A. SUPPLEMENTARY NOTES

This supplementary section contains some additional notes on the connections and differences between the Bayesian statistical approach vs. the Abductive statistical approach to model misspecification, robustness, and decision-making.

A1. The two types of model-uncertainty: Parametric and nonparametric

It is important to distinguish two main types of model uncertainty: *parametric* uncertainty and more general *nonparametric shape* uncertainty.

1) Parametric uncertainty is a classical scenario in which the model structure is assumed to be known but not the relevant parameter values. In this setup, it is implicitly assumed that the decision-maker is aware of the correct parameterized statistical models $f_0(\cdot;\theta)$, which is misspecified in terms of only θ.[26] We can therefore think of it as a *finite-dimensional* (statistical search) problem, for which we have a number of legacy theories, including Bayesian inference.

2) Under nonparametric shape uncertainty, we work with much deeper or more severe uncertainties about the shape of the data-generating model. It is a far more challenging *infinite-dimensional* (statistical search) problem for which there is no established general theory. This paper addresses this concern by offering a 'general theory of nonparametric model revision' whose foundation stands on two pillars: the density-sharpening principle and abductive inference. In our theoretical framework, we only have access to a probability model $f_0(x)$ that approximately encodes the decision-maker's beliefs about the distribution of the observations. Our theory then provides a systematic method for searching a useful class of alternative models $\mathrm{DS}(F_0, m)$ by 'correcting' (or sharpening) the hypothesized model f_0 in an automated data-driven manner.

[26] As a result, decision theory based on parametric uncertainty is predicated on a highly restrictive model uncertainty assumptions.

A2. Bayesian statistical approach vs. Abductive statistical approach

1) Bayesian inference is extremely effective for dealing with parametric model uncertainty. Bayes' law provides a principled method for updating beliefs about a model's parameters.

Bayes' law. Given observed data $x = (x_1, \ldots, x_n)$ and the parametrized likelihood function $f_0(x;\theta)$, Bayes' multiplicative rule updates belief about θ from the prior to posterior as follows

$$\widetilde{\pi}(\theta|x) \propto \pi(\theta) \prod_{i=1}^{n} \{f_0(x_i;\theta)\}. \tag{A.1}$$

For more details on standard methods for Bayesian parametric modeling, see Box (1980).[27]

Bayes decision rule. Optimal Bayes action is taken by minimizing the expected loss under the posterior:

$$\hat{a}_{\text{Bayes}} \;:=\; \text{argmin}_{a \in \mathbb{A}} \int_{\Theta} L_a(\theta)\, \tilde{\pi}(\theta|x)\, \mathrm{d}\theta.$$

For more detailed treatment refer to the standard textbooks like Berger (2013, Chapter 4).

2) Abductive inference, on the other hand, is a powerful mode of statistical reasoning for nonparametric model uncertainty problems. In particular, the proposed density-sharpening law provides a systematic rule for updating the *shape* of a probability density model.

The abductive decision analysts do not live in a fantasy world where decision-makers pretend to know the ideal parametrized model for the data. A new class of abductive inference-based decision-theoretic models called "dyadic models" are introduced in this paper (see Sec. 2), which allow the analyst to automatically generate a class of probable alternative models from data *without* imposing any prior structural constraints.

A3. Bayes' model synthesis process

The goal of the "model synthesis problem" is to answer the following question:

Given a set of observations, how to systematically go about searching for a model superior to the one the decision-maker initially guessed?

[27] For nonparametric Bayesian modeling see Ghosal and Van der Vaart (2017).

Bayes model synthesis process (contrast this with the abductive model synthesis process given in Sec. 3.1, method 2) takes into account the uncertainty of θ as follows:

1. Simulate $\theta_1, \ldots, \theta_B$ from the posterior distribution $\widetilde{\pi}(\theta|x)$, for some large B, say 1000.

2. Generate a set of plausible parametric models for the data $\{f(x|\theta_j)\}_{1\leqslant j\leqslant B}$.

3. Averaging over the posterior distribution: Compute the averaged density that accounts for the uncertainty of θ (compare this with density-sharpening based bootstrap averaging method, Eq. 23)

$$\bar{f}_\theta(x) = B^{-1}\sum_{j=1}^{B} f(x|\theta_j). \tag{A.2}$$

which is an approximation to the posterior predictive density

$$\bar{f}_\theta(x) \approx \int_\Theta f(x|\theta)\,\pi(\theta|x)\,\mathrm{d}\theta \tag{A.3}$$

The traditional frequentist point-estimate based $\hat{f} := f(x;\hat{\theta})$ underestimates the uncertainty inherent in θ, and as a result, it is much 'narrower' than the Bayes $\bar{f}_\theta(x)$. By averaging over the posterior distribution, $\bar{f}_\theta(x)$ restores the uncertainty lost when only a single $\hat{\theta}$ is used.

Remark 9. Also see Remark 7, where bootstrap is used as the poor man's Bayes posterior probability for each alternative model synthesized from the class $\mathrm{DS}(F_0, m)$.

A4. Awareness of Unawareness

In our abductive model synthesis process, as described in Sections 2 and 3.1, the crucial component is the dyadic model, which is founded on the density sharpening principle. One way to conceptualize dyadic models is as computational agents that are aware of their own *unawareness*. Using this model, decision-makers gain new previously unknown information that had been lurking in the shadows. As the analyst becomes aware of new facts, the belief is nonparametrically updated through the sharpening function $\widehat{d}_0(x)$. In other words, the sharpening function alerts decision-makers to their potential ignorance.

A5. Significance of abductive inference for decision making.

1) *Adaptability.* Reality always carries an element of surprise. To make effective decisions in a dynamic uncertain environment, it is critical to ensure that the decision model can withstand surprise; otherwise, it is unfit for use in the real world. The real advantage of using density-sharpening-based dyadic models is that they can recuperate from surprises through automated structural correction. As a result, an abductive decision rule based on dyadic models can adapt to surprises in the sense that if the true model deviates from the assumed one then still that decision works. This is achieved by averaging over the plausible alternative situations suggested by data, as described in section 3.

2) *Explainability.* Another advantage of abductive inference is that it provides an interpretable and transparent explanation of why and how the real world differs from decision-makers' initial belief about the model, which is of utmost importance when advising on decisions to policymakers.

A6. Bayes, Smooth Bayes, Sharp Bayes, and Robust Bayes

As an educated guess at the data-generation process, a decision analyst handpicks a class of parametric models $\{f_0(x;\theta) : \theta \in \Theta\}$ with quantile function $Q_0(u;\theta)$ and cdf $F_0(x;\theta)$.

Definition 5 (Parametrized sharpening kernel)**.** Define the sharpening kernel between the true generating process $f(x)$ and the assumed parametrized $f_0(x;\theta)$ as

$$d_\theta \;:=\; d_\theta(u; F, F_0(\cdot;\theta)) \;=\; \frac{f(Q_0(u;\theta))}{f_0(Q_0(u;\theta);\theta)}, \quad 0 < u < 1 \tag{A.4}$$

The corresponding sample estimate is given by

$$\widetilde{d}_\theta \;:=\; d_\theta(u; \widetilde{F}, F_0(\cdot;\theta)) = \frac{\widetilde{f}(Q_0(u;\theta))}{f_0(Q_0(u;\theta);\theta)}, \quad 0 < u < 1 \tag{A.5}$$

where $\widetilde{d}_\theta : \Theta \times [0,1] \to [0,\infty)$, which connects information in data to parameters of interest. With this definition in hand, we now present an important result.

Alternative representation of Bayes Rule. We express the likelihood-based standard Bayesian posterior update rule for θ as follows

$$\widetilde{\pi}(\theta|x) \propto \pi(\theta) \exp\big\{ -\int \widetilde{d}_\theta \log \widetilde{d}_\theta \big\}. \tag{A.6}$$

This is equivalent to Eq. (A.1) becuase of the following fact

$$\begin{aligned} -\int \widetilde{d}_\theta \log \widetilde{d}_\theta \;&=\; \int \log\{f_0(x_i;\theta)\}\, \mathrm{d}\widetilde{F} \;+\; \text{constant} && \text{(A.7)} \\ &\propto\; \sum_{i=1}^{n} \log\{f_0(x_i;\theta)\}. && \text{(A.8)} \end{aligned}$$

The Bayes rule is reformulated in terms of density sharpening kernel $\widetilde{d}_\theta$ because it offers a coherent and principled path for generalizing the belief update rule in situations where the probability model (likelihood function) is misspecified.

Smooth Bayes. Before delving into the Bayesian update rule under model misspecification, we describe "smooth" Bayes—an intriguing refinement of traditional Bayes. Substitute the noisy empirical $\widetilde{d}_\theta$ with the smoothed $\widehat{d}_\theta$ (following the method of Sec. 2.2) into Eq. (A.6) to get a smoothed version of the Bayes update rule:

$$\widehat{\pi}(\theta|x) \propto \pi(\theta) \exp\big\{ -\int \widehat{d}_\theta \log \widehat{d}_\theta \big\}, \tag{A.9}$$

Key assumption. Using the Savage axioms, the Bayesian update can be shown to be the rational way to make a decision when the guessed parametric family $\{f_0(x;\theta) : \theta \in \Theta\}$ *contains* the true data model $f(x)$. However, this is a very stringent requirement that is difficult to meet in practice. We prefer to operate under more realistic conditions, which allows for $f_0(x;\theta)$ to be misspecified.

Sharp Bayes. What if the analysts' a priori chosen family $\{f_0(\cdot;\theta) : \theta \in \Theta\}$ *does not* contain the actual data generating model $f(x)$? It is well known that Bayes' update exhibits undesirable characteristics under model misspecification.

Definition 6 (Generalized d-posteriors)**.** Define the following divergence-based generalized posterior density function

$$\pi(\theta|x) \propto \pi(\theta) \exp\big\{ - I_\psi(F, F_0(\cdot;\theta)) \big\} \tag{A.10}$$

where $I_\psi(F, F_0(\cdot;\theta))$ is the Csiszár class of divergence measure between the true data generator f and the assumed misspecified class $f_0(\cdot;\theta)$. We refer to (A.10) as d-posterior, because we can rewrite it using Eq. (16) as

$$\pi(\theta|x) \propto \pi(\theta) \exp\big\{ - \int \psi \circ d_\theta \big\}. \tag{A.11}$$

whose sample estimate is given by

$$\widetilde{\pi}(\theta|x) \propto \pi(\theta) \exp\big\{ - \int \psi \circ \widetilde{d}_\theta \big\}. \tag{A.12}$$

In a remarkable result, Bissiri et al. (2016)[28] showed that (A.12) provides a valid coherent rule for revising prior beliefs about the parameters of a model that is misspecified. Sharp-Bayes is the name given to this density-sharpening-based generalized Bayes update rule.

Remark 10 (The key idea)**.** The information in the observed data $x_1, \ldots, x_n$ is connected with the parameter of interest θ via functionals of the sharpening function $\widetilde{d}_\theta$, instead of the conventional likelihood function, whose precise probability form is never known in practice.

Robust Bayes. For outlier-resistant robust Bayesian analysis choose total variation divergence, a special case of Csiszár class with $\psi(x) = |x-1|$ in (A.12)

$$\widetilde{\pi}(\theta|x) \propto \pi(\theta) \exp\big\{ - \textstyle\int |\widetilde{d}_\theta - 1| \big\}. \tag{A.13}$$

Another particularly useful class of measures for robust Bayesian analysis is Rényi α-divergence,

[28]After some algebraic manipulation, it is not difficult to show that the main result of Bissiri et al. is equivalent to (A.12).

defined as

$$R_\alpha(F, F_0) \;=\; \frac{1}{\alpha(1-\alpha)}\Big(1 - \textstyle\int (f/f_0)^\alpha \, \mathrm{d}F_0\Big), \quad \alpha \in \mathcal{R}\backslash\{0,1\}. \tag{A.14}$$

It is a robust discrepancy measure between f and the imperfect f_0 whose nonparametric estimation can be done by expressing it as a functional of the sharpening function:

$$R_\alpha(\widetilde{F}, F_0(\cdot;\theta)) \;=\; \frac{1}{\alpha(1-\alpha)}\Big(1 - \textstyle\int \widetilde{d}_\theta^{\,\alpha}\Big), \quad \alpha \in \mathcal{R}\backslash\{0,1\}.$$

The following is the associated posterior belief update rule:

$$\widetilde{\pi}(\theta|x) \propto \pi(\theta)\exp\Big\{ -\frac{1}{\alpha(1-\alpha)}(1 - \textstyle\int \widetilde{d}_\theta^{\,\alpha})\Big\} \tag{A.15}$$

The value of $\alpha \in [0.50, 0.75]$ is commonly used to provide good robustness protection against outliers without losing too much efficiency.

Two major conclusions:

(1) Knowing the 'gap' between the sample distribution and the true data generator (as captured by $\widetilde{d}_\theta$) is sufficient to produce posterior beliefs, obviating the need to know the exact probabilistic form of the true likelihood function, which decision-makers almost never know in real-world scenarios.

(2) The fundamental object of statistical inference is not the guessed misspecified parametric model $f_0(\cdot;\theta)$ nor the unknown f, but the 'gap' between them, d_θ.

A7. Addressing Prior misspecification

The information-theoretic generalized Bayes rule, presented in the previous note, is still not fully satisfactory because it is rigidly based on assumed subjective prior $\pi(\theta)$.[29] Thus it is critical to investigate the robustness of statistical decisions in a reasonable neighborhood around the presumed prior, which can be operationalized through the density sharpening principle;

[29] For a more detailed account of "subjective" probability theory see the classic book by De Finetti (1975) and also Lad (1996).

see, for example, Mukhopadhyay and Fletcher (2018). Instead of making critical decisions based solely on analysts' vague subjective specifications, this allows for prior misspecification.

Notes A6 and A7 showcase how concept density-sharpening principles can unify both the classical and the most advanced versions of Bayesian inference using common terminology and notation – a novel contribution in and of itself.

REFERENCES

De Finetti, B. (1974–1975). *Theory of probability.* John Wiley, New York.

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis.* Springer Science & Business Media.

Bissiri, P. G., C. C. Holmes, and S. G. Walker (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*(5), 1103–1130.

Ghosal, S. and A. Van der Vaart (2017). *Fundamentals of nonparametric Bayesian inference*, Volume 44. Cambridge University Press.

Lad, F. (1996). *Operational subjective statistical methods: A mathematical, philosophical, and historical introduction.* Wiley, New York.

Mukhopadhyay, S. and D. Fletcher (2018). Generalized empirical Bayes modeling via frequentist goodness-of-fit. *Nature: Scientific Reports 8*(9983), 1–15.