The Metalog Distributions: Virtually Unlimited Shape Flexibility, Combining Expert Opinion in Closed Form, and Bayesian Updating in Closed Form

First version: 2020-7-15

This version: 2021-11-15

Thomas W. Keelin^{1,*}

Ronald A. Howard²

¹ Keelin Reeds Partners, Menlo Park, CA.

www.metalogs.org, www.keelinreeds.com.

*Correspondence: tomk@keelinreeds.com.

² Stanford University

Abstract: Users of probability distributions frequently need to convert data (empirical, simulated, or elicited) into a continuous probability distribution and to update that distribution when new data becomes available. Often, it is unclear which traditional probability distribution(s) to use, fitting to data is laborious and unsatisfactory, little insight emerges, and updating with Bayes rule is impractical. Here we offer an alternative -- a family of continuous probability distributions, fitting methods, and tools that: provide sufficient shape and boundedness flexibility to closely match virtually any probability distribution or data set; involve a single set of simple closed-form equations; stimulate potentially valuable insights when applied to empirical data; are simply fit to data with ordinary least squares; are easy to combine in closed form when weighting the opinion of multiple experts; and, under certain conditions, are easily updated in closed form according to Bayes rule when new data becomes available. The Bayesian updating method is presented in a way that is readily understandable as a fisherman updates his catch probabilities when changing the river on which he fishes. While metalog applications have been shown to improve decision-making, the methods and results herein are broadly applicable to virtually any use of continuous probability in any field of human endeavor. Diverse data sets may be explored and modeled in these new ways with freely available spreadsheets and tools.

Key Words: metalog, probability distribution, probability forecasts, quantile forecasts, Bayesian inference, expert elicitation, expert combination, linear opinion pooling, Vincentization, data-fitting, quantile distributions, quantile-parameterized distributions, decision analysis, applied statistics.

1. Introduction

Here we introduce two useful, complementary, and immediately practical results for representing data as a continuous probability distribution and updating that distribution when new data becomes available. In introducing the metalog distributions, Keelin (2016) showed they can improve decision-making in an actual decision analysis. The results of this paper build on this theme by introducing new and extended metalog methods.

In Section 2, we show that the metalog distributions have virtually unlimited flexibility to take on whatever shape and bounds the data and context may suggest. This is a major advance over the widely used Pearson family of distributions (Pearson 1893, 1895, 1901, 1916), developed over a century ago, which include the normal, uniform, beta, gamma, chi-squared, Student-t, and six others. Whereas the Pearson distributions can match up to the first four moments of the data, the metalogs can match any number of moments, as well as, in some cases, shapes for which moments do not exist. For a given set of moments, the Pearson distributions offer no choice of bounds, whereas the metalogs offer a choice of

unbounded, semi-bounded, and bounded distributions. While the Pearson distributions and other traditional distributions include complex equations and intractable integrals, the metalog equations are simple and closed form for both cumulative and density functions. Whereas the Pearson and other traditional distributions typically require non-linear, iterative methods for parameter estimation from data, the data *are the parameters* of metalog distributions, because metalogs belong to the class of quantile-parameterized distributions (Keelin and Powley 2011) that require only an ordinary least squares calculation to obtain coefficients.

While many advantages of metalogs were shown previously (Keelin 2016; Keelin, Chrisman, and Savage 2019), we show in Section 2 the new result that the metalog distributions are not just more shape flexible than the Pearson family and other traditional distributions but have virtually unlimited shape flexibility. We show further that they can approximate virtually any traditional distribution arbitrarily closely with a single set of simple equations, provide unexpected insights in empirical data exploration, and be considered as candidates for modeling diverse data sets in virtually any field. To illustrate the usefulness of this result, we offer three examples: eliciting and combining expert opinion, including particularly simple and appealing ways to do so with metalogs; exploring scientific or financial data, illustrated by the potential bimodality of returns on investing in gold; and the ability of metalogs to mimic classical distributions, suggesting that a single set of metalog equations can be used in their place with minimal loss of accuracy. With the capability enabled by freely available tools (<http://www.metalogdistributions.com/software.html> accessed 08.18.2021), the new implication is that these distributions, methods, and tools may be immediately advantageous for uses in engineering, finance, medicine, science, operations management, technology, and other fields.

In Section 3, we show that the metalog distributions, initially parameterized by data, can be updated according to Bayes' rule in closed form when new data emerges. Updating according to Bayes' rule is important because it ensures adherence to the fundamental logic of probability. Yet many updating procedures in practical use violate Bayes' rule because Bayes' rule has been difficult to implement in practice. For more than a century, the statistics community has dithered to little avail about subjective vs. objective probabilities and how, if at all, subjective prior probabilities can be avoided. In contrast and in parallel, scientific leaders of the last century, those faced with important real-world problems, have used Bayes rule to help achieve practical and critically important successes. Such successes include the breaking the German enigma code in World War II, demonstrating that smoking causes lung cancer, finding a missing H-bomb and enabling better business decision-making (McGrayne 2011). In our view, consistency with logic and value in practice win decisively over useless arguing about objectivity. The objectivist (frequentist) view only works when there is so much data that all prior information is irrelevant. Many practical applications do not meet this condition. The Bayesian view always works. Most fundamentally, probabilities are not a state of the world but a measure of how much we know about it. When new data improves our knowledge, it does not necessarily make prior data irrelevant (Berger 2013). Both must be appropriately considered.

So, while users of probability in any field should update in a way that is consistent with Bayes' rule whenever possible, the practical difficulties of doing so have been significant. First, probability distributions that can accommodate the actual shape and bounds of the data have not previously existed. The Section 2 result is a solution to this problem. Second, the parameters of most traditional distributions and most statistical models are non-linearly related to the data. The introduction of conjugate priors fifty years ago (Raiffa and Schlaifer 1961) was a major advance in closed form solutions to Bayesian updating, but these conjugate priors are also shape and bounds limited. Working around such limitations leads to shortcuts in parameter estimation with new data that are logically inconsistent with Bayes' rule, or to well-intentioned logically consistent approaches like Markov Chain

Monte Carlo (Wikipedia accessed 25.02.2020) that are computationally cumbersome. In Section 3, we offer a novel, simple, and practical alternative. When new data becomes available, update an “any shape” prior distribution on a phenomenon of interest to an “any shape” posterior using nothing more than the prior data, new data, and ordinary linear regression. We show that such updating is consistent with Bayes’ rule when using metalog distributions and certain practical and transparent assumptions.

2. Metalogs: Virtually Unlimited Shape Flexibility

The metalog (short for “meta-logistic”) is a generalization of the logistic distribution (Balakrishnan 1991; Logistic distribution Wikipedia accessed 25.02.2020). The quantile function of the unbounded metalog, $M(y|\mathbf{a})$, is a linear combination of a row vector of metalog basis functions $\mathbf{g}(y) = (g_1(y), \dots, g_k(y))$ and column vector of coefficients $\mathbf{a} = (a_1, \dots, a_k)$: $M(y|\mathbf{a}) = \mathbf{g}(y)\mathbf{a}$, where $y \in (0,1)$ is cumulative probability. Basis functions $\mathbf{g}(y)$, closed-form probability density function (PDF), and feasibility requirement are provided in Appendix A.

Analogous to the relationship between normal and lognormal distributions, if the log of a variable is metalog distributed, then that variable is log metalog distributed. The log metalog is a semi-bounded distribution. Similarly applying a logit transformation yields the logit metalog, a bounded distribution. All metalogs have closed-form quantile functions and PDFs, as shown in Appendix A, and may be conveniently fit to data with ordinary least squares.

Like a Taylor series, a metalog may have any number of terms k . The more terms, the more flexible it is to match a given shape. Unbounded, semi-bounded, and bounded metalogs with k terms have $k - 2$, $k - 1$, and k shape parameters, respectively. While there is no limit to the number of terms, 2-16 terms are typically used in practice.

Previously, it was shown (Keelin 2016) that metalogs parameterized by 105 cumulative distribution function (CDF)-data points from each of 30 traditional source distributions converge rapidly to those source distributions in terms of Kolmogorov-Smirnov (KS) distance (Wikipedia accessed 25.02.2020). Moreover, the source vs metalog PDFs for all 30 distributions become visually indistinguishable as the number of metalog terms k approaches ten. The KS distance to three significant figures is ≤ 0.001 in all cases. With the following theorem, we now show that this result is far more general.

Theorem 2.1. Metalog Flexibility: Any probability distribution with a continuous quantile function can be approximated arbitrarily closely by a metalog quantile function.

Proof: Let $\delta > 0$ be an arbitrarily small positive number and let $Q(y)$ be a continuous quantile function defined on the probability interval $y \in [\delta, 1 - \delta]$. By the Weierstrass approximation theorem (<https://mathworld.wolfram.com/WeierstrassApproximationTheorem.html> accessed 08.18.2021), for every $\varepsilon > 0$, there exists a polynomial $P(y)$ such that $|Q(y) - P(y)| < \varepsilon$ for all $y \in [\delta, 1 - \delta]$. By setting the a_j ’s to zero for all terms $a_j g_j(y)$ that include factor $\ln(y/(1 - y))$ in the metalog quantile function (Table A.1), $M(y)$ reduces to a polynomial. Therefore, there exists a metalog distribution $M(y)$ such that $|Q(y) - M(y)| < \varepsilon$ for all $y \in [\delta, 1 - \delta]$.

While this theorem guarantees the existence of a such a metalog for any continuous quantile function, it does not guarantee its feasibility. It does guarantee, however, that there exists a metalog that is everywhere within ε of being feasible for any arbitrarily small positive number ε , which should be

sufficient for most practical applications. In practice, feasibility must generally be checked rather than assumed.

Nor does this theorem provide a procedure for how to find the metalog α coefficients that correspond to an arbitrary continuous quantile function. But in practice, this is not an issue. The linearity of the metalog quantile function in α means that one can simply determine these coefficients from data in closed form by linear regression (ordinary least squares). Moreover, if one assumes the usual Gaussian error structure for linear regression, which we discuss Section 3, the ordinary-least-squares estimation of these coefficients is both maximum likelihood and unbiased. Thus, metalog distributions can be effectively fit to data without the time-consuming, non-linear, iterative parameter estimation methods required for many other distributions. While the metalog flexibility theorem is proven above for the unbounded metalog, it applies equally to semi-bounded and bounded metalogs since these are based on transformed data being unbounded-metalog distributed. More broadly, it applies equally to any QPD (quantile-parameterized distribution per Keelin and Powley's 2011 definition) that includes an unlimited number of polynomial terms.

Beyond showing the tight fit of metalogs to 30 traditional distributions, we previously used metalog shape-flexibility to explore various empirical data sets (Keelin 2016). In a fish-biology sample of the weights of 3,474 steelhead trout migrating up the Babine River in British Columbia to spawn, the best-fit metalog PDF was found to be bimodal. This bimodality is explained by the simultaneous presence in the river of first- and second-time spawners. In hydrology, based on 95 years of USGS maximum-gauge-height data for the Williamson River in Chiloquin, Oregon, the metalog shows an unusual PDF shape. This suggests further research to determine whether this shape is unique to the Williamson or a characteristic of multiple rivers, and if so why. In finance, we used Monte Carlo simulation to consider uncertainty in future returns from a portfolio of financial assets. We observed that the metalog PDFs parameterized by the simulation results for a bidding decision had unexpectedly short tails. This observation led to the unexpected discovery that three-branch discretization, a widely accepted and commonly used modelling assumption in decision analysis, was the source of the problem. Better decisions resulted by using metalogs to correct this problem. Here, we provide additional examples to illustrate Theorem 1.

2.1. Eliciting Expert Opinion

In many cases, three- to five-term metalog distributions provide sufficient shape flexibility to model expert opinion on a continuous uncertainty. In the original paper, Keelin (2016) used three term metalogs to model expert elicited probability distributions over 259 financial assets as part of a bidding decision analysis and showed that use of metalog's resulted in a better bidding decision compared to modeling these same uncertainties with three-branch discretization, a widely accepted and commonly used modelling assumption in decision analysis. Bratvold, et. al. (2020) used three-term log metalogs to model 56 oil companies' documented oil-field production uncertainties and observed systematic optimistic bias compared to actuals.

Statistics Canada (Dion, Galbraith, and Sirag 2020) used five-term metalogs to model expert opinion elicitation for future Canadian fertility rates. With an interactive spreadsheet, 17 experts were asked to adjust their quantile parameters with sliding bars until the resulting probability density graph accurately reflected their probabilistic beliefs about Canada fertility rate in 2043. This approach yielded the metalog quantile parameters and corresponding coefficients in Table 2.1 and the corresponding five-term metalog CDFs and PDFs in Fig 2.1.

In all 332 cases, metalogs provided sufficient shape and bounds flexibility to effectively model expert opinion and to be credibly used as a basis for subsequent analyses. There were no cases reported in which the assessed expert opinion violated the feasibility constraint of the three-five term metalogs selected for use.

2.2. Combining Expert Opinion

Using the weights (Table 2.1.C), self-assessed by the experts based on their knowledge level, Statistics Canada weighted the PDFs (Fig. 2.1.B) with using linear opinion pooling (Genest and Zidek 1986) to develop the combined-opinion PDF and CDF (Fig 2.1 orange curves). A shortcoming of this method, in according to Dion, Galbraith, and Sirag (2020), is “Despite the fact that experts’ responses are parametrized by metalog distributions, the resulting mixture distributions for fertility, mortality, and immigration are not metalog distributions, and do not belong to any defined parametric family. Characteristics such as central moments and quantiles are derived using numerical methods.”

Here we revisit that result by offering an alternative: a simple method for combining metalog expert opinions in such a way that, unlike linear-opinion pooling, the combined opinion is determined in closed form as a metalog.

	A					B					C
	Q0.05	Q0.25	Q0.50	Q0.75	Q0.95	a ₁	a ₂	a ₃	a ₄	a ₅	weight
expert 1	1.40	1.53	1.60	1.67	1.80	1.60	0.07	0.00	-0.01	0.00	0.07
expert 2	1.44	1.55	1.60	1.65	1.76	1.60	0.06	0.00	-0.07	0.00	0.06
expert 3	1.56	1.59	1.62	1.65	1.69	1.62	0.01	-0.00	0.10	0.04	0.04
expert 4	1.31	1.50	1.62	1.72	1.81	1.62	0.05	-0.04	0.20	0.00	0.04
expert 5	1.38	1.43	1.47	1.54	1.64	1.47	0.03	-0.00	0.08	0.21	0.04
expert 6	1.30	1.44	1.55	1.66	1.80	1.55	0.04	0.00	0.26	0.00	0.06
expert 7	1.34	1.54	1.60	1.76	2.15	1.60	0.21	-0.04	-0.49	0.99	0.06
expert 8	1.50	1.61	1.70	1.76	1.90	1.70	0.07	0.12	-0.00	-0.78	0.07
expert 9	1.25	1.50	1.65	1.76	1.95	1.65	0.12	0.00	-0.01	-0.26	0.07
expert 10	1.30	1.41	1.50	1.68	2.00	1.50	0.11	0.03	0.07	0.54	0.05
expert 11	1.40	1.55	1.65	1.73	1.88	1.65	0.08	0.05	0.03	-0.36	0.07
expert 12	1.39	1.60	1.70	1.80	2.01	1.70	0.13	0.00	-0.14	0.00	0.07
expert 13	1.31	1.40	1.49	1.59	1.68	1.49	0.01	0.00	0.33	0.00	0.07
expert 14	1.21	1.47	1.60	1.73	1.99	1.60	0.16	0.00	-0.18	0.00	0.06
expert 15	1.30	1.38	1.45	1.52	1.60	1.45	0.03	0.00	0.16	0.00	0.07
expert 16	1.40	1.44	1.52	1.63	1.80	1.52	0.04	0.06	0.21	0.00	0.07
expert 17	1.53	1.62	1.67	1.71	1.78	1.67	0.04	-0.01	0.00	0.00	0.05
weighted	1.37	1.50	1.59	1.68	1.84	1.59	0.08	0.01	0.03	-0.00	1.00

Table 2.1 Expert Elicited Fertility Rates and Corresponding Metalog Coefficients. A. Quantile uncertainty range for Canada fertility rate (%) in 2043 elicited from 17 experts and weighted average. **B.** Corresponding metalog coefficients and weighted average. **C.** Weights for the experts, which sum to 1.0.

Metalog distributions are particularly well-suited to combining expert opinion as a weighted average of their quantile functions. For example, this method simply posits that the combined median is the weighted average of the individual medians, that the combined quartiles are the weighted average of the individual quartiles, and so on. Known as Vincentization or Vincent Average (Wikipedia accessed 30.04.2021) after the biologist who first published it (Vincent 1912), this approach may be considered as

Fig. 2.1. Probability Distributions over Canada Fertility Rate, 2043. **A.** Five-term metalog CDFs for 17 experts (light gray) parameterized by five elicited quantiles (light gray dots) from Table 2.1.A; probability-weighted average (orange curve); and quantile-weighted average metalog (blue curve) parameterized by weighted average quantiles (blue dots) from the last row in Table 2.1.A. **B.** Corresponding PDFs.

an alternative to linear opinion pooling and other methods that weight expert-opinion *probabilities for a given quantile* rather than *quantiles for a given probability* (Genest and Zidek 1986). Research comparing these methods concluded that “averaging quantiles for a given probability is as good or better than averaging probabilities as a method for aggregating probability forecasts.” (Lichtendahl, Grushka-Cockayne, and Winkler 2013). Moreover, Vincentization has the key advantage that it is closed (Thomas and Ross 1980) under location-scale distribution families (Location-scale family, Wikipedia accessed 04.01.2021). That is, if the experts’ distributions were normal, Cauchy, exponential, or logistic, the same would be true of the Vincent consensus (Genest and Zidek 1986). Here we extend this result by showing that if the experts’ distributions are feasible k -term metalogs, then so is the Vincent consensus.

Theorem 2.2. Metalog Combination of Expert Opinion: Let $M_{k,i}(y|\mathbf{a}_i)$ be the k -term, feasible metalog quantile function of the i th expert and let w_i be the weight assigned to the opinion of that expert. Then, using Vincentization, the consensus-opinion metalog quantile function is $M_k(y|\mathbf{a})$ where $\mathbf{a} = \sum_i w_i \mathbf{a}_i$ and $M_k(y|\mathbf{a})$ is guaranteed to be feasible.

Proof: According to Vincentization, for every value of y , the quantile of the combined opinion is the weighted average of the experts’ quantiles. Thus, the consensus-opinion quantile function $M_k(y|\mathbf{a})$ is the weighted average of the expert quantile functions: $M_k(y|\mathbf{a}) = \sum_i w_i M_{k,i}(y|\mathbf{a}_i)$, which reduces to the above result. Like other QPDs, the set of feasible metalog coefficients is convex (Keelin and Powley 2011). Thus, so long as the individual experts’ metalogs are feasible, their convex combination, the consensus metalog $M_k(y|\mathbf{a})$, is also feasible.

Note that an analogous theorem also holds, more broadly and by the same logic, for any QPD (quantile-parameterized distribution that satisfies Keelin and Powley’s 2011 definition). Note also that such results can be equivalently stated in terms of quantiles. As illustrated in Table 2.1A, if $\mathbf{q}_i$ is the elicited quantile vector of the i th expert, then the consensus quantile vector is $\mathbf{q} = \sum_i w_i \mathbf{q}_i$. This follows from the linear relationship between quantile parameters and coefficients, $\mathbf{q} = \mathbf{Y}_r \mathbf{a}$, where $\mathbf{Y}_r$ is the invertible matrix defined in Section 3.1.

Applying this result to the 17 distributions from Table 2.1 yields the blue curves in Fig 2.1. Note that the resulting combined expert opinion appears reasonable as a consensus, though is it slightly different from the linear-opinion-pooling (orange curves in Fig. 2.1) result (Dion, Galbraith, and Sirag 2020). The key advantage of the former is its simple and exact closed form. In contrast, the latter, lacking a closed form, is an approximation based on a numerical weighting of densities or of Monte-Carlo-simulation histograms.

The importance of this result is that whenever the opinions of multiple experts are expressed as k-term metalogs (per Section 2.1) and weights can be assigned to the various experts, the combined opinion is a k-term metalog whose parameters are trivially easy to calculate in closed form.

2.3. Data Exploration

Fig. 2.2.A shows the empirical distribution of quarterly returns on investing in gold for the 20+ year period from March 31, 1998 through September 30, 2019. This data is provided in Appendix A. An eight-term metalog distribution, fit by ordinary least squares, closely matches this data. The corresponding metalog PDF in Fig. 2.2.B reveals an unexpected bimodal shape.

Fig. 2.2. Metalog exploration of empirical data: returns on investing in gold. A. Empirical data for quarterly returns on investing in gold, March 31, 1998 – September 30, 2019, and eight-term metalog CDF fit to that data. B. Corresponding metalog PDF revealing an unexpected bimodal shape.

If this bimodality were to be verified by further research (by methods such as bootstrapping, training vs. test data sets, expanded time periods, alternative time-period segmentations, and underlying analysis of central banks' buying and selling behaviors), it could have significant implications for investor trading decisions and design of derivative trading instruments.

Our purpose here is not to prove the bimodality of returns on investing in gold but rather to illustrate the use of metalogs as a data exploration and visualization tool that may provide new insights worthy of further exploration. The importance of this result is that the probabilistic shape of virtually any data set can be easily explored with metalogs. Appendix F provides guidelines to aid such explorations.

2.4. Mimicking Traditional Distributions

By Metalog Flexibility Theorem, any distribution with a continuous quantile function can be represented about as well by a metalog as by its own equations. While this theorem is a theoretical finding based on

an infinite number of metalog terms, in practice, eight-ten metalog terms are typically enough to make the source-distribution CDFs and PDFs visually indistinguishable from their metalog counterparts. Fig. 2.3 illustrates this with nine-term metalogs for a range of source distributions and parameter settings.

Since there is a metalog to mimic virtually any traditional distribution, there is no longer a need to “guess” a-priori which traditional distribution(s) to consider when fitting empirical data as in Section 2.3. When used as the first step in data exploration, metalog shapes may suggest other candidate distributions and/or insights for further exploration. As illustrated in Appendix C for the Pareto distribution, this result applies even to a range of highly skewed, heavy tailed distributions.

Fig 2.3. Lognormal, beta, gamma, and Weibull distributions and metalog representations thereof. Parameterized by nine CDF points from a range of source distributions, nine-term metalog PDFs (dashed orange curves) are visually indistinguishable from the source-distribution PDFs (blue curves). In all cases, the metalog bounds match those of the source distribution. For the lognormal distributions (upper left), $\mu = 0$; A-F correspond to $\sigma = (0.3, 0.5, 0.8, 1.0, 1.5, 2.0)$. For the gamma distributions (lower left), $\beta = 1$; A-F correspond to $\alpha = (1, 2, 3, 4, 5, 6)$. For the beta distributions (upper right), A-F correspond to (α, β) pairs $(3, 5)$, $(8, 8)$, $(5, 3)$, $(2, 2)$, $(1, 1)$, and $(0.8, 0.5)$ respectively. For the Weibull distributions, $\lambda = 1$; A-F correspond to $k = (0.5, 0.8, 1.0, 1.5, 2.0, 4.0)$. In each case, the metalog is parameterized by the nine quantiles of the source distribution that correspond to the y values $(0.001, 0.02, 0.10, 0.25, 0.5, 0.75, 0.90, 0.98, 0.999)$.

2.5. Universality and Computational Considerations

For all but a few extreme distributions like the Cauchy, a 16 term metalog is more than adequate to mimic them, and, depending on the desired degree of accuracy, many of the terms may have coefficients of 0. Thus, from a computational perspective, a single polynomial-based equation (the metalog quantile function) plus three simple transformations of it to model boundedness, if any, may be implemented

within any computational platform. Thereafter, the specification of most any continuous distribution is reduced to 16 α -coefficients and two optional bounds for a total of 18 numeric parameters. In contrast, a current catalogue of probability distributions (Wikipedia, List of Probability Distributions: https://en.wikipedia.org/wiki/List_of_probability_distributions, accessed 6.09.2021) lists roughly 115 non-metalog named continuous distributions, each with its own unique algebraic formula, restrictive properties, fitting procedures, and shape-flexibility limitations. Neither individually nor in reasonable combination do these come close to covering the shape and boundedness flexibility of the metalog distribution, which may be considered universal for this reason.

The metalogs can also play a unique role in Monte Carlo simulation and the cross-platform passing of simulation results. Given that they are closed-form quantile functions, they deliver streams of random variates when driven by uniform random samples. When the random number generator is also a cross-platform function, say with a four-dimensional seed, then a total of 22 numeric parameters translates into streams of random variates of any length which yield identical results, trial by trial in R, Python, Excel or any other computational platform.

2.6. Summary: Virtually Unlimited Shape Flexibility

In this section, we have shown that metalogs have virtually unlimited shape flexibility. That is, they can match any continuous quantile function arbitrarily closely. Moreover, metalogs can have whatever boundedness is appropriate, have simple closed form equations, and are simply fit to data with ordinary least squares. For use in practice, the metalog system provides an alternative to classical distributions, which are less flexible and typically more difficult to use.

3. Metalog Bayesian Inference

In a classic paper on the foundations of decision analysis, Howard (1970) showed how an uncertain long-run frequency of a coin coming up “heads” could be updated in light of new information according to Bayes rule. Specifically, if a decision maker’s prior probability distribution over p , the “long-run” frequency of the coin yielding “heads” on any given toss, can be described by a beta distribution with prior parameters n_0 and k_0 , then in light of new experimental information that k “heads” occurred in n tosses, that decision-maker’s posterior distribution over p would be a beta distribution with parameters $n_0 + n$ and $k_0 + k$. This well-known “beta binomial” Bayesian updating is consistent with Bayes rule because the beta distribution is a conjugate prior (Raiffa and Schlaifer 1961) of the binomial distribution.

Here we extend closed form Bayesian updating to the case where the uncertainty of interest to be updated is defined not by a scalar probability over a discrete event (like p), but by a probability density function over a continuous variable. Using metalogs, we show how the probability density function (PDF) over the uncertainty of interest can be updated in closed form in light of new data according to Bayes rule. And because this PDF is a metalog, it can take on whatever shape is dictated by combining the decision maker’s prior with new data, subject only to feasibility constraints and certain assumptions explained below.

We use a tangible situation to concretely illustrate key concepts. While fictitious, the settings and numbers are realistic. Consider a fly fisherman, Norman, named in honor of in honor of Norman Maclean, whose 1976 novel (Maclean 2009) later portrayed in a Robert Redford movie, inspired generations of fly fishermen and appropriate resource conservation. Norman has extensive experience in catch-and-release fly fishing on great rivers of Montana. To Norman, like most fly fishermen, size matters. Bigger trout are more interesting, wily, and challenging to catch than smaller ones.

Norman's plans include fly fishing on Oregon's Williamson River. He has not been to Oregon previously and knows little about the Williamson. In particular, Norman *does not know* that the trout on the Williamson (Fig. 3.1.B), due to genetics, food availability, and other environmental factors, typically grow bigger than those in the Montana rivers of his experience (Fig. 3.1.A). Anticipating this trip, Norman chooses to encode his probability distribution over the size of the next trout he will catch and release on the Williamson. We term this his *target* distribution. He is interested in this in its own right and further because it may help inform decisions such as his choice of fly rod, strength of leader material, and size of landing net.

3.1. Metalog Regression

Norman considers trout size data from his Montana experience (Fig. 3.2.B), which he has recorded in his personal journal over many years. To convert this data into a continuous distribution, he chooses a

Fig. 3.1 trout size comparison. A. Big Hole River, Montana. B. Williamson River, Oregon.

metalog distribution and calculates its $\mathbf{a}$ coefficients by ordinary least squares: $\mathbf{a} = (\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\mathbf{x}$, where $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ are the x- and y-coordinates of the data in Fig. 3.2.B, and $\mathbf{Y}$ is the $n \times k$ matrix whose (i, j) element is $g_j(y_i)$, and $y_i = i/(n + 1)$. Using a three-term metalog, this calculation yields $\mathbf{a} = (13, 1.8, 1.1)$ and the CDF plot, Fig. 3.2.A.

This same distribution is equivalently parameterized with the *quantile parameters* $\mathbf{q} = (10, 13, 18)$, Fig. 6D, that correspond to *reference probability vector* $\mathbf{y}_r = (.10, .50, .90)$, Fig. 3.2.C. Given $\mathbf{y}_r$, there is a linear one-to-one relationship between quantile parameters and *coefficient parameters* $\mathbf{a}$: $\mathbf{q} = \mathbf{Y}_r \mathbf{a}$ and $\mathbf{a} = \mathbf{Y}_r^{-1} \mathbf{q}$, where $\mathbf{Y}_r$ is the $k \times k$ matrix whose (i, j) element is $g_j(y_{r_i})$. $\mathbf{y}_r = (y_{r_1}, \dots, y_{r_k})$ may be selected as any k -length vector of strictly increasing elements in the open interval $(0, 1)$ such that $\mathbf{Y}_r$ is invertible. Once selected, we keep $\mathbf{y}_r$ fixed for a given application. This enables us to parameterize the prior target distribution with interpretable quantile parameters $\mathbf{q}$ in place of the uninterpretable coefficients $\mathbf{a}$. The metalog quantile function and metalog regression are equivalently expressed in terms of quantile parameters: $M(y|\mathbf{q}) = g(y)\mathbf{Y}_r^{-1}\mathbf{q}$ and $\mathbf{q} = \mathbf{Y}_r(\mathbf{Y}'\mathbf{Y})^{-1} \mathbf{Y}'\mathbf{x}$ respectively.

3.2. Target Prior

To quantify his target prior distribution, Norman begins by interpreting the quantile parameters $\mathbf{q} = (10, 13, 18)$ that resulted from metalog regression. He notes that these are consistent with his prior data and future Montana expectations: 10 percent of his Montana trout were 10 inches or less; 50 percent were 13 inches or less; and 90 percent were 18 inches or less. He is also satisfied that the shape and tails of this distribution are realistic. (If not, he could use a metalog with more terms, $k > 3$, and/or a semi-bounded or bounded metalog per Table A.2.) Since initially he has no reason to believe that Williamson trout differ in size from the Montana trout of his experience, he is comfortable adopting this distribution (Fig. 3.2.A) as his target prior.

Note that, if prior data as in Fig. 3.2.B are not available, the quantile parameters (Fig. 3.2.D) may be determined by direct assessment (Spetzler and Stall von Holstein 1975, Howard and Abbas 2016), simulation modelling, or any other method. To specify the target prior, only $\mathbf{q}$ is required, not prior data per se.

3.3. Updating According to Bayes theorem

Norman recognizes that the coefficients $\mathbf{a}$, and equivalently the quantile parameters $\mathbf{q}$, of his target prior are themselves uncertain and thereby subject to updating as he catches, measures, and releases Williamson trout. When new data are observed, assuming no change in the underlying random process, it would be natural to combine the new data with the prior data and calculate an updated metalog, parameterized by a posterior $\mathbf{a}$, via metalog regression on the combined data set. Under what conditions, if any, would updating in this simple intuitive way be Bayesian?

Fig. 3.2. Target distribution with quantile-parameter uncertainty. A. Target CDF. B. Data from which target CDF is determined (brown dots). C. Elements of reference probability vector $\mathbf{y}_r$. D. Elements of quantile parameter vector $\mathbf{q}$. E. Marginal PDFs over quantile-parameter elements. F. Encoded data prior (orange dots).

We now show that this updating procedure is Bayesian as a special case of the well-known Bayesian linear model (Bayesian linear regression, Wikipedia accessed 25.02.2020). We use the same Gaussian error structure, likelihood function, conjugate priors, and updating equations, interpreting metalog Bayesian inference as a special case thereof.

3.4. Gaussian Error Model

Using the Gaussian error model widely assumed for linear regression, we define the metalog random process by adding a normally distributed error term $N(0, \sigma^2)$ to $M(y|\mathbf{a}) = \mathbf{g}(y)\mathbf{a}$. In Norman's case, standard deviation σ is easy to interpret. Getting an exact length measurement of a wild, live, slippery fish is not easy. Even though he uses a landing net marked in 1-inch intervals on its mesh, Fig. 3.3, potential measurement error remains.

In addition, the error term must include any considerations that may be “left out” of the model $\mathbf{g}(y)\mathbf{a}$. For example, Norman's three-term metalog may not have sufficient shape flexibility to capture all nuances of the source distribution from which that data are drawn. So, the error term must also include the possible effect of having left out higher order metalog terms. In Norman's case based on Fig. 3.3, it might be appropriate to model standard deviation $\sigma \approx 1$ inch as deterministic, which we assume for convenience of exposition below. In other cases, σ is better treated as uncertain, which implies extensions to the equations per Appendix A.

Fig. 3.3. Measuring a live trout is subject to measurement error

3.5. Prior Over Parameters

The multivariate normal distribution over coefficients, $\mathbf{a} \sim \mathcal{N}(\bar{\mathbf{a}}, \sigma^2 \mathbf{A}^{-1})$, is a conjugate prior (Raiffa and Schlaifer 1961) to the likelihood function implied by the Gaussian error model, where $\bar{\mathbf{a}}$ is the mean of the coefficients, and $\mathbf{A}$ is a $k \times k$ precision matrix which determines covariance matrix $\sigma^2 \mathbf{A}^{-1}$ up to the multiplicative scalar constant σ^2 . Thus, if the uncertainty about $\mathbf{a}$ can be modelled in this way, Bayesian updating is closed form. Since the quantile parameters $\mathbf{q}$ are a linear transformation of $\mathbf{a}$, $\mathbf{q} = \mathbf{Y}_r \mathbf{a}$, the implied distribution over quantile parameters is also multivariate-normal, $\mathbf{q} \sim \mathcal{N}(\bar{\mathbf{q}}, \sigma^2 \mathbf{Y}_r \mathbf{A}^{-1} \mathbf{Y}_r')$, where $\bar{\mathbf{q}} = \mathbf{Y}_r \bar{\mathbf{a}}$. It makes sense to Norman that, for updating purposes in his case, the marginal distributions over his quantile parameters, Fig. 3.2.E, would be symmetric and centered around his prior deterministic estimates, Fig. 3.2.D. Thus, so long as $\mathbf{A}$ can be reasonably and conveniently determined, the multivariate normal model seems sensible to him.

3.6. Updating Equations

Given the Gaussian error model, prior values $\bar{\mathbf{a}}_0$ and $\mathbf{A}_0$ and n new data $(\mathbf{x}_n, \mathbf{y}_n)$, the coefficients update according to Bayes' theorem in closed form (Bayesian linear regression, Wikipedia accessed 25.02.2020): $\bar{\mathbf{a}}_n = \mathbf{A}_n^{-1}(\mathbf{A}_0 \bar{\mathbf{a}}_0 + \mathbf{Y}_n' \mathbf{x}_n)$ where $\mathbf{A}_n = \mathbf{Y}_n' \mathbf{Y}_n + \mathbf{A}_0$. If $\mathbf{A}_0$ and $\bar{\mathbf{a}}_0$ can be determined by any method, these updating equations apply. Here we determine them by a simple and simplifying assumption: the data prior.

Definition 3.1. Data Prior: a state of information that consists of equally likely prior data $\mathbf{x}_{n_0}$ and the belief that no other information is relevant to updating with new data.

Equivalently, a data prior is a prior data set, Fig. 3.2.B for example, combined with the assumption of an irrelevant (uninformative) prior to that prior. Assuming a data prior simplifies the updating equations.

Theorem 3.1. Updating the Means of the Parameters Given a Data Prior: Given a data prior $\mathbf{x}_{n_0}$ and new data $\mathbf{x}_n$,

- A. the Bayesian linear model equation for calculating the posterior mean of the coefficients $\bar{\mathbf{a}}_n$ reduces to metalog regression $\bar{\mathbf{a}}_n = (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{x}$, where column vector $\mathbf{x}$ is the combined, sorted $(\mathbf{x}_{n_0}, \mathbf{x}_n)$ data and $\mathbf{Y}$ is the corresponding $(n_0+n) \times k$ metalog-regression matrix, and
- B. the posterior mean of quantile parameters is $\bar{\mathbf{q}}_n = \mathbf{Y}_r \bar{\mathbf{a}}_n$.

Proof: To prove A, consider the state of information that consists of equally likely prior data $\mathbf{x}_{n_0}$ and new data $\mathbf{x}_n$. Let column-vector $\mathbf{x}$ be the combined, sorted $(\mathbf{x}_{n_0}, \mathbf{x}_n)$ data and let $\mathbf{Y}$ be the corresponding $(n_0+n) \times k$ metalog-regression matrix. If information prior to the data prior $\mathbf{x}_{n_0}$ is irrelevant, $\mathbf{A}_0$ and $\mathbf{A}_0 \bar{\mathbf{a}}_0$ must be zero in the above updating equations. Determining $\bar{\mathbf{a}}_n$ from the combined data $\mathbf{x}$ then reduces to $\bar{\mathbf{a}}_n = \mathbf{A}^{-1}(\mathbf{Y}'\mathbf{x})$ where $\mathbf{A} = \mathbf{Y}'\mathbf{Y}$, which further reduces to metalog regression on the combined data: $\bar{\mathbf{a}}_n = (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{x}$. To prove B, note that, as discussed in Section 3.1, coefficients may always be interpreted in terms of quantile parameters according to a linear transformation, $\mathbf{q} = \mathbf{Y}_r \mathbf{a}$. Since the expected value of a linear transformation of a random variable is the linear transformation of its expected value, it follows that $\bar{\mathbf{q}}_n = \mathbf{Y}_r \bar{\mathbf{a}}_n$.

Since linear regression packages are widely available, the importance of Theorem 3.1.A is that the Bayesian linear model mean-updating equations need not be separately programmed. The posterior mean of the coefficients can be calculated by metalog regression. Moreover, by Theorem 3.1.B, the posterior mean of the quantile parameters is calculated in closed form by a subsequent matrix multiplication.

Note that these equations for updating the mean do not depend on σ , the measurement-error standard deviation, regardless of whether σ is treated as certain or uncertain. So, if the sole purpose of an application is to update the target-prior parameters with the mean of quantile parameters and/or coefficients to attain those the target posterior, no further calculations are required. However, if one also wishes to update the probability distributions over the coefficients and quantile parameters, such updating proceeds simply and in closed form. Note that, given σ , both distributions are multivariate normal by the conjugacy property of the Bayesian linear model and that the means of these distributions are given by Theorem 3.1. Their covariance matrices are a further implication of this theorem.

Corollary 3.1. Updating the Covariance Matrices of the Parameters Given a Data Prior: Given a data prior $(\mathbf{x}_{n_0}, \mathbf{y}_{n_0})$, new data $\mathbf{x}_n$, the Bayesian linear model, and a multivariate normal distribution over the coefficients, $\mathbf{a} \sim \mathcal{N}(\bar{\mathbf{a}}, \sigma^2 \mathbf{A}^{-1})$,

- A. the posterior covariance matrix of the coefficients is $\sigma^2 \mathbf{A}^{-1}$, where $\mathbf{A} = \mathbf{Y}'\mathbf{Y}$, and $\mathbf{Y}$ is the $(n_0+n) \times k$ metalog-regression matrix that combines the prior and new data, and
- B. the posterior covariance matrix over the quantile parameters is $\sigma^2 \mathbf{Y}_r \mathbf{A}^{-1} \mathbf{Y}_r'$.

Proof: Proof of A is within the proof of Theorem 3.1.A above. Proof of B follows from the linear relationship between quantile parameters and coefficients discussed in Section 3.1, $\mathbf{q} = \mathbf{Y}_r \mathbf{a}$, and from noting that a linear transformation of a multivariate-normal-distributed variable is multivariate-normal-distributed with such a transformed covariance matrix.

The importance of this corollary is that the covariance matrices are easily updated in closed form without any additional assumptions. Moreover, since $\mathbf{Y}$ in metalog regression is determined solely by the number of data in that regression, so is $\mathbf{A}$. So given σ , the covariance matrices over coefficients and quantile parameters, $\sigma^2 \mathbf{A}^{-1}$ and $\sigma^2 \mathbf{Y}_r \mathbf{A}^{-1} \mathbf{Y}_r'$ respectively, need not be assessed. They are automatically determined by the number of data $n_0 + n$, where n_0 is further discussed in Section 3.7.

Analogous results hold when σ is treated as uncertain. If uncertainty on σ can be modeled as an inverse gamma distribution, the conjugacy property of the Bayesian linear model still holds; the marginal distributions over the coefficients and quantile parameters become multivariate t distributions instead of multivariate normal; and all parameters can be still updated on closed form. Moreover, these results can be extended in closed form to the cases where the probability distribution over the target variable, rather than being an unbounded metalog as assumed heretofore in Section 3, is a semi-bounded or bounded metalog. Appendix A provides the corresponding equations.

3.7. Specifying a Data Prior

An obvious example of a data prior is where prior data has been observed and, beyond that data, the observer believes that nothing else is relevant to the calculation of his posterior when considering new data. We term this an *empirical data prior*. For example, Fig. 3.2.B considered in the context of Norman's next Montana trout may be an empirical data prior.

In many situations, however, such empirical data may be unavailable, of partial relevance, from a changed random process, or otherwise deserving of a different weighting compared to new data. We use an *encoded data prior* for this situation. For example, Fig. 3.2.B considered in the context of Norman's first Williamson trout would require an encoded data prior. Since Oregon's Williamson River may differ from his past experience, Norman believes that new Williamson data should reasonably be weighted more heavily than his prior Montana data, though his prior Montana data remain relevant and thus must not be ignored.

We note that a data prior is comprised of two pieces of relevant information. First, the data $\mathbf{x}_{n_0}$ indicate the location, scale, and shape of the prior distribution over the target variable. This information is contained, for example, in Norman's target prior (Fig. 3.2.A), regardless of whether it was determined by metalog regression over empirical data from his journal (Fig. 3.2.B), by direct assessment (Spetzler and Stall von Holstein 1975, Howard and Abbas 2016) of its parameterizing quantiles (Fig. 3.2.D), or by any other method. The second piece of relevant information is the number of prior data n_0 , which is implicitly provided by empirical data, but which is missing from a target prior otherwise determined. n_0 is important because, akin to the concept *equivalent sample size* in Bayesian statistics (Berger 2013), it provides *strength of prior* information. In principle, a target prior such as Norman's could be based on a just a whim. Or it could be based on thousands of highly relevant prior observations. To encode a data prior, we must assess or otherwise determine n_0 .

n_0 controls the weight placed on prior data relative to new data. Specifically, given n_0 prior data and n new data, updating effectively puts $\frac{n_0}{n_0+n}$ of its weight on prior data and $\frac{n}{n_0+n}$ on the new data.

Before going to the Williamson, Norman considers the number of Williamson trout he would have to catch in order that the new data would be about equally informative as his prior experience. Even though Norman has caught hundreds of Montana trout, he knows that the Williamson may be different. He believes that having caught ten trout on the Williamson would be more informative about the size of his next Williamson trout than the sum of his Montana experience. In contrast, three Williamson trout would be helpful, but less informative than his prior experience. He settles on $n_0 = 6$ as the number of Williamson trout that would be about equally informative as his prior experience.

Given n_0 , we use location and shape information from the target prior to complete the encoding. Given the metalog target prior in Fig. 3.2A, defined by metalog quantile parameters $(10,13,18)$ where $\mathbf{y}_r = (.10, .50, .90)$, we naturally assign $\bar{\mathbf{q}}_0 = (10,13,18)$ and compute $\bar{\mathbf{a}}_0 = \mathbf{Y}_r^{-1}\bar{\mathbf{q}}_0 = (13, 1.8, 1.1)$. Given target prior $M(\mathbf{y}|\bar{\mathbf{a}}_0)$ and strength of prior n_0 , our encoded data prior is $\mathbf{x}_{n_0}$, where $x_i = M\left(\frac{i}{n_0+1} \middle| \bar{\mathbf{a}}_0\right)$ for $i = 1, \dots, n_0$. As illustrated in Fig. 3.2.F, Norman's encoded data prior consists of n_0 points from the target prior spaced according to the probabilities we would ordinarily assign to such data. One may correctly observe that this encoding method leaves out variability that sampling n_0 points from the target prior with measurement error would naturally produce. But, consistent with symmetry of the prior over $\mathbf{a}$, $\bar{\mathbf{a}}_n$ in the updating equations is independent of variability σ .

While we like the simplicity and transparency of this encoding method, we note that a data prior may be encoded in any other way that satisfies its definition. A well-selected data prior may be considered better and more authentic than available prior data itself, particularly if that data is noisy, subject to variable measurement error, or other factors that might be different from the process one believes will generate new data looking forward. The ultimate test is to confirm with the decision maker or expert that there is no material information beyond the proposed data prior that is relevant to updating.

3.8. Updating Illustrated

Fig. 3.4 illustrates the updating of Norman's target distribution as he catches and releases his first one hundred Williamson trout. Though the Williamson source distribution from which new data are sampled is a three-term metalog with $\mathbf{q} = (14, 18, 24)$ with Gaussian error $\sigma = 1$ inch, this source distribution is hidden from Norman. But as he updates with more and more Williamson data, he systematically learns that Williamson trout are typically larger than his prior Montana experience of $\mathbf{q} = (10, 13, 18)$.

Fig. 3.5 shows how the marginal distributions over the quantile parameters $\mathbf{q}$ narrow and ultimately converge to those of the source distribution as new data are revealed. As shown in Appendix D, inspection of the covariance matrix $\sigma^2 \mathbf{Y}_r \mathbf{A}^{-1} \mathbf{Y}_r'$ reveals that its magnitude is approximately inversely proportional to the total number of data $n_0 + n$. Since a given number of data contain more information about the median than the tails, marginal distributions farther from the median are wider.

While this example uses three-term metalogs, metalogs with more terms can be used when greater shape detail is needed. Appendix E provides an example where nine-term metalogs reveal a bimodal source distribution as they are updated with new data.

Fig. 3.4. Updating target distribution from prior to posterior with new data. A. Source distribution, $q = (14, 18, 24)$, unknown to Norman, is well to the right of and has longer right tail than the prior, $q = (10, 13, 18)$. B. Williamson trout one, larger than prior expectations at 15 inches, moves posterior slightly to the right. C. Trout two, smaller than prior expectations at 10 inches, moves posterior back near original prior. D. Trout three through six, at 24, 18, 18, and 17 inches respectively, are significant new information indicating that Williamson trout are likely larger than prior expectations. This moves posterior significantly to the right and increases the length of the right tail on the PDF. E. Based on having caught one hundred Williamson trout, the posterior converges to the source distribution and the prior is nearly irrelevant.

Fig. 3.5. Updating quantile-parameter distributions as new data are revealed

A. With one new datum, posterior marginal PDFs move and narrow only slightly. B. With six new data, posterior PDFs narrow and move significantly right. C. With 100 new data, posterior PDFs narrow greatly as they converge with near certainty to parameters of the source distribution.

3.9. Sufficient Statistics

Above, we have described a single “epoch” of Bayesian updating. In a subsequent epoch when additional new data is available, the previous epoch’s posterior becomes the new prior. Sufficient statistics to carry forward to define this prior are either 1) all previous data (including the previous prior and new data) or 2) the posterior quantile parameters $q_0 \leftarrow q_n$ and total number of data $n_0 \leftarrow n_0 + n$, the combination of which is sufficient to form a new encoded data prior. Option 1) has the advantage that the same final posterior is achieved regardless of how many prior epochs have been applied to a

given data set. Option 2) yields *approximately* the same final posterior but has the advantage of not having to preserve all previous data. Option 2) also provides the opportunity to adjust process parameters. For example, if the random process has changed such that a prior datum and a new datum are no longer equally relevant, one may choose an n_0 for a new epoch that does not equal the previous $n_0 + n$. Or one may choose to change the number of terms k if more or less shape detail is indicated.

3.10. Extension to QPDs

The Bayesian updating methods in this section apply not only to metalog distributions but also more broadly to any QPD that satisfies Keelin and Powley's 2011 definition.

4. Discussion

We have shown two major results. First, metalog distributions have virtually unlimited shape flexibility. Since they are also easy to fit to data with ordinary least squares, they may be immediately and practically useful for any application where distribution parameters are derived from data. Examples include gaining insight about the shape characteristics of empirical data, eliciting expert opinion, and combining expert opinion in closed form.

Second, when new data becomes available, metalog distributions can be conveniently and practically updated according to Bayes' rule. This result is based on four assumptions: 1) the target probability distribution to be updated is a metalog distribution (or more broadly a QPD); 2) coefficients are the mean of a multivariate normal (student t if σ is uncertain) distribution, which models their uncertainty for updating purposes; 3) measurement/model error is Gaussian as commonly assumed for linear regression; 4) relevant prior information can be expressed in terms of data. Under these assumptions, updating consistent with Bayes theorem is simply a linear regression on the combined set of prior and new data.

To apply either result, requires choosing which members of the metalog family to use, including boundedness and number of terms k , as well as the fitting of metalog distributions to data. Guidelines for metalog distribution selection and fitting to data are provided in Appendix F.

Appendix A.

Equations of the Metalog Distributions and the Bayesian Updating of their Parameters

A	Metalog quantile function	$M(y \mathbf{a}) = \mathbf{g}(y)\mathbf{a}$ where row vector $\mathbf{g}(y) = (g_1(y), \dots, g_k(y))$; $g_j(y)$ is given by $g_1(y) = 1$, $g_2(y) = \ln\left(\frac{y}{1-y}\right)$, $g_3(y) = (y - 0.5)\ln\left(\frac{y}{1-y}\right)$, $g_4(y) = y - 0.5$, $g_j(y) = (y - 0.5)^{\frac{j-1}{2}}$ for odd $j \geq 5$, and $g_j(y) = \ln\left(\frac{y}{1-y}\right)(y - 0.5)^{\frac{j-1}{2}}$ for even $j \geq 6$; y is cumulative probability, $0 < y < 1$; and $\mathbf{a}$ is a column vector of coefficients.
B	Metalog PDF	$m(y \mathbf{a}) = (\mathbf{g}'(y)\mathbf{a})^{-1}$ where $\mathbf{g}'(y)$ is the derivative of $\mathbf{g}(y)$ with respect to y .
C	Feasibility	$m(y \mathbf{a}) > 0$ for all $y \in (0,1)$.

Table A.1. Metalog equations. **A.** Metalog quantile function **B.** Metalog PDF. See Keelin (2016 for explicit basis-function derivatives. **C.** Metalogs are feasible over a wide range, but feasibility is not guaranteed for metalog-regression β corresponding to an arbitrary data set $(\mathbf{x}, \mathbf{y})$. Generally, feasibility must be checked rather than assumed. Fewer terms, more data, and smoother data help ensure feasibility.

A	distribution	metalog (unbounded metalog)	semi-bounded-low metalog (log metalog)	semi-bounded-high metalog (negative-log metalog)	bounded metalog (logit metalog)
B	transformation	$z(x) = x$	$z = \ln(x - b_l)$	$z = -\ln(b_u - x)$	$z = \ln\left(\frac{x - b_l}{b_u - x}\right)$
C	target quantile function	Table A.1.A	$M_{log}(y \mathbf{a}, b_l) = b_l + e^{M(y)}$ for $0 < y < 1$, $M_{log}(y; \mathbf{a}, b_l) = b_l$ for $y = 0$.	$M_{nlog}(y \mathbf{a}, b_u) = b_u - e^{-M(y)}$ for $0 < y < 1$, $M_{nlog}(y; \mathbf{a}, b_u) = b_u$ for $y = 1$.	$M_{logit}(y \mathbf{a}, b_l, b_u) = \frac{b_l + b_u e^{M(y)}}{1 + e^{M(y)}}$ for $0 < y < 1$, $M_{logit}(y \mathbf{a}, b_l, b_u) = b_l$ for $y = 0$, $M_{logit}(y \mathbf{a}, b_l, b_u) = b_u$ for $y = 1$.
D	target probability density function (PDF)	Table A.1.B	$m_{log}(y) = m(y) e^{-M(y)}$ for $0 < y < 1$, $m_{log}(y) = 0$ for $y = 1$.	$m_{nlog}(y) = m(y) e^{M(y)}$ for $0 < y < 1$, $m_{nlog}(y) = 0$ for $y = 1$.	$m_{logit}(y) = m(y) \frac{(1 + e^{M(y)})^2}{(b_u - b_l)e^{M(y)}}$ for $0 < y < 1$, $m_{logit}(y) = 0$ for $y = 0$ or $y = 1$.
E	bounds	unbounded except when $\beta_j = 0$ for all $j \in \{2, 3, \text{and even } j \geq 6\}$	lower bound b_l	upper bound b_u	lower bound b_l , upper bound b_u
F	shape parameters	$k - 2$	$k - 1$	$k - 1$	k
G	metalog regression	$\mathbf{a} = [\mathbf{Y}'\mathbf{Y}]^{-1}\mathbf{Y}'\mathbf{x}$	$\mathbf{a} = [\mathbf{Y}'\mathbf{Y}]^{-1}\mathbf{Y}'\mathbf{z}$ where $z = \ln(\mathbf{x} - b_l)$	$\mathbf{a} = [\mathbf{Y}'\mathbf{Y}]^{-1}\mathbf{Y}'\mathbf{z}$ where $z = -\ln(b_u - \mathbf{x})$	$\mathbf{a} = [\mathbf{Y}'\mathbf{Y}]^{-1}\mathbf{Y}'\mathbf{z}$ where $z = \ln\left(\frac{\mathbf{x} - b_l}{b_u - \mathbf{x}}\right)$
H	conversion to quantile parameters	$\mathbf{q} = \mathbf{Y}_r \mathbf{a}$	$\mathbf{q} = b_l + e^{\mathbf{Y}_r \mathbf{a}}$	$\mathbf{q} = b_u - e^{-\mathbf{Y}_r \mathbf{a}}$	$\mathbf{q} = \frac{b_l + b_u e^{\mathbf{Y}_r \mathbf{a}}}{1 + e^{\mathbf{Y}_r \mathbf{a}}}$
I	conversion to coefficients	$\mathbf{a} = \mathbf{Y}_r^{-1} \mathbf{q}$	$\mathbf{a} = \mathbf{Y}_r^{-1} \ln(\mathbf{q} - b_l)$	$\mathbf{a} = -\mathbf{Y}_r^{-1} \ln(b_u - \mathbf{q})$	$\mathbf{a} = \mathbf{Y}_r^{-1} \ln\left(\frac{\mathbf{q} - b_l}{b_u - \mathbf{q}}\right)$
J	quantile function over q_i for deterministic σ	$q_i \sim \mathcal{N}^{-1}$ where normal quantile function $\mathcal{N}^{-1} = \mathcal{N}^{-1}(y \bar{q}_{z_i}, \sigma^2 Q_{i,i})$, $\bar{q}_{z_i}$ is i th element of $\mathbf{Y}_r \bar{\mathbf{a}}$, and $Q_{i,i}$ is i th diagonal element of $\mathbf{Q} = \mathbf{Y}_r \mathbf{A}^{-1} \mathbf{Y}_r'$	$q_i \sim b_l + e^{\mathcal{N}^{-1}}$	$q_i \sim b_l + e^{-\mathcal{N}^{-1}}$	$q_i \sim \frac{b_l + b_u e^{\mathcal{N}^{-1}}}{1 + e^{\mathcal{N}^{-1}}}$
K	quantile function over q_i for probabilistic σ	$\mathbf{q} \sim t_v^{-1}$ where student t quantile function $t_v^{-1} = t_v^{-1}(y \bar{q}_{z_i}, \frac{b}{a} Q_{i,i})$, $v = 2a$, and $\bar{q}_{z_i}$ and $Q_{i,i}$ are as defined above	$q_i \sim b_l + e^{t_v^{-1}}$	$q_i \sim b_l + e^{-t_v^{-1}}$	$q_i \sim \frac{b_l + b_u e^{t_v^{-1}}}{1 + e^{t_v^{-1}}}$

Table A.2. Unbounded, semi-bounded and bounded metalog distributions and Bayesian updating of their parameters

Here we provide the metalog equations for target variables that are naturally unbounded, semi-bounded or bounded. For example, Norman's target variable of trout length could be modeled with a semi-bounded metalog with lower bound zero. The target-variable equations follow (I) and honor such bounds. **A.** Name of distribution. **B.** Transformation z of target variable x that is metalog distributed. b_l and b_u are lower and upper bounds, respectively. **C.** Target-variable quantile function **D.** Target-variable PDF **E.** Target-variable bounds. **F.** Number of target distribution shape parameters. **G.** Metalog regression equation. **H.** Conversion from coefficients to quantile parameters. **I.** Conversion from quantile parameters to coefficients. **J.** Quantile function over the i th quantile parameter q_i given deterministic σ . This is the normal quantile function for the unbounded metalog and transformations of this for other metalogs. **K.** Quantile function over q_i given probabilistic σ . This is the student-t quantile function for unbounded metalog and transformations of this for other metalogs. These transformations derive from the property that if $z(x) \sim T(y)$ where T is a quantile function and z is invertible and non-decreasing, then x is distributed according to the transformed quantile function (Gilchrist 2000): $x \sim z^{-1}(T(y))$.

We note that the well-known aphorism, “All models are wrong, but some are useful,” applies to our assumption of a multivariate normal (multivariate student t if σ is unknown) distribution over quantile parameters $\mathbf{q}$. This assumption is useful due to its conjugacy properties, but inherently contains small but non-zero probabilities of inconsistent $\mathbf{q}$, such as the median being less than the 0.1 quantile. For updating purposes, one must accept that such non-zero probabilities are immaterial.

A	Conditional conjugate prior over parameters	$p(\mathbf{a} \sigma) = \mathcal{N}(\bar{\mathbf{a}}, \sigma^2 \mathbf{A}^{-1})$
B	Conjugate prior over σ^2	$p(\sigma^2) = \Gamma^{-1}(\alpha, \beta)$
C	Conjugate prior over parameters	$p(\mathbf{a}, \sigma^2) = \mathcal{N} \cdot \Gamma^{-1}(\bar{\mathbf{a}}_0, \mathbf{A}_0, \alpha_0, \beta_0)$
D	Posterior over parameters	$p(\mathbf{a}, \sigma^2 \mathbf{x}, \mathbf{Y}) = \mathcal{N} \cdot \Gamma^{-1}(\bar{\mathbf{a}}_n, \mathbf{A}_n, \alpha_n, \beta_n)$
E	Bayesian updating	$\mathbf{A}_n = \mathbf{Y}'\mathbf{Y} + \mathbf{A}_0$ $\bar{\mathbf{a}}_n = \mathbf{A}_n^{-1}(\mathbf{A}_0\bar{\mathbf{a}}_0 + \mathbf{Y}'\mathbf{x})$ $a_n = a_0 + n/2$ $b_n = b_0 + (\mathbf{x}'\mathbf{x} + \bar{\mathbf{a}}_0'\mathbf{A}_0\bar{\mathbf{a}}_0 - \bar{\mathbf{a}}_n'\mathbf{A}_n\bar{\mathbf{a}}_n)/2$
F	Marginal distribution over parameters with uncertain σ	$p(\mathbf{a}) = t_\nu(\bar{\mathbf{a}}, \frac{b}{a}\mathbf{A}^{-1})$ where $\nu = 2a$
G	Specifying $\mathbf{A}_0$, a_0 and b_0 given data prior $(\mathbf{x}_{n_0}, \mathbf{y}_{n_0})$	$\mathbf{A}_0 = \mathbf{Y}'_{n_0}\mathbf{Y}_{n_0}$ $a_0 = (n_0 - k)/2$ Given an empirical data prior, $b_0 = (\mathbf{x}'_{n_0}\mathbf{x}_{n_0} - \bar{\mathbf{a}}_0'\mathbf{A}_0\bar{\mathbf{a}}_0)/2$, which is half the sum of squared residuals between this data and target-prior metalog. Figure A.1 provides a supplementary interpretation. Given an encoded data prior, b_0 may be estimated as explained in the legend.

Table A.3. Updating equations given probabilistic σ

A. Conditional on σ , the prior over $\mathbf{a}$ remains multivariate normal. **B.** Conjugate prior over σ^2 is an inverse-gamma distribution (Wikipedia accessed 25.02.2020) with parameters α and β . **C.** Implied joint prior over $\mathbf{a}$ is multivariate normal-inverse-gamma distribution (Wikipedia accessed 25.02.2020) with four parameters. **D.** Posterior over parameters is of the same functional form as the prior. **E.** Closed-form Bayesian updating of conjugate mean $\bar{\mathbf{a}}$ is the same as for deterministic σ . Additional equations update a and b . **F.** By integrating out σ uncertainty, marginal distribution over $\mathbf{a}$ is multivariate student-t. **G.** Formulas for specifying initial parameter values.

Fig. A.1. Interpreting σ or b in terms of marginal distribution width w . To apply the updating equations in the case of deterministic σ , its value must be assessed or otherwise determined.

Alternatively, for probabilistic σ , the conjugate prior over σ is an inverse-gamma-distribution of which parameter b_0 in Table A.3 must be determined. Given a data prior and prior values for the quantile parameters, there is a one-to-one relationship between σ or b and the width w of the marginal priors over the quantile parameters. Let w be the width of the marginal prior over q_i , the i th element of $\mathbf{q}$; α be a probability selected for ease of interpretation such that $0 < \alpha < 0.5$, and $Q_{i,i}$ be the i th diagonal element of $\mathbf{Q}$, where $\mathbf{Q} = \mathbf{Y}_r \mathbf{A}^{-1} \mathbf{Y}_r'$. For deterministic σ , $\sigma = -\frac{w}{2\Phi^{-1}(\alpha)\sqrt{Q_{i,i}}}$, where $\Phi^{-1}(\alpha)$ is the

quantile function of the standard normal distribution evaluated at α . For probabilistic σ , $b =$

$\frac{a}{4Q_{i,i}} \left(\frac{w}{t_v^{-1}(\alpha)} \right)^2$ where $a = \frac{n-k}{2}$ for total number of data n and $t_v^{-1}(\alpha)$ is the quantile function of the

standard univariate student t distribution with $\nu = 2a$ degrees of freedom evaluated at α . If, for example, $\alpha = 0.1$ and q_i is the median, then w is the 0.1-0.9 width of the marginal distribution over the median. These equations follow from observing that the multivariate-normal covariance matrix over the quantile parameters is $\sigma^2 \mathbf{Q}$. Therefore, the marginal distribution over q_i is normal with variance $\sigma^2 Q_{i,i}$. While the above equations for σ and b are for the unbounded metalog, a similar one-to-one relationship holds for semi-bounded and bounded metalogs. Based on the transforms in Table A.2.B, if σ is deterministic, the marginal distributions over q_i , as shown in Table A.2.J, are lognormal, negative lognormal, and logit-normal (Mead 1965) respectively. Similarly, if σ is probabilistic, the marginal distributions over q_i , as detailed in Table A.2.K, are log-student t, negative log-student t, and logit-student t respectively, where the logit-student t may be defined by substituting a student-t distribution for the normal distribution in Mead (1965). Thus, in all cases, there is a one-to-one relationship between σ or b and w , and the marginal distributions over the quantile parameters honor the natural bounds, if any, of the target variable.

Appendix B.
End of Quarter Returns on Investing in Gold at Start of Quarter Based on USD Spot Exchange Rate. (Parentheses denote negative number)

Date	Quarterly return on investment in XAU (gold)	Date	Quarterly return on investment in XAU (gold)	Date	Quarterly return on investment in XAU (gold)
9/30/2019	0.0447	6/29/2012	(0.0424)	3/31/2005	(0.0230)
6/28/2019	0.0907	3/30/2012	0.0660	12/31/2004	0.0483
3/29/2019	0.0077	12/30/2011	(0.0363)	9/30/2004	0.0609
12/31/2018	0.0769	9/30/2011	0.0824	6/30/2004	(0.0755)
9/28/2018	(0.0493)	6/30/2011	0.0475	3/31/2004	0.0265
6/29/2018	(0.0550)	3/31/2011	0.0076	12/31/2003	0.0781
3/30/2018	0.0174	12/31/2010	0.0862	9/30/2003	0.1124
12/29/2017	0.0180	9/30/2010	0.0533	6/30/2003	0.0265
9/29/2017	0.0307	6/30/2010	0.1160	3/31/2003	(0.0305)
6/30/2017	(0.0061)	3/31/2010	0.0148	12/31/2002	0.0757
3/31/2017	0.0886	12/31/2009	0.0886	9/30/2002	0.0289
12/30/2016	(0.1280)	9/30/2009	0.0876	6/28/2002	0.0390
9/30/2016	(0.0046)	6/30/2009	0.0078	3/29/2002	0.0850
6/30/2016	0.0723	3/31/2009	0.0423	12/31/2001	(0.0488)
3/31/2016	0.1618	12/31/2008	0.0127	9/28/2001	0.0827
12/31/2015	(0.0484)	9/30/2008	(0.0588)	6/29/2001	0.0500
9/30/2015	(0.0488)	6/30/2008	0.0093	3/30/2001	(0.0525)
6/30/2015	(0.0095)	3/31/2008	0.0998	12/29/2000	(0.0073)
3/31/2015	(0.0007)	12/31/2007	0.1212	9/29/2000	(0.0528)
12/31/2014	(0.0197)	9/28/2007	0.1446	6/30/2000	0.0374
9/30/2014	(0.0898)	6/29/2007	(0.0213)	3/31/2000	(0.0310)
6/30/2014	0.0337	3/30/2007	0.0426	12/31/1999	(0.0360)
3/31/2014	0.0685	12/29/2006	0.0642	9/30/1999	0.1381
12/31/2013	(0.0959)	9/29/2006	(0.0285)	6/30/1999	(0.0627)
9/30/2013	0.0765	6/30/2006	0.0552	3/31/1999	(0.0284)
6/28/2013	(0.2272)	3/31/2006	0.1289	12/31/1998	(0.0293)
3/29/2013	(0.0465)	12/30/2005	0.1016	9/30/1998	0.0000
12/31/2012	(0.0547)	9/30/2005	0.0776	6/30/1998	(0.0133)
9/28/2012	0.1094	6/30/2005	0.0167	3/31/1998	0.0412

Appendix C.

Using Metalogs For Highly-Skewed, Heavy-Tailed Empirical Data Sets

This appendix explores the metalog's ability to fit extremely skewed distributions and empirical data sets. It has long been observed that a wide range of social, quality-control, scientific, geophysical, actuarial, and many other types of observable phenomena follow power law distributions, a.k.a. Pareto distributions (Wikipedia accessed 16.11.2020, Newman 2005, Clauset, Shalizi, and Newman 2009). One such distribution formalizes the "80-20 rule," which, over a wide range of phenomena, suggests that 80% of the outcomes (or outputs) result from 20% of the causes (or inputs). In their widely-cited article, Clauset, Shalizi, and Newman (2009) consider 24 heavy-tailed interdisciplinary data sets and the extent to which power-law distributions and variations thereof may be considered good statistical models. Following Clauset et. al.'s notation, the power law PDF, CDF, and complementary CDF (CCDF) are given by $\left(\frac{a-1}{x_{min}}\right)\left(\frac{x}{x_{min}}\right)^{-a}$, $1 - \left(\frac{x}{x_{min}}\right)^{1-a}$, and $\left(\frac{x}{x_{min}}\right)^{1-a}$ respectively, where x_{min} is a lower-bound and scale parameter and where shape parameter α , with occasional exceptions, typically lies in the range $2 < \alpha < 3$ (Clauset, Shalizi, and Newman 2009, p. 2).

Here we explore two questions. Are the metalog distributions, which were unknown at the time of the Clauset article, sufficiently flexible to mimic the heavy-tailed power law distributions over their typical range? Are the metalog distributions worthy of consideration alongside power law distributions as candidate models for data that may exhibit power-law type behavior?

We answer the first question affirmatively. Table C.1 shows the exact quantiles for power law distributions over the wider range of α from 1.5 to 3.5 in 0.25 increments. These exact values correspond to the orange dots in the power law CCDF's in the left panel of Fig C.1. These power law CCDF's are exactly linear with slope $-\alpha$ when plotted in terms of the logs of variable of interest x and cumulative probability y on the horizontal and vertical axes respectively. Thus, the exact power law distribution (blue line) goes through each point exactly. If we parameterize a nine-term log metalog distribution (orange dashed curve) with these nine points, the log metalog also goes through each point exactly. The KS distance between the exact power law and the log metalog is < 0.001 in all cases illustrated.

power law parameter α (xMin=1)									
	1.5	1.75	2.	2.25	2.5	2.75	3.	3.25	3.5
0.001	1.002	1.0013	1.001	1.0008	1.0007	1.0006	1.0005	1.0004	1.0004
0.02	1.0412	1.0273	1.0204	1.0163	1.0136	1.0116	1.0102	1.009	1.0081
0.1	1.2346	1.1508	1.1111	1.0879	1.0728	1.0621	1.0541	1.0479	1.043
0.25	1.7778	1.4675	1.3333	1.2588	1.2114	1.1787	1.1547	1.1364	1.122
0.5	4.	2.5198	2.	1.7411	1.5874	1.486	1.4142	1.3608	1.3195
0.75	16.	6.3496	4.	3.0314	2.5198	2.2082	2.	1.8517	1.7411
0.9	100.	21.5443	10.	6.3096	4.6416	3.7276	3.1623	2.7826	2.5119
0.98	2500.	184.202	50.	22.8653	13.5721	9.3506	7.0711	5.6898	4.7818
0.999	$1. \times 10^6$	10000.	1000.	251.189	100.	51.7947	31.6228	21.5443	15.8489

(9-term log metalog CDF goes exactly though each point)

Table C.1. Power law quantiles for a range of shape parameters α . For nine selected cumulative probabilities y , corresponding power law quantiles are shown across a typical range of power law shape parameters α .

The power law PDFs, blue lines in the right panel of Fig. C.1, are also exactly linear in the logs of x and probability density y' . The exact power law PDFs are visually indistinguishable from the log metalog PDFs (orange dashed curves) parameterized by the Table 1 quantiles.

From this analysis, we conclude that if a variable is exactly power law distributed over the range of $1.5 < \alpha < 3.5$, then, for practical purposes, a metalog distribution can represent its uncertainty about equally as well as the power law distribution itself.

Fig C.1. Power law distributions and metalog representations thereof. This figure shows power law CCDFs (left panel) and PDF's (right panel) for a range of shape parameters α from 1.5-3.5 in 0.25 increments. These correspond to labels A-I respectively. Blue lines are the exact power law distributions. Orange dashed lines are the 9-term log metalog distributions parameterized by the power law quantiles in Table C.1

For illustration with empirical data, we consider six data sets analyzed by Clauset, Shalizi, and Newman (2009) for which the data is conveniently available (<http://tuvalu.santafe.edu/~aaronc/powerlaws/data.htm> accessed 16.11.2020):

- A. Words: the frequency of occurrence of unique words in the novel Moby Dick by Herman Melville
- B. Terrorism: the severity of terrorist attacks worldwide from February 1968 to June 2006, measured as the number of deaths directly resulting (Clauset, Young, and Gleditsch, 2007).
- C. Blackouts: the numbers of customers affected in electrical blackouts in the United States between 1984 and 2002.
- D. Cities: the human population of US cities in the 2000 US census.
- E. Flares: peak gamma ray intensity of solar flares between 1980 and 1989.
- F. Surnames: the frequencies of occurrence of U.S. family surnames in the 1990 US census.

Descriptors of these data sets are shown in Table C.2. In all cases, the data is heavily right-skewed, and the maximum value is ten or more standard deviations above the mean. Should metalogs be considered alongside power law distributions as candidate models for such data sets?

Such data sets rarely follow a power law over their entire range, but rather over a tail range of $x \geq x_{min}$, where the x_{min} values in Table C.2 are from Clauset et. al.'s (2009) detailed analysis; n_{tail} is the number of data that meet this condition; and $\hat{\alpha}$ is the corresponding maximum likelihood estimate, $\hat{\alpha} = 1 + n_{tail} \left[\sum_{i=n-n_{tail}+1}^n \ln \left(\frac{x_i}{x_{min}} \right) \right]^{-1}$, of power-law shape parameter α .

Fig C.2 shows the complementary cumulative CCDF distribution of the tail data along with the power-law (blue) and metalog (orange) fits. It is visually apparent that the metalog fits the data as well or better than the power law in most or all cases. This is no surprise because the metalog has more shape parameters. Nonetheless, at first glance, the metalog appears to be a credible alternative.

	n	mean	std dev	x_{min}	x_{max}	$\hat{\alpha}$	n_{tail}
words	18 855	11.14	148.33	7.	14086.00	2.02	2958
terrorism	9101	4.35	31.58	12.	2749.00	2.45	547
blackouts	211	253868.68	610308.58	230 000.	7.50×10^6	2.27	59
cities	19 447	9002.05	77825.05	52 457.	8.01×10^6	2.37	580
flares	12 773	689.41	6520.59	323.	231300.00	1.79	1711
surnames	2753	50591.20	113987.12	111 919.	2.50×10^6	2.49	239

Table C.2. Descriptors of some heavy-tailed data sets. For the heavy-tailed data sets A-F described in the text, this table shows the total number of data n , their mean and standard deviation, and their maximum x_{max} . x_{min} is the value determined by Clauset et. al. such that the conditional data set where $x \geq x_{min}$ is a reasonable candidate for study of power law-type behavior. n_{tail} is the number of data that meet this condition. $\hat{\alpha}$ is the corresponding maximum likelihood estimate of power-law shape parameter α .

Fig C.2. Power law and metalog fits to heavy-tailed data sets. Complementary cumulative (CCDF) distributions for data sets A-F described in the text and Table 2, plotted with $\log_{10}(x)$ and $\log_{10}(y)$ where y is cumulative probability, on horizontal and vertical axes respectively. Data (gray circles), power law fit (blue line), and metalog fit (orange curve) are shown for each data set.

This observation is further borne out by the fit statistics in Table C.4. As expected, since it has more shape parameters, the metalog in each case has a smaller KS distance and a higher log likelihood. For words and terrorism, the Bayes factors (difference between base-10 log likelihoods) are 97 and 28 orders of magnitude of likelihood, respectively, in favor of the metalog. Accounting for the additional complexity (shape parameters) of the metalog, the AIC (Akaike information criterion) favors the metalog in five of six cases. The Bayesian information criterion (BIC), which more heavily penalizes model complexity, favors the metalog in two of six cases.

Table C.4 provides the specifications for the metalogs in Fig. C.2 and Table C.3. The $\mathbf{a}$ coefficients are determined simply by linear regression. In most cases, similar results can be achieved by using a different number of metalog terms. In Appendix F, we discuss how we chose the number of terms, ways to avoid potential over- and under-fitting, and freely available tools to aid such investigations.

	KS _{metalog}	KS _{power law}	LLH _{metalog}	LLH _{power law}	Bayes factor	AIC _{min}	BIC _{min}
words	0.092	0.145	-5118.20	-5215.60	97.35	metalog	metalog
terrorism	0.083	0.117	-905.31	-932.82	27.51	metalog	metalog
blackouts	0.041	0.078	-355.94	-357.52	1.58	metalog	power law
cities	0.016	0.019	-3094.40	-3094.60	0.21	power law	power law
flares	0.008	0.010	-6151.80	-6157.30	5.47	metalog	power law
surnames	0.040	0.065	-1340.00	-1341.90	1.81	metalog	power law

Table C.3. Metalog vs. power law fit statistics. Goodness of fit statistics for data sets A-F described in the text and Table 1.3 for the metalog and power law fits shown in Fig C.2. KS (Kolmogorov-Smirnov) is the maximum distance between the curve and the data, which is less for the metalog in all cases. LLH is $\log_{10}(\text{likelihood of the data})$, which is greater for the metalog in all cases. Bayes factor, the difference in LLH, is the number of orders of magnitude by which the metalog is a more likely explanation of the data than the power law. AIC (Akaike information criterion) is a measure of information lost by the model, according to which the metalog is preferred in five of six cases. BIC (Bayesian information criterion) is an alternate measure of information loss, according to which the metalog is preferred in two of six cases.

	words	terrorism	blackouts	cities	flares	surnames
b_l	6.	11.	219 000.	52 360.	322.	109 432.
n_{terms}	4	5	8	6	10	4
a_1	1.91	2.14	12.14	10.46	6.13	11.18
a_2	0.55	0.43	11.90	0.71	-10.82	0.57
a_3	0.54	0.79	4.02	-0.21	-3.01	-0.08
a_4	1.88	2.08	-44.28	0.80	47.48	1.49
a_5	0.00	-4.49	-14.88	0.29	12.33	0.00
a_6	0.00	0.00	-37.74	0.49	55.25	0.00
a_7	0.00	0.00	106.46	0.00	-152.45	0.00
a_8	0.00	0.00	-8.89	0.00	10.37	0.00
a_9	0.00	0.00	0.00	0.00	-24.14	0.00
a_{10}	0.00	0.00	0.00	0.00	-44.16	0.00

Table C.4. Metalog specifications. The metalog curves in Fig C.2 are log metalog distributions with the number of terms, lower bound b_l , and $\mathbf{a}$ coefficients shown here. For the lower bound, we could not use $b_l = x_{\min}$ from Table C.2 because, for the log metalog, all data must be strictly $>$ (not $\geq$) b_l . So, avoiding more elaborate schemes for this illustration, we simply used $b_l = x_{n-n_{\text{tail}}}$, which is the largest x value not included in the tail.

Here we have made the case that metalog distributions might reasonably be considered as candidate models for data sets that might otherwise be modelled with a power law. More broadly, in light of their virtually unlimited shape and bounds flexibility, metalog models might similarly well serve as good candidates for diverse data sets in virtually any field of human endeavor.

Appendix D.

Covariance Magnitude Inversely Proportional to Number of Data

n	$\mathbf{A}$
3	$\begin{pmatrix} 3.0 & 0.0 & 0.5 \\ 0.0 & 2.4 & 0.0 \\ 0.5 & 0.0 & 0.2 \end{pmatrix}$
4	$\begin{pmatrix} 4.0 & 0.0 & 0.9 \\ 0.0 & 4.2 & 0.0 \\ 0.9 & 0.0 & 0.3 \end{pmatrix}$
5	$\begin{pmatrix} 5.0 & 0.0 & 1.3 \\ 0.0 & 6.1 & 0.0 \\ 1.3 & 0.0 & 0.6 \end{pmatrix}$
6	$\begin{pmatrix} 6.0 & 0.0 & 1.7 \\ 0.0 & 8.3 & 0.0 \\ 1.7 & 0.0 & 0.9 \end{pmatrix}$
10	$\begin{pmatrix} 10.0 & 0.0 & 3.5 \\ 0.0 & 17.7 & 0.0 \\ 3.5 & 0.0 & 2.3 \end{pmatrix}$
50	$\begin{pmatrix} 50.0 & 0.0 & 22.6 \\ 0.0 & 133.9 & 0.0 \\ 22.6 & 0.0 & 22.6 \end{pmatrix}$
100	$\begin{pmatrix} 100.0 & 0.0 & 47.3 \\ 0.0 & 290.0 & 0.0 \\ 47.3 & 0.0 & 50.9 \end{pmatrix}$

Table D.1. Covariance magnitude.

This table shows why the variances and covariances of the distribution over quantile parameters decline approximately in proportion n , the total number of prior and new data. While the illustration above is for $k = 3$ terms, the same result holds for any number of terms. Assuming a data prior, $k \times k$ matrix $\mathbf{A} = \mathbf{Y}'\mathbf{Y}$. The first column of $n \times k$ matrix $\mathbf{Y}$ consists of ones because the first metalog basis function $g_1(y) = 1$ per Table A.1. So, the (1,1) element of $\mathbf{A}$ always equals n . It is easily to confirm that matrix $\mathbf{C} = \mathbf{A}/n$ approaches a constant as n increases. In the case of deterministic σ per Table A.3.J, the implied multivariate normal covariance matrix over quantile parameters is $\sigma^2 \mathbf{Y}_r \mathbf{A}^{-1} \mathbf{Y}_r' = \sigma^2 \mathbf{Y}_r (n\mathbf{C})^{-1} \mathbf{Y}_r' = \frac{\sigma^2 \mathbf{Y}_r \mathbf{C}^{-1} \mathbf{Y}_r'}{n}$ where, except for n , all quantities are constant. So, the magnitude of the covariance matrix is inversely proportional to n . Similarly, for probabilistic σ , one can show that b/a in Table A.2.K approaches a constant as n increases. So, the same result holds for all eight cases in Table A.2.J-K.

Appendix E. Bayesian Updating Using Nine-Term Metalog for Increased Shape Flexibility

Fig. E.1. Updating quantile-parameter distributions as new data are revealed for a bi-model source distribution and nine-term metalog. Use of a nine-term metalog captures more nuances of shape flexibility. To illustrate with Bayesian inference, we consider a source distribution over the weight of migrating steelhead trout in British Columbia's Skeena River system based 3,474 empirical data gathered over 2010-2014. This source distribution is bimodal (Keelin 2016) likely because the fish population consists largely and discretely of first time and second time spawners, the latter of which are usually heavier. We use $\mathbf{q} = (4, 4.43, 5.13, 7.24, 9.83, 12.06, 15.28, 18.15, 21)$ lbs corresponding to $\mathbf{y}_r = (.02, .05, .10, .25, .5, .75, .90, .95, .98)$ to parameterize this bimodal nine-term metalog source distribution, the dashed-red curve in **A**. A less-informed observer might naively presume a unimodal prior distribution, $\mathbf{q} = (4.85, 5.58, 6.32, 7.79, 9.83, 12.4, 15.28, 17.31, 19.93)$ lbs, the dashed-orange curve in **A**, and adopt an encoded data prior with strength of prior $n_0 = 100$. This less-informed observer learns of the bi-modality by observing simulated samples from the source distribution and updating his prior with metalog regression. With 100 new data, the target posterior in **A** begins to take on the shape of the source. The marginal posterior distributions over quantile parameters, blue curves in **B**, have moved from their priors (orange-dashed curves in **B**) toward the source $\mathbf{q}$ parameters (red-dashed lines in **B**) and narrowed based on the additional data. Moreover, with 500 new data, posterior quantile parameters in **C** have nearly converged to the source. Accordingly, so also has the corresponding target posterior in **A**.

Fig. E.2. Quantile parameter convergence rates to the source. Based on simulation, prior quantile parameters converge to those of the source with increasing numbers of new data. **A.** Three-term metalog source distribution corresponding to Fig. 3.5. **B.** Nine-term metalog source distribution corresponding to Fig. E.1.

Appendix F. Guidelines for Metalog Distribution Selection and Fitting to Data

Selecting the number of metalog terms k for the metalog distribution requires judgment. In practice, $2 \leq k \leq 16$ is typical. Smaller numbers of terms are more likely to be feasible for a wide range of data sets, while larger numbers of terms provide more nuanced shape flexibility. We usually start with a smaller number of terms (2 to 5) and increase if more shape flexibility appears to be helpful or insightful. The metalog panel (I and Fig. S1) is useful in this regard. Metalog regression requires that $k \leq n$, where n is the total number of data. To avoid overfitting, k should generally be less than $0.1n$ to $0.2n$.

F.1 Fitting Metalog Distributions to Data

Given the k -term metalog quantile function $x = M(y|\mathbf{a})$ and CDF data where $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ where $n \geq k$, the coefficients are uniquely determined by metalog regression: $\mathbf{a} = [\mathbf{Y}'\mathbf{Y}]^{-1}\mathbf{Y}'\mathbf{x}$, where $\mathbf{Y}$ is the $n \times k$ matrix whose (i, j) element is $g_j(y_i)$ as defined in Table S1 and $[\mathbf{Y}'\mathbf{Y}]^{-1}$ is invertible. Assigning $y_i = i/(n + 1)$ to the i th largest sample x_i is consistent with y_i being the mean of the i th largest of independent samples from a uniform distribution. If one assumes that the x_i 's are randomly drawn, this probability assignment is natural. Other probability assignments such as $y_i = (i - 0.5)/n$ could be used instead (Q-Q Plot, Wikipedia accessed: 25.02.2020). For semi-bounded and bounded metalog distributions, where the log and logit transforms, respectively, of x are metalog distributed, the implied quantile functions and regression equations are given in Table A.2.C and A.2.G respectively.

For the various metalog distributions, feasibility may be checked visually by plotting the PDF $m(y)$ as shown in Table A.2.D and ensuring that this is positive everywhere over the domain of x . Typically, one makes this PDF plot by varying $y \in (0,1)$ parametrically and plotting $m(y)$ and $x = M(y|\mathbf{a})$ on the vertical and horizontal axes respectively. To automate the feasibility check, discretize $y \in (0,1)$ into a large number of points and test for $m(y) > 0$ at each point. An alternate method, implemented as a linear program, determines the $\mathbf{a}$ coefficients by minimizing the sum of absolute distances between the CDF and the data subject to feasibility constraint (Faber 2019).

F.2. Selecting Metalog Boundedness and Number of Terms k

Though higher numbers of terms may be occasionally useful, most metalog applications use $k = 2$ to $k = 16$ terms. Given four choices of boundedness (unbounded, semi-bounded low, semi-bounded high, and bounded), there are $4(16 - 2) = 60$ distinct and practical probability distributions within the metalog system. For many data sets, some these 60 distributions are infeasible and can be immediately eliminated.

Among feasible metalogs, one may use the same criterion (or criteria) for distribution selection that one typically uses for other distributions. Such criteria may include boundedness, KS distance, maximum likelihood, Bayes factor, AIC, BIC, and p-values, among others. Whatever criteria one chooses may be equally applied to the set of candidate models that consists of all feasible metalog distributions and all traditional and other distributions worthy of consideration.

When choosing among metalog distributions for a given data set, we typically begin by selecting the boundedness that corresponds to the natural boundedness of the situation. For example, one may choose the unbounded metalog to model measurement errors that could be unlimited in either direction; a semi-bounded metalog for a variable x such that $x \geq 0$ by definition; or a bounded metalog for uncertain fractions x of a population such that $0 \leq x \leq 1$ by definition.

For selecting the number terms k , in addition to other criteria, we often start with a “metallog panel” visual display metalog PDFs for a range of k . For example, Fig. F.1 shows the metalog panel for quarterly returns on investing in gold as parameterized by data in Appendix B and illustrated in Fig. 2.2. Blank cells in this panel are infeasible when fit with metalog regression. Among the feasible cells, we noted a relatively stable bimodal pattern for $k = 6 - 9$. Finding this interesting and potentially worthy of further investigation, we selected $k = 8$ for display in the main article. In contrast, if we hypothetically knew with near certainty (for example based on other research) that quarterly returns on gold are *unimodal*, then we might well have selected the unimodal $k = 5$ for display, which would reflect a belief that the indicated bimodality for $k = 6 - 9$ is due to statistical fluctuation and overfitting thereto. Keelin (2016, Sec. 6.3) provides further considerations for metalog distribution selection. Free metalog tools and software are available tools (<http://www.metalogdistributions.com/software.html> accessed 08.18.2021).

Fig. F.1. Metalog panel for quarterly returns on investing in gold 3/31/1998 – 9/30/2019. Unbounded metalog PDFs for various numbers of terms k parameterized by data in Appendix B. Blank cells in this panel are infeasible when fit with metalog regression.

References

- Balakrishnan, N. (1991) Handbook of the logistic distribution. CRC Press.
- Berger, J.O. (2013) Statistical decision theory and Bayesian analysis. Springer Science & Business Media. Introduction and Chapter 3.2.
- [Bratvold, R.B., Mohus, E., Petutschnig, D. and Bickel, E. \(2020\). "Production forecasting: Optimistic and overconfident—Over and over again." Society of Petroleum Engineers. doi:10.2118/195914-PA.](#)
- Clauset, A., Shalizi, C.R. and Newman, M.E. (2009) Power-law distributions in empirical data. *SIAM review*, 51(4), pp.661-703.
- Clauset, A., Young, M. and Gleditsch, K.S. (2007) On the frequency of severe terrorist events. *Journal of Conflict Resolution*, 51(1), pp.58-87.
- [Dion, P., Galbraith, N., Sirag, E. \(2020\). "Using expert elicitation to build long-term projection assumptions." In Developments in Demographic Forecasting, Chapter 3, pp. 43–62. Springer](#)
- Faber, I.J. (2019). Cyber Risk Management: AI-generated Warnings of Threats (Doctoral dissertation, Stanford University).
- Genest, C. and Zidek, J.V., 1986. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1(1), pp.114-135.
- Genest, C. (1992) Vincentization revisited. *The Annals of Statistics*, pp.1137-1142.
- Gilchrist, W. (2000) Statistical modelling with quantile functions. CRC Press.
- Howard, R.A., 1970. Decision analysis: Perspectives on inference, decision, and experimentation. *Proceedings of the IEEE*, 58(5), pp.632-643.
- Howard, R., & Abbas, A. E. (2016). Foundations of Decision Analysis, global edition. *Harlow, England: Pearson Education Limited*.
- Keelin, T.W. (2016) The metalog distributions. *Decision Analysis*, 13(4), pp.243-277.
- Keelin, T.W. and Powley, B.W. (2011) Quantile-parameterized distributions. *Decision Analysis*, 8(3), pp.206-219.
- Keelin, T.W., Chrisman, L. and Savage, S.L. (2019) December. The Metalog Distributions and Extremely Accurate Sums of Lognormals in Closed Form. In *2019 Winter Simulation Conference (WSC)* (pp. 3074-3085). IEEE
- Lichtendahl Jr, K. C., Grushka-Cockayne, Y., & Winkler, R. L. (2013). Is it better to average probabilities or quantiles?. *Management Science*, 59(7), 1594-1611.
- Maclean, N. (2009) A river runs through it and other stories. University of Chicago Press.
- McGrayne, S.B. (2011) *The theory that would not die: how Bayes' rule cracked the enigma code, hunted down Russian submarines, & emerged triumphant from two centuries of controversy*. Yale University Press.
- Mead, R., 1965. A generalised logit-normal distribution. *Biometrics*, 21(3), pp.721-732.
- Newman, M.E. (2005) Power laws, Pareto distributions and Zipf's law. *Contemporary physics*, 46(5), pp.323-351.

Pearson, K. (1893) Contributions to the Mathematical Theory of Evolution. Proceedings of the Royal Society of London, 54(326-330), pp.329-333.

Pearson, K. (1895) Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. Philosophical Transactions of the Royal Society of London. A, 186, pp.343-414.

Pearson, K. (1901) Mathematical contributions to the theory of evolution. X. Supplement to a memoir on skew variation. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 197, pp.443-459.

Pearson, K. (1916) Mathematical contributions to the theory of evolution. XIX. Second supplement to a memoir on skew variation. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 216, pp.429-457.

Raiffa, H. and Schlaifer, R. (1961) Applied statistical decision theory.

Spetzler, C.S. and Stael von Holstein, C.A.S., 1975. Exceptional paper—Probability encoding in decision analysis. Management science, 22(3), pp.340-358.

Thomas, E.A. and Ross, B.H., 1980. On appropriate procedures for combining probability distributions within the same family. *Journal of Mathematical Psychology*, 21(2), pp.136-152.

Vincent, Stella; Burnham (1912). "The function of the viborissae in the behavior of the white rat". **1**. Behavior Monographs.