On the Properties of the Metalog Distribution

Manel Baucells
Darden School of Business, University of Virginia. baucellsm@darden.virginia.edu

Lonnie Chrisman
Chief Technology Officer, Lumina Decision Systems. lchrisman@lumina.com

Thomas W. Keelin
Managing Partner, Keelin Reeds Partners. tomk@keelinreeds.com

Stephen Xu
Darden School of Business, University of Virginia. xuz25@darden.virginia.edu

There has been a longstanding interest among statisticians and analysts for parametric probability distributions (e.g., the Pearson or the Johnson families) flexible enough to match arbitrarily shaped, non-normal data sets. The metalog family emerged in 2016 as a new alternative and has been widely adopted for its tractability and flexibility. Metalogs can closely approximate any continuous quantile function, creating adaptable data-driven probability distributions. However, key theoretical properties such as the number and location of modes, explicit expressions for the moments, and feasibility guarantees have remained largely unexplored. In this paper, we define the metalog 2.0, which uniquely assigns coefficients to avoid degeneracy. We establish the possible number of modes, compute their location, and derive explicit formulas for moments and partial expectations, enabling more precision in applications. A major challenge when fitting data is that the metalog function may fail to be monotonic, rendering it invalid. We provide a mathematically exact feasibility test and introduce an algorithm to find the best feasible metalog fit with arbitrary precision.

Key words: Metalog distribution, Moments and Modes, Partial Expectation, Data fitting.

1. Introduction

A key step in many decisions is to develop a probabilistic forecast of the main uncertainties affecting the consequences. When the uncertain quantities are continuous in nature, decision analysts often resort to parametric distributions to fit past data, or capture their subjective forecasts (Theodossiou 1998). Most existing families of distributions such as the Pearson (which includes the Normal, Student, Beta, and Gamma) or the Johnson have a limited number of parameters and their shape is typically unimodal. When large data sets are available, such families may fall short of capturing important features of the data.

On the opposite extreme, kernel density estimators (KDE's), as a continuous analog to histograms, can be useful for capturing nuances of any-shape data (Chan et al. 2016, Bertschek and Kaiser 2004). The shapes of KDE's, however, depend on the arbitrary choice of bandwidth and their equations, which typically require as many terms as data, become increasingly unwieldy. The lack of tractability of the equations obscures the properties of KDE's, including a closed form expression for the moments. Similarly, splines employ piece-wise polynomials to track the data histogram or its empirical cdf (Wahba 1990), but depend on the arbitrary choice of knot locations and the smoothing criteria.

The metalog family, introduced by Keelin (2016), emerged as a new alternative possessing great flexibility while being highly tractable. The metalog takes the quantile function of the logistic distribution (the inverse of its cdf), and replaces the location and scale parameters by two polynomials. The resulting function M, if increasing, represents the quantile function of a probability distribution. Quantile functions possess many attractive properties (Gilchrist 2000): they are easy to sample or simulate, easy to model dependencies using copulas, and easy to transform. And because metalogs are linear in the parameters, any convex combination of metalogs is a metalog.[1]

The polynomial nature of metalogs allows them to uniformly approximate any continuous quantile function to arbitrary precision (Powell 1981, Ch. 6). In this sense, metalogs are universal approximators that could become an all-purpose family to model data. For example, metalogs with few parameters approximate known distributions very well (Keelin 2016, Tables 5-8), and provide closed form approximations of distributions such as the sum of lognormals (Keelin et al. 2019).

The original motivation for the Metalog was to aid the assessment of probabilities in decision analysis models. Since then it has been broadly adopted for management science applications, as well as physical sciences, engineering, economics and statistics. Published applications include wave patterns in astrophysics and cosmology (Iacovelli et al. 2022, Finke et al. 2021); scatter patterns in crystallography (Kohlbrecher and Breßler 2022); integrated circuit design and nanomaterials properties (Lai et al. 2022, Runolinna et al. 2023); river and volcanic discharges in earth science (Aspinall et al. 2023, Rongen et al. 2024); toxin levels in drinking water (Norling et al. 2022); conservation biology (Lloyd et al. 2023, Keating et al. 2023); electric vehicles, photovoltaics, and hydrogen production in renewable energy (Caban et al. 2024, Małek et al. 2024); cybersecurity risk mitigation (Wang et al. 2020); separating income classes using the metalog anti-modes (Bittencourt et al. 2025); combination of expert opinions in demographic forecasting (Dion et al. 2020); oil-field production forecasting (Bratvold et al. 2020, Nesvold and Bratvold 2022, 2023); machine learning in molecular biology (Stomma and Rudnicki 2024); cross-platform network simulations (Savage et al. 2019); conditional value at risk in quantitative finance (Zrazhevsky and Zrazhevska 2021); statistical estimation of quantiles from moments (Steins and Herty 2023); and Bayesian inference (Perepolkin et al. 2024, Redivo and Viroli 2024). We believe that such fast and widespread adoption is due to the metalog's properties of shape flexibility, choice of support, tractability of equations, ease of fitting to data, and ease of simulation.

The appeal of the metalog familiy, however, would increase even further if we were to have elementary expressions for the moments and partial expectations, a better understanding of the

[1] Suppose experts submit a feasible metalog expressing his or her probabilistic forecast. If we then aggregate opinions by averaging quantiles, as recommended by Lichtendahl Jr et al. (2013), then the aggregate is a feasible metalog.

number and exact location of its modes and anti-modes, and a way to address the problem of feasibility that often arises when fitting data. Our goal is to fill this gap.

In the original metalog, there is no apparent reason for the proposed assignment of polynomial terms to either the location or the scale polynomial. Here, we find that there is a unique assignment that avoids degeneracy when estimating the parameters from data using ordinary least squares. For seven terms or more, our assignment departs from the one originally proposed, hence we call ours the metalog 2.0. By justifying a unique assignment, we also settle a source of arbitrariness in the definition of the metalog family.

Flexible and tractable multi-modal distributions can greatly enhance the modeling of unusual but relevant events, i.e., probability concentrations away from the center. An obvious question, however, is to know the possible number of modes a metalog can have. We demonstrate that a metalog with k parameters can have up to $\lfloor \frac{k-1}{2} \rfloor$ modes. Thus, metalogs with 4 parameters or less are unimodal, 5- and 6-metalogs can possess up to two modes, and so on.

In Keelin (2016) the first four moments of M are computed using algebraic integration. Here, we show that M has finite moments of any order, and provide an elementary recursion to obtain all its moments. Importantly, we also deliver an elementary expression for the partial expectation, which promises to be very useful in applications such as finance (option valuation, value-at-risk analysis) or operations (expected sales and left-overs when inventory is limited).

A more serious issue that arises in data fitting is that of feasibility, ensuring that M is a proper quantile function. By their polynomial nature, metalogs possess a disturbing feature, namely, they may fail to be increasing. Keelin (2016) notices this possibility, but does not offer a way to guarantee a feasible metalog from data. In the subsequent years, it has become apparent that infeasibility can be pervasive, and that even diagnosing feasibility could be tricky.

In this paper, we provide a mathematically exact and computationally easy way to diagnose feasibility. A byproduct of this test identifies all modes and anti-modes of the metalog. Next, we develop a geometric approach to visualize the problem of finding the best feasible fit to the data, and show the solution exists and is unique. Finally we provide algorithm that approximates such solution to arbitrary precision. The python implementation, `https://github.com/Stephenxuu/metalog_algorithm.git`, takes the data and the desired number of terms as input and produces the coefficients of the best feasible fit metalog. It also reports the first four moments, together with the precise number and location of interior modes and anti-modes.

Metalogs, if not well understood, can behave in unexpected ways. By gaining a deeper understanding of their properties, however, we seek to tame their behavior in order to make statisticians and analysts more comfortable with their use, and facilitate the process of data modeling. In particular, by providing a way to address the thorny issue of feasibility, we hope to remove one of the major roadblocks in the adoption of the metalog.

2. Setup

The quantile function of the logistic distribution is given by $M(y) = \mu + s\ell(y)$, $y \in (0,1)$, where

$$\ell(y) = \ln \frac{y}{1-y}$$

denotes the logit function. The metalog quantile function introduced by Keelin (2016, Definition 1), or metalog for short, substitutes μ and s for a polynomial series around $y = 0.5$. The more terms in the polynomials, the greater the flexibility of the family. For example, the 6-metalog is given by $M(y) = \mu(y) + s(y)\ell(y)$, $y \in (0,1)$, where $\mu(y) = a_1 + a_4(y-0.5) + a_5(y-0.5)^2$ and $s(y) = a_2 + a_3(y-0.5) + a_6(y-0.5)^2$. In other words,

$$M(y) = a_1 + a_2\ell(y) + a_3(y-0.5)\ell(y) + a_4(y-0.5) + a_5(y-0.5)^2 + a_6(y-0.5)^2\ell(y).$$

Note that for k even the number of terms in $\mu(y)$ and $s(y)$ is the same.

2.1. The Metalog 2.0

We now define what could be called the metalog 2.0, which for $k \geq 7$ differs from the original 1.0 version on how it sequentially assign the terms. Our metalog 2.0 assigns a_1 to $\mu(y)$, a_2 and a_3 to $s(y)$, a_4 and a_5 to $\mu(y)$, a_6 and a_7 to $s(y)$ and so on until reaching the desired number of terms.[2] This assignment is the only one that avoids potential degeneracies (see §3.2). Accordingly, let $\mathbf{1}_{j\in\mu}$ denote the indicator function of $j(\bmod 4) \leq 1$ (i.e., $j = 1,4,5,8,...$) and $\mathbf{1}_{j\in s}$ the indicator of $j(\bmod 4) \geq 2$ (i.e., $j = 2,3,6,7,...$). Recall that $\lfloor . \rfloor$ is the integer part and that $0^0 = 1$.

Definition 1. Given $k \geq 2$ and $(a_1, a_2, ..., a_k) \in \mathbb{R}^k$, the k-metalog $M : (0,1) \to \mathbb{R}$ is defined as

$$M(y) = \sum_{j=1}^{k} a_j (y-0.5)^{\lfloor (j-1)/2 \rfloor} (\mathbf{1}_{j\in\mu} + \ell(y)\mathbf{1}_{j\in s}). \tag{1}$$

A 2-metalog is the symmetric logistic, with mean, median, and mode at a_1. A 3-metalog shifts the mean to $a_1 + a_3/2$, keeps the median at a_1 (all metalogs do), and allows for skewness by moving the mode to the $0.5 - 0.25a_3/a_2$ percentile. The 4-metalog retains the same mean, median, and mode, but allows us to manipulate also the kurtosis. And metalogs with $k \geq 5$ can be multi-modal.

Beware that the quantile function M is a map from $y \in (0,1)$ in the vertical axis of the cdf onto $\mathbb{R}$, its percentile in the horizontal axis. Figure 1 displays three metalog fits to the same data by plotting the cdf graph $(M(y), y)$ and the pdf graph $(M(y), 1/M'(y))$, $y \in (0,1)$, respectively.

Metalogs are differentiable, with the i^{th} derivative of M given by

$$\begin{aligned} M^{(i)}(y) &= \mu^{(i)}(y) + \sum_{j=0}^{i} \binom{i}{j} s^{(i-j)}(y)\ell^{(j)}(y), \text{ where} \\ \ell^{(i)}(y) &= (i-1)! \frac{y^i - (-1)^i(1-y)^i}{y^i(1-y)^i}, \ i \geq 1. \end{aligned} \tag{2}$$

[2] In contrast, the metalog 1.0 assigns a_7 to $\mu(y)$, a_8 to $s(y)$, a_9 to $\mu(y)$, a_{10} to $s(y)$ and so on. For $k \in \{7, 11, 15, 19, ...\}$ the functional forms of the metalogs 1.0 and 2.0 with the same number of terms are different.

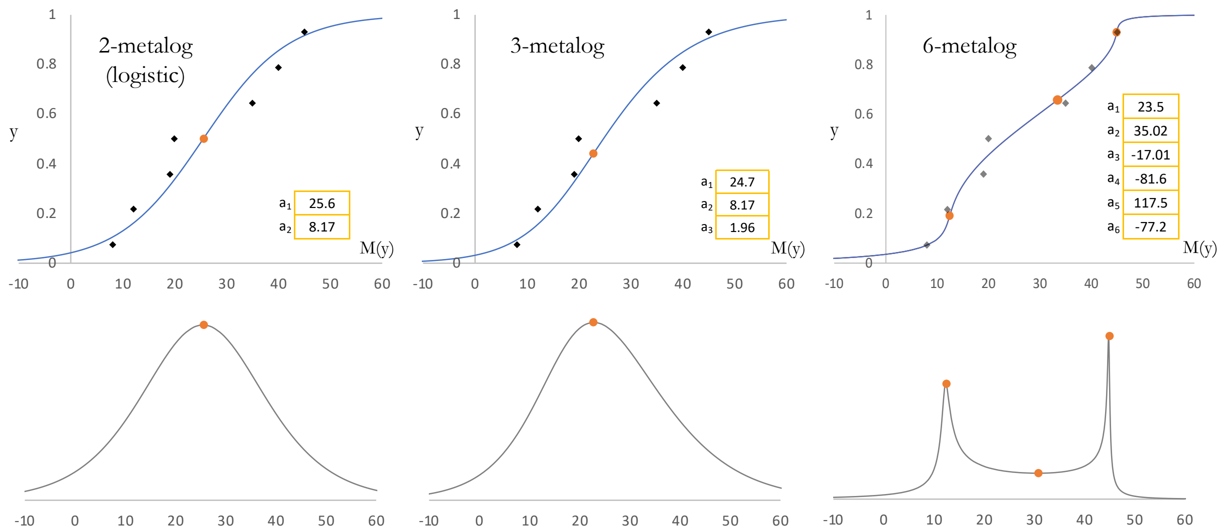

Figure 1 **cdf (top) and pdf (bottom) of a symmetric 2-metalog (left), a skewed 3-metalog (middle), and a bimodal 6-metalog (right) fitted to the data $\mathbf{x} = (8, 12, 19, 20, 35, 40, 45)$ and $y_i = (i - 0.5)/7$, $i = 1, ..., 7$. The orange dots represent the inflection points (modes and anti-modes).**

Some key properties of the metalog will result from having $\ell^{(i)}(y)$ be a quotient of polynomials with denominator > 0; helping justify the choice of logistic as the basis for the metalog.

2.2. Estimation of the Coefficients

Suppose we are given some cdf data $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times (0,1)^n$ comprising quantiles $x_1 \leq x_2 \leq ... \leq x_n$ and cumulative probabilities $0 < y_1 < y_2 < ... < y_n < 1$. The data points may result from subjective assessments (Keefer and Bodily 1983), or from observation. When $\mathbf{x}$ comes from observation, the cumulative probabilities are typically set to either $y_i = i/(n+1)$ (Weibull 1939) or $y_i = (i - 0.5)/n$ (Hazen 1914), $i = 1, ..., n$.

Given $\mathbf{y}$ and the desired number of parameters $k \leq n$, we compute the $n \times k$ basis matrix $\mathbf{Y}$ whose row i and column j entry is given by $(y_i - 0.5)^{\lfloor (j-1)/2 \rfloor}(\mathbf{1}_{j \in \mu} + \ell(y)\mathbf{1}_{j \in s})$, or

$$\mathbf{Y} = \begin{bmatrix} 1 & \ell(y_1) & (y_1 - 0.5)\ell(y_1) & (y_1 - 0.5) & (y_1 - 0.5)^2 & (y_1 - 0.5)^2\ell(y_1) & \dots \\ & & & \vdots & & & \\ 1 & \ell(y_n) & (y_n - 0.5)\ell(y_n) & (y_n - 0.5) & (y_n - 0.5)^2 & (y_n - 0.5)^2\ell(y_n) & \dots \end{bmatrix}.$$

Thus, $\mathbf{Ya}$ is the vector of predicted quantiles associated with the vector of parameters $\mathbf{a} \in \mathbb{R}^k$. Following Keelin (2016), we estimate $\mathbf{a}$ by minimizing the squared distance between the given and the predicted quantiles, $f(\mathbf{a}) = (\mathbf{x} - \mathbf{Ya})^\intercal(\mathbf{x} - \mathbf{Ya})$. If $\mathbf{Y}$ has rank k (see Proposition 2), then the best-fit coefficients are uniquely given by ordinary least squares (OLS)

$$\mathbf{a}^{\text{OLS}} = [\mathbf{Y}^\intercal \mathbf{Y}]^{-1}\mathbf{Y}^\intercal \mathbf{x}. \tag{3}$$

Figure 1 illustrates the OLS fit to the same data using $k = 2, 3$, and 6.

2.3. Support and Density of Feasible Matalogs

Definition 2. A metalog is *feasible* if and only if it is increasing, i.e., $M'(y) \ge 0$, $y \in (0,1)$.

A metalog is the quantile function of a probability distribution if and only if it is feasible. We can then invert M and obtain the cdf, $M^{-1}(x)$. Setting aside the constant metalog—which can only arise if $\mathbf{x}$ is constant—all other feasible metalogs are absolutely continuous with density and support given by[3]

$$m(x) = 1/M'(M^{-1}(x)),\ x \in \operatorname{supp}(M) = \begin{cases} (-\infty,\infty),\ \text{if } s(0) > 0 \text{ and } s(1) > 0, \\ (\mu(0),\infty),\ \text{if } s(0) = 0 \text{ and } s(1) > 0, \\ (-\infty,\mu(1)),\ \text{if } s(0) > 0 \text{ and } s(1) = 0,\ \text{and} \\ (\mu(0),\mu(1)),\ \text{if } s(0) = 0 \text{ and } s(1) = 0. \end{cases}$$

The proposed support follows from (see Lemma 2 in the Appendix)

$$\lim_{y\to 0} M(y) = \begin{cases} -\infty \text{ if } s(0) > 0, \\ \mu(0) \text{ if } s(0) = 0, \\ +\infty \text{ if } s(0) < 0; \end{cases} \quad \text{and} \quad \lim_{y\to 1} M(y) = \begin{cases} +\infty \text{ if } s(1) > 0, \\ \mu(1) \text{ if } s(1) = 0, \\ -\infty \text{ if } s(1) < 0. \end{cases}$$

It also follows that feasible metalogs necessarily satisfy $s(0), s(1) \ge 0$.

Feasible metalogs may have points where $M'(y) = 0$. E.g., the feasible metalog in Figure 2, right, exhibits two such points at $y_1 = 0.00019$ and $y_2 = 0.846$. At these points, the cdf has infinite slope and the pdf takes value infinity yet remains integrable. These feasible metalogs lie at the boundary of becoming infeasible.

3. Properties of the Metalog

3.1. Possible Number of Modes

Metalogs with more and more terms, even if restricted to be feasible, can uniformly approximate any target quantile function on any closed interval (Powell 1981, Ch. 6). But how fast does flexibility increase with k? Specifically, suppose the quantile we seek to approximate exhibits say 3 modes. What is the minimum number of terms required to capture this many modes?

The interior points where the pdf exhibits a local maximum (mode) or a local minimum (anti-mode) correspond precisely to the inflection points of the quantile function. The following is our cornerstone result—far from trivial—providing the maximum number of roots of a metalog and each of its derivatives, excluding the trivial case of a constant or a uniform metalog. Unless noted otherwise, the proof of all results is contained in the Appendix.

Proposition 1. *For $0 \le i \le \lfloor \frac{k+1}{2} \rfloor$, the i^{th} derivative of any non-linear metalog can have at most $2\lfloor \frac{k-1}{2} \rfloor - i + 1$ roots on $(0,1)$; and for $i \ge \lfloor \frac{k+1}{2} \rfloor$ the maximum number of roots is $i - 1$.*

[3] Note the polynomials $\mu(y)$ and $s(y)$ are always well defined on $[0,1]$.

Table 1 applies Proposition 1 to obtain the maximum number of inflection points. And there can be as many interior modes as the maximum possible number of inflection points plus one, then divided by two; resulting in at most $\lfloor \frac{k-1}{2} \rfloor$ modes. Bounded or semi-bounded metalogs could have modes at the extremes. These metalogs must meet $s(0)=0$ or $s(1)=0$, a constraint that reduces the number of interior roots, ensuring that the counting of modes never exceeds $\lfloor \frac{k-1}{2} \rfloor$.

For k even, it is easy to generate instances meeting the maximum number of modes. For k odd, the set of metalogs that achieve the maximal number of modes is much smaller. Still, we have been able to generate metalogs meeting $\lfloor \frac{k-1}{2} \rfloor$ for k up to 13.

Table 1 **Maximum number of extrema, inflection points, modes, and interior anti-modes of a feasible metalog.**

k	Roots $M'(y)$ $2\lfloor \frac{k-1}{2} \rfloor$	Roots of $M''(y)$ $2\lfloor \frac{k-1}{2} \rfloor - 1$	Modes $\lfloor \frac{k-1}{2} \rfloor$	Interior anti-modes $\lfloor \frac{k-1}{2} \rfloor - 1$
$3,4$	2	1	1	0
$5,6$	4	3	2	1
$7,8$	6	5	3	2
$9,10$	8	7	4	3
$11,12$	10	9	5	4
$13,14$	12	11	6	5

Sketch of the Proof. Set $i \geq \lfloor \frac{k+1}{2} \rfloor$ so that $\mu^{(i)}(y) = s^{(i)}(y) = 0$. For $k=4$, for example, compute $M''(y) = 2a_3\ell'(y) + (a_2 + a_3(y-0.5))\ell''(y)$. Having $\mu^{(i)}(y)=0$ ensures that $M^{(i)}(y)$ no longer involves μ-coefficients, and having $s^{(i)}(y)=0$ removes $\ell(y)$ (only ℓ', ℓ'', ... is involved).

Then, consider the function $y^i(1-y)^i M^{(i)}(y)$, which has the same zeros as $M^{(i)}(y)$ on $(0,1)$. Because $\ell^{(i)}(y)$ is a ratio of polynomials with denominator $y^i(1-y)^i$, we have that $y^i(1-y)^i M^{(i)}(y)$ is a polynomial. In our example,

$$\begin{aligned} y^2(1-y)^2 M''(y) &= 2a_3 y(1-y) + (a_2 + a_3(y-0.5))(2y-1) \\ &= 0.5a_3 + a_2(2y-1). \end{aligned}$$

Casual inspection of the first expression suggests a polynomial of degree 2. Surprisingly, the coefficients of the leading term y^2 cancel out, resulting in a polynomial of degree 1. In general, the degree of $y^i(1-y)^i M^{(i)}(y)$, $i \geq \lfloor \frac{k+1}{2} \rfloor$, could be as high as $i + \#s - 2$ (here, #s is the number of coefficients in $s(y)$). The challenge is to show that the $\#s - 1$ high order coefficients cancel out so that the degree drops to $i-1$. Moreover, this polynomial is proportional to $s(0)$ and $s(1)$ for $y=0$ and $y=1$, respectively. In our example, $0.5a_3 + a_2(2y-1) = 2s(0)(1-y) + 2s(1)y$. Thus, if $s(0)=0$ ($s(1)=0$), then $y=0$ ($y=1$) is a root. Hence $M^{(i)}(y)$, $i \geq \lfloor \frac{k+1}{2} \rfloor$, can have at most $i - 1 - \mathbf{1}_{s(0)=0} - \mathbf{1}_{s(1)=0}$ roots on $(0,1)$. For $i < \lfloor \frac{k+1}{2} \rfloor$, the integral function $M^{(i)}(y)$ can add at most one root to $M^{(i+1)}(y)$. It follows that $M^{(i)}(y)$ has at most $2\lfloor (k-1)/2 \rfloor - i + 1 - \mathbf{1}_{s(0)=0} - \mathbf{1}_{s(1)=0}$ roots on $(0,1)$. $\square$

3.2. The Assignment of Coefficients and the Rank Question

The metalog is balanced, in that the number of terms in $\mu(y)$ and in $s(y)$ is off by at most one. Balanced assignments produce better fits, as the polynomial terms adjust the interior portion of the function, and the logistic terms adjust the tails. Within balanced assignments, however, why assign a_7 to $s(y)$ (as we propose), and not to $\mu(y)$ (as originally proposed)? Our next result shows that there is only one balanced assignment ensuring that the basis matrix $\mathbf{Y}$ has full rank. To gain control on the structure of $\mathbf{Y}$, we restrict attention to probabilities in $\mathbf{y}$ being symmetric around 0.5, i.e., $y_i = 1 - y_{n-i+1}$, $i = 1, ..., n$. This case is eminently relevant when fitting elicited percentiles or observed data (Keefer and Bodily 1983).

Proposition 2. *Under symmetry, the only balanced assignment guaranteeing rank*$(\mathbf{Y}) = k$ *for all* $k \le n$ *assigns* a_1 *to* $\mu(y)$, a_2 *and* a_3 *to* $s(y)$, a_4 *and* a_5 *to* $\mu(y)$, a_6 *and* a_7 *to* $s(y)$ *and so on.*

In the case of k even, full rank always holds without requiring symmetry. To see this, use Proposition 1 to note that M itself can have at most $k-1$ roots. This contradicts that $\mathbf{Ya} = 0$, $\mathbf{a} \neq 0$, can hold for k rows. Thus, if $n \ge k$, full rank is guaranteed. For k odd, however, the number of roots could be k, so that $n \ge k+1$ is required for full rank.

If $n = k$ is odd, then $\mathbf{Y}$ may be rank deficient. Under symmetric percentiles, the assignment of coefficients matters for the rank. In particular, assigning all odd coefficients to $\mu(y)$ produces rank deficiency for $k \in \{3, 7, 11, 15, ...\}$, assigning all odd coefficients to $s(y)$ yields rank deficiency for $k \in \{5, 9, 13, 17, ...\}$, and the metalog 1.0 assignment in Keelin (2016) results in rank deficiency for $k \in \{7, 11, 15, 19...\}$. The only assignment that guarantees full rank is that of the metalog 2.0. The following example illustrates why assigning a_3 to $\mu(y)$ produces rank deficiency, whereas assigning a_3 to $s(y)$ guarantees full rank.

Example 1. Let $n = k = 3$, $y_2 = 0.5$, and $y_3 = 1 - y_1$. Use ℓ_i to denote $\ell(y_i)$. If a_3 is assigned to $\mu(y)$, then $\ell_2 = 0$, $\ell_1 + \ell_3 = 0$ and $y_3 - y_1 = 2(y_2 - y_1)$, so that Gaussian elimination produces

$$\mathbf{Y} = \begin{pmatrix} 1 & \ell_1 & y_1 - 1/2 \\ 1 & \ell_2 & y_2 - 1/2 \\ 1 & \ell_3 & y_3 - 1/2 \end{pmatrix} \to \begin{pmatrix} 1 & \ell_1 & y_1 - 1/2 \\ 0 & 1 & \frac{y_2 - y_1}{\ell_2 - \ell_1} \\ 0 & 0 & \frac{y_3 - y_1}{\ell_3 - \ell_1} - \frac{y_2 - y_1}{\ell_2 - \ell_1} = 0 \end{pmatrix} \to \ \text{rank}(\mathbf{Y}) = 2$$

Next, let $\hat{y}_i$ denote $y_i - 0.5$. If a_3 is assigned to $s(y)$, then

$$\mathbf{Y} = \begin{pmatrix} 1 & \ell_1 & \hat{y}_1\ell_1 \\ 1 & \ell_2 & \hat{y}_2\ell_2 \\ 1 & \ell_3 & \hat{y}_3\ell_3 \end{pmatrix} \to \begin{pmatrix} 1 & \ell_1 & \hat{y}_1\ell_1 \\ 0 & 1 & \frac{\hat{y}_2\ell_2 - \hat{y}_1\ell_1}{\ell_2 - \ell_1} \\ 0 & 0 & \frac{\hat{y}_3\ell_3 - \hat{y}_1\ell_1}{\ell_3 - \ell_1} - \frac{\hat{y}_2\ell_2 - \hat{y}_1\ell_1}{\ell_2 - \ell_1} \neq 0 \end{pmatrix} \to \ \text{rank}(\mathbf{Y}) = 3$$

The bottom-right entry cannot vanish if $y_1 < 0.5 < y_3$. Rank deficiency could still occur if the probabilities can be chosen arbitrarily, e.g., $\mathbf{y} \approx (0.01, 0.1, 0.196)$.

3.3. Moments

Our goal is to obtain the p^{th} central moment of a metalog, $\int_0^1 (M(y) - \mathbb{E}[M])^p dy$, $p \geq 1$, in closed form without resorting to algebraic integration or non-elementary expressions. The key challenge is to obtain the definite integral $I(m,u) = \int_0^1 (y-0.5)^m \ell^u(y) dy$, $m,u \geq 0$. By symmetry, note that $I(m,u)$ vanishes if m and u have different parity. The proof of the next lemma is long, and available in the online Supplement A1. As usual, we let $\sum_1^0 = 0$.

LEMMA 1. *For non-negative integers m and u having the same parity,*

$$I(m,u) = u! \sum_{n=0}^{m} \binom{m}{n} (-0.5)^{m-n} S_{n,u}, \text{ where} \tag{4}$$

$$S_{n,u} = \frac{n}{n+1} \sum_{k=1}^{u} [S_{n-1,k} - S_{n,k-1}] + \frac{1}{n+1} \sum_{k=1}^{u-1} [S_{n-1,k} - S_{n,k}] + \frac{1}{n} - \frac{1}{(n+1)^2}, \; n,u \geq 1,$$

with boundary values $S_{n,0} = \frac{1}{n+1}$ for $n \geq 0$, $S_{0,u} = 0$ for odd u, and $S_{0,u} = 2(1-2^{1-u})\zeta(u)$ for even $u \geq 2$, where the Riemann zeta function takes elementary values $\zeta(u) = \pi^u \eta_{u/2}$, with $\eta_1 = 1/6$ and

$$\eta_j = (-1)^{j+1} \frac{n}{(2j+1)!} + \sum_{i=1}^{j-1} (-1)^{i-1} \frac{\eta_{j-i}}{(2i+1)!}, \; j \geq 2.$$

PROPOSITION 3. *The k-metalog has finite moments of any order, with mean*

$$\begin{aligned} \mathbb{E}[M] &= \sum_{\substack{j=1,\ldots,k \\ j (\bmod 4)=1}} a_j I(\tfrac{j-1}{2}, 0) + \sum_{\substack{j=3,\ldots,k \\ j (\bmod 4)=3}} a_j I(\tfrac{j-1}{2}, 1) \\ &= a_1 + a_3 \tfrac{1}{2} + a_5 \tfrac{1}{12} + a_7 \tfrac{1}{12} + a_9 \tfrac{1}{80} + a_{11} \tfrac{23}{1440} + a_{13} \tfrac{1}{448} + a_{15} \tfrac{11}{3360} + \ldots \end{aligned} \tag{5}$$

and variance

$$\begin{aligned} \mathbb{V}[M] = & \sum_{\substack{1 \leq j \leq k \\ j (\bmod 4)=1}} a_j^2 (I(j-1,0) - I^2(\tfrac{j-1}{2},0)) + \sum_{\substack{1 \leq j \leq k \\ j (\bmod 4)=2}} a_j^2 I(j-2,2) \\ & + \sum_{\substack{1 \leq j \leq k \\ j (\bmod 4)=3}} a_j^2 (I(j-1,2) - I^2(\tfrac{j-1}{2},1)) + \sum_{\substack{1 \leq j \leq k \\ j (\bmod 4)=0}} a_j^2 I(j-2,0) \\ & + \sum_{\substack{1 \leq i < j \leq k \\ i,j (\bmod 4)=1}} 2a_i a_j (I(\tfrac{i+j-2}{2},0) - I(\tfrac{i-1}{2},0) I(\tfrac{j-1}{2},0)) + \sum_{\substack{1 \leq i < j \leq k \\ i,j (\bmod 4)=2}} 2a_i a_j I(\tfrac{i+j-4}{2},2) \\ & + \sum_{\substack{1 \leq i < j \leq k \\ i,j (\bmod 4)=3}} 2a_i a_j (I(\tfrac{i+j-2}{2},2) - I(\tfrac{i-1}{2},1) I(\tfrac{j-1}{2},1)) + \sum_{\substack{1 \leq i < j \leq k \\ i,j (\bmod 4)=0}} 2a_i a_j I(\tfrac{i+j-4}{2},0) \\ & + \sum_{\substack{1 \leq i < j \leq k \\ i (\bmod 4)=1 \\ j (\bmod 4)=3}} 2a_i a_j (I(\tfrac{i+j-2}{2},1) - I(\tfrac{i-1}{2},0) I(\tfrac{j-1}{2},1)) + \sum_{\substack{1 \leq i < j \leq k \\ i (\bmod 4)=2 \\ j (\bmod 4)=0}} 2a_i a_j I(\tfrac{i+j-4}{2},1) \\ & + \sum_{\substack{1 \leq i < j \leq k \\ i (\bmod 4)=3 \\ j (\bmod 4)=1}} 2a_i a_j (I(\tfrac{i+j-2}{2},1) - I(\tfrac{i-1}{2},1) I(\tfrac{j-1}{2},0)) + \sum_{\substack{1 \leq i < j \leq k \\ i (\bmod 4)=0 \\ j (\bmod 4)=2}} 2a_i a_j I(\tfrac{i+j-4}{2},1) \end{aligned}$$

$$= a_2^2\tfrac{\pi^2}{3} + a_3^2\tfrac{\pi^2+3}{36} + a_4^2\tfrac{1}{12} + a_5^2\tfrac{1}{180} + a_6^2\tfrac{\pi^2+20}{240} + a_7^2\tfrac{5\pi^2+84}{6720} + a_8^2\tfrac{1}{448} + a_9^2\tfrac{1}{3600} + \ldots$$

$$+a_2a_4 + a_2a_6\tfrac{\pi^2+12}{18} + a_2a_8\tfrac{1}{6} + a_4a_6\tfrac{1}{6} + a_4a_8\tfrac{1}{40} + a_6a_8\tfrac{23}{720} + \ldots$$

$$+a_3a_5\tfrac{1}{12} + a_3a_7\tfrac{\pi^2+10}{120} + a_3a_9\tfrac{7}{360} + a_5a_7\tfrac{13}{720} + a_5a_9\tfrac{1}{420} + a_7a_9\tfrac{1}{224} + \ldots$$

More generally, the p^{th} central moment is given by

$$\int_0^1 (M(y) - \mathbb{E}[M])^p dy = \sum_{n=0}^{p} \binom{p}{n} \left(\int_0^1 M^n(y)dy\right) (\mathbb{E}[M])^{p-n}, \text{ where}$$

$$\int_0^1 M^n(y)dy = \sum_{\substack{n_1+n_2=n \\ n_1,n_2\geq 0}} \sum_{\substack{\sum_{j\in\mu} c_j=n_1 \\ c_j\geq 0}} \sum_{\substack{\sum_{j\in s} d_j=n_2 \\ d_j\geq 0}} \frac{n!\left(\prod_{j\in\mu} a_j^{c_j}\right)\left(\prod_{j\in s} a_j^{d_j}\right)}{\prod_{j\in\mu} c_j! \prod_{j\in s} d_j!} I\left(\sum_{j\in\mu} c_j \left\lfloor \tfrac{j-1}{2}\right\rfloor + \sum_{j\in s} d_j \left\lfloor \tfrac{j-1}{2}\right\rfloor, n_2\right).$$

3.4. Partial Expectations

Consider a non-constant feasible metalog, M, and recall it possesses a density. Given $0 \leq y_0 < y_1 \leq 1$, let $x_0 = M(y_0)$ and $x_1 = M(y_1)$, $-\infty \leq x_0 < x_1 \leq \infty$, possibly defined using limits. The expected value of M conditional on $x_0 < x < x_1$ is given by

$$\mathbb{E}[M|x_0 < M(y) < x_1] = \frac{\int_{x_0}^{x_1} xm(x)dx}{M^{-1}(x_1) - M^{-1}(x_0)} = \frac{\int_{y_0}^{y_1} M(y)dy}{y_1 - y_0}.$$

The special case of $y_0 = 0$ or $y_1 = 1$ yields the one-sided conditional expectation, which is extremely useful in managerial applications such as value-at-risk analysis, pricing options, the expected sales and expected leftovers in a news-vendor problem, or in producing optimal stopping rules for search.

The challenge is to compute the partial expectation $\int_{y_0}^{y_1} M(y)dy$ in the numerator. Zrazhevsky and Zrazhevska (2021) provide expressions involving the non-elementary generalized hypergeometric function. Here, we provide an elementary expression.

Proposition 4. *For $0 \leq y_0 < y_1 \leq 1$, we have that*

$$\int_{y_0}^{y_1} M(y)dy = \sum_{j=1}^{k} a_j \left(\mathbf{1}_{j\in\mu} \int_{y_0}^{y_1} (y - \tfrac{1}{2})^{\lfloor \frac{j-1}{2}\rfloor} dy + \mathbf{1}_{j\in s} \int_{y_0}^{y_1} (y - \tfrac{1}{2})^{\lfloor \frac{j-1}{2}\rfloor} \ell(y)dy\right), \text{ where}$$

$$\int_{y_0}^{y_1} (y - \tfrac{1}{2})^m dy = \frac{(y_1 - \frac{1}{2})^{m+1} - (y_0 - \frac{1}{2})^{m+1}}{m+1} \text{ and}$$

$$\begin{aligned}\int_{y_0}^{y_1} (y - \tfrac{1}{2})^m \ell(y)dy = &\frac{[(y_1 - \frac{1}{2})^{m+1} - (-\frac{1}{2})^{m+1}]\ln(y_1)}{m+1} - \frac{[(y_0 - \frac{1}{2})^{m+1} - (-\frac{1}{2})^{m+1}]\ln(y_0)}{m+1} \\ &- \frac{[(y_1 - \frac{1}{2})^{m+1} - (\frac{1}{2})^{m+1}]\ln(1-y_1)}{m+1} + \frac{[(y_0 - \frac{1}{2})^{m+1} - (\frac{1}{2})^{m+1}]\ln(1-y_0)}{m+1} \\ &+ \frac{(1/2)^{m+1}}{m+1} \sum_{r=1}^{m+1} \frac{(2y_1-1)^r - (2y_0-1)^r}{r} - \frac{(-1/2)^{m+1}}{m+1} \sum_{r=1}^{m+1} \frac{(1-2y_1)^r - (1-2y_0)^r}{r},\end{aligned}$$

with the understanding that if $y_0 = 0$ or $y_1 = 1$, then $0 \cdot \ln(0) = 0$ (by taking the limit).

4. Feasibility

Figure 2 displays the OLS fit of a 5- and a 7-metalog to the same data seen in Figure 1. The 5-metalog upper tail is infeasible, and the 7-metalog has an infeasible lower tail and two infeasible Runge oscillations. The data is non-pathological, demonstrating that infeasibility can arise naturally. Our next goal is to diagnose and enforce feasibility, and hopefully restore a nice metalog fit to the data.

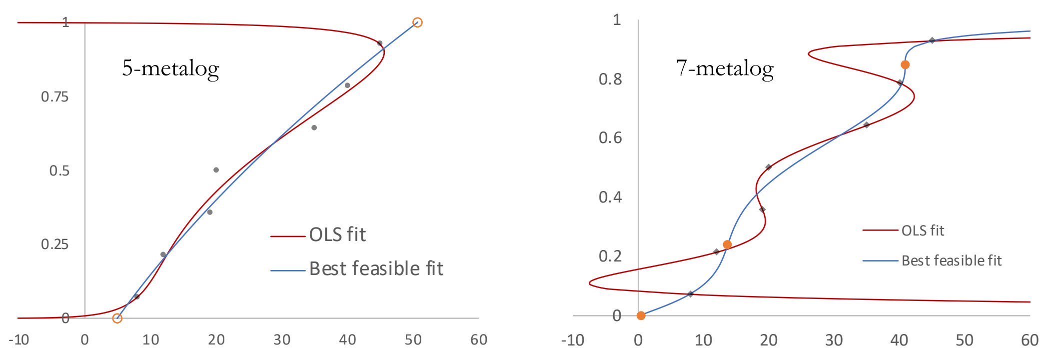

Figure 2 The ols fit of a 5- and 7-metalog to the data in Figure 1 produces an infeasible metalog (red line). Algorithm 2 will find the best feasible fit (blue line).

4.1. Diagnosing Feasibility and Identifying the Modes and Anti-Modes

Prior to our current work, infeasibility was dealt with by checking/enforcing $M'(y) \geq 0$ on a finite grid, say $y_i = i/(N+1)$, $i = 1, ..., N$. This approach is deficient because in between grid points the slope may turn negative. And near the edges, the tail may appear feasible, but actually become infeasible for values of y very close to 0 or to 1. This calls for a mathematically precise and computationally practical way to diagnose feasibility. Because limits preserve weak inequalities, a feasible metalog must necessarily be *tail feasible*, i.e., satisfy $\lim_{y\downarrow 0} M'(y) \geq 0$ and $\lim_{y\uparrow 1} M'(y) \geq 0$.

PROPOSITION 5. *A metalog is tail feasible if and only if*

$$\left\langle \begin{array}{l} s(0) > 0, \;\; or \\ s(0) = 0 \;\; and \;\; -s'(0) > 0, \;\; or \\ s(0) = s'(0) = 0 \;\; and \;\; \mu'(0) \geq 0 \end{array} \right\rangle \;\; and \;\; \left\langle \begin{array}{l} s(1) > 0, \;\; or \\ s(1) = 0 \;\; and \;\; s'(1) > 0, \;\; or \\ s(1) = s'(1) = 0 \;\; and \;\; \mu'(1) \geq 0 \end{array} \right\rangle .$$

And feasible if and only if, in addition, the slope at each inflection point on $(0,1)$ *is non-negative.*

Figure 3 visualizes the different feasibility conditions. Given $\mathbf{a}$, we can readily compute $s(0)$, $s(1)$, $s'(0)$, $\mu'(0)$... and check for tail feasibility. The condition that the slope at each inflection point be non-negative rules out Runge oscillations, but it requires us to find the roots of $M''(y)$. A simple approach to rapidly find the inflection points of a metalog is by means of a grid search, and identify the consecutive grid points where $M''(y)$ changes signs, and then pin-point the root within such interval. This grid approach is not air-tight because there could be an odd (even) number of inflection points on the intervals that (do not) change sign.

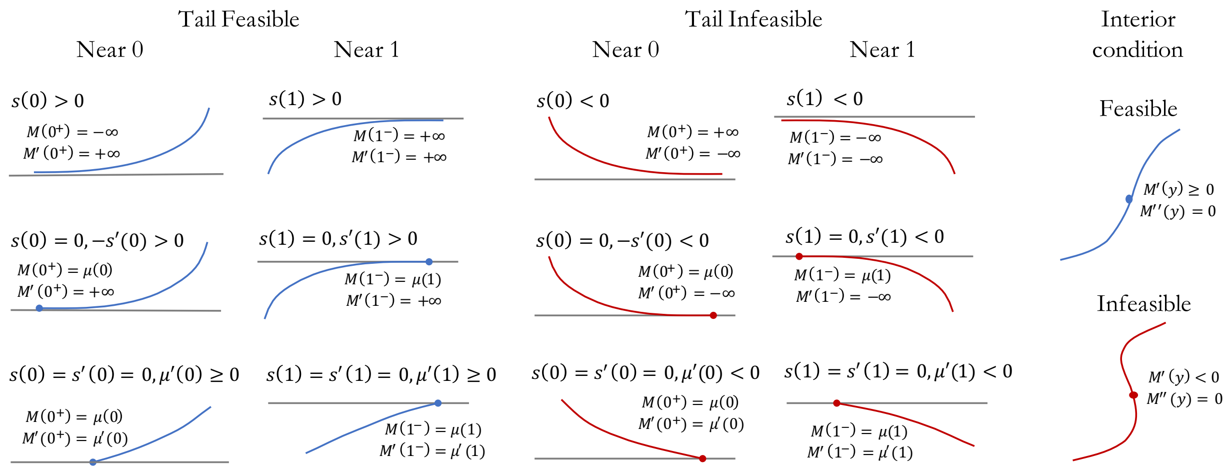

Figure 3 cdf Visualization of the Feasibility Conditions.

The alternative we propose is Algorithm 1, an air-tight procedure to find all inflection points of a metalog to arbitrary precision. It is based on the insight that for $i = \lfloor \frac{k+1}{2} \rfloor$, the function $y^i(1-y)^i M^{(i)}(y)$ is a polynomial of degree $i-1$ (see Table 2 in the Appendix), whose roots on $(0,1)$ can therefore be identified, and coincide with those of $M^{(i)}(y)$. The integral function $M^{(i)}(y)$ can only have one root in between two consecutive roots of $M^{(i)}(y)$. This allows us to use bisection followed by Newton and pin-point the roots of $M^{(i-1)}(y)$, then the roots of $M^{(i-2)}(y)$, until reaching $M''(y)$. Once done, it is straightforward to check that $M'(y) \geq 0$ whenever $M''(y) = 0$.

4.2. Cones of Feasible Coefficients

Next, we explore the geometric shape of the set of feasible metalogs. Mathematically, any k-metalog corresponds with a point $(a_1, a_2, ..., a_k)$ in $\mathbb{R}^k$, and vice-versa. We can therefore identify the space of all k-metalogs with $\mathbb{R}^k$. Our immediate goal is to determine the subset of coefficients describing all feasible k-metalogs, $\mathcal{A}_k \subset \mathbb{R}^k$, or *feasible set* for short. The following result confirms that $\mathcal{A}_k$ is a closed and convex cone, and applies Propositions 1 and 5 to describe such set for k up to 4. The online supplement A2 describes the cone $\mathcal{A}_5$.

Proposition 6. *For $k \geq 2$, the feasible set $\mathcal{A}_k$ is a convex and closed cone in $\mathbb{R}^k$. Moreover,*

$$\mathcal{A}_4 = \{\mathbf{a} \in \mathbb{R}^4 : \langle a_2 = a_3 = 0, a_4 \geq 0\rangle \text{ or } \langle a_2 > 0, -2a_2 < a_3 < 2a_2, a_4 \geq a_3 \ln \tfrac{2a_2+a_3}{2a_2-a_3} - 4a_2\rangle\}.$$

It follows that $\mathcal{A}_3 = \{(a_1,a_2,a_3) \in \mathbb{R}^3 : a_2 \geq 0, -\theta a_2 \leq a_3 \leq \theta a_2\}$, where $\theta \approx 1.667113$ is the unique positive solution of $\theta \ln \frac{2+\theta}{2-\theta} = 4$, and that $\mathcal{A}_2 = \{(a_1,a_2) \in \mathbb{R}^2 : a_2 \geq 0\}$.

To obtain the feasible cones requires us to find all the inflection points, and impose their slope to be non-negative. The case of $k \in \{3,4\}$ is straightforward because the one mode is at $y = 0.5 - 0.25a_3/a_2$. For $k = 5$, the up to three inflection points and two modes can be found as solutions of a quartic. For k larger than 5, however, the inflection points can only be obtained numerically.

Algorithm 1 Air-tight identification of the inflection points of a given metalog.

Require: coefficients $\mathbf{a} = [a_1, a_2, \ldots, a_k]$, $k \geq 3$.

1: $i \leftarrow 2, j \leftarrow \left\lfloor \frac{k+1}{2} \right\rfloor$
2: $\mathcal{R}^{(j)} = \{r_n^{(j)}\}_{n=1}^{\#(j)} \leftarrow$ The roots on $(0,1)$ of the polynomial $y^j(1-y)^j M^{(i)}(y)$ in (7)
3: **while** $j > i$ **do**
4: $\quad j \leftarrow j-1,\ \mathcal{R}^{(j)} = \emptyset, r_0^{(j+1)} \leftarrow 0, r_{\#(j+1)+1}^{(j+1)} \leftarrow 1, I(0) \leftarrow \operatorname{sign}\big(\lim_{y\to 0} M^{(j)}(y)\big)$
5: $\quad$ **for** $n = 0, 1, \ldots, \#(j+1)$ **do**
6: $\quad\quad$ **if** $n = \#(j+1)$ **then** $I(n+1) \leftarrow \operatorname{sign}\big(\lim_{y\to 1} M^{(j)}(y)\big)$ **else** $I(n+1) \leftarrow \operatorname{sign}\Big(M^{(j)}\big(r_{n+1}^{(j+1)}\big)\Big)$
7: $\quad\quad$ **if** $I(n) \cdot I(n+1) < 0$ **then**
8: $\quad\quad\quad \mathcal{R}^{(j)} \leftarrow \mathcal{R}^{(j)} \cup \{\text{The root of } M^{(j)} \text{ on } \big(r_n^{(j+1)}, r_{n+1}^{(j+1)}\big)\}$ ▷ Bisection & Newton
9: $\quad\quad$ **else if** $I(n) = 0$ **and** $n > 0$ **then** $\mathcal{R}^{(j)} \leftarrow \mathcal{R}^{(j)} \cup \{r_n^{(j+1)}\}$
10: $\quad\quad$ **end if**
11: $\quad$ **end for**
12: **end while**
13: **return** The inflection points $\{r_n^{(i)}\}_{n=1}^{\#(i)}$

Notes: Line 2 relies on a subroutine that finds the real roots of a polynomial. In Lines 4 and 6, use Lemma 2 to obtain the boundary signs $I(0)$ and $I(n+1)$. In Line 8, the derivatives to apply Newton are given in (6).

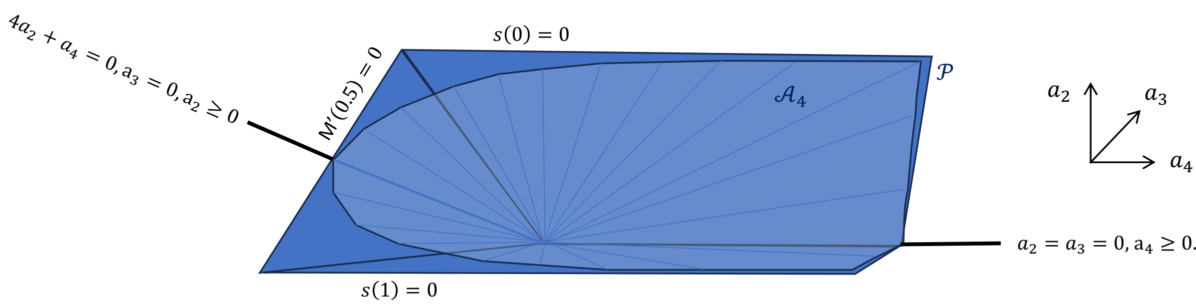

Figure 4 **The feasible set $\mathcal{A}_4$ is a closed and convex cone wrapped by $\mathcal{P} = \{\mathbf{a} \in \mathbb{R}^4 : s(0), s(1), M'(0.5) \geq 0\}$.**

4.3. Restrictions in the location of the Modes

When examining the feasible cones we learn that the location of the modes may be restricted. A 2-metalog is symmetric, hence its mode is at $y = 0.5$. A 3-metalog allows for skewness, but its mode must be located inside $[0.5 - 0.25\theta, 0.5 + 0.25\theta] \approx [0.0832, 0.9168]$. In other words, a feasible 3-metalog has at least 8.32% of mass to the left or to the right of its mode. For $k = 4$ feasibility is compatible with the one mode anywhere on $(0,1)$. For $k = 5$, the metalog could be bi-modal, but in this case precisely one mode must be inside the interval $(\frac{3-\sqrt{3}}{6}, \frac{3+\sqrt{3}}{6}) \approx (0.2113, 0.7887)$.

4.4. The Best Feasible Fit

Consider the cone $\mathcal{A}_k$ and the data-dependent point $\mathbf{a}^{\text{OLS}}$, both sitting in $\mathbb{R}^k$. If $\mathbf{a}^{\text{OLS}} \notin \mathcal{A}_k$, then the goal is to find the *best feasible fit*,

$$\mathbf{a}^* = \underset{\mathbf{a}\in\mathcal{A}_k}{\arg\min} f(\mathbf{a}), \text{ where } f(\mathbf{a}) = (\mathbf{x}-\mathbf{Ya})^{\mathsf{T}}(\mathbf{x}-\mathbf{Ya}).$$

The objective $f(\mathbf{a})$ can be rewritten as a constant plus the term $(\mathbf{a}-\mathbf{a}^{\text{OLS}})^{\mathsf{T}}\mathbf{Y}^{\mathsf{T}}\mathbf{Y}(\mathbf{a}-\mathbf{a}^{\text{OLS}})$. Geometrically, this implies that the iso-distance metalogs conform to the surface of an ellipsoid centered at $\mathbf{a}^{\text{OLS}}$ and with shape matrix $[\mathbf{Y}^{\mathsf{T}}\mathbf{Y}]^{-1}$ (Boyd and Vandenberghe 2004, §2.2.2). Altering the distance until the ellipsoid makes unique tangent contact with the boundary of $\mathcal{A}_k$ yields $\mathbf{a}^*$.

Proposition 7. *The best feasible fit, $\mathbf{a}^*$, exists and is unique.*

Recall that $\theta \approx 1.667113$. For $k=3$, the best feasible fit admits an explicit solution[4]

$$\mathbf{a}^* = \begin{cases} \mathbf{a}^{\text{OLS}}, \text{ if } -\theta a_2^{\text{OLS}} \le a_3^{\text{OLS}} \le \theta a_2^{\text{OLS}}, \\ \mathbf{a}^{\text{OLS}} - [\mathbf{Y}^T\mathbf{Y}]^{-1}\mathbf{C}^{\mathsf{T}}[\mathbf{C}\ [\mathbf{Y}^{\mathsf{T}}\mathbf{Y}]^{-1}\mathbf{C}^{\mathsf{T}}]^{-1}\mathbf{C}\mathbf{a}^{\text{OLS}} \text{ with } \mathbf{C} = (0,1,-1/\theta), \text{ if } a_3^{\text{OLS}} > \theta a_2^{\text{OLS}}, \text{ and} \\ \mathbf{a}^{\text{OLS}} - [\mathbf{Y}^T\mathbf{Y}]^{-1}\mathbf{C}^{\mathsf{T}}[\mathbf{C}\ [\mathbf{Y}^{\mathsf{T}}\mathbf{Y}]^{-1}\mathbf{C}^{\mathsf{T}}]^{-1}\mathbf{C}\mathbf{a}^{\text{OLS}} \text{ with } \mathbf{C} = (0,1,1/\theta), \text{ if } a_3^{\text{OLS}} < -\theta a_2^{\text{OLS}}. \end{cases}$$

For $k \ge 4$, we can only approximate $\mathbf{a}^*$ with arbitrary precision.

4.5. Reformulation as a Semi-Infinite Program

By its nature, the original feasibility problem $\min_{\mathbf{a}\in\mathbb{R}^k} f(\mathbf{a})$ s.t. $M'(y) \ge 0,\ y \in (0,1)$, involves a finite number of decision variables and an infinite number of constraints. Thus, the appropriate approach is to use semi-infinite programming (SIP). To enjoy the convergence to the best optimal fit using the exchange method (Hettich and Kortanek 1993, Theorem 7.2), however, requires the set of constraints to be continuous in y on a compact set. To achieve this, let

$$G(y) = \begin{cases} y(1-y)M'(y),\ y \in (0,1), \text{ and} \\ s(y),\ y \in \{0,1\}. \end{cases}$$

We verify that $G:[0,1]\to\mathbb{R}$ is continuous (to see this, apply L'Hôpital to obtain $\lim_{y\to 0,1} y(1-y)\ell(y) = 0$). Because limits preserve weak inequalities, we have that $M'(y) \ge 0$ on $(0,1)$ holds if and only if $G(y) \ge 0$ on $[0,1]$ does. Hence, the best feasible fit problem can be reformulated as

$$\mathbf{a}^* = \underset{\mathbf{a}\in\mathbb{R}^k}{\arg\min} f(\mathbf{a}) \text{ subject to } G(y) \ge 0,\ y \in [0,1].$$

[4] First, note that the feasible set is the polyhedron

$$\mathcal{A}_3 = \{\mathbf{a} \in \mathbb{R}^3 : M'(\underbrace{0.5-0.25\theta}_{\approx 0.0832}) \ge 0, M'(\underbrace{0.5+0.25\theta}_{\approx 0.9168}) \ge 0\}.$$

Thus, if the unconstrained OLS program is not feasible, then the best feasible fit can be found by running the remaining three constrained combinations: imposing $M'(0.0832)=0$, imposing $M'(0.9168)=0$, and imposing both. Each of these programs have an explicit solution using equality constraint OLS (Theil and Van de Panne 1960, p. 5). Imposing both equality constraints is unnecessary, as it produces the constant metalog. The constant metalog can only be optimal if $\mathbf{x}$ is constant, hence would have already been identified by the unconstrained $\mathbf{a}^{\text{OLS}}$ program.

We now argue that this program needs to be slightly modified due to a caveat of numerical nature. Indeed, when solving this SIP one may encounter a solution exhibiting $s(0)=0$ and $s'(0)>0$, or $s(1)=0$ and $s'(1)<0$, hence infeasible. To see that G must be negative on some interval $(0,\delta)$, or $(1-\delta,1)$, note that

$$\lim_{y\to 0} G'(y) = \begin{cases} +\infty \text{ if } s'(0)<0, \\ \mu'(0) \text{ if } s'(0)=0, \\ -\infty \text{ if } s'(0)>0; \end{cases} \quad \text{and} \quad \lim_{y\to 1} G'(y) = \begin{cases} +\infty \text{ if } s'(1)<0, \\ -\mu'(1) \text{ if } s'(1)=0, \\ -\infty \text{ if } s'(1)>0. \end{cases}$$

Numerically, however, δ may be smaller than the precision of the computer.[5] A viable way to approximate $\mathbf{a}^*$ within the precision of the computer is to solve

$$\mathbf{a}^*(\epsilon) = \arg\min_{\mathbf{a}\in\mathbb{R}^k} f(\mathbf{a}) \text{ subject to } G(y) \geq \epsilon,\ y \in [0,1].$$

Because $G(y)$ is continuous and linear with respect to the decision variables, and the objective is strictly convex, we have that $\mathbf{a}^*(\epsilon)$ is continuous in ϵ, hence the coefficients resulting from this SIP must be very close to $\mathbf{a}^*$, provided ϵ is small. Specifically, we meet the requirements of Dinh et al. (2012, Corollary 8.3), who establish that the optimal set correspondence is upper semi-continuous in ϵ. By strict convexity, such mapping is single-valued, hence continuous. It follows that $f(\mathbf{a}^*(\epsilon))$ is also continuous in ϵ, hence approximately equal to $f(\mathbf{a}^*)$.

4.6. The Algorithm

We now propose Algorithm 2, which is guaranteed to produce a feasible metalog that approximates $\mathbf{a}^*$ to arbitrary precision. The algorithm builds on SIP's exchange method, with some added features to deal with the nuances of the metalog. The python implementation is available at `https://github.com/Stephenxuu/metalog_algorithm.git`.

Line 6 of Algorithm 2 conducts an expensive air-tight test of interior feasibility using Algorithm 1. We only run Algorithm 1 whenever Line 5 fails to identify local minima with $G(y)<0$. Line 5 applies a fast, but not necessarily air-tight, search for the local minima of $G(y)$ by means of a grid. In our implementation, we set the grid at $\{5\cdot 10^{-i}, 10^{-i}, i=3,...,15\} \cup \{0, 0.01, 0.02, 0.03, ..., 0.98, 0.99, 1\} \cup \{1-5\cdot 10^{-i}, 1-10^{-i}, i=3,...,15\}$. If G's value at some grid point is less than or equal to that of its neighbors, then run bisection or Newton to get arbitrarily close to the point where $G'(y)=0$. The local minima must include 0 if $G(0)=s(0)<G(10^{-15})$; and include 1 if $G(1)=s(1)<G(1-10^{-15})$. This grid search is not air-tight, as G could take negative values in between grid points. Algorithm

[5] Should one then impose additional lexicographic tail constraints such as $s'(0)\leq 0$, or $s'(1)\geq 0$? The answer is no, because imposing these additional tail conditions will produce a feasible, but sub-optimal, solution. Indeed, the additional constraint induces a 'large' jump in the value of objective. By imposing $s(0), s(1)\geq \epsilon$ instead, we recover feasibility while producing a 'small' jump in the objective. In fact, the tail constraints down the lexicographic list, namely $s'(0)\leq 0$, $s'(1)\geq 0$, $\mu'(0)\geq 0$ and $\mu'(1)\geq 0$, are redundant.

Algorithm 2 Finding the best feasible fit

Require: $\mathbf{x}, \mathbf{y}$, $k \geq 4$, $\epsilon > 0$

1: $\mathcal{Y}^{(0)} \leftarrow \{\emptyset\}$, $S = \text{FALSE}$, $j \leftarrow 0$
2: **while** $S = \text{FALSE}$ **do**
3: $\quad \mathbf{a}^{(j)} \leftarrow \arg\min_{\mathbf{a} \in \mathbb{R}^k} f(\mathbf{a})$ s.t. $G(y) \geq \epsilon$, $y \in \mathcal{Y}^{(j)}$. ▷ Can add other linear constraints.
4: $\quad j \leftarrow j + 1$,
5: $\quad \mathcal{Y}^{(j)} \leftarrow \{y_1^{(j)}, ..., y_r^{(j)} \in [0,1] : y$ is a local minimum of G and $G(y) < 0\}$. ▷ Use a grid.
6: $\quad$ **if** $\mathcal{Y}^{(j)} = \{\emptyset\}$ **then** $\mathcal{Y}^{(j)} \leftarrow \{y_1^{(j)}, ..., y_r^{(j)} \in (0,1) | M''(y) = 0, M'(y) < 0\}$. ▷ Use Algorithm 1.
7: $\quad$ **if** $\mathcal{Y}^{(j)} = \{\emptyset\}$ **then** $S \leftarrow \text{TRUE}$, $\mathbf{a}^* \leftarrow \mathbf{a}^{(j)}$ **else** $\mathcal{Y}^{(j)} \leftarrow \mathcal{Y}^{(j-1)} \cup \mathcal{Y}^{(j)}$.
8: **end while**
9: **return** $\mathbf{a}^*(\epsilon)$

1 would then identify those points, hence fully guarantee feasibility without being too expensive. Thus, the algorithm only stops in Line 7 if all inflection points have non-negative slope. And that $G(0) = s(0) \geq \epsilon$ and $G(1) = s(1) \geq \epsilon$ ensure tail feasibility. Hence, the outputted metalog must be feasible.

Algorithm 2 can be made very efficient because some calculations need to be done just once. For example, the matrix $\mathbf{Y}$ and the inverse $[\mathbf{Y}^\intercal \mathbf{Y}]^{-1}$ need to be computed only once. Similarly, we can compute the grid values to identify the local minima of G using $\mathbf{G}\mathbf{a}^{(j)}$, where the basis matrix $\mathbf{G}$ of the grid needs to be computed only once. In the key optimization step in Line 3, note that for $j \geq 1$, $\mathbf{a}^{(j)}$ minimizes a quadratic objective under a finite set of linear inequality constraints. Thus, a standard quadratic programming (QP) solver can find its unique solution in a finite number of steps (Bertsekas 2016, p. 303).

Having $\epsilon > 0$ also ensures that the algorithm terminates in finite time. Because $G(y)$ is continuous, Theorem 7.2 in Hettich and Kortanek (1993) guarantees that the exchange method produces a sequence of solutions converging to $\mathbf{a}^*(\epsilon)$ and meeting $G(y) \geq \epsilon$. Thus, a solution meeting the less stringent stopping criterion of $G(y) \geq 0$ will necessarily arise in a finite number of steps. In our numerous trials using $\epsilon = 10^{-6}$, and for k as large as 16, Algorithm 2 typically terminates in 10 iterations or less, with a manageable list of constraints, and a run time of less than one second.

The solution can have at most $\lfloor \frac{k-1}{2} \rfloor$ binding points. If $G(y) \approx 0$ is a binding local minima in the interior, then it must exhibit $G'(y) = 0$. Together, this implies that $M''(y) \approx 0$ and $M'(y) \approx 0$. Hence, interior binding points produce modes where the cdf has infinite slope, and the pdf is unbounded yet integrable. The other points that could be binding are $s(0) = 0$ and $s(1) = 0$.

4.7. Illustrative Example of Algorithm 2

Let's describe how Algorithm 2 produces the feasible fits of a 5- and a 7-metalog shown in Figure 2. In both cases, the initial OLS fit is infeasible hence $G^{\text{OLS}}(y)$ takes negative values.

For $k=5$, the algorithm converges in two steps. The set of negatively valued local minima of G^{OLS} are $\mathcal{Y}^{(1)}=\{0.99999812, 1\}$. After forcing $G(y)\geq\epsilon$ at these two points, the resulting G still possesses a negatively valued local minima at $\mathcal{Y}^{(2)}=\{0\}$. Then, we impose $G(y)\geq\epsilon$ at $y\in\mathcal{Y}^{(1)}\cup\mathcal{Y}^{(2)}$, resulting in a feasible metalog with two binding points, $s(0)=0$ and $s(1)=0$. For a 5-metalog this implies that $a_2=a_3=0$, producing the quadratic $M(y)=24.47+45.75(y-0.5)+13.42(y-0.5)^2$ with support $(\mu(0),\mu(1))=(4.955, 50.71)$ seen in Figure 2, left.

For $k=7$, the algorithm converges in three steps. Initially, $\mathbf{a}^{\text{OLS}}$ exhibits an infeasible lower tail and two inflection points with negative slope. This is reflected in G^{OLS} having three negatively valued local minima at $\mathcal{Y}^{(1)}=\{0, 0.372, 0.829\}$. We run a QP forcing $G(y)\geq\epsilon$ at $y\in\mathcal{Y}^{(1)}$. The resulting G still has two negatively valued local minima at $\mathcal{Y}^{(2)}=\{0.000225, 0.844\}$. Next, we force $G(y)\geq\epsilon$ at $y\in\mathcal{Y}^{(1)}\cup\mathcal{Y}^{(2)}$, producing two negatively valued local minima at $\mathcal{Y}^{(3)}=\{0.0001943, 0.846039\}$. Finally, we run a QP forcing $G(y)\geq\epsilon$ at $y\in\mathcal{Y}^{(1)}\cup\mathcal{Y}^{(2)}\cup\mathcal{Y}^{(3)}$. The resulting G function has three local minima, all with non-negative values. To ensure $G(y)\geq 0$ on $[0,1]$ we run our air tight feasibility test that finds all the inflection points and verify that their slope is non-negative. The test passes and the algorithm terminates. As seen in Figure 2, right, the best feasible 7-metalog fit possesses 3 interior modes, two of them binding, and with one of the binding modes ($y_1=0.00019$) being quite close to zero.

4.8. Control of the Mean

Suppose we want the metalog to possess a pre-specified mean, b_m. The value of b_m could be the mean of the data, $\bar{\mathbf{x}}$, or some known theoretical value (e.g., to fit the residuals of a regression model, we may want to force the mean to zero).

Because the mean in (5) is linear in the coefficients, the requirement $\mathbb{E}[M]=b_m$ can be written as $\mathbf{Ca}=\mathbf{d}$, where $\mathbf{C}=(1,0,1/2,0,1/12,0,1/12,0,1/80,0,...)$ and $\mathbf{d}=(b_m)$. In the first pass of the algorithm, instead of $\mathbf{a}^{(0)}=\mathbf{a}^{\text{OLS}}$, the metalog with minimum distance $f(\mathbf{a})$ that also meets $\mathbf{Ca}=\mathbf{d}$ is given by (Theil and Van de Panne 1960, p. 5)

$$\mathbf{a}^{(0)}=\mathbf{a}^{\text{OLS}}-[\mathbf{Y}^T\mathbf{Y}]^{-1}\mathbf{C}^{\mathrm{T}}\boldsymbol{\Lambda},\ \text{where } \boldsymbol{\Lambda}=[\mathbf{C}\ [\mathbf{Y}^{\mathrm{T}}\mathbf{Y}]^{-1}\mathbf{C}^{\mathrm{T}}]^{-1}\left(\mathbf{Ca}^{\text{OLS}}-\mathbf{d}\right).$$

If this program fails to be feasible, then Algorithm 2 works the same, provided we always add the equality constraint $\mathbf{Ca}=\mathbf{d}$ in Line 3 of the algorithm. The final result will be the best feasible fit among the metalogs having mean b_m.

4.9. Control of the Support

Many variables have natural lower and upper bounds. E.g., prices and lengths have a zero lower bound, and percentages are often bounded between zero and one. Directly accommodating this situation was the motivation for the metalog transforms in Keelin (2016, §4). We now show how to stay within the untransformed metalog realm while controlling the interval support.

To illustrate, suppose we seek to set the support to (b_l, b_u). Recall that if $s(0) = 0$ and $s(1) = 0$, then the end points of the metalog are $\mu(0)$ and $\mu(1)$, respectively. Hence, if feasible, such metalog will have support $(\mu(0), \mu(1))$. Thus, to enforce (b_l, b_u) as the support, it suffices to run Algorithm 2 with Line 3 having the added constraints that $s(0) = \epsilon$, $\mu(0) = b_l$, $s(1) = \epsilon$, and $\mu(1) = b_u$. In post-processing, the parameters of the s polynomial could be tweaked to produce $s(0), s(1) = 0$. Note that these four constraints are linear in the coefficients, and can be written as $\mathbf{Ca} = \mathbf{d}$, where

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & -0.5 & 0 & 0 & 0.25 & -0.125 & 0 \\ 1 & 0 & 0 & -0.5 & 0.25 & 0 & 0 & -0.125 \\ 0 & 1 & 0.5 & 0 & 0 & 0.25 & 0.125 & 0 \\ 1 & 0 & 0 & 0.5 & 0.25 & 0 & 0 & 0.125 \end{pmatrix} \text{ and } \mathbf{d} = \begin{pmatrix} \epsilon \\ b_l \\ \epsilon \\ b_u \end{pmatrix}.$$

If one seeks the semi-bounded support (b_l, ∞), then only the first two rows of $\mathbf{C}$ and $\mathbf{d}$ must be used; and to impose support $(-\infty, b_u)$, only the last two rows must be used. The final result will be the best feasible fit among the metalogs having the desired support. Of course, adding constraints reduces the flexibility, and one may want to increase k to compensate for that.

4.10. Feasible Fit of Transformed Metalogs

An alternative way to set the support to (b_l, ∞), $(-\infty, b_u)$, or (b_l, b_u) is by means of transforms. Specifically, Keelin (2016, §4) proposes $b_l + e^M$, $b_u - e^{-M}$, and $\frac{b_l + b_u e^M}{1 + e^M}$, respectively. These transforms increase shape flexibility while retaining a smaller k (Keelin 2016, Figure 7). Unfortunately, the transform of a metalog generally lacks a closed form expression for the moments and the partial expectation, and does not necessarily preserve the number of modes.

Feasibility, however, is preserved under transformations. Indeed, for some $-\infty \le b_l < b_u \le \infty$ let $h : \mathbb{R} \to (b_l, b_u)$ be continuous, strictly increasing, and differentiable. By the chain rule, $dh(M(y))/dy = h'(M(y))M'(y)$, we have that $h(M(y))$ is feasible if and only if $M(y)$ is feasible. By implication, $h(M(y))$ and $M(y)$ share the same feasible set $\mathcal{A}_k$. Thus, if a metalog passes the feasibility test, then so do any of its monotonic transforms.

Finally, Algorithm 2 can successfully estimate a feasible metalog transform, provided the data is consistent with the intended support, i.e., $\mathbf{x} \subset (b_l, b_u)$. As done in linear regression, the idea is to find the best feasible fit M to the transformed data $(h^{-1}(\mathbf{x}), \mathbf{y})$ by means of Algorithm 2. Then, let $h(M)$ be the quantile function that fits the original data.[6] For example, to fit e^M to $(\mathbf{x}, \mathbf{y})$, we set M to be the best feasible fit to the data $(\ln \mathbf{x}, \mathbf{y})$.

[6] In purity, one should minimize the distance between $\mathbf{x}$ and $h(M(y))$, but this approach is very impractical because this objective may fail to be convex.

5. Practical Examples

5.1. Detecting a Clinically Relevant Mode with a 16-metalog

Clinical results of post-kidney-transplant for 246 children with chronic kidney disease are shown as black dots in Figure 5, where numbers on the x-axis are the annual rate of change of the log-scaled glomerular filtration rate (Bae et al. 2023).

While the OLS-feasible 5-metalog provides a reasonably good cdf fit to this data, an important subtlety remains hidden: a small mode on the far left near -0.28, inherent in the data, that corresponds to the clinically relevant fraction of children for whom the transplant was unsuccessful. To portray these data with visual and clinical accuracy, this mode must be represented. Metalogs with $k \geq 12$ are a viable way of doing so.

Setting $k = 16$ produces an OLS-infeasible metalog. Algorithm 2 amends the problem and yields the pdf seen in Figure 5, right, with the desired mode near -0.28, missing for $k = 5$. The 16-metalog also identifies a small accumulation in the data near -0.08.

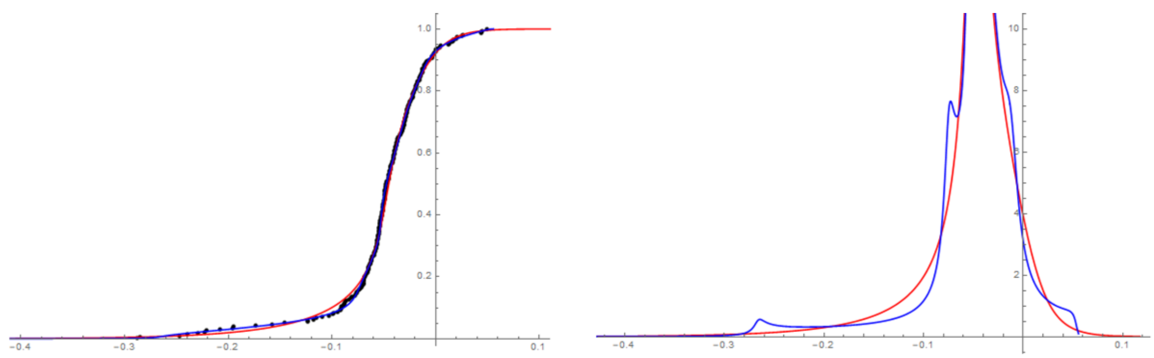

Figure 5 **Left: cdf data of the children's kidney function rate of change post kidney transplant (black dots, $N = 246$), the best fit of a 5-metalog (red line), and a 16-metalog (blue line). Right: pdf of the two metalog fits.**

5.2. R&D Project: Fitting a Mixed Distribution

How well can feasible metalogs approximate the cdf of a mixed distribution? Consider an R&D project with a 80% probability of technical failure, yielding a payout of 0. In the case of success, the result of commercialization has a 15% probability of an attractive upside and a 5% probability of a modest downside, both distributed continuously.

We employ 100 data points to model the present value distribution for the project (see black dots in Figure 6). A metalog fit to this data is OLS-infeasible for all $k = 3, \ldots, 16$. Applying Algorithm 2 results in increasingly tighter feasible fits, as shown for $k \in \{4, 10, 16\}$ in Figure 6, left. The tighter fits are enabled, in part, by the presence of additional binding modes. By zooming-in on the x-axis to the narrower range of ± 3, Figure 6, middle, shows that the optimal fit for $k = 4, 10$, and 16 exhibits 1, 2, and 4 modes, respectively.

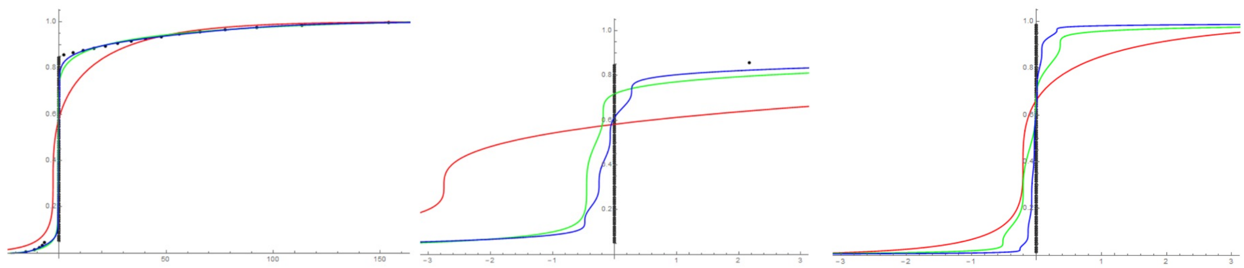

Figure 6 Left: cdf of an R&D project having 80% chance of failure, and the best fit of a 4- (red), 10- (green), and 16-metalog (blue). Middle: near 0 the metalog uses multiple modes to fit the cdf jump. Right: If the probability of failure increases to 98%, then the metalog fit uses all its available modes (1, 4, and 7, respectively) in order to approximate this big jump.

Taking the above example further, we consider the case of a larger vertical segment by reducing the technical success probability from 20% to 2% by setting $x_1 = -5$, $x_{100} = 25$, and $x_i = 0$ otherwise. Again, this data set is OLS-infeasible for all $k = 3, ..., 16$. Algorithm 2 continues to yield optimal solutions that are tight cdf fits, as shown in Figure 6, right. In this case, the optimal fit for $k = 4$, 10, and 16 employs the maximum number of possible modes, 1, 4, and 7, respectively.

5.3. Developing a New Product

Our final example is a real life application where lack of feasibility posed a serious impediment to using the metalog. Figure 7 shows disguised NPV data from an actual analysis of the prospects for developing a new product, an improved oven. The decision analysts modeled this using 24 scenarios with various probabilities, yielding the cdf data (black dots) in the figure. For subsequent simulation analysis and display, the decision analysts desired a continuous distribution to represent this uncertainty. For $k \leq 4$, the OLS fit is feasible, but insufficiently accurate. For $5 \leq k \leq 16$, the OLS fit is infeasible. By employing Algorithm 2, increasingly accurate feasible fits are achieved with increasing k. The fits for $k = 10, 16$ are illustrated, which satisfied the need.

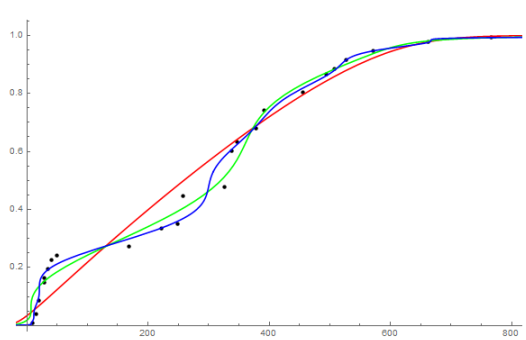

Figure 7 Metalog Fits to Data using $k = 4$ (red), $k = 10$ (green) and $k = 16$ (blue).

6. Conclusions

In this paper, we have explored the properties of the metalog family of distributions, focusing on their modes, moments, partial expectations, and feasibility. We provide elementary expressions for the moments and partial expectations, and establish that metalogs with k parameters can have up to $\lfloor \frac{k-1}{2} \rfloor$ modes. Having pinned down the maximum number of modes provides a good sense of how flexible the metalog becomes as k increases. This flexibility allows metalogs to model complex, multi-modal data effectively. We also provide a new method to control the upper or lower bounds of the distribution, yielding an alternative to introducing transformations.

Regarding the critical issue of feasibility—i.e., ensure that the metalog function is increasing—we have introduced a mathematically precise and computationally efficient method to diagnose feasibility and an algorithm to find the best feasible fit to data. This resolves a significant challenge in the practical application of metalogs, making them more accessible for statistical modeling and decision analysis. Our numerical experiments indicate that the convergence rate of the best fit metalog towards a target distribution is very fast, prompting us to conjecture it is exponential. With feasibility ensured, it is straightforward to model dependencies and create multi-variate metalog distributions using standard copula methods (Perepolkin et al. 2025, §4).

In this paper, the selection of the order of the polynomials, k, is left as a choice of the analyst. It could also be made data-driven via cross-validation, the use of information criterion AIC or BIC, or regularization. Ridge regularization in particular preserves the quadratic nature of the objective. Algorithm 2 would then find the best regularized feasible fit.

Another line of inquiry is to explore the properties of quantile functions with kernels other than logistic (Keelin and Powley 2011). The latter task appears quite challenging, however, given that many of our results hinge on the tractable nature of the logit function

Appendix: Proofs

As a preliminary, consider the behavior of the metalog and all its derivatives near the edges.

LEMMA 2. *Let* $i \geq 0$. *If* $s(0), s'(0), ..., s^{(i)}(0) = 0$, *then* $\lim_{y\downarrow 0} M^{(i)}(y) = \mu^{(i)}(0)$. *Otherwise,* $M^{(i)}(y)$ *goes towards* $+\infty$ $(-\infty)$ *if the following holds (fails):*

$$\begin{array}{l} (-1)^{i+1}s(0) > 0, \ \textit{or} \\ s(0) = 0 \ \textit{and} \ (-1)^{i}s'(0) > 0, \ \textit{or} \\ s(0), s'(0) = 0 \ \textit{and} \ (-1)^{i-1}s''(0) > 0, \ \textit{or} \\ \cdots \\ s(0), s'(0), ..., s^{(i-1)}(0) = 0 \ \textit{and} \ (-1)s^{(i)}(0) > 0. \end{array}$$

And if $s(1), s'(1), ..., s^{(i)}(1) = 0$ *then* $\lim_{y\uparrow 1} M^{(i)}(y) = \mu^{(i)}(1)$. *Otherwise,* $M^{(i)}(y)$ *goes towards* $+\infty$ $(-\infty)$ *if the following holds (fails):*

$$\begin{array}{l} s(1) > 0, \ \textit{or} \\ s(1) = 0 \ \textit{and} \ s'(1) > 0, \ \textit{or} \\ s(1), s'(1) = 0 \ \textit{and} \ s''(1) > 0, \ \textit{or} \\ \cdots \\ s(1), s'(1), ..., s^{(i-1)}(1) = 0 \ \textit{and} \ s^{(i)}(1) > 0. \end{array}$$

Proof of Lemma 2. Recall that $M^{(i)}(y) = \mu^{(i)}(y) + \sum_{j=0}^{i} \binom{i}{j} s^{(i-j)}(y)\ell^{(j)}(y)$, where $\ell^{(j)}(y) = (j-1)!\frac{y^j - (-1)^j(1-y)^j}{y^j(1-y)^j}$, $i, j \geq 1$. As $y \to 0$, the numerator of $\ell^{(j)}(y)$ approaches $(-1)^{j+1}$; whereas the denominator goes to 0 at a speed that grows faster with j. If $s(0) \neq 0$, and by L'Hôpital's rule, then the limit is determined by the i^{th} term, $s^{(0)}(y)\ell^{(i)}(y)$, and is plus or minus infinity depending on the sign of $(-1)^{i+1}s(0)$, yielding the first logical condition. If $s(0) = 0$ and $s'(0) \neq 0$, then the limit is driven by the previous term, $s^{(1)}(y)\ell^{(i-1)}(y)$, and is plus or minus infinity depending on the sign of $(-1)^{i}s'(0)$. If $s(0) = s'(0) = 0$ and $s''(0) \neq 0$, then we proceed in a similar fashion. The case of $y \to 1$ is similar, except that the numerator of $\ell^{(j)}(y)$ always approaches 1. $\square$

Before establishing **Proposition 1** we establish two key lemmas. Relabel the coefficients as follows. Let $\check{a}_0 = a_1$, $\check{a}_1 = a_4$, $\check{a}_2 = a_5$,... be the μ-terms; $\hat{a}_0 = a_2$, $\hat{a}_1 = a_3$, $\hat{a}_2 = a_6$,... the s-terms; and $\#\mu$ and $\#s$ the number of terms in each polynomial. Then, $\mu(y) = \sum_{t=0}^{\#\mu-1} \check{a}_t(y-0.5)^t$ and $s(y) = \sum_{t=0}^{\#s-1} \hat{a}_t(y-0.5)^t$. Apply the binomial expansion of $(y-0.5)^t$ in both cases, and note in (2) that the numerator of $\ell^{(i)}(y)$ is a polynomial, to obtain

$$M^{(i)}(y) = \underbrace{\sum_{t=i}^{\#\mu-1} \check{a}_t \sum_{u=0}^{t-i} \frac{t!(-0.5)^{t-i-u}}{(t-i-u)!u!} y^u}_{\mu^{(i)}(y)} + \ell(y) \underbrace{\sum_{t=i}^{\#s-1} \hat{a}_t \sum_{u=0}^{t-i} \frac{t!(-0.5)^{t-i-u}}{(t-i-u)!u!} y^u}_{s^{(i)}(y)} + \underbrace{\frac{\sum_{t=0}^{\#s-1} \hat{a}_t \sum_{u=0}^{i+t-1} b_{i,t,u} y^u}{y^i(1-y)^i}}_{\sum_{j=1}^{i}\binom{i}{j} s^{(i-j)}(y)\ell^{(j)}(y)}. \tag{6}$$

The next result derives the coefficients $b_{i,t,u}$.

Lemma 3. *For $i \geq 1$, $b_{i,t,u}$ is equal to*

$$\sum_{j=0}^{\min\{u,i-1\}} \sum_{v=0}^{\min\{u-j,i-1\}} P_{i-v}(i,j) \frac{i!t!(-1)^{j+1+i-v}(-0.5)^{t-u+v}}{j!(i-v)!(t-u+v)!(u-j-v)!}, \ u \leq t, \ \textit{and}$$
$$\sum_{j=0}^{\min\{t,i-1\}} \sum_{v=0}^{\min\{t-j,i+t-u-1\}} P_{i+t-u-v}(i,j) \frac{i!t!(-1)^{j+1+i+t-u-v}(-0.5)^{v}}{j!(i+t-u-v)!v!(t-j-v)!}, \ u \geq t,$$

where

$$P_w(i,j) = \sum_{n=1}^{w} c(w,n) \sum_{m=0}^{n-1} j^m i^{n-m-1},$$

and $c(w,n)$ denote the signed Stirling numbers of the first kind.

Proof of Lemma 3. Let $H^{(i)}(y) = y^i(1-y)^i \sum_{j=0}^{i-1} \binom{i}{j} s^{(j)}(y) \ell^{(i-j)}(y)$ be the numerator of the third term in (6).

Step 1. Substitute $\ell^{(i-j)}(y)$ into $H^{(i)}(y)$, and note that $s^{(i)}(y) = 0$ if $i \geq \#s$, to obtain

$$H^{(i)}(y) = \sum_{j=0}^{\min\{i,\#s\}-1} \binom{i}{j} (i-j-1)! [y^i(1-y)^j - (-1)^{i-j} y^j (1-y)^i] s^{(j)}(y).$$

Step 2. The term $y^i(1-y)^j - (-1)^{i-j} y^j(1-y)^i$ is equal to

$$\sum_{u=j}^{i+j-1} \left[(i-j) \sum_{n=1}^{i+j-u} c(w,n) \sum_{m=0}^{n-1} j^m i^{n-m-1} \right] \frac{(-1)^{i+1-u}}{(i+j-u)!} y^u.$$

To see the equivalence, use $(1-y)^i = -\sum_{n=0}^{i} \binom{i}{n} (-1)^{n+1} y^n$ and the change of variable $u = j+n$ to obtain $-(-1)^{i-j} y^j (1-y)^i = \sum_{n=0}^{i} \binom{i}{n} (-1)^{n+i-j+1} y^{j+n} = \sum_{u=j}^{i+j} \binom{i}{u-j} (-1)^{i+1-u} y^u$. Similarly, the change of variable $u = i+n$ yields $y^i(1-y)^j = -\sum_{n=0}^{j} \binom{j}{n} (-1)^{n+1} y^{i+n} = -\sum_{u=i}^{i+j} \binom{j}{u-i} (-1)^{i+1-u} y^u$. Thus, $y^i(1-y)^j - (-1)^{i-j} y^j (1-y)^i$ is equal to

$$\sum_{u=j}^{i+j} \binom{i}{u-j} (-1)^{i+1-u} y^u - \sum_{u=i}^{i+j} \binom{j}{u-i} (-1)^{i+1-u} y^u$$
$$= \sum_{u=j}^{i-1} \binom{i}{u-j} (-1)^{i+1-u} y^u + \sum_{u=i}^{i+j} \left[\binom{i}{u-j} - \binom{j}{u-i} \right] (-1)^{i+1-u} y^u.$$

Note that for $u = i+j$ the bracketed term vanishes, so the second sum need only run until $i+j-1$. For $u = i, ..., i+j-1$, the bracketed term in the second sum becomes

$$\binom{i}{u-j} - \binom{j}{u-i} = \frac{1}{(i+j-u)!} \left[\prod_{n=0}^{i+j-u-1} (i-n) - \prod_{n=0}^{i+j-u-1} (j-n) \right].$$

This expression is also equal to $\binom{i}{u-j}$, $u = j, ..., i-1$, in the first sum because $\prod_{n=0}^{i+j-u-1} (j-n) = 0$. Finally, use $\prod_{n=0}^{w-1} (i-n) = \sum_{n=1}^{w} c(w,n) i^n$, $w \geq 1$, and $i^n - j^n = (i-j) \sum_{m=0}^{n-1} j^m i^{n-m-1}$ to obtain

$$\prod_{n=0}^{i+j-1-u} (i-n) - \prod_{n=0}^{i+j-1-u} (j-n) = (i-j) \sum_{n=1}^{i+j-u} c(w,n) \sum_{m=0}^{n-1} j^m i^{n-m-1}.$$

Step 3. We have that

$$H^{(i)}(y) = \sum_{t=0}^{\#s-1} \hat{a}_t \sum_{j=0}^{\min\{t,i-1\}} L(i,j)S(t,j), \text{ where}$$

$$L(i,j) = \frac{i!}{j!} \sum_{u=j}^{i+j-1} \left[\sum_{n=1}^{i+j-u} c(w,n) \sum_{m=0}^{n-1} j^m i^{n-m-1} \right] \frac{(-1)^{i+1-u}}{(i+j-u)!} y^u, \text{ and}$$

$$S(t,j) = \sum_{u=0}^{t-j} \frac{t!(-0.5)^{t-j-u}}{(t-j-u)!u!} y^u.$$

To see this, note that the previous step yieds $H^{(i)}(y) = \sum_{j=0}^{\min\{i,\#s\}-1} L(i,j) s^{(j)}(y)$. Next, that $(y-0.5)^{t-j} = \sum_{u=0}^{t-j} \binom{t-j}{u} (-0.5)^{t-j-u} y^u$ implies $s^{(j)}(y) = \sum_{t=j}^{\#s-1} \hat{a}_t S(t,j)$, resulting in $H^{(i)}(y) = \sum_{j=0}^{\min\{i,\#s\}-1} L(i,j) \sum_{t=j}^{\#s-1} \hat{a}_t S(t,j)$. Changing the order of summation produces the result.

Step 4. Note that $L(i,j)$ is a polynomial of degree $i+j-1$, and $S(t,j)$ of degree $t-j$. Hence, each product inside the second sum is a polynomial of degree up to $i+t-1$, which we can write as

$$\sum_{j=0}^{\min\{t,i-1\}} L(i,j)S(t,j) = \sum_{u=0}^{i+t-1} b_{i,t,u} y^u.$$

The expression for the $b_{i,t,u}$ coefficients results from the sum of products of the polynomials $L(i,j)$ and $S(t,j)$, after separating the case of $u \le t$ and $u \ge t$. $\square$

Table 2 **For $i = \lfloor (k+1)/2 \rfloor$, $y^i(1-y)^i M^{(i)}(y)$ is a polynomial of degree $i-1$.**

k	$i = \lfloor (k+1)/2 \rfloor$	$y^i(1-y)^i M^{(i)}(y)$
3,4	2	$-a_2 + 0.5a_3 + 2a_2 y$
5,6	3	$2a_2 - a_3 + 0.5a_6 + (-6a_2 + 2a_3 - 0.5a_6)y + (6a_2 + 0.5a_6)y^2$
7,8	4	$-6a_2 + 3a_3 - 1.5a_6 + 0.75a_7 + (24a_2 - 10a_3 + 4a_6 - 1.5a_7)y$ $+(-36a_2 + 10a_3 - 3a_6 + 1.5a_7)y^2 + (24a_2 + 2a_6)y^3$
9,10	5	$24a_2 - 12a_3 + 6a_6 - 3a_7 + 1.5a_{10}$ $-(120a_2 - 54a_3 + 24a_6 - 10.5a_7 + 4.5a_{10})y$ $+(240a_2 - 90a_3 + 34a_6 - 13.5a_7 + 6a_{10})y^2$ $-(240a_2 - 60a_3 + 20a_6 - 9a_7 + 3a_{10})y^3$ $+(120a_2 + 10a_6 + 1.5a_{10})y^4$
11,12	6	$-120a_2 + 60a_3 - 30a_6 + 15a_7 - 7.5a_{10} + 3.75a_{11}$ $+(720a_2 - 336a_3 + 156a_6 - 72a_7 + 33a_{10} - 15a_{11})y$ $-(1800a_2 - 756a_3 + 318a_6 - 135a_7 + 58.5a_{10} - 26.25a_{11})y^2$ $+(2400a_2 - 840a_3 + 312a_6 - 126a_7 + 54a_{10} - 22.5a_{11})y^3$ $-(1800a_2 - 420a_3 + 150a_6 - 63a_7 + 22.5a_{10} - 11.25a_{11})y^4$ $+(720a_2 + 60a_6 + 9a_{10})y^5$

LEMMA 4. *If $u \ge i \ge t+1$, then $b_{i,t,u} = 0$. Consequently, if $i \ge \lfloor \frac{k+1}{2} \rfloor$, then*

$$y^i(1-y)^i M^{(i)}(y) = \sum_{t=0}^{\#s-1} \hat{a}_t \sum_{u=0}^{i-1} b_{i,t,u} y^u = \sum_{u=0}^{i-1} y^u \sum_{t=0}^{\#s-1} \hat{a}_t b_{i,t,u} \tag{7}$$

is a polynomial of degree $i-1$ (see Table 2).

Proof of Lemma 4. *Step 1.* First note that if $i \geq \#\mu$, then $\mu^{(i)}(y) = 0$; and if $i \geq \#s$, then $s^{(i)}(y) = 0$, and that $\lfloor \frac{k+1}{2} \rfloor = \max\{\#s, \#\mu\}$. Thus, if $i \geq \lfloor \frac{k+1}{2} \rfloor$, then it follows from (6) that $y^i(1-y)^i M^{(i)}(y) = \sum_{t=0}^{\#s-1} \hat{a}_t \sum_{u=0}^{i+t-1} b_{i,t,u} y^u$.

Step 2. Let $u \geq t$ and $i \geq t+1$. In this case, $b_{i,t,u}$ can be written as

$$b_{i,t,u} = i! \sum_{j=0}^{t} \binom{t}{j} (-1)^{j+1} \sum_{v=0}^{\min\{t-j, i+t-u-1\}} P_{i+t-u-v}(i,j) \binom{t-j}{v} \frac{(-1)^{i+t-u-v}(-0.5)^v}{(i+t-u-v)!}.$$

Here, $P_{i+t-v-u}(i,j)$ is a polynomial in j of degree $i+t-v-u-1$; and $\binom{t-j}{v} = (t-j)(t-j-1)...(t-j-v+1)/v!$ is a polynomial in j of degree v. Thus, for all v, the product inside the second sum is a polynomial in j of degree $i+t-1-u$. Hence, for some coefficients A_n, $n = 0, ..., i+t-1-u$, that may depend on (i,t,u), but not on j,

$$b_{i,t,u} = i! \sum_{j=0}^{t} \binom{t}{j} (-1)^{j+1} \sum_{n=0}^{i+t-1-u} A_n j^n = i! \sum_{n=0}^{i+t-1-u} A_n \sum_{j=0}^{t} \binom{t}{j} (-1)^{j+1} j^n.$$

Step 3. By the orthogonality of binomial sums when the polynomial degree is less than the summation index, we have that $\sum_{j=0}^{t} \binom{t}{j} (-1)^{j+1} j^n = 0$ for $0 \leq n \leq t-1$. It follows that if $i+t-1-u \leq t-1$ (equiv. $u \geq i$), then the second sum vanishes.

Step 4. Note that $\hat{a}_0$ multiples $(i-1)![y^i - (-1)^i(1-y)^i] = (i-1)! \sum_{u=0}^{i-1} \binom{i}{u} (-1)^{i-u+1} y^u$, that is,

$$b_{0,u} = \binom{i}{u} (i-1)! (-1)^{i-u+1} \neq 0, \; u = 0, ..., i-1.$$

Thus, the degree of $H^{(i)}(y)$ is at least $i-1$. And for $i \geq \#s$, the degree must be exactly $i-1$. $\square$

Proof of Proposition 1. Let $i \geq \max\{\#s, \#\mu\}$ so that $y^i(1-y)^i M^{(i)}(y)$ is a polynomial of degree $i-1$. We verify that $y^i(1-y)^i M^{(i)}(y)$ is equal to $(-1)^{i-1}(i-1)!s(0)$ for $y = 0$ and equal to $(i-1)!s(1)$ for $y = 1$. Thus, if $s(0) = 0$ or $s(1) = 0$, then the polynomial exhibits a root at 0 or 1, respectively. Hence, $M^{(i)}(y)$ can have at most $i - 1 - \mathbf{1}_{s(0)=0} - \mathbf{1}_{s(1)=0}$ roots in $(0,1)$. For i less than $\max\{\#\mu, \#s\}$, the integral function $M^{(i-1)}(y)$ could add one more root. It follows that $M^{(i)}$ can have at most $\max\{\#\mu, \#s\} + \max\{\#\mu, \#s\} - i - 1 - \mathbf{1}_{s(0)=0} - \mathbf{1}_{s(1)=0}$ roots. $\square$

Proof of Proposition 2. Assume k is even and $n \geq k$, or k is odd and $n > k$. Suppose $\text{rank}(\mathbf{Y}) < k$. Then, there is a coefficient vector $\mathbf{a} \neq 0$ such that $\mathbf{Ya} = 0$. Thus, the associated metalog exhibits $M(y_i) = 0$, $i = 1, ..., n$, and has n distinct roots. By Proposition 1, however, $M(y)$ can have at most $2\max\{\#\mu, \#s\} - 1$ roots, a number equal to $k-1$ when k is even, and equal to k when k is odd; a contradiction. Hence, $\text{rank}(\mathbf{Y}) = k$.

Next, consider the case of $k = n$ odd under symmetry. For any assignment, let $\#\mu^{odd}$ denote the number of coefficients in $\mu(y)$ with $(y-0.5)$ having an odd exponent, and $\#s^{even}$ seven the number of coefficients in $s(y)$ with $(y-0.5)$ having an even exponent, including $a_2(y-0.5)^0\ell(y)$. For all balanced assignments, note that $\#\mu^{odd} + \#s^{even}$ is either $(k-1)/2$ or $(k+1)/2$.

Favorable case $\#\mu^{\text{odd}} + \#s^{\text{even}} = (k-1)/2$. Take $\mathbf{Y}$ and re-order the columns, by keeping the constant term first, followed by the $(k-1)/2$ terms with $\#s^{\text{even}}$ and $\#\mu^{\text{odd}}$ first, followed by the rest of $(k-1)/2$ terms with $\#s^{\text{odd}}$ and $\#\mu^{\text{even}\geq 2}$. To illustrate, we use $k=5$. For conciseness, ℓ_i and $\hat{y}_i$ refer to $\ell(y_i)$ and $(y-0.5)$, respectively.

$$\mathbf{Y} = \left(\begin{array}{c|cc|cc} 1 & \ell_1 & \hat{y}_1 & \hat{y}_1 & \hat{y}_1^2 \\ 1 & \ell_2 & \hat{y}_2 & \hat{y}_2\ell_2 & \hat{y}_2^2 \\ 1 & \ell_3 & \hat{y}_3 & \hat{y}_3\ell_3 & \hat{y}_3^2 \\ 1 & \ell_4 & \hat{y}_4 & \hat{y}_4\ell_4 & \hat{y}_4^2 \\ 1 & \ell_5 & \hat{y}_5 & \hat{y}_5\ell_5 & \hat{y}_5^2 \end{array}\right) \to \left(\begin{array}{c|cc|cc} 0 & \ell_1 & \hat{y}_1 & \hat{y}_1\ell_1 & \hat{y}_1^2 \\ 0 & \ell_2 & \hat{y}_2 & \hat{y}_2\ell_2 & \hat{y}_2^2 \\ \hline 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 2\hat{y}_1\ell_1 & 2\hat{y}_1^2 \\ 0 & 0 & 0 & 2\hat{y}_2\ell_2 & 2\hat{y}_2^2 \end{array}\right) \to \left(\begin{array}{c|cc|cc} 1 & 0 & 0 & 0 & 0 \\ \hline 0 & \ell_1 & \hat{y}_1 & \hat{y}_1\ell_1 & \hat{y}_1^2 \\ 0 & \ell_2 & \hat{y}_2 & \hat{y}_2\ell_2 & \hat{y}_2^2 \\ \hline 0 & 0 & 0 & \ell_1 & \hat{y}_1 \\ 0 & 0 & 0 & \ell_2 & \hat{y}_2 \end{array}\right)$$

By symmetry, if $i+j=n+1$, then $\ell_i + \ell_j = 0$, $\hat{y}_i + \hat{y}_j = 0$, $\hat{y}_i^2\ell_i + \hat{y}_j^2\ell_j = 0$, $\hat{y}_i^3 + \hat{y}_j^3 = 0$, ... We use this fact to replace each row $j = n, n-1, ..., (n+3)/2$ with the sum of itself plus row $i = 1, 2, ..., (n-1)/2$, respectively, resulting in the second matrix. Symmetry also implies that for the middle row $i = (n+1)/2$, $y_i = 0.5$, $\hat{y}_i = 0$ and $\ell_i = 0$. The third matrix moves this middle row to the top, and divides rows $j = n, n-1, ..., (n+3)/2$ by $2\hat{y}_i \neq 0$.

This third matrix contains two identical $\frac{k-1}{2} \times \frac{k-1}{2}$ sub-matrices involving only $\#s^{\text{even}}$ and $\#\mu^{\text{odd}}$ terms (see colored cells in the example). A linear combination of any of its rows takes the form

$$H(y) = a_2\ell(y) + a_4(y-0.5) + a_6(y-0.5)^2\ell(y) + a_8(y-0.5)^3 + ...$$

which can be seen as a metalog with $k-1$ parameters, some set to zero (i.e., no intercept, or term $\hat{y}\ell(y)$, or $\hat{y}^2$, etc....). Because H has no intercept, or $\hat{y}^2$, or $\hat{y}\ell$ terms, etc... it satisfies $H(y) = -H(1-y)$. This implies that 0.5 is always a root of $H(y)$, and that if $0 < y < 0.5$ is a root, then so is $1-y > 0.5$. Thus, $H(y)$ can have at most $(k-3)/2$ roots on $(0, 0.5)$. But if the sub-matrix were to be rank deficient, this would imply that it has $(k-1)/2$ roots, a contradiction. Hence, the sub-matrix has full rank. Applying Gaussian elimination to the two sub-matrices, while embedded in the larger matrix, shows that $\mathbf{Y}$ has full rank.

Unfavorable case $\#\mu^{\text{odd}} + \#s^{\text{even}} = (k+1)/2$. The same construction has more terms on the $\#\mu^{\text{odd}} + \#s^{\text{even}}$ side of the matrix than on the $\#s^{\text{odd}} + \#\mu^{\text{even}\geq 2}$ side. To visualize this, consider the $k=5$ case but with the 5^{th} being an s-term. The same process as before produces:

$$\mathbf{Y} = \left(\begin{array}{c|ccc|c} 1 & \ell_1 & \hat{y}_1 & \hat{y}_1^2\ell_1 & \hat{y}_1\ell_1 \\ 1 & \ell_2 & \hat{y}_2 & \hat{y}_2^2\ell_2 & \hat{y}_2\ell_2 \\ 1 & \ell_3 & \hat{y}_3 & \hat{y}_3^2\ell_3 & \hat{y}_3\ell_3 \\ 1 & \ell_4 & \hat{y}_4 & \hat{y}_4^2\ell_4 & \hat{y}_4\ell_4 \\ 1 & \ell_5 & \hat{y}_5 & \hat{y}_5^2\ell_5 & \hat{y}_5\ell_5 \end{array}\right) \to \left(\begin{array}{c|ccc|c} 0 & \ell_1 & \hat{y}_1 & \hat{y}_1^2\ell_1 & \hat{y}_1\ell_1 \\ 0 & \ell_2 & \hat{y}_2 & \hat{y}_2^2\ell_2 & \hat{y}_2\ell_2 \\ \hline 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 2\hat{y}_1\ell_1 \\ 0 & 0 & 0 & 0 & 2\hat{y}_2\ell_2 \end{array}\right) \to \left(\begin{array}{c|ccc|c} 1 & 0 & 0 & 0 & 0 \\ \hline 0 & \ell_1 & \hat{y}_1 & \hat{y}_1\ell_1 & \hat{y}_1^2 \\ 0 & \ell_2 & \hat{y}_2 & \hat{y}_2\ell_2 & \hat{y}_2^2 \\ \hline 0 & 0 & 0 & 0 & \ell_1 \\ 0 & 0 & 0 & 0 & \ell_2 \end{array}\right)$$

From what we know, the two colored sub-matrices have full rank. Hence, the rank of $\mathbf{Y}$ is $k-1$. Nothing in this logic depends on using $k=5$ as example, and holds for all $k=n$ odd. $\square$

Proof of Proposition 3. We begin by deriving the expression for $\int_0^1 M^n(y)\,dy$. Letting $\hat{s}(y) = s(y)\ell(y)$ and applying the multinomial theorem to expand $M^n(y) = (\mu(y)+\hat{s}(y))^n$ produces

$$(\mu(y)+\hat{s}(y))^n = \sum_{\substack{n_1+n_2=n\\ n_1,n_2\geq 0}} \frac{n!}{n_1!n_2!}\mu^{n_1}(y)\hat{s}^{n_2}(y),$$

Next, we use the multinomial expansion of $\mu^{n_1}(y)$ to obtain

$$\mu^{n_1}(y) = \sum_{\substack{\sum_{j\in\mu} c_j=n_1\\ c_j\geq 0,\ j\in\mu}} \frac{n_1!}{\prod_{j\in\mu} c_j!}\left(\prod_{j\in\mu} a_j^{c_j}\right)(y-0.5)^{\sum_{j\in\mu} c_j \lfloor \frac{j-1}{2}\rfloor}.$$

Similarly, the multinomial expansion of $\hat{s}^{n_2}(y)$, where $\hat{s}(y) = \sum_{j\in s} a_j (y-0.5)^{\lfloor \frac{j-1}{2}\rfloor}\ell(y)$, produces

$$\hat{s}^{n_2}(y) = \sum_{\substack{\sum_{j\in s} d_j=n_2\\ d_j\geq 0,\ j\in s}} \frac{n_2!}{\prod_{j\in s} d_j!}\left(\prod_{j\in s} a_j^{d_j}\right)(y-0.5)^{\sum_{j\in s} d_j \lfloor \frac{j-1}{2}\rfloor}\ell(y)^{\sum_{j\in s} d_j}.$$

To simplify the exponents, we let $m_1 = \sum_{j\in\mu} c_j \left\lfloor \frac{j-1}{2}\right\rfloor$ and $m_2 = \sum_{j\in s} d_j \left\lfloor \frac{j-1}{2}\right\rfloor$, and recall that $\sum_{j\in s} d_j = n_2$. Then,

$$\begin{aligned}
\mu^{n_1}(y)\hat{s}^{n_2}(y) &= \left(\sum_{\substack{\sum_{j\in\mu} c_j=n_1\\ c_j\geq 0,\ j\in\mu}} \frac{n_1!\left(\prod_{j\in\mu} a_j^{c_j}\right)(y-0.5)^{m_1}}{\prod_{j\in\mu} c_j!}\right) \times \left(\sum_{\substack{\sum_{j\in s} d_j=n_2\\ d_j\geq 0,\ j\in s}} \frac{n_2!\left(\prod_{j\in s} a_j^{d_j}\right)(y-0.5)^{m_2}\ell(y)^{n_2}}{\prod_{j\in s} d_j!}\right)\\
&= \sum_{\substack{\sum_{j\in\mu} c_j=n_1\\ c_j\geq 0,\ j\in\mu}} \sum_{\substack{\sum_{j\in s} d_j=n_2\\ d_j\geq 0,\ j\in s}} \frac{n_1!n_2!}{\prod_{j\in\mu} c_j!\prod_{j\in s} d_j!}\left(\prod_{j\in\mu} a_j^{c_j}\right)\left(\prod_{j\in s} a_j^{d_j}\right)(y-0.5)^{m_1+m_2}\ell(y)^{n_2}.
\end{aligned}$$

Taking the integral, while recalling that $I(m,n) = \int_0^1 (y-0.5)^m \ell(y)^n\,dy$, yields

$$\int_0^1 \mu^{n_1}(y)\hat{s}^{n_2}(y)\,dy = \sum_{\substack{\sum_{j\in\mu} c_j=n_1\\ c_j\geq 0,\ j\in\mu}} \sum_{\substack{\sum_{j\in s} d_j=n_2\\ d_j\geq 0,\ j\in s}} \frac{n_1!n_2!}{\prod_{j\in\mu} c_j!\prod_{j\in s} d_j!}\left(\prod_{j\in\mu} a_j^{c_j}\right)\left(\prod_{j\in s} a_j^{d_j}\right) I(m_1+m_2,n_2).$$

Substituting into the raw moment expression and cancelling $n_1!n_2!$ produces

$$\begin{aligned}
\int_0^1 M^n(y)\,dy &= \sum_{n_1+n_2=n} \frac{n!}{n_1!n_2!}\int_0^1 \mu^{n_1}(y)\hat{s}^{n_2}(y)\,dy\\
&= \sum_{\substack{n_1+n_2=n\\ n_1,n_2\geq 0}} \sum_{\substack{\sum_{j\in\mu} c_j=n_1\\ c_j\geq 0,\ j\in\mu}} \sum_{\substack{\sum_{j\in s} d_j=n_2\\ d_j\geq 0,\ j\in s}} \frac{n!}{\prod_{j\in\mu} c_j!\prod_{j\in s} d_j!}\left(\prod_{j\in\mu} a_j^{c_j}\right)\left(\prod_{j\in s} a_j^{d_j}\right) I\left(m_1+m_2,n_2\right).
\end{aligned}$$

Recalling that $m_1+m_2 = \sum_{j\in\mu} c_j \left\lfloor \frac{j-1}{2}\right\rfloor + \sum_{j\in s} d_j \left\lfloor \frac{j-1}{2}\right\rfloor$, the result follows.

That all moments of a k-metalog are finite depends on $I(m,u)$ being bounded. This follows from Results 1 and 2 in the Online Supplement A1.

For $n=1$, the mean, note that if $j (\bmod 4)=0$, then $\lfloor \frac{j-1}{2} \rfloor$ is odd and $I(\lfloor \frac{j-1}{2} \rfloor, 0)=0$. Similarly, if $j (\bmod 4)=2$, then $I(\lfloor \frac{j-1}{2} \rfloor, 1)=0$. Also, that $j (\bmod 4) \in \{1,3\}$ implies $\lfloor \frac{j-1}{2} \rfloor = \frac{j-1}{2}$. Thus,

$$\int_0^1 M(y)dy = \sum_{\substack{j=1,\ldots,k \\ j (\bmod 4)=1}} a_j I(\tfrac{j-1}{2}, 0) + \sum_{\substack{j=3,\ldots,k \\ j (\bmod 4)=3}} a_j I(\tfrac{j-1}{2}, 1).$$

For $n=2$, the variance, let $\diamond$ and $\bullet$ be short for $2\lfloor \frac{j-1}{2} \rfloor$ and $\lfloor \frac{i-1}{2} \rfloor + \lfloor \frac{j-1}{2} \rfloor$, respectively. Then,

$$\int_0^1 M^2(y)dy = \sum_{1 \leq j \leq k} a_j^2 (I(\diamond, 0)\mathbf{1}_{j \in \mu} + I(\diamond, 2)\mathbf{1}_{j \in s}) + 2 \sum_{1 \leq i < j \leq k} a_i a_j (I(\bullet, 0)\mathbf{1}_{i,j \in \mu} + I(\bullet, 1)\mathbf{1}_{\substack{i \in s, j \in \mu \\ || j \in s, i \in \mu}} + I(\bullet, 2)\mathbf{1}_{i,j \in s}).$$

Using $\mathbb{V}[M] = \int_0^1 M^2(y)dy - \mathbb{E}[M]^2$, that $I(j,u)=0$ if j and u have different parity, that $\lfloor \frac{j-1}{2} \rfloor = \frac{j-1}{2}$ for $j (\bmod 4) = 1, 3$, and that $\lfloor \frac{j-1}{2} \rfloor = \frac{j-2}{2}$ for $j (\bmod 4) = 2, 0$, yields the proposed expression. Note that all terms involving a_1 cancel out because $I(0,0)=1$. $\square$

Proof of Proposition 4. The change of variable $x = y - 0.5$ and $\ell(y) = \ln y - \ln(1-y)$ yields

$$\int_{y_0}^{y_1} (y-0.5)^m \ell(y)dy = \int_{y_0-0.5}^{y_1-0.5} x^m \ln(x+0.5)dx - \int_{y_0-0.5}^{y_1-0.5} x^m \ln(0.5-x)dx.$$

These two integrals admit an explicit form (Zwillinger 2018, eq. 481) producing the result. $\square$

Proof of Proposition 5. The tail feasibility conditions are an immediate consequence of Lemma 2, as applied to the first derivative.

Suppose M is feasible. Because $M'(y) \geq 0$ at all $y \in (0,1)$, then the slope at all inflection points must be positive. And because weak inequalities are preserved under limits, feasibility also implies tail feasibility (i.e., if tail feasibility fails, then there is $y \in (0,1)$ with $M'(y) < 0$).

Finally, suppose to the contrary that M is tail feasible and that all its inflection points have positive slope, but it is infeasible. Thus, $M'(z) < 0$ for some $z \in (0,1)$. Because all inflection points have positive slope, it cannot be that $M''(z) = 0$. Consider the case a) $M''(z) > 0$. We have two sub-cases, a1) $M''(y) > 0$ for all $0 < y < z$ and a2) $M''(y) = 0$ for some $0 < y < z$. In a1) the slope remains negative and lower tail feasibility fails, a contradiction. In a2), because M is convex on (y,z), we have that $M'(y) < 0$, contradicting that all inflection points have positive slope. Next, consider case b) $M''(z) < 0$ and the sub-cases b1) $M''(y) < 0$ for all $z < y < 1$ and b2) $M''(y) = 0$ for some $z < y < 1$. In b1), we get that M is upper tail infeasible, a contradiction. In b2) that M is concave on (z,y) implies $M'(y) < 0$, again a contradiction. $\square$

Proof of Proposition 6. In the space of coefficients, the constraint $M'(y) \geq 0$ constitutes a hyperplane going through the origin. The set of feasible coefficients is then given the intersection of infinitely many such hyperplanes, one for each $y \in (0,1)$. And the arbitrary intersection of hyperplanes that pass through the origin is always a closed and convex cone.

For $k = 2$, feasibility trivially holds iff $a_2 \geq 0$. For $k = 3$, the feasible sets results from $\mathcal{A}_4$ after setting $a_4 = 0$. Let $k = 4$ so that $s(y) = a_2 + a_3(y - 0.5)$. The first tail feasibility condition, $s(0), s(1) \geq 0$ yields $-2a_2 \leq a_3 \leq 2a_2$, which can only hold if $a_2 \geq 0$. If $s(0) = 0$ ($a_3 = 2a_2$), then $s'(0) = a_3 \leq 0$, which coupled with $a_3 = 2a_2 \geq 0$ imposes $a_2 = a_3 = 0$. That $s'(0) = 0$ then requires the third condition, $\mu'(0) = a_4 \geq 0$. And if $s(1) = 0$ ($a_3 = -2a_2$), then $s'(1) = a_3 \geq 0$, which coupled with $a_3 = -2a_2 \leq 0$ imposes $a_2 = a_3 = 0$. That $s'(1) = 0$ then requires the third condition, $\mu'(1) = a_4 \geq 0$. Thus, tail feasibility holds if and only if either $a_2 = a_3 = 0$ and $a_4 \geq 0$ (constant or uniform), or $a_2 > 0$ and $-2a_2 < a_3 < 2a_2$ (a convex but not closed cone).

Let $a_2 > 0$. Recall that $y^2(1-y)^2 M''(y) = a_2(2y-1) + 0.5a_3$, which we set to zero to produces the inflection point candidate $\mathring{y} = 0.5 - 0.25a_3/a_2$. That the slope be positive imposes $M'(\mathring{y}) = a_2\ell'(\mathring{y}) + a_3(\mathring{y} - 0.5)\ell'(\mathring{y}) + a_3\ell(\mathring{y}) + a_4 \geq 0$. Let $b = a_3/a_2$ so that $\ell'(\mathring{y}) = \frac{16}{(2-b)(2+b)}$, $\ell(\mathring{y}) = \ln \frac{2-b}{2+b}$. Then, after dividing by a_2 and noticing that $\ell'(\mathring{y}) + b(\mathring{y} - 0.5)\ell'(\mathring{y}) = 4$, we conclude that $M'(\mathring{y}) \geq 0$ becomes equivalent to $a_4 \geq a_3 \ln \frac{2a_2+a_3}{2a_2-a_3} - 4a_2$. $\square$

Proof of Proposition 7. Given $\mathbf{a}^{\text{OLS}}$ in (3), it is well known that $(\mathbf{x} - \mathbf{Ya})^{\mathsf{T}}(\mathbf{x} - \mathbf{Ya}) = \mathbf{x}^{\mathsf{T}}\mathbf{x} - (\mathbf{Ya}^{\text{OLS}})^{\mathsf{T}}\mathbf{Ya}^{\text{OLS}} + (\mathbf{a} - \mathbf{a}^{\text{OLS}})^{\mathsf{T}}\mathbf{Y}^{\mathsf{T}}\mathbf{Y}(\mathbf{a} - \mathbf{a}^{\text{OLS}})$. Because the first two terms do not depend on $\mathbf{a}$, the iso-distance contours are the surface of an ellipsoid centered at $\mathbf{a}^{\text{OLS}}$. Because $\mathbf{Y}$ has full rank, the Gram matrix $\mathbf{Y}^{\mathsf{T}}\mathbf{Y}$ is positive definite. Next, let $\mathcal{T}_k = \{\mathbf{a} \in \mathbb{R}^k : (\mathbf{x} - \mathbf{Ya})^{\mathsf{T}}(\mathbf{x} - \mathbf{Ya}) \leq \mathbf{x}^{\mathsf{T}}\mathbf{x}\}$ be the set of coefficients that improve the fit relative to the trivial metalog $\mathbf{a} = \mathbf{0}$. Clearly, $\mathcal{T}_k$ is a closed and convex ellipsoid, hence compact. Because the trivial metalog is feasible, $\mathcal{A}_k \cap \mathcal{T}_k$ is non-empty, and must contain $\mathbf{a}^*$. The original program can be recasted as the minimization of a continuous function over a compact set, $\min_{\mathbf{a} \in \mathcal{A}_k \cap \mathcal{T}_k} (\mathbf{a} - \mathbf{a}^{\text{OLS}})^T \mathbf{Y}^{\mathsf{T}}\mathbf{Y}(\mathbf{a} - \mathbf{a}^{\text{OLS}})$, hence an optimal solution exists. By the projection theorem (Bertsekas 2009, Prop. 1.1.9) the optimal solution is characterized by $(\mathbf{a}^{\text{OLS}} - \mathbf{a}^*)^{\mathsf{T}}\mathbf{Y}^{\mathsf{T}}\mathbf{Y}(\mathbf{a} - \mathbf{a}^*) \leq 0$ for all $\mathbf{a} \in \mathcal{A}_k$, and uniqueness follows. $\square$

References

Abramowitz M, Stegun IA (1948) *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 (US Government printing office).

Aspinall W, Sparks R, Hill B, Costa A, Connor C, Inakura H, Hasenaka T, Miyoshi M, Kiyosugi K, Tsuji T, Ushioda M (2023) Aso volcano, japan: assessing the 100-year probability of a new caldera-forming eruption based on expert judgements with bayes net and importance sampling uncertainty analysis. *Journal of Applied Volcanology* 12(1):5.

Bae S, Schwartz GJ, Mendley SR, Warady BA, Furth SL, Muñoz A (2023) Trajectories of eGFR after kidney transplantation according to trajectories of eGFR prior to kidney replacement therapies in children with chronic kidney disease. *Pediatric Nephrology* 38(12):4157–4164.

Bertschek I, Kaiser U (2004) Productivity effects of organizational change: microeconometric evidence. *Management science* 50(3):394–404.

Bertsekas D (2009) *Convex optimization theory*, volume 1 (Athena Scientific).

Bertsekas D (2016) *Nonlinear Programming*, volume 4 (Athena Scientific).

Bittencourt R, Pereira HBdB, Moret MA, Da Cunha Lima IC, Galam S (2025) A novel evaluation of income class boundaries using inflection points of probability density functions: A case study of brazil. *Entropy* 27(2):186.

Boyd S, Vandenberghe L (2004) *Convex optimization* (Cambridge University Press).

Bratvold RB, Mohus E, Petutschnig D, Bickel E (2020) Production forecasting: Optimistic and overconfident—over and over again. *SPE Reservoir Evaluation & Engineering* 23(03):0799–0810.

Caban J, Małek A, Kroczyński D (2024) A method for assessing the technical condition of traction batteries using the metalog family of probability distributions. *Energies* 17(13):3096.

Chan TC, Demirtas D, Kwon RH (2016) Optimizing the deployment of public access defibrillators. *Management science* 62(12):3617–3635.

Dinh N, Goberna MA, López MA (2012) On the stability of the optimal value and the optimal set in optimization problems. *Journal of Convex Analysis* 19(4):927–953.

Dion P, Galbraith N, Sirag E (2020) Using expert elicitation to build long-term projection assumptions. *Developments in demographic forecasting*, 43–62 (Springer).

Finke A, Foffa S, Iacovelli F, Maggiore M, Mancarella M (2021) Cosmology with ligo/virgo dark sirens: Hubble parameter and modified gravitational wave propagation. *Journal of Cosmology and Astroparticle Physics* 2021(08):026.

Gilchrist W (2000) *Statistical modelling with quantile functions* (CRC Press).

Hazen A (1914) Storage to be provided in impounding municipal water supply. *Transactions of the American society of civil engineers* 77(1):1539–1640.

Hettich R, Kortanek KO (1993) Semi-infinite programming: theory, methods, and applications. *SIAM review* 35(3):380–429.

Iacovelli F, Mancarella M, Foffa S, Maggiore M (2022) Forecasting the detection capabilities of third-generation gravitational-wave detectors using gwfast. *The Astrophysical Journal* 941(2):208.

Keating L, Randall L, Stanton R, McCormack C, Lucid M, Seaborn T, Converse S, Canessa S, Moehrenschlager A (2023) Using decision analysis to determine the feasibility of a conservation translocation. *Decision Analysis* 20(4):295–310.

Keefer DL, Bodily SE (1983) Three-point approximations for continuous random variables. *Management Science* 29(5):595–609.

Keelin TW (2016) The metalog distributions. *Decision Analysis* 13(4):243–277.

Keelin TW, Chrisman L, Savage SL (2019) The metalog distributions and extremely accurate sums of lognormals in closed form. *2019 Winter Simulation Conference (WSC)*, 3074–3085 (IEEE).

Keelin TW, Powley BW (2011) Quantile-parameterized distributions. *Decision Analysis* 8(3):206–219.

Kohlbrecher J, Breßler I (2022) Updates in sasfit for fitting analytical expressions and numerical models to small-angle scattering patterns. *Journal of Applied Crystallography* 55(6):1677–1688.

Lai Y, Das Sarma S, Sau J (2022) Quality factor for zero-bias conductance peaks in majorana nanowire. *Physical Review B* 106(9):094504.

Lichtendahl Jr KC, Grushka-Cockayne Y, Winkler RL (2013) Is it better to average probabilities or quantiles? *Management Science* 59(7):1594–1611.

Lloyd N, Keating L, Friesen A, Cole D, McPherson J, Akçakaya H, Moehrenschlager A (2023) Prioritizing species conservation programs based on iucn green status and estimates of cost-sharing potential. *Conservation Biology* 37(3):e14051.

Małek A, Dudziak A, Caban J, Stoma M (2024) Strategic model for yellow hydrogen production using the metalog family of probability distributions. *Energies* 17(10):2398.

Nesvold E, Bratvold R (2023) Field features do not explain greenfield production forecasting bias. *SPE Journal* 28(03):1290–1307.

Nesvold E, Bratvold RB (2022) Debiasing probabilistic oil production forecasts. *Energy* 258:124744.

Norling M, Clayer F, Gundersen C (2022) Levels of nitramines and nitrosamines in lake drinking water close to a co2 capture plant: A modelling perspective. *Environmental Research* 212:113581.

Perepolkin D, Goodrich B, Sahlin U (2024) Hybrid elicitation and quantile-parametrized likelihood. *Statistics and Computing* 34(1):11.

Perepolkin D, Lindsröm E, Sahlin U (2025) Quantile-parameterized distributions for expert knowledge elicitation. *Decision Analysis* Forthcoming.

Powell M (1981) *Approximation Theory and Methods* (Cambridge: Cambridge University Press), ISBN 978-0521295147.

Redivo E, Viroli C (2024) Bayesian estimation of a quantile-based factor model. *Journal of Statistical Computation and Simulation* 94(17):3892–3932.

Rongen G, Morales-Nápoles O, Kok M (2024) Using the classical model for structured expert judgment to estimate extremes: a case study of discharges in the meuse river. *Hydrology and Earth System Sciences* 28(13):2831–2848.

Runolinna M, Turnquist M, Teittinen J, Ilmonen P, Koskinen L (2023) Extreme path delay estimation of critical paths in within-die process fluctuations using multi-parameter distributions. *Journal of Low Power Electronics and Applications* 13(1):22.

Savage S, Doheney S, Smith C (2019) Advanced tutorial: networking simulations across platforms and enterprises with virtual sips. *2019 Winter Simulation Conference (WSC)*, 105–117 (IEEE).

Steins E, Herty M (2023) Error propagation for quantile estimation via combining polynomial chaos expansions and metalog distributions. *Probabilistic Engineering Mechanics* 73:103478.

Stomma P, Rudnicki W (2024) Hcs—hierarchical algorithm for simulation of omics datasets. *Bioinformatics* 40(2):ii98–ii104.

Theil H, Van de Panne C (1960) Quadratic programming as an extension of classical quadratic maximization. *Management Science* 7(1):1–20.

Theodossiou P (1998) Financial data and the skewed generalized t distribution. *Management science* 44(12-part-1):1650–1661.

Wahba G (1990) *Spline models for observational data* (SIAM).

Wang J, Neil M, Fenton N (2020) A bayesian network approach for cybersecurity risk assessment implementing and extending the fair model. *Computers & Security* 89:101659.

Weibull W (1939) A statistical theory of strength of materials. *IVB-Handl.* .

Zrazhevsky G, Zrazhevska V (2021) Quintile regression based approach for dynamical var and cvar forecasting using metalog distribution. *System research and information technologies* 1(1):139–150.

Zwillinger D (2018) *CRC standard mathematical tables and formulas* (CRC press).

Online Supplement

A1. Proof of Lemma 1

Step #1. Expressions (4). For non-negative integers n and u, let $S_{n,u} = \sum_{i=0}^{u} J^n_{i,u-i}$, where

$$J^n_{i,u} = \frac{(-1)^u}{i!u!} \int_0^1 y^n \ln^i(y) \ln^u(1-y) dy.$$

Substituting the binomial expansions of $(y-0.5)^j$ and of $\ell^u(y) = (\ln(y) - \ln(1-y))^u$ produces

$$\begin{aligned}\int_0^1 (y-0.5)^m \ell^u(y) dy &= \sum_{n=0}^{m} \binom{m}{n} (-0.5)^{m-n} \int_0^1 y^n \ell^u(y) dy \\ &= \sum_{n=0}^{m} \binom{m}{n} (-0.5)^{m-n} u! \underbrace{\sum_{i=0}^{u} \frac{(-1)^{u-i}}{i!(u-i)!} \int_0^1 y^n \ln^i(y) \ln^{u-i}(1-y) dy}_{S_{n,u}}.\end{aligned}$$

Step #2. We have that

$$\begin{aligned}J^n_{i,0} &= \frac{(-1)^i}{(n+1)^{i+1}},\ i, n \geq 0, \\ J^n_{0,u} &= J^n_{0,0} + \frac{n}{n+1} \sum_{m=1}^{u} (J^{n-1}_{0,m} - J^n_{0,m-1}),\ u, n \geq 1, \\ J^n_{i,u} &= J^n_{i,0} + \frac{n}{n+1} \sum_{m=1}^{u} (J^{n-1}_{i,m} - J^n_{i,m-1}) + \frac{1}{n+1} \sum_{m=1}^{u} (J^{n-1}_{i-1,m} - J^n_{i-1,m}),\ i, u, n \geq 1.\end{aligned}$$

That $J^n_{i,0} = \frac{1}{i!} \int_0^1 y^n \ln^i(y) dy = \frac{(-1)^i}{(n+1)^{i+1}}$, $i, n \geq 0$, follows from Zwillinger (2018, eq. 674).

For the second recursion, use integration by parts:

$$\begin{aligned}J^n_{0,u} &= \frac{(-1)^u}{u!} \int_0^1 y^n \ln^u(1-y) dy = y^n J^0_{0,u}(y)\Big|_0^1 - \int_0^1 n y^{n-1} J^0_{0,u}(y) dy \\ &= 1 - \int_0^1 n y^{n-1} dy - \int_0^1 n(y^n - y^{n-1}) \sum_{m=0}^{u} \frac{(-1)^m \ln^m(1-y)}{m!} dy \\ &= n J^{n-1}_{0,0} + n \sum_{m=0}^{u} J^{n-1}_{0,m} - n \sum_{m=1}^{u} J^n_{0,m-1} - n J^n_{0,u}.\end{aligned}$$

Moving $nJ^n_{0,u}$ to the left side, dividing by $n+1$, and noting $\frac{n}{n+1} J^{n-1}_{0,0} = J^n_{0,0}$ produces the result.

Similarly for the third recursion:

$$\begin{aligned}J^n_{i,u} &= \frac{(-1)^u}{u!i!} \int_0^1 y^n \ln^i(y) \ln^u(1-y) dy = \frac{1}{i!} y^n \ln^i(y) J^0_{0,u}(y)\Big|_0^1 - \int_0^1 \left(\frac{n \ln^i(y)}{i!} + \frac{\ln^{i-1}(y)}{(i-1)!} \right) y^{n-1} J^0_{0,u}(y) dy \\ &= -\int_0^1 \left(\frac{n \ln^i(y)}{i!} + \frac{\ln^{i-1}(y)}{(i-1)!} \right) \left(y^{n-1} + (y^n - y^{n-1}) \sum_{m=0}^{u} \frac{(-1)^m \ln^m(1-y)}{m!} \right) dy \\ &= -n J^n_{i,u} - J^n_{i-1,0} + \sum_{m=1}^{u} (n J^{n-1}_{i,m} - n J^n_{i,m-1} + J^{n-1}_{i-1,m} - J^n_{i-1,m}).\end{aligned}$$

Move $-nJ^n_{i,u}$ to the left, divide by $(n+1)$, and use $-\frac{1}{n+1} J^n_{i-1,0} = J^n_{i,0}$ produces the desired result.

Step #3. Recursive Formula for $S_{n,u}$, $n,u \geq 1$. Applying the three recursions from Step #2 to each term in $S_{n,u} = J^n_{u,0} + J^n_{0,u} + \sum_{i=1}^{u-1} J^n_{i,u-i}$ and grouping terms produces

$$S_{n,u} = \sum_{i=0}^{u} J^n_{i,0} + \frac{n}{n+1} \sum_{i=0}^{u-1} \sum_{m=1}^{u-i} (J^{n-1}_{i,m} - J^n_{i,m-1}) + \frac{1}{n+1} \sum_{i=1}^{u-1} \sum_{m=1}^{u-i} (J^{n-1}_{i-1,m} - J^n_{i-1,m})$$

Next, re-indexing the first double sum using $k = m+i$ yields

$$\sum_{i=0}^{u-1} \sum_{m=1}^{u-i} (J^{n-1}_{i,m} - J^n_{i,m-1}) = \sum_{k=1}^{u} \sum_{i=0}^{k-1} (J^{n-1}_{i,k-i} - J^n_{i,k-i-1}) = \sum_{k=1}^{u} (S_{n-1,k} - J^{n-1}_{k,0} - S_{n,k-1}).$$

Similarly, re-indexing the second double sum using $k = m+i-1$ yields

$$\sum_{i=1}^{u-1} \sum_{m=1}^{u-i} (J^{n-1}_{i-1,m} - J^n_{i-1,m}) = \sum_{k=1}^{u-1} \sum_{i=1}^{k} (J^{n-1}_{i-1,k-i+1} - J^n_{i-1,k-i+1}) = \sum_{k=1}^{u-1} (S_{n-1,k} - J^{n-1}_{k,0} - S_{n,k} + J^n_{k,0}).$$

Substituting these expressions results in

$$S_{n,u} = \sum_{k=0}^{u} J^n_{k,0} + \frac{n}{n+1} \sum_{k=1}^{u} \left(S_{n-1,k} - J^{n-1}_{k,0} - S_{n,k-1}\right) + \frac{1}{n+1} \sum_{k=1}^{u-1} \left(S_{n-1,k} - J^{n-1}_{k,0} - S_{n,k} + J^n_{k,0}\right).$$

Lastly, that $J^n_{i,0} = \frac{(-1)^i}{(n+1)^{i+1}}$, $i,n \geq 0$, implies that $\frac{-n}{n+1} \sum_{k=1}^{u} J^{n-1}_{k,0} = \frac{1}{n+1} \sum_{k=0}^{u-1} J^{n-1}_{k,0}$ and $\frac{1}{n+1} \sum_{k=1}^{u-1} J^n_{k,0} = -\sum_{k=2}^{u} J^n_{k,0}$. Thus, the boundary terms can be simplified to

$$\begin{aligned}
&\sum_{k=0}^{u} J^n_{k,0} - \frac{n}{n+1} \sum_{k=1}^{u} J^{n-1}_{k,0} - \frac{1}{n+1} \sum_{k=1}^{u-1} J^{n-1}_{k,0} + \frac{1}{n+1} \sum_{k=1}^{u-1} J^n_{k,0} \\
=&\sum_{k=0}^{u} J^n_{k,0} + \frac{1}{n+1} \sum_{k=0}^{u-1} J^{n-1}_{k,0} - \frac{1}{n+1} \sum_{k=1}^{u-1} J^{n-1}_{k,0} - \sum_{k=2}^{u} J^n_{k,0} \\
=&J^n_{0,0} + J^n_{1,0} + \frac{1}{n+1} J^{n-1}_{0,0} = \frac{1}{n+1} - \frac{1}{(n+1)^2} + \frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{(n+1)^2}.
\end{aligned}$$

Step #4. Boundary Values. Clearly, $S_{n,0} = \int_0^1 y^n dy = \frac{1}{n+1}$, $n \geq 0$. Next, note that $I(0,u) = u! S_{0,u}$, or $S_{0,u} = I(0,u)/u!$. If u is odd, then $I(0,u) = 0$ and $S_{0,u} = 0$. If $u \geq 2$ is even, then $I(0,u) = \int_0^1 \ell^u(y) dy$ are the moments of the standard logistic distribution, given by

$$\begin{aligned}
\int_0^1 \ell^u(y) dy &=_1 \int_{-\infty}^{\infty} \frac{w^u}{e^{-w} + e^w + 2}\, dw =_2 2 \int_0^{\infty} \frac{w^u}{e^{-w} + e^w + 2}\, dw \\
&=_3 2 \int_0^{\infty} w^u \left[\sum_{k=0}^{\infty} (-1)^k (k+1) e^{-(k+1)w} \right] dw \\
&=_4 2 \sum_{k=0}^{\infty} (-1)^k (k+1) \int_0^{\infty} w^u e^{-(k+1)w}\, dw \\
&=_5 2u! \sum_{k=0}^{\infty} \frac{(-1)^k}{(k+1)^u} =_6 2u!(1 - 2^{1-u}) \zeta(u).
\end{aligned}$$

Thus, $S_{0,u} = I(0,u)/u! = 2(1 - 2^{1-u}) \zeta(u)$.

The different equalities above come about as follows.

$=_1$: Apply the change of variable $y = e^w/(1+e^w)$ or $\ell(y) = w$.

$=_2$: Note that the integrand is an even function.

$=_3$: Let $x \in (0,1)$, differentiate the geometric series equality $\frac{1}{1+x} = \sum_{k=0}^{\infty}(-1)^k x^k$, to obtain $-\frac{1}{(1+x)^2} = \sum_{k=1}^{\infty}(-1)^k k x^{k-1}$, which then produces $\frac{x}{(1+x)^2} = \sum_{k=0}^{\infty}(-1)^k(k+1)x^{k+1}$. Note that $\frac{x}{(1+x)^2} = \frac{1}{x+1/x+2}$ and substitute $x = e^{-w}$ to obtain $\frac{1}{e^{-w}+e^{w}+2} = \sum_{k=0}^{\infty}(-1)^k(k+1)e^{-(k+1)w}$.

$=_4$: Interchange sum and integral, after using $\int_0^\infty w^u e^{-(k+1)w}\, dw = \frac{u!}{(k+1)^{u+1}}$ (Zwillinger 2018, eq. 641) to verify that $\sum_{k=0}^{\infty}\int_0^\infty |w^u(-1)^k(k+1)e^{-(k+1)w}| dw = u! \sum_{k=0}^{\infty}\frac{1}{(k+1)^u} < \infty$, as $u \geq 2 > 1$.

$=_5$: Again, apply $\int_0^\infty w^u e^{-(k+1)w}\, dw = \frac{u!}{(k+1)^{u+1}}$ and note that $(k+1)$ cancels out.

$=_6$: By Abramowitz and Stegun (1948, p. 807, Eq. 23.2.19), $\sum_{k=0}^{\infty}\frac{(-1)^k}{(k+1)^u} = (1-2^{1-u})\zeta(u)$. $\square$

A2. The Feasible Cone for the 5-metalog

For $a_5 \neq 0$, let $\bar{a}_i = a_i/a_5$, $i=1,...,4$, $p = -\left(\frac{\bar{a}_3}{4}+\frac{1}{12}\right)$, $q = -\left(\frac{\bar{a}_2^2}{8}+\frac{\bar{a}_3}{24}+\frac{1}{108}\right)$,

$$\Delta = \frac{q^2}{4}+\frac{p^3}{27}, \text{ and}$$
$$x = \begin{cases} \sqrt[3]{-q/2+\sqrt{\Delta}}+\sqrt[3]{-q/2-\sqrt{\Delta}}-1/12, & \Delta > 0, \\ 2\sqrt{-p/3}\cos\left(\frac{1}{3}\arccos\left(\frac{3q}{2p}\sqrt{-3/p}\right)\right)-1/12, & \Delta \leq 0. \end{cases}$$

Then, the potential loci of the inflection points are:

$$y_1 = \frac{1}{2}\left(-\sqrt{2x+\tfrac{1}{2}}-\sqrt{-2x+\tfrac{1}{2}+\frac{2\bar{a}_2}{\sqrt{2x+1/2}}}\right)+\frac{1}{2},$$
$$y_2 = \frac{1}{2}\left(-\sqrt{2x+\tfrac{1}{2}}+\sqrt{-2x+\tfrac{1}{2}+\frac{2\bar{a}_2}{\sqrt{2x+1/2}}}\right)+\frac{1}{2},$$
$$y_3 = \frac{1}{2}\left(\sqrt{2x+\tfrac{1}{2}}-\sqrt{-2x+\tfrac{1}{2}-\frac{2\bar{a}_2}{\sqrt{2x+1/2}}}\right)+\frac{1}{2}, \text{ and}$$
$$y_4 = \frac{1}{2}\left(\sqrt{2x+\tfrac{1}{2}}+\sqrt{-2x+\tfrac{1}{2}-\frac{2\bar{a}_2}{\sqrt{2x+1/2}}}\right)+\frac{1}{2}.$$

Proposition 8. *For $k=5$, $\mathbf{a} \in \mathcal{A}_5$ if and only if either*

- $a_2 = a_3 = 0$, $a_4 \geq 0$ *and* $-a_4 \leq a_5 \leq a_4$, *or*
- $a_2 > 0$, $-2a_2 < a_3 < 2a_2$, $a_5 = 0$ *and* $a_4 \geq a_3 \ln \frac{2a_2+a_3}{2a_2-a_3} - 4a_2$, *or*
- $a_2 > 0$, $-2a_2 < a_3 < 2a_2$, $a_5 \neq 0$, *and* $a_4 \geq a_5 T(\bar{a}_2, \bar{a}_3)$, *where*

$$T(\bar{a}_2,\bar{a}_3) = \begin{cases} g(y_2), & a_5 > 0, \Delta > 0, \\ \max\{g(y_2), g(y_4)\}, & a_5 > 0, \Delta \leq 0, \\ \min\{g(y_3), g(y_1)\}, & a_5 < 0, \Delta \leq 0, \\ g(y_3), & a_5 < 0, \Delta > 0, \text{ and} \end{cases}$$
$$g(y) = -\frac{\bar{a}_2+\bar{a}_3(y-0.5)}{y(1-y)} - \bar{a}_3\ell(y) - 2(y-0.5).$$

Moreover, $a_5 T(\bar{a}_2, \bar{a}_3) = |a_5| T(a_2/|a_5|, a_3/a_5)$.

Proof of Proposition 8. Any feasible solution satisfies $s(0), s(1) \geq 0$, or $-2a_2 \leq a_3 \leq 2a_2$ and $a_2 \geq 0$. Consider the case that either $s(0) = a_2 - 0.5a_3 = 0$ or $s(1) = a_2 + 0.5a_3 = 0$. Feasibility then requires, respectively, that $-s'(0) = -a_3 \geq 0$ or $s'(1) = a_3 \geq 0$. Either case can only hold if $a_2 = a_3 = 0$. That both $s(0) = s'(0) = 0$ and $s(1) = s'(1) = 0$ hold imposes that both $\mu'(0) = a_4 - a_5 \geq 0$ and $\mu'(1) = a_4 + a_5 \geq 0$ must hold, resulting in $-a_4 \leq a_5 \leq a_4$.

Next, focus on the case $a_2 > 0$ and $-2a_2 < a_3 < 2a_2$. If $a_5 = 0$, then the feasible set given by $\mathcal{A}_4$. For $a_5 \neq 0$, the inflection points solve $y^2(1-y)^2 M''(y) = 0$, which for $k=5$ results in the quartic $2a_5 y^2(1-y)^2 + a_2(2y-1) + 0.5a_3 = 0$, yielding the four candidates. Apply the change of variable $y = u + 0.5$ to produce the depressed quartic $u^4 + Au^2 + Bu + C$ with $A = -1/2$, $B = \bar{a}_2$, and $C = \bar{a}_3/4 + 1/16$. Because $a_2 > 0$ and $a_5 \neq 0$, we have that $B \neq 0$, allowing us to rule out the biquadratic case. It is well known that y_1 to y_4 are the four solutions of the quartic in $\mathbb{C}$, where x is a solution of the cubic equation $(2x - A)(x^2 - C) - B^2/4 = 0$ (which we find by applying the change of variable $x = t - 1/12$ and solving the depressed cubic $t^3 + pt + q = 0$).

Because $-2a_2 \leq a_3 \leq 2a_2$, then at least two solutions are real, and one of the solutions is always outside $[0, 1]$. If $\Delta \leq 0$, then there are four real solutions (with multiplicity if $\Delta = 0$). If $\Delta > 0$ then the solution on $(0, 1)$ is necessarily a mode, and given by y_2 when $a_5 > 0$ and y_3 when $a_5 < 0$. And if $\Delta \leq 0$, then the solutions on $(0, 1)$ are y_2, y_3, and y_4 when $a_5 > 0$ (of which y_2 and y_4 are modes), or y_1, y_2, and y_3 when $a_5 < 0$ (of which y_1 and y_3 are modes). Finally, we impose the mode to be positive at these inflection points. Say y_2 is an inflection point and $a_5 \neq 0$. The requirement that $M'(y_2) \geq 0$ becomes $a_4 \geq -\frac{a_2 + a_3(y - 0.5)}{y(1-y)} - a_3\ell(y) - 2a_5(y + 0.5)$, or $a_4 \geq a_5 g(y_2)$. Hence, if in addition y_4 is an inflection point, then $M'(y_2), M'(y_4) \geq 0$ becomes $a_4 \geq a_5 g(y_2)$ and $a_5 g(y_4)$. If $a_5 > 0$, then this is equivalent to $a_4 \geq a_5 \max\{g(y_2), g(y_4)\}$; and if $a_5 < 0$, to $a_4 \geq a_5 \min\{g(y_2), g(y_4)\}$. $\square$

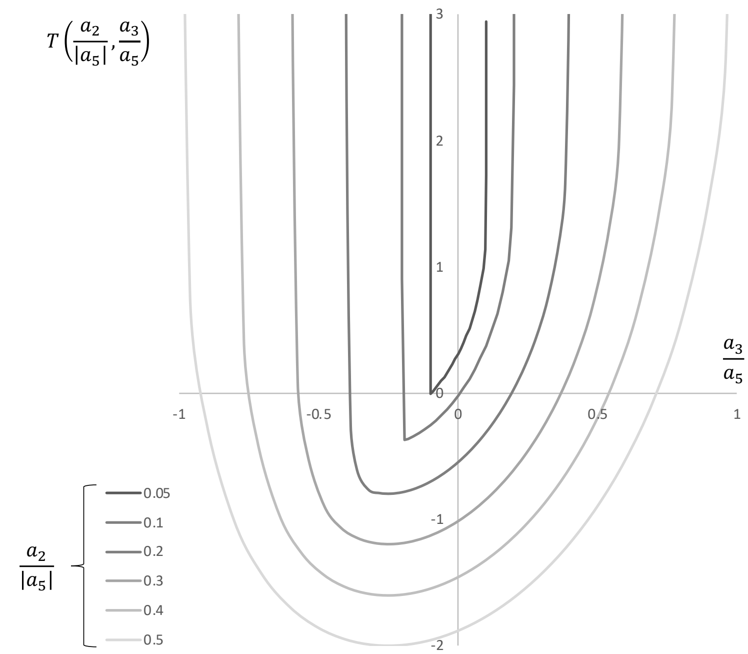

Figure 8 **A 5-metalog with $a_2 > 0$ and $a_5 \neq 0$ is feasible if and only if $-2a_2 < a_3 < 2a_2$ and $a_4 \geq |a_5| T\left(\frac{a_2}{|a_5|}, \frac{a_3}{a_5}\right)$.**