Median Stable Distributions and Schröder's Equation

Gib Bassett

August 7, 2018

Abstract: Median stable distributions are an extension of traditional (mean) stable distributions. The extension is part of a research program begun many years ago with Roger Koenker. The program seeks generalizations of well-known statistics to new domains.(If least squares is the linear model version of the sample mean, then where are the linear model versions of other statistics, like the quantiles). In the current implementation, the traditional definition of stability (in terms of *sums* of iid random variables) is recast as a condition on the sampling distribution of an estimator. For the traditional (mean) stable distribution, the sample mean's (rescaled) sampling distribution is identical to the distribution of the iid data. Median stable distributions are defined similarly by replacing the sample mean with the sample median. Since the sampling distribution of the median is a functional its stable distribution is the solution to a functional equation. It turns out that this defining functional equation is an instance of a famous equation due to Schröder from 1870. The fame of the equation is due to the way it incorporates iteration of functions, a key feature of what many years later would become dynamic systems analysis. The current paper reviews median stable distributions in light of the connection to Schröder's functional equation.

1 Introduction: Median Stable Distributions

Median stable distributions are an extension of the idea of (mean) stable distributions [1]. The traditional definition of stability (in terms of *sums* of iid random variables) is recast as a condition on the sampling distribution of an estimator.[1] This makes the traditional analysis a special case of the more general problem of the conditions under which an estimator has the same distribution as the data. This extension continues a research program begun many years ago with Roger Koenker where well known statistics are generalized to new domains; [2], [7].

Let $\hat{X} = \hat{X}(F)$ denote the sample mean based on iid data, $X_i(F) = X(F)$ whose common cdf is F. A mean stable distribution is a cdf H so that (after centering at a μ and scaling by a λ) the distribution of the estimator and the data are the same:

$$\lambda(\hat{X}(H) - \mu) \overset{d}{=} X(H) - \mu \tag{1}$$

Familiar examples of mean stable distributions are the normal (with μ as the expected value and λ=$\sqrt{n}$), and the Cauchy (with μ as the median and $\lambda = 1$). Replacing the sample mean with the sample median in (1) defines a median stable distribution.

To fix ideas consider the simplest case of the median with $n = 3$ i.i.d observations. Let $\hat{X} = \hat{X}(F)$ denote the sample median and its sampling cdf by $M(x) = M(x : F) = \Pr[\hat{X} < x]$, which is given by,

$$M(x) = G(F(x)) \tag{2}$$

where the ”G-function” is[2],

$$G(w) = 3w^2 - 2w^3, \quad w \in [0,1]. \tag{3}$$

The G-function is depicted in Figure 1 along with schematics illustrating how G maps

[1] Identifying a stable distribution in terms of an estimator is not standard. The usual presentation, motivated by the sample mean, is in terms of *sums* of i.i.d random variables; for example, [6]. Reinterpreting the standard definition in terms of the sample mean has the advantage of generalization to new contexts.

[2] $\hat{X} < x$ if: (i) two-out-of-three, or (ii) three-out-of-three of the X_i are less than x. Two-out-of-three has probability, $F(x)^2(1-F(x))$, and can occur in three-choose-two equals 3 ways. Three-out-of-three can occur in 1 way, which has probability, $F(x)^3$. So, $M(x) = F(x)^3 + 3F(x)^2(1-F(x)) = 3F(x)^2 - 2F(x)^3 = G(F(x))$.

F to M. (Note that $G(w)$ is the distribution of the sample median when the data is uniformly distributed on $[0,1]$).

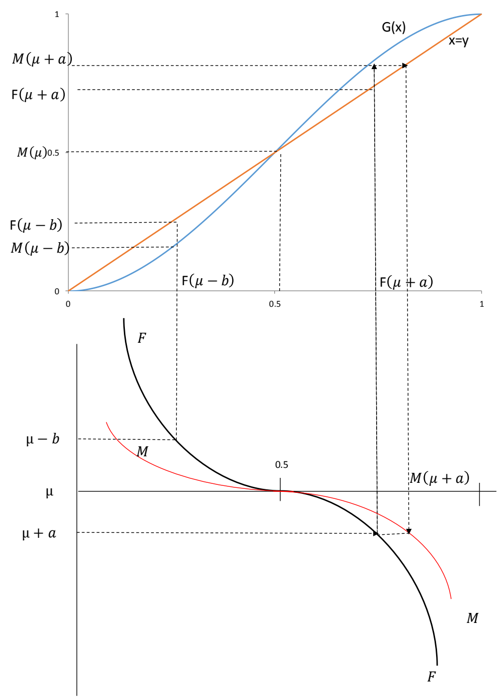

Figure 1: Mapping from F to M, via G

It is readily verified that the median of $G(F(x))$ and $F(x)$ are the same. The sample median is median-unbiased similar to the sample mean being mean-unbiased. Without loss of generality the data is hereafter centered so that either its unique median is 0, or 0 is in the interval of medians, $[\mu^-, \mu^+]$.

A median stable distribution therefore is a cdf H with a λ-scaling factor satisfying (1), or, $H(x) = M[x : H(\lambda^{-1}x)]$, or on substituting for M,

$$H(x) = G(H(\lambda^{-1}x)). \tag{4}$$

The solutions to this functional equation define the $n = 3$ version of median stable distributions. If the data is distributed like H then so too is the median.

Median stable distributions for $n > 3$, $n = 2r + 1$ are defined similarly: substitute G_r for G in (4) where G_r is the distribution of the sample median given iid uniformly distributed data;

$$G_r(w) = C_r \int_0^w g(t)^r dt. \tag{5}$$

where g is the derivative of G_3, and the constant C_r makes $G(1) = 1$; $C_r = 6^{-r} \frac{(2r+1)!}{r!r!}$; see, e.g., [5]. The sampling cdf of the median is $G_r(F(x))$.

Schröder. I recently discovered that the defining functional equation for median stability is an instance of an old and famous functional equation due to Schröder in 1870, [13]. As in our case, the "given" is G, and the Schröder problem is to solve the functional equation (4) for the unknown H.

The fame of the equation is due its central role in what would many years later become complex dynamic systems. It arises because of the neat way that the equation features iteration of functions.

To illustrate, suppose H has an inverse (in our case, quantile) function H^{-1} so that G in (4) expressed in terms of H and H^{-1} is:

$$G(w) = H(\lambda H^{-1}(w)). \tag{6}$$

(Notice that this shows what G looks like for a given H; see Schröder's "Opposite Approach" below).

Composing $G(w)$ twice gives,

$$G^{(2)}(w) = G(G(w)) = H(\lambda H^{-1}(H(\lambda H^{-1}(w)) = H(\lambda^2 H^{-1}(w))$$

and continuing,

$$G^{(k)}(w) = H(\lambda^k H^{-1}(w)). \tag{7}$$

This k-composition works when k is a positive integer. But it also works for any real k thus defining composition for non integer k. For example, $G^{(\frac{1}{2})}(w) = H(\lambda^{\frac{1}{2}} H^{-1}(w))$ (because $G^{(\frac{1}{2})}(G^{(\frac{1}{2})}(w)) = G(w)$). Alternatively: $G^{(\sqrt{2})}(w) = H(\lambda^{\sqrt{2}} H^{-1}(w))$, and $G^{(\sqrt{-1})}(w) = H(\lambda^{\sqrt{-1}} H^{-1}(w))$.

Newton's method for approximating the roots of a polynomial was the problem Schröder was working on that led to his functional equation. A polynomial determines an iterating Newton "G"-function whose fixed point is a root of the polynomial. Under the right conditions on the polynomial and a starting w_0 the iteration $G^{(k)}(w_0)$ converges to a root of the polynomial.

Solving Schröder's equation(4) is not straightforward. In fact, despite introducing the functional equation "Schröder never succeeded in finding methods that guaranteed solutions.... He more or less admitted defeat, settling instead for what one might call the "Opposite Approach", wherein one begins with [H] and λ, and then defining [G] [via (6)], thereby obtaining ready made examples of "pre-solved" Schröder equations"([4], p.214).

It was not until 1884 that Koenigs [9] showed how to get H for a given G. There is typically no closed form solution and H is expressed as the limit of a functional composition operation involving the known G function.[3]

Discussion of Schröder's Equation and median stable distributions is continued in the next section. The equation is used to derive properties of H around zero and in the tails.

Section 4 discusses connections between median stable distributions and the remedian, an estimator defined recursively as the median of medians; see [12]. Many well known features of mean stable distributions have direct extensions to the remedian. This is because the defining recursive feature of the remedian is a property of the sample mean (the mean of means is a mean), and many features of mean stable distributions derive from this recursive property.

The final section departs from the median problem with its G_r and takes up Schröder's "Opposite Approach" in which G-functions are derived from known H distributions. In particular, the G functions corresponding the the Pareto and Laplace distribution are derived.It thus reveals how the distributions are encapsulated in the associated G

[3]Many results are summarized in [15] and [10]. For an historical overview of the role of the functional equation in analyzing complex systems see, [8], and its role in the history of composition operators see, [14]. The canonical problem in the literature has a fixed point of 0 (rather than our fixed point of $1/2$), and the known function corresponds to what is our $G^{-1}(\frac{1}{2}+w)$; see [15], p. 209-10.

function.

2 Schröder's Equation

An (H,λ) solution to (4) inherits features of the G function and will be: increasing, symmetric about 0, with $H(0) = \frac{1}{2}$, convex (concave) for $0 < w < 1/2$ $(1/2 < w < 1)$.

Further, an (H, λ) determines a scale family of distributions. If $(H(x), \lambda)$ is a solution then so is, $H(\sigma x) = G(H(\lambda^{-1}\sigma x))$, $\sigma > 0$.

In addition, an $(H(x), \lambda_0)$ for a given λ_0 determines solutions for all $\lambda > 1$. If $(H(x), \lambda_0)$ solves (4) then so does $(H(x^\alpha), \lambda_0^{\alpha^{-1}})$, $\alpha > 0$.

(Let $H(x|\lambda_0)$ denote the solution with λ_0. Consider $L(x) = H(x^\alpha|\lambda_0)$ so that, $L(x) = G(H(\lambda_0^{-1}x^\alpha)) = G(H((\lambda_0^{-1/\alpha}x)^\alpha)) = G(L(\lambda^{-1/\alpha}x))$, which says, $L(x)$ with $\lambda_0^{-1/\alpha}$ solves the functional equation).

Since a solution for a $\lambda > 1$ determines solutions for all λ we focus on the case, $\lambda = g(\frac{1}{2}) = s$, the slope of $G(w)$ at the fixed point, $w = \frac{1}{2}$.

This particular case of the functional equation with $\lambda = s$ is then:

$$H(x) = G(H(s^{-1}x)). \tag{8}$$

With $\lambda = s$ the density of H at $x = 0$ is 1. (The derivative of (4) is: $h(x) = g(H(\lambda^{-1}x)h(\lambda^{-1}x)\lambda^{-1}$, so $h(0) = g(\frac{1}{2})h(0)\lambda^{-1}$. If $0 < h(0) < \infty$, then $\lambda = g(\frac{1}{2}) = s$.

Schröder Solutions. Repeated substitution of the LHS of (8) into the RHS, gives, $H(x) = G^{(k)}(H(s^{-k}x)) = G^{(k)}(\frac{1}{2} + h(0)s^{-k}x) + o(s^{-k}x))$. Koenigs [9] showed that the solution to (8) (with $h(0) = 1$) is,

$$H(x) = \lim_{k\to\infty} G^{(k)}(\tfrac{1}{2} + s^{-k}x) \tag{9}$$

For an $H(x)$ (with $h(0) = 1$) the associated $G(w)$ with $g(\frac{1}{2}) = s$ is given by,[4]

$$G(w) = H(sH^{-1}(w)) \tag{10}$$

[4]Analogous to the *standard* normal distribution, $h(0) = 1$ defines the *standard* cdf solution to the functional equation.

3 Properties of Median Stable Distributions

Several features of H follow from (9) and (10). The first concerns the density $h(x)$ around zero, which like the normal is proportional to e^{-x^2}.

Theorem 1. *$h_r(x) = \exp(-\alpha(r)x^2 h_r(0)^2 + o(x))$*
where, $\alpha(r) = \frac{-4r}{s^2-1}$.

Proof. The derivative of (8) (suppressing the r-subscript) is $h(sx) = s^{-1}g(H(x))h(x)$, where $g(w) = s[1-4(w-1/2)^2]^r$, so: $\log(h(sx)) - \log(h(x)) = r\log(1-4(H(x)-1/2)^2)$. Dividing by x^2 and taking limits, $x \to 0$ gives, for the RHS:$-4rh(0)^2$, while the LHS is:

$$\lim_{x\to 0} \tfrac{s^2\log(h(x))}{x^2} - \tfrac{\log(h(sx))}{s^2x^2} = (s^2-1)\lim_{x\to 0}\tfrac{\log(h(x))}{x^2}.$$

So,

$$\lim_{x\to 0}\tfrac{\log(h(x))}{x^2} = -\alpha(r)h(0)^2,$$

where $\alpha(r) = \frac{-4r}{s^2-1}$. □

The normal (with density equal to one at the origin) has density, $e^{-\pi x^2}$. The density of the median stable distribution (with density equal to one at the origin) is $e^{-\alpha(r)x^2}$. With $n = 3$ ($r = 1$), the coefficient on x^2 is $\alpha(1) = 3.2$–close to the normal value of π. As $r \to \infty$, $\alpha(r)$ decreases to π.

The next result describes the tail of H. The tail of $H(x)$ depends on $G(w)$ near $w = 0$ (or, near $w = 1$). To motivate, consider the simplest case of $r = 1$, $n = 2r+1 = 3$ where $G(w) = H(sH^{-1}) = 3w^2 - 2w^3$ or

$$\frac{H(sH^{-1})}{3w^2} = 1 - \frac{2}{3}w$$

or changing to $x = H^{-1}$ and taking limits,

$$\lim_{x\to-\infty} H(sx)(3H(x)^2)^{-1} = 1$$

and this means

$$\lim_{x\to-\infty} -|x|^{-\alpha}\log 3H(x) = 1$$

where $s^\alpha = 2$. With $s = \frac{3}{2}$, $\alpha \approx 1.71$, which compares to the Normal tail rate of $\alpha = 2$.

For the general case with $r > 1$ the leading term of $G_r(w)$ is $\frac{(2r+1)!}{r!(r-1)!}w^{r+1}$, and $s_r = 4^r \frac{(2r+1)!}{r!r!}$ so

Theorem 2.

$$\lim_{x \to -\infty} -|x|^{-\alpha} \log cH_r(-x) = 1$$

where, $c = \frac{(2r+1)!}{r!(r-1)!}$, *and* $s_r^\alpha = r + 1$.

Proof. The proof is similar to the above $r = 1$ case and is omitted. □

A final property for H follows from the the well known normal limiting distribution of the median: as $n \to \infty$, the ($\sqrt{n} = \sqrt{2r+1}$-scaled) median given iid uniform data is normal with mean zero and standard deviation $\frac{1}{2}$. In our notation,

$$= \lim_{r \to \infty} G_r(\tfrac{1}{2} + \frac{1}{\sqrt{2r+1}}x) = N(0:0, \tfrac{1}{2}).$$

In terms of our functional equation this means,

$$\lim_{r \to \infty} G_r(N(s_r^{-1}x : 0, \tfrac{1}{2})) = N(x : 0, \tfrac{1}{2})$$

(Verify noting

$$\lim_{n \to \infty} s_n^{-1}\sqrt{n} = \sqrt{\pi/2}$$

and $N((\pi/2)^{1/2}x : 0, \sigma_M) = N(x : 0, \frac{1}{2}))$. Hence as $n = 2r + 1 \to \infty$ the median stable distributions approach the normal.

4 Remedian Stable Distributions

Mean stable distributions do not depend on the number of observations. The normal, Cauchy, or any other mean stable distribution is stable for any n. The number of observations only determines the appropriate λ_n scaling factor; for the normal, $\lambda_n = \sqrt{n}$; for the Cauchy, $\lambda_n = 1$. Mean stable distributions also have a domain-of-attraction feature concerning the limiting distribution of the sample mean. Since mean stability holds for all n, the limiting distribution of the sample mean with stable data will be the same stable distribution. But even with nonstable data the limiting distribution will be a stable distribution. The best known example is the normal. With normal data, the limiting distribution of the sample mean is normal. But with non-normal data (and finite variance) the limiting distribution will be normal. (If the variance is not finite, the limit will be a different stable distribution). Mean stable distributions not only make the sample mean reproduce the distribution of the data, but they are also the limiting distributions of the sample mean when data is not stable.

Mean stable distributions are the same for all n, and they are the limit distributions of the sample mean. These features of mean stability are a consequence of the recursiveness of the sample mean; the mean of a collection of means is a mean. This recursive feature does not work for the sample median-the median of a collection of medians is not a median. As a result, and in contrast to mean stability, median stable distributions are n dependent, and the domain of attraction property does not hold for the usual median.

While the properties do not hold for the median, they do hold for a different estimator, the remedian. The remedian and mean share the recursive property; the remedian is defined as the median of medians; see [12]. Its base-r version-(with $(2r+1)^k$ observations) is defined recursively as the median of "data", each of which is itself a median, $(\hat{X}_{1\hat{X}_{(k-1)}}, \hat{X}_{2\hat{X}_{(k-1)}}, ..., \hat{X}_{(2r+1)\hat{X}_{(k-1)}})$. Because of this recursion, the sampling distribution is the k^{th} iterate of the median G_r function; the sampling distribution of the base $2r+1$ remedian with $n=(2r+1)^k$ observations is, $G_r^{(k)}(F(x))$, [12].

Consider the scaled-by s_r^k-distribution of the remedian; $G_r^{(k)}(F(s_r^{-k}x))$. If the data has median stable distribution, H_r, then the distribution of the remedian is the same as the data, $G_r^{(k)}(H_r(s_r^{-k}x)) = H_r(k)$. In other words, the base r remedian has a remedian stable distribution that is the same as the median stable distribution H_r. Since the remedian, like the mean, is recursive, this remedian stable distribution H_r is remedian stable for all $n=(2r+1)^k$, $k=1,2,....$

Further, while the domain of attraction feature fails for the median, it works for the remedian. Let the data have distribution, $F(x)$, where, $f(x)>0$ in a neighborhood of zero. The distribution of the remedian is, $G_r^{(k)}(F(s_r^{-k}x))$, or $G_r^{(k)}(\frac{1}{2}+f(\omega s_r^{-k}x)x)$, for $|\omega|<1$, which as $k\to\infty$ is $H_r(f(0)w)$, the median stable distribution with scale, $1/f(0)$.

(More generally, let $F(x)=\frac{1}{2}+c(x)x^{\alpha}+o(x)$, where $c(x)>0$ in a neighborhood of 0. Then, as in the above, the limiting distribution of the remedian is, $H_r(c(0)x^{\alpha})$).

Thus, like mean stability, the limiting distributions of the sample remedian are the remedian stable distributions.

5 Pareto and Laplace G functions

This section presents two examples of presolved functional equations. It is reminiscent of Schröder's opposite approach approach in which, being unable to solve for H in terms of G, he starts with an H and derives its G. The first example is the Pareto distribution for which the resulting G is piecewise linear. The second example has H as the Laplace distribution and derives its associated G. It is interesting to see how the properties of the Laplace and Pareto are encapsulted in their associated G functions.

Pareto. Consider the polyhedral G function,

$$\begin{aligned} G(\tfrac{1}{2}+w) &= \tfrac{1}{2}+sw & 0 \leq w < w_0 < \tfrac{1}{2} \\ &= \tfrac{1}{2}+sw_o+b(w-w_o) & w_o \leq w \leq \tfrac{1}{2} \end{aligned}$$

where $s > 1$, $b = \frac{1-2sw_o}{1-2w_o}$, and $G(\frac{1}{2}-w)$ is given by symmetry.

The associated $H(x)$ is a polyhedral–linear in-between the kinks–cdf whose value at the kink, $x_i = s^i w_o$, is:

$$H(s^i w_o) = G^{(i)}(\tfrac{1}{2}+w_o) = \tfrac{1}{2}+sw_o+(s-1)w_o\tfrac{b-b^i}{1-b}$$

(Verify using (9), and $H_k(x) = G^{(k)}(\frac{1}{2}+s^{-k}x)$, where $H_k(s^i w_o) = G^{(i)}(\frac{1}{2}+w_o) = \frac{1}{2}+sw_o+(s-1)w_o\frac{b-b^i}{1-b}$). Evaluating the density/slope of $H(x)$ gives

$$h(x) = (\frac{b}{s})^i$$

for, $s^i w_o < x < s^{i+1} w_o$, a linear interpolation of the Pareto distribution; that is, $h(x) = Ax^{-\beta}$, $\beta = 1 - \frac{\log b}{\log s}$.

Laplace. Let $H(x)$ be the standard ($h(0) = 1$) Laplace distribution and $H^{-1}(w)$ the associated inverse/quantile function:

$$\begin{aligned} H(x) &= 1-\tfrac{1}{2}e^{-2x} & x > 0 \\ H^{-1}(w) &= -\tfrac{1}{2}ln[2(1-w)] & w > \tfrac{1}{2} \end{aligned} \tag{11}$$

The distribution for $x < 0$ is given by symmetry, $(H(x) = 1 - H(-x))$, as is the inverse for $w < \frac{1}{2}$; $H^{-1}(w) + H^{-1}(1-w) = 1$. For the associated G-function:

$$\begin{aligned} G(w) &= 1 - \tfrac{1}{2}[2(1-w)^{\lambda}] \qquad \tfrac{1}{2} < w < 1 \\ &= 2^{\lambda-1} w^{\lambda} \qquad\qquad 0 < w < \tfrac{1}{2}. \end{aligned} \tag{12}$$

The G functions for the Pareto and Laplace distributions are compared with each other as well as the median G ($r = 1, n = 3$) in Figure 5. The fat tails of the Pareto are reflected in the slower approach of its G function to 0 and 1. The values near 0 and 1 for the median G are seen to fall between the the Laplace and Pareto thus indicating the tails of the median stable distribution fall between the Laplace and Pareto.

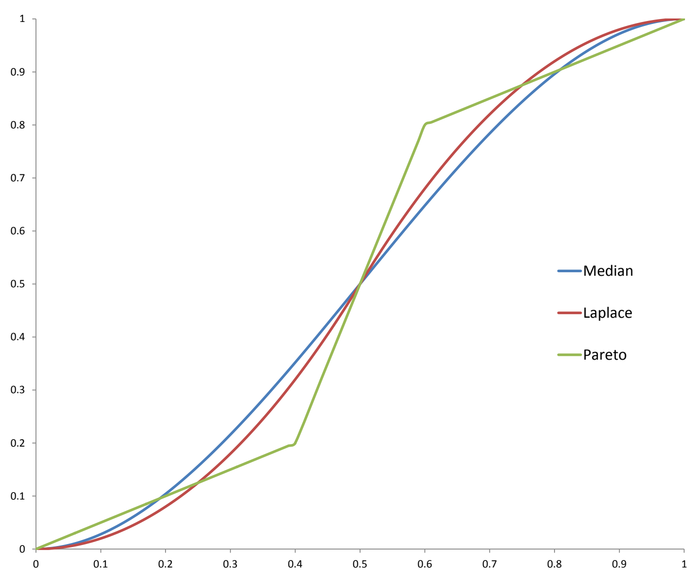

Figure 2: G functions

References

[1] Bassett, G. W. (2016). ”Median Stable Distributions”, in Robust Rank-based and Nonparametric Methods, by Regina Liu and Joseph W. McKean. Springer, (2016),

Chapter 14: 233-244.

[2] Bassett, G.W., Koenker, R.: The Asymptotic Theory of Least Absolute Error Regression. J. Amer. Statist. Assoc. **73**, 618-622 (1978).

[3] Carleson, Lennart; Gamelin, Theodore W. (1993). Complex Dynamics. Textbook series: Universitext: Tracts in Mathematics. New York: Springer-Verlag. ISBN 0-387-97942-5.

[4] Curtright, Thomas, and Cosmas Zachos (2009). "Evolution profiles and functional equations." J. Phys. A: Math. Theor. 42.ANL-HEP-PR-09-80 (2009).

[5] David H. A. (1981). Order Statistics. Wiley, New York (1981).

[6] Feller, W. (1971). An Introduction to Probability Theory and Its Applications. Wiley, New York (1971).

[7] Koenker, R., Bassett, G.W. (1978) Regression Quantiles, Econometrica. 46:33-50.

[8] Gamelin, T. W. (1996). "A History of Complex Dynamics, from Schrder to Fatou and Julia. By Daniel S. Alexander. Wiesbaden (Vieweg)." Historia Mathematica 23.1 (1996): 74-84.

[9] Koenigs, G. (1884). "Recherches sur les intgrales de certaines quations fonctionelles". Annales Scientifiques de l'cole Normale Suprieure 1 (3, Supplment): 341.

[10] Kuczma, Marek (1968). Functional equations in a single variable. Monografie Matematyczne. Warszawa: PWN Polish Scientific Publishers. ASIN: B0006BTAC2

[11] Kuczma, Marek (1964). "Note on Schrder's functional equation." Journal of the Australian Mathematical Society 4.02 (1964): 149-151.

[12] Rousseeuw, P.J., Bassett, G.W. (1990). The Remedian: A Robust Averaging Method for Large Data Sets. J. Amer. Statist. Assoc. **85**, 97-104 (1990).

[13] Schrder, Ernst (1870). ”Ueber iterirte Functionen”. Math. Ann. 3 (2): 296322. doi:10.1007/BF01443992.

[14] Shapiro, Joel H. (1998). ”Composition operators and Schrder’s functional equation.” Contemporary Mathematics 213 (1998): 213-228.

[15] Szekeres, G. (1958). ”Regular iteration of real and complex functions”. Acta Mathematica 100 (34): 361376. doi:10.1007/BF02559539. [1]