Best Data Approximation Science

Ph. D. & Dr. Sc. Lev G. Gelimson (AICFS)

Mathematical Journal of the "Collegium" All World Academy of Sciences, Munich (Germany), 15 (2015), 7.

By classical estimation, approximation, and data processing [1], the common least square method and other ones provide: The worse data, the greater influence. **Best data approximation science** [2-8] reliably ensures for the first time: The better data, the greater influence. Hence also outliers are fully adequately considered.
Supplemented data half-division theory considers (e.g., in the 1D case) ordered data $x_1 \le x_2 \le \dots \le x_n$ with the mean interval length $(x_n - x_1)/(n - 1) > 0$, $n > 1$, and supplements them with $x_0 = -\infty = -\omega$ and $x_{n+1} = +\infty = \omega$ (the countable cardinality). Then the reals set $R = (-\infty, +\infty) = |-\omega, \omega| = \Sigma_{j=0}^{n} |x_j , x_{j+1}| = \Sigma_{j=0}^{n} \{_{1/2}x_j + (x_j , x_{j+1}) + {}_{1/2}x_{j+1}\}$.
Quantile-variance theory includes multidimensional generalization using either distances by data rotation invariance or coordinates otherwise. **The quantile method** uses rationally selected data quantiles, e.g., quartiles $q_{1/4}$, $u = q_{1/2}$ (median), and $q_{3/4}$ (at 0.6745σ from the mode of a normal distribution), then determines left σ_L $(x \le u)$ and right σ_R $(x \ge u)$ standard deviations σ: $\sigma_L = (q_{1/2} - q_{1/4})/0.6745$, $\sigma_R = (q_{3/4} - q_{1/2})/0.6745$. **The variance method** directly determines (about any u, e.g., mean x_m or median $q_{1/2}$) $\sigma^2 = \Sigma_{j=1}^{n} (x_j - u)^2/n$, $\sigma_L^2 = \Sigma_{x(j)\le u} (x_j - u)^2/(n/2)$, and $\sigma_R^2 = \Sigma_{x(j)\ge u} (x_j - u)^2/(n/2)$.
Binormal weight theory naturally weights data via binormal probability density:
$f(x) = (2/\pi)^{1/2}/(\sigma_L + \sigma_R) \exp[-(x_j - u)^2/(2\sigma_L^2)]$, $f(x) = (2/\pi)^{1/2}/(\sigma_L + \sigma_R) \exp[-(x_j - u)^2/(2\sigma_R^2)]$.
Variable-variance weight theory uses c with sign $c = \mathrm{sign}(\sigma_R - \sigma) = \mathrm{sign}(\sigma - \sigma_L)$ and
$\sigma_L(x) = \sigma + 2(\sigma_L-\sigma)/\pi \arctan[c(x-u)/(\sigma_L-\sigma)]$, $\sigma_R(x) = \sigma + 2(\sigma_R-\sigma)/\pi \arctan[c(x-u)/(\sigma_R-\sigma)]$.
Local weight theory generally weights any non-unimodal distributions $(a > 0, b > 0)$:
$x_G = \Sigma_{j=0}^{n} 0.5(x_j + x_{j+1})\exp\{-a[(n-1)(x_{j+1} - x_j)/(x_n - x_1)]^b\}/\Sigma_{j=0}^{n} \exp\{-a[(n-1)(x_{j+1} - x_j)/(x_n - x_1)]^b\}$ $= \Sigma_{j=1}^{n-1} 0.5(x_j + x_{j+1})\exp\{-a[(n-1)(x_{j+1} - x_j)/(x_n - x_1)]^b\}/\Sigma_{j=1}^{n-1} \exp\{-a[(n-1)(x_{j+1} - x_j)/(x_n - x_1)]^b\}$.
For $n = 2$, $x_G = (x_1 + x_2)/2$. For $x_1 < x_2 = \dots = x_n$ and $a = \ln(n/2)/(n - 1)^b$, $x_G = [x_1 + (n -1)^2 x_2]/[1 + (n-1)^2]$. For $b = 2$, $a = \ln(n/2)/(n - 1)^2$. See 1D & 2D data processing, Figs. 1, 2:

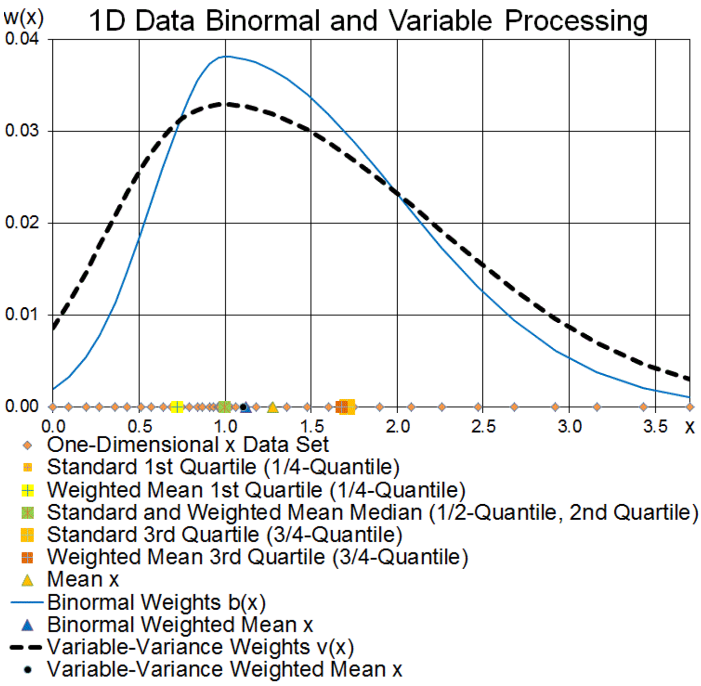

Figure 1. One-dimensional data processing

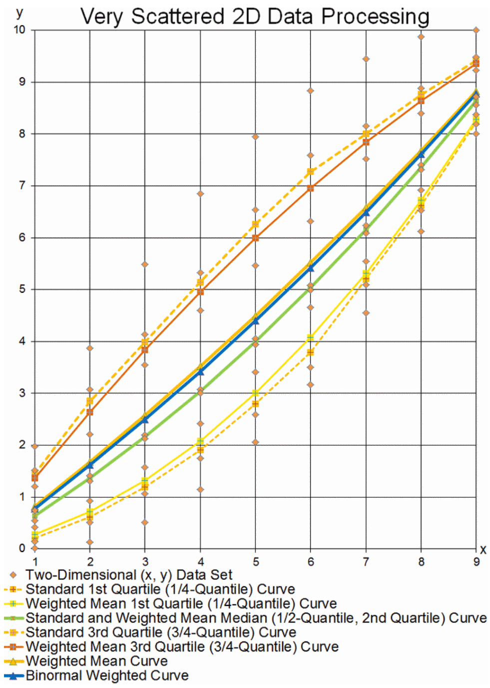

Figure 2. 2D data processing

In Fig. 1, mean $x_m = 1.28$, $u = q_{1/2} = 1$, $u_{binormal} = 1.1229$, $u_{variable} = 1.1057$, $u_{local} = 1.0652$.

Very asymmetric/scattered data also in aeronautical fatigue are adequately fitted.

Acknowledgements to Anatolij Gelimson for our constructive discussions on coordinate system transformation invariances and his very useful remarks.

Keywords: Ph. D. & Dr. Sc. Lev Gelimson, "Collegium" All World Academy of Sciences, Academic Institute for Creating Fundamental Sciences, Mathematical Journal, Universal Mathematics, perfectly sensitive unimathematics, Unimathematik, Best Data Approximation Science, classical estimation, approximation, data processing, least square method, outlier overinfluence, best data basis, supplemented data half-division theory, ordered data, mean interval length, real number set, quantile-variance theory, multidimensional generalization, distance, data rotation invariance, coordinate, quantile method, rationally selected data quantile, quartile, median, mode, normal distribution), one-side standard deviation, variance method, binormal weight theory, natural weight, binormal probability density, variable-variance weight theory, local weight theory, general weight, non-unimodal distribution, one-dimensional data processing, 2D data processing, very asymmetric scattered data, aeronautical fatigue, adequate fitting, coordinate system transformation invariance, Quantianalysis, Uninumber, Quantioperation, Quantiset, Multiquantity, Uniquantity, aeronautical fatigue, General Problem Theory, Elastic Mathematics, General Strength Theory, Universal Probabilistic Science, Universal Statistical Science, Universal Data Processing Science with Multiple-Sources Intelligent Iteration.

[1] Encyclopaedia of Mathematics. Ed. Michiel Hazewinkel. Volumes 1 to 10. Supplements I to III. Kluwer Academic Publ., Dordrecht, 1987-2002
[2] Lev Gelimson. Basic New Mathematics. Drukar Publishers, Sumy, 1995
[3] Lev Gelimson. Quantianalysis: Uninumbers, Quantioperations, Quantisets, and Multiquantities (now Uniquantities). Abhandlungen der WIGB (Wissenschaftlichen Gesellschaft zu Berlin), 3 (2003), Berlin, 15-21
[4] Lev Gelimson. General Problem Theory. Abhandlungen der WIGB (Wissenschaftlichen Gesellschaft zu Berlin), 3 (2003), Berlin, 26-32
[5] Lev Gelimson. Elastic Mathematics. General Strength Theory. The "Collegium" All World Academy of Sciences Publishers, Munich, 2004
[6] Lev Gelimson. Providing Helicopter Fatigue Strength: Flight Conditions. In: Structural Integrity of Advanced Aircraft and Life Extension for Current Fleets – Lessons Learned in 50 Years After the Comet Accidents, Proceedings of the 23rd ICAF Symposium, Claudio Dalle Donne (Ed.), 2005, Hamburg, Vol. II, 405-416
[7] Lev Gelimson. Universal Probabilistic Science. In: Review of Aeronautical Fatigue Investigations in Germany during the Period April 2011 to April 2013, Ed. Dr. Claudio Dalle Donne, EADS Innovation Works, CTO/IW-MS-2013-069, International Committee on Aeronautical Fatigue, ICAF 2013, 2013, 30-32
[8] Lev Gelimson. Universal Data Processing Science with Multiple-Sources Intelligent Iteration. In: Review of Aeronautical Fatigue Investigations in Germany during the Period April 2011 to April 2013, Ed. Dr. Claudio Dalle Donne, EADS Innovation Works, CTO/IW-MS-2013-069, International Committee on Aeronautical Fatigue, ICAF 2013, 2013, 34-35