800	0.647
900	0.645
1000	0.643

Table 1. *a* for d_i close to the mean *k/M*

When $\left|d_i - k/M\right| < M^{a-1}$ for all these distances, for TFG, when $k < 2M^a$, the upper bound is $2M^a$; when $k \ge 2M^a$, the upper bound is:

$$
\begin{aligned}
&\sum_{i=1}^{k}\left[d_i^p-(d_T-d_i)^p\right]/d_T^p=\sum_{i=1}^{k}\left[\left(\frac{d_i}{d_T}\right)^p-\left(1-\frac{d_i}{d_T}\right)^p\right]\\
&\le\sum_{i=1}^{k-M^a}\left[\left(\frac{(i+M^a)/M}{k/M}\right)^p-\left(1-\frac{(i+M^a)/M}{k/M}\right)^p\right]+M^a\\
&\le\sum_{i=1}^{k-M^a}\left(\frac{i+M^a}{k}\right)^p-\sum_{i=0}^{k-M^a}\left(\frac{i}{k}\right)^p+M^a\\
&=\sum_{i=k-M^a+1}^{k}\left(\frac{i}{k}\right)^p-\sum_{i=0}^{M^a}\left(\frac{i}{k}\right)^p+M^a\\
&<2M^a
\end{aligned}
\tag{22}
$$

So the upper bound for TFG is $2M^a$.

For TSG, we have:

$$
\begin{aligned}
&\sum_{i=k+1}^{M}\left[d_i^p-(d_i-d_T)^p\right]/d_T^p=\sum_{i=k+1}^{M}\left[\left(\frac{d_i}{d_T}\right)^p-\left(\frac{d_i}{d_T}-1\right)^p\right]\\
&\le\sum_{i=k+M^a}^{M}\left[\left(\frac{i-M^a}{d_T}\right)^p-\left(\frac{i-M^a}{d_T}-1\right)^p\right]+M^a\\
&(because\quad x^p-(x-1)\quad is\quad descreasing)\\
&\le\int_{k+1}^{M+1-M^a}\left[\left(\frac{x}{d_T}\right)^p-\left(\frac{x}{d_T}-1\right)^p\right]dx+M^a\\
&=\frac{1}{(p+1)d_T^p}\left((M+1-M^a)^{p+1}-(M+1-M^a-d_T)^{p+1}-(k+1)^{p+1}\right)+M^a\\
&\le\frac{1}{(p+1)d_T^p}\left((M+1-M^a)^{p+1}-(M+1-M^a-d_T)^{p+1}-d_T^{p+1}\right)+M^a
\end{aligned}
\tag{23}
$$

Because

$$
\begin{aligned}
&-(M+1-M^a-d_T)^{p+1}\\
&=-\sum_{i=0}^{\infty}\binom{p+1}{i}(M+1-M^a)^{p+1}\left(-d_T/(M+1-M^a)\right)^i\\
&=-(M+1-M^a)^{p+1}+(p+1)(M+1-M^a)^p d_T+many\quad negative\quad items
\end{aligned}
\tag{24}
$$

, TSG is no larger than:

$$
(M+1-M^a)^p d_T^{1-p}-d_T/(1+p)+M^a \tag{25}
$$

When $d_T=(1-p^2)^{1/p}M$, it has a maximum of

$$p\left(1-p^2\right)^{(1-p)/p}\left(M+1-M^a\right)+M^a \quad (26)$$

Theorem 2. When d_i has a uniform distribution, statistically, $L^p\,(p<1)$ location estimator is strict super robust for large *M* and small *p*.

The table below shows the super robustness numerically:

p	*n/M*
0.50	0.60
0.45	0.57
0.40	0.54
0.35	0.51
0.30	0.48
0.25	0.44
0.20	0.41
0.15	0.38
0.10	0.34
0.05	0.30

Table 1. The numerical relation of *p* and *n/M* when *M*=1000 and *a*=0.643.

3.2 Super Robustness on More General Noise Distribution

Before sorted, d_i has a distribution function $f(x)$ and a cumulative distribution function $F(x)$. Based on order statistics, the distribution of *k*-th smallest value [5] is:

$$C_M^k F(x)^k \left(1-F(x)\right)^{k-1} f(x) \quad (27)$$

For this distribution, the probability that x falls in the interval $[a,b]$ is:

$$\int_a^b C_M^k F(x)^k \left(1-F(x)\right)^{k-1} f(x)dx = \int_{F^{-1}(a)}^{F^{-1}(b)} C_M^k x^k \left(1-x\right)^{k-1} dx \quad (28)$$

Thus, we have:

$$P\left(F^{-1}(k/M-M^a) < d_k < F^{-1}(k/M+M^a)\right) > 1-2e^{-2M^{2a-1}}/M \quad (29)$$

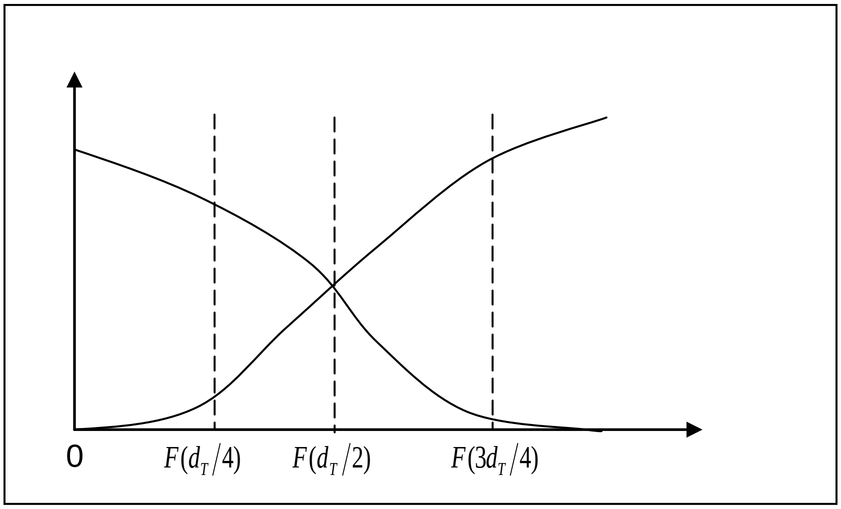

Fig. 1 The relation of $\left(\frac{d_i}{d_T}\right)^p$ and $\left(1-\frac{d_i}{d_T}\right)^p$

For TFG,

$$\sum_{i=1}^{k}\left[d_i^p-(d_T-d_i)^p\right]/d_T^p=\sum_{i=1}^{k}\left[\left(\frac{d_i}{d_T}\right)^p-\left(1-\frac{d_i}{d_T}\right)^p\right] \tag{31}$$

Let's divide it into four groups as shown in Fig. 1: (1) when $d_i \le F(d_T/4)$,

$$\frac{F^{-1}(d_i)}{d_T}\le 1-\frac{F^{-1}(d_i)}{d_T}-\frac{1}{2},$$

or

$$\left(\frac{F^{-1}(d_i)}{d_T}\right)^p-\left(1-\frac{F^{-1}(d_i)}{d_T}\right)^p\le -\frac{1}{2^p};$$

when $d_i \in [F(d_T/4), F(d_T/2)]$,

$$\frac{F^{-1}(d_i)}{d_T}\le 1-\frac{F^{-1}(d_i)}{d_T},$$

or

$$\left(\frac{F^{-1}(d_i)}{d_T}\right)^p\le\left(1-\frac{F^{-1}(d_i)}{d_T}\right)^p;$$

when $d_i \in [F(d_T/2), F(3d_T/4)]$,

$$\frac{F^{-1}(d_i)}{d_T}\le \frac{3}{2}-\frac{F^{-1}(d_i)}{d_T},$$

or

$$\left(\frac{F^{-1}(d_i)}{d_T}\right)^p-\left(1-\frac{F^{-1}(d_i)}{d_T}\right)^p\le \frac{1}{2^p};$$

Thus, the overall upper bound of TFG is :

$$\left(F(d_T)-F(3d_T/4)\right)M-F(d_T/4)M/2^p+\left(F(3d_T/4)-F(d_T/2)\right)M/2^p . \quad (32)$$

For TSG, when $F(d_T)>0.5$, the upper bound is $M-MF(d_T)$, in this case, we notice that the TFG upper bound is much smaller than $MF(d_T)$; when $F(d_T)\in[0.25,0.5]$, because the function $x^p-(x-1)^p$ is decreasing, the upper bound is $\left((3/2)^p-(1/2)^p\right)\left(1-F(3d_T/2)\right)M+\left(F(3d_T/2)-F(d_T)\right)M$; when $F(d_T)\in[0,0.25]$, the upper bound is

$\sum_{i=1}^{2}\left((i+1)^p-i^p\right)\left(F((i+1)d_T)-F(id_T)\right)M+\left(1-F(4d_T)\right)M+\left(F(2d_T)-F(d_T)\right)M$. For small d_T, we can obtain a lower upper bound

$\sum_{i=1}^{\lfloor 1/d_T\rfloor-1}\left((i+1)^p-i^p\right)\left(F((i+1)d_T)-F(id_T)\right)M+\left(1-F(\lfloor 1/d_T\rfloor d_T)\right)M+\left(F(2d_T)-F(d_T)\right)M$

by repeating the same estimation approach.

Summarizing the above, we have:

Theorem 3. When d_i has same distribution function, statistically, $L^p\,(p<1)$ location estimator is strict super robust for large *M* and small *p*.

Specially, when all the components of the error vector has a uniform distribution, d_i has a distribution of $f(x)\sim(K+1)x^K$ and $F(x)$ is x^{K+1}, for which, statistically, $L^p\,(p<1)$ location estimator is strict super robust for large *M* and small *p*.

4. Conclusions

Statistically, the maximum likelihood location estimator of the exponential power distribution, or L^p location estimation, when $p<1$, is strict super robust.

5. References

[1]. *Robust Statistics*, Peter. J. Huber, Wiley, 1981 (republished in paperback, 2004)

[2]. *Robust Statistics - The Approach Based on Influence Functions*, Frank R. Hampel, Elvezio M. Ronchetti, Peter J. Rousseeuw and Werner A. Stahel, Wiley, 1986 (republished in paperback, 2005).

[3]. http://en.wikipedia.org/wiki/Beta_distribution

[4].http://en.wikipedia.org/wiki/Binomial_distribution#Cumulative_distribution_function

[5]. http://en.wikipedia.org/wiki/Order_statistic

[6]. http://arxiv.org/abs/1206.5057 : The Robustness and Super-Robustness of L^p Estimation, when p < 1

Remark: This paper is the extraction of the results on location estimation published in the reference [6], the most matured theoretic pieces. When a logistic transform is done on the original data, scaling transform becomes translation so it is possible to estimate scaling on very noise data. It needs more works on the super robustness of more complicated transformations.