Friedrich-Alexander-Universität Erlangen-Nürnberg

Wirtschafts-und Sozialwissenschaftliche Fakultät

Diskussionspapier
23 / 1998

On Idempotent Estimators of Location

Ingo Klein

Lehrstuhl für Statistik und Ökonometrie
Lehrstuhl für Statistik und empirische Wirtschaftsforschung
Lange Gasse 20 · D-90403 Nürnberg

IDEMPOTENT ESTIMATORS OF LOCATION

Ingo Klein
Lehrstuhl für Statistik und Ökonometrie
Universität Erlangen–Nürnberg
Lange Gasse 20
D–90403 Nürnberg
Germany
E-mail: ingo.klein@wiso.uni-erlangen.de

Abstract

Idempotence is a well–known property of functionals of location. It means that the value of the functional at a singular distribution must be identically to the mass point of this distribution. First, we explain the role of idempotence in the known axiomizations of location functionals. Then we derive the distribution of idempotent and sufficient statistics. In the special cases of parametric families of location we get the so–called power–n–distributions. Power–n–distributions again are distributions with a parameter of location and can be derived from every location family for which the density is constrained. Additionally we show that the completeness of the populations family insures the completeness of the family of power–n–distributions. And at last, we give a further, now very easy proof that the normal distribution is the only one for which a idempotent, sufficient and unbiased estimator attains the Cramér–Rao–lower bound.

1 Introduction

Idempotence is a well–known property of functionals of location. It means that the value of the functional at a singular distribution must be identically to the mass point of this distribution.

First, we explain the role of idempotence in the known axiomizations of location functionals. Then we derive the distribution of idempotent and sufficient statistics. In the special cases of parametric families of location we get the so–called power–n–distributions. Power–n–distributions again are distributions with a parameter of location and can be derived from every location family for which the density is constrained. Additionally we show that the completeness of the populations family insures the completeness of the family of power–n–distributions. And at last, we give a further, now very easy proof that the normal distribution is the only one for which a idempotent, sufficient and unbiased estimator attains the Cramér–Rao–lower bound.

2 Statistical measures of location

2.1 Statistical functionals

Let $\mathcal{D}$ be the set all univariate statistical distribution functions and $F \in \mathcal{D}$. Furthermore, let $X_1, \ldots, X_n$ be a sample from the population with distribution function F and $T_n = T_n(X_1, \ldots, X_n)$ be a statistic. If T_n can be written as $T_n) = T(F_n)$, where T does not depend on n and F_n is the empirical distribution function of $X_1, \ldots, X_n$, then T will be called a statistical functional (see Fernholz (1983), p. 5). As domain of t will be considered a set of distribution functions $\mathcal{F}$ that contains the empirical distribution functions F_n for all $n \geq 1$ and the population distribution function F. Instead of $T(F)$ we write $T(X)$ if X is distributed according to F.

2.2 Idempotence

Let $\mathcal{F}$ contain the singular distribution δ_x with mass point $x \in \mathbb{R}$ what means $\delta_x(y) = I_{[x,\infty)}(y)$, $y \in \mathbb{R}$ with indicator function

$$I_A(y) = \begin{cases} 1 & \text{for } y \in A \\ 0 & \text{for } y \notin A \end{cases}$$

for $A \subseteq \mathbb{R}$. T will be called idempotent if $T(\delta_x) = x$.

Instead of idempotence Fleming & Wallace (1986) speak about "reflexivity" and Aczél (1990) about "agreement". Reflexivity or agreement are one of the characterizing properties of so–called merging functions that aggregate different ratio-scaled measurements.

Idempotence can also be used for the characterization of measures of location. If all the measurements of a variable have the same value x a location measure should have the value x, too (see Klein (1984), p. 136).

2.3 Idempotence and the axiomatization of location

There are several systems of conditions (axioms) to characterize the specific properties of a statistical functional of location (see f.e. Bickel & Lehmann (1975), Oja (1981), Dabrowska (1985)). All these approaches are very similar.

The axiomatization of Bickel & Lehmann starts with two special relations on $\mathcal{D}$: The first binary relation is the well–known stochastical ordering $\preceq$. If $F, G \in \mathcal{D}$ F will be called stochastically not larger than G (shortly $F \preceq G$) if $F(x) \geq G(x)$ for all $x \in \mathbb{R}$.

The second relation $R \subseteq \mathcal{D} \times \mathcal{D}$ also is binary. Let X be the random variable with distributon functions F. We say $(F, G) \in R$ if G is the distribution function of the random variable $-X$.

Due to Bickel and Lehmann $T : \mathcal{F} \longrightarrow \mathbb{R}$ is a measure of location, if

(1) $\quad F, G \in \mathcal{F}$ and $F \preceq G \Longrightarrow T(F) \geq T(G)$

(2) $\quad F, G \in \mathcal{F}$ and $(F, G) \in R \Longrightarrow T(F) = -T(G)$

(3) $\quad F, F \circ g^{-1} \in \mathcal{F}$ for $g(x) = a + bx,\ a, b \in \mathbb{R} \Longrightarrow T(F \circ g^{-1}) = g(T(F))$

The third property will be called "equivariance". Bickel and Lehmann (1975), p. 1047 show the independence of these axioms.

For symmetric distributions the symmetry point determines the value of every measure of location (see also Bickel & Lehmann (1975), p. 1047, Theorem 1.1). Because singular distributions are symmetric with the mass point as symmetry point location functionals have to be idempotent (see Bickel & Lehmann (1975), p. 1047, Theorem 1.2).

There a two alternative ways to show the idempotence of functionals. First, we can prove this property in a direct way. Second, we only have to show that a functional is a measure of location in the sense of Bickel & Lehmann.

Obviously, a lot of popular distributional measures are idempotent. The most important is the arithmetic mean $\int_0^1 F^{-1}(p)dp$. This also is a measure of location on the set of distribution functions for which the integral exist. $F^{-1}(.)$ is the well–known quantile function

$$F^{-1}(u) = \inf\{x \in \mathbb{R} | F(x) \geq u\} \text{ for } u \in [0,1].$$

The quantiles $F^{-1}(p)$, $0 < p < 1$ are idempotent but not neccessarily a measure of location in the sense of Bickel & Lehmann. Further idempotent functionals are considered in Klein (1999).

In this paper we will show the consequences of idempotence on such important statistical concepts as sufficiency, completeness and the Rao–Cramér-inequality.

3 Idempotence and sufficiency

An estimator T_n will be called sufficient for a parameter θ, if the conditional distribution of $X_1, \ldots, X_n$

$$f_{X_1,\ldots,X_n|T_n}$$

does not depend on θ (see for example Casella & Berger (1990), p. 247).

Due to the factorization theorem in the case of i.i.d. random variables $X_1, \ldots, X_n$, T_n is sufficient for a parameter of location θ if and only if

$$\prod_{i=1}^{n} f_{X_1}(x_i;\theta) = f_{T_n}(t(x_1,\ldots,x_n);\theta)h_n(x_1,\ldots,x_n)$$

for all $x_1, \ldots, x_n$ belonging to support $D(X_1)^n$ if $X_1, \ldots, X_n$ and $\theta \in \Theta \subseteq \mathbb{R}$. $h_n(.)$ may not depend on θ.

If T_n is idempotent the factorization theorem gives

$$f_{T_n}(x;\theta) = \frac{f_X(x-\theta)^n}{h_n(x,\ldots,x)}$$

for $x \in D(X_1)$ and $\theta \in \Theta$. Instead of $h_n(x,x,\ldots,x)$ we write $h_n(x)$.

The following examples consider some parametric family of distributions for which the sufficient statistic is well–known.

Example 3.1 *Let $X_1, \ldots, X_n$ be independent and identically bernoulli–distributed with parameter p then it is easy to see by the factorization theorem that $\sum_{i=1}^n X_i$ is a sufficient statistic. The 1–1-function $\overline{X}_n$ of $\sum_{i=1}^n X_i$ also is sufficent such that*

$$f_{\overline{X_n}}(x;p) = \frac{p^{nx}(1-p)^{n-nx}}{h_n(x)}$$

for $x \in \{0,1\}$ and $p \in (0,1)$. For $f_{\overline{X}_n}$ to be a probability function it is

$$h_n(x) = \binom{n}{nx}.$$

$h_n(.)$ depends on x.

Example 3.2 *Let $X_1, \ldots, X_n$ be independent and identically distributed with the density*

$$f(x;\lambda) = \frac{1}{\lambda}e^{-\lambda x}$$

for $x > 0$ and $\lambda > 0$. λ is a scale parameter that can be estimated sufficiently by the sample mean $\overline{X}_n$ with density

$$f_{\overline{X}_n}(x;\lambda) = \frac{1/\lambda^n e^{-nx/\lambda}}{h_n(x)}.$$

It is well–known that $\sum_{i=1}^n X_i$ has the density of a gamma–distribution

$$f_{\sum_{i=1}^n X_i}(x;n,\lambda) = \frac{\lambda^{-n}}{\Gamma(n)}x^{n-1}e^{x/\lambda}$$

for $x > 0$. Hence, it is

$$f_{\overline{X}}(x;n,\lambda) = \frac{\lambda^{-n}}{n\Gamma(n)}(nx)^{n-1}e^{-nx/\lambda}$$

such that

$$h_n(x) = \frac{n\Gamma(n)}{(nx)^{n-1}}$$

is a function of x.

Example 3.3 *Let $X_1, \ldots, X_n$ be independent and identically distributed with*

$$f_X(x;\theta) = \frac{1}{\theta}I_{(0,\theta}(x)$$

for $\theta > 0$. The smallest order statistic $X^{(1)}$ is sufficient and idempotent. The corresponding density is

$$f_X(x;\theta)^n/h_(x) = \frac{1}{\theta^n}I_{(0,\theta)}(x)^n/h_n(x).$$

It is well–known that the density is

$$f_{X^{(1)}}(x;\theta) = nf_X(x;\theta)F_X(x;\theta)^{n-1} = n\frac{1}{\theta}\left(\frac{x}{\theta}\right)^{n-1}I_{(0,\theta)}(x)$$

such that $h_n(x) = 1/(nx^{n-1})$ depends on x.

Example 3.4 *Let $X_1, \ldots, X_n$ be independent, identically distributed with density*

$$f_X(x-\theta) = e^{-(x-\theta)}I_{(\theta,\infty)}(x)$$

for $\theta \in \mathbb{R}$. In this case , it is

$$f_X(x-\theta)^n/h_n(x) = e^{-n(x-\theta)}I_{(\theta,\infty)}(x)/h_n(x).$$

If we set $h_n(x) = n$, we get an exponential distribution with parameters θ und $1/n$. The smallest order statistic is an idempotent function of $X_1, \ldots, X_n$ that has this distribution and is sufficient for θ. Notice that $h_n(.)$ is independent of x.

4 Power–n–distribution

The last example is special because we consider a a parametric family of location. In this case the sufficient and idempotent statistic also has a distribution with location parameter. This is due to the fact that $h_n(.)$ is constant.

This result can be generalized to every parametric family of location. Let X be distributed with density $f_X(.-\theta)$ having support $D(X)$. Then we define

$$f_n(x;\theta) := f_X(x-\theta)^n / \int_{D(X)} f_X^n(y-\theta)dy$$

for $x \in D(X)$ and $\theta \in \mathbb{R}$. This only can be a well–defined density if

$$h_n := \int_{D(X)} f_X^n(y-\theta)dy = E(f_X^{n-1}(X-\theta)) < \infty.$$

A sufficient but not very restrictive condition is that $f_X(x;\theta) < \infty$ for all $x \in \mathbb{R}$. Because $Z = X - \theta$ is pivotal, ($D(Z)$ and f_Z do not depend on θ), we get

$$E(f_X^{n-1}(X-\theta)) = E(f_Z^{n-1}(Z))$$

what does not depend on θ. Therefore, f_n is a density with location parameter θ. We write $f_n(.-\theta)$ and call f_n "power–n–density of f_X. If there is a sufficient and idempotent estimator of a location parameter this estimator is distributed according to the power–n–distribution.

Example 4.1 *Ferguson (1962) showed that under some conditions of regularity the only parametric family of location belonging to the exponential family distribution – and therefore admitting a sufficient statistic – is given by*

$$f_X(x-\theta) = |\gamma| \frac{r^r}{\Gamma(r)} \exp\left(-re^{\gamma(x-\theta)} + r\gamma(x-\theta)\right).$$

Special cases are the normal, the exponential– and as a limit case the uniform distribution. This parametric exponential family of location can be derived from an Γ–distribution with parameters r and λ via a logarithmic transformation. Therefore, we call the members of this exponential family of location loggamma–distributions. The parameters r, γ und θ are related by

$$\theta = 1/\gamma \log(r/\lambda).$$

For loggamma–distributions, it is

$$f_X(x-\theta)^n = |\gamma|^n \frac{r^{nr}}{\Gamma(r)^n} \exp\left(-nre^{\gamma(x-\theta)} + nr\gamma(x-\theta)\right).$$

Hence, we have

$$h = \int_{-\infty}^{\infty} f_X(x-\theta)^n dx = |\gamma|^{n-1} n^{nr} \frac{\Gamma(nr)}{\Gamma(r)^n}$$

and the power–n–density is

$$f_n^*(x-\theta) = |\gamma| \frac{(nr)^{nr}}{\Gamma(nr)} \exp\left(-nre^{\gamma(x-\theta)} + nr\gamma(x-\theta)\right).$$

This is a loggamma–distribution with parameters γ, θ and nr.

Power–n–distributions can be derived from every location familiy, not only from families for which a sufficient and idempotent estimator exists. In the following example we consider the t–distribution for which the order statistics are minimal–sufficient (see Cox & Hinkley (1974), p. 34).

Example 4.2 *Let X be distributed according to the t–distribution with parameter v*

$$f_X(x-\theta) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)} \frac{1}{\sqrt{v\pi}} \left(\frac{1}{1+x^2/v}\right)^{v/2+1}$$

for $x \in \mathbb{R}$, $\theta \in \mathbb{R}$ and $v > 0$. Again, we have

$$h_n = \left(\frac{\Gamma((v+1)/2)}{\Gamma(v/2)}\right)^n \left(\frac{1}{\sqrt{v\pi}}\right)^n \frac{\Gamma(v^*/2)}{\Gamma((v^*+1)/2)} \sqrt{v\pi}$$

with $v^* = n(v+1) - 1$ and

$$f_n^*(x-\theta) = \frac{\Gamma((v^*+1)/2)}{\Gamma(v^*/2)} \frac{1}{\sqrt{v\pi}} \left(\frac{1}{1+x^2/v}\right)^{(v^*+1)/2}.$$

If Y_n^* again is the random variable belonging to this distribution then Y_n^* obviously is not t–distributed. Instead, the transformation $Z_n^* = \sqrt{v^*/v}Y_n^*$ gives an t–distribution with $n(v+1)-1$ degrees of freedom. Hence, the distribution of Y_n^* only is closely related to a t–distribution.

5 Completeness of the power–n–distribution

A parametric family of location $\{f_X(x-\theta)|\theta \in \mathbb{R}\}$ will be called complete if for any function $u(X)$ with $E(u(X)) = 0$ for all $\theta \in \mathbb{R}$ there must be $P(g(X) = 0) = 1$ for all $\theta \in \mathbb{R}$ (see f.e. Casella & Berger (1990), p. 260). An estimator is complete if its parametric family is complete.

Now it is easy to see that the completeness of the underlying parametric family of location implies the completeness of the location family of power–n–distributions: Let X be a complete random variable and Let Y_n be the random variable corresponding to the density f_n. Then, we have

$$E(u(Y_n)) = E\left(u(X_1) \cdot \frac{f_X(X)^{n-1}}{K_n}\right).$$

Now set $v(x) = u(x)f_X(x)^{n-1}/h_n$ for $x \in \mathbb{R}$. If $E(v(X)) = 0$ for all $\theta \in \mathbb{R}$ then $P(v(X) = 0) = 1$ for all $\theta \in \mathbb{R}$ because of the completeness of X. It is $f_X(x)^{n-1} > 0$ for $x \in D(X)$ Hence, we have $P(u(X) = 0) = 1$ for all $\theta \in \mathbb{R}$. This shows the completeness of Y_n for all $n \in \mathbb{N}$.

If T_n is a sufficient location estimator and especially idempotent then T_n is complete and sufficient (and therefore minimal sufficient) if the populations distribution is complete.

Example 5.1 *Due to Lehmann & Scheffé ((1950), p. 314) the family of Cauchy–distributions with location parameter θ is complete. Therefore, the family of distributions with density*

$$f_n(x-\theta) = \frac{\Gamma(n)}{\Gamma((2n-1)/2)} \frac{1}{\sqrt{\pi}} \left(\frac{1}{1+x^2}\right)^n.$$

is complete.

Example 5.2 *Due to Patel et al. ((1976), p. 166) the exponential distributions with location parameter are complete. As we have shown the first order statistic has the corresponding power–n–distribution that therefore has to be complete. Hence, the first order statistic is sufficient and complete. Additionally, it is unbiased. Then, due the theorem of Lehmann and Scheffé the first oder statistic is an "uniformly mininum variance unbiased estimator" (UMVUE) for the location parameter.*

6 Cramér–Rao–lower–bound and idempotence

It is well known that under some regularity conditions the variance of an unbiased sufficient statistic T_n attains the Cramér–Rao–lower–bound if there are functions $K(\theta, n)$ and $t(x_1, \dots, x_n)$ with

$$\sum_{i=1}^{n} \frac{\partial \ln f_X(x_i - \theta)}{\partial \delta} = K(\theta, n)(t(x_1, \dots, x_n) - \theta)$$

(see Mood, Graybill & Boes (1974), pp. 315). Hence, we have the unbiased and sufficient statistic $T_n = t(X_1, \dots, X_n)$. If additionally T_n is idempotent, then it is

$$-n \frac{\partial \ln f_X(x - \theta)}{\partial x} = K(\theta, n)(t(x, \dots, x) - \theta).$$

This implies for the populations distribution f_X

$$\ln f_X(x - \theta) = -\frac{K(\theta, n)}{n} \frac{1}{2} (x - \theta)^2,$$

such that

$$f_X(x - \theta) = \frac{1}{2\pi} e^{-1/2(x-\theta)^2}$$

is the density of the normal distribution. This is a very simple proof that under conditions of regularity the normal distribution is the only one having an idempotent, unbiased and sufficient estimator for the parameter of location that attains the Cramér–Rao–lower–bound.

7 Summary

Idempotence is a well–known property of functionals of location. It means that the value of the functional at a singular distribution must be identically to the mass point of this distribution.

First, we explained the role of idempotence in the known axiomizations of location functionals. Then we derived the distribution of idempotent and sufficient statistics. In the special cases of parametric families of location we got the so–called power–n–distributions. Power–n–distributions again are distributions with a parameter of location and can be derived from every location family for which the density is constrained. Additionally we show that the completeness of the populations family insures the completeness of the family of power–n–distributions. And at last, we gave a further, now very easy proof that the normal distribution is the only one for which a idempotent, sufficient and unbiased estimator attains the Cramér–Rao–lower bound.

References

1. Aczél, J. (1990). Merging relative scores: our correct ways and a wrong way. *Journal of Computing and Information* **1**, 1-23.

2. Bickel, P.J. & Lehmann, E.L. (1975). Descriptive statistics for nonparametric models. I. Introduction. II. Location. *Annals of Statistics* **3**, 1038–1069.

3. Casella, G. & Berger, R.L. (1990). *Statistical inference.* Duxberry Press, Belmont.

4. Cox, D.R. & Hinkley, D.V. (1974). *Theoretical Statistics.* Chapman & Hall, London.

5. Dabrowska, D. (1985). Descrptive parameters of location, dispersion and stochastic dependence. *Statistics* **16**, 63–88.

6. Ferguson, T.S. (1962). Location and scale parameters in exponential families of distributions. *Annals of Mathematical Statistics* **33**, 986–1001.

7. Fernholz, L.T. (1983). *Von Mises Calculus for statistical functionals.* Lecture Notes in Statistics **19**. Springer, New York.

8. Fleming, P.J. & Wallace, J.J. (1986). How not to lie with Statistics: the correct way to summarize bechmark results. *Communications of the ACM* **26**. 218–221.

9. Huber, P.J. (1964). Robust estimation of location parameters. *Annals of Mathematical Statistics* **35**, 73–101.

10. Klein, I. (1999). On the axiomatization of location. In preparation.

11. Lehmann, E.L. (1983). *Theory of point estimation.* Wiley & Sons, New York.

12. Lehmann, E.L. & Scheffé, H. (1959). Completeness, similar regions, and unbiased estimation–Part 1. *Sankhya* **10**, 305–340.

13. Mood, A.M., Graybill, F.A. & Boes, D.C. (1974). *Introduction to the theory of statistics.* McGraw–Hill, Auckland.

14. Jureckova, J. & Sen, P.K. (1997). Asymptotic representations and interrelations of robust estimators and their applications. In: Maddala, G.S. & Rao, C.R. (eds.). *Handbook of Statistics 15.* North–Holland, Amsterdam, 467–512.

15. Oja, H. (1981). On Location, scale, skewness and kurtosis of univariate distributions. *Scandinavian Journal of Statistics* **8**, 154–168.

16. Patel, J.K., Kapadia, C.H. & Owen, D.B. (1976). *Handbook of statistical distributions.* Marcel Dekker, New York.

17. Serfling, R.J. (1980). *Approximation theorems of mathematical statistics.* Wiley & Sons, New York.