Modelplasticity and abductive decision making

Subhadeep Mukhopadhyay[1] (ID)

Abstract

'All models are wrong but some are useful' Box (Robustness in statistics, Elsevier, pp 201–236, 1979). But, how to find those useful ones starting from an imperfect model? How to make informed data-driven decisions equipped with an imperfect model? These fundamental questions appear to be pervasive in virtually all empirical fields—including economics, finance, marketing, healthcare, climate change, defense planning, and operations research. This article presents a modern approach (builds on two core ideas: abductive thinking and density-sharpening principle) and practical guidelines to tackle these issues in a systematic manner.

1 The uncertainty principle

How to make decisions under uncertainty? Decision-making under uncertainty relies mainly on how efficiently we can extract *useful* knowledge from the data that were

---

✉ Subhadeep Mukhopadhyay
  deep@unitedstatalgo.com

1   United Analytics and Computational Intelligence, Inc., California, USA

⦣ Springer

previously *unknown* to the decision-maker[1] C. R. Rao, in his Rao (1996) article[2] on 'Uncertainty, Statistics, and Creation of New Knowledge' provided an exquisite description of the mechanics of decision-making under uncertainty using a simple logical formula:

$$\boxed{\text{Uncertainty of knowledge}} + \boxed{\text{Knowledge of uncertainty}} = \boxed{\text{Usable knowledge.}} \quad (1)$$

A decision analyst confronts data $X_1, \ldots, X_n$ equipped with a tentative (imprecise and uncertain) probabilistic model $f_0(x)$ of the underlying phenomena. The challenge then boils down to effectively using the misspecified model $f_0(x)$ to learn from data and to apply that knowledge for informed decision-making. Rao's uncertainty principle suggests the following three-staged approach, which we call the '*model-building triad*':

`Stage 1`. *Model Elicitation*. The first step of decision making is formulating a probability model of the phenomena of interest, from which economic agents derive their initial expectations. A simple parametrized model-0 $f_0(x)$ is usually formed based on either gut instinct or the scientific context of the investigation. The *uncertainty* of $f_0(x)$ arises due to the lack of perfect knowledge[3] about the underlying probability law. Accordingly, the modeler has to start the analysis by acknowledging the uncertainty of the initial knowledge model $f_0(x)$.

`Stage 2`. *Model Uncertainty Quantification*. Before making decisions based on the provisional model $f_0(x)$, it is crucial to investigate its uncertainty (blind spots) in light of the new data. It's always a good practice to inspect expert opinions based on hard empirical facts by asking.[4]: what's new in the data that can't be explained by the assumed model? Discovering surprising and previously *unknown* facts can prompt decision makers to consider other alternative actions.

`Stage 3`. *Model Rectification and Risk Management*. Finally, we incorporate the learned uncertainty into the uncertain model $f_0(x)$ to produce a rectified model for making empirically guided informed decisions. It is important to *sharpen* the "judgment component" (intuition based on past experiences) in light of the new data before it gets outdated.

The purpose of this article is to describe a general statistical theory that permits us to implement this three-staged model-building procedure for data analysis and decision-making.

2 Learning with imperfect model

'*All analysts approach data with preconceptions. The data never speak for themselves. Sometimes preconceptions are encoded in precise models. Sometimes they*

---

[1] "Anything that gives us *new* knowledge gives us an opportunity to be more rational" – Herbert Simon.

[2] The existence of this article is barely known.

[3] what Keynes (1937) called 'uncertain knowledge.'

[4] Those who ignore experts' knowledge and only trust data are *empirical-fools* Those who ignore data and only trust their gut-instinct are *emotional-fools* (Tversky and Kahneman 1974). Expert decision-makers always use empirically guided intuition by *appropriately* combining both data and available knowledge.

are just intuitions that analysts seek to confirm and solidify. A central question is how to revise these preconceptions in the light of new evidence.'

— Heckman and Singer (2017)

Empirical scientific inquiry typically starts with a simple yet believable model of reality (model-0) and aims to *sharpen* existing knowledge by gathering new observations.

We observe a random sample $X_1, \ldots, X_n \overset{\cdot}{\sim} F_0$. By "$\overset{\cdot}{\sim}$" we mean that F_0 is an 'approximately correct' structured provisional model for X that is given to us by subject-matter experts. We like to extract new knowledge from the data by smartly leveraging existing knowledge[5] that is encoded in the initial approximate model $f_0(x)$.[6]

Creating knowledge-guided statistical models. The core mechanism of our process involves: (i) inspecting whether the structured provisional model-0 is still a good fit in light of fresh data; (ii) if not, then we like to know what's new in the data that cannot be tackled by the current model; and, finally, (iii) repair the current misspecified model in order to cope with the new reality. However, the question remains as to how can we design an inference machine that can offer these successively fine-grained insights? To address this question, we will describe a new statistical model building principle, called the 'density-sharpening principle.'

2.1 A dyadic model

We introduce a dyadic model with two interrelated subsystems that accommodates the decision maker's concern for misspecification of the starting expert-guided model.

Definition 1 (Dyadic model). X be a general (discrete, continuous, or mixed) random variable with true unknown density $f(x)$ and cdf $F(x)$. Let $f_0(x)$ represents a simple approximate model for X with cdf $F_0(x)$, whose support includes the support of $f(x)$.[7] Then the following dyadic density decomposition formula holds:

$$f(x) = f_0(x) \, d\big(F_0(x); F_0, F\big), \tag{2}$$

here $d(u; F_0, F)$ is defined as

$$d(u; F_0, F) = \frac{f(Q_0(u))}{f_0(Q_0(u))}, \quad 0 < u < 1, \tag{3}$$

---

[5] Model amendment principle: the starting model $f_0(x)$ is incomplete but not useless. It contains valuable background knowledge. Rather than throwing this vital information, we want to build a model by smartly taking clues from it. The goal is to amend model-0, not to abandon it completely.

[6] As for notation: by $F_0(x)$, we denote the cumulative distribution function (cdf) of the starting model-0; $f_0(x)$ is the probability density function (pdf) and quantile function is denoted by $Q_0(u)$. The expectation with respect to $f_0(x)$ will be abbreviated as $\mathbb{E}_0(x)$.

[7] For dealing with truly zero-probability events see Coletti and Scozzafava (2002), Ch. 11.

where $Q_0(u) = \inf\{x : F_0(x) \geq u\}$ for $0 < u < 1$ is the quantile function. The function $d(u; F_0, F)$ is called 'comparison density' because it *compares* the initial model-0 $f_0(x)$ with the true $f(x)$ and it integrates to one:

$$\int_0^1 d(u; F_0, F) \, du = \int_x d(F_0(x); F_0, F) \, dF_0(x) = \int_x (f(x)/f_0(x)) \, dF_0(x) = 1.$$

However, we will interpret the d-function as the density-sharpening function (DSF), since it plays the role of "sharpening" the initial model-0 to hedge against its potential misspecification. To simplify the notation, $d(F_0(x); F_0, F)$ of Eq. (2) will be abbreviated as $d_0(x)$.

A few remarks on density-sharpening law:

1. The model building mechanism of Definition 1 provides a statistical process of *transforming and refining* a crude initial model into a useful one for better decision-making.
2. Note that if $d(u; F_0, F) \neq 1$, i.e., if $d(u; F_0, F)$ deviates from uniform distribution then *change* of probability assignment is needed to embrace the current scenario. The density sharpening mechanism of (2) prescribes how to revise the old probability assignments in light of new evidence.
3. Similar to Rao's uncertainty law (1), we can also write down a simple logical equation that captures the essence of the density-sharpening based model building principle (Definition 1):

$$\boxed{\text{Misspecified model-0}} \times \boxed{\text{Knowledge of misspecification}} = \boxed{\text{Upgraded model-1}} \quad (4)$$

Interpretation of the components: the first component is the starting imprecise model $f_0(x)$, coming from expert knowledge. The second component $d_0(x)$ is the quality-assurer of the model that manages the risk of misspecification of the initial $f_0(x)$. $d_0(x)$ *sharpens* the decision-makers initial mental model by extracting knowledge from data that is previously unknown, which justifies its name—density sharpening function (DSF). Finally, the model-0 is "stretched" by $d_0(x)$ following Eq. (2) (only when the ideal scenario is different from the expected one) to incorporate the newly discovered information into the revised model. The class of d-sharp distributions turns the uncertain knowledge-distribution $f_0(x)$ into a usable distribution by properly sharpening using $d_0(x)$. Also, see Supp. A2, where a comparison between the traditional Bayes' rule and the density-sharpening-based multiplicative model update rule is presented.

2.2 Comparison coding

The density-sharpening law provides a mechanism of building a model $f(x)$ for the data $X_1, \ldots, X_n$ by *inheriting* knowledge from the assumed working model $f_0(x)$. To apply the formula (2), we need to estimate $d_0(x)$ from data.[8] And we call this learning

---

[8] To keep the theory of estimation simple, we will mainly focus on the X continuous case. A detailed account for the discrete case can be found in Mukhopadhyay (2021).

process 'comparison coding' because $d_0(x)$ codes how surprising the current situation is in light of the model-0 by contrasting expectations with reality.

Since the density-sharpening function $d_0(x) := d(F_0(x); F_0, F)$ is a function of $F_0(x)$, we can approximate it by a linear combination of polynomials that are function of $F_0(x)$ and orthonormal with respect to the base-model $f_0(x)$. One such orthonormal system is the LP-family of polynomials (Mukhopadhyay and Parzen 2020; Mukhopadhyay 2017, 2021), which can be constructed as follows. For an arbitrary continuous F_0, define the first-order LP-basis function as *standardized* $F_0(x)$:

$$T_1(x; F_0) = \sqrt{12}\{F_0(x) - 1/2\}. \tag{5}$$

Note that $\mathbb{E}_0(T_1(X; F_0)) = 0$ and $\text{Var}_0(T_1(X; F_0)) = 1$. Next, apply Gram–Schmidt procedure on powers of the first-order LP-basis functions $\{T_1^2(x; F_0), T_1^3(x; F_0), \ldots\}$ to construct a higher-order LP orthogonal system $T_j(x; F_0)$:

$$T_2(x; F_0) = \sqrt{5}\{6F_0^2(x) - 6F_0(x) + 1\} \tag{6}$$

$$T_3(x; F_0) = \sqrt{7}\{20F_0^3(x) - 30F_0^2(x) + 12F_0(x) - 1\} \tag{7}$$

$$T_4(x; F_0) = \sqrt{9}\{70F_0^4(x) - 140F_0^3(x) + 90F_0^2(x) - 20F_0(x) + 1\}, \tag{8}$$

and so on. Compute these polynomials by performing the Gram–Schmidt process numerically, which can be done using readily available computer packages like R or python.

Definition 2 (Comparison coding). Expand comparison density in the LP-orthogonal series

$$d_0(x) := d(F_0(x); F_0, F) = 1 + \sum_j \text{LP}[j; F_0, F] \, T_j(x; F_0). \tag{9}$$

To estimate the unknown LP-Fourier coefficient, note that:

$$\begin{aligned}
\text{LP}[j; F_0, F] &= \int T_j(x; F_0) d_0(x) f_0(x) \, \mathrm{d}x \\
&= \int T_j(x; F_0) f(x) \, \mathrm{d}x \\
&= \mathbb{E}_F[T_j(X; F_0)].
\end{aligned} \tag{10}$$

Replacing $\text{LP}[j; F_0, F]$ with its plug-in estimator in (9) we get

$$\widetilde{d}_0(x) = 1 + \sum_j \widetilde{\text{LP}}[j; F_0, F] \, T_j(x; F_0), \tag{11}$$

where

$$\widetilde{\text{LP}}[j; F_0, F] = \mathbb{E}_{\widetilde{F}}[T_j(X; F_0)] = \frac{1}{n} \sum_{i=1}^n T_j(x_i; F_0). \tag{12}$$

Although (11) provides a robust nonparametric comparison-coding procedure, it has one drawback: the estimated $\widetilde{d}$ may be unsmooth due to the presence of a large number of small noisy LP-coefficients. To avoid unnecessary ripples in $\widetilde{d}$, we need to isolate the small number of non-zero LP-coefficients. Our denoising strategy goes as follows (Mukhopadhyay 2023): sort the empirical $\widehat{\text{LP}}[j; F_0, F]$ in descending order based on their absolute value and compute the penalized *ordered* sum of squares. This Ordered PENalization scheme will be referred as OPEN model-selection method:

$$\text{OPEN}(m) = \text{ Sum of squares of top } m \text{ LP coefficients } - \frac{\gamma_n}{n}m. \tag{13}$$

Throughout, we use AIC penalty with $\gamma_n = 2$. Find the m that maximizes the OPEN(m). Store the selected indices j in the set $\mathcal{J}$. The OPEN-smoothed LP-coefficients will be denoted by $\widehat{\text{LP}}_j$. Finally, return the following *smoothed* estimate:

$$\widehat{d}_0(x) = 1 + \sum_{j\in\mathcal{J}} \widehat{\text{LP}}[j; F_0, F]\, T_j(x; F_0). \tag{14}$$

Remark 1 (The scientific value of sparse d). The DSF $d_0(x)$ is the *bridge* between the theoretical world (idealized model) and the empirical world (real observations). A meaningful way to measure the simplicity of a model is the number of "new" statistical parameters that it contains *beyond* the given scientific parameters—that is, the parsimony (number of parameters) of d.[9] A sparse $\widehat{d}$ provides an intelligent and parsimonious way to elaborate the model-0 (not an indiscriminate, brute-force elaboration) to produce a 'sophisticatedly simple' model. Simplicity is vital to make the model usable and interpretable by decision-makers, who like to understand *how* to change the initial model to explain the data.

2.3 A deep dive into model uncertainty

Understanding the deficiency of the current model is an essential part of the process of iterative model building and refinement: Have we overlooked something? Where are our knowledge gaps? This section provides a comprehensive understanding and exploratory tool for representing and assessing potential model misspecifications. Also, see Supp. note A1, which explains the distinctions between parametric and nonparametric model uncertainties.

2.3.1 Graphical exploration of model uncertainty

Example 1 Consider the following scenario: Fig. 1 displays the data that a physicist just collected from an experiment. The blue curve is the physics-informed background distribution $f_0(x)$, which, in this case, is an exponential distribution with $\lambda_0 = 25$, and the red curve is the true unknown probability distribution. The physicist is mainly

---

[9] It selects only a handful of reasonable (rival) hypotheses out of a vast collection of possibilities. By 'reasonable,' we mean hypotheses with a high OPEN(m) score, which balances complexity (number of parameters of d) and accuracy (of explaining the surprising phenomenon).
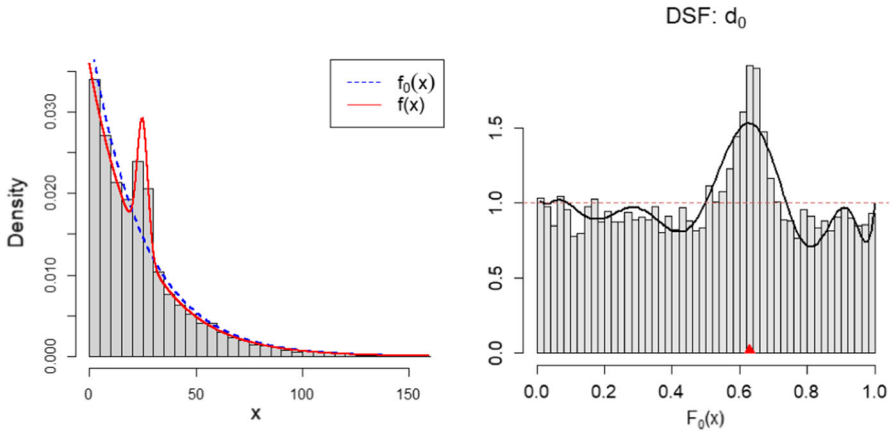
Fig. 1 How closely does reality resemble the model? $10,000$ samples are generated from the true (unknown to the analyst) model $0.9\mathrm{Exp}(\lambda_0) + 0.1\mathcal{N}(25, 2.5^2)$. The right plot shows the graph of $\widehat{d}_0(x)$, which acts as a 'magnifying glass' that forces us to examine what extra information data are willing to reveal *beyond* the known assumed model. The crucial point here is: model uncertainty, as captured by $\widehat{d}_0(x)$, has a *shape*—it is not a single numeric number. It helps physicists narrow down a large set of potential rival theories to a few plausible ones

interested in knowing whether there is any *new physics* hidden in the data, i.e., anything new in the data that was overlooked by existing theory. If so, what is it? How does the theory (f_0) relate to practice? This will help the physicist to come up with some scientific explanations and potential alternative theories.

The Shape of Uncertainty. The researcher ran the density-sharpening algorithm of the previous section with $m = 10$, and the resulting $\widehat{d}_0(x)$ is displayed in the right of Fig. 1 as a function of $F_0(x)$. Few conclusions: (i) **Model appraisal**: The non-uniformity of $\widehat{d}$ tells us that the "shape of the data" is inconsistent with the presumed model-0. (ii) **Model amendment**: The shape of $\widehat{d}$ also informs the scientist about the nature of deficiency of the old model—i.e., what are the most worrisome aspects of the presumed model? In this example, the most consequential unanticipated pattern is the presence of a prominent 'bump' (excess mass) around $F_0^{-1}(0.63) \approx 24.85$, which might be indicative of new physics. This newly discovered pattern can now be used to improve the background exponential model.

Remark 2 (Visual explanatory decision-aiding tool) One of the unique abilities of our exploratory learning is its ability to generate explanations on *why and how* the model-0 is incomplete.[10] Thus, the graph of $\widehat{d}(u; F_0, F)$ explicitly addresses decision-makers model misspecification concerns. It digs into the observations to uncover the "blind spots" of the current model that can ultimately drive discovery (locating novel hypotheses) and better decisions.

---

[10] Explanation-based statistical reasoning is at the core of abductive inference, as discussed later.

2.3.2 Measure of model uncertainty

A general measure of the *degree* of model misspecification is defined using the Csiszár information divergence class.

Definition 3 For $\psi : [0, \infty) \mapsto \mathbb{R}$ a convex function with $\psi(1) = 0$, define the Csiszár class of statistical divergence measure between F and F_0:

$$I_\psi(F, F_0) = \int_{-\infty}^{\infty} \psi\left(\frac{f(x)}{f_0(x)}\right) f_0(x) \, dx \tag{15}$$

We prefer to represent it in terms of density-sharpening function as follows:

$$I_\psi(F, F_0) = \int_{-\infty}^{\infty} \psi \circ d(F_0(x); F_0, F) \, dF_0(x)$$
$$= \int_0^1 \psi \circ d(u; F_0, F) \, du, \quad \text{where } u = F_0(x). \tag{16}$$

One can recover popular divergence measures by appropriately choosing the ψ-function:

- KL-divergence: $\psi(x) = x \log(x)$; $I_{KL}(F, F_0) = \int d \log d$.
- Total variation divergence: $\psi(x) = |x - 1|$; $I_{TV}(F, F_0) = \int |d - 1|$.
- Squared Hellinger distance: $\psi(x) = (1 - \sqrt{x})^2$; $I_H(F, F_0) = \int (1 - \sqrt{d})^2$.
- χ^2-divergence: $\psi(x) = (x - 1)^2$; $I_{\chi^2}(F, F_0) = \int (d - 1)^2 = \int d^2 - 1$.

One can quickly estimate the χ^2-model misspecification index by expressing it in terms of LP-Fourier coefficients (applying Parseval's identity to equation 9):

$$I_{\chi^2}(F, F_0) = \int d^2 - 1 = \sum_{j=1}^{m} \left| LP[j; F_0, F] \right|^2. \tag{17}$$

$I_{\chi^2}(F, F_0)$ quantifies the uncertainty of the preliminary model $f_0(x)$ in light of the given data—i.e., whether $f_0(x)$ is catastrophically wrong or slightly wrong. Estimate it by plugging the empirical LP-coefficients (12) into (17). Since, under $H_0 : F = F_0$, the sample LP-coefficients have the following limiting null distribution (see Theorem 2 of Mukhopadhyay (2017)):

$$\sqrt{n}\widetilde{LP}[j, F_0, F] \xrightarrow{d} \mathcal{N}(0, 1), \quad \text{i.i.d for all } j,$$

$n\widetilde{I}_{\chi^2}(F, F_0)$ follows χ_m^2 under null. One can use this to compute the p-value. Applying this measure to example 1, we get a p-value of practically zero—indicating that the background exponential model is badly damaged and should be *repaired* before making a decision.

Remark 3 It is interesting to contrast our theory with Hansen and Sargent (2022), Sec. 4 and 5.1, keeping in mind that their $m(x)$ is exactly our sharpening function $d_0(x)$. This reinforces our belief that our theory can be applied broadly to econometrics and decision-making under uncertainty.

2.4 *d*-Sharp models

Definition 4 $DS(F_0, m)$ stands for **D**ensity-**S**harpening of $f_0(x)$ using m-term LP-series approximated $d_0(x)$, given by:

$$f(x) = f_0(x)\Big[1 + \sum_{j=1}^{m} LP[j; F_0, F] \, T_j(x; F_0)\Big], \tag{18}$$

obtained by replacing (9) into (2). $DS(F_0, m)$ generates a relevant class of plausible models in the neighbourhood of the postulated $f_0(x)$ that are worthy of consideration.

A few additional points on density-sharpening:

1. The $DS(F_0, m)$-based density-sharpening principle provides a mechanism for *exploring* data by *exploiting* the uncertain background knowledge model. It starts with data and an approximate model $f_0(x)$—and produces a more refined picture of reality following (18).
2. The process of density-sharpening suitably 'stretches' the theory-informed model to create a class of robust empirico-scientific models. Moreover, it shows how new models are born out of data-driven *mutation* of pre-existing ones.
3. The truncation point m indicates the radius of the neighborhood around the elicited $f_0(x)$ to create permissible models. $DS(F_0, m)$ models with higher m entertain alternative models of higher complexity. However, to maintain conceptual appeal and interpretability, it is advisable to focus on the vicinity of f_0 by choosing an m that is not too large. Substituting the smooth estimates $\widehat{LP}[j; F_0, F]$ of Eq. (14) into the formula (18), we get the most economical model (among competing alternatives around f_0) that best explains the empirical surprise.[11]
4. It provides an architecture of an 'intelligent agent' that *simultaneously* possesses the ability to: learn (what's new can we learn from the data), reason (how to explain the surprising empirical findings), and plan (how to self-modify to adapt in the new situations).

Example 2 (Glomerular filtration data) We are given glomerular filtration rates[12] for 211 kidney patients. The experiment was done at Dr. Bryan Myers' Nephrology research lab at Stanford University. The dataset was previously analyzed in Efron and Hastie (2016).

---

[11] It brings our theory close to Gilbert Harman's "Inference to the best explanation" idea; see Harman (1965). This is an area that merits further research.

[12] Glomerular filtration rate (GFR) measures how much blood is filtered through the kidney to remove excess wastes and fluids. Low `gfr` value indicates that the kidneys are not functioning as well as they should.
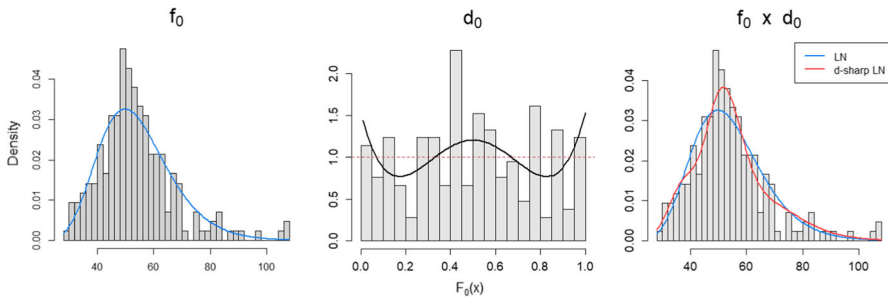
Fig. 2 Glomerular filtration data modeling. Left: the fitted lognormal distribution. Middle: The estimated density-sharpening function d_0 provides an economical description of the empirical surprise, thereby supplying clues for forming new explanatory hypotheses of the data. Right: the d-sharp lognormal with heavier tails and sharper peak

The blue curve on the left plot of Fig. 2 shows the best-fitted lognormal (LN) distribution. We start our analysis by asking whether the parametric LN model needs to be refined to fit the data. The middle panel displays the density-sharpening function, which provides insights into the nature of misspecification of the LN model: the peak and the tails of the initial LN distribution need repairing; LN underestimates the peak and neglects the presence of heavier tails. The repaired LN model (displayed on right-hand side of Fig. 2) is given by

$$\widehat{f}(x) \,=\, f_0(x)\big[1 \,+\, 0.18 T_4(x; F_0)\big], \tag{19}$$

where $f_0(x)$ is $\mathrm{LN}(\mu_0, \sigma_0)$, with $\mu_0 = 4$ and $\sigma_0 = 0.24$. The part in the square bracket comes from $d_0(x)$, which provides recommendations on how to suitably elaborate the LN-model to capture the *unexplained* shape. The point of this example was to show how the density-sharpening principle (DSP) allows an analyst to explicitly perform model formulation, fitting, checking, and repairing—all seamlessly combined into one workflow.

It is interesting to compare our d-sharp LN-model (the red curve) with the seven-parameter exponential family fit shown in Fig. 5.7 of Efron and Hastie (2016). The most noticeable difference lies in the right tail. Efron's seven-parameter exponential family model shows weird spikes on the extreme-right tail. The main reason for this is that it is based on polynomials of raw x: $(x, x^2, \ldots, x^7)$, which are not robust. That is to say, these traditional bases are unbounded and highly sensitive to 'large' data points. In contrast, our LP-polynomials are functions of $F_0(x)$, not raw x, and thus robust by design. The other operational difference between our approach and Efron's exponential family approach is that we model the "gap" between lognormal and the data, which is often far easier to approximate nonparametrically (only required one parameter, see Eq. 19) than modeling the data from scratch.[13]

---

[13] There is an easy way to see that: compare the shapes of the histograms of the left two plots of Fig. 2.
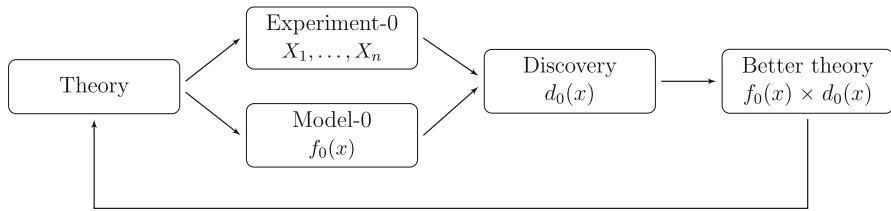
Fig. 3 Architecture of abductive inference machine (AIM). Density-sharpening principle provides a systematic process of inserting the new information into the existing knowledge model to resolve empirical surprise and inconsistency. This continuous cycle of iterative model sharpening is called *abductive learning*, which facilitates the emergence of new theories from data. As George (Box 1980) said, 'The statistician's role is to assist this evolution'

2.5 Modelplasticity and abductive inference machine

Not the smallest advance can be made in knowledge beyond the stage of vacant staring, without making an abduction at every step. — C. S. Peirce (1901)

Modelplasticity—Models ability to modify and adapt itself in response to new data. The density-sharpening principle enables the model to develop *new shapes* in the face of change.

Density-sharpening and model evolution. Modeling is a continual process, not a one-time data-fitting exercise. The density sharpening mechanism allows us to combine new observations with a priory expected model to generate new insights, as depicted in Fig. 3.

Statistical law of model evolution. Density-sharpening supports this dynamic process of recursive model upgrading: $f_k(x) = f_{k-1}(x) d_{k-1}(x)$, for $k = 1, 2, \ldots$, by allowing the model to constantly evolve and reshape itself with fresh sets of data—going from a simple approximate model to a much more mature, accurate model of reality.

Abduction and creation of new knowledge. Abduction is the creative part of an inferential process that aims at producing new theories from data. It builds upon what we know to discover new facts about nature. Abductive learning is concerned with the following questions: What *new* can we learn from the data? How to *change* the prior hypothetical model to explain the current situation? Which alternative classes of models are worthy of being entertained?

'*Does statistics help in the search for an alternative hypothesis? There is no codified statistical methodology for this purpose. Text books on statistics do not discuss either in general terms or through examples how to elicit clues from data to formulate an alternative hypothesis or theory when a given hypothesis is rejected.*'

— C. R. Rao (2001)

Charles Sanders Peirce (1837–1914) was the pioneer of abductive reasoning; see Stigler (1978) and Mukhopadhyay (2023) for more details on the Peircean view of statistical modeling. The goal of Abductive Inference Machine[14] or AIM is to provide

---

[14] They are not traditional pattern recognition (or matching) engine, they are pattern *discovery* engine.

a learning framework that endows a model with this ability to learn, grow and change with new information.

Remark 4 The density-sharpening process plays an essential role for abductive inference, which provides the computational machinery for generating novel hypotheses with explanatory merit and selecting specific ones for further examinations.

Remark 5 (Abductive inference $\neq$ Hypothesis testing) Any scientific inquiry begins with observations and some initial hypotheses. Classical statistical inference develops tools to test the validity of the null model in light of the data. Since all scientific theories are incomplete, accepting or rejecting a particular hypothesis is a pointless exercise. The real question is not whether the null hypothesis is true or false. The real question is: how far is the reality from the postulated model? In which direction(s) should we search to find a better model? Density-sharpening law provides a process of progressive refinement of yesterday's hypothesis.

2.6 Attention mechanism

> *We often neglect how we get rid of the things that are less important...And oftentimes, I think that's a more efficient way of dealing with information.*
> —Duje Tadin[15]

Attention is the prerequisite of gaining new knowledge. Intelligent learners have the ability to quickly *infer* where to focus attention to gain knowledge. In our modeling framework $d_0(x)$ draws analyst's attention quickly and efficiently to the new informative part by suppressing boring details; verify it from the graphs of $d_0(x)$ in Figs. 1 and 2. It acts as a 'gating mechanism' that filters out the new interesting (surprising) aspects of the data, and ignores the dull and unsurprising part—thereby sharpening the model's intelligence by guiding where to pay attention for information processing.

> "*The whole function of the brain is summed up in: error-correction*"
> — W. Ross Ashby, English psychiatrist and a pioneer in cybernetics.

Remark 6 In the brain, a dedicated circuit (or system) performs information-filtering similar to what $d_0(x)$ does for our dyadic model. The existence of such a brain circuit was first hypothesized by Francis (Crick 1984)—he called it 'The Searchlight Hypothesis.' Since then, significant progress has been made to hunt down the brain region, what is now called basal ganglia, that suppresses irrelevant inputs. For more details see Halassa and Kastner (2017) and Gu et al. (2021). Basal ganglia help us focus on what's important and tune out the rest. The mechanics of our model-building mimic the brain's cognitive process that uses existing knowledge to sieve out the new information for correcting the error (sharpening) of the earlier mental model.

---

[15] Jordana Cepelewicz (2019) To Pay Attention, the Brain Uses Filters, Not a Spotlight *Quanta Magazine*, https://www.quantamagazine.org/to-pay-attention-the-brain-uses-filters-not-a-spotlight-20190924.

3 Decision-making with imperfect model

> *How should a decision maker acknowledge model misspecification in a way that guides the use of purposefully simplified models sensibly?*
>
> — Cerreia-Vioglio et al. (2020)

This section demonstrates how practicing abductive inference based on the density-sharpening principle can enable better decision-making in highly uncertain environments.

3.1 Abductive model of decision making

Abduction is the process of generating and revising a model *before* choosing the optimal action. An abducer makes decisions in a dynamic uncertain environment by allowing for potential model misspecification.[16] Abductive decision-making is about knowing *when* to change course and *how* to change it.

How can a decision-maker *abduct*? The mechanics of abductive decision-making consist of three steps: (i) generating a set of plausible alternative models based on new evidence; (ii) constructing a 'robust' model (by choosing the least favorable alternative model or by averaging the alternative models with proper weights); and (iii) selecting an action that maximizes expected utility under the newly revised model. Two modes of abductive decision-making under uncertainty are presented below.

Notation. A decision-maker (DM) has to take an action a from the set of available actions $\mathbb{A} = \{a_1, \ldots, a_q\}$ based on observed outcome $X_1, \ldots, X_n$ from an unknown probability distribution $f(x)$, representing some natural or social phenomenon. The DM selects the optimal action that minimizes expected loss (or risk) under the assumed model-0:

$$\widehat{a}_0 := \mathrm{argmin}_{a \in \mathbb{A}} \int L_a(x) \, dF_0(x), \tag{20}$$

where $f_0(x)$ is the DM's posited probability distribution over outcomes. However, as an abducer, the DM is completely aware that the uncertainty about the outcomes may not be fully captured by a single, rigidly defined probability distribution $f_0(x)$ and thus wants to choose the best decision by accommodating the uncertainty of model-0.

Decision making based on density sharpening principle. To account for the imperfect nature of model-0, the most natural thing to do is to work with an enlarged class of plausible distributions around the vaguely acceptable $f_0(x)$:

$$\Gamma_M = \left\{ f : f \in \mathrm{DS}(F_0, m), \ m \leq M \right\} \tag{21}$$

within a certain reasonable neighbourhood, say $M = 10$. We like to use this enlarged class of distributions Γ_M for robust decision-making. Two such strategies are discussed below.

---

[16] The importance of model uncertainty in economics, finance, and business is beautifully illustrated in Hansen and Sargent (2014), although from a different perspective.

Method 1. A cautious DM selects an action by its expected loss under the least favourable distribution within the set Γ_M:

$$\check{f}_{a,M} \; = \; \arg\sup_{F\in\Gamma_M} \int L_a(x) \, \mathrm{d}F(x). \tag{22}$$

We call this an *abductive-minimax* procedure. Our proposal is partly inspired by the 'local-minimax' idea of Hansen and Sargent (2001a, b).

Method 2. We now describe another robust decision-making procedure that takes into account the uncertainty in the analyst's elicited probability model of future states. Two key concepts are: bootstrap model averaging and action-profile function.

Step 1. We use bootstrap to explore $f \in \Gamma_M$ in an intelligent way. Draw n samples with replacement from the original data. Denote the bootstrap empirical cdf as $\widetilde{F}_*^{(1)}$. Perform density-sharpening algorithm based on $\widetilde{F}_*^{(1)}$, and denote the estimated d-sharp model as $f_*^{(1)}$.

Step 2. Use $f_*^{(1)}$ to select the best action from the given set of q-actions $\{a_1, \ldots, a_q\}$. Denote the selected action as $a_*^{(1)}$.

Step 3. Repeat steps 1–2, B times (say $B = 1000$ times). And return:

- The sample bootstrap *distribution* p_A of optimal actions $\{a_*^{(1)}, \ldots, a_*^{(B)}\}$—which we call the *action profile* of the decision problem.
- Bootstrap systematically generates probable alternative models $\{f_*^{(1)}(x), \ldots, f_*^{(B)}(x)\}$ from the class Γ_M that can explain the data. Compute bootstrap model averaged distribution[17]:

$$\bar{f}(x) \; = \; \frac{1}{B} \sum_{j=1}^{B} f_*^{(j)}(x). \tag{23}$$

By averaging all plausible alternatives, $\bar{f}(x)$ becomes robust to model uncertainty. In this strategy, the policymaker does not have to put his/her complete faith in a single alternative distribution to assign probabilities. Bootstrap density exploration generates and weights different alternatives (from the class Γ_m) appropriately to create a realistic model. Figure 4 shows the bootstrap-generated densities for the gfr data of example 2. The light blue curves are the plausible alternative models, and the dark blue is the averaged density that takes into account all likely scenarios. It's worth contrasting $\bar{f}(x)$ with more traditional parametric uncertainty-based Bayesian predictive density; see supplementary A1.

Remark 7 Two remarks: (i) Rather than assuming that alternative probable models are given to us *a priori*, we use density-sharpening-based abductive inference method to *synthesize* them, allowing us to account for a much broader range of model uncertainty than is possible with conventional approaches like Bayesian model averaging;

---

[17] This is also known as bagging (Breiman 1996) or bootstrap smoothing (Efron 2014).
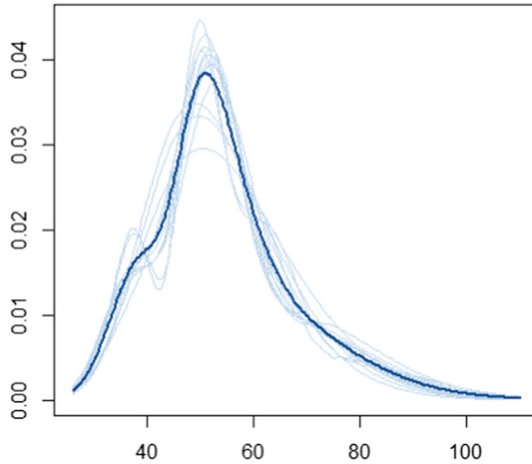
Fig. 4 The light blue curves are the bootstrap generated gfr densities, which try to present a landscape of plausible scenarios as suggested by the data. The dark blue denote the estimated model-averaged distribution $\bar{f}(x)$, which takes into account the uncertainty surrounding the initially postulated model $f_0(x)$ as captured by the class Γ_M

see Supp. note A4. (ii) Furthermore, in our scheme, bootstrap provides a way to *automatically estimate* the posterior probability of accepting each synthesized alternative model (light blue curves in Fig. 4), eliminating the need for arbitrary subjective probabilities over models. To compute the $\bar{f}$, bootstrap-derived model posterior probabilities are used to weigh the various alternatives from the density-set Γ_M; see Supp. note A3.

Step 4. This bootstrap scheme can also be used to approximately compute the least favourable distribution, defined in (22):

$$\check{f}_{a,M} = \operatorname{argmax}_j \int L_a(x) \, \mathrm{d}F_*^{(j)}(x).$$

The decision-maker can use this estimated model to carry out the proposed abductive-minimax procedure (method 1).

Step 5. **Robust procedure**[18]: A pragmatic[19] decision-maker chooses an action (or ranks the actions) that minimizes expected loss (or maximizes the expected utility) with respect to the averaged-distribution: $\widehat{a}_{\text{robust}} := \operatorname{argmin}_{a \in \mathbb{A}} \int L_a(x) \, \mathrm{d}\bar{F}(x)$. Our strategy prescribes action that is robust across a wide range of plausible alternative models. It could be especially powerful for dealing with "deep uncertainty" in making robust policies. For a comprehensive overview on this subject, see Marchau et al. (2019).

Step 6. Quantifying the 'robustness' of the action (or decision rule): How much does the optimal action change when a model is selected from a reasonable

---

[18] Our philosophy of robustness is in complete agreement with Huber (1977), who advocated *distributional* robustness: "one would like to make sure that methods work well not only at the [idealized parametric] model itself, but also in a neighborhood of it."

[19] Pragmatism is the logic of abduction.

neighborhood of the assumed initial opinion, i.e. from the Γ_m class? The shape of the action profile distribution can be used to determine how robust the optimal action is to model perturbation. In particular, the entropy of the action profile distribution can be used to assess the robustness (or stability) of the inference to potential model misspecification:

$$\text{Entropy}[p_A] = -\sum_{i=1}^{q} p_A(i) \log p_A(i) = -\sum_{i=1}^{q} \Pr(A = a_i) \log \Pr(A = a_i). \quad (24)$$

Uniform probability over possible actions yields maximum uncertainty—indicating that the decision is highly non-robust (sensitive) to model misspecification.

3.2 Quantile decision analysis

Until now, we have assumed experts can precisely formulate their opinion in a probabilistic form $f_0(x)$. However, for complex real-world decision-making problems, experts might only have *incomplete* information about the uncertainty distribution of the target variable. A decision-analyst often elicit their partial knowledge about an uncertain quantity as a set of quantile-probability (QP) pairs $\{x_i, F(x_i)\}$, for $i = 1, \ldots, \ell$. The job of an analyst is to find a simple, flexible, and parameterizable density that honors the assessed percentiles.

Model ambiguity due to incomplete information. The task of eliciting an expert's probability distribution from a small set of QP pairs is a vital yet nascent topic in decision analysis; see Powley (2013), Keelin and Powley (2011), Hadlock (2017). In this section, we present an algorithm called Q2D (stands for quantile to distribution) that provides a systematic approach to deduce a reliable expert distribution from ℓ arbitrary QP-specifications.

Probability-gap Approximation. The main theoretical idea behind Q2D algorithm: Recall our $\text{DS}(F_0, m)$ model

$$f(x) = f_0(x)\left[1 + \sum_{j=1}^{m} \text{LP}[j; F_0, F] \, T_j(x; F_0)\right] \quad (25)$$

Integrating from minus infinity to x on both sides, we have

$$\int_{-\infty}^{x} (f(z) - f_0(z)) \, dz = \sum_{j=1}^{m} \text{LP}[j; F_0, F] \int_{-\infty}^{x} S_j(F_0(z)) \, dF_0(z),$$

where $S_j(u) = T_j(Q_0(u); F_0)$ is defined over the unit interval $[0, 1]$ and $Q_0(u)$ is the quantile function of the distribution f_0. This leads to

$$F(x) - F_0(x) = \sum_{j=1}^{m} \text{LP}[j; F_0, F] \int_{0}^{F_0(x)} S_j(u) \, du. \quad (26)$$

Given a set of arbitrary ℓ quantile-probability data $(x_i, F(x_i))$, for $i = 1, \ldots, \ell$, we can rewrite (26) compactly as a matrix equation

$$v = S_0 \boldsymbol{\beta} \qquad (27)$$

where $v_i = F(x_i) - F_0(x_i)$, $\beta_i = \mathrm{LP}_j$, and $S_0 \in \mathcal{R}^{\ell \times m}$, $S_0[i, j] = \int_0^{F_0(x_i)} S_j(u)$. The desired parameters are $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)$, where β_j is shorthand for $\mathrm{LP}[j; F_0, F]$.

For $m \leq \ell$, we can uniquely estimate $\boldsymbol{\beta}$ using the least-square method

$$\widetilde{\boldsymbol{\beta}} = \mathrm{minimize}_{\boldsymbol{\beta}} \, \|v - S_0 \boldsymbol{\beta}\|^2 = (S_0^T S_0)^{-1} S_0^T v. \qquad (28)$$

For large ℓ (say, $\ell \geq 5$), a better, more stable estimate can be found through regularization

$$\widehat{\boldsymbol{\beta}} = \mathrm{minimize}_{\boldsymbol{\beta}} \, \|v - S_0 \boldsymbol{\beta}\|^2 + \lambda \|\boldsymbol{\beta}\|_1 \qquad (29)$$

where $\| \cdot \|_p$ is the ℓ_p norm, and $\lambda > 0$ is the regularization parameter. The lasso (Tibshirani 1996) penalized $\widehat{\boldsymbol{\beta}}$ yields a sparse estimate and counters over-fitting. This penalized estimate provides a tradeoff between accuracy and interpretability.[20] Finally, plug the estimated LP-Fourier coefficients β_j into the primary equation (18) to get the expert distribution.

Remark 8 The expert quantile specifications should not be viewed as a 'gold standard'—they are nothing but a preliminary guess (prone to errors of judgment or hindsight bias) whose purpose is to steer the analyst in the right direction.[21] For that reason, we recommend the smoothed (denoised) regularized $\widehat{\boldsymbol{\beta}}$ over the naive $\widetilde{\boldsymbol{\beta}}$, since it makes little sense to find an exact fit to the noisy QP-data.

Example 3 (Bimodal Distribution) We are given the following quantile judgments:

Quantile: x_i	-3.40	-2.53	-1.20	0	2.0	2.83	3.60
Probability: $F(x_i)$	0.04	0.15	0.39	0.50	0.75	0.90	0.97

In our Q2D algorithm, we choose F_0 (an initial approximate shape) to be normal distribution. To estimate the parameters μ_0 and σ_0 of the normal distribution, note that the quantile function $Q(u) \approx \mu_0 + \sigma_0 \Phi^{-1}(u)$. Thus one can quickly get a rough estimate by simply performing a linear regression[22] on $(\Phi^{-1}(u_i), Q(u_i))$; see Fig. 5. The estimated normal distribution is shown in the right panel, along with the Q2D-estimated density.

---

[20] Note that due to regularization, Eq. (29) can even tackle cases with $m > \ell$. In such scenarios, the OLS is ill-posed, with an infinite number of solutions.

[21] Winkler (1967) emphasized that the expert does not have some 'true' density function waiting to be elicited, only a 'satisficing' initial distribution that the policymaker is 'content to live with at a particular moment of time.'

[22] This technique will work for any location-scale family $f_0(x)$, e.g. normal, Laplace, logistic, etc.
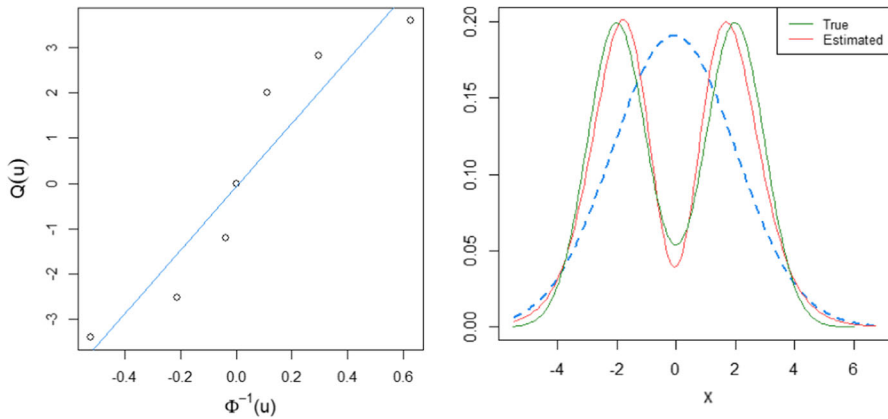
Fig. 5 Learning from incomplete information. Left panel: Regression on the scatter $(\Phi^{-1}(u_i), Q(u_i))$, $i = 1, \ldots, 7$. Right panel: The blue dotted line is the estimated f_0: normal distribution. Q2D estimated curve is shown in red, which is impressively close to the true mixture density $\frac{1}{2}\mathcal{N}(-2, 1) + \frac{1}{2}\mathcal{N}(2, 1)$

Example 4 (U.S. Navy data) Fig. 6 shows a histogram of 122 repair times (in hours) for a component of a U.S. Navy weapons system. The dataset was analyzed in Law (2011). Imagine that for privacy and other reasons, we do not have access to the full data. The goal is to infer a probability distribution that faithfully represents the following quantiles:

Quantile: x_i	0.12	1.30	3.00	7.00	26.17
Probability: $F(x_i)$	0.01	0.20	0.50	0.80	0.99

We start with exponential distribution as our initial guess, which is often taken as a 'default' distribution (model-0) in reliability analysis. For $X \sim \text{Exp}(\lambda)$, we have

$$\text{Median}(X) = \lambda \ln(2), \quad \text{where } \lambda = \mathbb{E}(X).$$

From the quantile table we get $\hat{\lambda} = 3/\ln(2) = 4.32$. Next, we apply the Q2D algorithm to derive the LP-parameters with $f_0 = \text{Exp}(4.32)$. The resulting density sharpening function and the final d-sharp exponential are shown in Fig. 6. The red curve on the right plot shows an excellent fit to the data, which was derived by the Q2D algorithm simply by utilizing the five quantile-probability pairs.

3.3 Decision-making based on multiple experts

High-stakes decision-making (say, COVID-19 pandemic or climate change) is often based on multiple experts' opinions instead of putting all bets on a single rigidly defined probability model. The challenge is to aid data-driven decision-making by appropriately combining several experts' models. We describe one possible way to
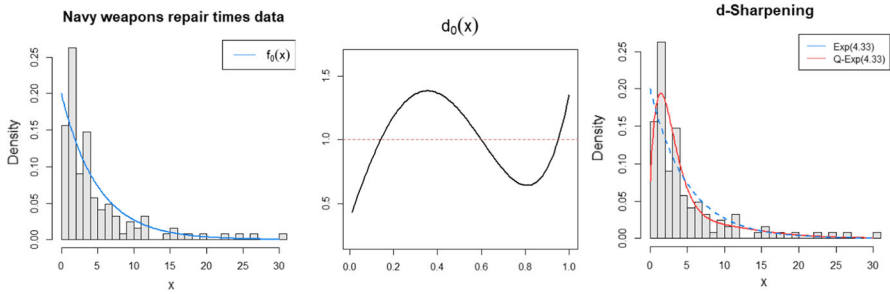
Fig. 6 Left: histogram of 122 repair times for a component of a U.S. Navy weapons system. The blue curve is the $f_0 = \text{Exp}(4.32)$. Our analysis only used five QP-data (not the full data), whose outputs are shown in the middle and right-hand panels. The inferred density-sharpening function tells that the peak and the tail of the exponential model need correction. The repaired exponential model is displayed in red

build a '*consensus committee model*' that can be used as a possible model-0 within an abductive decision-making framework.

Learning from multiple expert distributions. Given k competing probability models $\{f_{01}, \ldots, f_{0k}\}$, which may differ markedly in shape, define the following model-weights:

$$\text{Relevance weight:} \quad w_\ell = \frac{1}{1 + \sum_j |\text{LP}_{j|\ell}|^2}, \quad \text{for } \ell = 1, \ldots, k \tag{30}$$

where $\text{LP}_{j|\ell}$ is the LP-Fourier coefficients of the ℓth model:

$$d_{0\ell}(x) := d(F_{0\ell}(x); F_{0\ell}, \widetilde{F}) = 1 + \sum_j \text{LP}_{j|\ell} \, T_j(x; F_{0\ell}). \tag{31}$$

Note that the *relevance weight* for the ℓth model is always $0 < w_\ell \leq 1$, and

$$w_\ell = 1 \quad \text{if and only if} \quad \text{LP}_{j|\ell} = 0, \ \forall \, j.$$

$\text{LP}_{j|\ell} = 0$ for all j when $f_{0\ell}$ fully explain the data and there is no need to sharpen it further (i.e., $d_{0\ell} = 1$). In that sense, w_ℓ's are data-driven weights (which will keep changing as we get more and more fresh data), computed based on the degree of agreement between the observed data and expert model $f_{0\ell}$. Define mixture expert distribution as

$$f_{\text{mix}}^0(x) = \sum_{\ell=1}^k \pi_\ell \, f_{0\ell}(x), \tag{32}$$

where $\pi_\ell = w_\ell / \sum_\ell w_\ell$. This model serves two purposes: it tries to resolve conflicting opinions based on data and at the same time encourages one to include as much diverse information as possible.

Additional layer of model uncertainty. If the analyst believes that the correct model might not be among the collection of models being considered, then use the combined expert model $f_{\mathrm{mix}}^0(x)$ as a model-0 in the subsequent density-sharpening-based learning and abductive decision-making process.

4 Model management science

How should an analyst use imperfect models to learn from data?[23] What should be the output of such an analysis that can ultimately aid informed decision-making? We address these questions by introducing a general inferential framework for statistical learning and decision-making under uncertainty—which builds on two core ideas: abductive thinking and density-sharpening principle. Some of the defining features of our approach for data analysis, scientific discovery, and decision-making are highlighted below:

- *Data analysis and science of model management*: No model is perfect, irrespective of how cunningly it is designed. The central problem of statistical model developmental process is to understand how a relatively simple model can evolve into a more complex and mature one in the presence of a new data environment. The principle of density-sharpening assists this model evolution process (thereby helping empirical scientists to abduct): by abductively generating explanations on *why* the presumed model-0 is unfit for the data [playing the role of a quality inspector] and also providing recommendations on *how* to fix the misspecification issues [serving as a policy adviser] in order to make better decisions in new circumstances.

- *Discovery and creation of new knowledge*: Abductive data analysts are less interested in *testing* a particular working model. They are mainly interested in conceptual innovation: discovering new pursuit worthy hypotheses based on surprising empirical evidence.[24] The density-sharpening function $d(u; F_0, F)$ picks out 'what's new' in the data beyond the current scientific knowledge encoded in $f_0(x)$, thereby helping the scientist to uncover *new unexpected* knowledge from the data using graphical tools. The density-sharpening principle (DSP) provides a learning mechanism that isolates the 'known' from the 'unknown' and allows us to focus on the newfound pattern in the data, which is the basis for knowledge-creation.[25]

- *Abductive inference and decision-making*: The proposed theory of abductive decision-making tackles model uncertainty induced by imprecise, ambiguous, and incomplete knowledge about the underlying probabilistic structure. An abductive-decision support system automatically discovers and explicitly articulates the possible alternatives to the analysts, which forces them to rethink their choices before taking impulsive action; see Supp. note A5. This style of empirical reasoning and adaptive

---

[23] The challenge of learning from uncertain knowledge is also a fundamental issue in the development of intelligent systems.

[24] A largely unexplored topic relative to the vast literature on hypothesis testing. As noted by George E. P. Box (2001): "Much of what we have been doing is adequate for testing but not adequate for discovery."

[25] Curious readers are invited to read the paper "Nobel Turing Challenge: creating the engine for scientific discovery" by Kitano (2021), where he argued that the single-most-important mission of AI is to accelerate scientific discovery.

decision-making could be especially valuable in situations where strategic planners need to take quick action in the face of uncertainty, equipped with approximate subject-matter knowledge.

Supplementary material

For more discussion on how our abductive statistical approach compares to the more traditional Bayesian statistical approach for model misspecification, robustness, and decision-making, see the Supplementary section.

Supplementary Information   The online version contains supplementary material available at https://doi.org/10.1007/s10203-023-00390-5.

Code and data availability   All the datasets and R-code written for the analysis are available upon request to the author.

Declarations

Conflict of interest   The Author declares that there is no conflict of interests.

References

Box, G.: Statistics for discovery. J. Appl. Stat. **28**(3–4), 285–299 (2001)

Box, G. E.: Robustness in the strategy of scientific model building. In: Robustness in Statistics, pp. 201–236. Elsevier (1979)

Box, G.E.: Sampling and Bayes' inference in scientific modelling and robustness. J. R. Stat. Soc.: Ser. A (General) **143**(4), 383–404 (1980)

Breiman, L.: Bagging predictors. Mach. Learn. **24**(2), 123–140 (1996)

Cerreia-Vioglio, S., Hansen, L.P.,. Maccheroni, F., Marinacci, M.: Making decisions under model misspecification. University of Chicago, Becker Friedman Institute for Economics Working Paper (2020-103) (2020)

Coletti, G., Scozzafava, R.: Probabilistic logic in a coherent setting, Volume 15. Springer (2002)

Crick, F.: Function of the thalamic reticular complex: the searchlight hypothesis. Proc. Natl. Acad. Sci. **81**(14), 4586–4590 (1984)

Efron, B.: Estimation and accuracy after model selection. J. Am. Stat. Assoc. **109**(507), 991–1007 (2014)

Efron, B., Hastie, T.: Computer Age Statistical Inference, vol. 5. Cambridge University Press, Cambridge (2016)

Gu, Q.L., Lam, N.H., Wimmer, R.D., Halassa, M.M., Murray, J.D.: Computational circuit mechanisms underlying thalamic control of attention. bioRxiv 2020–09 (2021)

Hadlock, C.C.: Quantile-parameterized methods for quantifying uncertainty in decision analysis. Ph.D. thesis (2017)

Halassa, M.M., Kastner, S.: Thalamic functions in distributed cognitive control. Nat. Neurosci. **20**(12), 1669–1679 (2017)

Hansen, L.P., Sargent, T.J.: Acknowledging misspecification in macroeconomic theory. Rev. Econ. Dyn. **4**(3), 519–535 (2001)

Hansen, L.P., Sargent, T.J.: Robust control and model uncertainty. Am. Econ. Rev. **91**(2), 60–66 (2001)

Hansen, L.P., Sargent, T.J.: Uncertainty within Economic Models, vol. 6. World Scientific, Singapore (2014)

Hansen, L.P., Sargent, T.J.: Risk, ambiguity, and misspecification: Decision theory, robust control, and statistics. University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2022-157 (November 28, 2022), SSRN: 4287610 (2022)

Harman, G.H.: The inference to the best explanation. Philos. Rev. **74**(1), 88–95 (1965)

Heckman, J.J., Singer, B.: Abducting economics. Am. Econ. Rev. **107**(5), 298–302 (2017)

Huber, P.J.: Robust Statistical Procedures. SIAM, Philadelphia (1977)

Keelin, T.W., Powley, B.W.: Quantile-parameterized distributions. Decis. Anal. **8**(3), 206–219 (2011)

Keynes, J.M.: The general theory of employment. Q. J. Econ. **51**(2), 209–223 (1937)

Kitano, H.: Nobel turing challenge: creating the engine for scientific discovery. NPJ Syst. Biol. Appl. **7**(1), 1–12 (2021)

Law, A.M.: How to select simulation input probability distributions. In: Proceedings of the 2011 Winter Simulation Conference (WSC), pp. 1389–1402 (2011) IEEE

Marchau, V.A., Walker, W.E., Bloemen, P.J., Popper, S.W.: Decision Making Under Deep Uncertainty: From Theory to Practice. Springer Nature, Cham (2019)

Mukhopadhyay, S.: Large-scale mode identification and data-driven sciences. Electron. J. Stat. **11**(1), 215–240 (2017)

Mukhopadhyay, S. . Density sharpening: Principles and applications to discrete data analysis, pp. 1–51. Technical Report (2021). arXiv:2108.07372

Mukhopadhyay, S.: Abductive inference and C. S. Peirce: 150 years later. J. Quant. Econ., 1–27 (2023) (**in press**)

Mukhopadhyay, S., Parzen, E.: Nonparametric universal copula modeling. Appl. Stochastic Models Bus. Ind. Spec. Issue "Data Science" **36**(1), 77–94 (2020)

Peirce, C.S.: The proper treatment of hypotheses: A preliminary chapter, toward an examination of Hume's argument against miracles, in its logic and in its history. MS **692**, 890–904 (1901)

Powley, B.W.: Quantile function methods for decision analysis. Ph. D. thesis (2013)

Rao, C.R.: Uncertainty, statistics, and creation of new knowledge. Chance **9**(4), 5–11 (1996)

Rao, C.R.: Statistics: reflections on the past and visions for the future. Commun. Stat.-Theory Methods **30**(11), 2235–2257 (2001)

Stigler, S.M.: Mathematical statistics in the early states. Ann. Stat., 239–265 (1978)

Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc.: Ser. B (Methodol.) **58**(1), 267–288 (1996)

Tversky, A., Kahneman, D.: Judgment under uncertainty: Heuristics and biases. Science **185**(4157), 1124–1131 (1974)

Winkler, R.L.: The quantification of judgment: some methodological suggestions. J. Am. Stat. Assoc. **62**(320), 1105–1120 (1967)