What Can We Learn from *Correct* Calculation of Histograms? (*Revised Title*)

James S. Weber, Ph.D.[1]
[1] Recent employers: University of Illinois Chicago; University of Wisconsin Milwaukee.

Revised Abstract. Defining "good" histograms often has focused on asymptotic results about bin width, and various integrated measures of error, among them mean integrated squared error, "MISE." MISE leads to a statistical objective function to optimize with data leading to a "best" histogram density and equivalent frequency histogram. MISE optimal histograms have been approximations, not exactly correct MISE histograms.

Now an exact calculating procedure using histogram shape level sets shows that bin edge discontinuity significantly interferes with achieving useful approximations. Exact MISE and other kinds of good histograms can be obtained so that error in approximated MISE histograms for small to moderate samples size can be assessed.

Apparently little thought has been given to exact calculation of histograms, for example, MISE minimizing histograms, because histograms often are so easy to construct. However exact calculations show that *reasonable approximations* are very far from exact, even from the perspective of just shape. If there is any interest in histograms that really correspond to various statistical criteria, then, exact - *correct* - procedures should be used for calculating histograms that are optimal for various statistical criteria such as MISE, maximum likelihood, method of moments, shape stability, other least squares and some model parameter estimates using uniformly grouped data, for example min chi-squared estimates of normal parameters. (K. Smith, 1916.)

1. Introduction

Histograms are widely used for teaching, exploratory data analysis, looking at residuals, and reports of all kinds. But what is a *good* histogram?

For *roughly* the last 66 years the focus in finding "good" histograms has been on bin width. D. W. Scott (1992) and several of Scott's cited sources explain why. In this environment, although not a bin width rule, (actually a bin width *and location* rule), MISE histograms often have often been considered. Also maximum likelihood, ML, has been considered, although not extensively and not in comparison with MISE.

" 'Exact' Calculation…" in the title of my JSM 2016 contributed paper may not have been the best word, although it is not far off the mark: Replace "Exact" with "*Correct*." Many decimal places in computer calculations suggest that calculated numbers usually are connected sufficiently to initial values to be considered *exact**. Reasonable, sensible histogram calculation procedures lead to results that are exact in this sense and expected to be *good-enough approximations*. But in fact they are not *good enough* because of bin edge discontinuity and arbitrary restrictions of domains of optimization that exclude true optima. (*Ensuring that bin edges do not equal data values ensures this kind of exactness.)

1.1 *Whatever the question*? …Exact..? or …Correct..? – *What are the answers*?

- In the absence of a maximum bin width restriction, MISE can over smooth – sometimes only one bin
- ML for prescribed max number of bins seems to work pretty well, *does not over smooth*
- For some data, MISE and ML histogram *densities* are identical – densities, not just shape, and global, not local
- (Almost all shape-local MISE and ML histogram densities are identical. Rank orders differ, so global optimized MISE and ML densities often, *but not always, differ*.)
- Half-open bin structure and bin edge discontinuity matter. Minimum bin width does not exist. Focus on infimum bin width, inf MISE obj ftn value, sup ML.
- Can't omit optimal points from domains of optimization. Current practice does.
- Other answers not included in JSM 2016 presentation include elegant, *exact* shape stability analysis of bin widths for a range of bin width values, not approximate grid search and polynomial model of stability, as presented by Simonoff, J. and Udina, 1997.
- Many kinds of *good* histograms can be calculated exactly via shape level sets: **i** MISE, **ii** ML, **iii** Method of moments, **iv** shape stability. Uniform bin width histogram shape reversal relates to data symmetry. Mode inversion can result from different uniform bin width and location. (Table 2b)
- **Do other *data aggregating graphic* optimizations overlook boundary discontinuity similar to bin edge discontinuity? Are there other arbitrary restrictions of optimization domains? How should data aggregating graphics be compared and correctness assessed?**

1.2 Three red flag DEFICIENCIES in current practice:

1. Search domain for $h \leq$ range of data excludes some optima.
2. Bin edges $\neq$ any data values, so some optima are not found.
3. All shapes not considered.

Too often these lead to approximations that are not even close to exact correct values.

1.3 MISE and ML objective functions.

Histogram shape: Consider data x_i, uniform bin width, h, location anchor point, t_o ($x_{min} - h < t_o \leq x_{min}$) histogram half-open bins, $[t_o + (k-1)h, t_o + kh)$, $k = 1$ to $\boldsymbol{K}$, with $x_{max} < t_o + \boldsymbol{K}h$. The shape of the resulting histogram is the list of bin counts, ν_k, $k = 1$ to $\boldsymbol{K}$, beginning with the first positive bin count, ν_1, and ending with $\nu_{\boldsymbol{K}}$. For a shape, ν_k, $k = 1$ to $\boldsymbol{K}$, and bin width, h, an MISE histogram objective function is given by (1), Scott (1992), p. 77, (3.52), as well as others.

$$\textbf{UCV}(\boldsymbol{h}) \equiv \text{MISE} \qquad (\mathbf{1})$$

$$2/(n-1)h - [(n+1)/n^2(n-1)h]\sum_{k=1}^{K} \nu_k{}^2 = (2/(n-1) - [(n+1)/n^2(n-1)]\sum_{k=1}^{K} \nu_k{}^2)/h$$

Almost always (1) *is negative* and **UCV**(h) would be *minimized by minimum bin width* if minimum bin width existed. Due to bin edge discontinuity and right open bins, for all shapes, minimum bin width, h_{min}, does not exist. So (1) is infimized by infimum h, h_{inf}. (See Appendix **B.2**.) (Sometimes it is helpful explicitly to emphasize all dependencies in (1) by UCV(t_o, h; x_i) and $\nu_k(t_o, h; x_i)$.)

Elementary likelihood, (3), is the familiar product of density evaluations, (2), at each sample point. In (2) $\nu_{k[x_i)}$ is the bin count for the bin that contains x_i. For some bins, $\nu_k^{\nu k} = 0^0 \equiv 1$.

$$f(x_i; t_o, h) \equiv \nu_{k[x_i)} / nh \quad \textbf{(2)}$$

$$\boldsymbol{L}(x_i; t_o, h) \equiv \prod_{i=1}^{n} f(x_i; t_o, h) = \prod_{i=1}^{n} \frac{\nu_{k[x_i)}}{nh} = (\frac{1}{nh})^n \prod_{i=1}^{n} \nu_{k[x_i)} = (\frac{1}{nh})^n \prod_{k=1}^{K} (\nu_k)^{\nu_k}, \ \nu_k \geq 0 \ \textbf{(3)}$$

For a shape, likelihood, (3), is positive. Like MISE, $\boldsymbol{L}(x_i; t_o, h)$ would be optimized (*maximized*) *for each shape by minimum bin width, if minimum h existed*. Instead, infimum h, h_{inf}, leads to supremum $\boldsymbol{L}(x_i; t_o, h)$.

Notice that infimum bin width for a shape is achieved for uniform width bin edges that equal at least two data values. To see this, note that if bin edges equal at most one data value, then there exist smaller width uniform width bins leading to the same bin counts. Current practice often excludes or fails to include explicitly bin edges that equal any data values. Hence optimizing values are missing from the domain of optimization. And this has consequences: Reasonable approximation methods that often work don't work here and are not even close.

1.4 Estimation error, Calculation error

Before explaining exact – correct - calculation via shape level sets, we give examples to motivate pursuing exact correct calculations. First we review ideas about error and how to compare approximations of optimal MISE (and ML) histograms with exact optimum infimum for (1) (and supremum for (3)).

To be very clear, first distinguish between *estimation error* and *calculation error*. Estimation error relates to estimator variability and associated or assumed distribution, or some understanding of estimator variability and connection to a sampling distribution. Estimation error can lead to estimators, such as MISE histograms, for example.

In contrast, calculation error often is assumed, apparently, to be insignificant on account of many decimal places of computer accuracy, or intractable, or one way or another not worth considering *for calculating histograms*. A remainder term or some numerical analysis handle on *calculating* error is missing. However now, using uniform bin width histogram shape level sets, MISE, ML, MOM and shape stability histograms can be calculated correctly, exactly. (I.e. Correct calculations should be done exactly, whereas exact calculations are not necessarily correct.)

1.5 Assessing calculation error via shape and density rank

Approximate histogram density and frequency histograms may be compared with exact by ranking histogram densities via values for objective functions for MISE, (1), and maximum likelihood (3). Consider a list of histogram density shapes, bin location, width values, and data: (ν_k^q/nh^q), (t_o^q, h^q), $q = 1$ to Q, x_i, $i = 1$ to n. Calculate objective function values $(1)^q$, $(3)^q$ and then sort histogram densities ν_k^q/nh^q, according to $(1)^q$, $(3)^q$. For MISE smaller $(1)^q$, more negative is better; for ML larger $(3)^q$, more positive is better.

Each *shape* corresponds to many bin location and width values, namely, a shape level set. Since (1), (3) are strictly monotone in h, they lead to optimal density limiting values, (t_o^{hinf}, h_{inf}), for each shape. Infimum bin width, h_{inf} is associated with unique location, t_o^{hinf}.

Histogram approximation error can be approached primitively, simply as the number of *shapes* ranked ahead of an approximate density, leading to the exact optimal histogram density and shape. There are no assumptions or knowledge about distributions of approximation rank error. *Error* is shown as the number of shapes by which an approximation differs from exactly correctly determined estimates, ranked by a statistical objective function value.

Rank each *shape* according to its best limiting value for (1) and (3). This leads to *shape rank*. The shape rank of a histogram density approximation is at least as good and almost always better than the *density rank* because MISE (or ML) histogram approximations almost never are the infimum MISE value (or sup ML value) because bin edges rarely equal any data values. For an approximate MISE histogram, since the optimization domain often is truncated arbitrarily (by excluding bin edges that equal data values), the optimal objective function that corresponds to only the shape will be better than the objective function value for the shape-sub-optimal histogram density. That is, bin width for a shape MISE sub-optimal density will be greater than the infimum bin width for the shape.

MISE *shape rank* of a histogram density is the rank of the infimum MISE value for the shape, not the rank of the MISE value. The MISE density rank of a density is simply the rank among all MISE shape infimum MISE objective function values. The MISE histogram for data, x_i, is the density for the shape with minimum infimum MISE value (1) among all uniform bin width optimal histogram MISE values for various shapes.

Tables **1abc, 2ab** show MISE and ML densities as equivalent frequency histograms, (ν_k) instead of density histograms (ν_k/nh). Often this is done, is current practice.

Tables **1abc** focus on a sample having identical MISE and ML histogram densities. (Note that almost all *shape-local* MISE and ML histogram *densities* are *identical*.)

Tables 2ab show two one-bin MISE histograms. Beginning a search with the bin width at most ½ the data range is a restriction of the optimization domain that can lead to significantly sub-optimal MISE histograms and mislead regarding over smoothing for MISE histograms. (Note that one bin histograms are the easiest situations for understanding that minimum bin width does not exist, only infimum.)

Tables 1abc, 2ab are *easy* to verify and challenge claimed superiority of MISE over ML histograms.

Table 1a. Identical MISE and ML uniform bin width densities
Shape comparisons and rankings for Rice Stats website MISE (1)
Scott et al (circa 2000); Shimazaki MISE (2007); Exact MISE (1) & Exact ML (3)

1. Silica *sample*, Scott 1992 p 279 ***n* = 22**

Histogram Calculator	Histogram Shape	MISE *Shape* Rank	MISE *Density* Rank
A. Rice… MISE approx **(1)[12]**	**(4,0,7,2,8,1)**	**14**	**26**
B. Shimazaki approx **MISE[15]**	**(3,1,1,6,2,1,8)**	**3**	**≥ 3**
C. Exact MISE (1)	**(1,3,0,6,3,1,8)**	**1**	**1**
D. Exact UBW ML (3)	**(1,3,0,6,3,1,8)**	**1**	**1**

Scott 1992, [13] p 279, *n* = 22 "**B.7 SILICA DATA Percentage of silica in 22 chondrites meteors.**" 20.77 22.56 22.71 22.99 **26.39** 27.08 27.32 27.33 27.57 27.81 28.69 29.36 30.25 31.89 32.88 33.23 33.28 33.40 33.52 33.83 33.95 **34.82**
Source Scott, D. W. 1992, from Ahrens (1965) and Good and Gaskins (1980).
(**Bold underlined** values are also bin edges for exact MISE and ML bins.)

Table 1b. Approximate and Exact MISE
Bin counts, bin parameter values, and MISE Obj Ftn Values

	Shape - bin counts - $\mathbf{v_k}$	t_o	h	MISE Obj ftn* val
A. Rice… MISE	(4,0,7,2,8,1)	14.8734	2.807	-0.0741
B. Shimazaki MISE	(3,1,1,6,2,1,8)	20.77	2.00714	-0.08333
C&D *Exact MISE*(&ML)	(1,3,0,6,3,1,8)	**20.0675**	2.1075	-0.08366

Note. The objective function (1) for Rice…, [12] & Appendix C is (3.52), p 77, Scott 1992. Rice… MISE shape, $\mathbf{v_k}$, (t_o, h) bin parameter values and objective function values are from Appendix C, that shows a Rice Stats website display. For Shimazaki, objective function (1) Is evaluated from Shimazaki website values for $\mathbf{v_k}$; (t_o, h), and (1), *not* an expression in [15]. Silica Data

Table 1c. Histogram *shapes* that rank ahead of Rice Stats website MISE shape among 435 shapes of at most seven bins for the Silica data.

Row		Shape	MISE Obj ftn(1)	Shape % #1obj ftn value				
C, D	1.	**(1, 3, 0, 6, 3, 1, 8)**	**-0.0837**	**100.00%**	**EXACT**	**MISE(1)**	**& M L (3)**	**ML #1**
	2.	(4, 0, 7, 2, 9)	-0.08334	99.62%				ML #5
B	3.	(3, 1, 1, 6, 2, 1, 8)	-0.08333	99.61%	**Shimazaki**	**MISE [15]**		ML #4
	4.	(1, 3, 1, 7, 1, 8, 1)	-0.08292	99.12%				ML #16
	5.	(1, 3, 1, 6, 2, 1, 8)	-0.08185	97.85%				ML #7
	6.	(4, 0, 8, 2, 8)	-0.0807	96.46%				ML #8
	7.	(2, 2, 1, 6, 2, 1,8)*	-0.08063	96.38%				ML#10*
	8.	(4, 0, 6, 3, 1, 8)	-0.08012	95.77%				ML #2
	9.	(4, 18)	-0.07997	95.59%				ML > #21
	~~10*.~~	~~(2, 2, 1, 6, 2, 1,8)*~~	~~-0.07964~~	~~95.20%~~				~~ML#12*~~
	10.	(5, 17)	-0.0795	95.03%				ML > #21
	11.	(5,16,1)	-0.07902	94.46%				ML > #21
	12.	(4,17,1)	-0.079	94.43%				ML > #21
	13.	(4, 0, 6, 2, 2, 8)	-0.07899	94.42%				ML #3
A	**14.**	**(4, 0, 7, 2, 8, 1)**	-0.07878	**94.17%**	**Rice…**	**MISE(1)**	**[12]**	ML#11

(* #10* same as #7; from a level set *edge*. Level set *edge* min *h* > level set min *h*.)

Table 2a. One bin MISE example 1

Shape comparisons and rankings for Rice Stats website MISE (1)
Scott et al (circa 2000); Shimazaki MISE (2007); Exact MISE (1) & Exact ML (3)

Weber 2008 Data #2 **n = 22**

Histogram Calculator	Histogram Shape	MISE *Shape* Rank	MISE *Density* Rank
A. Rice… MISE apprx (1) [12]	**(4,6,0,6,4)**	**7**	**21**
B. Shimazaki apprx MISE [15]	**(10,10)**	**6**	**≥ 6**
C. Exact MISE (1)	**(20)**	**1**	**1**
D. Exact UBW ML (3)	**(5,5,0,5,5)**	**18**	**18**

Weber 2008 Data #2, n = 20 - exactly symmetric data: **2.05**, 2.27, 2.50, 2.95, 3.18, 3.41, 3.64, 3.86, 4.09, 4.32, 7.73, 7.50, 7.05, 6.82, 6.59, 6.36, 6.14, 5.91, 5.68, **7.95**

Exact MISE bins, **Weber 2008 Data #2**

t_o 2.05
h 5.9

Bin edges: **2.05** **7.95** 13.85

Data Includes x_{min} = **2.05,** **7.95** = x_{max}, h ↔ range for an 1 bin MISE Shape

Table 2b. One bin MISE example 2

Shape comparisons and rankings for Rice Stats website MISE (1)
Scott et al (circa 2000); Shimazaki MISE (2007); Exact MISE (1) & Exact ML (3)

Weber 2008 Data #3 ***n* = 12**

Histogram Calculator	Histogram Shape	MISE *Shape* Rank	MISE *Density* Rank
A. Rice… MISE apprx (1) [12]	**(6,6)**	**6**	**10**
B. Shimazaki apprx MISE [15]	**(6,6)**	**6**	**≥ 6**
C. Exact MISE (1)	**(12)**	**1**	**1**
D. Exact UBW ML (3)	**(3,0,3,1,2,3)**	**53**	**53**

Weber 2008 Data #3, n = 12 - shapes include (1,2,3,3,2,1) and (3,2,1,1,2,3):
0.37, 1.13, 1.23, 2.25, 2.35, 2.45, 3.37, 4.37, 4.47, 5.37, 5.47, **5.61**

Exact MISE bins, **Weber 2008 Data #3**

t_o 0.37
h 5.24

Bin edges: **0.37** **5.61** 10.85

Data Includes x_{min} = **0.37,** **5.61** = x_{max}, h ↔ Range for a 1 bin MISE Shape

2. Uniform Bin Width Histogram Shapes, Shape level sets, inf MISE, sup ML

Obtaining the above examples requires correct calculation of MISE and ML enabled via uniform bin width histogram shape level sets. Uniform bin width histogram shape level sets are the core tool for correct, exact calculation of MISE and maximum likelihood uniform bin width histograms. Also an optimum cannot be expected unless *all* histogram shapes are considered. How can we be sure of considering every shape?

2.1 All possible UBW shapes, Histogram shape level sets, inf MISE, sup ML

Appendix **B.1** describes a compact subset, a domain of optimization, $\boldsymbol{D}_o \subset \{(t_o, h) \mid h > 0\}$. $\boldsymbol{D}_o$ includes subsets* of (t_o, h) values that lead to shapes of one or two bins, and all of the (t_o, h) values for shapes of three and at most $\boldsymbol{K}$ bins. (Compactness is *not* necessary here. Think of compactness here as nothing more than capitalizing the first letter of a sentence. Further, compactness here does *not* ensure the existence of maximum or minimum values for MISE and ML objective functions, which I already have pointed out do not exist.) (*Arbitrarily wide bins are excluded by an upper bound that exceeds the range of the data.)

Convex polygon shape level sets partition $\boldsymbol{D}_o$. Partitioning $\boldsymbol{D}_o$ ensures that every shape is represented by a shape level set. For data, x_i, the list of $\boldsymbol{S}$ possible uniform bin width histogram shapes is determined from selecting an interior point from the s^{th}-shape-level-set, s = 1 to $\boldsymbol{S}$, $(t_o^{int,s}, h^{int,s})$, and calculating the histogram for data x_i with half-open bins $[t_o^{int,s} + (k-1)h^{int,s}, t_o^{int,s} + kh^{int,s})$, k = 1 to $\boldsymbol{K}$.

How can h_{inf} be determined for each shape? The shape level sets are specified by vertices of the convex polygon shape level sets, so it is as simple as selecting the bin width from the vertex with smallest bin width. Since this is an infimum bin width, not minimum, the associated shape is determined from a shape level set interior point, for example the average of the vertices, or average of any three or average of any two non adjacent

vertices. Using h_{inf}, t_o^{hinf}, and bin counts, v_k^s for the shape leads to inf MISE and sup ML objective function values.

2.2 Shape Level Sets

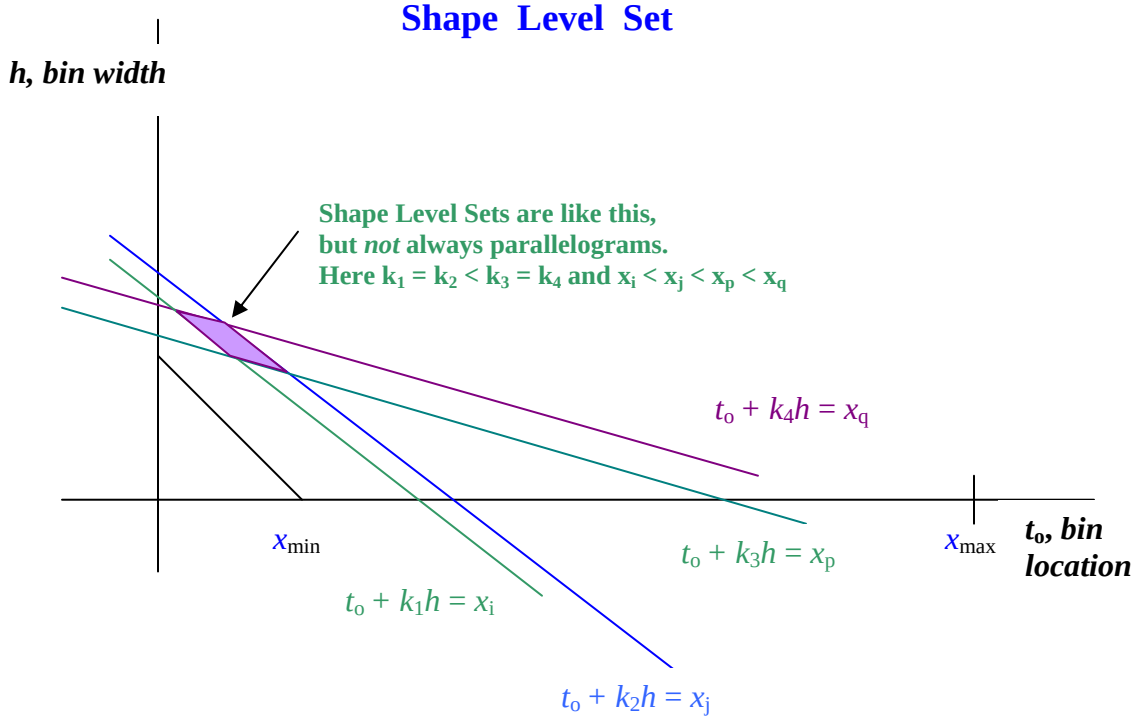
Figure 1, below, illustrates a shape level set. How are shape level sets obtained? Appendix **B.1** describes subset $\boldsymbol{D}_o$ containing (t_o, h) values leading to uniform width bins that put x_{min} in a first bin, **etc**. $\boldsymbol{D}_o$ is partitioned into shape level sets by n **x** $\boldsymbol{K}$ shape level set boundaries, (4ab), for $\boldsymbol{K} \equiv$ a max number of bins, n = sample size.

$$\text{data value} = \text{bin edge} \qquad (4a)$$

$$x_i = t_o + kh,\ k = 1 \text{ to } \boldsymbol{K},\ i = 1 \text{ to } n \qquad (4b)$$

Figure 1

Shape Level Set



Each shape level set *vertex* corresponds to (t_o, h) values leading to at least two bin edges that equal data values.

This leads to convex polygon shape level sets specified by their vertices, $(t_o, h)_{s,v}$, for the s^{th} shape, $s = 1$ to $\boldsymbol{S} \equiv$ number of shapes, and $v = 1$ to $\boldsymbol{V}_s \equiv$ number of vertices for the s^{th} shape. Shapes and level set vertices are organized and presented *lexicographically* sorted on the number, $\boldsymbol{K}_s$, of bins actually required for the data, then bin counts, $v_{s,k}$. Concatenating $\boldsymbol{K}_s$, $v_{s,k}$ with vertices $(t_o, h)_{s,v}$, $v = 1$ to $\boldsymbol{V}_s$, leads to (5).

(5) $$\{(\boldsymbol{K}_s, v_{s,k}, (t_o, h)_{s,v}) \mid k = 1 \text{ to } \boldsymbol{K}_s,\ v = 1 \text{ to } \boldsymbol{V}_s,\ s = 1 \text{ to } \boldsymbol{S}\}$$

$\boldsymbol{S} \equiv$ number of shapes, $\boldsymbol{K} \equiv$ max number of bins

$\boldsymbol{K}_s \equiv$ number of bins for s^{th} shape $\equiv$ index of bin for x_{max}, $\boldsymbol{K}_s \le \boldsymbol{K}$)

$\boldsymbol{V}_s \equiv$ number of vertices for s^{th} shape

Matrix (5) is right ragged $\boldsymbol{S}$ **x** $(1 + \boldsymbol{K}_s + \boldsymbol{V}_s)$, $\boldsymbol{S}$ rows and $(1 + \boldsymbol{K}_s + 2\boldsymbol{V}_s)$ entries in each row.

Summary of Shape Level Set Procedure.
To obtain *all shapes* for a data collection, construct $\boldsymbol{D}_o$, use the lines (4b) to calculate shape level set vertices, $(t_o,h)_{s,v}$, then level set interior points, $(t_o, h)_{s,int}$, then bins $[t_{o,s,int} +(k-1)h_{s,int}, t_{o,s,int} + kh_{s,int})$, and bin counts, $\nu_{s,k}$ for all of the level sets. This determines $\boldsymbol{K}_s$, $\nu_{s,k}$, $(t_o, h)_{s,v}$ for (5). ($(t_o, h)_{s,int}$ is the easiest way to calculate $\nu_{s,k}$ without details about half open bins, *min h, inf h*, etc.)

Lists of all shapes show extreme histogram shape variability and the challenge of defining "good" histograms.

3 Concluding Notes

3.1 Histogram Objects and Methods

Histogram calculations and shape level sets should be viewed in a larger context shown by Table **3**. The main focus has been situation **A** and exact calculation of MISE and ML histograms. Situation **D** is the easy, well known histogram construction. Situations **B** & **C** do not solve any pressing problems but can be used to construct examples or confirm infeasibility of combinations of UBW histogram shapes. For example, Weber 2008 Data #3 for Table 2b was created via method **B**. (That is, there are numbers leading to shapes that include (1,2,3,3,2,1) and (3,2,1,1,2,3). But the linear program leading to uniform histogram shapes that include (1,2,3,4,4,3,2,1) and (4,3,2,1,1,2,3,4) is not feasible.)

Table 3 Histogram objects and exact methods

Known – have	Unknown – want	Method – use
A. data x_i	**All uniform bin width Shapes, v_k**	**A**. **Shape level set algorithm**
B. {shape**s** v_k}	Data x_i;(t_o,h) & bins	**B**. Simplex algorithm
C. data x_i; shapes vk	(t_o, h) & bins	**C**. **Level set** or simplex **algorithm**
D. data x_i; (t_o, h)	Bin frequencies v_k	**D**. Familiar histogram construction

3.2 Conclusion

It is a puzzle that this elementary but embracing analysis did not appear earlier. Upon noticing many distinguished authors cited in Scott 1992, etc., a tempting explanation is simply Little, R. J. (2013) "In Praise of Simplicity not Mathematistry! Ten Simple Powerful Ideas for the Statistical Scientist" Journal of the American Statistical Assoc. 108:502, 359-360. DOI: 10.1080/01621459.2013.787932. (*JASA* Most Read 4/2015.).

This effort began with a simple question: "For data, x_i, what uniform bin width histogram shapes are possible?" That is, what are the possible uniform width bin counts? I expected to look up something, or at least find related prior work. Most references below were examined with the narrow objective of finding the answer to this question or relevant discussion. The usual criterion of using a result does not apply to almost all of these sources. I looked at them and none address the question "What shapes are possible?" My answer to that question naturally lead to "So what?" Exactly calculating many kinds of histograms and discovering that reasonable approximations do not work addresses "So what?" Or as Don Saari might abbreviate: WGAD (Saari, 1995, pp 102, 228.)

Regarding Saari, 1995, this elementary exploration into calculating histograms has been inspired at least a little by the general theory achieved via elementary methods (i.e. simple linear algebra) presented in Saari, 1995. Another general and relatively *transparent* result obtained by elementary mathematics (i.e. calculus) is Weber, J.S. 1991.

Regarding my 2002 voting theory paper, "'How many voters are needed for paradoxes?" Economic Theory, **20**, 41-355 (2002)", situation **B**, Table **3** constructing histogram examples via the simplex algorithm is similar to constructing voter preference profiles via integer programming.

Regarding my 1991 paper, it extends discussion of densities for random variables x such that $f(x)$ has a prescribed density, where f is an arbitrary "nice" function, $\boldsymbol{R^1 \to R^1}$. *Possibly* the first such result is characterization of densities that x can have if $f(x) = x^2$ is *chi-square* 1 *d.f.* The core tool is inverse images of intervals for which f is strictly monotone, $f^{-1}(a, b) = (f^{-1}(a), f^{-1}(b))$ or $(f^{-1}(b), f^{-1}(a))$, or interval inverse images of points where f is constant. If f is constant on an interval, then $a = b$ and the inverse image of a is an interval. Needed along with inverse images are generalized inverse functions: $\boldsymbol{R^1 \to R^1}$: $f(f^{-1}(f(x))) = f(x)$. This leads to generalizations of results of cited papers of S. Geisser (1966, 1973), C. Roberts (1966, 1971); and, overlooked in my 1991 paper, H. W. Block (1975); Funk & Rodine (1975), B. Ramachandran (1975), maybe E. S. Key (1994). My generalizations are non-trivial: **i** a symmetry requirement is eliminated, **ii** gamma families of densities are replaced by arbitrary densities.

Why consider only piecewise differentiable functions that are strictly monotone or constant on intervals? Assigning the uniform density on [0,1] to the absolute value function on [–1, +1] and appealing to a familiar non measurable set found in many introductions to measure theory leads to a non-measurable density (with apologies for the term "non-measurable density"). For this reason focusing on absolute continuity seems arbitrary. Either be as general as possible and include non-measurable densities, *or* keep the discussion as simple as possible and work with piecewise differentiability and the simplest kinds of integrals.

***And the point is*?** The core tool for exact – correct – calculation of histograms is shape level sets. Shape level sets are *inverse images* of uniform bin width shapes – bin counts – in $\{(t_o, h)\}$. Is the use of inverse images to attack problems with calculation of histograms, and the main question I consider in my 1991 paper as far as the similarity goes? Don't know. For now, this is just a noticeable coincidence.

Regarding my 1991 paper, not being distracted by absolute continuity etc. gives a view of transformations of one random variable via contractions, dilations and step functions. That is *more transparent* than (or maybe just different from) functional analysis approaches.

Acknowledgements

Occasional emails, brief discussions with Oscar H. M. Padilla, Albert Madansky, Mark Pinsky, Don Saari, Stan Sclove, Linus Schrage, Stephen Stigler, Richard Stong (*not* Strong), Sandy Zabell, journal editors and anonymous reviewers are gratefully acknowledged, especially since this is not an active research area for these individuals (except possibly anonymous reviewers). Errors, misleading claims, other problems are the responsibility of the author.

References

Almost all references were examined with the narrow objective of finding the answer or thoughts about the question "What uniform bin width histograms shapes are possible for data?" The usual criteria that some result in a paper is used does not apply to most of these sources. A feature of many is that I looked at them and *they do not address the question "What uniform bin width histogram shapes are possible*?"

1 Ahrens, I.S. (1965). "Observations on the Fe-Si-Mg Relationship in Chondrites"*Geochemica et Cosmochimica Acta* 29 801-806.

2 Anonymous Editors. (2006). "How Many Histograms Can a Data Set Have?" Problem Solution 11126, *The American Mathematical Monthly*, **113**, 850. [3]

3 Austin, J. D.; Scott, D. W. (1991). Beyond Histograms: Average Shifted Histograms. *Mathematics and Computer Education*, **25**(1), 42-52. [3]

4 Bargagliotti, A. E; Saari, D. G. (2010) "Symmetry of nonparametric statistical tests on three samples." *Journal of Mathematics and Statistics*, 6, 4, 395-408. [3]

5 Bargagliotti, A. E. (2009) "Aggregating and decision making using ranked data." *Mathematical Social Sciences*, 58, 3, 354-366. [3]

6 Birge, L, Rozenholc, Y. 2002. "How many bins should be put in a regular histogram" Prepublication no. 721, Laboratoire de Probabilities at Modeles Aleatoires, CNRS-UMR 7599, Universite Paris VI & Universite Paris VII, 4 place Jussiue, Case 188, F-75252 Paris Cedex 05. [http://www.proba.jussieu.fr/mathdoc/textes/PMA-721.pdf] [3]

7 Bowman, A. W. (1984) "An Alternative Method of Cross-Validation for the Smoothing of Density Estimates." *Biometrika*, 71, 353-360

8 _____ (1985) "A Comparative Study of Some Kernel-Based Nonparametric Density Estimates," *Journal of Statistical Computation and Simulation*, 21, 313-327.

9 Boyd, S. http://stanford.edu/class/ee364b/lectures/seq_slides.pdf - nonconvex optimization

10 Brown, L. D., and Hwang, J. T. G. (1993) "How to Approximate a Histogram by a Normal Density," The American Statistician, 47:4, 251-255. http://dx.doi.org/10.1080/00031305.1993.10475992 [3]

11 Cooper, L. L.; Shore, F. S. (2008). "Students' Misconceptions in Interpreting Center and Variability of Data Represented via Histograms and Stem-and-Left Plots." Journal of Statistics Education, v16n2 [3]

12 Doane, D. P. (1976), "Aesthetic frequency classifications," *The American Statistician*, **30**, 181-183. [3]

13 Farnsworth,D.L.(2000)"The Case Against Histograms," *Teaching Statistics* **22**,81-85.

14 Fisher, R. A. (1970) *Statistical Methods for Research Workers*. 14th ed., rev. Darien, CO: Hafner 1970. [3]

15 Fisher, R. A. (1916). Correspondence. (See Pearson,E. 1968; Stigler, S.M. 2005.) [3]

16 Good, I.J. and Gaskins, R.A. (1980). "Density Estimation and Bump-Hunting by the Penalized Likelihood Method Exemplified by the Scattering and Meteorite Data (with Discussion)" *JASA* **75** 42 – 73.

17 Graunt, J. (1662) *Natural and Political Observations Made Upon the Bills of Mortality*. Martyr, London.

18 Haunsperger, D. (1992). "Dictionaries of paradoxes for statistical tests on k-samples." *Journal of the American Statistical Association* **87**, 249-272.

19 Haunsperger, D.; Saari, D.G. (1991). "The lack of consistency for statistical decision procedures." The American Statistician **45**, 252-255.

20 Hodges, J. L., and Lehman E. L (1956). "The Efficiency of Some Nonparametric Competitors of the *t*-test" *Ann. Math. Statist.* **27** 324-335. [3]

21 Huff, D. (1954), How To Lie With Statistics. W. W. Norton, New York.

22 Kendall, M.G.; Stuart A. (1969). The advanced theory of statistics, Vol. 1, Distribution theory 3rd ed. C. Griffin, London UK
23 Le-Rademacher, J.; Billard, L. (2011) "Likelihood functions and some maximum likelihood estimators for symbolic data." *Journal of Statistical Planning and Inference*.
24 Lindgren, B.W.(1968). Statistical Theory. 2nd Ed. The Macmillan Co. New York.
25 Little, R. J. (2013) "In Praise of Simplicity not Mathematistry! Ten Simple Powerful Ideas for the Statistical Scientist" Journal of the American Statistical Association 108:502, 359-360. DOI: 10.1080/01621459.2013.787932. (*JASA* Most Read 4/2015.)
26 Majumder, M., Hofmann, H., Cook, D. (2013), "Validation of Visual Statistical Inference, Applied to Linear Models, JASA, 108:503, 942-956, DOI: 10.1080/01621459.2013.808157.
27 Albert Madansky, December 2014: Larry Brown et al *Amer Stat* paper: normal densities that minimize integrated squared error between it and a histogram density.
28 Marron, J. S.; M. P. Wand. (1992), "Exact Mean Integrated Squared Error," The A*nnals of Statistics* 20, 712-736.
29 Moore, D. S.; McCabe, G. P.; Duckworth, W. M.; Alwan, L. C. (2009). *The Practice of Statistics: Using Data for Decisions*. Second Edition. W. H. Freeman and Company, New York.
30 Pearson, E. S. (1968). "Some early correspondence between W. S. Gossett, R. A. Fisher, and Karl Pearson, with notes and comments." *Biometrika* **55** 445-457.
31 Pearson, K. (1894). "Contributions to the Mathematical Theory of Evolution" Philosophical Trans. Royal Society London (A) **185** 71-110.
32 Pearson, K. (1895). "Contributions to the Mathematical Theory of Evolution. II. Skew Variation Homogeneous Material" Philosophical Transactions of the Royal Society A: **186** 343-414. Bibcode:1895RSPTA.186.343P (http://adabs.harvard.edu / abs/1895rspta.186.343P), doi:101098rsta.1895.0010(http;dx.doi.org/10.1098%2Frsta.1895.0010).
33 Rudemo, M. (1982). "Empirical Choice of Histograms and Kernel Density Estimators." *Scandinavian Journal of Statistics* **9**, 65-78.
34 Saari, D. G. (1995), *Basic Geometry of Voting*, Springer-Verlag, Berlin, Heidelberg.
35 Sain, S. R.; Baggerly, K. A.; Scott, D. W.; Scott, W. R. (1992). "Cross-Validation of Multivariate Densities."*Journal of the American Statistical Association* **89**,pp807-817.
36 Scott, D. W. *et al, University of Houston – Clear Lake, Tufts University*. (April 5 2012, January 2014) website (http://onlinestatbook.com/statsim/histogram/index.html)
37 Scott, D. W.; Scott, W. R. (2008). "Smoothed Histograms for Frequency Data on Irregular Intervals." *The American Statistician* **62**, 256-261.
38 Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, Inc. New York
39 Scott, D.W. (1988). "Comment" *JASA*, **83:401** 96-98.
40 Scott, D.W. (1979). "On optimal and data-based histograms,"*Biometrika*, **66** 605-610.
41 Scott, D.W., G. R. Terrell (1987). "Biased and Unbiased Cross-Validation in Density Estimation,"*JASA*, **82:400**, 1131-1146.
42 Sheppard, W. F. (1898). On the calculation of most probable values of frequency-constants, for data arranged according to equidistant divisions of a scale. *Proc. London Math. Soc*. **29**, 353-380.
43 Shimazaki, H.; S. Shinomoto. (http://176.32.89.45/~hideaki/res/histogram.html). (April 6 2012) website. [http://toyoizumilab.brain.riken.jp/hideaki/res/histogram.html]
44 Shimazaki, H.; S. Shinomoto. "A method for selecting the bin size of a time histogram." *Neural Computation* (2007) Vol. 19(7), 1503-1527.

45 Silverman, B. W. (1978). "Choosing the window width when estimating a density,"*Biometrika* **65**, 1-11.
46 Silverman, B. W. (1986). "Density estimation for Statistics and Data Analysis." Chapman Hall.
47 Simonoff, J. S.; F. Udina. (1997) Measuring the Stability of Histogram Appearances When the Anchor Position is Changed. *Computational Statistics and Data Analysis*.
48 Stigler, S. M. 1986. "The History of Statistics: The measurement of uncertainty before 1900." Belknap Press, Harvard University Press, Cambridge, Massachusetts.
49 Stigler, S. M. 1999. *Statistics on the Table: The history of Statistical Concepts and Methods*. Harvard University Press, Cambridge, Massachusetts.
50 Stigler, S. M. 2005a. email re: K. Smith, 1916 and R. A. Fisher/Karl Pearson disputes.
51 Stigler, S. M. 2005b."Fisher in 1921." *Statistical Science* 2005.Vol20,No.1,pp32-49.
52 Stodden, V., D. H. Bailey, J. Borwein, R. J. LeVeque, W. Rider and W. Stein "Setting the Default to Reproducible: Reproducibility in Computational and Experimental Mathematics" (http://www.davidhbailey.com/dhbpapers/icerm-report.pdf) (also Stodden, V., J. Borwein and D. H. Bailey "Setting the Default to Reproducible in Computational Science Research." SIAM News: Volume 46, Number 5, June 2013.)
53 Sturges, H. A. (1926). "The choice of a class interval." *Journal of the American Statistical Association*, **21**, 65-66.
54 Thompson, J. R.; Tapia, R. A. (1990), *Nonparametric Function Estimation, Modeling, and Simulation*. SIAM, Philadelphia.
55 Ward, J. H. (1963) "Hierarchical Grouping to Optimize an Objective Function." *JASA Journal of the American Statistical Association*, 58, 236-244.
56. Weber, J. S. 1991. "Generalized transformations of random variables." Statistics and Probability Letters 12 (1991) pp 161-166. North Holland.
57. Weber, J. S. 2002. "How many voters are needed for paradoxes?" Economic Theory, **20**, 341-355 (2002)
58 Weber, J. S. 2008. Small Sample Histogram Possibilities and Paradoxes. In *JSM Proceedings*. Alexandria, VA: American Statistical Association.
59, 60, 61, 62 US patent Numbers 7,392,156, 7,603,254, 8,131,503, Application Number 13/411,974 allowed October 6, 2014
62 Wikipedia http://en.wikipedia.org/wiki/histogram (last modified 5 December 2014)

Appendices

Appendix A Using histogram shape level sets for many histogram estimators

Appendix B: Some details, some proofs
- **Appendix B.1 Domain D_0**
- **Appendix B.2 Showing minimum bin width does not exist, use infimum bin width**
- **Appendix B.3 Over-smoothing, MISE and ML - Bin width, number of bins rules**
- **Appendix B.4 Data symmetry, shape reversals**

Appendix A Using histogram shape level sets for various histogram estimators

In the main text we focus on MISE histograms provided by two easily accessible website MISE histogram calculators and compare with correctly calculated MISE and maximum likelihood uniform bin width histograms. This is meant to show that inaccuracies in current practice are too great to ignore. However uniform bin width histogram shape level sets are widely applicable to exact calculation of many kinds of histograms. Examining the Rice Stats website and the Shimazaki website suggests that there probably would be unacceptable inaccuracy in other *reasonable approximations*. For example, the

1916 dissertation of K. Smith is suspect because of the reasonableness of R. A. Fisher's objection as well as the clear inability in 1916 to do easily approximate calculations like those performed by the Rice Stats website and Shimazaki websites. Another example, Simonoff and Udina shape stability analysis emphasizes how simple exact calculations are via shape level sets. (The situations considered below connect to published or evident histogram situations.)

MISE – The MISE objective function depends on bin width and shape is almost always negative and infimized at the infimum bin width. The global MISE histogram density is associated with the minimum of all of the *local* shape MISE infima.(For each shape, the *inf* bin width associated with a unique location (unique up to integer multiples of bin width.)

Exact calculations identify the *inf* bin width (t_o, h) vertex for each shape level set. *Shape* is determined from any level set *interior* point. Also see note in Shape Stability discussion combining shape stability with MISE.

ML – The ML situation is identical to MISE except that the ML objective function is always positive. Like MISE, ML is optimized for a shape for *inf bin width* for a shape. The global ML density is the density associated with the maximum among the local shape ML objective function suprema. (J. R. Thompson and R. Tapia (1990) show that among step function densities the usual histogram density is maximum likelihood.)

MOM – Method of Moments may be more detailed than expected.
See: https://arxiv.org/ftp/arxiv/papers/1606/1606.04891.pdf. (Recently a journal editor wrote that Table 1ab is unclear. So anyone looking at my arxiv MOM paper must figure out Table 1ab. That's not the whole story, but certainly a big part. Warning: My arxiv paper may state incorrectly that sometimes bin width minima exist. If so, that's an error that I have not corrected.

In the same two dimensional {(t_o, h)} plane, the MOM mean and *variance* constraints are piecewise *straight lines*. (The variance *level curves* – graphs – are straight lines. The *spacing of the lines is nonlinear* in h but does not depend on t_o.) *Shape level sets* make the situation clear for UBW histogram densities, relative frequency histograms, and frequency histograms

MOM is not the same for frequency histograms, relative frequency histograms and histogram densities. For these three kinds of MOM histograms, for each shape, the mean and variance straight line level curves may intersect the associated shape level set, or not, and may intersect each other inside or outside of the associated shape level set. A mean or variance level curve outside of a shape level set corresponds to (t_o, h) values that lead to a different shape. For a shape, if a mean or variance level curve intersects its shape level set, then there is an adjustment of the bins so that grouped data, multinomial relative frequency or histogram density mean or variance or both equal the sample mean and variance. This does not quite clarify all of the possibilities.

(MOM density histograms, frequency histograms and relative frequency histograms do *not* freely scale to one another because the definition of moments is not the same for histogram densities, multinomial models, and grouped data frequency histograms. This is different from the MISE and ML *presentation* of density histograms as frequency histograms. With that said, I suppose a MOM histogram density could be presented as a

frequency histogram with a foot note that the moments correspond to the density histogram, not frequency histogram grouped data moments.)

Even though *two* parameters, bin width, h, and bin location, t_o, determine uniform width bins, I focus on the first *three* moments. For histogram densities and frequency histograms, for each shape, constraining t_o and h, beyond the constraints imposed by the shape, MOM leads to straight-line relationships between t_o and h for the mean and variance, but not skewness. These straight lines may or may not intersect a shape level set interior + boundary, or just a boundary or vertex. Calculated examples may be the easiest way to explain all of the combinations of *mean and variance consistency with shape*, and that all of the combinations actually occur. Further, examples show that for *many* shapes, there are unique values for (t_o, h) that lead to histogram density or frequency histogram grouped data moments that equal the sample mean and variance. (The grouped data mean depends on bin location and width. Variance depends only on bin width, hence, grouped variance level curves are horizontal straight lines in $\{(t_o, h)\}$.)

Both gamma and Fisher-Pearson measures of skewness depend only on shape. Also gamma and Fisher-Pearson are monotonically related. Consequently skewness rankings of frequency, relative frequency and density histograms for gamma sn Fisher-Pearson are identical. (Also see Weber 2008, *although in that note, there is misunderstanding of skewess and shape*.) Having the histogram skewness equal to the sample skewness rarely happens except for exactly symmetric data and symmetric UBW histogram shape. Symmetric data does not guarantee symmetric histogram shapes – see Appendix B.4.

For data collections comprised of finitely many rational number values there are countable numbers of bin width and location values for which *all grouped data moments* equal the data collection moments. This extends what is already known (D. W. Scott email) that histogram moments converge to sample moments as bin width becomes arbitrarily small. Further, this is foreshadowed by Anon (2006) [actually Anon. ed. = R. Stong, assoc. editor, *Monthly* problems. Thank you Professor Stong.]

Min *Chi-squared* was considered by K. Smith, circa 1916, in her Ph.D. dissertation. This is not about "good" histograms, but estimating normal parameters from grouped data, from uniform bin width histograms, via min *chi-squared*. K Smith was K. Pearson's student. R. A. Fisher objected to Smith's procedure because different bins lead to different bin counts and different normal parameter estimates. (Stigler 2005)

This suggests three related problems.
1. For a fixed histogram ***density***, determine the normal mean and variance parameters that minimize the *chi-squared* statistic for actual bin frequency and expected count for a normal model.
2. For a fixed histogram ***shape***, determine the UBW histogram bin parameter values and normal mean and variance parameter values for the normal density that minimize the *chi-squared* statistic for actual bin frequency and expected count for a normal model.
3. For a data collection, determine UBW t_o and h parameter values and the normal mean and variance parameter values that that minimize the *chi-squared* statistic for the UBW histogram bin frequency and expected count for a normal model. *Tentatively*, these problems can be solved with shape level sets and LaGrange multipliers. [What about Sheppard's correction?]

Like method of moments, if the LaGrange multiplier or other solution is outside of the shape level set, then the associated (t_o, h) values do not lead to the same shape. Then maybe the optimum is on the boundary, possibly on a minimum length path that reaches the shape level set from the unrestricted optimum.

The last problem involves solving the second problem for every UBW histogram shape and then selecting the minimum among all of the shape minima. Four variables are involved: bin parameters, t_o, h; and normal parameters, mean and variance. For t_o, h in a shape level set, the shape, i.e. observed frequencies, are the same. The expected counts are given by the usual normal expected values for bin intervals $[t_o + (k-1)h, t_o + kh)$, for various normal mean and variance parameter values. For each shape there is a minimum (or infimum) chi squared value, for bin parameter values in a shape level set and various normal parameter values. The minimum of all of the shape minima or infima solves the third problem. A final note: location may not be unique unless some bin edges equal data values. Otherwise, a translation of both the bin edges and normal mean will lead to the same min *chi-squared* calculation.

Approximating a histogram density with a normal density to minimize integrated squared deviation between the densities. – L. Brown and G. Hwang, (1993) *JASA*. This suggests two kinds of problems.
1. For a fixed histogram density, determine the normal mean and variance parameters for the normal density that minimizes the integrated squared deviation between the normal density and the histogram density.
2. For a data collection, determine the UBW t_o and h parameter values and the normal mean and variance parameter values that minimize integrated squared deviations between the histogram density and the normal density.

Maybe both problems can be solved with shape level sets and LaGrange multipliers.

Note that the integrated squared deviation of the normal curve from the histogram density is unchanged when both the histogram density and the normal density are translated (via t_o and the normal mean) within a shape level set.

So a unique solution to the more general problem is not assured, unless translation of t_o and the normal mean is restricted.

Brown & Hwang address problem 1, apparently, do not change the histogram, only adjust normal parameters. Overall, this calls to mind Smith's 1916 effort.

Shape stability
Shape stability could be considered the most primitive and intuitive rule for selecting bin width, even if it does not have a distinct theoretical description beyond selecting among bin widths that have the fewest number of shapes due to translation of bins. And the exact partitioning of a range for bins can be done exactly, some might say elegantly, with shape level sets.

J. Simonoff and F. Udina 1997 consider bin widths that have the fewest number of different shapes associated with each bin width and develop a bin width index based on the number of shapes and an application of the econometric Gini index.

The Simonoff – Udina approximation method involves estimating the number of shapes for each of a set of bin width values via a fixed number of uniformly spaced translations for each bin width. Then a polynomial model of shape stability leads to bin widths

associated with the fewest number of shapes. We do not know of a website implementation of this procedure. Tentatively, the polynomial models of numbers of shapes for various bin widths – a model of shape stability – ironically will be very unstable for different sets of bin widths and different numbers of uniform translations for each bin width. In contrast, the exact method described below is …exact and correct. One result emerges every time. There is no algorithm variability due to different initial values or parameter settings.

Shape level sets afford exact description of bin width shape stability. Each shape level set has a vertex for the min or inf bin width and a vertex for the max or sup bin width for a shape. Using a bin width range $R_h \equiv$

$$[(\text{Min } \{|x_i\text{-}x_j| \mid x_i \neq x_j\})/2, \quad (x_{max} - x_{min}) + (\text{Min } \{|x_i\text{-}x_j| \mid x_i \neq x_j\})/2],$$

the set of points $\{h_{min,s}, h_{max,s} \mid s = 1 \text{ to } S = \text{number of shapes}\}$ partitions R_h into open cells, $(h_{\text{min or max,si}}, h_{\text{min or max,sj}})$, … $h_{\text{min or max,si}} < h_{\text{min or max,sj}}$ … and cell end points, each of which is associated with a fixed set and fixed count of shapes, for the bin width intervals and interval end points, etc.

A similar partition of the range of the objective function shows which sets of shapes can be incorrectly ranked on account of approximate bin widths. The range of MISE values may be partitioned by the extreme values for each shape: {inf MISE(shape *s*), sup MISE(shape *s*) | *s* = 1 to ***S***}, wherein inf usually occurs for inf bin width, sup for sup bin width, for each shape. Just as MISE and ML do not lead to the same ordering of shapes, even though all local extrema occur for the same bin parameter values, the partition of the MISE range over all shapes by {inf MISE(shape *s*), sup MISE(shape *s*) | *s* = 1 to ***S***} will give cells of MISE values that correspond to a set of shapes that all have a common range of MISE values within the range of each range for each shape. Clearly a similar partition can be based on ranges of likelihood values, (3), for shapes, etc.

Number of bins rules The shape level set algorithm provides shape level sets for a range of numbers of bins, from $1 \leq k_1 \leq k_2 \equiv \boldsymbol{K} < \infty$. If a number of bins rule prescribes exactly ***k*** bins, then $k_1 = k_2 = \boldsymbol{k}$. If a range of numbers of bins is prescribed, then obviously k_1 = min number of bins and k_2 = max number of bins, etc. For example, for the Sturges Rule, $k_1 \leq log_2(n) \leq k_2$, k_1, k_2 integers $\leq \boldsymbol{K}$. The all histogram shapes and densities associated with Sturges numbers of bins can be obtained.

Bin width rules Exact determination of shape stability partitions the bin range into open cells and cell end points. The same set of shapes arises for any bin width in a cell interior or cell end point. The width prescribed by a bin width rule will be contained in a cell interior or equal a cell end point. This leads to all of the shapes and densities that satisfy a bin width rule point estimate.

Appendix B: Some details, some proofs

Appendix B.1 Domain $\boldsymbol{D}_o$

Uniform bin width frequency histogram "shape" is *the* *list* of bin counts, ν_k, $k = 1$ to $\boldsymbol{K}$, *beginning with the first positive count*, ν_1. A shape level set (Fig.1, §2.2) is the set of bin parameter values for t_o, h, that lead to the same shape for the same data. (Shape level sets are identical for frequency, relative frequency and density histograms.) The first bin, $\mathbf{[}t_o + (k - 1)h, t_o + kh\mathbf{)}$, $k = 1$, must contain the data minimum, x_{min}. As a consequence, each level set is *one* convex polygon in $\boldsymbol{D}_o$ in the $\{(t_o, h)\}$ plane. $\boldsymbol{D}_o$ has enough (t_o, h) points to lead to all UBW histogram shapes that data can have. As previously noted, this is more restrictive than is customary, Scott (1992, p 49, etc.), Weber (2008).

$\boldsymbol{D}_o$ is defined by the following three constraints:

1 x_{min} contained in a first bin means $t_o \le x_{min} < t_o + h \leftrightarrow$
1a: $t_o \le x_{min}$
1b: $x_{min} < t_o + h$, $x_{min} - h < t_o$

2 At most $\boldsymbol{K}$ bins $\leftrightarrow$
2a: $x_{max} < t_o + \boldsymbol{K}h$
Exactly $\boldsymbol{K}$ bins so x_{max} is in the last bin $[t_o + (\boldsymbol{K}–1)h, t_o + \boldsymbol{K}h) \leftrightarrow$
2b: $t_o + (\boldsymbol{K}–1)h \le x_{max}$ and 2a

3 $\boldsymbol{D}_o$ is bounded: $h < \boldsymbol{B} \equiv (x_{max} - x_{min}) + \delta$, $0 < \delta$

These lead to the following boundaries for $\boldsymbol{D}_o$:

1a: $t_o = x_{min}$
1b: $x_{min} = t_o + h$
2a: $x_{max} = t_o + \boldsymbol{K}h$
3: $h = (x_{max} - x_{min}) + \delta$

The above boundaries lead to four $\boldsymbol{D}_o$ vertices, clockwise:
vertex $_1$: (1a,3): $(x_{min}, (x_{max} - x_{min}) + \delta)$
vertex $_2$: (1a,2): $(x_{min}, (x_{max} - x_{min})/\boldsymbol{K})$
vertex $_3$: (1b,2): $(x_{min} - (x_{max} - x_{min})/(\boldsymbol{K} - 1), (x_{max} - x_{min})/(\boldsymbol{K} - 1))$
vertex $_4$: (1b,3):$(x_{min} –((x_{max} –x_{min}) + \delta)),(x_{max} –x_{min}) + \delta)$

Since D_o depends on $\boldsymbol{K}$, δ, x_{min}, x_{max}, one might write $\boldsymbol{D}_o(\boldsymbol{K}, \delta, x_{min}, x_{max})$ or $\boldsymbol{D}_o{}^{\boldsymbol{K},\delta}{}_{[x_{min}, x_{max}]}$

Normalizing data $\mathbf{[}x_{min}, \ldots \ldots, x_{max}\mathbf{]}$ ***to*** **[0, 1] and letting** $\boldsymbol{K} \to \infty$, $\delta \to 0$ **brings** ***vertex*** $_2$ and ***vertex*** $_3$ **together, leading to** $\boldsymbol{D}_o{}^{\infty}{}_{[0,1]}$**:**

vertex $_1$: (1a,3): $(x_{min}, (x_{max} - x_{min}))$; $(0, 1 + \delta)$ $\to$ (0, 1)
vertex $_2$: (1a,2): $(x_{min}, 0)$; $(0, 1/\boldsymbol{K})$ $\to$ (0, 0)
vertex $_3$: (1b,2): $(x_{min}, 0)$; $(–1/(\boldsymbol{K} – 1), 1/(\boldsymbol{K} – 1))$ $\to$ (0, 0)
vertex $_4$: (1b,3): $(x_{min} –((x_{max} – x_{min}))),(x_{max} – x_{min}))$; $(–(1 + \delta), 1 + \delta)$ $\to$ (–1, 1)

$\boldsymbol{D}_o{}^{\infty}{}_{[0,1]}$ is the triangle with vertices: **(0,1), (0,0), (-1,1)**.

Shape level set boundaries (4b): $x_i = t_o + kh$, $k = 1$ to $\boldsymbol{K}$, $i = 1$ to n *partition* $\boldsymbol{D}_o$ or $\boldsymbol{D}_o{}^{\infty}{}_{[0,1]}$ into shape level sets like the one shown in Figure 1. All of the lines $x_i = t_o + kh$, $i = 1$ to n, $k = 1$ to $\boldsymbol{K}$ intersect $\boldsymbol{D}_o$. The shape level set algorithm generates the lines $x_i = t_o + kh$ one by one, determines intersection vertices with $\boldsymbol{D}_o$ and already calculated existing polygons. Including the last line, $x_n = t_o + \boldsymbol{K}h$, completes a level set partition.
The level sets can be specified and organized by sorting (5b) lexicographically on $\boldsymbol{K}_s$, $\nu_{s,k}$; and sorting (5c), (5d) on MISE-(1) values ascending, ML-(3) values descending.

$\{(\nu_{s,k}\,,\,(t_o,\,h)_{s,v}) \mid s = 1 \text{ to } \boldsymbol{S},\ k = 1 \text{ to } \boldsymbol{K},\ v = 1 \text{ to } \boldsymbol{V}_s\}$ (B.1.1)
$\{(\nu_{s,k}\,,\,(t_o,\,h)_{s,v}) \mid s = 1 \text{ to } \boldsymbol{S},\ k = 1 \text{ to } \boldsymbol{K}_s\,,\ v = 1 \text{ to } \boldsymbol{V}_s\}$ (B.1.1*)
$\{(\boldsymbol{K}_s,\,\nu_{s,k}\,,\,(t_o,\,h)_{s,v}) \mid s = 1 \text{ to } \boldsymbol{S},\ k = 1 \text{ to } \boldsymbol{K}_s\,,\ v = 1 \text{ to } \boldsymbol{V}_s\}$ (5) ≡ (B.1.2)
$\{(\text{MISE-(1)},\,\boldsymbol{K}_s,\,\nu_{s,k}\,,\,(t_o,\,h)_{s,v}) \mid s = 1 \text{ to } \boldsymbol{S},\ k = 1 \text{ to } \boldsymbol{K}_s\,,\ v = 1 \text{ to } \boldsymbol{V}_s\}$ (B.1.3)
$\{(\text{ML-(3)},\,\boldsymbol{K}_s,\,\nu_{s,k}\,,\,(t_o,\,h)_{s,v}) \mid s = 1 \text{ to } \boldsymbol{S},\ k = 1 \text{ to } \boldsymbol{K}_s\,,\ v = 1 \text{ to } \boldsymbol{V}_s\}$ (B.1.4)

$\boldsymbol{S}$ ≡ number of shapes
$\boldsymbol{K}$ ≡ max number of bins
$\boldsymbol{K}_s$ ≡ number of bins for s^{th} shape ≡ index of bin for x_{max}
$\boldsymbol{V}_s$ ≡ number of vertices for s^{th} shape
(B.1.1), (B.1.1*), (5) == (B1.1.2), (B1.1.3), (B1.1.4) all are right ragged matrices (from $\boldsymbol{V}_s$, $\boldsymbol{K}_s$) with $\boldsymbol{S}$ rows, although obviously can be made rectangular with zeroes.

As is typical, Scott (1992) p 49 does *not* require bin indexing so that the first bin contains the minimum data value. Weber (2008) and Visual Basic implementation of histogram level set exact calculation of various kinds of histograms begins with partitioning $\boldsymbol{D}_o$.

Appendix B.2 Showing minimum bin width does not exist, so use infimum bin width

Proof: Referring to Fig. 1, the closure of a *SLS*, $\overline{SLS}$, has a minimum bin width, h_{min}, at the intersection of

$$t_o + k_2h = x_j \text{ and } t_o + k_3h = x_p. \text{ That is, } h_{min} = (x_p - x_j)/(k_3 - k_2).$$

Keep in mind that in a *SLS*, in Fig. 1, t_o and h are variable and x_j, x_p, k_2, k_3 are not variable. Question: Is h_{min} ε *SLS*?
SLS are associated with half-open bins, $[t_o + (k-1)h,\ t_o + kh)$. (Kendall & Stuart, various editions; D. W. Scott, 1992).

$$t_o + k_2h = x_j < x_p = t_o + k_3h.$$

So a left, closed bin edge, $t_o + k_2h = x_j$, and a right, open bin edge, $t_o + k_3h = x_p$, is associated with the bins for the shape associated with Fig. 1 *SLS*.
For $h = h_{min}$, x_j ε $[t_o + k_2h\ ,\ t_o + (k_2+1)h))$.
For $h > h_{min}$, x_p ε $[t_o + (k_3-1)h\ ,\ t_o + k_3h))$.
However for $h = h_{min}$, x_p is *not contained in* $[t_o + (k_3-1)h,\ t_o + k_3h))$.
That is, $x_p = t_o + k_3h_{min}$, so x_p *is not less than* $t_o + k_3h_{min}$.
So shape for $h = h_{min}$ is not the same as the shape for the *SLS*, for h close-to-but > h_{min}.
So $(t^{\,hmin}{}_o,\ h_{min})$ is not contained in the *SLS*.
We must focus on $h_{inf} = inf\ h$, not h_{min}, and inf MISE, sup ML.
Fortunately $h_{min}(\overline{SLS})$ is easy to calculate and $h_{inf}\ (SLS) = h_{min}(\overline{SLS})$. *Inf* MISE and *sup* ML are easily calculated with the shape bin counts for $(t_o,\ h)$ ε *SLS* and $h_{inf} = h_{min}(\overline{SLS})$.
The easiest example of infimum, not minimum bin width is for the one-bin histogram: for the bins $[x_{min},\ x_{max} + \delta)$, $0 < \delta$, $h_{inf} = (x_{max} - x_{max})$

Appendix B.3 Over-smoothing, MISE and ML - Bin width, number of bins rules

Table B.3 *and discussion* show that six over-smoothing rules do not prevent shape reversals, using data example #5 (Weber 2008 data #2).

Table B.3 Over-smoothing rules and bin translation skew shape reversals

Uniform Bin width Histogram Shapes				Satisfies over-smoothing conditions (Yes or No), Scott 1992, pp 73 to 75. n = 20, range = 5.9, s_x = 1.9705, IQR=3.41 Rule			
v_1, v_2 …	**t_0**	**h**	**# Bins**	**(3.42) # Bins ≥ 4**	**(pre3.42) h < 1.725**	**(3.43) h < .707**	**(3.44) h < 3.72**
C: 10, 9, 1	1.425	3.208	3				
D: 1, 9, 10	-1.048	3.208	3	No	No	No	Yes
E: 8, 4, 7, 1	1.977	1.979	4				
F: 1, 7, 4, 8	0.108	1.979	4	Yes	No	Yes	Yes
G: 6, 4, 4, 5, 1	1.983	1.475	5				
H: 1, 5, 4, 4, 6	0.642	1.475	5	**Yes**	**Yes**	**Yes**	**Yes**
I: 4,6,0,5,4,1	1.976	1.181	6				
J: 1,4,5,0,6,4	0.9391	1.181	6	**Yes**	**Yes**	**Yes**	**Yes**

Data : 2.05, 2.27, 2.50, 2.95, 3.18, 3.41, 3.64,
3.86, 4.09, 4.32, 5.68, 5.91, 6.14, 6.36,
6.59, 6.82, 7.05, 7.50, 7.73, 7.95

Source: Weber, J.S. -2008 JSM Proc.

In addition to the four rules above, Scott 1992 §3.3.1, p 55 includes the Freedman – Diaconis rule: $h \leq 2\text{IQR}n^{-1/3} = 2(3.41)/(20)^{-1/3} = 2.5125$. Histogram bins E, F; G, H; I, J satisfy this rule. (Freedman, D. and Diaconis P., 1981) Also the Sturgis Rule, Scott 1992 p 48, $K = 1 + \log_2 n$ leads to $1 + \log_2(20) = 1 + 4.321928 = 5.321928$, five or six bins. Histogram bins G, H; I, J satisfy this rule. Thus if two additional columns of rules are added to Table 7 then G & H; I & J satisfy all six rules. (Calculations for Table 7 are available upon request.) (Since the data are symmetric, Doane (1976) adjustment(s) of the Sturges' rule do not apply.)

How well does MISE satisfy over-smoothing rules? From Table 1*, exact MISE clearly over-smoothes in three out of six examples. Overall, MISE via Rice Stats website, Shimazaki *and exact calculation* appears to over-smooth in nine out of eighteen instances in six examples. *However, ML does not violate any over-smoothing rules in these examples.*

Over-smoothing observations:

a. Over-smoothing rules do not prevent shape reversals.
b. Exact and approximate MISE (1a) over-smoothes in 50% of my examples.
c. ML does not over-smooth in any of my examples.

Appendix B.4 Data symmetry, shape reversals

Over-smoothing and shape reversals can be approached via data symmetry. (The data for Table 2a are symmetric.) The ordinary meaning of "smoothing" suggests that more smoothing should reduce graphic irregularities, such as reversal of non-symmetric histogram shapes. But the situation appears opposite for uniform bin width histograms. Part **B** $\Rightarrow$ **A** below means that exactly symmetric data leads to reversal of ***asymmetric*** shapes only for sufficiently wide bins.

The following lemma connects data symmetry and asymmetric shape reversals.

Lemma Data Symmetry and Asymmetric Shape Reversals.
A. and **B**. below are equivalent.
A. Data are symmetric.
B. For every UBW histogram shape, there are translations of the bins leading to the reversal of shape, *including asymmetric shapes* such as (1, 7, 4, 8) and (8, 4, 7, 1).

Sketch of *proof*:
A $\Rightarrow$ **B** *Reflection* across the data axis of symmetry *of the bins and the data* reverses the shape. Due to data symmetry, data values and value frequencies are unchanged. Reflection of *UBW* bin edges can be achieved via a translation. So translation of the bins leads to a reversal of the bin counts, such as (1, 7, 4, 8) to (8, 4, 7, 1). (Data and bin edges are reflected, but bin orientation is not reflected or reversed. Bins are *defined* from bin edges *a* and *b* as [*a*, *b*). Reflection of "[*a*, *b*)" is not "(*b*, *a*]".)

B $\Rightarrow$ **A** Small bin width "shape reversals" involve *isolated individual data values. So small enough bin width shape reversals imply symmetric data value* ***frequencies****. Also,* ***arbitrarily small*** *bin width symmetric shape reversals imply symmetric data* ***values****. Symmetric data value frequencies and symmetric data values mean symmetric data*. (Weber 2008; Supplement 3)

So for symmetric data *only over smoothing wide bins* can lead to reversals of *asymmetric* shapes. (Simonoff & Udina 1997, §2.3.3, opine that symmetric data may be *better* from the perspective of bin width shape stability. This Lemma relates to that thought.)

After word. … This effort includes, begins with an elementary solution to a problem that has been on the table since 1662: For data, x_i, what uniform bin width histogram shapes (i.e. bin counts) are possible? An elementary solution to an old problem often raises a red flag: Is it *Junk Science*? That it claims to correct errors in what has become common knowledge is yet another red flag. About all I can suggest regarding this worry is revisit R. J. Little (2013) *JASA*, with an open mind.